Fuzzy Numbers in Cost Range Estimating

Ahmed A. Shaheen, Ph.D.[1]; Aminah Robinson Fayek[2]; and S. M. AbouRizk[3]

Abstract: Range estimating is a simple form of simulating a project estimate by breaking the project into work packages and approximating the variables in each package using statistical distributions. This paper explores an alternate approach to range estimating that is grounded in fuzzy set theory. The approach addresses two shortcomings of Monte Carlo simulation. The first is related to the analytical difficulty associated with fitting statistical distributions to subjective data, and the second relates to the required number of simulation runs to establish a meaningful estimate of a given parameter at the end of the simulation. For applications in cost estimating, the paper demonstrates that comparable results to Monte Carlo simulation can be achieved using the fuzzy set theory approach. It presents a methodology for extracting fuzzy numbers from experts and processing the information in fuzzy range estimating analysis. It is of relevance to industry and practitioners as it provides an approach to range estimating that more closely resembles the way in which experts express themselves, making it practically easy to apply an approach.

DOI: 10.1061/(ASCE)0733-9364(2007)133:4(325)

CE Database subject headings: Fuzzy sets; Cost estimates; Monte Carlo method; Construction management.

Introduction

Range estimating is one of the most commonly used forms of Monte Carlo simulation in construction practice. This technique is a simple form of simulating a project estimate by breaking the project into work packages and approximating the variables in each package using statistical distributions. During simulation, these distributions are sampled, and a total project cost is aggregated and statistically analyzed to derive proper cost indicators with various probabilities of achieving them. Such an approach is useful in quantifying uncertainties with high risk work packages, thus leading to better decisions regarding the project budget.

This paper explores an alternate approach to range estimating that is grounded in fuzzy set theory. The approach addresses two shortcomings of Monte Carlo simulation. The first is related to the analytical difficulty associated with fitting statistical distributions to subjective data, and the second relates to the required number of simulation runs to establish a meaningful estimate of a given parameter at the end of the simulation.

Fuzzy set theory enables us to subjectively elicit information about parameters of interest in each work package from an estimator without having to sacrifice accuracy, as such elicitation is a cornerstone of fuzzy set theory (as opposed to approximating a statistical distribution from subjective information). Fuzzy set theory requires only one pass of calculations to establish an estimate of the parameter of concern, as opposed to the multiple runs required in Monte Carlo simulation. This paper demonstrates that comparable results to Monte Carlo simulation can be achieved using the fuzzy set theory approach. In addition, it presents a methodology for extracting fuzzy numbers from experts and processing the information in fuzzy range estimating analysis.

Modeling Range Estimating Using Monte Carlo Simulation

Ahuja et al. (1994) define range estimating as a simulation modeling process performed after an estimate is made, e.g., an estimate of duration or cost, to reflect the degree of uncertainty associated with an estimate. Using cost range estimating as an example, the process can be summarized as follows:

1. Identify the major work components in the form of major cost packages and their related subcategories, which can be restricted to the major items that affect the total cost bottom line by a certain percentage;
2. Identify the uncertain items (e.g., cost items whose values are not deterministic);
3. Use statistical distributions (i.e., triangular and/or uniform distributions) to model the variability of each uncertain item;
4. Use Monte Carlo simulation to provide the final outputs; and
5. Collect statistics on the mean, standard deviation, and minimum/maximum values of the output.

The Monte Carlo simulation technique has been used to model the risk analysis of cost and time in construction applications. Using this technique, the modeled system takes inputs in the form of random variables. The process continues by performing experiments with many variations of the input and then collects sets of outputs in the form of statistical distributions, which are analyzed

[1]Project Engineer, Lockerbie & Hole, Inc., 53367 Range Rd. 232, Sherwood Park AB, Canada T8A 4V2. E-mail: ashaheen@lockerbiehole.com

[2]Professor, Dept. of Civil and Environmental Engineering, Hole School of Construction Engineering, 3-013 Markin/CNRL, NREF, Univ. of Alberta, Edmonton AB, Canada T6G 2W2 (corresponding author). E-mail: aminah.robinson@ualberta.ca

[3]Professor, Dept. of Civil and Environmental Engineering, Hole School of Construction Engineering, 3-014 Markin/CNRL, NREF, Univ. of Alberta, Edmonton AB, Canada T6G 2W2. E-mail: abourizk@ualberta.ca

Note. Discussion open until September 1, 2007. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on June 12, 2006; approved on November 13, 2006. This paper is part of the ***Journal of Construction Engineering and Management***, Vol. 133, No. 4, April 1, 2007. ©ASCE, ISSN 0733-9364/2007/4-325–334/$25.00.

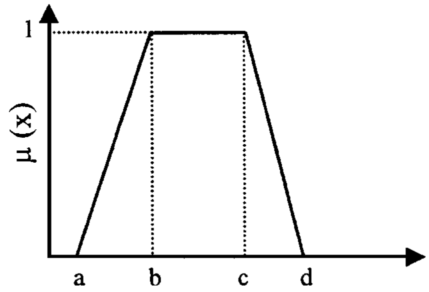

Fig. 1. Figure 1 is a trapezoidal fuzzy number

to provide the measure of uncertainty and risk. The main steps followed in Monte Carlo simulation are (Ahuja et al. 1994):

1. Generate reproducible random numbers;
2. Generate random variates reflecting the true nature of the modeled item (i.e., duration and cost estimate). This step is referred to as "input modeling," which requires the modeled item be modeled by an appropriate probabilistic distribution that best represents the item;
3. Run the model and calculate the desired output parameters;
4. Repeat Steps 1 to 3 for a large number of iterations; and
5. Terminate after a specified number of iterations, and analyze the collected output statistics.

Range estimating can be an effective tool for modeling the uncertainty inherent in cost or duration estimates. However, this technique is based entirely on probabilistic and statistical modeling techniques that model the randomness of the problem. In addition, the process requires a large number of iterations in order to reach a reliable output. Selecting the appropriate statistical distribution that best models the inputs is an important issue in random modeling.

Modeling Uncertainty Using Fuzzy Numbers

Introduction to Fuzzy Numbers

Fuzzy set theory deals with a set of objects characterized by a membership (characteristic) function that assigns to each object a grade of membership ranging between zero (no membership) and one (full membership) (Zadeh 1965). The concept of "unsharp boundaries" that fuzzy set theory represents mimics the human way of thinking, which works with shades of gray rather than black and white. A fuzzy number is a fuzzy membership function that is both convex and normal. Fuzzy numbers are written in the form of a domain value and its corresponding confidence level. Fig. 1 shows an example of a trapezoidal convex and a normal fuzzy number. A fuzzy number can be considered a generalization of the concept of interval of confidence. Therefore, the mathematical operations of fuzzy numbers (i.e., addition, subtraction, multiplication, and division) can be processed using the concepts of the interval of confidence.

Fuzzy Numbers versus Probability Distributions

Fuzziness and probability represent different aspects of uncertainty. According to Kaufmann and Gupta (1985), a fuzzy number is not a random variable. The random variable is defined in terms of the theory of probability, whereas a fuzzy number is a subjective datum that is defined by the theory of fuzzy sets. In order to show the difference between the two types of uncertainty, the summation operation is taken as an example. Fig. 2 shows that the distribution of summation of probability distributions will eventually achieve a Gaussian shape (normal) that obeys the Laplace-Gauss central limit theorem. On the other hand, the addition of two fuzzy numbers that are similar in shape will result in a larger fuzzy number that maintains the same shape. This example shows how the two approaches differ in the way they process uncertainty.

According to Ferson (2002), some of the disadvantages of Monte Carlo methods are computational burden, sensitivity to uncertainty about input distribution shapes, and the need to assume correlations among all inputs. On the other hand, modeling uncertainty using fuzzy arithmetic is computationally simple, not very sensitive to moderate changes in the shapes of input distributions, and does not require the analyst to assume particular correlations among inputs. However, the results generated by fuzzy arithmetic are conservative and may overestimate uncertainty. As illustrated in Fig. 2, the summation operation of two fuzzy numbers generated an output that is wider in range.

Fuzzy set theory can be used as an effective alternative to the random modeling of uncertainty. It is a very attractive alternative because it is more capable of extracting and representing the required information from experts by effectively capturing their linguistic and subjective evaluations. In addition, the calculations involved are much easier and faster compared to the probabilistic approach.

In order to minimize the effect of overestimation in fuzzy

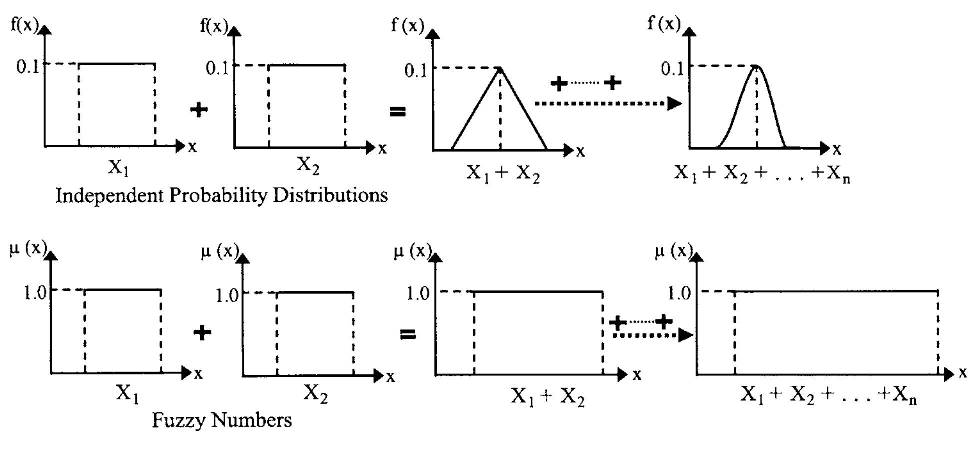

Fig. 2. Comparison between probability and fuzzy convolution

modeling, this paper will introduce concepts that must be considered when modeling uncertainty using fuzzy arithmetic. The following sections present some of the concepts that need to be incorporated and how they will be used to model uncertainty in cost range estimating.

Crisp Representation of a Fuzzy Number

A fuzzy number can be defined by a crisp quantity that represents the "defuzzified" or "expected" value of the fuzzy number. As indicated earlier, fuzzy numbers are a generalization of the concept of the interval of confidence. As we are dealing with a number represented by an interval, ranking this number is not a straightforward process. Therefore, calculating the "expected value" of the fuzzy number will render the fuzzy number ranking and comparison much easier. Different methodologies have been developed to capture an expected value of a fuzzy number.

One of the most common defuzzification methods is the Center of area (COA), which is calculated as

$$y^* = \frac{\int x\mu_i(x_i)}{\int \mu_i(x_i)} \tag{1}$$

where y^*=defuzzified value; $\mu_i(x)$=aggregated membership function; and x=output variable. Eq. (1) represents the centroid of the fuzzy number.

In probability theory, the mean value and variance correspond, respectively, to the centroidal distance and central moment of inertia of an area (Ang and Tang 1975). The centroidal distance (x_0) of a unit area is calculated by

$$x_0 = \frac{\int_{-\infty}^{\infty} xf(x)dx}{\text{area}} = \int_{-\infty}^{\infty} xf(x)dx \tag{2}$$

Eq. (2) is also the first moment (about 0) of the irregular-shaped area. The moment of inertia about the vertical centroidal axis (I_y) is

$$I_y = \int_{-\infty}^{\infty} (x-x_0)^2 f(x)dx \tag{3}$$

In probability, the mean value of a continuous random variable X with a probability density function $f_X(x)$ is

$$E(X) = \int_{-\infty}^{\infty} xf_X(x)dx \tag{4}$$

The variance of a continuous random variable X with a probability density function $f_X(x)$ is:

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)dx \tag{5}$$

Comparing Eqs. (2) and (3) with (4) and (5), respectively, the mean value is equivalent to the centroidal distance, and the variance is equivalent to the centroidal moment of inertia of an area (Ang and Tang 1975).

Therefore, as the probabilistic mean value is equivalent to the centroidal distance of an area, the COA defuzzification method of any fuzzy number will generate a defuzzified value that is identical to the probabilistic mean of the normalized fuzzy number. Therefore, when a fuzzy number maintains the same range and the same shape of a bounded or a truncated probabilistic distribution, the defuzzified value of the fuzzy number using COA method will be equivalent to the mean value of the probabilistic distribution. The equations for calculating the expected values (EV) using the COA method of common fuzzy numbers used in this study are as follows [the fuzzy numbers are represented by a four element notation (a,b,c,d) as shown in Fig. 1]:

- Uniform fuzzy numbers (a,a,b,b)

$$\text{EV}_{\text{Uniform}} = \frac{a+b}{2} \tag{6}$$

- Triangular fuzzy numbers (a,b,b,c)

$$\text{EV}_{\text{Triangular}} = \frac{a+b+c}{3} \tag{7}$$

- Trapezoidal fuzzy numbers (a,b,c,d)

$$\text{EV}_{\text{Trapezoidal}} = a + \frac{2(c-b)(b-a)+(c-b)^2+(b-a)(d-a)+(c-b)(d-a)+(d-a)^2}{3(c-b+d-a)} \tag{8}$$

Eqs. (6)–(8) calculate the expected values of the uniform, triangular, and trapezoidal fuzzy numbers, respectively, which are equivalent to the mean values of uniform, triangular, and trapezoidal probability distributions, respectively.

The variances of the common fuzzy numbers used in the study can be calculated using the probabilistic definition of variance as follows:

- Uniform fuzzy numbers (a,a,b,b)

$$\text{Variance}_{\text{Uniform}} = \frac{(b-a)^2}{12} \tag{9}$$

- Triangular fuzzy numbers (a,b,b,c)

$$\text{Variance}_{\text{Triangular}} = \frac{a^2+b^2+c^2-ab-ac-bc}{18} \tag{10}$$

- Trapezoidal fuzzy numbers (a,b,c,d), according to van Dorp and Kortz (2003)

$$\begin{aligned}\text{Variance}_{\text{Trapezoidal}} = {} & \frac{(b-a)}{(d+c-b-a)}\left(\frac{1}{6}(a+b)^2+\frac{1}{3}b^2\right) \\ & + \frac{1}{(d+c-b-a)}\left(\frac{2}{3}(c^3-b^3)\right) \\ & + \frac{(d-c)}{(d+c-b-a)}\left(\frac{1}{3}c^2+\frac{1}{6}(c+d)^2\right) \\ & - (\text{EV trapez})^2\end{aligned} \tag{11}$$

Eqs. (9)–(11) calculate the variances of the uniform, triangular, and trapezoidal fuzzy numbers, respectively, which are equivalent to the variance values of uniform, triangular, and trapezoidal probability distributions, respectively.

Therefore, when a fuzzy number maintains the same range and the same shape of a bounded or a truncated probabilistic distribution, the variance of the fuzzy value will be equal to the variance of the related probability distribution.

Summation of Fuzzy Numbers

As indicated earlier, fuzzy numbers are considered a generalization of intervals, and the mathematical operations on fuzzy numbers (i.e., summation) can be processed using the concepts of the interval of confidence. Therefore, the methodology adopted for the summation of fuzzy numbers is the α-cut method and interval analysis. α-cut ($\alpha \in [0,1]$) is a discretization technique applied on the continuous membership function to generate a discrete set of variables in the form of intervals (a,b). The α-cut technique is based on the "extension principle" which implies that algebraic operations on real numbers can be extended to fuzzy numbers (Zadeh 1965). Interval arithmetic is used to analyze the generated intervals. The α-cut method can be used on different types and shapes of fuzzy numbers. The calculation process using α-cut method is carried out through the following steps:

1. Select a particular α-cut value ($0 \leqslant \alpha \leqslant 1$);
2. Find the corresponding intervals of the selected α-cuts;
3. Use the interval operations to calculate the summation; and
4. Repeat the steps for as many α-cuts as needed.

The types of fuzzy numbers used in this study (uniform, triangular, and trapezoidal) can be considered special cases of the trapezoidal shape. Therefore, they can be represented by four variables; trapezoidal$=(a,b,c,d)$, uniform$=(a,a,b,b)$, and triangular$=(a,b,b,d)$. The summation of these numbers can be performed by adding the bases and peaks of the numbers ($\sum$ first variables, $\sum$ second variables, $\sum$ third variables, and $\sum$ fourth variables).

Fuzzy Summation and Probabilistic Central Limit Theorem

According to the central limit theorem, the sum of independent random variables tends to the normal distribution as the number of random variables, regardless of their distributions, increases without limit (Ang and Tang 1975). Therefore, to prove that using fuzzy numbers in range estimating can yield comparable results to the probabilistic approach, we need to prove that the defuzzified expected value of the fuzzy output using COA is comparable to the mean of the probabilistic output. The probabilistic output is represented by a Gaussian distribution based on the central limit theorem. In addition, we need to prove that the summation of variances of fuzzy inputs is comparable to the summation of variances of probabilistic distributions.

Boswell and Taylor (1987) investigated the concept of the fuzzy random variable, which is a fuzzy set consisting of a membership function and a basic set whose components are an ordinary mapping (real random variables) from a probability space. In addition, the random variables exhibit an infinite number of distributional types whose summation or average is a fuzzy random variable with a membership function of its own and a basic set of random variables. The conclusion of their study is that the summation of independent fuzzy random variables converges, in the limit, to a fuzzy Gaussian random variable, providing a fuzzy equivalence of the central limit theorem of classical probability theory.

Boswell and Taylor's study (1987) provided a mathematical proof of the similarity between the summation of fuzzy random variables and the central limit theorem of classical probabilistic summation. In this study, we are only dealing with fuzzy numbers, which are not random. Therefore, the analysis is much easier and it is considered a special case of the summation of fuzzy random variables.

In addition to the mathematical proof provided by Boswell and Taylor (1987), an experiment was conducted using Microsoft EXCEL and the @RISK (Palisade, NY) add-in, in order to show how the summation of fuzzy numbers behaves as compared to the probabilistic range estimating approach. The experiment is designed to compare the outputs of the summation of fuzzy inputs and probabilistic distributions. The experiment is designed using the following assumptions:

1. The two approaches will share the same boundaries. The boundaries are randomly created between the following ranges:
   - Minimum range (MIN_{range})=Random (0–1,000); and
   - Maximum range (MAX_{range})=Random ([Min+100] to [Min+1,000]).
2. Fuzzy numbers used are:
   - Uniform (a,b): $a=(MIN_{range})$ and $b=(MAX_{range})$;
   - Triangular: (a,b,c): $a=(MIN_{range})$, $c=(MAX_{range})$, and b is randomly created between (a) and (c); and
   - Trapezoidal (a,b,c,d): $a=(MIN_{range})$, $d=(MAX_{range})$, and b and c are randomly created between (MIN_{range}) and (MAX_{range}), provided that $b \neq c$.
3. Probability distributions used are:
   - Uniform (min, max): min(continuous boundary parameter)=(MIN_{range}) and max(continuous boundary parameter)=(MAX_{range}).
   - Triangle (min, most likely, max): min(continuous boundary)= parameter(MIN_{range}), max(continuous boundary parameter)=(MAX_{range}), most likely (continuous mode parameter) is randomly created between (MIN_{range}) and (MAX_{range}).
   - PERT (min, most likely, max), which is an approximation of the Beta distribution: min(continuous boundary parameter)=(MIN_{range}), max(continuous boundary parameter)=(MAX_{range}), most likely (continuous parameter) is randomly created between (MIN_{range}) and (MAX_{range}).
   - Generalized beta ($\alpha_1,\alpha_2,\min,\max$): min(continuous boundary parameter)=(MIN_{range}), max(continuous boundary parameter)=(MAX_{range}), and α_1 (continuous shape parameter) and α_2 (continuous shape parameter) are both randomly created between 2 and 25. The range of the shape parameters is selected between 2 and 25 to restrict the shape of the beta distributions to unimodal distributions. In order to deal with bimodal beta distribution, the distribution must be divided into two distributions (each with one local maximum). Two fuzzy numbers are created in order to match the parameters of the divided distributions. The scope of this experiment only covers the unimodal distributions.
   - Normal truncated (μ,σ): The distribution is truncated between (MIN_{range}) and (MAX_{range}), μ (mean) is randomly created between (MIN_{range}) and (MAX_{range}), and σ (standard deviation) is randomly created between (MIN_{range}) and (MAX_{range}).

- Exponential truncated (β): the distribution is truncated between (MIN_{range}) and (MAX_{range}), β (continuous scale parameter) is randomly created between (MIN_{range}) and (MAX_{range}).
- Lognormal truncated (μ,σ): the distribution is truncated between (MIN_{range}) and (MAX_{range}), μ (mean) is randomly created between (MIN_{range}) and (MAX_{range}), and σ (standard deviation) is randomly created between (MIN_{range}) and (MAX_{range}).

4. 1,000 randomly selected combinations of fuzzy numbers and related probabilistic distributions are generated.
5. 6 inputs are used in each of the 1,000 combinations.
6. The probabilistic summations of each combination are performed using a Monte Carlo simulation that has the following assumptions:
   - Total number of iterations=1,000 and
   - Random generator seed=1.
7. The goal of the experiment is to calculate the absolute error between the expected values and variances of the outputs, as follows:

$$\text{Absolute error (of expected values EV)} = 100(|EV_{fuzzy} - EV_{probabilistic}|)/EV_{probabilistic} \quad (12)$$

$$\text{Absolute error (of standard deviation)} = 100\left(\left|\sqrt{\sum \text{fuzzyVariances}} - \sqrt{\sum \text{probabilisticVariances}}\right|\right)\Big/\sqrt{\sum \text{probabilisticVariances}} \quad (13)$$

8. Statistics collected for the absolute errors are:
   - Mean absolute error for all 1,000 combinations; and
   - Standard deviation of absolute error for all 1,000 combinations.

The results of the experiment were as follows:

1. After running the experiment for 1,000 iterations:
   - The mean absolute error of the "expected values" comparison is 5.2% with a standard deviation of 4%; and
   - The mean absolute error of the "standard deviation" comparison is 13.5% with a standard deviation of 11.6%.
2. When the beta distribution is not included in the analysis, the results after 1,000 iterations are improved:
   - The mean absolute error of the expected values comparison is 5.0% with standard deviation of 3.8%; and
   - The mean absolute error of the standard deviation comparison is 9.6% with a standard deviation of 6.9%.

From the results generated the following conclusions were made:

- The fuzzy approach generated very comparable results to the probabilistic approach. This was proven experimentally when bounded and truncated probability distributions were used in the analysis.
- When the beta distribution was removed from the experiment, better results were obtained, which indicates that the beta distribution in general requires some parameter control to obtain better results.

Fuzziness and Ambiguity Measures

When experts use fuzzy numbers in modeling uncertainty, it is very important to have some measures to assess uncertainty and evaluate how vague or precise the experts' estimates are. There are different definitions and terms that describe "uncertainty." Uncertainty definitions can be categorized into two terms: vagueness and ambiguity. According to Klir and Folger (1988), vagueness is related to the difficulty of providing a sharp and precise distinction of a specific incident or phenomenon due to its vagueness. As for the term "ambiguity," it describes the situation in which there is difficulty in making a specific selection or decision between alternatives due to their ambiguity. In fuzzy set theory, the concept of "fuzziness measure" deals with the first type of uncertainty (vagueness). The fuzziness measure is related to the degrees to which an arbitrary element of the universal set (X) belongs to the individual crisp subsets of X (Klir and Folger 1988). The other measure, the "ambiguity measure," deals with the lack of precision in determining the exact value of a magnitude (Delgado et al. 1998b).

The fuzziness measure adopted in this paper is the one developed by Klir and Folger (1988). Their approach defines the fuzziness of a set in terms of the lack of distinction between the set and its complement. In other words, the more a set is different from its complement, the fuzzier it is. In order to calculate the fuzziness measure using this approach, a fuzzy complement approach and a distance function (e.g., Hamming distance) are utilized. For a given fuzzy set $A(x)$, the measure of fuzziness, $F(A)$, is obtained using the following equation:

$$F(A) = \sum_{x \in X} (1 - |2A(x) - 1|) \quad (14)$$

Eq. (14) is only applicable to finite fuzzy sets, but it can be modified to fuzzy sets defined on infinite but bounded subsets. For example, when $X=[a,b]$, the measure of fuzziness is then obtained by

$$F(A) = \int_a^b (1 - |2A(x) - 1|)dx = b - a - \int_a^b |2A(x) - 1|dx \quad (15)$$

The fuzziness of a crisp number or a fuzzy uniform number is zero because the lack of distinction between a fuzzy uniform number or a crisp number and their complements is zero. As for the ambiguity measure, the approach developed by Delgado et al. (1998a) is selected as the ambiguity measure used in this paper. According to Delgado et al. (1998a), ambiguity (AG) is obtained by the following formula:

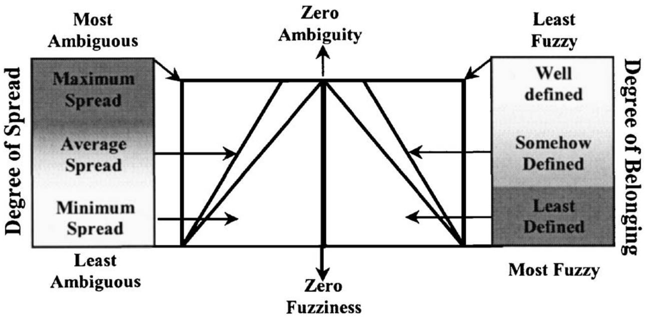

Fig. 3. Fuzziness and ambiguity measures definition

$$\mathrm{AG}(\mu) = \int_0^1 r[R(r) - L(r)]dr \qquad (16)$$

where (μ)=fuzzy number with r-cut representation $(L(r), R(r))$. The term $[R(r) = L(r)]$=length of the r-cut interval $(L(r), R(r))$. Therefore, AG(μ) can be considered as a "global spread" of the fuzzy number. The ambiguity AG(μ) can be calculated for some of the most common fuzzy numbers as follows:

1. For a trapezoidal fuzzy number (a_1, a_2, a_3, a_4)

$$\mathrm{AG}(\mu)_{\mathrm{Trapezoidal}} = (a_3 - a_2)/2 + [(a_4 - a_3) + (a_2 - a_1)]/6 \qquad (17)$$

2. For a triangular fuzzy number (a_1, a_2, a_3)

$$\mathrm{AG}(\mu)_{\mathrm{Triangular}} = [(a_3 - a_2) + (a_2 - a_1)]/6 \qquad (18)$$

3. For a uniform fuzzy number (a_1, a_2)

$$\mathrm{AG}(\mu)_{\mathrm{Uniform}} = (a_2 - a_1)/2 \qquad (19)$$

4. For a crisp number (a_1)

$$\mathrm{AG}(\mu)_{\mathrm{Crisp}} = 0 \qquad (20)$$

To clarify the difference between the fuzziness and ambiguity measures, Fig. 3 provides a comparative illustration of the two measures for three types of fuzzy numbers (uniform, triangular, and trapezoidal).

For comparative purposes, the three fuzzy numbers are defined on the same range. Fig. 3 shows that the uniform fuzzy number has the least fuzziness measure, and the triangular fuzzy number has the highest fuzziness measure, because the degree of belonging of the first is well defined (in terms of intervals) while the latter has a "fuzzily" defined degree of belonging depicted by the sloped lines that form the triangle. As for the trapezoidal fuzzy number, although it has sloped lines similar to those of the triangular fuzzy number, it is less fuzzy than the triangular fuzzy number because it contains a defined "flat or uniform range." This unique shape of the trapezoidal fuzzy number renders its fuzziness somewhere between that of the uniform fuzzy number and the triangular fuzzy number.

As for the ambiguity measure, Fig. 3 shows that the triangular fuzzy number has the least ambiguity, and that the uniform fuzzy number is most ambiguous. A crisp number has zero fuzziness and zero ambiguity. Taking a crisp number as the reference, it is noted that when the overall spread of any fuzzy number around this crisp number is minimized, the ambiguity is decreased. This explains why the uniform fuzzy number would certainly have the highest measure of ambiguity, and the triangular fuzzy number would have the least. Again, the trapezoidal fuzzy number falls in between the two extremes, simply because it possesses, to some extent, the characteristics of both of the other two fuzzy numbers.

In order to provide a measure that combines the effect of both the fuzziness and ambiguity measures, a "fuzzy number quality index" (FNQI) is calculated using the weighted average of both measures. For a fuzzy set A, the fuzzy quality index is defined measured as:

$$\mathrm{FNQI} = [W_F F(A) + W_{\mathrm{AG}} \mathrm{AG}(A)]/[W_F + W_{\mathrm{AG}}] \qquad (21)$$

where W_F and W_{AG}=weights of the fuzziness measure and ambiguity measure respectively and $F(A)$ and AG(A) are the fuzziness and ambiguity measures of fuzzy number (A), respectively. In this paper, equal weights are assumed for both measures. As there is no reference index against which the fuzzy quality measure can be compared, an acceptance level can be set by the user to reject the estimates with clearly high (FNQI) compared to the others. Table 1 shows the evaluation process of the different measures.

As shown in Table 1, depending only on the fuzziness measure to assess the "quality" of the estimates is not usually sufficient. Although some fuzzy numbers have low measures of fuzziness, they still lack the necessary precision in determining the exact value, which renders them "ambiguous." A good example of this is uniform fuzzy numbers (sometimes referred to as crisp intervals). Although they will always have a zero measure of fuzziness, they have high measures of ambiguity. Therefore, it is important to assess the quality of the estimates using two measures (fuzziness and ambiguity).

Table 1. Comparison of Fuzziness Measure, Ambiguity Measure, and Fuzzy Quality Index

Type of fuzzy number	Fuzzy parameters				Fuzziness measure	Ambiguity measure	Fuzzy quality index
	a	*b*	*c*	*d*			
Trapezoidal	2.0	7.0	10.0	20.0	7.5	4.0	5.8
Trapezoidal	3.0	7.0	9.0	15.0	5.0	2.7	3.9
Trapezoidal	10.0	10.0	15.0	18.0	1.5	3.0	2.3
Triangular	12.0	16.0	16.0	25.0	6.5	2.2	4.4
Triangular	7.0	12.0	12.0	12.0	2.5	0.8	1.7
Triangular	7.0	8.0	8.0	10.0	1.5	0.5	1.0
Uniform	6.0	6.0	30.0	30.0	0.0	12.0	6.0
Uniform	2.0	2.0	7.0	7.0	0.0	2.5	1.3
Uniform	6.0	6.0	7.0	7.0	0.0	0.5	0.3

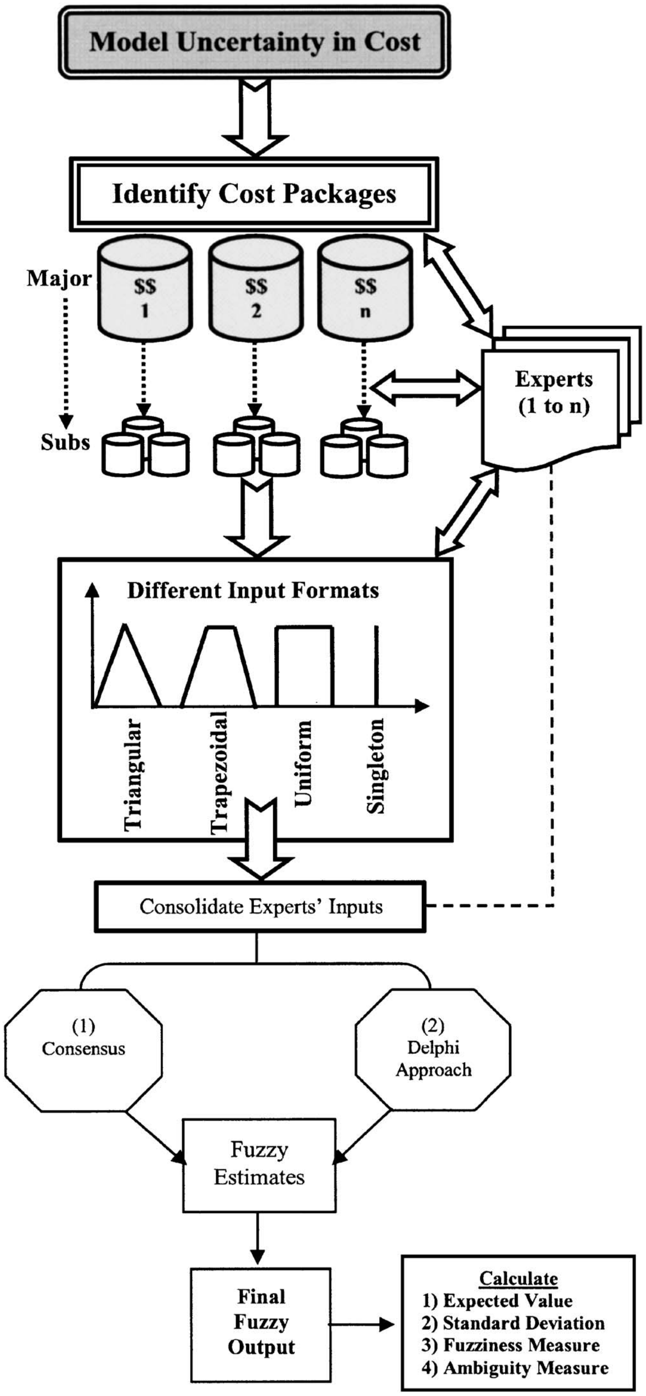

Fig. 4. Proposed methodology flowchart

Proposed Fuzzy Range Estimating Model

In this section, a detailed description of the model developed for fuzzy range estimating is provided. The model utilizes the different concepts discussed in the previous sections. Fig. 4 shows a flowchart of the proposed model for modeling uncertainty in cost range estimating using fuzzy arithmetic.

The following steps are required in the proposed model:

1. By consulting the experts involved in the process, the problem under investigation is identified. In the case of uncertainty in cost estimates, the major cost packages and their subgroups have to be clearly identified by the experts.
2. Each expert is then required to give his/her best estimate based on his/her experience choosing from the different input formats (triangular, trapezoidal, uniform, singleton). The logical meaning of these input formats is explained as follows:
   - Triangular fuzzy numbers: Chosen when the expert believes that the estimate of the item has a "most likely" or "most plausible" point that is between a maximum and a minimum boundary.
   - Trapezoidal fuzzy number: Chosen when the expert believes that the estimate has a most plausible range or interval that lies between a maximum and a minimum boundary.
   - Uniform fuzzy numbers (sometimes referred to as a "crisp interval"): Chosen when the expert believes that the estimate should take the form of an interval that has a maximum and a minimum point, and when he/she finds it difficult to assign a most plausible estimate to a point or a range.
   - Crisp number (singleton): Chosen when the expert is 100% confident or certain that the estimate is a deterministic one that has zero uncertainty.
3. The next step is to consolidate the experts' inputs. The consolidation process can be done using three different approaches as follows:
   - Consensus: The final estimates are made by consensus between the experts. This approach usually involves round-table discussions between the experts participating in the estimation process. Generally, the greater the number of experts involved in the process, the more time required to reach a consensus. This approach can be used in the estimation process within the same company or entity.
   - Fuzzy Delphi approach: The Delphi method is a systemic approach used for long range forecasting in which the estimates of experts are made to converge using statistical analysis (Kaufmann and Gupta 1988). It is conducted by interviewing highly qualified experts to obtain their opinions regarding specific issues, such as cost and time estimates. Since it is difficult in most cases to provide precise and crisp estimates of cost or time, a fuzzy representation of the process can model this uncertainty. The fuzzy Delphi method can be an effective method for extracting subjective information from experts. According to Kaufmann and Gupta (1988), the fuzzy Delphi method consists of the following steps:

1.

n number of experts are asked to provide three estimates of a specific value (e.g., estimating activity duration) by determining the minimum, the most plausible, and the maximum estimate. The estimates are presented in the form of triangular fuzzy numbers

$$A_i = (a^i, b^i, c^i) \tag{22}$$

where A=triangular fuzzy number; a, b, and c=first, second, and third elements of the fuzzy number, respectively; and i= ith expert (i=1,2,...n).

2. The estimates are averaged. For each expert, the deviation from the average is calculated as follows:

$$F_{\text{avg}}(a_{\text{avg}}, b_{\text{avg}}, c_{\text{avg}}) = \left(\frac{1}{n}\sum_{i=1}^{n} a^i, \frac{1}{n}\sum_{i=1}^{n} b^i, \frac{1}{n}\sum_{i=1}^{n} c^i \right) \tag{23}$$

$$F_{\text{avg}} - A_i = \left(\frac{1}{n}\sum_{i=1}^{n} a^i - a^i, \frac{1}{n}\sum_{i=1}^{n} b^i - b^i, \frac{1}{n}\sum_{i=1}^{n} c^i - c^i \right) \quad (24)$$

where F_{avg}=fuzzy average; and a_{avg}, b_{avg}, and c_{avg}=first, second, and third elements of the fuzzy number, respectively.

3. The deviations in the estimates are sent back to the experts for revision. Each expert provides a new triangular fuzzy number. Steps 1–3 are repeated until two successive averages become reasonably close based on the decision maker's stopping criterion. The way the technique is developed restricts the input option to one input type (triangular fuzzy number), which limits the expert's freedom in providing his or her estimate in different formats (i.e., trapezoidal fuzzy number). Therefore, in this paper, a modified fuzzy Delphi method is presented where experts are given the choice of representing their estimate in either uniform, triangular, or trapezoidal fuzzy numbers. The proposed fuzzy Delphi method has the same steps explained before. For easier fuzzy arithmetic processing, uniform and triangular fuzzy numbers can be represented by the trapezoidal four-element format because they are considered particular cases of the trapezoidal fuzzy number, as follows:

- Trapezoidal fuzzy number=(a,b,c,d);
- Triangular fuzzy number=(a,b,c) can be represented as (a,b,b,c);
- Uniform fuzzy number=(a,b,c) can be represented as (a,a,b,b); and
- Crisp number=(a) can be represented as (a,a,a,a).

The general presentation of the four-element fuzzy number is $A=(e_1,e_2,e_3,e_4)$, where e_1, e_2, e_3, and e_4=first, second, third, and fourth elements of the fuzzy numbers, respectively.

The average fuzzy number F_{avg} is calculated as follows:

$$F_{\text{avg}} = \left(\frac{1}{n}\sum_{i=1}^{n} e_1^i, \frac{1}{n}\sum_{i=1}^{n} e_2^i, \frac{1}{n}\sum_{i=1}^{n} e_3^i, \frac{1}{n}\sum_{i=1}^{n} e_4^i \right) \quad (25)$$

The deviation from the average is calculated as follows:

$$F_{\text{avg}} - A_i = \left(\frac{1}{n}\sum_{i=1}^{n} e_1^i - e_1^i, \frac{1}{n}\sum_{i=1}^{n} e_2^i - e_2^i, \frac{1}{n}\sum_{i=1}^{n} e_3^i - e_3^i, \frac{1}{n}\sum_{i=1}^{n} e_4^i - e_4^i \right) \quad (26)$$

The expected value of the average can be calculated using Eq. (8).

4. The process proposed in this paper is a structured process for extracting information from the experts in the form of fuzzy numbers. Depending on the application, the process can work for both cost and activity duration evaluation. The cost range estimating application is presented in this paper. In cost range estimating, the next step in the process after extracting the necessary information from the experts is to calculate the total estimated cost of the problem under consideration, which is performed using simple summation. The operation starts by adding up the cost of the subitems for each work package. The final total cost estimate is then obtained by adding up the costs of the work packages. The final evaluation, representing the total estimated cost of the modeled packages, is in the form of a fuzzy number. The following information will be calculated for the final fuzzy output:
   - Expected mean value of the output calculated using one of the equations from Eqs. (6)–(8);
   - Standard deviation calculated using one of the equations from Eqs. (9)–(11);
   - Fuzziness measure using Eq. (15);
   - Ambiguity measure using one of the equations from Eqs. (17)–(20); and
   - FNQI using Eq. (21).

   This information is important in assessing the precision and quality of the output when compared to other outputs obtained from different estimating techniques. The following section provides an illustrative example comparing the proposed fuzzy range estimating process to the probabilistic range estimating technique.

Illustrative Example Comparing Fuzzy and Probabilistic Approaches

To illustrate the proposed approach, a case study of a tunneling project using a tunnel boring machine is used. The North of Edmonton Sanitary Trunk (NEST) project, conducted by the City of Edmonton had a maximum budgeted cost of $8.8 million and a preliminary estimated cost of $6 million. The City of Edmonton had concerns regarding the budget and wanted to know the chances of exceeding the preliminary estimated cost and being within the total budgeted cost. The main cost packages and their subcategories are shown in Table 2.

A Monte Carlo simulation study was conducted to estimate the chances of meeting the budgeted cost (AbouRizk et al. 2006). In the study, the writers used Simphony's (University of Alberta, Alberta) range estimating template to do the analysis. Simphony is a specialized simulation tool that supports the Monte Carlo simulation technique in discrete event simulation and range estimating (AbouRizk and Mohamed 2000). The model inputs are listed in Table 2.

After 500 iterations, the following statistics were collected using the Monte Carlo simulation model:

- Low estimated cost: $5,486,345;
- High estimated cost: $6,840,657;
- Mean estimated cost: $6,059,263;
- Standard deviation: $280,249.7; and
- 80th percentile: $6,300,000.

In this example, only the comparison between the final outputs of the probabilistic and fuzzy approach is performed. Therefore, for the fuzzy range estimating approach, the inputs represented by the fuzzy numbers are given by the same shape of probability distributions used (i.e., a triangular fuzzy number is used when a probabilistic triangular distribution is used). The fuzzy output was generated after one iteration only by summing up all the fuzzy inputs involved in the analysis. The fuzzy output is a trapezoidal fuzzy number that has the following characteristics:

Fuzzy trapezoidal number parameters: a_1=$5,038,248.8, a_2=$5,697,250.2, a_3=$5,732,250.2, and a_4=$7,417,863.4;

- Expected mean value: $6,054,474 [using Eq. (8)];
- Standard deviation: $501,046.33 [using Eq. (11)];
- Fuzziness measure: 1,172,307.1 (11.7 scaled) [using Eq. (15)];
- Ambiguity measure: 408,269.1 (4.08 scaled) [using Eq. (17)]; and
- FNQI: 7.89 (scaled) [using Eq. (21)].

Table 2. Data for the NEST Case Study Project

Item	Description	Optimistic $	Most likely $	Pessimistic $
1	**Main work shaft**			
1.1.1	Mobilization—move in	40,000	70,000	100,000
1.1.2	Power installation		89,000	
1.1.3	Power-156 Str.	15,000		50,000
1.2	Excavate work shaft	97,600	122,000	146,400
1.3	Excavate undercut	200,000	269,000	350,000
1.4	Excavate tail tunnel to east	100,000	123,000	150,000
1.5	Form and pour undercut		80,000	
1.6	Form and pour tail undercut		39,000	
1.7	Form and pour shaft	100,000	120,000	150,000
2	**Access manhole**			
2.1	Excavate access shaft		16,000	
2.2	Backfill shaft and install AMH		44,000	
3	**Tunneling (866 m)**			
3.1	Tunnel and install segments—866 m price per m	2,254	2,474	3,360
3.2	Patch and rub tunnel crown	80	134	140
3.3	Patch and rub tunnel-final cleanup	161	188	215
3.4	Spoil removal	5.4	8.1	9.7
4	**Access manhole shaft**		61,000	
5	**Tunneling (756 m)**			
5.1	Tunnel and install segments—756 m price per m	2,254	2,474	3,360
5.2	Patch and rub tunnel crown	80	134	140
5.3	Patch and rub tunnel-final cleanup	161	188	215
5.4	Spoil removal	5.4	8.1	9.7
6	**Removal shaft**		101,000	

The fuzzy output is shown in Fig. 5. The fuzziness, ambiguity, and FNQI measures were scaled (divided by 100,000) in order to make the comparison more readable and easier to grasp. The quality measures (fuzziness, ambiguity, and FNQI) provide more meaningful information when used as relative comparative indices, and can not be used to assess the quality of the fuzzy output unless more alternatives exist. Therefore, when more than one run is done and, hence, more than one fuzzy output is obtained, the fuzziness and ambiguity measure can be used to assess the quality of these outputs and check which one has the least FNQI.

When comparing the outputs of both the probabilistic and fuzzy input approaches the following observations are made:

- The difference between the probabilistic mean and the fuzzy expected value is 0.07% (the fuzzy output is less by 0.07%).
- A unique concept in fuzzy set theory is the law of possibility or possibility measure (Kaufmann and Gupta 1985). Using the possibility measure, one can determine which value is more plausible or possible. This type of information that fuzzy set theory provides is unique; probability theory does not support this concept, because the probability for a specific random variable to take place is close to zero (Lorterapong and Moselhi 1996). Therefore, it is easy to determine which specific variable is more possible and plausible using this measure. For example, the possibility measure for the $6 million project estimated cost equals 0.84 which means that the project cost can reach $6 million with a 0.84 possibility as shown in Fig. 5. In addition, the most possible and plausible variable in a normal fuzzy number is the one that has a possibility measure of 1.0. Therefore, the most possible and plausible output is ($5,697,250, $5,732,250).

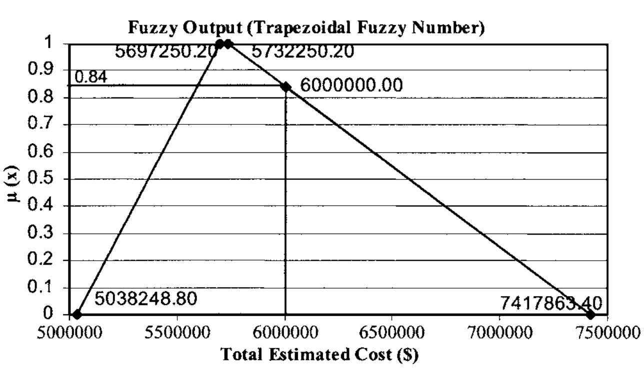

Fig. 5. Fuzzy output

Conclusions and Recommendations

In this paper, modeling uncertainty using fuzzy set theory was shown to be as effective as the probabilistic approach. Fuzzy set theory has the advantage of providing easier and faster to obtain outputs. An effective approach for extracting subjective information from experts was presented. Using the approach, experts are free to express their knowledge in different formats, e.g., triangular or trapezoidal fuzzy numbers, which suits the way in which they think and better captures their subjective assessments.

Modeling range estimating using fuzzy arithmetic yields very comparable outputs when compared to the probabilistic approach.

The fuzzy approach has the advantage of being faster and easier to process because it only takes one iteration to generate the output, while it takes Monte Carlo simulation a number of iterations to generate a reasonable and reliable output. The final fuzzy output is assessed by quality measures that can be used to compare the fuzziness, ambiguity, and the combined effect of both of the output. These quality measures can be used to compare multiple final estimates by showing how vague or precise the experts were in providing their estimates.

Further, the possibility measure in fuzzy set theory is considered a unique concept that evaluates the plausibility for a specific variable within the fuzzy number to take place. Probability theory does not support a similar measure. The focus of this study is modeling cost range estimating using fuzzy set theory. Fuzzy range estimating in scheduling can also be investigated. The same methodology developed for the fuzzy cost range estimating problem can be utilized in fuzzy scheduling range estimating. However, different fuzzy arithmetic operations were utilized to calculate the forward and backward pass calculations and criticality measurements. This paper has laid the foundation for future research on incorporating fuzzy numbers in range estimating applications.

Notation

The following symbols are used in this paper:

$A(x)$ = fuzzy number A of a variable x;
$AG(A)$ = ambiguity measure of fuzzy number (A);
$E(X)$ = mean value of a continuous random variable X;
e_1 = first element of a fuzzy number;
e_2 = second element of a fuzzy number;
e_3 = third element of a fuzzy number;
e_4 = fourth element of a fuzzy number;
$F(A)$ = fuzziness measure of fuzzy number (A);
F_{avg} = average fuzzy number;
$f_X(x)$ = probability density function;
I_y = moment of inertia about the vertical centroidal axis;
max = maximum;
MAX_{range} = maximum range;
min = minimum;
MIN_{range} = minimum range;
$Var(X)$ = variance of a continuous random variable X;
W_{AG} = weight of ambiguity measure;
W_F = weight of fuzziness measure;
x_0 = centroidal distance of a unit area;
y^* = defuzzified value of fuzzy number;
$\mu(x)$ = membership function of variable x; and
$\mu_i(x)$ = aggregated membership function of output variable (x).

References

AbouRizk, S., and Mohamed, Y. (2000). "Simphony—An integrated environment for construction simulation." *Proc., 2000 Winter Simulation Conf.*, Orlando, Fla., 32, 1907–1914.
AbouRizk, S., Mohamed, Y., Fernando, S., and Shahin, A. (2006). "Simulation analysis of North Edmonton sanitary trunk." *Can. J. Civ. Eng.*, in press.
Ahuja, H. N., Dozzi, S. P., and AbouRizk, S. M. (1994). *Project management: Techniques in planning and controlling construction project*, 2nd Ed., Wiley, New York.
Ang, A. H.-S., and Tang, W. H. (1975). *Probability concepts in engineering planning and design: Volume I—Basic principles*, Wiley, New York.
Boswell, S. B. and Taylor, M. S. (1987). "A central limit theorem for fuzzy random variables." *Fuzzy Sets Syst.*, 24(3), 331–334.
Delgado, M., Vila, M. A., and Voxman, W. (1998a). "On a canonical representation of fuzzy numbers." *Fuzzy Sets Syst.*, 93(1), 125–135.
Delgado, M., Vila, M. A., and Voxman, W. (1998b). "A fuzziness measure for fuzzy numbers: Applications." *Fuzzy Sets Syst.*, 94(2), 205–216.
Ferson, S. (2002). *Ramas risk calc 4.0 software: Risk assessment with uncertain numbers*, CRC, Boca Raton, Fla.
Kaufmann, A., and Gupta, M. M. (1985). *Introduction to fuzzy arithmetic, theory and applications*, Van Nostrand Reinhold, New York.
Kaufmann, A., and Gupta, M. M. (1988). *Fuzzy mathematical models in engineering and management science*, Elsevier Science, Amsterdam, The Netherlands.
Klir, G., and Folger, T. (1988). *Fuzzy sets, uncertainty, and information*, Prentice-Hall, Englewood Cliffs, N.J.
Lorterapong, P., and Moselhi, O. (1996). "Project-network analysis using fuzzy sets theory." *J. Constr. Eng. Manage.*, 122(4), 308–318.
van Dorp, J. R., and Kotz, S. (2003). "Generalized trapezoidal distributions." *Metrika*, 58(1), 85–97.
Zadeh, L. A. (1965). "Fuzzy sets." *Inf. Control.*, 8, 338–353.