

Computation of Temporal Uncertainties in Construction Project Management Using Imprecise Probability

Mehdi Modares^{1*}, Michael Desch¹, Rafi L. Muhanna², and Robert L. Mullen³

¹*Department of Civil, Architectural and Environmental Engineering, Illinois Institute of Technology, USA,*

²*School of Civil and Environmental Engineering, Georgia Institute of Technology, USA*

³*Department of Civil and Environmental Engineering, University of South Carolina, USA*

**Corresponding author: mmodares@iit.edu*

Abstract. In construction project management, scheduling plays a crucial role in a project's success. Construction scheduling and monitoring schemes are used to predict the project's total duration. For this prediction, the Critical Path Method (CPM) and the Program Evaluation and Review Technique (PERT) are widely used. These methods predict the overall project's completion time based on the duration of each task and the allowable float in each task's starting time thru a network of tasks.

However, there exist variations and uncertainties in the duration of each event. As such, deterministic construction scheduling schemes are incapable of quantifying and computing these uncertainties. Although precise probability approaches are utilized for the enumeration and analysis of uncertainties in construction scheduling, their accuracy highly depends on the availability of significant data. Due to limitations of the available data, those precise probability approaches may yield erroneous results.

In this work, a new forward-backward formulation for CPM and PERT is introduced, through which, uncertainties are defined based on the concepts of imprecise probability. To represent the imprecise probability structures, a probability box (P-box) formulation is utilized. Thru the network, in the forward formulation, the uncertainty in the duration of each task is represented by an independent P-box, leading to a determination of the resulting P-box for the project's total time. Conversely, in the backward path, given the P-box for the project's total time, the float time P-box for each task is determined. The forward path calculation is performed using P-box arithmetic; whereas, the backward path calculation is achieved through solving an inverse problem. For illustrating the applicability of this method, an example problem is presented and compared with conventional CPM, PERT, and interval approaches.

Keywords: Imprecise Probability, Construction Scheduling, Critical Path Method

1. Introduction

A network for planning, scheduling, and monitoring construction and industrial projects is based on a set of activities, with prescribed durations, that constructs a directed acyclic graph (Fortin et al. 2010). The Critical Path Method (CPM) is a widely accepted method utilized to determine the characteristics of these networks. Using CPM, the total duration of a project, as well as the project's critical activities, are determined.

The numerical calculations performed in conventional CPM are deterministic. However, if there are uncertainties in the duration of activities, the conventional CPM scheme does not quantify or compute their effects. Methods that consider the uncertainties in scheduling are mainly based on probabilistic approaches (Abourizk et al. 1991, 1994; Ahuja et al. 1994). As such, they require sufficient information to choose the Probability Density Function (PDF) for each uncertain duration. For example, in Program Evaluation and Review Technique (PERT), the duration of each activity is defined in terms of the minimum, expected, and maximum durations (Moder et al. 1983; Cottrell 1999). PERT is performed presuming the duration of each activity is related to the Beta distribution. Additionally, possibilistic methods are used to enumerate the uncertainties in scheduling. For example, in the fuzzy approach to scheduling, the durations of activities are defined as fuzzy variables with predefined membership functions (Lessmann et al. 1994; Fetz et al. 1999; Lu and AbouRizk 2000).

The selection of a probability distribution for the duration of each activity needs inter-connected, multi-activity data. Alternatively, if sufficient data to determine these distributions are unavailable, a range of probability distributions that represent the possible activity durations may be considered. Imprecise probability theory is one of the approaches for considering uncertainty in a system without the assumption of a prescribed PDF. Imprecise probability theory is useful when the available information about a system is not sufficient enough such that the PDF of input parameters cannot be accurately chosen. Contrary to fuzzy set theory, using the imprecise probability approach can sharpen the description of its parameters as more information becomes available.

In this work, a method is developed that is capable of treating temporal uncertainties in project scheduling. Imprecise probability theory is used to describe the duration of each activity in a project. As input parameters, the duration of each activity is defined as a Probability-box (P-box). These are sets of cumulative distribution functions (CDFs), defined by the upper and lower bounding CDFs. For the forward path analysis thru the network, direct P-box arithmetic is used. Conversely, for the backward path thru the network, the float times for each task are calculated based on the solution of an inverse-like problem.

2. Background

This section provides the background information that is foundational to the method developed in this work. This includes overviews of the CPM, PERT, and P-box based Imprecise Probability.

2.1. CRITICAL PATH METHOD

The critical path method is an approach to project scheduling that identifies: 1) the total project duration, 2) a 'critical path' of activities that cannot be delayed without increasing the total project duration, 3) the earliest and latest times each activity can be scheduled for without increasing the total project duration (early start/finish and late start/finish), 4) how long each activity can be delayed without affecting other activities (free float), and 5) how long each activity can be delayed without affecting the critical path (total float). To perform this method, first, a list of project tasks is developed, and it is determined which tasks are reliant upon each other. A network of the activities is then developed. Activities are considered nodes and the links between nodes represent the reliance between tasks. The duration of each task is also determined. The following steps are followed to perform the analysis (Hinze 2012):

Forward Pass

- The early start (ES) time of each starting node, ES_S , is set as 0
 $ES_S = 0$
- The early finish (EF) time of each starting node, EF_S , is set as the early start time plus the duration (Dur) of the starting node, Dur_S .
 $EF_S = ES_S + Dur_S$
- The early start time of all other nodes in the network is defined as the maximum value of the early finish time of all directly preceding nodes.
 $ES = \max(EF \text{ of each predecessor})$
- The early finish of each node is then defined as the early start plus the duration of the node.
 $EF = ES + Dur$

Backward Pass

- A lag value is calculated for each link in the network. For a link showing node A precedes node B, the lag can be calculated as the early start of node B minus the early finish of node A
 $L_{A \rightarrow B} = ES_B - EF_A$
- The free float (FF) of each node is then calculated as the minimum value of the lags of each of the links leaving the node.
 $FF = \min(\text{lag of each link leaving node})$
- The free float and total float (TF) of the end node is set to 0.
 $FF_E = 0 \text{ and } TF_E = 0$
- The total float of all other nodes is defined as the lag to a decedent node plus the free float of the descendent node. If a node has multiple descendants, then the minimum sum is taken.
 $TF = \min(\text{Lag to descendant} + FF \text{ of descendant})$
- The late start (LS) time of a node is then calculated as the early start plus the total float
 $LS = ES + TF$
- The late finish (LF) time of a node is then calculated as the early finish plus the total float
 $LF = EF + TF$

In this approach, the end of day convention is used where all values represent the end of a day. Therefore, the early start time of the starting node is set to zero, $ES_S = 0$. The total project duration time will be the finish day (EF or LF of the end node). The critical path is found as the path from the network start to end with zero total float. At least one such path will exist. If any task on this path is delayed or extended, the project time will increase.

2.2. PROGRAM EVALUATION AND REVIEW TECHNIQUE

Program Evaluation and Review Technique (PERT) is also used for project scheduling. PERT has a similar approach to the CPM; however, using PERT, the duration of each activity is defined in terms of the minimum, expected and maximum durations.

PERT has an underlying probabilistic assumption related to the Beta distribution. The minimum, expected, and maximum durations for each activity are used to calculate the probabilistic parameters for each activity. The mean (μ_{PERT}) and standard deviation (σ_{PERT}) are calculated as:

$$\mu_{PERT} = \frac{\text{minimum} + 4 \times \text{expected} + \text{maximum}}{6} \quad (1)$$

$$\sigma_{PERT} = \frac{\text{maximum} - \text{minimum}}{6} \quad (2)$$

2.3. IMPRECISE PROBABILITY STRUCTURES BASED ON P-BOXES

2.3.1. Continuously Bounded P-boxes

Define the CDF for the random variable X as $F(x)$. If the distribution parameters are uncertain, for every x , an interval $[\underline{F}(x), \overline{F}(x)]$ generally can be constructed that is capable of bounding the possible values of $F(x)$, therefore, $\underline{F}(x) \leq F(x) \leq \overline{F}(x)$. The two CDFs $\underline{F}(x)$ and $\overline{F}(x)$ are known as "probability bounds" (Williamson 1990, Ferson 2003). Figure 1 shows a continuous probability box schematically.

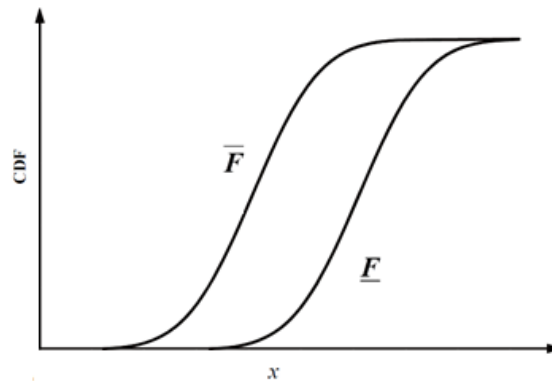


Figure 1. A general continuous probability box

2.3.2. Discrete Bounded P-boxes

Alternatively, the bounded P-boxes can be constructed by the use of discrete, interval based P-box structures. This discrete P-box structure includes a collection of interval values, with corresponding associated probability. Figure 2 shows a uniformly discretized P-box that encloses the original continuous P-box schematically.

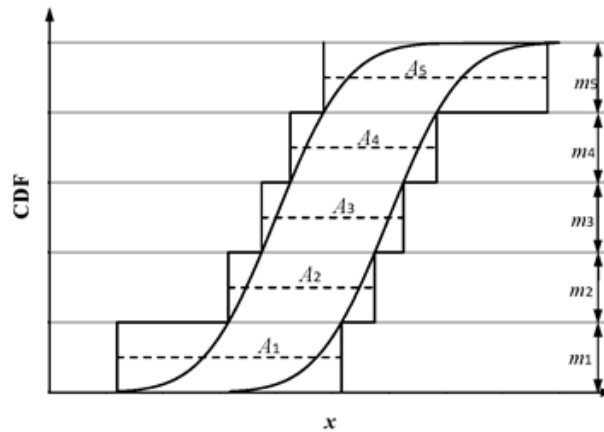


Figure 2. A general uniformly discretized probability box (Enclosure Discretization)

3. Methodology

3.1. FORMULATION OF IMPRECISE PROBABILITY PROJECT SCHEDULING

The general algorithm for Imprecise Probability Project Scheduling (IPPS), which is based on the CPM, is given below.

1. Determine the imprecise probability structure, based on P-boxes, for each activity. For each activity:
 - a) Construct an independent imprecise probability structure for the uncertain duration.
 - b) Construct a P-box structure for the uncertain duration.
2. Perform forward pass and backward pass analyses on the network using the constructed P-box structures for the uncertain duration of each activity using either a combinatorial or simulation approach.
 - a) In the forward pass, for each activity, determine the P-box structures of 1) the early start time and 2) the early finish time by using P-box addition.
 - b) In the backward pass, for each activity, determine the P-box structures of 1) the lag time(s) associated with the activity, 2) the free float time, 3) the total float time, 4) the late start time, and 5) the late finish time.
3. Identify the critical path(s) as the path(s) from the start through the end of the network where every activity has zero total float.
4. Identify the P-box structure for the total project duration.

3.2. IPPS COMPUTATIONS PROCEDURE

Using IPPS, the interval-based computations for P-box structures are performed using interval arithmetic operations for all combinations of intervals corresponding to the P-boxes. For one combination and considering two P-box structures \mathbf{X} and \mathbf{Y} with intervals $\mathbf{X}_i = [\underline{x}_i, \overline{x}_i]$ and $\mathbf{Y}_j = [\underline{y}_j, \overline{y}_j]$ corresponding to their i and j discretization levels, respectively:

1. In the forward path for addition:

$$\mathbf{X}_i + \mathbf{Y}_j = [\underline{x}_i, \overline{x}_i] + [\underline{y}_j, \overline{y}_j] = [\underline{x}_i + \underline{y}_j, \overline{x}_i + \overline{y}_j] \quad (3)$$

2. In the backward path for subtraction :

$$\mathbf{X}_i - \mathbf{Y}_j = [\underline{x}_i, \overline{x}_i] - [\underline{y}_j, \overline{y}_j] = [\underline{x}_i - \overline{y}_j, \overline{x}_i - \underline{y}_j] \quad (4)$$

As the backward path is an inverse problem of the forward path, the input and output interval variables are dependent. As such, the application of Eq. 4 may yield results with overestimation. To reduce this overestimation, the interval variables are defined based on their midpoints and radii, as:

$$\mathbf{X}_i = \text{mid}(\mathbf{X}_i) + \text{rad}(\mathbf{X}_i) \times \boldsymbol{\varepsilon} \quad (5)$$

$$\mathbf{Y}_j = \text{mid}(\mathbf{Y}_j) + \text{rad}(\mathbf{Y}_j) \times \boldsymbol{\varepsilon} \quad (6)$$

in which, $\boldsymbol{\varepsilon} = [-1, 1]$. Therefore, Eq. 4 can be rewritten as:

$$\mathbf{X}_i - \mathbf{Y}_j = \text{mid}(\mathbf{X}_i) - \text{mid}(\mathbf{Y}_j) + (\text{rad}(\mathbf{X}_i) - \text{rad}(\mathbf{Y}_j)) \times \boldsymbol{\varepsilon} \quad (7)$$

in which, the overestimation is reduced because of the single representation of $\boldsymbol{\varepsilon}$ in Eq. 7.

IPPS Summary

In IPPS, the interval solution approach is used to compute the P-box results of the CPM procedure. An interval is chosen from each of the input duration P-boxes, and the network is solved. Eqs. 3 and 7 are used where additions and subtractions are needed. This gives each network output as an interval value. All possible combinations of interval network results are computed and then used to formulate each output as a P-box.

4. Numerical Example

4.1. PROBLEM DEFINITION

The numerical example, based on a model from Chu (2008), is used to illustrate the developed method. This model, a network with seven tasks and three parallel paths, is defined by Table I. Each task has three durations attributed to it, an optimistic (minimum) time to completion, an expected (mean) time to completion, and a pessimistic (maximum) time to completion.

Table I. Network Definition				
<i>Task</i>	<i>Minimum Duration</i>	<i>Expected Duration</i>	<i>Maximum Duration</i>	<i>Preceding Task(s)</i>
A	7	10	16	-
B	4	5	6	-
C	26	30	50	A
D	35	40	58	B
E	8	12	14	B
F	30	36	48	C and D
G	8	9	12	E

Figure 3 depicts this network schematically.

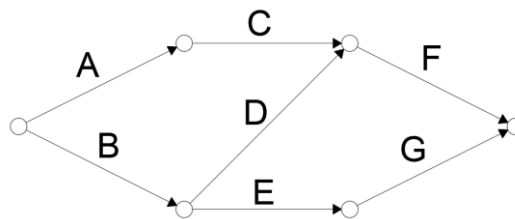


Figure 3. Network Graph, as defined by Chu (2008), with Tasks on Links

4.2. PROBLEM SOLUTION

The network graph in Figure 3 depicts the tasks on the links. As the algorithm used for the developed method places tasks on the nodes, Figure 4 is drawn to depict the same network but with tasks on the nodes. A dummy “END” node with zero duration is added to satisfy the requirement that the network terminates on a single ending node.

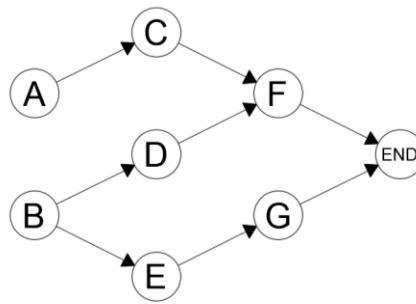


Figure 4. Network Graph – Tasks on Nodes

The network is solved using the following methods:

1. CPM - The duration of each task is the expected duration. $Dur_i = expected_i$
2. PERT - The duration of each task is the PERT mean. $Dur_i = \mu_{PERT_i}$
3. IPPS - The duration of each task is a normal distribution with mean and standard deviation equal to the PERT mean and PERT standard deviation, respectively $Dur_i = N(\mu_{PERT_i}, \sigma_{PERT_i})$.
4. Interval Analysis - The duration of each task is the interval constructed between the task's minimum duration and maximum duration $Dur_i = [min_i, max_i]$.

IPPS Analysis

In the IPPS method, for each task, no variation is considered in either the mean or standard deviation, $\bar{F}(x)$ coincides with $\underline{F}(x)$. The imprecise probability structure of each activity is discretized into ten P-box intervals with equal probability mass. A combinatorial analysis (10^7 combinations) is performed (seven-node network with each node with ten P-box intervals). As the normal CDF function is continuous, the tails are truncated at $P = 0.005, P = 0.995$ in order to establish finite bounds for the lower-most and upper-most P-box intervals. Figure 5 shows the seven P-box interval envelopes for the imprecise probability durations.

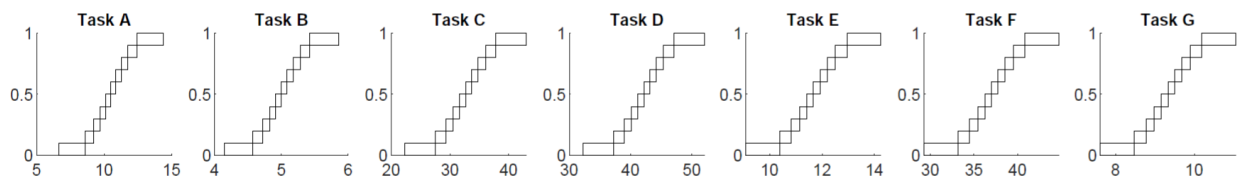


Figure 5. P-box durations

4.3. ANALYTICAL RESULTS

The problem is solved using the four analysis schemes and the detailed results are presented in Table II. The results displayed for the IPPS is the interval corresponding to a probability $P = 0.5$.

Computation of Temporal Uncertainties in Construction Project Management Using Imprecise Probability

TABLE II. Results for each task, comparing CPM, PERT, IPPS, and Interval approaches

<i>Task</i>	<i>Approach</i>	<i>Duration</i>	<i>Early Start</i>	<i>Early Finish</i>	<i>Free Float</i>	<i>Total Float</i>	<i>Late Start</i>	<i>Late Finish</i>
A	CPM	10	0	10	0	5	5	15
	PERT	10.50	0.00	10.50	0.00	4.00	4.00	14.50
	IPPS	[10.50, 10.50]	[0.00, 0.00]	[10.50, 10.50]	[0.00, 0.00]	[3.67, 4.33]	[3.67, 4.33]	[13.83, 15.17]
	Interval	[7, 16]	[0, 0]	[7, 16]	[0, 0]	[0, 6]	[0, 6]	[7, 22]
B	CPM	5	0	5	0	0	0	5
	PERT	5.00	0.00	5.00	0.00	0.00	0.00	5.00
	IPPS	[5.00, 5.00]	[0.00, 0.00]	[5.00, 5.00]	[0.00, 0.00]	[0.00, 0.00]	[0.00, 0.00]	[5.00, 5.00]
	Interval	[4, 6]	[0, 0]	[4, 6]	[0, 0]	[0, 0]	[0, 0]	[4, 6]
C	CPM	30	10	40	5	5	15	45
	PERT	32.67	10.50	43.17	4.00	4.00	14.50	47.17
	IPPS	[32.67, 32.67]	[10.50, 10.50]	[42.33, 44.00]	[3.67, 4.33]	[3.67, 4.33]	[13.83, 15.17]	[46.31, 49.76]
	Interval	[26, 50]	[7, 16]	[33, 66]	[0, 6]	[0, 6]	[7, 22]	[33, 72]
D	CPM	40	5	45	0	0	5	45
	PERT	42.17	5.00	47.17	0.00	0.00	5.00	47.17
	IPPS	[42.17, 42.17]	[5.00, 5.00]	[46.62, 47.71]	[0.00, 0.00]	[0.00, 0.00]	[5.08, 5.17]	[47.17, 49.00]
	Interval	[35, 58]	[4, 6]	[39, 64]	[0, 2]	[0, 2]	[4, 8]	[39, 66]
E	CPM	12	5	17	0	55	60	72
	PERT	11.67	5.00	16.67	0.00	58.17	63.17	74.83
	IPPS	[11.67, 11.67]	[5.00, 5.00]	[16.49, 16.84]	[0.00, 0.00]	[58.02, 60.09]	[62.93, 65.18]	[74.36, 77.08]
	Interval	[8, 14]	[4, 6]	[12, 20]	[0, 0]	[49, 82]	[53, 88]	[61, 102]
F	CPM	36	45	81	0	0	45	81
	PERT	37.00	47.17	84.17	0.00	0.00	47.17	84.17
	IPPS	[37.00, 37.00]	[47.19, 48.75]	[83.54, 86.60]	[0.00, 0.00]	[0.00, 0.00]	[47.19, 48.75]	[83.54, 86.60]
	Interval	[30, 48]	[39, 66]	[69, 114]	[0, 0]	[0, 0]	[39, 66]	[69, 114]
G	CPM	9	17	26	55	55	72	81
	PERT	9.33	16.67	26.00	58.17	58.17	74.83	84.17
	IPPS	[9.33, 9.33]	[16.49, 16.84]	[25.61, 26.39]	[58.02, 60.09]	[58.02, 60.09]	[74.36, 77.08]	[83.53, 86.60]
	Interval	[8, 12]	[12, 20]	[20, 32]	[49, 82]	[49, 82]	[61, 102]	[69, 114]
END	CPM	0	81	81	0	0	81	81
	PERT	0.00	84.17	84.17	0.00	0.00	84.17	84.17
	IPPS	[0.00, 0.00]	[83.54, 86.60]	[83.54, 86.60]	[0.00, 0.00]	[0.00, 0.00]	[83.54, 86.60]	[83.54, 86.60]
	Interval	[0, 0]	[69, 114]	[69, 114]	[0, 0]	[0, 0]	[69, 114]	[69, 114]

Figure 6 shows the P-box output for project duration calculated as the Early Finish of task F.

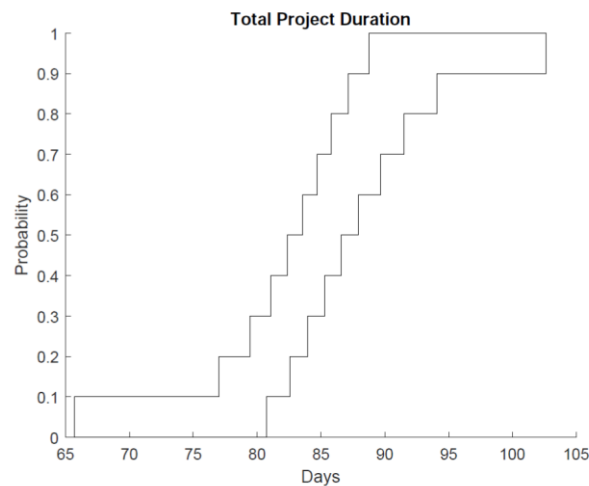


Figure 6. P-box output for project total duration (early finish of activity F)

4.4. OBSERVATIONS

For the CPM, PERT, and IPPS, path B-D-F is immediately identified at the critical path. In the interval approach, it can be seen that node D has the possibility of some total float. This indicates that with no assumption on the distribution of the uncertainty, there is sufficient uncertainty in the network tasks to have multiple possible critical paths.

It can be seen that in every case in Table 2 the interval results enclose the IPPS results. As well, the IPPS results enclose the deterministic PERT results. This outcome is expected as imprecise probabilities make fewer presumptions than deterministic values, and intervals make no presumptions within their bounds at all. Differences between the interval bounds and the first/last p-box bounds can be observed (as seen in Figure 6). This can be accounted for as the effect of the difference between the interval bounds and the truncation points of the imprecise probability CDFs. The truncations at $P = 0.005, 0.995$ correspond to ± 2.576 standard deviations, while the interval solution corresponds to ± 3 standard deviations.

4.5. COMPUTATION TIME

The network results are obtained using MATLAB R2019a running on a processor clocked at 1.70GHz with 8Gb of memory. The computation time for CPM was 0.02 sec, for PERT was 0.02 sec, for IPPS was 918.25 sec, and for the interval method was 0.03 sec. The significantly longer computation time for IPPS is attributed to the combinatorial solution. Using a simulation approach for IPPS will significantly reduce the computational effort.

5. Summary and Conclusions

In this work, a new formulation for construction project scheduling for a network with temporal uncertainty based on the concepts of imprecise probability (called IPPS) is introduced. IPPS can obtain the P-box structures for the early start, early finish, free float, total float, late start, and late finish for each activity as well as the critical path and the overall project duration. The results obtained from IPPS have a higher level of confidence and robustness because of the objective determination of uncertainties in the parameter distributions. Although using the combinatorial approach in IPPS method yields exact results, this approach can be extremely computationally expensive. Therefore, for larger networks, using a simulation approach in IPPS is more desirable. The IPPS capability to more robustly enumerate uncertainties makes it attractive to introduce imprecise probability concepts in the field of construction project scheduling and management.

References

- Abourizk, S.M., D. W. Halpin, and J. R. Wilson. "Visual interactive fitting of beta distributions", *Journal of Construction Engineering and Management*, 117, pp. 589-605. (1991).
- Abourizk, S.M., D. W. Halpin, and J. R. Wilson, "Fitting beta distributions based on sample data", *Journal of Construction Engineering and Management*, 120, pp. 288-305. (1994).
- Ahuja, H.N., S. P. Dozzi, and S. M. Abourizk, "Project management: techniques in planning and controlling construction projects", John Wiley, (1994).
- Chu, C. "Analysis of effect factors-based stochastic network planning model," *Journal of Systems Engineering and Electronics*, 19(1), pp. 112-118, (2008), DOI: 10.1016/S1004-4132(08)60054-7.
- Cottrell, W. D. "Simplified program evaluation and review technique (PERT)", *Journal of Construction Engineering and Management* 125, pp. 16-22 (1999).
- Ferson, S., V. Kreinovich, L. Ginzburg, and F. Sentz. "Constructing probability boxes and Dempster-Shafer structures" (2003).
- Fetz, T., M. Oberguggenberger, J. Jager, D. Koll, G. Krenn, H. Lessmann, and R.F. Stark. "Fuzzy Models in Geotechnical Engineering and Construction Management," *Computer-Aided Civil and Infrastructure Engineering*, 14 pp. 93-106. (1999).
- Fortin, J., P. Zielinski, D. Dubois, and H. Fargier. "Criticality analysis of activity networks under interval uncertainty" *J. Sched*, 13 pp. 609-627. (2010), DOI: 10.1007/s10951-010-0163-3.
- Hinze, J.W. *Construction Planning and Scheduling*. 4th Ed. Prentice Hall, Upper Saddle River, NJ. (2012).
- Lessmann, H., J. Muhlogger, and M. Oberguggenberger. "Netzplantechnik mit unscharfen Methoden (Network techniques with fuzzy methods)" *Bauingenieur*, 69 pp. 469-478. (1994).
- Lu, M. and S. M. Abourizk, "Simplified CPM/PERT simulation model", *Journal of Construction Engineering and Management* 126, 219-226 (2000).
- Moder, J.J., C. R. Phillips, and E. W. Davis. "Project management with CPM, PERT and precedence diagramming" Van Nostrand Reinhold, (1983).
- Williamson, R. and T. Downs. "Probabilistic arithmetic: numerical methods for calculating convolutions and dependency bounds", *International Journal of Approximate Reasoning*, 4 pp. 89-158 (1990).