

Cloud Based Interactive Probabilistic Simulation for AEC Industry

Jing Du, Ph.D.¹, Rui Liu, Ph.D.² and Yilmaz Hatipkarasulu, Ph.D.³

¹ Department of Construction Science, The University of Texas at San Antonio, San Antonio, TX; PH (210) 458-3053; email: jing.du@utsa.edu

² Department of Construction Science, The University of Texas at San Antonio, San Antonio, TX; PH (210) 458-3054; email: rui.liu@utsa.edu

³ Department of Construction Science, The University of Texas at San Antonio, San Antonio, TX; PH (210) 458-3099; email: yilmaz.karasulu@utsa.edu

ABSTRACT

Probabilistic simulation is a popular approach to analyze uncertainties in AEC (Architecture, Engineering and Construction) Industry including Monte Carlo simulations of cost, PERT simulations, and probabilistic risk assessment. However, the application inefficiencies in the simulation process make it susceptible to problems of collaborativity, usability, and difficulties in interactive decision-making. This paper proposes a Cloud-based apparatus to improve the efficiency, usability, and collaborativity of probabilistic simulation in tackling a range of AEC problems using a Real-time Interactive Simulation Kit (RISK) tool. This tool may ultimately lead to interactive probabilistic simulation, which ensures a seamless integration between simulation and decision-making process. A simulation technology called “*Simulate As You Operate*” (SAYO) was developed to realize “*What You See Is What You Get*” (WYSIWYG) type of probabilistic simulation. The paper also includes two RISK applications to illustrate the tool; PERT simulation and risk analysis. RISK presents an effective analysis and communication tool for the uncertainty analysis of AEC problems.

INTRODUCTION

Construction projects are constantly faced with uncertainty, ambiguity, and variability. To address AEC (Architecture, Engineering and Construction) problems with uncertainties, the construction industry and the construction research community have a long history of applying probabilistic simulation, which includes Monte Carlo simulation of construction costs (Chou et al. 2009), project duration simulation (Lee et al. 2006), probabilistic risk assessment (Sadeghi et al. 2010), project portfolio investment decision (Smith 1994), resource allocation (Lu et al. 2008) and stochastic model of building simulation (Yi and Malkawi 2012). Compared to the deterministic approaches that were traditionally applied in the industry, probabilistic simulation is able to capture the uncertainties coherently, which means a full reflection of the probabilistic rules (Mooney 1997).

However, the process of probabilistic simulation has several application limitations. The first limitation is the inefficiency of probabilistic simulation process.

In a typical probabilistic simulation, the input uncertainties are propagated into uncertainties in output by a large number of repetitive simulations (Mooney 1997) referred to as trials. For each trial, the input variables are sampled from the specified probability density functions (PDFs), and yield just one result. The new trial occurs only when the previous one is completely over. The system is ultimately simulated through time such that the probability distributions of possible outcomes can be obtained by aggregating results of all trials. This process carries a significant computational overhead for problems with large number of variables. As the number of variables increases and computational speed decreases, real-time simulation and interactive decision making becomes extremely difficult. The second limitation is related to the platform of collaboration. Without a cross-functional protocol, different functional units in a construction project are likely to perform probabilistic simulations separately which makes it is difficult to exchange models across organizations or to ensure the authenticity of the models or model information received from another source. In addition, there are no proven methods to benchmark simulation results against other industry peers. The last limitation is the functionality of the available simulation software (e.g., @RISK and Crystal Ball). The set-up, modeling, and interpretation are rather time-consuming and require expertise in both probability theories and simulation. The quality of simulation is usually limited to the capacity of the analysts.

This paper proposes an approach called “*Simulate As You Operate*” (SAYO) which can realize “*What You See Is What You Get*” (WYSIWYG) type of probabilistic simulation. A Cloud based apparatus, Real-time Interactive Simulation Kit (RISK), was built upon SAYO to improve the efficiency, collaborativity and usability of probabilistic simulations in the AEC. The core of RISK is a Cloud based computing unit that performs the real-time computation, synthesizes the simulation results, and returns the results to users instantaneously via web browsers.

SAYO: SIMULATION AS YOU OPERATE

In general, probabilistic simulation time can be divided into four components:

- 1. Parameterization time (PT or pt_i):** The time spent by the user to parameterize the model, for example, defining the PDFs for model input;
- 2. Random number generation time (GT or gt_i):** The time spent by the system to generate or retrieve random numbers for the simulation according to the arbitrary distributions defined by the user;
- 3. Simulation time (ST or st_i):** The pure time spent by the system to perform the actual simulation tasks; and
- 4. Overhead (OH or ot_i):** The time spent for data fetching, processing operations and transferring, such as transferring data between central processing unit (CPU) and the main memory, between CPU and secondary storage (hard disk), off-line storage and tertiary storage (e.g. tape drives), and among different hierarchical levels of the memory system.

These four components can be executed in parallel to improve the efficiency. The strategy of parallel execution is called “*Simulate As You Operate*” or SAYO. As illustrated in Figure 1, for *completely divisible* problems, the parameterization,

generation, sub-simulation (and its corresponding overhead) can be executed concurrently. Divisible problem means part of the model will be ready for simulation after the user parameterized and defined part of the model while completely divisible means a sub-simulation is feasible between two variables.

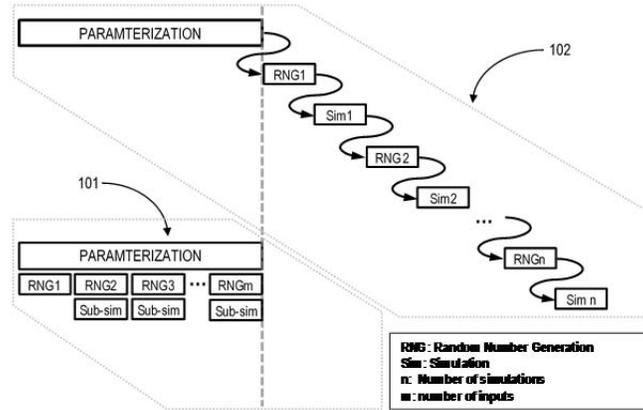


Figure. 1 Traditional Simulation Method and SAYO

For each parameterized variable, n random numbers (n =number of simulation trials) are generated together at one time, which becomes a random number tuple (RNT) for sub-simulations (while traditional method generates only one random number at one time). The time needed for each sub-simulation depends on the maximum of the above three. Assuming the parameterization takes the longest time then total time required for the simulation of a *completely divisible* problem is:

$$TT_{cd} = \sum_{i=1}^m \max\{pt_i, gt_i, (st_{i+1} + ot_{i+1})\} = \sum_{i=1}^m pt_i = PT \tag{1}$$

where, m equals to the number of model variables and PT denotes the total time for parameterization. Using the traditional probabilistic simulation method, such as Monte Carlo simulation, the random number generation process and simulation won't be started until all the model variables have been parameterized. Simulation is then divided into n trials, wherein one random number is generated for each model variable in each simulation trial. The generated random numbers of m variables will then be used to perform one simulation and yield one result of the model. This process will be repeated for n times and statistical inferences may be made upon n simulation results. In this process, each trial requires one instance of overhead time and thus the total overhead time needed for traditional method is $(n \times OH)$, where OH is the overhead time for one simulation trial. Therefore, the total time required for traditional method can be defined as:

$$TT_{tra} = PT + GT + ST + n \times OH \tag{2}$$

As shown in Equation 2, SAYO approach saves $GT + ST + n \times OH$ in simulation time over traditional method.

RISK: REAL-TIME INTERACTIVE SIMULATION TOOL

As illustrated in Figure 2, RISK tool structure includes a *DigitSource* which contains true random numbers generated by physical processes such as Quantum process; a *DigitBank* that stores user's history of parameterized models, model inputs and model simulation outputs; a *Model Evaluation* module that assigns the modeling, parameterizing and updating tasks to the other modules, and divides an entire problem to a set of sub-problems for enabling parallel and instant computation using grid computing; a distribution filter that can converted uniformly distributed random numbers to random numbers that follow arbitrary distributions; a *Temporary Storage* or cache that stores the sub-models and corresponding variables; a *Cloud based Grid Computing* facility that finish the computing tasks assigned; a *Synthesizing Module* that synthesizes the simulation results of sub-problems; and finally a web based user interface which is either in tabular or click-and-point format.

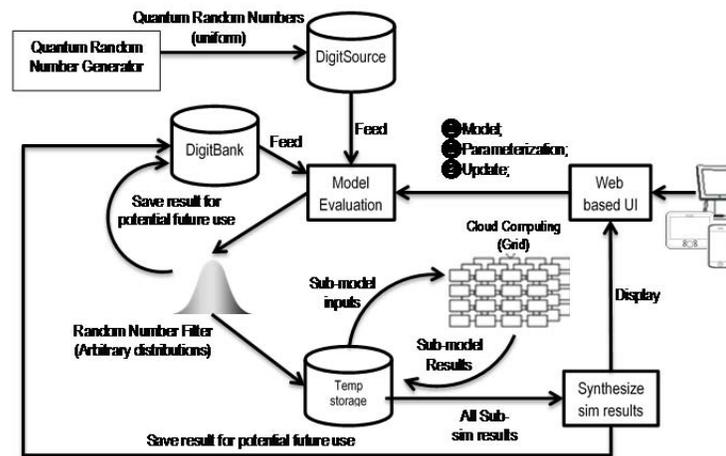


Figure 2. RISK Tool Architecture

In RISK tool, the user first defines the model for a given problem through the web based UI. Then the user parameterizes at least one model variable (a random variable), which will be sent to the Model Evaluation module instantaneously. After receiving the distribution parameters (parameterization), the Model Evaluation module first checks if there are any random number tuples (RNTs) in the DigitBank from past modeling and simulation, and follows the defined distribution. If any tuples are found, Model Evaluation module moves the RNTs to the temporary storage for future computation. If there are no existing RNTs that follow the defined distribution, Model Evaluation module utilizes the source random numbers, DigitSource, to generate random numbers following the defined distribution and saves them as RNTs in the temporary storage for future computation. The uniformly distributed true random numbers will be converted into random numbers that follow arbitrary distributions through a Random Number Filter, based on existing random number generation methods such as inverse F method, Acceptance-Rejection method, or Markov Chain Monte Carlo method. If the problem is *completely divisible*, after the user defined and parameterized the first two variables, a simulation starts immediately in the cloud based grid computing facility while the user is parameterizing the third

variable, and result will be saved as a random number tuple for variable 1 and 2, and RNT_{1and2} . Then the random tuple of the third parameterized variable, or RNT_3 , will be aggregated with RNT_{1and2} , which gives $RNT_{1and2and3}$, while the user is still parameterizing the fourth variable. This process will be repeated until $RNT_{1and2and3and\dots andM}$ is obtained, when the user is very much likely just done with the parameterization of the last variable m .

The prerequisite of a perfect SAYO is a linear model, or in other words, for any variable i and $i+1$, it is possible to perform a simulation, such as the case of risk register simulation. If not, the random number tuple for variable i , or RNT_i , is saved until it can be simulated. Another special case of SAYO is: for any variable i , the simulation depends on variable p and q , where $q > p$ and $p - i > 1$. Then a simulation can be performed between variable i and p first as an interim simulation. When RNT of q is ready, the result from the temporary simulation is updated.

It is important to note that all the interim simulation and updating processes are executed concurrently in parallel with the user parameterization process. If all the model variables have been simulated, the temporary simulation outputs will be synthesized to the final simulation result and shown in the web based UI. Even if the problem is *indivisible*, all the required random number tuples can still be generated in parallel with user parameterization process; and all the random number tuples are prepared all together before the simulation. Once the parameterization is completed, the request of simulation is sent to Cloud based Grid computing module automatically, where simulation is performed. Computational overhead for simulation and database operation will be significantly reduced. According to equation (2), it saves $(n-1) \times OH$ compared to traditional simulation method.

The users may select to modify only part of the model, such as the cases in a what-if scenario analysis. In this case, RISK will hold the model information such as random number tuples used in the simulation and the simulation results of unaffected part of the model fixed, and only repeat the process described above on the affected part of the model. The final results will be synthesized to reflect the change in state of the model.

IMPLEMENTATION EXAMPLES

Two examples are presented to demonstrate how RISK help realize WYSIWYG type of probabilistic simulation in addressing AEC problems.

PERT Simulation Example. Figure 3 shows a simple scheduling problem with only five activities (A, B, C, D and E) and corresponding table of random number tuples. The random number tuples are duration RNT ($RNT_{ACTIVITY}$), starting date RNT ($RNT_{ACTIVITY/S}$), and finish date RNT ($RNT_{ACTIVITY/F}$), which contain, for example, 1,000 random numbers each. In this example, if we assume a start date of 1/7/2013 for activity A, the tool generates a random number tuple for the starting date of activity A (RNT_{AS}) immediately. After the scheduler parameterized distribution, a random number tuple for the duration of activity A (RNT_A) is generated instantly, the finish date of activity A (RNT_{AF}) is calculated by ($RNT_{AS} + RNT_A$). Based on the network, RNT_{AF} is transferred to B and D as their random number tuples of start dates respectively, namely RNT_{BS} and RNT_{DS} . Following the same logic, RNT_B , RNT_{BF} ,

RNT_{CS} , RNT_C , RNT_{CF} , RNT_D , RNT_{DF} are transferred, generated and calculated, while the scheduler is still parameterizing and modeling the schedule.

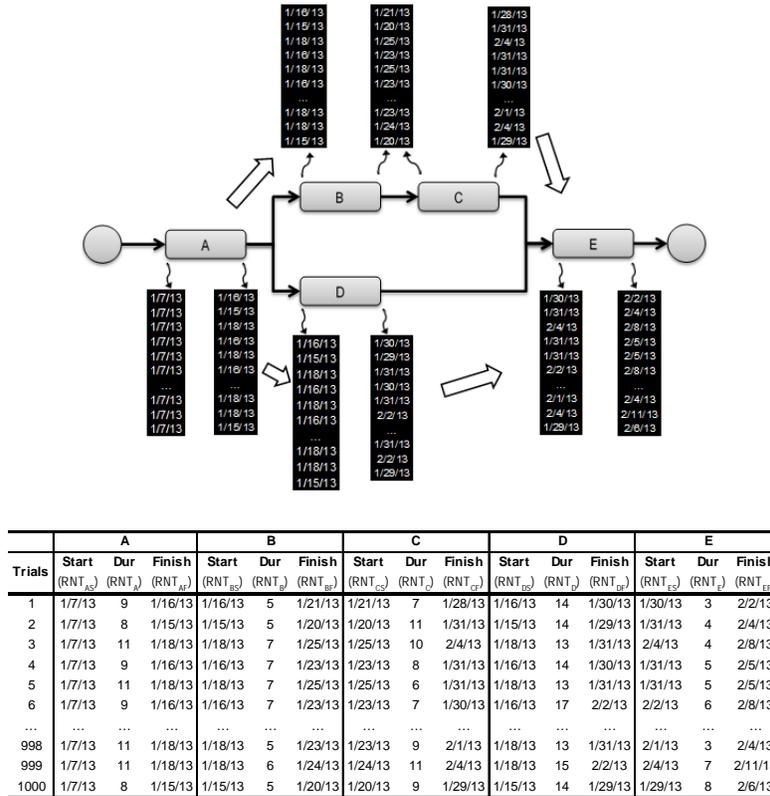


Figure 3. PERT Simulation Example using RISK

The random number tuple RNT_{ES} equals to the maximum of RNT_{CF} and RNT_{DF} and thus a maximum calculation is done and RNT_{ES} is obtained. RNT_{EF} , which is also the random number tuple of the project finish date is then calculated by $RNT_{ES} + RNT_E$. Since all the critical random number tuples have been generated, transferred and calculated concurrently when the scheduler is parameterizing and modeling the schedule, the final project completion date distribution is obtained immediately after the scheduler parameterized activity E.

Risk Analysis Example

The second example of RISK application is the simulation of Risk Registers, a common practice of risk management in the industry. The Construction Industry Institute Research Project 280 recommends a three-level risk management framework (Construction Industry Institute, 2012), which includes:

1. **Risk identification:** to identify a list of risk items from a risk register;
2. **Deterministic risk analysis:** to develop a single-point estimate of the risks' impacts and to calculate the expected value.
3. **Probabilistic risk analysis:** to analyze risk through probabilistic distribution estimates of potential impacts.

Realizing the level-3 risk analysis requires Monte Carlo simulation that is often supported by third-party software such like @RISK or Oracle Crystal Ball. The set-up of these software packages is rather cumbersome, requiring expertise in statistics as well as simulation. As a result, although the level-3 risk analysis is considered more desirable, it is rarely being used in real-world projects (Construction Industry Institute, 2012).

RISK tool can be applied to perform the level-3 risk analysis, without any third-party software. As shown in Figure 4, once the user identifies a risk item and specifies the impact of the risk item on either cost or schedule as a three-point estimate, a random number tuple will be generated following triangular distribution. By SAYO the impacts in the form of distribution curves will be displayed instantaneously, and critical values under provided significance level α will be calculated. Then the user starts to identifies/selects the second risk item.

Risk 1		Risk 2		Risk N		Project
\$ 2,578		\$ 1,072		\$ 11,572		\$195,772
\$ 2,102		\$ 1,071		\$ 12,692		\$195,774
\$ 1,727		\$ 846		\$ 11,479		\$193,834
\$ 1,866		\$ 987		\$ 11,050		\$201,031
\$ 1,192		\$ 908		\$ 6,564		\$228,611
\$ 2,462		\$ 967		\$ 14,042		\$207,115
...	
\$ 2,190	+	\$ 856	+ ... +	\$ 10,180	=	\$193,150
\$ 1,858		\$ 970		\$ 19,280		\$208,523
\$ 1,856		\$ 865		\$ 13,384		\$193,664
\$ 2,070		\$ 993		\$ 9,394		\$213,259

Figure 4. Random Number Tuples Risk Register Simulation

If there are correlations among risk items, methods for preserving the correlations will be used such as Cholsky decomposition. Given the likelihood of the risk items that are provided by the user, a threshold may be applied to each risk item to filter out partial set of the generated random number tuple. Then the overall impacts of two risk items are calculated and displayed instantaneously by aggregating two random number tuples. This process will be repeated for the rest risk items. Note that the computation is performed concurrently and updated on a timely basis, i.e., every time when any part of the risk register is updated. Once the user finishes the selection of the last risk item, the impacts of the entire risk register will be displayed instantaneously.

RISK may also maintain the PDFs of all commonly used risk items, such as what have been identified by the industry (for example; Construction Industry Institute, 2007). The PDFs may be obtained from historical data, or surveys to the domain experts. The random number tuples can be generated in advance and stored on DigitBank.

DISCUSSION AND CONCLUSION

AEC problems are becoming more complex. The variables of construction decision-making are increasing dramatically, and exposed to uncertainty, ambiguity, and

variability. Probabilistic simulation has long been used by the industry and construction research community to deal with the uncertainties. Despite the great success (compared to conventional deterministic approaches), there remains an urgent need to make probabilistic simulation more practical and efficient.

This paper attempts to offer a solution with “Simulate As You Operate” (SAYO) method and an apparatus (RISK) based on a Cloud infrastructure. The parameterization, random number generation, and simulation processes are carefully integrated to reduce unnecessary waiting time and overhead in a probabilistic simulation. It ends up with the interactive probabilistic simulation, and in turn, seamless decision-making process. This paper also proposes a Cloud environment that allows remote project participants to work on the same platform. For the scheduling example, schedulers at different locations may access the same schedule and edit the variables concurrently. Updates to the schedule can be shown instantaneously in a WYSIWYG fashion for all the participants. With predefined security levels participants may view, edit or override the simulation process at real-time. Collaborativity is realized in advanced analysis and transparency is ensured.

With the innovative simulation technology and corresponding tools probabilistic simulation becomes a practical tool with easier input and output functions. The model set-up, random number generation, and simulation processes are removed completely from user operations. As a result, user interface may be realized in portable devices such as smart phones or tablets, which may be more convenient for construction project participants.

REFERENCES

- Chou, J.-S., Yang, I., and Chong, W. K. (2009). "Probabilistic simulation for developing likelihood distribution of engineering project cost. " *Automation in Construction*, 18(5), 570-577.
- Construction Industry Institute, C. (2007). *Contracting to Appropriately Allocate Risk*: Construction Industry Institute.
- Construction Industry Institute, C. (2012). *Applying Probabilistic Risk Management in Design and Construction Projects*: Construction Industry Institute.
- Lee, S. H., Pena-Mora, F., and Park, M. (2006). "Dynamic planning and control methodology for strategic and operational construction project management. " *Automation in Construction*, 15(1), 84-97.
- Lu, M., Lam, H.-C., and Dai, F. (2008). "Resource-constrained critical path analysis based on discrete event simulation and particle swarm optimization. " *Automation in Construction*, 17(6), 670-681.
- Mooney, C. Z. (1997). *Monte Carlo simulation*: Sage Publications, Inc.
- Sadeghi, N., Fayek, A. R., and Pedrycz, W. (2010). "Fuzzy Monte Carlo simulation and risk assessment in construction. " *Computer-Aided Civil and Infrastructure Engineering*, 25(4), 238-252.
- Smith, D. (1994). "Incorporating risk into capital budgeting decisions using simulation. " *Management Decision*, 32(9), 20-26.
- Yi, Y. K., and Malkawi, A. M. (2012). "Site-specific optimal energy form generation based on hierarchical geometry relation. " *Automation in Construction*, 26, 77-91.