Probabilistic risk assessment of excavation performance in tunnel projects using Bayesian networks: a case study

O. Špačková  
*Czech Technical University in Prague, Czech Republic*

D. Straub  
*Engineering Risk Analysis Group, TU Munich, Germany*

**ABSTRACT:** A model for probabilistic assessment of excavation performance of tunnel projects is presented. The model is based on Dynamic Bayesian Networks (DBN) and enables to consider the quality of the design and construction process. It is applied on a case study, the excavation of a road tunnel by means of the New Austrian Tunnelling Method. The influence of main model parameters and assumptions (e.g. quality, distribution of unit time) is assessed through a sensitivity analysis.

*Keywords: Tunnel, Risk, Excavation, Bayesian networks.*

1 INTRODUCTION

Tunnel projects are prone to escalations of construction costs and duration. On average, final construction costs of tunnel and bridge construction projects are 34% above original estimates, and there has been no improvement over the past seventy years (Flyvbjerg et al. 2004). New techniques, which would improve the accuracy of these estimates and which would enable a systematic learning from past projects, are therefore needed.

Construction costs and duration are usually estimated by means of expert judgements. While these are irreplaceable, they should be underpinned by objective models, which enable a better quantification of the uncertainties associated with these predictions. Existing probabilistic models for tunnel projects are mostly based on Monte Carlo (MC) simulation of the construction process (Min et al. 2008, Chung et al. 2006, Ruwanpura & Ariaratnam 2007). Other models use Bayesian networks (Sousa 2010), artificial neural networks (Benardos & Kaliampakos 2004) or analytical solutions (Isaksson & Stille 2005). The models are mostly able to describe in detail the uncertainties in the prediction of geotechnical conditions and common variations of performance rates or unit costs, but in general they fail to consider the impact of other factors. These include extraordinary events (e.g. cave-in collapse, fires, flooding) as well as human and organizational factors. In particular the latter lead to a significant increase in the uncertainty of the final project cost and duration and should be included in a realistic model.

The proposed probabilistic model of tunnel projects aims to overcome the above-mentioned gaps. It utilizes dynamic Bayesian networks (DBNs) to model the process of tunnel construction, particularly the time needed to execute the excavation with regard to uncertain geotechnical conditions, varying unit time and quality of design and construction. In contrast to the DBNs models presented in Sousa (2010), the proposed model includes the full probability distributions of random variables such as unit time and cumulative (total) time even if it discretizes them.

The suggested model is applied to a case study which was taken from Min (2003). In the original work, the Decision Aids for Tunnelling (DAT) tool based on MC simulation was used for probabilistic assessment of construction time and costs. The DAT model has been developed since the 80s, and has been applied to a number of projects. In the case study described in this paper, the same assumptions as in the original model are first utilized in order to verify the results of the DBN model. In the second step, additional aspects (such as the influence of the human factor, adjustments of the unit time distributions, variable length of excavation cycles) are included in the DBN model and their effect on the final estimate is described within a sensitivity analysis.
Bayesian networks (BNs) are directed acyclic graphical models that represent the joint probability distribution of a set of random variables. The nodes of the BN are random variables, while the directed links between them characterize their dependencies. Because of their graphical nature, BNs can be highly efficient for modeling and communicating complex models involving large numbers of random variables. BNs have recently found a number of applications in engineering (Faber et al. 2002, Friis-Hansen 2004, Grêt-Regamey & Straub 2006, Langseth & Portinale 2007, Straub 2009). Detailed introductions to BN can be found in (Jensen & Nielsen 2007).

An example of a BN is depicted in Figure 1a. Here, the random variables are geology (G), construction time (T) and cost (C), and the causal dependence between them is represented by the links in the BN. In BN terminology, G is called a parent of T and C, whereas C is called a child of G and T. Each node is described by its probability distribution conditional on its parents. As an example, the distribution of construction time T is described conditional on the geology G; this conditional distribution is denoted as \( p(T|G) \). Applying the chain rule, the joint probability distribution of this BN is obtained as

\[
p(G, T, C) = p(G)p(T|G)p(C|T, G)
\] (1)

In general, for any BN it holds that the joint probability distribution of the whole network is defined as the product of the conditional probabilities of all the nodes given their parents. In this way, the BN efficiently decomposes the joint probability distribution into local (conditional) probability distributions.

Stochastic processes describing the development of a system in time or space can be modelled by Dynamic Bayesian Networks (DBNs), an example of which is shown in Figure 1b. The \( i \)th slice of the DBN represents the state of the system in time/position \( i \), here consisting of the two random variables \( G_i \) and \( T_i \). The joint probability of \( G_i \) and \( T_i \) is obtained as

\[
p(G_i, T_i) = \sum_{G_{i-1}} p(G_{i-1})p(G_i|G_{i-1})p(T_i|G_i)
\] (2)

where \( p(G_{i-1}) \) is the marginal probability distribution of random variable \( G_{i-1} \) describing the geology in slice \( (i - 1) \), \( p(G_i|G_{i-1}) \) is conditional probability describing changes of geology between neighbouring slices of the DBN and \( p(T_i|G_i) \) is the conditional probability describing the construction time for given geological conditions.

3 MODELLING TUNNEL EXCAVATION PROCESS USING BAYESIAN NETWORKS

In the following, we present a BN model of the tunnel excavation process for a specific tunnel that was previously investigated by other researchers. This application facilitates the validation of the BN model, while it is sufficiently general to draw conclusions on the applicability of the model and to investigate the influence of the model assumptions.
3.1 Tunnel specifics

The Suncheon-Dolsan road tunnel is located in the south of South Korea between the towns Suncheon and Dolsan. The project and its modelling were described in Min (2003), Min et al. (2003), Min et al. (2005) and Min et al. (2008). The tunnel consists of two tubes with length of 1.9 km, of which we consider only one tube. The tunnel was constructed from both tunnel ends, the respective sections are denoted as section A (of length 610m) and section B (of length 1290m). In this paper, only results for section A are presented. The geometry and geotechnical zones as taken from Min (2003) are shown in Figure 2. The NATM with drill and blast technology was applied for excavation. Geological conditions in the area are good, consisting mostly of Micrographic Granite and Diorite. Based on the available investigations (borehole drilling, electrical resistivity survey and seismic exploration), five rock classes were defined by means of three parameters (RMR, Resistivity and Q-value) for modelling purposes.

3.2 Bayesian network model of tunnel excavation process

A DBN model is developed to represent the various uncertain factors influencing the tunnel excavation process. Each slice in the DBN consists of random variables describing the uncertain geotechnical conditions and construction process variables in a tunnel segment of length $\Delta l$. The $i$th slice represents a tunnel segment from position $(i-1)\Delta l$ to position $i\Delta l$ along the tunnel axis. Within one slice, all random variables are modelled as constant, i.e. the states of the variables are fixed over the interval $\Delta l$. For this reason, $\Delta l$ must be chosen in order to best represent the real excavation process, as discussed in Section 3.6.

Two alternative DBN models are shown in Figure 3. The variables of the models are described in Table 1. DBN (a) corresponds to the DAT model originally used in Min (2003). It should be remembered that the DAT does not use BN, yet every probabilistic model can be interpreted as a BN. DBN (b) displays an enhanced model, including additional variables and dependences in the construction process. Both DBN models are discrete-space Markov chain models. They are inhomogeneous, i.e. the conditional probability distributions of the variables are varying along the tunnel axis. Both DBN models are introduced in detail in the following sections.
3.3 DBN model of geotechnical conditions

The geotechnical conditions within a slice $i$ of the tunnel are described by the random variables rock class $R_i$, height of the overburden $O_i$, ground class $G_i$ and, in the extended model, zone $Z_i$. For modelling the rock class $R_i$, the tunnel is first divided into zones within which the rock class can be modelled by the same conditional probability distribution. (In statistical terminology, rock class is a homogenous process within a zone.)

As evident from Figure 3, the spatial variability of rock class along the tunnel axis is modelled as a Markov process in both DBN models. The suitability of Markov processes for modelling of geotechnical parameters (including rock class, degree of jointing) along the tunnel axis was shown already in Chan (1981). Since then, the DAT model is based on continuous Markov process models. In the DBN model, the Markov process is discretized into a Markov chain (i.e. transformed to a discrete space represented by slices of the DBN). The rock class of each slice is described by a conditional probability table (transition probabilities), an example of which is given in Table 2. These conditional probabilities are derived from the parameters of the continuous Markov process reported in Min (2003), assuming that changes in rock class occur as a Poisson process, in accordance with Chan (1981).

Table 1. Overview of DBN model variables

<table>
<thead>
<tr>
<th>Id.</th>
<th>Variable</th>
<th>Type</th>
<th>States of discrete/ type of continuous distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>Rock class</td>
<td>Random/Discrete</td>
<td>I, II, III, IV, V</td>
</tr>
<tr>
<td>O</td>
<td>Overburden</td>
<td>Determ./Discrete</td>
<td>Low, Medium, High</td>
</tr>
<tr>
<td>E</td>
<td>Geometry</td>
<td>Determ./Discrete</td>
<td>1 (begin/end), 2 (typical), 4 (chem.plant), 5 (EPP)</td>
</tr>
<tr>
<td>T</td>
<td>Unit time</td>
<td>Random/Cont.</td>
<td>Triangular</td>
</tr>
<tr>
<td>Q</td>
<td>Quality</td>
<td>Random/Discrete</td>
<td>Poor, good, excellent</td>
</tr>
<tr>
<td>Z</td>
<td>Zone</td>
<td>Random/Discrete</td>
<td>1, 2, ..., 17</td>
</tr>
</tbody>
</table>

Table 2. Conditional probability table for Markov model of rock classes in zone 1, for a DBN slice with length of $\Delta l = 4m$.

<table>
<thead>
<tr>
<th>$R_i$</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.1353</td>
<td>0.2149</td>
<td>0.2149</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>0.5707</td>
<td>0.3679</td>
<td>0.4172</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>III</td>
<td>0.2940</td>
<td>0.4172</td>
<td>0.3679</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IV</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

In the enhanced DBN, the locations of borders of zones with statistically homogeneous rock class conditions are modelled as random by introducing the random variable $Z_i$. Let $Pr(Z_i = k)$ denote the probability that the $i$th slice of the DBN is part of zone $k$ and $Pr(Z_{i-1} = k)$ the probability that the $(i-1)$th slice lies in zone $k$. Furthermore, let $F_{Bk}(x)$ be the known cumulative probability distribution function (CDF) of the location of the border between zones $k$ and $k+1$. Assuming that probability distributions of zone borders are non-overlapping, the probability of the $i$th slice being in zone $k$ can be determined as

$$Pr(Z_i = k) = 1 - F_{Bk}(\Delta l - \frac{\Delta l}{2})$$

(3)

The corresponding conditional probabilities are

$$Pr(Z_i = k | Z_{i-1} = k) = \frac{Pr(Z_i = k)}{Pr(Z_{i-1} = k)}$$

(4)

$$Pr(Z_i = k + 1 | Z_{i-1} = k) = 1 - \frac{Pr(Z_i = k)}{Pr(Z_{i-1} = k)}$$

(5)

$$Pr(Z_i = k + 1 | Z_{i-1} = k + 1) = 1$$

(6)
The height of overburden $O_i$ is modelled deterministically. The ground class $G_i$ is defined deterministically for given $R_i$ and $O_i$. As evident from Table 1, each $G_i$ corresponds to a specific combination of $R_i$ and $O_i$.

3.4 DBN model of construction performance

Construction performance in a slice $i$ of the tunnel is described by the variables cross section geometry $E_i$, construction method $M_i$, excavation time $T_i$, and, in the extended model, design/construction quality $Q_i$.

The deterministic variable geometry $E_i$ enables the variations of the cross section to be distinguished along the tunnel (e.g. typical cross section vs. extended cross section for emergency parking places EPP). It is also used to consider the special requirements for the construction process at the beginning/end of the tunnel and in the area where the tunnel passes underneath an existing chemical plant.

The applied construction method $M_i$ describes the excavation type and the related support pattern and is determined conditional on the ground class $G_i$ and tunnel geometry $E_i$. The modelling of $M_i$ is taken from Min (2003), where the details of the different construction methods are described.

For every construction method $M_i$, the excavation time $T_i$ is defined by a probability distribution $F(t_i|m_i)$. For representation in the DBN, the continuous distribution is discretized as described in Straub (2009). By not including a direct link between $M_{i-1}$ and $M_i$, the model assumes full flexibility in changing construction methods from one slice to the next. In addition, it is assumed that changes of construction patterns are not connected with additional switch-over time. This neglects the fact that changes in the excavation technology and the support structure can be demanding with respect to both time and costs (Sousa 2010).

In the extended model, the excavation time $T_i$ is furthermore dependent on the design/construction quality $Q_i$. This additional variable represents the uncertain quality of design and construction works, which introduces dependence among the performance in each slice of the tunnel. The quality $Q_i$ is in one of the three states “poor”, “good” or “excellent” throughout the entire tunnel construction, i.e. the variable is fully dependent from one slice to the next and the conditional probability matrix $p(q_i|q_{i-1})$ in each slice is thus the 3x3 identity matrix. This simple model reflects the fact that the quality of a tunnel project cannot be directly measured and can only be deduced from the average performance over long sections of the tunnel project (Špačková et al. 2010). The quality influences the conditional distribution of the excavation time $T_i$: the better the construction quality, the lower the variability of $T_i$. For each state of $Q_i$ and construction method $M_i$, a different distribution $F(t_i|m_i, q_i)$ is defined for the excavation time $T_i$.

3.5 Calculation of the distribution of the total excavation time in the DBN

In the application presented in this paper, the main output is the estimate of the total excavation time. In the DBN model, this is obtained by introducing the cumulative time $T_{\text{cum},i}$ in each slice, defined as $T_{\text{cum},i} = T_{\text{cum},i-1} + T_i$. $T_{\text{cum},i}$ represents the time needed for excavation of the tunnel from the beginning to location $i\Delta l$. Because exact inference algorithms are used for evaluating the DBN, in particular the Frontier Algorithm (Murphy 2002), both $T_i$ and $T_{\text{cum},i}$ must be discretized. Due to the required fine discretization of $T_{\text{cum},i}$, and associated large number of states, the definition of the conditional probability table of $T_{\text{cum},i}$ becomes impracticable. For this reason, the frontier algorithm was modified using a convolution function that allows defining the conditional probability table to be avoided. The new algorithm is computationally efficient (computations shown here take in the order of 20 - 180 seconds on a standard computer).

3.6 Influence of slice length in the DBN model

By choosing a slice length $\Delta l$ in the DBN model, we make implicit assumptions about dependences among the variables along the tunnel. In the model, changes of conditions can only occur between slices. Therefore, $\Delta l$ must be sufficiently small to capture the variability of geotechnical conditions along the tunnel axis, in particular the rock class $R_i$. This can be assessed by the probability that a change of $R_i$ occurs within one slice. Using the Poisson assumption, this probability is $\Pr(\text{Change}) = 1 - \exp(-\Delta l/l_R)$ where $l_R$ is the mean length of a particular rock class along the tunnel axis. For the Dolsan tunnel, $l_R$ is in the range 1.5m – 43m. As a rule of thumb, a value of $\Delta l \leq l_R$ provides reasonable accuracy for modelling changes in the geotechnical conditions along the tunnel. This requirement must be considered along with other criteria.
Because the model assumes that the construction method in slice $i$ is determined purely based on the geotechnical conditions (ground class $G_i$) and the cross section geometry $E_i$, it implies full flexibility in changing construction methods between slices. In reality, construction methods are only changed between excavation cycles. Therefore, for the model to be realistic, the slice length should not be shorter than the length of the excavation cycles. Unless otherwise specified, the calculations in this paper are based on $\Delta l = 4\text{m}$.

The conditional distribution used to specify the variables in the DBN must be adjusted for the slice length. In particular, the conditional probability table of $R_i$ (as shown exemplarily in Table 2) must be calculated specifically for a given value of $\Delta l$ using the Poisson assumption. Furthermore, the excavation time $T_i$ depends directly on $\Delta l$. If the mean and variance of the time $T_{ref}$ for a reference length $\Delta l_{ref}$ are known, then the mean and variance of $T_i$ are

\[
E[T_i] = \frac{\Delta l}{\Delta l_{ref}} E[T_{ref}],
\]

\[
\text{Var}[T_i] = \frac{\Delta l}{\Delta l_{ref}} \text{Var}[T_{ref}].
\]

It is assumed that the probability distribution of $T_{ref}$ is a triangular distribution. However, the choice of the distribution type has a little effect on the final results, due to the fact that the cumulative time is obtained as the sum of a larger number of $T_{ref}$’s.

4 NUMERICAL INVESTIGATIONS

4.1 Original model - DBN in Figure 3a

To validate the DBN model, the DBN in Figure 3a is constructed with the same assumptions as used in the DAT model presented in Min (2003). The resulting DBN is then applied to compute the total excavation time $T_{tot}$ for the Dolsan A tunnel. Unlike in Min (2003), the delay between excavation of heading and bench was not considered in the DBN as it has little impact on total construction time and the excavation of the tunnel portal was not modelled because necessary data were not available. Even with these differences, the calculated mean value of $T_{tot}$ is within 3% of the value given in Min (2003) and the standard deviation of $T_{tot}$ is within 10% of the value given in Min (2003), as seen from Table 3.

The results presented in Min (2003) are based on an inconsistent definition of the probability distributions of the advance rates (it ignores the fact that the advance rate is defined as an average over certain lengths and that the variance of this average advance rate thus depends on the corresponding length of the tunnel). To overcome this inconsistency, the results presented in the following use the assumption that the variances of the advance rates given in Min (2003) are valid for 10m of tunnel excavation. With this assumption, the resulting $T_{tot}$ is as given in Table 4. It can be observed that the variance of $T_{tot}$ decreases compared to the results with the original definitions, where the given distributions of average advance rates are applied to sections that have lengths $10\text{m} – 120\text{m}$.

<table>
<thead>
<tr>
<th>Simulation type</th>
<th>DOLSAN A</th>
<th>(using the original assumptions of DAT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total constr. time (days)</td>
<td>Mean</td>
<td>St.dev.</td>
</tr>
<tr>
<td>DAT acc. to (Min 2003)</td>
<td>195</td>
<td>3.39</td>
</tr>
<tr>
<td>MC – discrete space</td>
<td>191</td>
<td>3.17</td>
</tr>
<tr>
<td>DBN</td>
<td>191</td>
<td>3.06</td>
</tr>
</tbody>
</table>

Table 3. Comparison of results from DAT and DBN model

<table>
<thead>
<tr>
<th>Simulation type</th>
<th>DOLSAN A</th>
<th>(using the modified assumptions of DAT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total excavation time (days)</td>
<td>Mean</td>
<td>St.dev.</td>
</tr>
<tr>
<td>MC – discrete space</td>
<td>190</td>
<td>1.66</td>
</tr>
<tr>
<td>DBN</td>
<td>190</td>
<td>1.64</td>
</tr>
</tbody>
</table>

Table 4. Comparison of results from MC simulation and DBN model (using the modified assumptions of DAT)

4.2 Extended model - DBN in Figure 3b

In the extended DBN shown in Figure 3b, variables $Q_i$ describing the design/construction quality and variables $Z_i$ describing the uncertainty in the position of zone borders are introduced. The probability of different quality classes were assigned based on engineering judgement as $\text{Pr}(Q = \text{excellent}) = 0.1$, $\text{Pr}(Q = \text{good}) = 0.6$ and $\text{Pr}(Q = \text{poor}) = 0.3$. The probability distributions of the excavation times $T_i$ are now defined conditional on the quality; for projects with excellent quality, the distributions from the DAT model used above are applied. For projects with good and poor quality, distributions with higher variances are used. These models are based on data from a tunnel project in Czech Republic, which also used NATM, indicating that the variances of $T_i$ are considerably higher than those given in Min (2003).
(It is pointed out that the available data does not allow a representative statistic, but the observations correspond with general experience on tunnel projects in the Czech Republic.) The conditional distributions of \( T_{tot} \), from which the distributions of \( T_i \) are calculated (see Par. 3.6), are shown exemplarily for a particular construction method in Figure 4. The calculations were performed under two different assumptions: (a) the mean value of the excavation times \( T_i \) is not dependent on the quality and is as in Min (2003) and (b) the mean value of the excavation times \( T_i \) is increased by a factor of 1.07 in the case of good quality and by a factor of 1.15 in the case of poor quality.

The comparison of the total excavation time \( T_{tot} \) for the Dolsan A tunnel with 610m length as calculated by means of the DBN model with the original assumptions and the extended DBN model is displayed in Figure 5. The variance of the \( T_{tot} \) is significantly higher with the extended model, in particular when including a dependence of the mean excavation time on the quality (case b), and is likely to more represent realistically the true uncertainties in the predictions of the tunnel construction process. The influence of the main assumptions in the extended DBN model is further studied in a sensitivity analysis.

4.3 Sensitivity analysis

Figure 6 displays the results of the sensitivity analysis performed with the extended DBN model (a). In Figure 6a, the influence of the spatial discretization is shown. With increasing slice length \( \Delta l \), the variance of \( T_{tot} \) slightly increases. This is due to the assumption that the construction method can be freely selected for each slice. With the choice of a large \( \Delta l \), a limited flexibility of the construction technology is assumed, which leads to a higher variance of \( T_{tot} \). Figure 6b shows the influence of including the design/construction quality \( Q_i \) in the model. If \( Q_i \) is known to be excellent, the variance of \( T_{tot} \) is smaller than in the case of unknown \( Q_i \). Finally, Figure 6c illustrates the effect of including the variables \( Z_i \), which allow the position of the geotechnical zones to be modelled as random, in the DBN model. For this application, it is found that the consideration of this randomness has a negligible effect on the estimate of \( T_{tot} \). However, this effect might be larger if the excavation times \( T_i \) would vary more strongly between different construction methods.

5 CONCLUDING REMARKS

A novel model for tunnel excavation processes based on Dynamic Bayesian Networks (DBN) was introduced in this paper. The main new feature of this model is that it explicitly includes the quality of the design and construction process. Because the quality is expected to be similar over the entire process, the uncertainty in the quality leads to an increased uncertainty (variance) of the estimate of the total construc-
tion time, which appears to reflect more realistically the actual uncertainties in tunnel projects. While it is not included in this paper, the modelling of the excavation cost can be performed analogically.

The main inputs to the DBN model, like to any other model of tunnel construction process, are the probability distributions of the excavation times (advance rates) for given construction methods and the geological conditions. When determining these probability distributions, due attention must be paid to the definition of these variables, since their variance is a direct function of the reference length (i.e. the length over which the advance rates are averaged). In our experience, estimates of the variances made by experts are not generally reliable (unlike estimates of the mean excavation times). Therefore, a next step will be to obtain more realistic estimates of these variances based on data from past tunnel projects.

The DBN model will be further developed along several lines. On the one hand, additional factors will be included in the model for assessing the full project risks. These include the switch-over time and cost as well as extraordinary events (e.g. cave-in collapses). On the other hand, the automatic updating of the model with observations made during the geological investigations and during the tunnel construction process will be facilitated. To this end, the algorithms for evaluating the DBN are currently further developed.

ACKNOWLEDGEMENTS

Funding by DAAD through the programme “Research grants for doctoral candidates” is gratefully acknowledged. Furthermore, this work has received financial support from the project SGS10/020/0HK1/1T/11 at the Czech Technical University in Prague and project No. TA01030245 of the Technology Agency of the Czech Republic.

REFERENCES


Min, S.Y. 2003. The application of decision Decision Aids for Tunneling (DAT) to the Sucheon tunnel in Korea. M.Sc. Thesis, Massachusetts Institute of Technology, Cambridge, USA.


