

SIMULATION INPUT UPDATING USING BAYESIAN TECHNIQUES

Tae Hwan Chung
Yasser Mohamed
Simaan AbouRizk

Department of Civil & Environmental Engineering
220 Civil and Electrical Building
University of Alberta
Edmonton, AB, T6G 2G7, CANADA

ABSTRACT

Simulation built on assumption and approximation has been traditionally utilized to make predictions prior to construction. Although there are many benefits of simulation such as its capability of multiple experiments with various scenario assumptions, it may lead to erroneous predictions when simulation input data are not accurate. Long-term repetitive projects such as tunnel construction provide opportunities to fine-tune the simulation input parameters based on real project progress. Bayesian updating techniques represent a very effective approach to enhance the quality of the estimates based on the observed data. This paper outlines some benefits that can be achieved using Bayesian updating techniques. The major benefits of these techniques includes more accurate simulation outcome even at the early stage of the project.

1 INTRODUCTION

Simulation can give many benefits to repetitive projects such as tunnel construction. Multiple simulation experimentations will enable tunnel project planners to effectively plan the schedule and cost before construction.

In order to achieve the successful application of simulation into the actual projects, accurate input data should be used for the simulation model. Due to uniqueness of each construction project, it is hard to accumulate good quality input data. In many cases, subjective inputs from industry experts are mainly used for simulation inputs. Consequently, these subjective inputs may produce inaccurate simulation results leading to erroneous schedule and cost prediction of the project.

Bayesian updating techniques can produce a balanced estimation by combining subjective data and observed data. Based on the small number of data sets collected even at the early stage of project lifetime, the techniques can produce better simulation results by considerably enhancing the quality of subjective input data. In this paper, Bayesian updating techniques are demonstrated through a

case study of an actual tunnel project, North Edmonton Sanitary Trunk (NEST) project.

This paper demonstrates how Bayesian techniques can be used to update the distributions of the input parameters as the project progresses. The use of these techniques shows that the quality of projection for the simulation model can be considerably improved.

2 BACKGROUND ON BAYESIAN TECHNIQUES

If the value of a parameter is assumed to be continuous with underlying probability density function (PDF), the prior assumptions on the parameter can be formally updated through Bayes' theorem as experimental results become available. After the prior distribution $f'(\theta)$ is revised in light of the experimental results, the posterior distribution $f''(\theta)$ is expressed as follows (Ang and Tang 1975):

$$f''(\theta) = kL(\theta)f'(\theta) \quad (1)$$

where θ = the random variable for the parameter of a distribution; the normalizing constant $k = \left[\int_{-\infty}^{\infty} L(\theta)f'(\theta)d\theta \right]^{-1}$; and $L(\theta)$ = the likelihood of observing the experimental outcome assuming a given θ .

If a prior distribution is a conjugate of the distribution of the underlying random variable, a posterior distribution can be conveniently obtained as the same mathematical form as the prior (Ang and Tang 1975). An example of such derivation is demonstrated for normal distribution as follows.

For the normal population with known standard deviation σ , the likelihood function for the parameter μ , is

$$L(\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2 \right] = \prod_{i=1}^n N_{\mu}(x_i, \sigma) \quad (2)$$

where $N_{\mu}(x_i, \sigma)$ denotes the density function of μ with mean value x_i and standard deviation σ . According to Ang and Tang (1975) the product of m normal density functions with respective means μ_i and standard deviation σ_i is also a normal density function with mean μ^* and variance σ^* that can be calculated as in Eq. (3).

$$\mu^* = \frac{\sum_{i=1}^m (\mu_i / \sigma_i^2)}{\sum_{i=1}^m 1 / \sigma_i^2} \quad \text{and} \quad (\sigma^*)^2 = \frac{1}{\sum_{i=1}^m 1 / \sigma_i^2} \quad (3)$$

The likelihood function $L(\mu)$ thus becomes

$$L(\mu) = N_{\mu} \left(\frac{\sum_{i=1}^n x_i / \sigma^2}{\sum_{i=1}^n 1 / \sigma^2}, \frac{1}{\sum_{i=1}^n 1 / \sigma^2} \right) = N_{\mu} \left(\frac{(1/\sigma^2) \sum_{i=1}^n x_i}{n/\sigma^2}, \frac{1}{\sum_{i=1}^n 1 / \sigma^2} \right) = N_{\mu} \left(\bar{x}, \frac{\sigma}{\sqrt{n}} \right) \quad (4)$$

where \bar{x} is the sample mean.

If $f'(\mu)$ is assumed to be $N(\mu', \sigma')$ with likelihood function of the equation (4), the posterior distribution of μ becomes

$$f''(\mu) = kL(\mu)f'(\mu) = kN_{\mu} \left(\bar{x}, \frac{\sigma}{\sqrt{n}} \right) N_{\mu}(\mu', \sigma') \quad (5)$$

which is a product of two normal density functions. The obtained posterior distribution is also a normal distribution and has the following parameters:

Mean,

$$\mu^* = \left[\frac{1}{(\sigma')^2} \right] \mu' + \left[\frac{n}{\sigma^2} \right] \bar{x} = \frac{\bar{x}(\sigma')^2 + \mu'(\sigma^2/n)}{(\sigma')^2 + (\sigma^2/n)} \quad (6)$$

Standard Deviation,

$$\sigma^* = \sqrt{\frac{1}{\frac{1}{(\sigma')^2} + \frac{n}{\sigma^2}}} = \sqrt{\frac{(\sigma')^2 (\sigma^2/n)}{(\sigma')^2 + (\sigma^2/n)}} \quad (7)$$

(Ang and Tang 1975)

Based on the updating techniques described in this section, subjective input assumption in a simulation model due to lack of data can be improved once sets of actual data become available as the project progresses. The next section discusses the description of actual project, NEST, as a case study and the overview of simulation models used in this paper.

3 OVERVIEW OF NEST PROJECT AND SIMPHONY SIMULATION MODEL

The existing sanitary sewer system in the north area of Edmonton has reached the limit of its capacity. The construction of five major sanitary trunks from the developing areas to the Gold Bar Wastewater Treatment Plant and the Capital Region Sewage Treatment Plant was recommended from various studies. In 1994, the overall plan for North Edmonton Sanitary Trunk (NEST) was developed as the first recommended trunk. The purpose of this plan was to increase the capacity of the existing sewage system and allow continued growth in North Edmonton.

The first section of NEST (NC1) was proposed as the first stage construction of NEST. The proposed NC1 section of NEST is served as a temporary outlet, which will be used for storage during the wet weather flow and for conveyance during the dry weather flow. The NC1 section of NEST is a 1538-meter tunnel having a 2.94-meter finished diameter lined with pre-cast concrete segments (Report 2: NEST design report 2002).

A detailed simulation analysis at pre-construction stage had been conducted in order to improve the productivity estimates for the NEST project. The tunnel was modeled using a tunneling template, one of the special purpose simulation (SPS) tools provided within the Symphony platform. This template consists of 16 modeling elements through which tunneling activities in various stages can be defined (TBM Tunnel Simulation Template User's Guide 2000).

Figure 1 shows the major components of the Symphony model for the NC1 section of NEST. A base model consists of an undercut, soil segments, and a removal shaft. The soil segments are used to model the changes in soil conditions and tunnel geometry. Every soil segment has an element for modeling the TBM excavation and lining processes in addition to an element for modeling surveying activities. Users need to enter basic input data such as the length, soil type, and TBM penetration rate of each section.

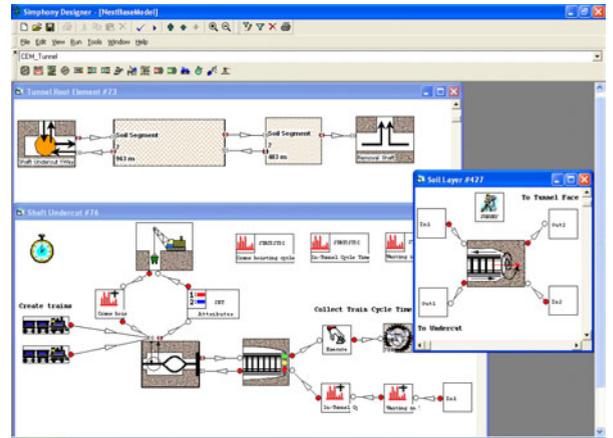


Figure 1: Components of the Symphony Model for the NC1 section of NEST

For simulation analysis at pre-construction stage, some input data were obtained from assumptions based on industry expert’s opinion. These subjective inputs can be served as a good starting reference but may yield inaccurate simulation results.

4 BAYESIAN UPDATING PROCESS OF MODEL INPUTS

4.1 Data Collection

Actual project data such as soil types and TBM penetration rates were recorded during the construction of this project. For the collection of TBM penetration rates, a custom designed device was anchored to the conveyor traveling on the segmental liner behind the TBM to monitor and record the actual advancement of the TBM.

4.2 Comparisons of Distributions between Assumption and Actual Sample

Statistical distributions of major input parameters such as TBM penetration rates, time between TBM breakdowns, time to repair the TBM, and survey time were fit by using commercial distribution fitting software. These parameters, which were available from the data collection process or the project progress report, were believed to play important roles in the simulation modeling and affect the overall productivity.

The distributions fit based on the sample data were compared with the ones used for the original simulation model. The comparisons of these two distributions of each parameter show some discrepancies. For the TBM penetration rate, the original model assumed a uniform distribution with a mean of 3 m/hour while the actual distribution turned out to be normally distributed with a mean of 5.04 m/hour. For both “Time to Repair the TBM” and “Survey Time”, triangular distributions with lower means were better fit than uniform distributions assumed for the original model. As the tunneling construction proceeds, how input parameters are effectively updated with observed data and used for the simulation model is of great interest.

4.3 Bayesian Updating Process

For the input parameter updates for simulation, a systematic approach is required to combine original assumption with actual sample data. The Bayesian techniques can be a useful methodology to update these parameters. The information to be updated can be either objective data based on the sample or subjective judgments from the experts’ opinions.

For the purpose of illustrating how these techniques can be applied to the simulation input updates, one of the major input parameters, TBM penetration rates, was selected. Two soil segment elements for simulation model were used for the different time frames. One segment is

used for the sections completed at a specific day during the construction process while the other is used for the remaining sections. For instance, on November 15, 2001, the first soil segment represents the completed 963-meter-long section while the second segment represents the remaining 483-meter-long section as shown in Figure 2. The actual information can be used for the first modeling segment while the updated information based on Bayesian techniques can be used for the second modeling segment.

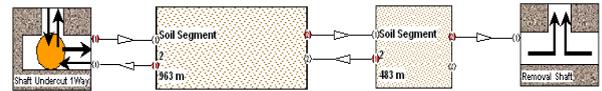


Figure 2: Simulation Modeling for NEST

The simulation input parameters are updated twice a month, in the middle and end of each month. Since actual TBM penetration rates for some time periods are missing, updating cannot be done during those periods and updating dates each month may slightly vary. By using Bayesian updating, prior information for TBM penetration rate is to be updated as the tunnel construction proceeds. The updated information is called posterior information.

Figure 3 shows the actual overall productivity and the progress chart for NEST project. The learning curve effect is supported with a finding that the lower productivity was recorded at the early stage of the project and the overall productivity becomes stable from August 31, 2001 when about 19 percent of the entire tunnel section was completed.

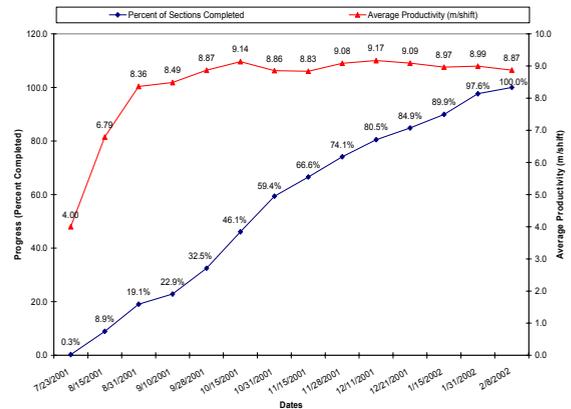


Figure 3: Percent Completion and Average Productivity for NEST Project

Dates appeared in this chart indicate the overall milestone for updating the distributions for simulation input parameters. It should be noted that there were some periods when valid sample TBM penetration rates were not available and therefore updating was not done during that period. The first updating was done on August 15, 2001 when about 9 percent of tunnel sections were completed. There were 71 sample data for TBM penetration rates and commercial software was used to fit the distribution. The result shows

that normal distribution was one of the best distributions suited for the TBM penetration rates.

4.4 Updating TBM Penetration Rates

Since the original model assumed a uniform distribution for the TBM penetration rate ranging from 2 to 4 m/hour, it needs to be transformed into normal distribution for prior information in order to mathematically simplify the Bayesian updating process. It is assumed that the mean TBM penetration rate lies between 2 and 4 m/hour with a 99 % confidence interval. The mean of this distribution is 3 m/hour, and the variance can be calculated by using a standard normal distribution table and the confidence interval. Since the corresponding Z-value is 2.58, the standard deviation σ of this normal distribution can be calculated as follows:

$$\frac{4 - \mu}{\sigma} = Z$$

$$\frac{4 - 3}{\sigma} = 2.58$$

$$\sigma = 0.39$$

Figure 4 shows the normal distribution with a mean of 3 m/hr and a standard deviation of 0.39, which was transformed from the uniform distribution ranging from 2 to 4 m/hr. Since the prior distribution is assumed to be a normal distribution, the posterior distribution obtained from Bayesian updating can be the same mathematical form as the prior. The formulas for calculating posterior distribution parameters are obtained from the following equations:

Prior information (subjective): $\mu' = 3.0$ and $\sigma' = 0.39$
 Likelihood function (sample on August 15, 2001): $\bar{x} = 4.74$, $\sigma = 1.37$, and $n = 71$

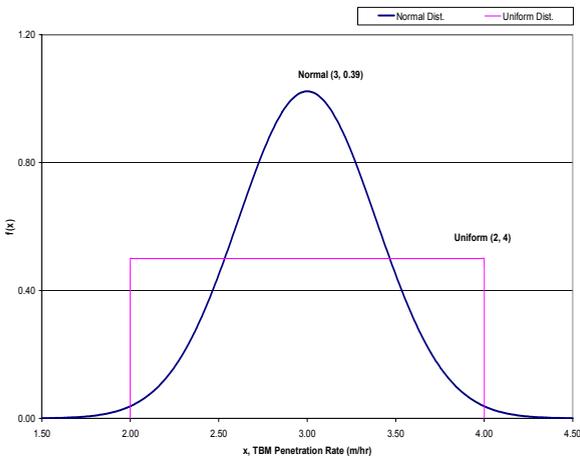


Figure 4: Transforming Uniform into Normal Distribution for TBM Penetration Rate

Using equations,

$$\mu'' = \frac{\bar{x}(\sigma')^2 + \mu'(\sigma^2/n)}{(\sigma')^2 + (\sigma^2/n)} = \frac{4.74 \times 0.39^2 + 3.0 \times (1.37^2/71)}{0.39^2 + (1.37^2/71)} = 4.48$$

$$\sigma'' = \sqrt{\frac{(\sigma')^2 (\sigma^2/n)}{(\sigma')^2 + (\sigma^2/n)}} = \sqrt{\frac{0.39^2 \times (1.37^2/71)}{0.39^2 + (1.37^2/71)}} = 0.15$$

Therefore, the posterior distribution updated on August 15, 2001 is the normal distribution with a mean of 4.48 and a standard deviation of 0.15. Figure 5 shows the comparisons of prior, likelihood function, and posterior distribution updated on August 15, 2001.

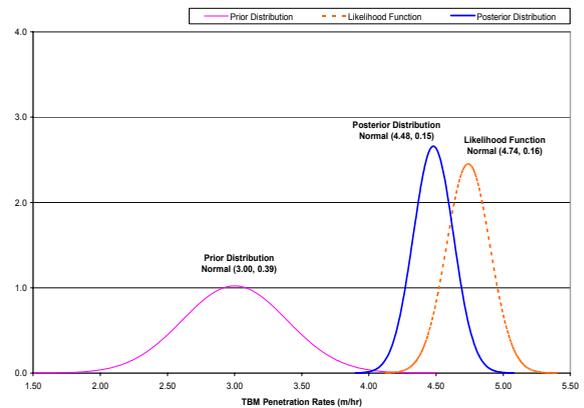


Figure 5: Updating TBM Penetration Rates on August 15, 2001

5 SIMULATIONS WITH UPDATED PARAMETERS

Simulations were conducted on the specified time intervals and the simulation results were compared with actual mean TBM penetration rates and productivities as shown in Figure 6. The comparisons of the actual and updated mean

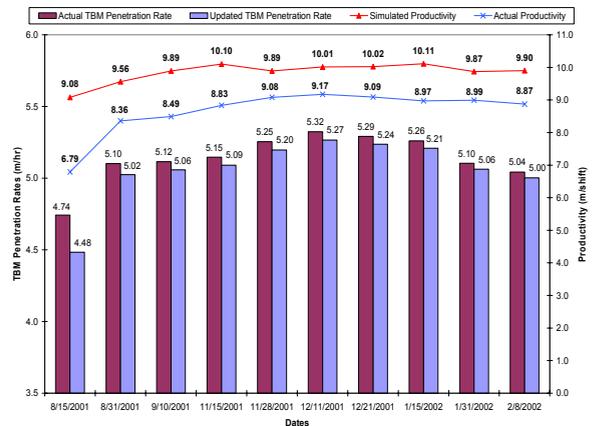


Figure 6: Comparisons of TBM Penetration Rates and Productivities between Actual and Simulation Results

TBM penetration rates show that the differences between two values tend to decrease as the time passes by. The updated average TBM penetration rate at the early stage of the project (on August 15, 2001) was 5.5% less than the actual ones. But, the rate was reduced to 0.8% at the end of the project (on February 8, 2002).

The results show that the productivity predicted from simulation tends to be somewhat higher than the actual one. These results are probably due to some factors affecting productivity such as rock drilling since these modeling elements are not included in the current tunnel simulation template and were not modeled for this project.

5.1 Simulation Results for the Remaining Sections

Figure 7 compares the actual and simulation productivity for the remaining sections. For simulation, TBM penetration rates were updated by Bayesian techniques. Two trend-lines on time intervals compare the remaining durations at the actual project with the ones predicted from the simulations. When the simulation was conducted before the project, input parameters for simulation were assumed without any update. This initial simulation experiment predicts the total duration of 177 shifts with the average productivity of 8.17 m/shift while the actual duration was 163 shifts with the average productivity of 8.87 m/shift. On August 15, 2001, the first simulation with the updated input parameters predicts 145 shifts for the remaining section. This predicted result is very close to the actual duration of 144 shifts. Thus, the difference between actual and simulation duration for the remaining section was considerably reduced from the initial simulation conducted before the project. The comparisons of two trend-lines on later dates do not show any significant deviation between these two results except for the 9-shift difference on September 10, 2001. Simulations conducted during construction predict better project durations than the initial one conducted before the project. These results indicate that Bayesian techniques were successfully applied to update the distribution of the input parameter.

Another interesting fact was also found that updating the input parameters even at the very early stage of the project can provide good simulation results. Based on the average productivity obtained from simulation, the progress charts were updated and the results were compared with the actual progress. These results were obtained from three simulations: on early assumption (pre-construction), August 15, and August 31. The calculation of percent complete was based on the average productivity obtained from simulation. Progress charts on August 15 and August 31 were continuously updated from the original one. The comparisons of these progress charts indicate that updated progress charts on both August 15 and August 31 become very close to actual one. This result leads to the belief that even early updates during construction can significantly improve the prediction of the project performance by eliminating the uncertainty contained in the original assumption.

6 CONCLUSION

For the long-term repetitive projects such as tunnel construction, simulation can be utilized as a powerful tool to experiment with multiple scenarios instead of actual costly experimentation in the field. Since simulation has been traditionally used to make predictions prior to construction, some input data are mainly obtained from assumptions rather than actual data. Thus, simulation results may lead to erroneous predictions for the project when inaccurate input data are assumed for the simulation model. A proper updating process for simulation during construction can improve the simulation prediction by reducing uncertainties. The renewed predictions can be served to improve overall project control such as schedule and cost.

This paper presents Bayesian techniques to update the distributions of input parameters for the tunnel simulation. One of the tunnel projects recently completed by the City of Edmonton Public Works Department, NEST, was used as a case study for this research study. Bayesian techniques show a formal approach to combine original assumption with the obtained sample data as the project progresses. TBM penetration rates were selected as an example to show how the distribution for the simulation input parameter is updated. The simulation results show that even early updates during construction can significantly improve the prediction of the project performance by eliminating the uncertainty contained in the original assumption. In this project, it is determined that the earliest time to properly update the distribution for TBM penetration rates is when approximately 9 percent data of the total tunnel section is completed and more than 50 sample are gathered. These results can be used as a guideline for similar tunnel projects when simulation is applicable and proper simulation updates are required.

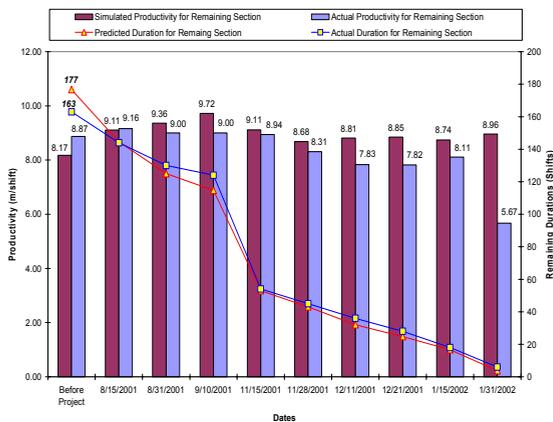


Figure 7: Comparisons of Productivity and Durations for the Remaining Sections

REFERENCES

- Ang, A. H-S., and Tang, W. H. (1975). Probability Concepts in Engineering Planning and Design. Vol. 1, John Wiley & Sons, New York, N.Y.
- Report 2: The City of Edmonton, Asset Management and Public Works, Drainage Services Design and Construction (2002). "North Edmonton Sanitary Trunk (NEST) Stage NC1." Design Report.
- TBM Tunnel Simulation Template User's Guide (2000), NSERC/Alberta Construction Industry Research Chair

AUTHOR BIOGRAPHIES

TAE HWAN CHUNG is a PhD student in the department of Civil Engineering at the University of Alberta. He received his MSE in Civil Engineering from Purdue University in 2000. His research interests focus on Bayesian updating application into simulation.

YASSER MOHAMED is a research and technology transfer engineer in the department of Civil Engineering at the University of Alberta. He received his BSCE and MSCE in Civil Engineering from Zagazig University, Egypt, in 1990 and 1996. He received his Ph.D. degree from University of Alberta in 2002. His research interests are focused on decision support of construction management using simulation and AI techniques. His e-mail address is <yaly@ualberta.ca>.

SIMAAN ABOURIZK is a professor in the Department of Civil Engineering at the University of Alberta. He received his BSCE and MSCE in Civil Engineering from Georgia Institute of Technology in 1984 and 1985, respectively. He received his Ph.D. degree from Purdue University in 1990. His research interests focus on the application of computer methods and simulation techniques to the management of construction projects. His e-mail address is <abourizk@ualberta.ca>.