Proceedings of International Structural Engineering and Construction
Holistic Overview of Structural Design and Construction
Edited by Vacanas, Y., Danezis, C., Singh, A., and Yazdani, S.

ISSN: 2644-108X

QUANTIFYING UNCERTAINTY IN SIMULATION MODELING

JACE THIBAULT and SIMAAN ABOURIZK

Dept of Civil and Environmental Engineering, University of Alberta, Edmonton, Canada

Uncertainty can be defined as a state of either incomplete or otherwise bounded knowledge. Simulation models, and the engineering systems that they represent, often contain various types of uncertainty. Different approaches and theories can be applied to model these various types of uncertainty with a range of degrees in difficulty and accuracy. The objective of this paper is to explain the various types of uncertainty found in simulation models and to examine where uncertainty can be better represented or potentially reduced. To achieve this objective, a Monte Carlo Simulation model called the As-Planned Model is developed to estimate both cost and schedule using a risk-based approach for a simplified, Light Rail Transit construction project. After the project is complete, the As-Planned model is then compared to the project's actual results. The resulting conclusions about various types of uncertainty are derived through both output comparison as well as uncertainty analysis.

Keywords: Monte Carlo, Cost, Schedule, Risk, Estimating.

1 INTRODUCTION AND LITERATURE REVIEW

This paper investigates how various types of uncertainty are quantified in simulation models and identifies which types of uncertainty can be reduced. Different approaches and theories can be used to represent uncertainty with varying efficacy. When representing an engineering system using simulation, various types of uncertainties are propagated through the model and cumulatively affect results. Uncertainty itself has been widely discussed in systems literature. In engineering systems literature specifically, types of uncertainty and the various theories available to represent them have been identified. Ayyub (2003) has classified the main sources of information uncertainty in engineering systems as ambiguity, likelihood, approximations, and inconsistency. Methods for estimating model uncertainty include common approaches, such as the Taylor Series Method and Monte Carlo Simulation (Brown and Heuvelink 2006), as well as novel approaches such as machine learning techniques (Solomatine and Shrestha 2009). For the purposes of uncertainty analysis, Monte Carlo Simulation is preferred because it can be applied more generally and requires fewer assumptions and user inputs (Brown and Heuvelink 2006). Brown and Heuvelink (2006) define uncertainty propagation as a scenario where uncertainties in input data and models lead to uncertainties in the model output, and they indicate that Monte Carlo Simulation is a very useful method for approaching this problem. The Monte Carlo Simulation method randomly samples from the joint distribution of possible inputs and models uncertainty, generating a set of realizations of the system as model output. Statistics generated from the sample set describe the extent of uncertainty in the simulation results. This paper combines, in a Monte Carlo Simulation model, various concepts discussed in the literature, including the representation of uncertainty and its potential reduction.

Identifying different sources of uncertainty, strategically reducing cognitive sources of uncertainty by acquiring knowledge, and accounting for non-cognitive sources of uncertainty will allow engineers to optimize their models and better understand how uncertainty propagates within them. *Simphony.NET* is used to create a model that integrates risk with cost and with schedule uncertainty to estimate costs and the schedule for a project (As-Planned model). After completing the project, the model is updated to reflect the actual data accumulated (As-Built model). Results of the As-Planned and As-Built model are then compared to investigate the role of uncertainty in the simulation and determine where improvements can be made. The illustrative example is a theoretical, simplified Light Rail Transit (LRT) capital construction project. With sufficient data, the methodology could be applied in a case study to any completed construction project. Verification and validation of the model are also discussed.

2 METHODOLOGY

The overall approach is intended for construction projects with definable tasks (work packages) and is summarized in Figure 1.

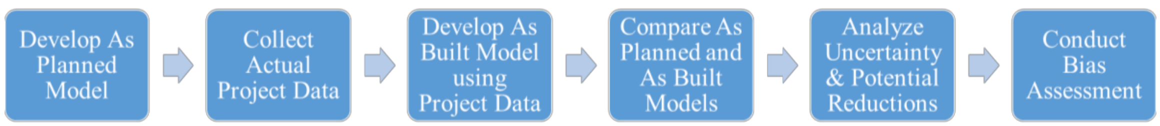

Figure 1. Summary of the approach.

To develop the As-Planned model, the project work packages and their anticipated relationships are required. Additionally, estimated duration and costs of tasks are required. Estimates can be constant values or represented using a distribution if uncertainty exists. Risks that affect the project cost or schedule need identification and quantification. Particularly, a description of the risk, its likelihood, the impact, and the relationships to work packages (if any) are necessary. Risks can be divided into two broad categories: risks with potential schedule impacts and risks that do not impact schedule. Uncertain risks should be represented using distributions. Distributions and their parameters should be selected on their ability to represent uncertainty as accurately as possible since these selections can significantly affect model results (AbouRizk 2013). Uncertainty about whether a selected distribution and parameters accurately represent uncertainty is a form of meta-uncertainty caused by approximation; Section 3 discusses this in more detail. Estimated escalation rates (for both hard and soft costs), as well as profit and overhead for the project as a whole, are required. Estimates of profit and overhead can be a fixed percentage of construction costs or represented using a distribution if uncertainty exists. For hard and soft cost escalation, the percentage input can be a compounded percentage applied over the entire project term to simplify the model. The simulation model is shown at a summary level in Figure 2. Each work package for the project is represented using a Composite element, which contains the modelling elements required to track time and collect costs for the work package.

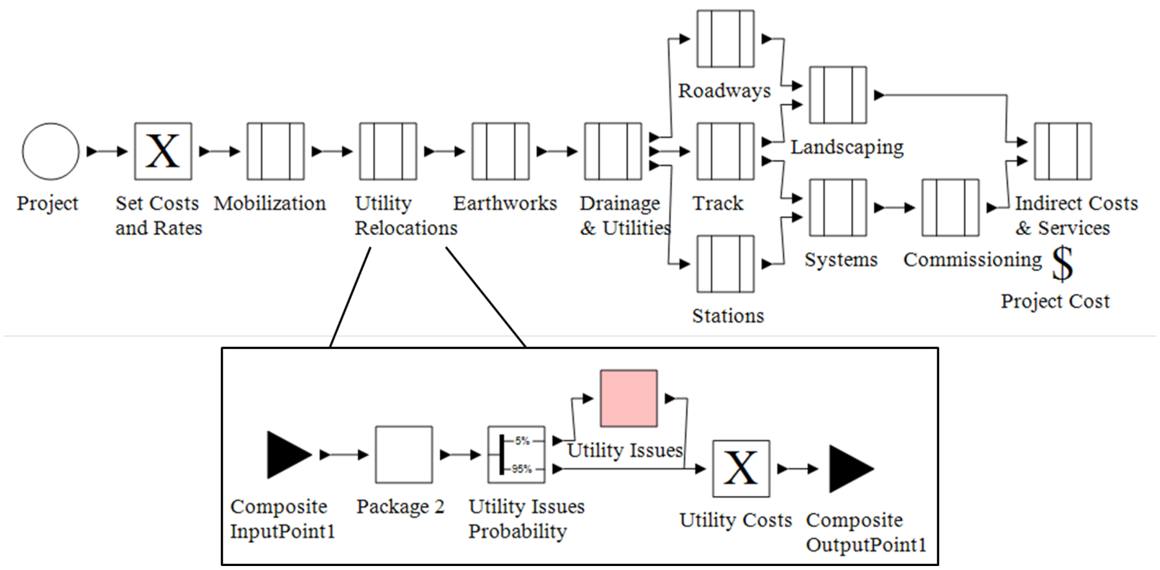

Figure 2. A view of the As-Planned Model in *Simphony.NET*.

Task elements contained within Composite elements represent the completion of work packages (shown in white). The consequences of realized schedule risks (in red) are shown by holding the entity for a period of time. The simulation includes several Execute elements, which run user-written code when an entity passes through the Execute element. For this simulation, various code has been written within Execute elements to set global variables, track the time required to complete work packages while accounting for risk events, collect costs, and account for non-schedule risks. The Probabilistic Branch element is used to model risks realization by routing entities based on the probabilities assigned to the element. The Cost element simply acts as a repository of costs collected by the Execute elements. This simulation performs 10,000 iterations to ensure uncertain aspects of the model are sampled sufficiently.

To develop the As-Built Model, actual project data are required, including relationships, durations, costs for each work package, realized risks and their impacts, and relationships to work packages (if any). The actual escalation rates, as well as profit and overhead, are also required. Two key outputs are generated by the As-Planned and As-Built models: the project duration and detailed cost report. For the As-Planned model, outputs include the mean, standard deviation, minimum, and maximum of the project duration and costs. For the As-Built model, the project duration and costs are deterministic, as the output is known. With the model outputs, one can compare the data, and improve the understanding of uncertainty in the model.

3 ILLUSTRATIVE EXAMPLE

The illustrative example is a theoretical, simplified LRT construction project. Using the simulation approach described in Section 2, the As-Planned model indicates an estimated mean duration of 570 days with a standard deviation of 43 days. Over 10,000 iterations, the minimum and maximum durations were 447 and 764 days, respectively. The As-Planned model indicates an estimated mean overall project cost of $155M with a standard deviation of $13M. The distribution of overall cost from the As-Planned model is shown in Figure 3.

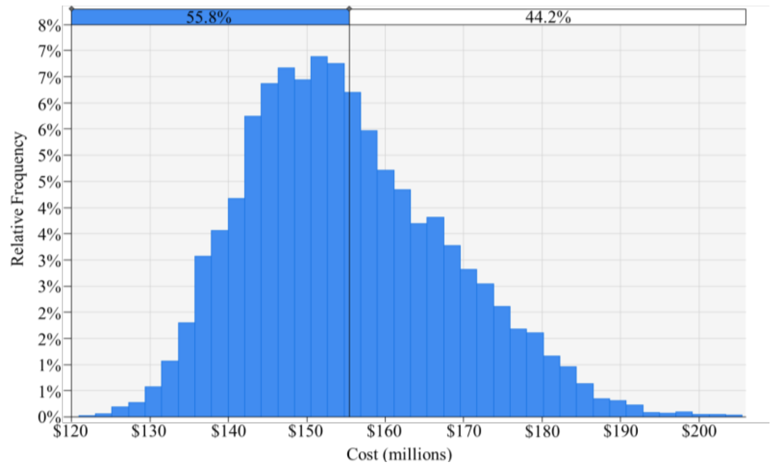

Figure 3. Overall cost distribution generated from the As-Planned model.

The As-Built model indicates an actual duration of 626 days and cost of $155M. In the illustrative example, the actual duration and mean-estimated duration differed by 8.9% while the actual and mean-estimated cost differed by less than 0.1%. The detailed costs from the As-Planned and As-Built models can be compared at the package-level to highlight items where uncertainty substantially contributed to deviations between estimated and actual outcomes. To address deviations and optimize the modelling approach, uncertainty should be identified and reduced with uncertainty analysis. For the illustrative example, uncertainties are identified and categorized according to the literature (Ayyub 2003).

Ambiguity arises from incompletely or incorrectly identifying possible outcomes (Ayyub 2003). In this illustrative example, the degree of ambiguity was relatively low: a number of risks were identified during development of the As-Planned model, and only a subset of these was realized in the As-Built model. Projects with a high degree of ambiguity can encounter the opposite scenario: risks are not identified or adequately modelled but then are realized. To reduce ambiguity, substantial effort is required to sufficiently account for relevant possible outcomes. Cost-benefit analysis is recommended to compare the cost of risk identification and management.

Likelihood is related to non-cognitive sources of uncertainty such as physical randomness and statistical uncertainty caused by sampling (Ayyub 2003). For the illustrative example, physical randomness was broadly accounted for in the uncertainty of work package durations. For example, weather and geotechnical conditions were factors in determining pessimistic and optimistic durations for work package tasks. To reduce likelihood uncertainty, physical characteristics including historical weather data and geotechnical investigations can be assessed and modelled using a variety of approaches. Modelling physical characteristics with a higher degree of certainty also may reduce uncertainty in durations for work package tasks. To reduce statistical uncertainty caused by sampling, a sufficient number of samples should be collected using Monte Carlo Simulation. Byrne (2013) indicates that a confidence interval width of 0.01 can be achieved assuming a 95% confidence level using as many as 9,604 iterations.

Simplifications and assumptions can reduce the difficulty of modelling a system and are sometimes necessary to effectively model a complex process. Assumptions and simplifications should be deliberately made and identified. Cost-benefit analysis is recommended to compare the cost of avoiding simplifications and assumptions. For the illustrative example, an assumption was that soft cost escalation applied to a single work package. The assumption was found to be valid in the As-Built model, but the uncertainty in that assumption was not accounted for in the As-Planned model. Statistical distribution could be used to account for uncertainties related to simplifications/assumptions. Sources of vagueness include parameter definition, biases, and comprehension of complex systems (Ayyub 2003). Where parameters are defined linguistically, fuzzy set theory can reduce uncertainty (Naderi 2008). Qualitative verbal expressions for likelihood can also be represented using statistical distributions (AbouRizk 2013). If linguistic parameters such as risk likelihood are defined and accounted for using statistical distributions or fuzzy set theory, uncertainty can be more accurately represented. Where subjectivity is required, a variety of individual perspectives should be captured. Residual bias can be captured by conducting a bias assessment. Inconsistent information stems from both human and organizational errors (Ayyub 2003). Where uncertainty is associated with inconsistency, bias random variables can account for the resulting errors (Ayyub 2003). Remaining uncertainty can be collected in a bias assessment. Non-abstracted aspects of a system are the aspects not included in a model (Ayyub 2003). While abstracted aspects can account for some non-abstracted aspects, some non-abstracted uncertainty remains. Unknown aspects of a system can result from unawareness of something unknown. Bias assessment is based on two implicit assumptions: "(1) the value of the variable or parameter for the real system is known or can be accurately assessed from historical information or expert judgment; and (2) the state of knowledge about the real system is complete and reliable" (Ayyub 2003). For the illustrative example, the first assumption is valid since the actual duration and cost are known. As Ayyub (2003) states, the second assumption cannot generally be validated; however, the bias ratio remains a valuable indicator of uncertainty attributable to non-abstracted and unknown aspects of a system. For the illustrative example, the bias was found to be virtually non-existent for cost uncertainty (0.9991) and low for schedule uncertainty (0.9105).

4 VERIFICATION AND VALIDATION

Model verification ensures both the program and implementation is correct; model validation ensures the model has a satisfactory range of accuracy for its use (Sargent 2007). The As-Built model was verified by confirming that the outputs were identical to the system outputs. The similarly structured As-Planned model is verified by extension. Two validation tests were used. First, the bias ratio (Section 3) indicated that the cost and schedule predicted by the As-Planned model were similar to the As-Built model. Second, the quantile-quantile (Q-Q) plots (Figure 4) indicate that output parameters of the As-Planned model do not significantly deviate from normality. The model was verified and validated by an independent third party with knowledge of LRT construction and simulation modelling.

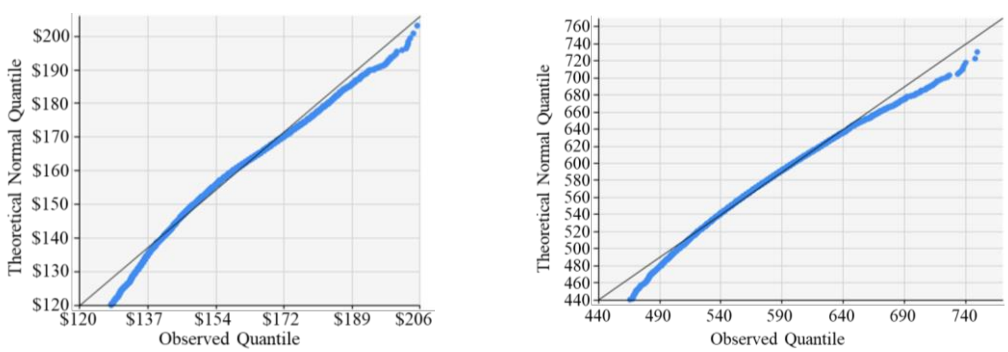

Figure 4. Q-Q plots comparing As-Planned model costs (left) and durations (right) to a normal distribution.

5 CONCLUSIONS

The purpose of this paper is to investigate uncertainty quantification in simulations and which types of uncertainty can be reduced in engineering systems. *Simphony.NET* was used to create an As-Planned model to estimate cost and schedule for a construction project using a risk-based approach. After completing the project, actuals were captured in an As-Built model. Results from the two models were compared; uncertainty analysis was used to identify where uncertainty could be reduced. Bias assessment was used to account for non-abstracted/unknown aspects of the system. A simplified LRT construction project served as an example. The various types of uncertainty were identified as ambiguity, likelihood, approximations, inconsistency, and non-abstracted/unknown aspects. Approaches for reducing uncertainty varied. Some types of uncertainty (ambiguity and likelihood) can be reduced by enhancing knowledge. Uncertainty related to inconsistency, or non-abstracted/unknown aspects, are best handled using bias assessment. There are opportunities to improve how uncertainty is accounted for in simulations. Additional research is required to understand the propagation of uncertainty throughout a model, particularly when different types of uncertainty are represented in a simulation model.

References

AbouRizk, H., *Understanding and Improving Input for Quantitative Risk Analysis in the Construction Industry*, M.Sc. Thesis, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada. https://doi.org/10.7939/R39C6SB73. 2013.

Ayyub, B. M., *Risk Analysis in Engineering and Economics*, Boca Raton: Chapman and Hall/CRC, 2003.

Brown, J. D. and Heuvelink, G.B., *Assessing Uncertainty Propagation through Physically Based Models of Soil Water Flow and Solute Transport*, Encyclopedia of Hydrological Sciences, Anderson, M. G. and McDonnell J. J., (eds.), 2006.

Byrne, M., *How Many Times Should a Stochastic Model Be Run? An Approach Based on Confidence Intervals*, International Conference on Cognitive Modelling (ICCM) 2013, Carleton University, Ottawa, Canada, July 14, 2013.

Naderi, M., *Fuzzy Logic Application in Risk Analysis of Construction Management*, M.Sc. Thesis, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada, 2008.

Sargent, R., *Verification and Validation of Simulation Models*, Proceedings of the 2007 Winter Simulation Conference, Henderson, S. G., Biller, B., Hsieh, M.-H., Shortle, J., Tew, J. D., and Barton, R. R., (eds.), Washington, DC, 124-137, December 9-12, 2007.

Solomatine, D. P. and Shrestha D. L., *A Novel Method to Estimate Model Uncertainty Using Machine Learning Techniques*, Water Resources Research 45(W00B11), January, 2009.