Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

IMPROVING TUNNELING SIMULATION USING BAYESIAN UPDATING AND HIDDEN MARKOV CHAINS

Michael Werner
Wenying Ji
Simaan AbouRizk

Department of Civil and Environmental Engineering
University of Alberta
Edmonton, Alberta, CANADA

ABSTRACT

Ground conditions remain an uncertain factor in tunneling projects, complicating the ability of practitioners to reliably estimate project productivity and, in turn, duration. This study proposes a Bayesian-based approach to incorporate real-time project data into simulation-based ground prediction models to improve prediction accuracy. Changes in ground conditions are modeled using a Hidden Markov Model, which is updated with actual project data using the Baum-Welch algorithm. The prediction model is then incorporated in *Simphony.NET* to enhance simulation of tunneling construction operations. A case study conducted in Edmonton, Canada, demonstrates that the proposed approach is capable of incorporating real-time data in a manner that resulted in enhanced duration prediction accuracy.

1 INTRODUCTION

Productivity of tunneling operations depends, in part, on the geological conditions encountered during project delivery. While geotechnical investigations (i.e., intermittent borehole samples taken across the tunnel alignment) are routinely conducted during the planning stages of tunnel construction, these investigations are primarily concerned with the identification of major geotechnical risks, such as voids, water pockets, or other unexpected ground conditions. Sampling at a frequency that would provide the data required to reliably estimate productivity is extremely costly and, in some cases, unachievable (e.g., underground conditions where only certain portions of the project site are accessible for investigation). In practice, ground conditions used to estimate productivity are often assumed from past projects or by domain experts. This can result in inaccurate predictions and, consequently, in unexpected project delays and expenditures.

While Markov process models have been developed to predict the uncertainty of geological conditions in the past (Chan 1981; Ioannou 1987; Liu et at. 2009; Sutanto 1997), these models failed to consider dependency between observation variables. In recent years, academic models for predicting ground conditions have been developed using Hidden Markov Model (HMM)-based approaches to achieve more accurate predictions (Leu and Adi 2011; Zhang et al. 2015). HMMs are suitable for predicting ground conditions because they can consider uncertain ground parameters together with their location information. Despite improvements for predicting conditions in the planning phase, these models are unable to appropriately incorporate relevant, real-time data that are acquired during the execution phase of construction. A prediction model capable of improving the accuracy of HMM parameters, by incorporating dynamically updated information, is expected to enhance tunneling project planning and delivery.

Updating parameters of a model using new information has been systematically investigated by researchers. Bayesian-based approaches have been successfully applied to derive analytical posterior distributions in construction simulation for normal (Chung et al. 2006) and beta distributions (Ji and

AbouRizk 2017). However, the set of observations in a HMM is a sequence of various observable outcomes and, therefore, traditional Bayesian updating techniques are not as directly applicable. With HMMs, both the unique instances in the data sequence and the transition from one outcome to the next must be accounted for when updating model parameters. The observations, as well as the transition information, become required inputs in the updating process. Although analytical methods cannot be applied for updating HMMs, numerical techniques can, potentially, be used for reasonable HMM updating (Rabiner 1989).

Here, a Baum-Welch algorithm (Baum and Petrie 1966) is implemented to allow HMMs in tunneling simulation models to be updated with real-time data. The HMM, the Baum-Welch algorithm-based parameter updating technique, and the creation of the simulation environment are detailed, and a case study is performed to demonstrate the feasibility and applicability of the proposed approach.

2 METHODOLOGY

The proposed methodology is comprised of three primary components: (1) the HMM, which is used in this study for stochastic modeling purposes, (2) the Baum-Welch algorithm, which is used to achieve parameter updating of the HMM, and (3) the specialized simulation environment, which is created using the construction simulation platform, *Simphony.NET* (AbouRizk et al. 2016). The methodology is detailed as follows.

2.1 Hidden Markov Model

A Hidden Markov Model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (i.e., hidden) states. A graphical representation of a HMM is illustrated in Figure 1.

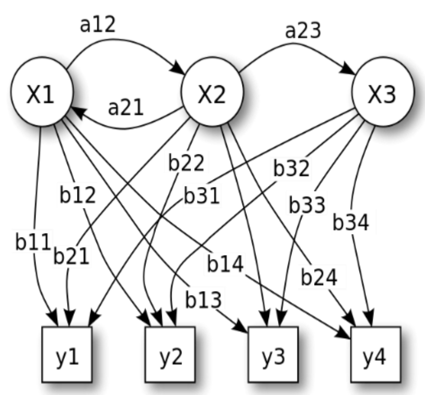

Figure 1: Graphical example of a HMM, where (x) represents a hidden state, (y) represents an observable state, (a) represents a transition probability from a hidden state to another, and (b) represents the emission probability from a hidden state to an observable state.

The first step for estimating a Markov model using a Bayesian approach involves determining the inputs. A HMM includes initialized parameters, namely prior state density matrix-π, transition matrix-A, and observation matrix-B. Rabiner and Juang (1993) suggested using uniformly distributed probabilities for initializing prior state density matrices and transition matrices, as these values are unknown and every possibility is equally likely to occur. Alternatively, Rabiner and Juang (1993) have also recommended making reasonable assumptions in the initialization of the observation matrix. The notation used for HMM parameters, their corresponding matrices and vectors, and their elements, based on Rabiner and Juang (1986) and Rabiner (1989), are detailed as follows:

$X = \{x_1, x_2, \dots, x_T\}$: A set of all possible hidden states
$Y = \{y_1, y_2, \dots, y_T\}$: A set of all possible observable states
$A = \{a_{ij}\}, a_{ij} = P(s_{t+1} = j \mid s_t = i)$: State transition probability matrix
$B = \{b_{im}\}, b_{im} = P(x_t = m \mid s_t = i)$: Emission probability matrix
$\pi = \{\pi_i\}, \pi_i = P(s_1 = i)$: Initial state distribution or prior distribution
$\theta = \{A, B, \pi\}$: HMM parameters
$U = \{u^{(A)}, u^{(B)}, u^{(C)}\}$: Hyperparameters used to define the prior over θ

2.2 Baum-Welch Algorithm

Bayesian statistics allow for the modeling of changes in probabilistic values within a random system. In the case of Markov modeling, Bayesian statistics allow users to easily update predictions as new information becomes available. In essence, the posterior prediction is an integration of prior predictions and newly collected information. However, determining a method to adjust HMM parameters (i.e., A, B, π) to maximize the probability of the observation sequence from a given model is extremely difficult; indeed, there is no way to analytically solve such a problem (Rabiner 1989). Rather, parameter derivation can be achieved using iterative numeric methods, such as the expectation modification method (Dempster et al. 1977) or gradient techniques (Levinson et al. 1983).

The Baum-Welch algorithm uses the well-known Expectation Maximization algorithm to determine the maximum likelihood estimate of HMM parameters from a set of observed feature vectors. The Baum-Welch algorithm, first proposed by Leonard E. Baum and Lloyd R. Welch in a series of articles in the late 1960s (Baum and Petrie 1966; Baum and Sell 1968; Baum et al. 1970), is numeric in nature and is based on the forward-backward procedure described by Rusian Stratonovich (1960). The Baum-Welch method was selected for implementation in this study due to its ease of application and to its ability to be implemented into the *Simphony.NET* modeling environment.

The algorithm involves using a set of observed outcomes to update the parameters of a HMM. These observed outcomes may be a simple set containing one observation array or may be comprised of a complex series of observation arrays. The algorithm uses these sets of observations to generate parameters, which are referred to as updating parameters. These updating parameters represent probability values and are used to update the parameters of the HMM. The updating parameters output by the algorithm include: α, β, ξ and ϒ. The ϒ and ξ updating parameters are computed based on the α and β parameters. ϒ and ξ are the only parameters used in the final updating step. The formulae for calculating these parameters are summarized in Equations 1 to 6.

$$\alpha_1(j) = \pi_j b_j(O_1) \tag{1}$$

$$\alpha_{t+1}(j) = b_j(O_{t+1}) \sum_{i=1}^{N} \alpha_t(i) a_{ij} \tag{2}$$

$$\beta_T(j) = 1.0 \tag{3}$$

$$\beta_t(j) = \sum_{i=1}^{N} \beta_{t+1}(i) a_{ji} b_i(O_{t+1}) \tag{4}$$

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} \tag{5}$$

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} \tag{6}$$

After the above parameters are calculated, updated probabilities for the A and B matrix are calculated using Equations 7 and 8.

$$a_{ij}^* = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{7}$$

$$b_i^*(v_k) = \frac{\sum_{t=1}^{T} \gamma_t(i)\psi}{\sum_{t=1}^{T} \gamma_t(i)} \tag{8}$$

2.3 Simphony Modeling Environment

The majority of simulation environments available today provide graphical modeling constructs to visually represent the logic flow of systems, operations, and processes that have been abstracted for the purposes of simulation-based analysis. Here, *Simphony.NET* was used as the simulation environment for achieving the proposed approach. Specific Markov chain elements were developed for use in the *Simphony.NET* simulation environment general template, enabling the definition and emulation of Markov chains in *Simphony.NET*, and enhancing flexibility of the types of models that can be constructed.

Markov elements were designed to have one-to-one mapping between constructs that appear in the graphical layout of a Markov chain model on paper and one developed using the envisioned *Simphony.NET* environment. Markov chain states are represented by a state modeling visualization and, for the case of hidden Markov chains, are represented by observations. Transitions between hidden states and emission from hidden states to observable states are represented using directional arrows, with a probability value assigned to each. The higher-level "MarkovModel" modeling element is responsible for controlling and managing Markov chain model parameters from the model layout, simulating the Markov chain, and tracking Markov chain behavior. The "MarkovTransition" modeling element was created to trigger the Markov chain to be stepped during the simulation. Elements implemented in the modeling environment and the Markov model created within the *Simphony.NET* user interface are illustrated in Figure 2.

3 CASE STUDY AND RESULTS

To demonstrate the feasibility and applicability of the proposed approach, a real project from Edmonton, Canada, is utilized to perform the simulation analysis. The case project was a three-phase drainage improvement project; this case study focuses on two of these phases including a 1km tunnel (Project A) and a 500m tunnel (Project B). Both tunnels were 2340mm diameter storm tunnels completed using Tunnel

Boring Machine (TBM) excavation. Based on borehole data collected prior to construction, the majority of tunneling was expected to involve clay till (with inter-till sand zones), a mix of sandy clay, and the possibility of sand pockets.

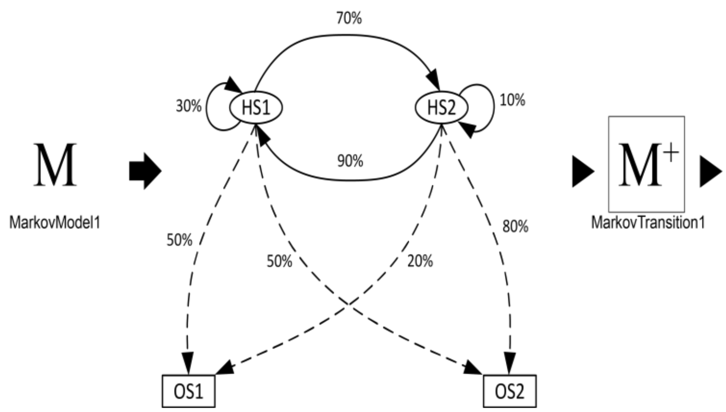

Figure 2: Markov model elements embedded within *Simphony.NET*.

3.1 Data Preparation and Model Inputs

Beginning in September 2015, TBM excavation proceeded along Project A; 978m were completed at the end of August 2016. Throughout this phase, geotechnical conditions, daily excavation progress, and shift length were monitored on a daily basis. Project A data are summarized in Table 1.

Table 1: Actual project data from Project A.

Ground Conditions	Count (m)	Length of Persistence (m)			Effective Duration (hr./m)*	
		Average	Minimum	Maximum	Average	Std. Dev.
Clay	207	13.8	3	44	2.16	0.16
Clay Sand	712	37.5	3	174	2.10	1.45
Sand	59	6.6	2	10	2.34	2.07

**Effective Duration calculated as shift time divided by meters excavated each day.*

Actual project duration data for Project B, at 100m intervals, are detailed in Table 2. These data will be compared to simulation results.

Table 2: Actual project data from Project B.

	0-100m	100-200m	200-300m	300-400m	400-500m	**Total**
Clay Clay Sand	2.10 ± 0.43 n/a	2.18 ± 0.59 2.03 ± 0.06	2.73 ± 1.61 2.44 ± 0.65	2.33 ± 0.69 n/a	2.70 ± 1.51 n/a	**2.41 ± 1.07** **2.35 ± 0.58**
Actual Duration (hrs.)	211	208.50	232.50	229.50	270	**1152**
Total Actual Duration (hrs.)	211	419.5	652	881.5	1152	

**Values are average ± standard deviation.*

3.2 Model and Assumptions

A tunneling construction simulation model, developed based on a previously proposed model (AbouRizk et al. 2016), was developed. Due to the similarity of the present case study, the proposed model was used as a baseline. This model was then embellished with the Markov chain elements as well as updated project parameters for the case study. The model schematic is illustrated in Figure 3.

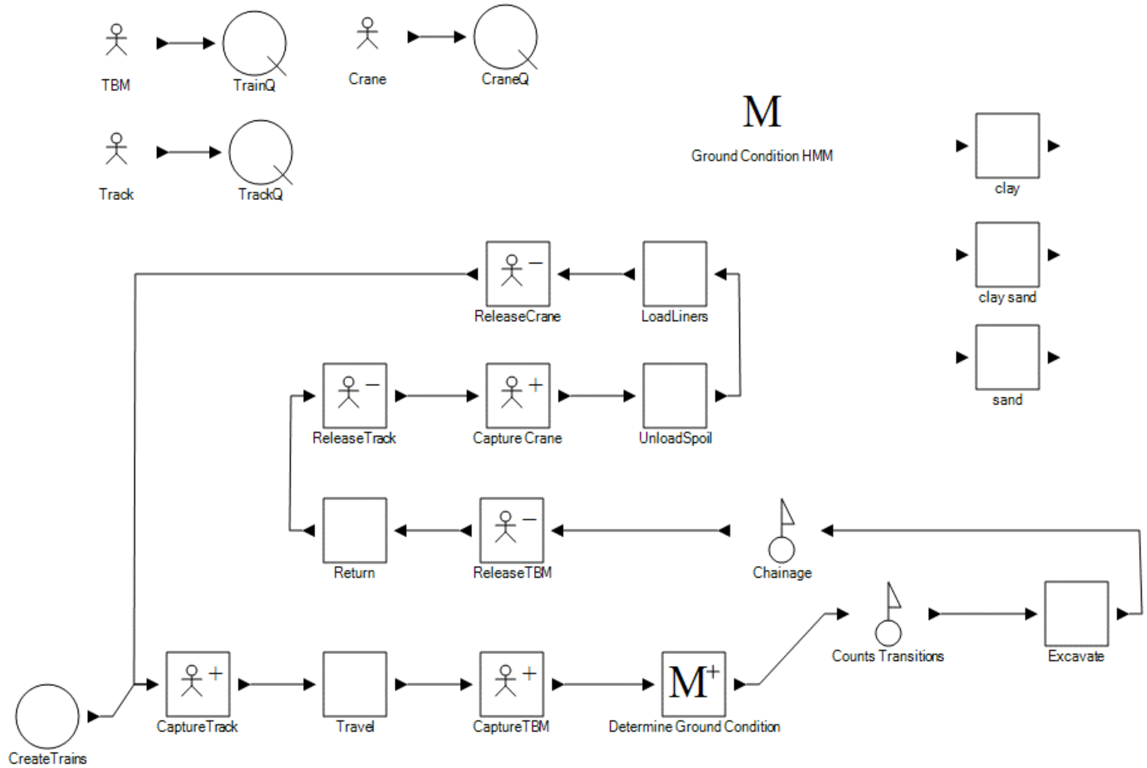

Figure 3: Case study tunneling model schematic.

Excavation duration is dependent on the ground condition (i.e., output) of the HMM chain. The pseudo code below provides details of how excavation duration was determined:

```
If Ground Condition HMM has an output of "clay"
    Duration = Sample of Normal Distribution for clay (Average duration for clay,
standard deviation for clay)
If Ground Condition HMM has an output of "clay sand"
    Duration = Sample of Normal Distribution for clay sand (Average duration for clay
sand, standard deviation for clay sand)
If Ground Condition HMM has an output of "sand"
    Duration = Sample of Normal Distribution for sand (Average duration for sand,
standard deviation for sand)
```

Task elements (shown on the right side of Figure 3) are used to store the duration distributions of each ground condition. Bayesian updating occurs, here, through the distribution fitting interface provided within *Simphony.NET*. Notably, certain assumptions were made to facilitate the clear implementation of the proposed updating and prediction techniques: duration data were assumed to be normally distributed, since the *Simphony.NET* modelling environment has existing capabilities to handle normal distributions.

3.3 Simulation Modeling and Run Details

A total of 6 simulation runs were completed to assess the effectiveness of the proposed HMM and Bayesian updating components in the tunneling model. As illustrated, the initial values of transition and emission probabilities were uniformly set. The tunneling model in Figure 3 was then trained using Project A data (Table 1) to develop the baseline HMM using the implemented Baum-Welch algorithm as illustrated in Figure 4.

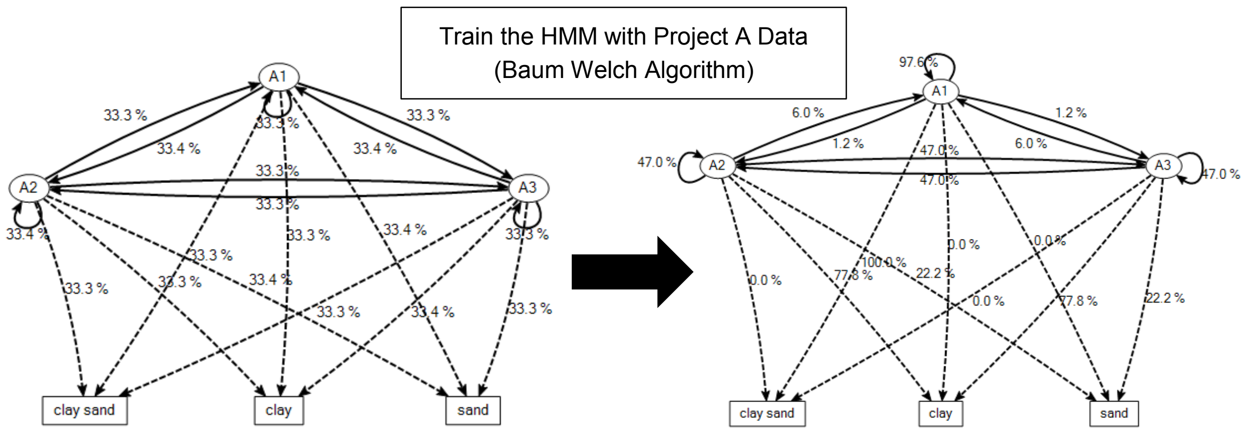

Figure 4: Baum-Welch training step of HMM.

A simulation run, which simulated project updates at 100m intervals for the 500m project, was completed to demonstrate how actual construction project planning and control would occur. Prior to project execution, a model would be created using basic assumptions and best guesses to estimate production rates and durations. Notably, the preliminary model could be improved using historical data of similar projects or expert knowledge. An initial run would simulate tunneling of the entire project. In practice, the execution phase of construction would begin, and real data would be collected and used to update the model and duration prediction. This process would be repeated at specified intervals until the project is complete. Here Project A data were used to "train" the Project B model. The first run (Run 1) simulated tunneling of the entire 500m. Predicted duration of Project B was recorded. At this point, actual Project B data were used to update the model and to predict the duration for the remaining 400m (as Run 2). This process was repeated at 100m intervals. Duration of each run was analyzed by adding the actual duration to date to the average simulated duration for the remaining length of tunnel. Each run was performed 1000 times to ensure a meaningful sample of results was obtained. Results are summarized in Table 3.

Table 3: Details of simulation model runs.

Run Number	Simulated Length (m)	Actual Data Length (m)
1	500	0
2	400	100
3	300	200
4	200	300
5	100	400
6	0	500

Since it uses actual project data to simulate the entire project, Run 6 (i.e., the final run) can be compared to the actual project duration to determine accuracy of the model as well as the updating process.

3.4 Results

For illustrative purposes, the results of three runs are summarized. The HMM chain, predicted ground conditions, and duration distributions of each run are presented. Evolution of the HMM from Run 1 to 5, and 5 to 6, are illustrated in Figure 5.

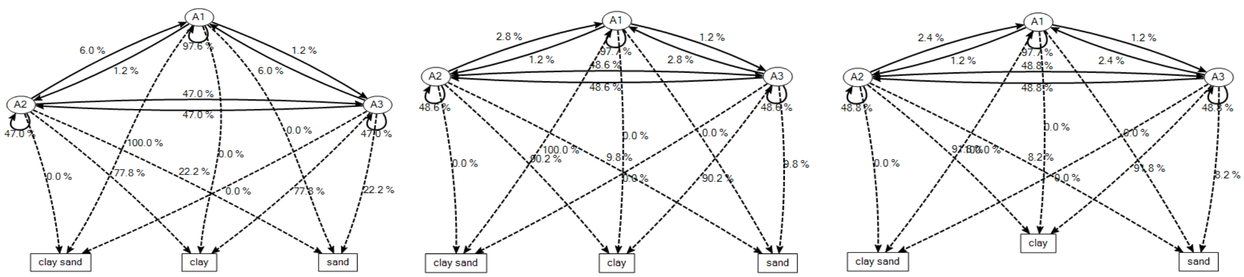

Figure 5: HMM chain for runs 1, 5, and 6 (from left to right, respectively).

In Run 1, hidden state A1 dominates the chain with a very low likelihood of transitioning to either A2 or A3. As is often observed in HMM chains, each hidden state is generally more associated with an observation—in this case, a specific ground condition. For example, A1 is primarily associated with clay sand and has a self-transition probability of 97.6%, dominating the HMM chain and approaching what is known as an absorbing state. As the HMM is updated, the likelihood of observing clay increases for all hidden states. This behavior is expected: while previous Project A data were predominantly clay sand, actual data from Project B were mostly clay (particularly through the first 200m). Sand was not observed during Project B excavation, as is reflected by the decreasing likelihood of transition to sand. Visual representations of the simulated ground conditions encountered in Runs 1, 5, and 6 are illustrated in Figure 6.

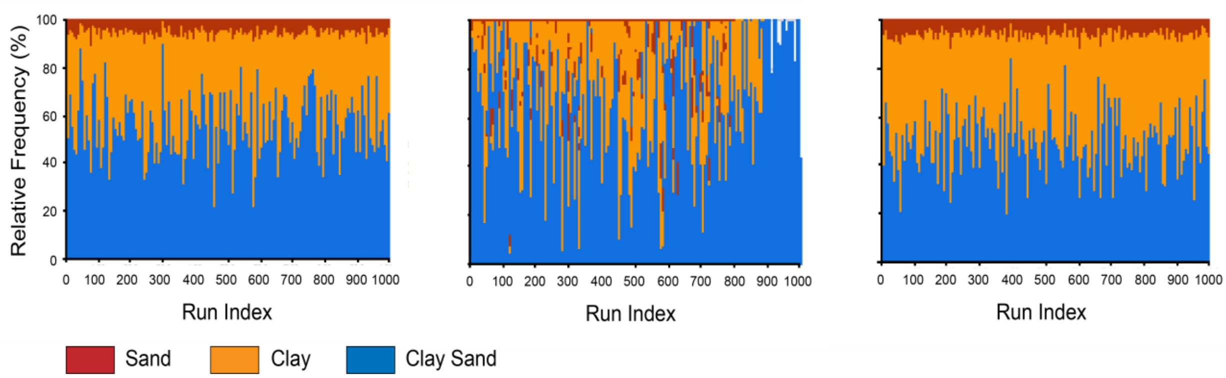

Figure 6: Simulated ground conditions encountered for runs 1, 5, and 6 (from left to right, respectively).

Run 1 is dominated by Project A data and, consequently, clay sand is the most commonly encountered ground condition in this run. In Runs 5 and 6, the influence of Project B data emerge, as evidenced by the greater and lower frequency of clay and sand encountered, respectively. Project durations for each run are summarized in Table 4.

Table 4: Summary of run durations.

Run Number	Simulation Average Duration (hrs.)	Actual Duration (hrs.)	Total Duration (hrs.)
1 (500m)	1491	0	1491
2 (400m)	1185	211	1396
3 (300m)	790	420	1210
4 (200m)	527	652	1179
5 (100m)	264	882	1146
6 (500m)	1331	0	1331

Actual duration for Project B was 1152 hours (or 144 days) of construction, which is used as a baseline for evaluating the effectiveness of the proposed methodology. Results indicate that duration prediction accuracy was improved after each addition of actual data, beginning with a prediction of 1491 hours (or 186 days) in Run 1 and ending with a prediction of 1146 hours (or 143 days) in Run 5. Compared to Run 1, which was based on historical information alone, Run 6 simulated project duration using a combination of both historical data and information from Project B. This resulted in a simulated duration of 1331 hours (or 166 days), which is more accurate than Run 1. Notably, the model is not calibrated to mirror specific projects; rather, it was developed as a means of incorporating real-time information into tunneling simulation. Indeed, Runs 4 and 5, with information from 300 and 400m of actual data, were able to most reliably predict actual project duration, at 1179 and 1146 hours, respectively. Altogether, the results demonstrate that the updating process was able to incorporate real-time data in a manner that resulted in marked improvements in duration prediction accuracy.

4 CONCLUSIONS

The present study was focused on improving simulation-based prediction of tunneling project durations by incorporating new, periodically-generated project data acquired during project execution. This research focused on predicting the physical attribute of the environment that is interacting with the work face—specifically, ground conditions. The approach included the use of HMMs to model the stochasticity of ground conditions. Parameters for the HMM were updated in a real-time manner using Baum-Welch algorithm-based updating techniques. Simulation and parameter updating were achieved in an automated fashion using an improved *Simphony.NET* environment.

The feasibility and applicability of the proposed approach was demonstrated through a case study of a real tunneling project conducted in Edmonton, Canada. Application of the proposed approached improved accuracy of project duration predictions and was found capable of producing results that were similar to the actual project. While additional, real cases must be simulated to ensure the validity of the proposed method, and methods for determining prior information should be systematically developed, preliminary results indicate that the proposed approach has the potential to improve planning and delivery of tunneling projects by enhancing the reliability of ground condition predictions.

ACKNOWLEDGMENTS

This work was supported by a Collaborative Research and Development Grant (CRDPJ 492657) from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

AbouRizk, S., S. Hague, and R. Ekyalimpa. 2016. *Construction Simulation: An Introduction Using Simphony*. 1st ed. Edmonton, Canada: University of Alberta.

Baum, L. E. and T. Petrie. 1966. "Statistical Inference for Probabilistic Functions of Finite State Markov Chains." *The Annals of Mathematical Statistics*, 37(6): 1554-1563.

Baum, L. E. and G. R. Sell. 1968. "Growth Transformations for Functions on Manifolds." *Pacific Journal of Mathematics*, 27(2): 211-227.

Baum, L. E., T. Petrie, G. Soules, and N. Weiss. 1970. "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains." *The Annals of Mathematical Statistics*, 41(1): 164-171.

Chan, M.H. 1981. *A Geological Prediction And Updating Model in Tunneling*. Master's thesis, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts. http://hdl.handle.net/1721.1/15886.

Chung, T. H., Y. Mohamed, and S. AbouRizk. 2006. "Bayesian Updating Application into Simulation in the North Edmonton Sanitary Trunk Tunnel Project." *Journal of Construction Engineering and Management*, 132(8): 882-894.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society*, 39(1): 1-38.

Ioannou, P.G. 1987. Geologic Prediction Model for Tunneling. *Journal of Construction Engineering and Management*, 113(4): 569–590.

Ji, W. and S. AbouRizk. 2017. "Credible Interval Estimation for Fraction Nonconforming: Analytical and Numerical Solutions." *Automation in Construction*, 83: 56-67.

Leu, S. S. and T. J. W. Adi. 2011. "Probabilistic Prediction of Tunnel Geology using a Hybrid Neural-HMM." *Engineering Applications of Artificial Intelligence*, 24(4): 658-665.

Levinson, S. E., L. R. Rabiner, and M. M. Sondhi. 1983. "An Introduction to the Application of the Theory of Probabilistic functions of a Markov Process to Automatic Speech Recognition." *The Bell System Technical Journal*, 62(4): 1035-1074.

Liu, D.H., Y. Q. Zhou, S. Wang, and Y.L. Zhang. 2009. "Stochastic Simulation and Risk Analysis of Water Tunnel TBM Construction Scheduling Based on Geologic Prediction Using Markov Process." *Journal of System Simulation*, 21(2): 558–562.

Rabiner, L. R. 1989. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." *Proceedings of the IEEE*, 77(2): 257-286.

Rabiner, L. R. and B. H. Juang. 1986. An Introduction to Hidden Markov Models. *IEEE ASSP Magazine*, 4-16.

Rabiner, L. and B. H. Juang. 1993. *Fundamentals of Speech Recognition*. 1st ed. Upper Saddle River, New Jersey: Prentice Hall.

Stratonovich, R. L. 1960. "Conditional Markov Processes." *Theory of Probability and its Applications*, 5(2): 156-178.

Sutanto, A. 1997. *Geologic Prediction Model With Real Time Bayesian Series Analysis*. Master's thesis, Department of Construction Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan. http://etheses.lib.ntust.edu.tw/cgi-bin/gs32/gsweb.cgi?o=dstdcdr&s=id=%22G00M9405808%22.&searchmode=basic.

Zhang, L., R. Ekyalimpa, S. Hague, M. Werner, and S. AbouRizk. 2015. "Updating Geological Conditions using Bayes Theorem and Markov Chain." In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz et al. ,3367-3378, Piscataway, New Jersey: IEEE.

AUTHOR BIOGRAPHIES

MICHAEL WERNER recently received his MSc in Construction Engineering and Management from the University of Alberta. He is interested in developing methods for improved tunneling simulation and project delivery. His e-mail address is mwerner@ualberta.ca.

WENYING JI received his PhD in Construction Engineering and Management from the University of Alberta. A postdoctoral fellow at the University of Alberta until July 2018, Dr. Ji will transition into the role of Assistant Professor at George Mason University in August 2018. Dr. Ji is an interdisciplinary scholar focused on the integration of advanced data analytics, complex system simulation, and construction management to enhance decision-support processes of construction operations. His e-mail address is wji2@gmu.edu.

SIMAAN ABOURIZK is a Distinguished University Professor and Tier 1 Canada Research Chair in Operations Simulation in the Department of Civil and Environmental Engineering at the University of Alberta. His research efforts have focused on advancing simulation modeling and analysis of construction processes for improved construction planning, productivity improvement, constructability review, and risk analysis. His e-mail address is abourizk@ualberta.ca.