LESSONS LEARNED FROM SCHEDULE ESTIMATION USING REAL-TIME DATA IN A CONCRETING OPERATION

Sangwon Han* and Pan Hao

[1] *School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia*
** Corresponding author (s.han@unsw.edu.au)*

ABSTRACT: The unique, complex and interdependent nature of modern construction operations means that the use of previous datasets in simulating a new project may raise validity issues. To address this problem, the use of real-time data in simulation studies has been introduced and practiced in recent years. However, little is known about the details of how to use real-time data to better estimate the performance of an operation. In particular, there is little awareness of the timing of data collection required to yield a valid sample on which to base subsequent plans and schedules for the entire operation. Using data from a concreting operation in a Dubai construction project, this paper examines the validity and reliability of forecasted results using different segments of the available data. For this purpose, the datasets are statistically analyzed and then segmented based on the percentage of the operation completed. A discrete event simulation model is developed to estimate the schedule performance using each cumulative data segment and the simulation results are compared to actual project data. Through the case study, it was found that the statistical distribution identified at an early stage in the operation may not successfully represent the statistical properties of the whole dataset due to local extreme variation; and that it is important to identify the bottleneck resource (e.g., pumps in this study) and pay particular attention to variation in its cycle time in order to successfully estimate and control the performance of an operation.

***Keywords**: Real-time Data, Discrete Event Simulation, Concrete Delivery, Schedule Prediction*

1. INTRODUCTION

Discrete event simulation (DES) has been utilized extensively for planning and analysis of construction operations in order to address complexities and uncertainties inherent in construction environments [1]. DES modeling subdivides a construction operation (e.g., earthmoving) into manageable smaller work tasks (e.g., load, haul, dump and return) to directly replicate the complex logistics of the operation. Then, in order to incorporate uncertainties, it applies statistical distribution (e.g., Normal or Triangular) to the work tasks, based on the analysis of previous similar operations. An underlying assumption here is that these work tasks are common so that the past experience of project managers can be easily migrated [2]. Modern construction operations are increasingly complex and interdependent by nature. For this reason, validity of data for the planning and analysis of an operation may be questionable when historical project data is used, which may have been executed in a vastly different environment from the operation in hand.

To address this, several researchers including [3] and [4] suggest the idea of real-time simulation, integrated with automatic data collection techniques (e.g., sensors, RFID, GPS, and so forth). These researchers developed innovative frameworks where real-time data are continuously fed into the simulation model in order to improve predictability of performance estimation. However, little is known about the details of how to use the real-time data to better estimate the performance of an operation. In particular, there is little awareness of the timing of data collection required to acquire a valid sample on which to base subsequent plans and schedules for the entire operation.

For the purpose of performance estimation, it is important to establish when a reliable estimation will be produced, as well as how reliable it is. Performance estimation near the end of an operation, where almost all datasets are obtained, would be very reliable, but less meaningful to project managers because only a small amount of further planning would be required. On the other hand, performance estimation at the very beginning of an operation can be very helpful to project managers, if it is accurate, but its reliability is questionable. Therefore, finding an earliest point in time at which a reliable estimation can be obtained would contribute to the effective application of real-time simulation for performance estimation of construction operations.

To address this, this paper analyzes a concreting operation in a Dubai construction project. A discrete event simulation model is developed and the simulation results using different segment of the concreting operation are compared to actual project data to examine the timing and reliability of the performance estimation.

2. CASE EXAMPLE: CONCRETING OPERATION

The Tameer Towers project in Abu Dhabi was selected as a case study. The project consists of four residential towers: a commercial skyscraper, a luxurious hotel and 10 podiums of retail underneath the entire complex. In order to find an earliest point at which reliable schedule estimation can be made, the concrete delivery and placement process for the Tower 1 raft foundation was chosen as a case study. The dimension of the Tower 1 raft foundation is 144m by 144m with 200mm thickness, which requires 4,147m^3 of concrete. This process involves 502 duty cycles, taking about 20 hours (from 10:35 AM to 6:12 AM next morning) using 84 different trucks. The ready mixed concrete was supplied from a company which has five plants. Due to space restriction of the project site, 5 pumps were used for the Tower 1 raft foundation slab. Trucks arrive at the site with information including the time of departure, concrete loading time, temperature and initial slump at the plant. Site data including trucks' arrival times and unloading start and finish times were collected by two groups (one from the contractor and the other from the consulting company) for the purpose of data consistency. Based on this information, travel, loading and waiting durations of each duty cycle were calculated.

3. INPUT DATA MODELLING

In order to find the earliest point at which reliable schedule estimation can be made, the data of 502 duty cycles was classified into 10 segments (A to J in Table 1). As shown in Table 1, segment F's average travel time is much longer than those of the other segments. This is mainly attributed to a traffic jam experienced in collecting data for segment F. Also, the average waiting time for segment A is much longer than those of the other segments. This confirms that the process includes a transient period at segment A before reaching a steady state. Finally, though relatively constant, unloading time shows an increasing trend on average.

The current dataset relating to an operation at hand can be used to estimate future performance, as it addresses uniqueness of the operation. Thus, it is assumed that an estimation of schedule performance can be made based on data that has been accumulated. This means we can use the first 10% of the operation's data (segement A in Table 1) when we estimate the future performance at 10% progress. In order to make an estimation of future performance at a certain moment, we need to find a suitable statistical distribution which can best fit the data that has accumulated so far. To this end, EasyFit was utilized to find the best fit statistical distribution curves and to estimate parameters of the statistical distributions. To assess the fitness between the actual data and the statistical distributions, the chi-square test was conducted.

Table 2 shows the best fit statistical distribution curves at each cumulative 10% of progress. For example, Lognormal (2.78, 0.71, 7.24) was identified to best fit the first 10% of duty cycle data (segment A). Table 2 shows a general trend that the fit statistics (χ^2) for both travel and unload time increase as progress through the operation increases, which means a lower goodness-of-fit between the current dataset and the identified statistical distribution. This is not surprising because as more data is collected, the variance of the data increases, unless the data is highly consistent. This implies that the dataset collected from the case process has some inconsistency between the segments, which supports the proposition that real-time data collection is needed for this process.

Table 3 shows the fitness of each identified distribution to the whole dataset. That is, it shows how well the first x% of dataset represents the statistical properties of the whole dataset; and hence, how well it would estimate the entire process schedule performance. It should be noted that the fit statistics (χ^2) in Table 2 show goodness-of-fit between the cumulative data at x% of progress and the identified distribution that best describes the statistical properties of the cumulative data. On the other hand, the fit statistics (χ^2) in Table 3 show goodness-of-fit between the cumulative data at x% of progress and the identified distribution that best describes the statistical properties of the entire dataset. Therefore, fit statistics (χ^2) in Table 3 are much higher than those in Table 2. For example, the fit statistic (χ^2) of Lognormal (2.78, 0.71, 7.24) to the first 10% of Travel Time data is 0.57; but the fit statistic (χ^2) to the whole Travel Time data is 68.37. This is due to inconsistency of the data, which indicates that the statistical properties of the remaining 90% of data are quite different from those of the first 10%. For this reason, the deviation of fit statistics (χ^2) between Table 2 and 3 decreases as more data accumulates. Note that fit statistics (χ^2) at 100% in Table 2 and 3 are the same.

Table 1. Dataset of 502 duty cycles

Segment	Duty Cycles	Travel Time (minute)			Waiting Time (minute)			Unload Time (minute)		
		Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
A	1-50	27.67	10	59	18.11	4	51	5.19	3	8
B	51-100	25.51	14	53	2.51	1	7	5.51	3	8
C	101-150	22.69	13	60	2.37	0	8	5.39	3	7
D	151-200	27.13	12	45	5.61	1	16	5.37	3	9
E	201-250	28.96	10	59	6.53	1	22	5.47	3	10
F	251-300	40.41	13	60	5.76	1	14	5.76	3	10
G	301-350	28.13	12	54	4.82	1	12	5.85	3	10
H	351-400	32.26	17	55	6.74	1	19	6.08	3	13
I	401-450	33.93	14	61	4.98	1	11	6.62	3	10
J	451-502	28.06	16	59	5.15	1	17	7.17	4	10

Table 2. Fit statistics (χ^2) of the best fit distributions to each segment

Progress	Segment	Travel Time		Unload Time	
		Best Fit Distribution	χ^2	Best Fit Distribution	χ^2
10%	A	Lognormal (2.78, 0.71, 7.24)	0.57	Pert (2.58, 10.52, 4.46)	1.31
20%	A-B	Cauchy (5.02, 22.86)	4.54	Cauchy (0.875, 5.10)	2.96
30%	A-C	Cauchy (4.43, 22.34)	12.26	Cauchy (0.86, 5.2)	6.39
40%	A-D	Erlang (5.11, 3, 8.92)	9.62	Exponential (0.42, 3.0)	7.04
50%	A-E	Gamma (3.17, 5.58, 8.67)	9.48	Cauchy (0.84, 5.10)	22.15
60%	A-F	Erlang (6.73, 3, 8.63)	18.05	Cauchy (0.833, 5.132)	12.23
70%	A-G	Erlang (6.5, 3, 8.6)	15.34	Cauchy (0.862, 5.15)	14.37
80%	A-H	Lognormal (3.19, 0.42, 2.64)	11.20	Cauchy (0.912, 5.23)	25.0
90%	A-I	Lognormal (3.19, 0.43, 3.03)	13.02	Normal (1.645, 5.69)	46.99
100%	A-J	Lognormal (3.12, 0.46, 4.38)	17.58	Cauchy (1.03, 5.51)	23.98

Table 3. Fit statistics (χ^2) of the best fit distributions to the entire dataset

Progress	Segment	Travel Time		Unload Time	
		Best Fit Distribution	χ^2	Best Fit Distribution	χ^2
10%	A	Lognormal (2.78, 0.71, 7.24)	68.37	Pert (2.58, 10.52, 4.46)	-
20%	A-B	Cauchy (5.02, 22.86)	166.21	Cauchy (0.875, 5.10)	34.71
30%	A-C	Cauchy (4.43, 22.34)	187.48	Cauchy (0.86, 5.2)	34.94
40%	A-D	Erlang (5.11, 3, 8.92)	143.3	Exponential (0.42, 3.0)	99.78
50%	A-E	Gamma (3.17, 5.58, 8.67)	69.34	Cauchy (0.84, 5.10)	35.13
60%	A-F	Erlang (6.73, 3, 8.63)	32.91	Cauchy (0.833, 5.132)	35.26
70%	A-G	Erlang (6.5, 3, 8.6)	28.17	Cauchy (0.862, 5.15)	34.87
80%	A-H	Lognormal (3.19, 0.42, 2.64)	23.94	Cauchy (0.912, 5.23)	34.3
90%	A-I	Lognormal (3.19, 0.43, 3.03)	23.87	Normal (1.645, 5.69)	30.33
100%	A-J	Lognormal (3.12, 0.46, 4.38)	17.58	Cauchy (1.03, 5.51)	23.98

4. SIMULATION MODEL DEVELOPMENT

In order to test the schedule performance against the best fit distribution identified at each segment, a discrete event simulation model is developed using AnyLogic 6.4 University version. By applying the best fit distribution curves to the simulation model, the difference between the simulated schedule performance and the actual duration is measured.

4.1. Model Description

Figure 1 shows the ready mixed concrete production, delivery, and placement process described in the Anylogic environment. A duty cycle of a concrete delivery truck begins only when a truck (‘Truck’) and an order (‘Order’) are both available. If there is no delivery order, available trucks wait at ‘TruckQueue’ and conversely, if there are no available trucks, an order waits at ‘OrderQueue’ until a truck becomes available. When both an order and a truck are available, the truck can begin to load concrete (‘Load’) at the concrete plant (‘Plant’). Once finishing the loading, the truck can travel to the construction site (‘TravelToSite’). In order to pour concrete, a pump (‘Pump’) is required. Therefore, on arrival of a truck, if no pump is available, the truck waits at the construction site (‘WaitToPour’) until a pump is available. Similarly, if there are no available trucks, a pump waits for a truck to arrive at the site for pouring (‘PumpQueue’). Once both a pump and a truck are available, the truck can approach the pump (‘Maneuvering’) and unload concrete through the pump (‘Unload’). After unloading concrete, the pump is released from the truck and is able to serve the next available truck. The empty truck washes its body (‘Wash’) and returns to the plant for the following duty cycle (‘Return’).

4.2 Model Validation

In order to estimate the schedule performance using the best fit distribution identified at each segment, the simulation model first needs to be validated. For this, the process logistics were tested and the simulation results acquired by using the best fit distribution curve identified at 100% of progress was compared to and the actual schedule performance. In order to statistically validate simulation results, the model was run 1,000 times to obtain stochastic results for the total operation time (TOT) and the average truck waiting time (ATWT). The simulation model yielded results quite close to the actual performance. The actual TOT was 1,180 minutes and the average of the TOT simulation results was 1,181 minutes. The actual ATWT was 6.2 minutes and the average of simulated ATWT results was 5.5 minutes. In this way, the model was validated for use in measuring the predictability of best fit distributions identified at each segment.

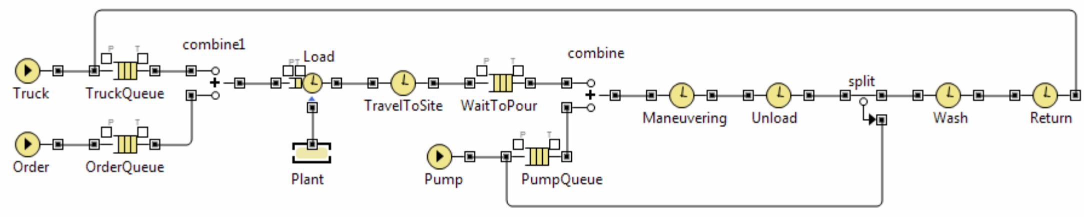

Figure 1. Simulation Model

5. SIMULATION RESULT ANALYSIS

In order to examine the impact of using each cumulative data segment on the total duration, the developed model was simulated on each dataset 100 times (Table 4). Before the simulation, it was expected that the first 30% progress datasets (A-C segments) would bring the greatest deviation as its fit statistics (χ^2) were the largest in Travel Time. However, surprisingly, it was found that there was no significant difference in terms of the Total Time. The greatest difference (at only 1.95% or 23 minutes) was found when using the first 40% progress dataset. This means that the total time is not very sensitive to the variation of travel time. Careful analysis of the simulation results revealed that the process is highly governed by the production rate of the pumps, as opposed to the trucks' cycles; and that the provision of enough trucks (84 trucks) offset the variation in travel time (refer to the average travel time of segment F: about 10 minutes longer than those of other segments). This can be proved by calculating trucks' and pumps' production rates. In this process, a pump's average cycle time is 10.3 minute (4.5 minutes for maneuvering and 5.8 minutes for unloading). As 5 pumps are assigned, if there is no restriction in terms of providing enough trucks, the total time to serve the 502 duty cycles would be:

502 cycles * (5 pumps / 10.3 minutes) = 1,034 minutes.

This is quite close to the actual duration of 1,180 minutes. The difference of 146 minutes (1,180 – 1,034) is largely attributed to randomness of the duty cycle and the pumps' waiting time at the initial stage before reaching at the equilibrium state. This confirms that the pump is the "bottleneck resource" that restricts the overall performance of this operation, and that some trucks are waiting at the construction site while the pumps work continuously. Because of the waiting trucks, the impact of variation in the travel time is minimized. This can be further validated by Figure 2.

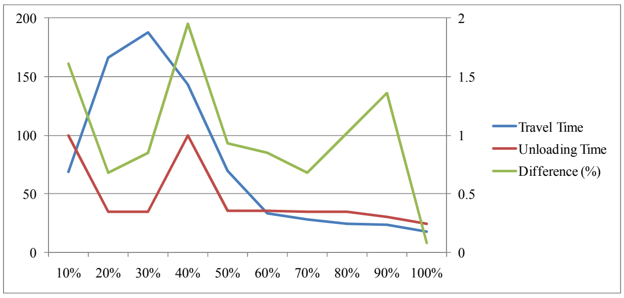

Figure 2. Comparison of Simulation Results

Figure 2 shows that the difference between the actual duration and the simulated duration is more closely related with the fit statistics (χ^2) of unloading time, rather than that of travel time. This confirms that the overall performance of this operation is highly restricted by the pumps rather than the trucks. For this reason, contrary to initial expectations, variation in travel time did not significantly affect the overall schedule performance. Also, as shown in Table 1, there is a general increasing trend in the unload time as time goes by, however unloading time is relatively constant. Because of this, the simulation results do not fluctuate very much, and the maximum deviation is less than 2% at 40% progress in this operation.

Through this case study, two important lessons are learned regarding performance estimation using real-time data. Firstly, it is a general trend that collecting more data increases the goodness-of-fit of the real-time data to the entire dataset; however, the goodness-of-fit may decrease due to sudden variation, particularly at the beginning stage. For example, as shown in Table 3, the fit statistic (χ^2) of travel time at the 30% is higher than 20% (i.e., low reliability). This is attributed to the fact that the average travel time at segment C is 22.69 minutes (Table 1). In comparison to other segments, this is exceptionally small. The outlying nature of this observation distorts the statistical properties of the dataset. It is seen in this case that real-time data collected before 40-50% may not properly address the statistical properties of the whole dataset.

Secondly, there is a general trend that the fit statistics (χ^2) used in a simulation model has a strong relationship with the prediction errors of the model. However, it is of utmost importance to identify and understand the bottleneck resource which constrains the operation performance. As shown in Figure 2, the prediction error is more sensitive to the fit statistics (χ^2) of unload time (included in the cycle time of the pumps which are the bottleneck resource in this operation), rather than that of travel time (included in the cycle time of the trucks, which are not a bottleneck resource). As the bottleneck resource can be identified at the initial planning stage, Figure 2 suggests that construction managers need to pay the greatest attention to variations in the cycle time of the bottleneck resource (e.g., unloading time in this case), to forecast the performance. This also suggests that construction managers can increase the probability of completing an operation within its planned duration by carefully controlling the cycle time of the bottleneck resource. Therefore, in order to increase the predictability of an operation, it would be effective to set the bottleneck resource to be a resource whose cycle time is highly controllable. For example, as shown in the case study, it would be desirable to set the pumps as the bottleneck since it is easier for construction managers to control unloading time (i.e., major component of the pump cycle time) than haul time (i.e., major component of the truck cycle time).

Table 4. Simulation Results

Progress	Segment	Travel Time		Unload Time		Total Time	Difference
		Best Fit Distribution	χ^2	Best Fit Distribution	χ^2		
Actual	-	-	-	-	-	1,180	-
10%	A	Lognormal (2.78, 0.71, 7.24)	68.37	Pert (2.58, 10.52, 4.46)	-	1,161	1.61%
20%	A-B	Cauchy (5.02, 22.86)	166.21	Cauchy (0.875, 5.10)	34.71	1,172	0.68%
30%	A-C	Cauchy (4.43, 22.34)	187.48	Cauchy (0.86, 5.2)	34.94	1,170	0.85%
40%	A-D	Erlang (5.11, 3, 8.92)	143.3	Exponential (0.42, 3.0)	99.78	1,157	1.95%
50%	A-E	Gamma (3.17, 5.58, 8.67)	69.34	Cauchy (0.84, 5.10)	35.13	1,169	0.93%
60%	A-F	Erlang (6.73, 3, 8.63)	32.91	Cauchy (0.833, 5.132)	35.26	1,170	0.85%
70%	A-G	Erlang (6.5, 3, 8.6)	28.17	Cauchy (0.862, 5.15)	34.87	1,172	0.68%
80%	A-H	Lognormal (3.19, 0.42, 2.64)	23.94	Cauchy (0.912, 5.23)	34.3	1,168	1.02%
90%	A-I	Lognormal (3.19, 0.43, 3.03)	23.87	Normal (1.645, 5.69)	30.33	1,164	1.36%
100%	A-J	Lognormal (3.12, 0.46, 4.38)	17.58	Cauchy (1.03, 5.51)	23.98	1,181	-0.08%

6. CONCLUSIONS

Simulation results are highly affected by the quality of input data. Due to the unique nature of a construction project, the idea of real-time simulation has been raised by several researchers. However, little is known about how to utilize the real-time data to estimate performance of a construction operation.

This paper analyzed a concreting operating in a Dubai construction project to address this issue. This paper examined the goodness-of-fit of real-time data to represent the statistical properties of the entire dataset. Then, a discrete-event simulation model was developed to confirm the relationship between the goodness-of-fit of real-time data and predictability of the model.

Through the case study, it was found that a statistical distribution identified before 40-50% progress may not successfully represent the statistical properties of the entire dataset. More importantly, it was revealed that it is important to identify the bottleneck resource and pay the greatest attention to variation in its cycle time, in order to successfully estimate and control the performance of an operation.

While this paper shows an initial direction of how to effectively use real-time data for planning and analysis of a construction operation, the research needs to be further validated by assessing more case studies. Another research direction linked to this research is the study of how to use real-time data in an operation where resources are well balanced.

ACKNOWLEGMENT

The authors express their sincere appreciation to Mr. Reza Sabetnia for his assistance with the data collection for this paper.

REFERENCES

[1] Halpin, D. W., Jen, H., and Kim, J., “A Construction Process Simulation Web Service.” *Proceedings of the 2003 Winter Simulation Conference*, pp. 1503-1509, 2003.

[2] Peña-Mora, F., Han, S., Lee, S., and Park, M., “Strategic-operational construction management: Hybid system dynamics and discrete event approach.” *Journal of Construction Engineering and Management*, ASCE, Vol. 134(9), pp. 701-710, 2008.

[3] Lu, M., Dai, F., and Chen, W., “Real-time decision support for planning concrete plant operations enabled by integrating vehicle tracking technology, simulation, and optimization algorithm.” *Canadian Journal of Civil Engineering*, Vol. 34, pp.912-922, 2007.

[4] Song, L., Ramos, F., and Arnold, K., “A framework for real-time simulation of heavy construction operations.” *Proceeding of the 2008 Winter Simulation Conference*, pp. 2387-2395, 2008