S7-1

IMPROVING THE USABILITY OF STOCHASTIC SIMULATION BASED SCHEDULING SYSTEM

Tae-Hyun Bae[1], Ryul-Hee Kim[1], Kyu-Yeol Song[2] and Dong-Eun Lee[3]

[1] Graduate student, Kyungpook National University, School of Architecture & Civil Engineering Daegu, KOREA

[2] Director, Dongwoo Engineering & Construction Co., LTD. Daegu, KOREA.

[3] (Corresponding) Assistant professor, Kyungpook National University, School of Architecture & Civil Engineering, Daegu, KOREA.
Correspond to dolee@knu.ac.kr

ABSTRACT: This paper introduces an automated tool named Advanced Stochastic Schedule Simulation System (AS4). The system automatically integrates CPM schedule data exported from Primavera Project Planner (P3) and historical activity duration data obtained from a project data warehouse, computes the best fit probability distribution functions (PDFs) of historical activity durations, assigns the PDFs identified to respective activities, computes the optimum number of simulation runs, simulates the schedule network for the optimum number of simulation runs, and estimates the best fit PDF of project completion times (PCTs). AS4 improves the reliability of simulation-based scheduling by effectively dealing with the uncertainties of the activities' durations, increases the usability of the schedule data obtained from commercial CPM software, and effectively handles the variability of the PCTs by finding the best fit PDF of PCTs. It is designed as an easy-to-use computer tool programmed in MATLAB. AS4 encourages the use of simulation-based scheduling because it is simple to use, it simplifies the tedious and burdensome process involved in finding the PDFs of the many activities' durations and in assigning the PDFs to the many activities of a new network under modeling, and it does away with the normality assumptions used by most simulation-based scheduling systems in modeling PCTs.

Keywords: Project completion time; simulation; scheduling; PERT.

1. INTRODUCTION

Existing simulation-based scheduling methods estimate the probability distribution functions (PDFs) of activities' durations retrieved from a project data warehouse, model a new network by assigning the PDFs identified to the activities of the network under study, and compute project completion times (PCTs) by simulating the network for many iterations defined by the user. The existing methods, which compute the probability to complete a project within a deadline defined by the user, provide an important measure to the construction scheduling community because they improve predictability by modeling the randomness and uncertainty of activity durations and handle the variability of PCTs. However, the existing methods are neither willingly employed nor efficiently used because they suffer from several deficiencies, as discussed in detail in later sections.

This study aims to develop a new simulation-based scheduling system that eliminates the deficiencies of existing simulation-based scheduling methods. The research activities were conducted in four steps. First, the performance of existing simulation-based scheduling systems was investigated. Second, a strategy was set to eliminate the deficiencies of existing systems. Then, a totally automated system, which accommodates the strategy, was developed. Third, a detailed illustration of the new system was demonstrated using a small network. Fourth, the capability of the new system to handle a large network was verified.

2. EXISTING TECHNIQUES FOR SIMULATION-BASED SCHEDULING

Detailed descriptions of existing simulation-based scheduling methods are provided in other publications (e.g., Sculli 1983, Barraza et al 2004, Lee 2005, Lee and Arditi 2006). The existing methods are involved in separately operating a commercial statistics package to identify the PDFs of activity durations, assigning the PDFs to the activities, and computing the PDF of PCTs. Sculli (1983) claims that existing methods are superior to PERT in accuracy. However, the accuracy of a system depends on the accuracy of the PDFs of activity durations and on the PDF of PCTs (Vose 2000, Yao and Chi 2007).

The existing simulation-based scheduling systems are classified into two categories, namely extended systems which are adopted from or added into existing operations research oriented discrete event simulation systems (i.e., CYCLONE, ABC, UM-CYCLONE, etc), and complete systems. Examples of extended systems include CYCLONE-CPM (Halpin, 1990), ABC-CPM (Shi, 1999), and CPM add-in for STROBOSCOPE (Martinez and Ioannou, 1997), while complete systems include Monte Carlo for Primavera (2008), Palisade @RISK for Project (2008), S3 (Lee and Arditi, 2006), and Cristal Ball for schedule risk analysis which is an add-in to MS Excel (Cristal Ball, 2004). The deficiencies of existing simulation-based scheduling systems are discussed below. First, in definition of precedence relationships, the user must define an activity's predecessors and successors one by one manually in the spreadsheet or manually convert the data structure of the document exported from a CPM package to make them usable by existing simulation-based scheduling software. This inconvenience is exacerbated when these methods are used with large networks consisting of many hundreds of activities. There are of course systems that define network models in a syntax environment (Halpin 1990, Martinez and Ioannou 1997), but these systems lack practicality when dealing with large networks, because in syntax format too, precedence relationships must be defined manually.

Second, most existing simulation-based scheduling systems depend on commercial statistics software packages often external (all extended systems and S3) but sometimes internal (e.g., Crystal Ball, and Palisade @RISK for Project) to the systems. One should estimate the PDFs of activity durations by manually and independently operating the software packages and should manually assign in the network under study. Again, this inconvenience is more acute when one deals with large networks. This inconvenience seriously limits the usability. Of all the existing simulation-based scheduling systems, only Monte Carlo for Primavera ®attempts to cope with this inefficiency, but it does this at the expense of compromising the accuracy of the prediction.

Third, the usability of existing simulation-based scheduling systems is limited because none offer an algorithm that automatically combine a schedule data obtained from CPM software and historical activity duration data from a project data warehouse and automatically convert this information into a format that is compatible for simulation operations.

Fourth, existing systems use graphical user interfaces to facilitate the modeling of the precedence relationships between activities. However, this facilitation comes at a cost. It increases computation time because computer resources are eaten up by the demands of the interface. In addition, these systems are confined by their user interface to consider only a few activities at a time when modeling the precedence relationships between activities. Specifically, S3 (Lee and Arditi 2005, 2006) and ABC-CPM (Shi 1999) are afflicted with this problem.

Fifth, existing simulation-based scheduling systems compromise the accuracy of prediction by assuming implicitly that PCTs are normally distributed on account of the central limit theorem. But it is clear that the normality assumption is nothing but an approximation. Nevertheless, existing systems assume that the sum of critical activities' durations on the critical path follows a normal distribution. It is presumably because for easy to obtain the mean and variance, even though discrete event simulation easily accommodates the use of different PDFs.

Finally, the existing scheduling methods such as (1) deterministic CPM, (2) probabilistic PERT, and (3) normal-PDF-based simulation scheduling has been in existence in separate standalone applications. But it is not convenient to use three different software packages when one wants to make use of all three of these methods for comparison purposes. A built-in capability that allows the system to generate results using these methods all at once could provide opportunities for comparison in an instant and could improve credibility relative to schedule analysis.

3. ADVANCED STOCHASTIC SCHEDULE SIMULATION SYSTEM (AS4)

The system retrieves schedule data from a commercially available CPM software such as P3, converts the data into a manageable format, estimates the best-fit-PDFs of activity durations, assigns the PDFs so identified to the respective activities, generates random variates based on the individual PDFs assigned to each activity, runs CPM for a predefined number of iterations, estimates the PDF of PCTs, and makes a prediction relative to project completion time. The method described below was coded into an automated system by using MATLAB programming. The algorithm of AS4 is presented in Figure 1. The detail descriptions are delayed due to lack of space.

The schedule data read in step ① in Figure 1 are converted into an appropriate data structure for simulation runs. The conversion algorithm in Figure 2 is used. The best-fit-PDFs and their parameters are computed using the algorithm presented in Figure 3 and stored in Section Ⓒ of Table 2. In this paper, detailed descriptions of the two algorithms are omitted due to lack of space.

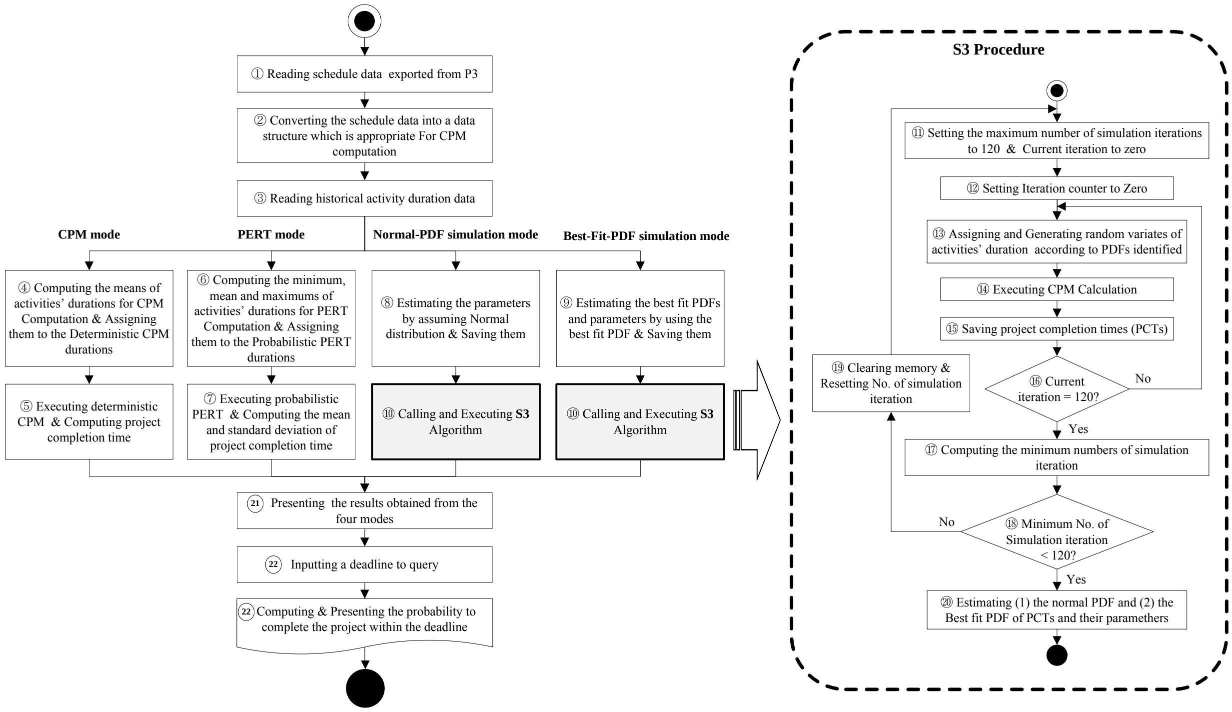

Figure 1. The advanced stochastic simulation based scheduling system

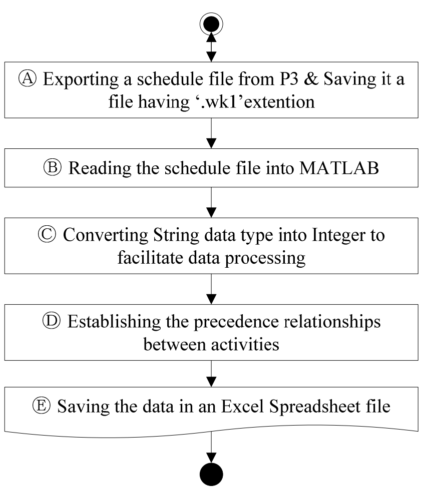

Figure 2. The conversion algorithm of P3 schedule data

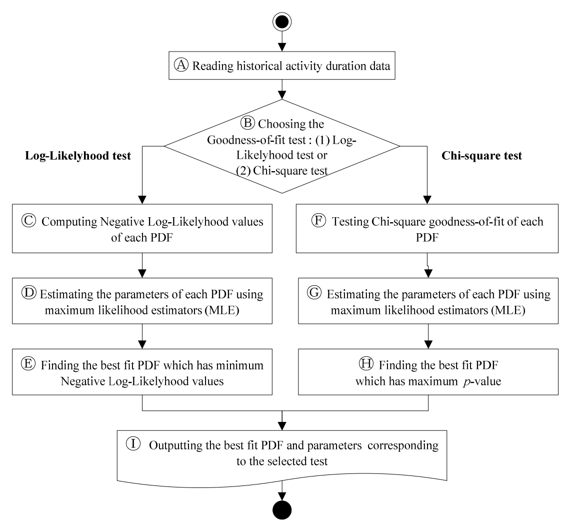

Figure 3. The algorithm estimating the best fit PDF

Table 2. Historical activity duration table

Activity Duration	Iterations																														Normal PDF		Best Fit PDFs			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	mu	sigma	Optimum PDF	MLE(1)	MLE(2)	MLE(3)
1	24	21	34	33	42	34	13	43	32	42	48	23	43	32	24	35	38	37	35	32	33	35	34	32	42	33	34	34	24	43	33.57	7.64	GEV PDF	-0.43	8.04	31.45
2	25	24	17	23	15	56	23	52	31	11	51	46	21	41	30	37	26	21	20	19	23	25	25	17	19	35	20	19	29	41	28.07	11.72	GEV PDF	0.13	7.89	22.38
3	33	45	21	45	46	25	34	38	33	37	33	35	34	46	45	19	21	28	25	35	38	33	41	42	27	24	35	36	26	21	33.37	8.19	GEV PDF	-0.47	8.63	31.28
4	20	24	15	25	26	16	15	19	17	25	20	30	23	15	18	19	25	25	20	19	27	15	28	19	24	27	21	24	20	22	21.33	4.37	GEV PDF	-0.30	4.25	19.86
5	30	34	33	35	32	32	25	25	26	34	31	35	32	27	29	29	38	34	24	32	27	28	35	33	32	39	25	29	29	30	30.77	3.95	GEV PDF	-0.31	3.90	29.45
6	24	25	28	19	18	23	22	23	16	24	21	22	19	33	30	25	26	31	24	22	27	22	24	25	23	25	26	26	25	27	24.17	3.69	Normal PDF	24.17	3.69	
7	18	20	25	17	16	20	22	13	18	17	20	15	26	33	34	21	17	15	20	19	22	24	16	18	25	20	23	17	14	16	20.03	4.97	GEV PDF	0.10	3.41	17.71
8	24	28	15	25	26	25	26	19	18	18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	18	18	16	27	28	24	30	32	24	25	22	[illegible]	[illegible]	22.57	4.23	GEV PDF	[illegible]	3.94	20.97
9	25	25	34	40	34	23	22	23	27	19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	34	33	15	19	25	24	25	23	25	24	21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10	33	45	44	27	32	32	35	35	36	34	31	35	32	35	34	35	28	42	53	35	41	35	37	35	34	32	42	17	24	18	34.20	7.43	Normal PDF	34.20	7.43	
11	20	11	23	30	31	20	24	21	26	23	18	24	26	16	18	32	16	34	35	18	17	31	15	32	20	24	18	33	21	25	23.40	6.54	GEV PDF	-0.26	6.18	21.07
12	30	37	32	21	40	23	12	41	44	24	32	18	14	16	22	24	33	35	43	44	35	46	24	22	30	34	33	45	37	15	30.20	10.11	GEV PDF	-0.53	10.92	27.94
13	24	21	33	15	18	24	22	26	25	34	24	22	30	21	20	17	19	34	31	25	22	25	24	22	21	33	21	40	41	22	25.20	6.48	GEV PDF	0.04	4.78	22.22
14	18	20	25	17	16	20	25	17	19	35	20	19	29	42	13	16	35	23	12	22	35	45	24	20	12	19	34	22	16	17	25.10	8.44	GEV PDF	0.15	5.48	19.01
15	18	12	13	35	20	21	11	12	10	20	12	35	13	19	17	20	15	28	38	24	22	23	38	41	22	18	20	11	14	13	20.30	6.37	GEV PDF	0.20	5.44	15.96
16	30	35	34	32	41	32	41	11	23	35	34	32	42	43	45	42	32	35	16	22	35	35	38	34	31	35	32	35	20	18	32.37	6.62	EV PDF	35.31	7.11	
17	24	34	25	21	33	23	12	21	11	12	19	20	18	25	34	21	25	28	26	28	26	27	23	35	23	65	23	12	24	35	25.03	10.10	GEV PDF	0.01	7.35	20.78
18	18	14	15	30	24	22	12	31	25	28	18	19	24	30	22	24	29	12	15	22	15	11	34	33	21	24	23	25	12	28	[illegible]	[illegible]	GEV PDF	-0.35	6.61	19.80

Ⓐ Section Ⓑ Section Ⓒ Section

4. CASE STUDIES

4.1 Case I

The network shown in Figure 4 was reproduced from Feng et al.'s (1997) and Hegazy's (1999) work to demonstrate the procedure described in the preceding section. It consists of an activity-on-node network composed of 18 activities and a "Start" node which has duration of zero days.. This small network is reused in this case study in order to illustrate the potential of AS4 in the context of a small network that endured rigorous testing in earlier research, albeit for different objectives.

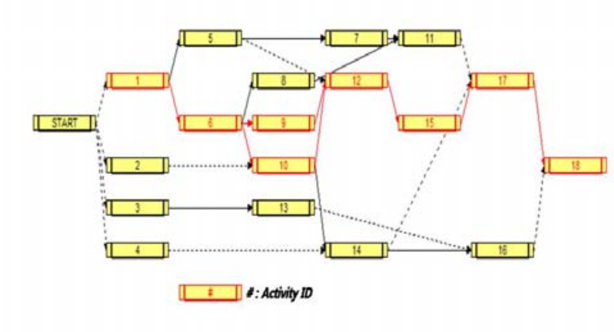

Figure 4. Network for Case I (PERT view of P3)

Using the schedule data exported from P3 and the fictitious historical activity duration data (Section Ⓐ in Table 2) assumed to be sampled from a project data warehouse, the activities' durations are calculated for the four modes as follows; (1) deterministic CPM, (2) probabilistic PERT, (3) normal PDF simulation mode which is based on a normal PDF of activities' durations and PCTs, and (4) best-fit-PDF simulation mode which is based on best-fit-PDF of activities' durations and PCTs. Then, the outputs obtained in the four modes are analyzed by the algorithm shown in Figure 1.

In CPM mode, a set of deterministic durations obtained from a CPM package are assigned to the activities. In PERT mode, the minimum, mean, and maximum values of the historical activity durations are calculated and assigned to respective activities as optimistic, most-likely and pessimistic times. In the normal-PDF simulation mode, the parameters of the normal PDFs which describe the historical activities' durations are estimated and are written in Section Ⓑ of Table 2. In the best-fit-PDF simulation mode, the best-fit-PDFs of the historical activities' durations are estimated and written in Section Ⓒ of Table 2.

Unlike the existing systems that rely mostly on manually operating commercial statistics packages for estimating PDFs, AS4 provides a built-in facility that computes the best-fit-PDFs of the many activities even in the largest of networks in only a few seconds. AS4 automatically executes the modeling and analysis in all the four modes, and perform CPM calculations for 120 iterations (or as many iterations as specified by the program), and obtain as many sets of PCTs.

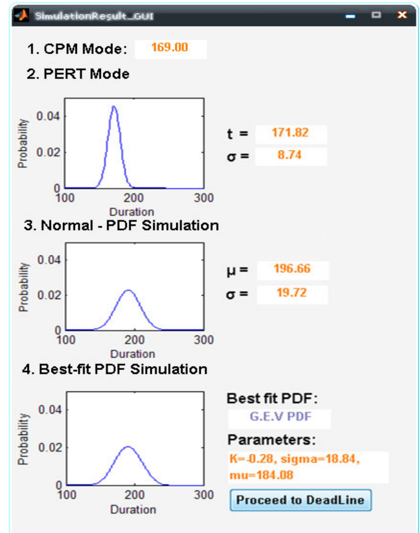

Figure 5. Statistics obtained from the four modes

As shown in Figure 5, the system finds a total project duration of 169 days in CPM mode, (mu = 171.82 days, sigma = 8.74 days) in PERT mode, (mu = 196.66 days, sigma = 19.72 days) in the normal-PDF simulation mode, and (mu = 184.08 days, sigma = 18.84 days, k-value = -0.28) in the best-fit-PDF simulation mode (in this case, Generalized Extreme Value distribution).

Figure 6. Probability to complete the project within a deadline

As shown in Figure 6, the probability to complete the project within 169 days was queried. Table 3 presents the summary results obtained in the four different modes and Figure 7 shows the variability of PDFs obtained in the four modes. The findings indicate that the generalized extreme value (GEV) distribution is the best fitting distribution for the PCTs of the network. Therefore, if one uses the CPM PCT of 169 days as benchmark, the normal-PDF simulation mode and the best-fit-PDF simulation mode lead to probabilities of occurrence of 8% and 11%, respectively. This finding indicates in this case that the normal-PDF simulation mode is 3% more conservative than the best-fit-PDF simulation mode. It takes only 28.12 seconds to obtain the complete set of analysis results in all four modes.

Table 3. The results of Case I obtained by AS4

Modes	The PDF of PCTs	PCTs		Prob. to complete the project within 169 days
		Mean	Std. Dev	
CPM	Deterministic	169 days		100%
PERT	PERT-Beta	172	9	37%
Normality-based Sm.	Normal PDF	197	20	8%
Best-fit-PDF-based Sm.	GEV PDF	184	19	11%

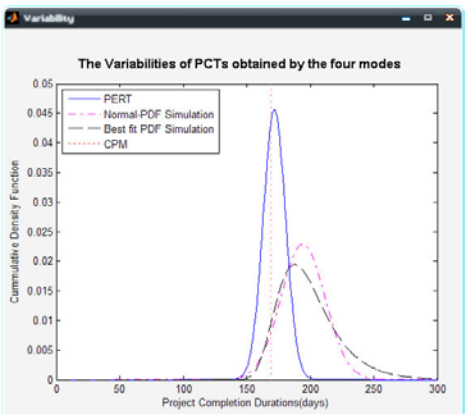

Figure 7. Variability of PCT for Case I

4.2 Case II

The network distributed with the P3 program by Primavera, Inc. is used in Case II to verify the effectiveness of AS4 in dealing with a large network. Case II is a plant expansion and modernization project that is named BASE, and has 134 activities. There are no constraints and lags attached to activities. Only finish-to-start relationships are used. A fictitious historical activity duration table which is similar to Section Ⓐ in Table 2 is prepared and used. The table is not shown for lack of space. However, it is available upon request.

The four modes lead to a project completion time of 415 days in CPM mode, (mu = 420.71 days, sigma = 11.77 days) in PERT mode, (mu = 426.17 days, sigma = 15.21 days) in normal-PDF simulation mode, and (mu = 414.50 days, sigma = 18.05 days, k-value = -0.21) in best-fit-PDF simulation mode (in this case, Generalized Extreme Value distribution). The probabilities to complete the project within the PCT dictated by deterministic CPM (i.e., 415 days in this example) were computed by the four modes as shown in Table 4.

The findings indicate that the generalized extreme value (GEV) distribution is the best fitting distribution for the PCTs of the network. Therefore, if one uses the CPM PCT of 415 days as benchmark, the normal-PDF simulation mode and the best-fit-PDF simulation mode lead to probabilities of occurrence of 23% and 38%, respectively. This finding indicates in this case that the normal-PDF simulation mode is 15% more conservative than the best-fit–PDF simulation mode. It takes only 76.43 seconds to obtain the complete set of analysis results in all four modes.

Table 4. The results of Case II obtained by AS4

Modes	The PDF of PCTs	PCTs		Prob. to complete the project within 169 days
		Mean	Std. Dev	
CPM	Deterministic	169 days		100%
PERT	PERT-Beta	422	12	31%
Normality-based Sm.	Normal PDF	426	15	23%
Best-fit-PDF-based Sm.	GEV PDF	415	18	38%

5. BENEFITS AND LIMITATIONS

The benefits of using AS4 can be summarized as follows:

First, AS4 makes use of schedule data generated by commercially available CPM software that have been used by practitioners for many years. Therefore, AS4's appeal is stronger than existing methods.

Second, AS4 does away with the existing methods' cumbersome and time consuming process. Instead, it introduces a built-in algorithm that automates the process.

Third, AS4 automatically generates the PCT and the probability of completing a project by a deadline only if two tables are provided: (1) schedule data generated by a CPM package and (2) historical activity duration data in spreadsheet format.

Fourth, the system automatically generates results in four modes. This analysis allows the user to compare the schedule risk relative to using different scheduling methods. The system assists the user in calibrating the probability to complete a project more accurately.

The limitations of using AS4 are related to testing and development issues. First, fictitious data are used in the case studies for historical activity durations rather than actual data retrieved from actual project data warehouses. It would be desirable to develop an algorithm that automatically retrieves the historical activity durations from a project data warehouse. Second, no constraints or lags are considered in the case studies and the precedence relationships between activities are assumed to be only of the finish-to-start type. It would be a good addition, if other types of relationships are accommodated in a later version of the system.

6. CONCLUSIONS

AS4 entirely automates the processes relative to estimating the best-fit-PDFs and parameters of the many activities' durations (simulation input data), making use of the schedule data exported from commercially available deterministic CPM software packages in addition to historical duration data obtained from a project data warehouse, and estimating the PDF of PCTs.

The main contribution of this study is the development of an easy-to-use computerized tool named AS4 that is a welcome addition to the field of stochastic simulation-based scheduling. This automated tool is developed by using the facilities of MATLAB. AS4 achieves lateral and vertical integration. Lateral integration is achieved by analyzing a schedule in four different modes, namely deterministic CPM, probabilistic PERT, simulation based on a normal-PDF of PCTs and simulation based on a best-fit-PDF of PCTs. Vertical integration is achieved by performing several operations in an automated process that includes: (1) consolidating schedule data generated by a commercial CPM package and historical activity duration data retrieved from a project data warehouse; (2) estimating the PDFs of historical activity durations using a built-in facility and assigning the PDFs to the activities; (3) simulating the network for many iterations, (4) estimating the PDF of PCTs obtained from the simulation runs, and (5) calculating the probability of completing the project by a given deadline.

This tool reduces the time and effort spent for data processing relative to estimating the best-fit-PDFs of the many activities in a network, converts automatically the schedule data of a large network having many hundreds or even thousands of activities into a format that is appropriate for stochastic schedule simulation, enhances the reliability of network modeling and analysis by estimating the best-fit-PDFs not only of activity durations, but also of PCTs, and computes the probability to complete a project within a specific deadline.

REFERENCES

[1] Ang, A. H-S., and Tang, W. H., *Probability Concepts in Engineering Planning and Design–Analytical models of random phenomena*. Vol. 1, Wiley, New York, NY, 1975.

[2] Barraza, A. G., Back, W. E., and Mata, F., "Probabilistic Forecasting of Project Performance using Stochastic S Curves", *Journal of Construction Engineering and Management*, ASCE, 130(1), 25-32, 2004.

[3] Chaua K.W., Caob, Y., Ansona, M. and Zhangb, J., "Application of data warehouse and Decision Support System in construction management", *Automation in Construction*, 12(2), 213-224, 2003.

[4] Crystal Ball, "Crystal Ball's Critical Path /Time to Market Analysis", www.crystall.com, 2004.

[5] Feng, C., Liu, L., and S., "Using Genetic Algorithms to solve construction time-cost trade-off problems", *Journal of Computing in Civil Engineering*, ASCE, 11(3), 184-189, 1997.

[6] Forrest, W. Y., "Lecture Notes on Statistics and Data Analysis with Vista", Department of Psychology, University of North Carolina, Chapel Hill, NC, http://forrest.psych.unc.edu/research, Accessed 2006.

[7] Grubbs, F. E., "Attempts to Validate Certain PERT Statistics or 'Picking on PERT'", *Operations Research*, 10(6), 912-915, 1962.

[8] Halpin, D. W., *MICROCYCLONE user's manual*. Division of Construction Engineering and Management, Purdue University, West Lafayette, IN., 1990.

[9] Halpin, D. W. and Riggs, L. S., *Planning and analysis of construction operations*, John Wiley & Son, Inc, New York. 270-283, 1992.

[10] Hegazy, T., "Optimization of construction time-cost trade-off analysis using Genetic Algorithms", *Canadian Journal of Civil Engineering*, 26, 685-697, 1999.

[11] Lee, D.-E. and Arditi, D., "Automated Statistical Analysis in Stochastic Project Scheduling Simulation", *Journal of Construction Engineering and Management*, ASCE, 132(3), 268-277, 2006.

[12] Lee, D.-E., "Probability of Project Completion Using Stochastic Project Scheduling Simulation (SPSS)", *Journal of Construction Engineering and Management,*, ASCE, 131(3), 310-318, 2005.

[13] Lu, M. and AbouRizk, S. M., "Simplified CPM/PERT Simulation Model", *Journal of Construction Engineering and Management*, ASCE, 126(3), 219-226, 2000.

[14] Martinez, J. C. and Ioannou P.G., "State-Based Probabilistic Scheduling Using Stroboscope's CPM Add-On", *Proceedings, Construction Congress V*, 438-445, Minneapolis, MN, Oct 4-8, 1997.

[15] Monte Carlo for Primavera® (Accessed 2008), http://www.primavera.com.

[16] Palisade (Accessed 2008) Palisade @RISK for Project - advanced risk analysis for project management, http://www.palisade.com/reviews/rprj_or.asp.

[17] Rujirayanyong, T. and Shi, J. J., "A project-oriented data warehouse for construction". *Automation in Construction*, 15(6), 800-807, 2006.

[18] Sculli, D., "The Completion Time of PERT Networks", *Journal of the Operational Research Society*, 34(2), 155-158, 1983.

[19] Shi, J. J., "Activity-based construction (ABC) modeling and simulation method", *Journal of Construction Engineering and Management*, ASCE, 125(5), 354-360, 1999.

[20] Vose, D., *Quantitative Risk Analysis: A Guide to Monte Carlo Simulation Modeling.* John Wiley and Sons, New York, NY., 1996.
[21] Vose, D., *Risk Analysis: a Quantitative Guide*, Second edition, John Wiley & Sons, New York, NY., 2000.
[22] Yao, M. J. and Chu, W. M., "A new approximation for obtaining the probability distribution function for project completion time", *Computers and Mathematics with Applications*, 54(2), 282-295, 2007.