algorithms

MDPI

Article

Fuzzy-Based Multivariate Analysis for Input Modeling of Risk Assessment in Wind Farm Projects

Emad Mohamed [1], **Parinaz Jafari** [1] **and Simaan AbouRizk** [1,2,*]

1 Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB T6G 2W2, Canada; ehmohame@ualberta.ca (E.M.); parinaz@ualberta.ca (P.J.)
2 5-080 Markin/CNRL Natural Resources and Engineering Facility, Edmonton, AB T6G 2W2, Canada
* Correspondence: abourizk@ualberta.ca

Received: 29 October 2020; Accepted: 2 December 2020; Published: 4 December 2020

check for updates

Abstract: Currently, input modeling for Monte Carlo simulation (MSC) is performed either by fitting a probability distribution to historical data or using expert elicitation methods when historical data are limited. These approaches, however, are not suitable for wind farm construction, where—although lacking in historical data—large amounts of subjective knowledge describing the impacts of risk factors are available. Existing approaches are also limited by their inability to consider a risk factor's impact on cost and schedule as dependent. This paper is proposing a methodology to enhance input modeling in Monte Carlo risk assessment of wind farm projects based on fuzzy set theory and multivariate modeling. In the proposed method, subjective expert knowledge is quantified using fuzzy logic and is used to determine the parameters of a marginal generalized Beta distribution. Then, the correlation between the cost and schedule impact is determined and fit jointly into a bivariate distribution using copulas. To evaluate the feasibility of the proposed methodology and to demonstrate its main features, the method was applied to an illustrative case study, and sensitivity analysis and face validation were used to evaluate the method. The results demonstrated that the proposed approach provides a reliable method for enhancing input modeling in Monte Carlo simulation (MCS).

Keywords: Monte Carlo simulation; input modeling; fuzzy logic; risk analysis and assessment; multivariate distribution; marginal Beta; copula

1. Introduction

Wind and solar energy are expected to lead the future transformation of the global electricity sector, with onshore and offshore wind predicted to produce about 35% of total electricity demands by 2050 [1]. To reach the targeted installation capacity, considerable investments in the construction of renewable energy infrastructure are being made [1]. In Alberta, Canada, $3.6 billion will be invested through the renewable electricity program (REP) to add 5000 megawatts of renewable energy by 2030 [2,3].

As a relatively novel type of infrastructure, wind farm construction is characterized by a lack of relevant literature and a scarcity of historical data. The development of risk management plans for these types of projects, therefore, are highly dependent on the collection of expert knowledge [4]. While the boom in the wind energy industry has encouraged new contractors to engage in the construction of these projects, a lack of data represents a challenge for new contractors when conducting risk management. Inadequate risk identification and assessment can have a detrimental impact on these large-scale projects, resulting in negative effects on cost, time, quality, and safety, while simultaneously discouraging contractors from engaging in wind farm construction.

Risk assessment is necessary during all phases of a wind farm project, including design, construction, operation, maintenance, and life cycle planning [5]. Although there are several types of qualitative and quantitative methods for reliability-based risk assessment, determining which approach to apply in each phase of the project's life cycle will depend on the amount and/or type of data available at a particular phase [5]. Qualitative approaches (e.g., failure mode and effect analysis (FMEA), fault tree analysis (FTA), event tree analysis (EVA), and risk matrices), are better suited to the planning and early construction phases of a project when data are limited [5]. However, as projects progress and as more data are gathered, quantitative methods (e.g., analytical methods, stochastic methods, and Bayesian approaches) are favored because of their comprehensive capabilities [5].

Monte Carlo simulation (MCS) is a widely applied stochastic quantitative approach for risk assessment. It is an extremely powerful tool used for understanding and quantifying the potential effects of uncertainty on a project [6], and has been widely applied to simulate cost and time in construction [7]. As with many quantitative methods, however, the application of MCS is constrained by the need for variables to be input as probability density functions, limiting its use in the planning and early construction phases of a project. Currently, there are two primary approaches for developing the probability density functions that are input into MCS-based risk assessment models (Figure 1).

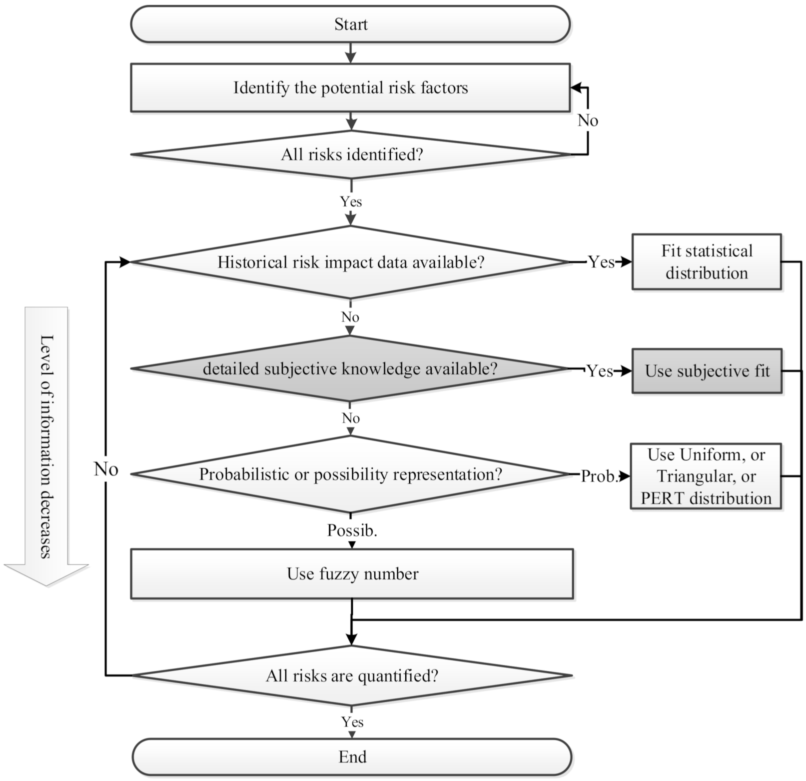


Figure 1. Levels of data available and accompanying methods used for input modeling.

As classified by Biller and Gunes [8], the two approaches are categorized based on data availability. The first approach can be adopted when there is sufficient historical data available for a particular variable [8]. In this case, a probability distribution is fit to the data and is then input into the risk assessment model. The second approach is adopted when there is an absence of data. This approach uses elicitation methods to construct the input distribution [8], where the risk analyst can decide to use either (1) a probabilistic approach by choosing either triangular, uniform, or PERT distributions [9,10], or (2) a possibility approach, where fuzzy numbers are used to represent the impact of a risk factor [11].

Wind farm construction does not have sufficient historical data available to complete a risk assessment using the first approach. Nevertheless, there is a substantial amount of detailed subjective knowledge available [4]. The second approach, however, cannot make use of this knowledge, thereby missing an opportunity to enhance the reliability of risk assessment results. Indeed, the current state-of-the-art lacks methods that can derive an appropriate probability distribution function from detailed subjective expert knowledge.

An additional difficulty experienced when conducting risk assessment modeling in wind farm construction is the consideration of risk factors affecting schedule and cost as independent. Delays in project schedule will often result in increased project costs, where impacts on cost are generally accompanied by project delays [12]. Treating these cost and schedule impacts as dependent during risk assessment modeling will help generate more realistic results. Methods for modeling the dependence of a risk factor's impact on cost and schedule, however, remain relatively unexplored.

To address these limitations, this paper proposes a methodology that is designed to enhance risk assessment outcomes by (1) assisting risk analysts in fitting appropriate distributions for detailed subjective knowledge of the cost and schedule impacts of risk factors (see dark grey area, Figure 1) and (2) adapting existing methods to model the dependence between the cost and schedule impact of a risk factor in a Monte Carlo simulation (MCS)-based risk assessment model. Fuzzy logic is used to process and quantify subjective knowledge, which is then fit to a probabilistic distribution function that represents the marginal distribution for the impact on either cost or schedule. Then, copula-based modeling of multivariate distributions [13] is used to model the dependence between the cost and schedule impact of a risk factor. The remaining sections of this paper are organized as follows: MCS in risk assessment, input modeling for MCS, and correlation and dependence between input distributions are first discussed in a literature review section. Then, the research methodology is explained. Next, an illustrative case study is presented to demonstrate the functionality of the proposed method, and a sensitivity analysis is performed to establish its validity. The final section summarizes conclusions and future research directions.

2. Literature Review

2.1. MCS for Risk Assessment and Input Modeling

Risk assessment is conducted by evaluating the probability of occurrence and the impact of risk factors to determine their severity on project outcomes. Mathematically, this is accomplished by multiplying probability (P) by impact (I) [14,15] and summing the results to obtain an overall effect of risk factor, n, on project cost and time, as per Equation (1):

$$S = \sum_{i=1}^{n} P_i * I_i \tag{1}$$

MCS is a probabilistic technique for the quantitative analysis of risks in the construction industry [16]. MCS makes use of probability distributions rather than deterministic values to model the uncertainty associated with a particular input [9]. In a MCS risk analysis experiment, the term I_i in Equation (1) is replaced with a probability distribution function representing the impact on cost or schedule.

A MCS experiment for assessing project risks is performed as follows. First, the baseline cost and schedule of the individual activities are prepared [9]. Then, risk factors affecting the project are identified, and the cost impact, schedule impact, probability of occurrence, and affected work-packages are determined for each individual risk factor [9]. If a risk occurs while running the simulation experiment, the cost and schedule impacts are added to the affected activities. An example project, consisting of three activities (A, B, C) and one risk factor affecting Activity A, is presented in Figure 2.

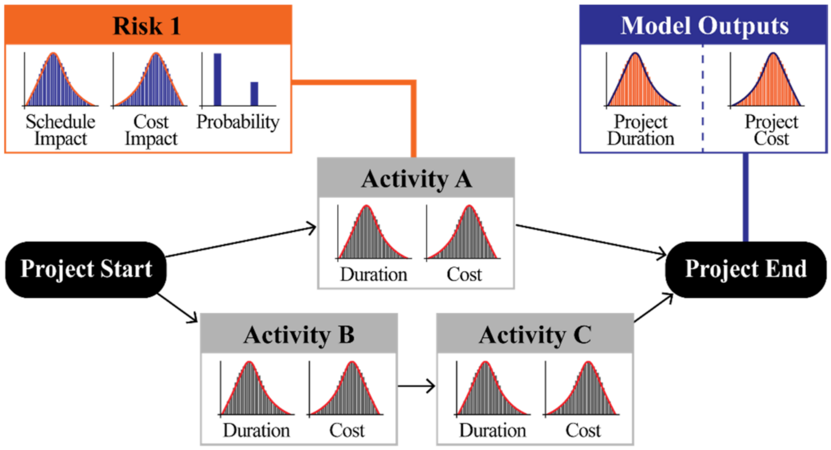


Figure 2. Monte Carlo (MC) simulation experiment for risk assessment.

A key limitation for the practical application of MCS in risk assessment is the development of risk impact probability distributions that are not readily available because of insufficient historical data. These distributions, therefore, must be derived from other existing information and knowledge [17,18]. The development of such distributions, known as input modeling, is widely discussed in literature due to its impact on simulation outputs and, consequently, on the quality of decisions made based on the simulation results [19]. Input modeling in the case of absent data is known as an elicitation process. Here, expert judgment, recognized as a type of scientific data [20], is elicited and used to construct probability distributions [21]. Elicitation of expert judgment can take three forms, where experts are asked to specify (1) the cumulative distribution function, (2) the density distribution function, or (3) to provide partial information about the distribution such as mean, standard deviation, or several quintiles of the distribution [22]. The elicitation process typically involves the elicitation of the most likely, maximum, and minimum values, or estimating the mean and variance from experts [23]. Several studies have provided a review and guidelines of the statistical methods used to elicit probability distributions [22,24,25]. For example, Galway [26] concluded that multiple experts should be asked to provide upper, lower, and most likely values for an uncertain quantity, which can then be fitted to a triangular distribution. In contrast, Morris et al. [21] developed a web-based tool that has five optional methods for the elicitation process, including a roulette method, quartile method, tertile method, probability method, and hybrid method.

Another challenge of elicitation for risk assessment is the introduction of biases arising from the inherent subjectivity of risk evaluation. Although experts are considered to be knowledgeable and experienced, their judgment may, as a result of biases, be inaccurate—especially when judging probability [27]. There are two major types of biases: cognitive and motivational. Cognitive biases are defined as “systematic deviations [of expert evaluation] from logic, probability, or rational choice theory” [28] and are often associated with heuristic judgment processes [29]. Examples of cognitive biases include overconfidence (i.e., excessive confidence in one’s own answers to questions), anchoring (i.e., the tendency to rely too heavily on one piece of information when making decisions), and availability bias (i.e., the tendency to overestimate the likelihood of events with greater availability in memory). Various debiasing methods, such as decomposition, multiple experts, and exploration of the extremes of a target variable, can be applied to reduce cognitive biases [28].

Motivational biases are defined as “those in which judgments are influenced by the desirability or undesirability of events, consequences, outcomes, or choices” [30]. One example of motivational bias is the underestimation of project costs to provide more competitive bids. A strategy for overcoming motivational biases, proposed by Montibeller and Winterfeldt [30], is the decomposition of a target variable into component variables and events (i.e., root causes). It is expected that the evaluation of a specific root cause by an expert will be more precise than the evaluation of a risk factor as a whole.

As such, analysts are encouraged to adopt a decomposition strategy when eliciting subjective risk evaluations from experts.

While recommendations for successfully inputting uniform, triangular, or PERT distributions into Monte Carlo risk assessment models have been proposed in literature, other forms of input modeling remain limited by a need for a large number of experts or by an inability to integrate their detailed knowledge [9,10]. The fitting of expert opinion regarding the impact and probability of occurrence of risks into a probability distribution has been proposed by Li et al. [31]. However, their approach requires a large number of experts to be involved in the assessment, which is not feasible for novel projects, such as onshore wind farms, where the number of experts at a construction company is limited. Furthermore, their approach does not allow experts to express their detailed knowledge about a risk factor, potentially limiting the accuracy and representativeness of the risk assessment results. In 2003, Nasir et al. [24] proposed a methodology to enhance the input modeling of Monte Carlo risk assessment using belief networks to estimate the boundaries (i.e., minimum and maximum) of Beta-Pert distributions for activity durations. In their method, they assumed that, while experts can estimate the most-likely activity duration, the boundaries are more difficult to estimate using traditional methods [24]. The Bayesian belief network was, therefore, applied to integrate the risk factors affecting activity duration when calculating the optimistic and pessimistic boundaries of Beta-Pert. However, their model did not distinguish between the variability in activity duration and the risk impact. Furthermore, their model was built using multiple questionnaires, making it difficult for construction companies to implement this approach because of the number of experts and time required to complete the questionnaires.

2.2. Correlation and Dependence in Risk Assessment

A common source of error in MCS is the assumption of independence between the input variables of a model [32]. If two random variables are modeled as independent probability distributions in a simulation model, the sampled random variates of the two distributions will exhibit one of the following: (1) one variable is high, while the other is low, (2) both are high, or (3) both are low [25]. Disregarding the dependence or correlation in large risk models can result in estimation errors through under- or overestimation [25]. To improve risk analysis and subsequent decision-making, relationships between random variables must be considered [33].

Multiple research studies investigating the effect of correlation in Monte Carlo risk assessment models have been conducted. Touran and Wiser [32] presented a methodology that can account for the correlation between cost components in probabilistic cost estimation models. It is important to note, however, that their model only considered variability in the cost of and correlation between project work packages without considering external risk factors in their model. Touran [34] later extended the methodology by proposing a method that can account for subjective correlation between cost components by experts when historical data are absent.

Van Dorp and Duffey [25] proposed a method that can account for dependencies between activity durations when developing a project schedule network. The authors suggested that activities affected by the same external risk factor should be dependent and, therefore, that correlations between project activities can be determined from common risk factors that are shared. However, the correlation between cost and time impact of a risk factor was not considered in their model. Ökmen and Öztaş [35] developed a heuristic method for correlating schedule-risk analysis in construction schedule planning. Their model considered correlations between activities that are affected by the same risk factors, as well as correlations between the risk factors themselves. However, only the impact of risks on the schedule were modeled. Moret and Einstein [36] presented a model that considers correlations between the costs of activities in rail line construction when calculating the cost of a rail project. They then extended this by developing a comprehensive MCS model that considers (1) variability in activity duration and cost, (2) correlations between the cost of the activities, and (3) external risk factors affecting the project [37]. However, their model did not consider the correlation between the cost and schedule impact of a risk factor. Other research studies have attempted to use multivariate distributions for presenting the output of a simulation model for integrated cost and schedule-risk analysis [38,39]. While certain commercial software, such as @risk developed by the Palisade Group [40], have correlations implemented, dependencies between the cost and schedule impact of a risk factor have not been addressed by these previously developed models.

Copulas have been used in many applications to model dependencies between random variables. A copula-based joint distribution can be constructed using assessed rank-order correlations and marginal distributions, thereby reducing the effort required for assessments and to search for conditional independence [33]. Copulas are a flexible method, as they do not have any restrictions on the type of marginal distributions that can be used [41]. Using copulas to "couple" the marginal distribution requires two steps, namely (1) modeling the marginal distributions and (2) modeling dependencies between random variables [33]. Copulas have been used in risk management literature to successfully model the dependencies between variables that can affect the decision to purchase a used aircraft [33], in turn determining if the purchase of a used aircraft would generate more profit than using funds for other investments. Copulas have also been used to model dependencies between the activities of a project during the construction and scheduling of a project network [25].

2.3. Construction Risk Assessment in Onshore Wind Project and Its Challenges

Installing a wind turbine onto its foundation and completing final assembly appears on the surface to be straightforward. However, constructing a wind farm involves a long list of civil engineering and electrical work that require high levels of project management and coordination [42]. Once permits, approvals, and project finance are secured, the rigorous management of a complex series of engineering, logistical, and electrical processes must occur to reduce uncertainties and risks [42]. While many studies have been conducted to enhance risk management in wind projects, most of these studies have focused on the exploration and identification of risk factors affecting onshore wind projects. For example, while several studies have investigated which risk factors affected the planning, construction, and operation phases of wind projects [43,44], few studies have focused on methods for assessing the risk factors. Kucukali [45] developed a methodology for assessing the overall risk severity in wind projects based on a linguistic subjective scale. Rolik [46] proposed a Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis approach to assess the risk level in wind energy projects. Mohamed and colleagues [47] proposed a simulation-based approach to assess the severity of risk factors on the cost and time of onshore wind projects. Because of the lack of historical data, triangular and uniform distributions were used to depict the cost and schedule impact of the identified risk factors [47]. Notably, none of the aforementioned studies were capable of (1) incorporating large amounts of expert knowledge into simulation-appropriate input distributions and (2) considering the dependency between cost and schedule.

2.4. Fuzzy Logic

Fuzzy logic is often used to solve problems characterized by subjective uncertainty, ambiguity, and vagueness [48]. It has been widely applied by researchers in construction to incorporate the influence of factors that are linguistically assessed into an uncertain variable quantity [49,50]. The application of fuzzy logic provides a means of quantifying subjective evaluations and converting this information into a probability distribution function for input into MCS-based risk assessment models. The application of fuzzy logic for the linguistic assessment of root causes of risk factors in onshore wind projects, however, remains relatively unexplored.

3. Materials and Methods

This research proposes a method that is capable of addressing the current input modeling limitations of MCS risk assessment (Figure 3). The methodology has three main components, namely input data, data processing, and multivariate representation, which are detailed as follows.

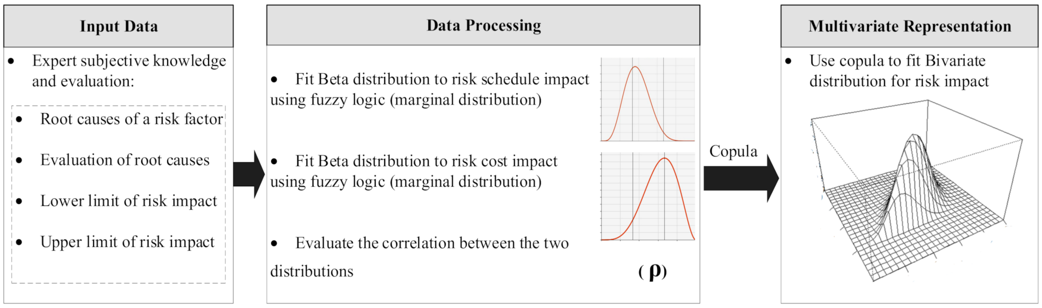


Figure 3. Research methodology.

3.1. Input Data

The input data component details the information required from the expert to successfully apply the methodology. Here, an expert provides detailed information about the impact of risks on schedule or cost, including (1) the minimum potential value of the risk impact (lower limit, A), (2) the maximum potential value of the risk impact (upper limit, B), and (3) root causes of the risk factor along with their evaluation.

To obtain this information, a root cause analysis, defined as "a structured investigation that aims to identify the true cause of a problem and actions necessary to eliminate it" [51], is performed to identify the potential root causes of different risk events [52,53]. Root cause analysis consists of five steps [51]: problem understanding, problem-cause brainstorming, data gathering, data analysis, and root-cause identification. Several tools and techniques can be applied for the identification of root causes, such as cause-and-effect charts, matrix diagrams, five whys, fault tree analysis, and failure mode and effect analysis. Here, the analyst may choose any root cause analysis method, provided that it is suited to their application.

After the comprehensive identification of the root causes, possible scenarios that may occur as a result of these root causes are then defined, ensuring that all possible combinations that can lead to the primary risk factor are captured. All defined root causes/scenarios are assessed in terms of frequency and adverse consequences on the overall schedule or cost. This is often described in subjective terms, such as "if the root cause 1 is very severe, it will significantly impact the total schedule or cost, and this is very likely." As discussed previously, because of the subjectivity of the problem domain, current methods are not able to consider root causes when deriving a probability distribution for the impact.

3.2. Data Processing

3.2.1. Marginal Distributions

Fuzzy set theory is then used to scientifically quantify the combined influence of the root causes to derive the marginal probability distribution functions that will be input into the MCS models. The workflow for deriving the marginal distributions is illustrated in Figure 4, and a step-by-step procedure of the proposed method is detailed in Supplement SA.

This study has chosen a Beta distribution to represent the marginal distribution of schedule or cost impact. A Beta distribution was selected because (1) its specialized form, the PERT distribution, is commonly applied in risk assessment studies [54], (2) it is a bounded distribution with finite limits, making it intuitively plausible to many decision-makers in risk analysis [54], (3) it is flexible, taking any shape according to its shape parameters (i.e., ß and α) [55,56], and (4) it is frequently used to describe variability or uncertainty over a fixed (i.e., bounded) range [18]. The steps for integrating subjective knowledge into the Beta distributions are detailed as follows.

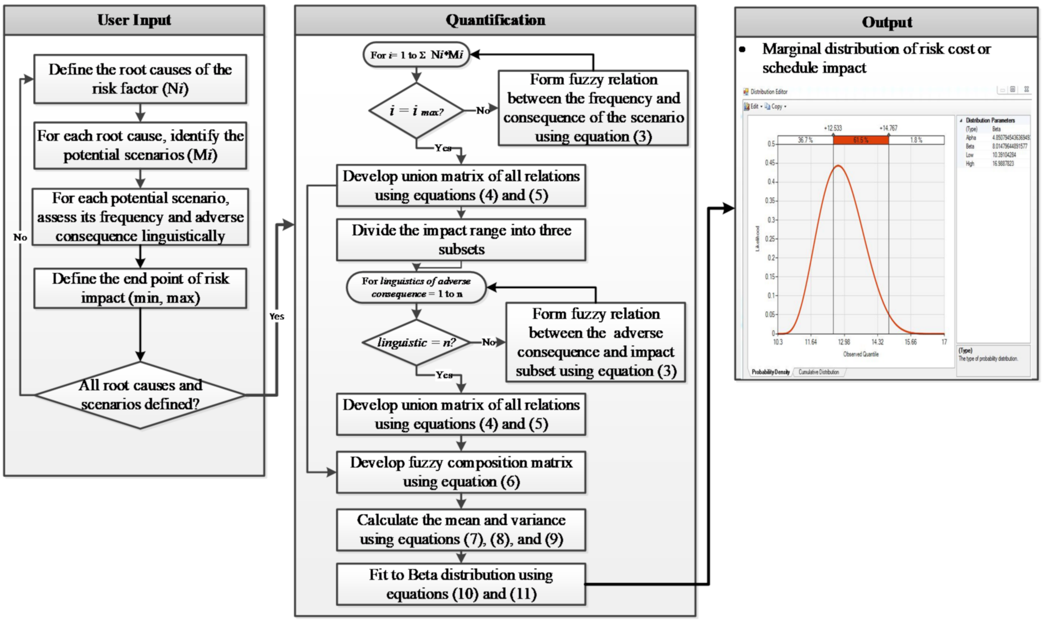


Figure 4. Quantification steps for marginal distribution.

A membership function is usually used in fuzzy sets to represent the relationship between a range of possible values and a linguistic term [57]. The membership function assigns a membership degree, in which the relation of these values to the linguistic term is defined within the interval [0, 1], representing no and full membership, respectively. In general, the membership of a fuzzy set, A, in the case of a discrete universe of discourse, X, is usually expressed as follows:

$$A = \sum_{1=1}^{n} \frac{\mu_i}{x_i} = \frac{\mu_1}{x_1} + \frac{\mu_2}{x_2} + \ldots\ldots + \frac{\mu_n}{x_n} = \sum \mu_A(x)/x, \tag{2}$$

where $\mu_i = \mu_i(x_i)$ is the degree of belonging of element x_i to set A, and n = the number of elements in set A. In construction applications, triangular and trapezoidal fuzzy numbers are usually used to represent the membership function. The development of the membership function, including the selection of linguistic terms and the range of values that a linguistic term represent, is determined by the analyst.

Once root causes are identified and assessed linguistically, the quantification analysis can be conducted. The fuzzy logic quantification algorithm combines the adverse consequence (C) (i.e., the contribution of the root cause to the overall cost or schedule-risk impact as a percentage) and the frequency of occurrence (F) for each root cause scenario. This is accomplished by calculating a fuzzy relation matrix between (F) and (C), resulting in R (F, C), which is the Cartesian product $F \times C$. The elements of R (F, C) are computed as follows:

$$\mu_R(x_i, y_i) = min[\mu_F(x_i), \mu_C(y_i)], \tag{3}$$

where x_i = an element of universe X; y_i= an element of universe Y; $\mu_R(x_i, y_i)$ = the membership value of element (x_i, y_i) in the fuzzy relation R; min = the minimum values of both elements x_i and y_i; $\mu_F(x_i)$ = the membership value of element x_i in fuzzy set F; and $\mu_C(y_i)$ = the membership value of element y_i in fuzzy set C.

Once all of the fuzzy relation matrices have been calculated, the fuzzy logic quantification algorithm calculates the union matrix of all the fuzzy relation matrices, thereby representing the combined effect of all root cause scenarios. The union of two fuzzy relations, for example S and Z, is denoted by $S \bigcup Z$, and the membership function is calculated as follows:

$$\mu_{S \cup Z}(x_i, y_i) = max[\mu_s(x_i, y_i), \mu_z(x_i, y_i)], \tag{4}$$

where *max* = the maximum value of both relations *s* and *z*.

Union *U*, between the fuzzy relation matrices *R* (*F*, *C*), is then computed as:

$$U = max\ [(F_1 \times C_1) \cup (F_2 \times C_2) \cup (F_3 \times C_3)\ \ldots\ldots\ldots\ldots.. \cup (F_k \times C_k)], \tag{5}$$

where *max* = the maximum value of the two relations.

The next step uses expert knowledge to subjectively assess the relationship between the adverse consequence of a root cause and the overall cost or schedule impact of a risk factor. For example, "if the adverse consequence of root 1 is large, then the overall risk impact is medium." The range of a risk factor impact (i.e., between minimum and maximum) is then calibrated by mapping the range to a predefined scale using the concept of membership values. The mapping values represent the confidence level with which the expert believes that a particular value belongs to the set [58] and that the impact will be in a certain range. To facilitate mapping, the risk impact range is divided into three equidistant subsets. Notably, Beta distributions can take multiple shapes, of which three cases, as shown in Figure 5, are of interest for calibration. These are shapes that are (1) skewed to the upper limit, (2) skewed to the lower limit, or (3) symmetric. Cases where the distribution is skewed toward the lower limit are mapped to the small impact range; cases where the distribution is symmetric are mapped to the medium impact range; while cases where the distribution is skewed toward the upper limit are mapped to the large impact range.

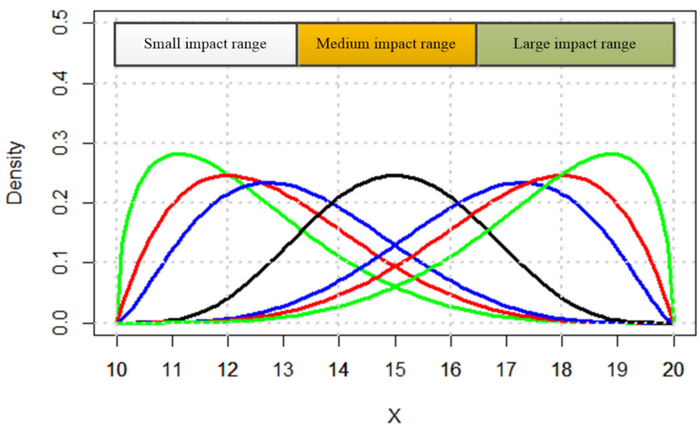


Figure 5. Calibration of risk impact range.

Once all of the adverse consequences of root causes are related to the overall risk impact, a fuzzy relation Q (*C*, *I*), which is the Cartesian product $C \times I$ between fuzzy subset *C*, representing the adverse consequence, and fuzzy subset *I*, representing the overall risk impact as per Equation (3), is computed. After forming all fuzzy relations matrices, a union matrix, *V*, of all relations is calculated using Equation (5).

Finally, a fuzzy composition of the union matrices *U* and *V* is calculated to assess the overall combined impact of all root causes. The composition is defined by Equation (6), which is the standard max-min composition as follows:

$$U \circ V(x_i, z_k) = \max_{y_j}\left\{min\left[\mu_U(x_i, y_i), \mu_V\left(y_j, z_k\right)\right]\right\}, \tag{6}$$

where $U \circ V(x_i, z_k)$ = membership value of element (x_i, z_k) in composition matrix between *U* and *V*; $\mu_U(x_i, y_i)$ = membership value of element (x_i, y_i) in union matrix *U*; and $\mu_V\left(y_j, z_k\right)$ = membership value of element $\left(y_j, z_k\right)$ in union matrix *V*.

A fuzzy subset from the composition matrix is used to represent the overall uncertain variable under study [46]. A fuzzy subset (e.g., subset = *O*), or one row from the matrix, will be selected such that the product of the row summation and the corresponding frequency of occurrence is the maximum. Then, the selected fuzzy subset (e.g., subset = O) will be used to calculate the mean, μ_I, and the variance, $\sigma^2{}_I$, of the marginal risk impact according to the following equations derived from [50,59–61]:

$$P(R_I = z_k) = \frac{\mu_o(z_k)}{\sum_1^m \mu_o(z_k)}, \tag{7}$$

$$\mu_I = \sum\nolimits_{k=1}^{m} (z_k) * P(R_I = z_k), \tag{8}$$

$$\sigma^2{}_I = \left[\sum\nolimits_{k=1}^{m} (z_k)^2 * P(R_I = z_k)\right] - \mu_I{}^2, \tag{9}$$

where R_I = risk impact; z_k = element of the risk impact; $P(R_I = z_k)$ = probability of occurrence of the risk impact to be element z_k; $\mu_o(z_k)$ = membership value of element z_k in subset O; and m = number of risk impact elements in subset O.

Once the mean and variance are calculated, they are used, along with the minimum (A) and the maximum (B), to derive a generalized Beta distribution. Equations (8) and (9) together with terms A and B are used to estimate the shape parameters, α and ß, of the generalized Beta distribution as follows [55]:

$$\alpha = \frac{\mu_I - A}{B - A}\left[\frac{(\mu_I - A)(B - \mu_I)}{\sigma^2{}_I} - 1\right] \text{and} \tag{10}$$

$$ss = \alpha\left[\frac{B - \mu_I}{\mu_I - A}\right] \tag{11}$$

The probability density function is then visualized using shape parameters α and ß together with the end points of distributions A and B.

3.2.2. Correlation of Dependent Variables

Using copula-based multivariate modeling requires the correlation between dependent variables to be evaluated. When historical data are lacking, a subjective evaluation of the correlation from experts is acquired. Here, the Spearman correlation (ρ) is used to measure the association between random variables because of its ability to capture relationships through a pairwise measure of dependence [25,33]. While the Spearman correlation coefficient varies between −1 and +1, most correlations between cost and schedule in construction are positive, limiting the values in this application to between 0 and 1 [34]. Correlations are classified into three categories, as proposed by Touran [34], to reflect the linguistic representation of correlations often used by experts, namely weak (0–0.3), moderate (0.3–0.6), and strong (0.6–1). The midpoint of each interval is chosen to represent the interval (i.e., 0.15 for weak, 0.45 for moderate, and 0.8 for strong). Consequently, the correlation matrix (R) will take one of the following forms:

$$R_{\text{weak}} = \begin{bmatrix} 1 & 0.15 \\ 0.15 & 1 \end{bmatrix}, \tag{12}$$

$$R_{\text{moderate}} = \begin{bmatrix} 1 & 0.45 \\ 0.45 & 1 \end{bmatrix}, \text{ and} \tag{13}$$

$$R_{\text{strong}} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix} \tag{14}$$

3.3. Multivariate Representation

The last component of the methodology is the multivariate representation of the marginal distributions of the schedule- and cost-risk impacts that is achieved using copulas [13]. There are many classes of copulas, including elliptical, Archimedean, and Marshall-Olkin [41]. Selection criteria for specific classes of copulas have yet to be established [36]; therefore, the class of copula selected is up to the discretion of the analyst.

4. Application and Results

An illustrative case study is presented to demonstrate the functionality of the proposed approach. The example risk factor chosen for illustration is public obstruction during the construction phase of the project, which is a critical risk factor that is known to cause project delays and financial losses [62].

4.1. Input Data

Four root causes were identified, namely noise due to construction activities, harm to business and farming activities of the local community, traffic disturbance due to logistic and supply chain to the construction site, and poor communication with the local community. Scenarios/root causes that may occur are detailed in Table 1.

Table 1. Root causes of risk factors with accompanying scenarios and assessment.

No.	Root Cause/Scenario	Frequency of Occurrence (*F*)	Adverse Consequence (*C*)
1	Construction noise is low	Likely	Very small
2	Construction noise is medium	Likely	Large
3	Construction noise is high	Unlikely	Large
4	Harm to activities is low	Unlikely	Small
5	Harm to activities is medium	Somewhat likely	Large
6	Harm to activities is high	Unlikely	Very large
7	Traffic disturbance is low	Very likely	Very small
8	Traffic disturbance is medium	Somewhat likely	Large
9	Traffic disturbance is high	Unlikely	Very large
10	Poor communication	Unlikely	Medium

4.2. Data Processing

4.2.1. Marginal Distributions

The assessment (Table 1) includes the frequency of occurrence (*F*) and adverse consequence (*C*) of the scenario/root cause in linguistic terms. The linguistic terms were then represented using a membership function chosen by the analyst. In this example, the following membership functions, adopted from [58], were used to represent (*F*) and (*C*), as shown in Tables 2 and 3, respectively:

Table 2. Membership function for frequency of occurrence (*F*).

Element of Linguistic Variable	Frequency of Occurrence (*F*)				
	Very Unlikely	Unlikely	Somewhat Likely	Likely	Very Likely
0	1	0	0	0	0
0.1	0.8	0.8	0	0	0
0.2	0.2	1.0	0	0	0
0.3	0	0.8	0.5	0	0
0.4	0	0	0.8	0	0
0.5	0	0	1	0.5	0
0.6	0	0	0.8	0.8	0
0.7	0	0	0.5	1.0	0.5
0.8	0	0	0	0.8	0.8
0.9	0	0	0	0.6	0.9
1.0	0	0	0	0	1

Table 3. Membership function for adverse consequence (C).

Element of Linguistic Variable	Adverse Consequence (C)				
	Very Small	**Small**	**Medium**	**Large**	**Very Large**
0	1	1	0	0	0
0.1	0.81	0.9	0	0	0
0.2	0.25	0.5	0	0	0
0.3	0	0	0.2	0	0
0.4	0	0	0.8	0	0
0.5	0	0	1	0	0
0.6	0	0	0.8	0	0
0.7	0	0	0.2	0	0
0.8	0	0	0	0.5	0.25
0.9	0	0	0	0.9	0.81
1.0	0	0	0	1	1

Once the linguistic assessment was conducted for all scenarios, fuzzy relations were constructed using Equation (3). Table 4 shows the fuzzy relation of the first scenario (i.e., construction noise is low). Remaining scenarios were represented similarly (data not shown).

Table 4. Fuzzy relation R (F, C).

Frequency of Occurrence (F)	Adverse Consequence (C)		
	0	**0.1**	**0.2**
0.5	0.5	0.5	0.25
0.6	0.8	0.8	0.25
0.7	1.0	0.81	0.25
0.8	0.8	0.8	0.25
0.9	0.6	0.6	0.25

After all relations were determined, a fuzzy union matrix U of all relations was established using Equations (4) and (5). Table 5 shows the fuzzy union matrix of all relations.

Table 5. Fuzzy union matrix (U) of all relations of the scenarios.

Frequency of Occurrence (F)	Adverse Consequence (C)										
	0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
0	0	0	0	0	0	0	0	0	0	0	0
0.1	0.8	0.8	0.5	0.2	0.8	0.8	0.8	0.2	0.5	0.8	0.8
0.2	1.0	0.9	0.5	0.2	0.8	1.0	0.8	0.2	0.5	0.9	1
0.3	0.8	0.8	0.5	0.2	0.8	0.8	0.8	0.2	0.5	0.8	0.8
0.4	0	0	0	0	0	0	0	0	0.5	0.8	0.8
0.5	0	0	0	0	0	0	0	0	0.5	0.9	1
0.6	0	0	0	0	0	0	0	0	0.5	0.8	0.8
0.7	1.0	0.81	0.25	0	0	0	0	0	0.5	0.9	1
0.8	0.8	0.8	0.25	0	0	0	0	0	0.5	0.8	0.8
0.9	0.9	0.81	0.25	0	0	0	0	0	0.5	0.5	0.5
1.0	1	0.81	0.25	0	0	0	0	0	0	0	0

Then, the end points of the cost-risk impact distribution were identified by the analyst as minimum = $ 10,000 and maximum = $ 40,000. The range was divided into three equidistant sections, each with the value of $ 10,000. Then, at the analyst's discretion, each range impact was divided into five elements. A membership degree for each element in each impact range was assigned by the analyst. The elements in each range along with the membership degrees are provided in Figure 6.

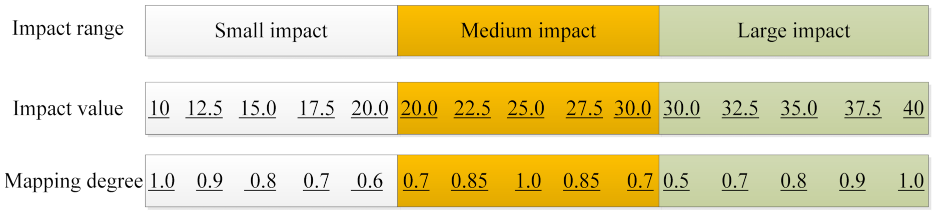


Figure 6. Ranges of risk impact along with values and their membership degree.

The adverse consequence of root causes was related to the overall risk impact subjectively (i.e., if the consequence is very small, the impact will be small) by the expert, as shown in Table 6.

Table 6. Relationships between adverse consequence (*C*) and impact (*I*).

No.	Adverse Consequence (*C*)	Impact (*I*)
1	Very small	Small
2	Small	Small
3	Medium	Medium
4	Large	Large
5	Very large	Large

Once assessed, each relationship was represented using a fuzzy relation matrix as per Equation (3). An example of fuzzy relation for row No. 3, Q (C_{medium}, I_{medium}), is provided Table 7.

Table 7. Fuzzy relation Q (*C*, *I*) between medium adverse consequence (*C*) and medium impact (*I*).

Adverse Consequence (*C*)	Impact (*I*)				
	20.0	22.5	25.0	27.5	30.0
0.3	0.2	0.2	0.2	0.2	0.2
0.4	0.7	0.8	0.8	0.8	0.7
0.5	0.7	0.85	1	0.85	0.7
0.6	0.7	0.8	0.8	0.8	0.7
0.7	0.2	0.2	0.2	0.2	0.2

After establishing all relations, a fuzzy union matrix, *V*, between all relations was established, as detailed in Table 8, using Equations (4) and (5).

Table 8. Fuzzy union matrix *V* of all relationships between adverse consequence (*C*) and impact (*I*).

Adverse Conseq.	Impact $\times 10^3$												
	10	12.5	15	17.5	20	22.5	25	27.5	30	32.5	35	37.5	40
0	1.0	0.9	0.8	0.7	0.6	0	0	0	0	0	0	0	0
0.1	0.9	0.9	0.8	0.7	0.6	0	0	0	0	0	0	0	0
0.2	0.5	0.5	0.5	0.5	0.5	0	0	0	0	0	0	0	0
0.3	0	0	0	0	0.2	0.2	0.2	0.2	0.2	0	0	0	0
0.4	0	0	0	0	0.7	0.8	0.8	0.8	0.7	0	0	0	0
0.5	0	0	0	0	0.7	0.85	1.0	0.85	0.7	0	0	0	0
0.6	0	0	0	0	0.7	0.8	0.8	0.8	0.7	0	0	0	0
0.7	0	0	0	0	0.2	0.2	0.2	0.2	0.2	0	0	0	0
0.8	0	0	0	0	0	0	0	0	0.5	0.5	0.5	0.5	0.5
0.9	0	0	0	0	0	0	0	0	0.5	0.7	0.8	0.9	0.9
1.0	0	0	0	0	0	0	0	0	0.5	0.7	0.8	0.9	1.0

Fuzzy composition matrix $U \circ V(x_i, z_k)$ of the union matrices *U* and *V* (Tables 5 and 8, respectively) was calculated using Equation (6). The resulting matrix is provided as Table 9. The multiplication of the rows' summation and frequency of occurrence is provided. Row No. 9 (Table 9, grey) was selected, as it provided the maximum value for the product.

Table 9. Composition matrix $U \circ V$ of all relationships between adverse consequence (C) and impact (I).

Frequency of Occurrence	**Impact $\times 10^3$**													**Row Sum**	**Multiplication**
	10.0	**12.5**	**15.0**	**17.5**	**20.0**	**22.5**	**25.0**	**27.5**	**30.0**	**32.5**	**35.0**	**37.5**	**40.0**		
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.1	0.8	0.8	0.8	0.7	0.7	0.8	0.8	0.8	0.7	0.7	0.8	0.8	0.8	10.0	1.00
0.2	1.0	0.9	0.8	0.7	0.7	0.85	1	0.85	0.7	0.7	0.8	0.9	1.0	10.9	2.18
0.3	0.8	0.8	0.8	0.7	0.7	0.8	0.8	0.8	0.7	0.7	0.8	0.8	0.8	10.0	3.00
0.4	0	0	0	0	0	0	0	0	0.5	0.7	0.8	0.8	0.8	3.6	1.44
0.5	0	0	0	0	0	0	0	0	0.5	0.7	0.8	0.9	1.0	3.9	1.95
0.6	0	0	0	0	0	0	0	0	0.5	0.7	0.8	0.8	0.8	3.6	2.16
0.7	1.0	0.9	0.8	0.7	0.6	0	0	0	0.5	0.7	0.8	0.9	1.0	7.9	5.53
0.8	0.8	0.8	0.8	0.7	0.6	0	0	0	0.5	0.7	0.8	0.8	0.8	7.3	5.84
0.9	0.9	0.9	0.8	0.7	0.6	0	0	0	0.5	0.5	0.5	0.5	0.5	6.4	5.76
1.0	1.0	0.9	0.8	0.7	0.6	0	0	0	0	0	0	0	0	4.0	4.00

After selecting the subset, row No. 9, that represents the cost-risk impact, the probability of each element in the cost-risk impact subset was calculated, according to Equation (7), as follows:

$$P(R_I = 10) = P(R_I = 12.5) = P(R_I = 15) = P(R_I = 35) = P(R_I = 37.5) = P(R_I = 40)\tfrac{0.8}{7.3} = 0.1095,$$
$$P(R_I = 17.5) = P(R_I = 32.5) = \tfrac{0.7}{7.3} = 0.0958$$
$$P(R_I = 22.5) = P(R_I = 25) = P(R_I = 27.5) = \tfrac{0}{7.3} = 0,$$
$$P(R_I = 30) = \tfrac{0.5}{7.3} = 0.0685,$$
$$P(R_I = 20) = \tfrac{0.6}{7.3} = 0.082$$

The mean and the variance were calculated, according to Equations (8) and (9), as follows:

$$\begin{aligned}\mu_I = [(10 * 0.1095) \quad & +(12.5 * 0.1095) + (15 * 0.1095) + (17.5 * 0.0958) \\ & +(20 * 0.082) + (22.5 * 0) + (25 * 0) + (27.5 * 0) + (30 * 0.0685) \\ & +(32.5 * 0.0958) + (35 * 0.1095) + (37.5 * 0.1095) \\ & +(40 * 0.1095)] * 10^3 = 24910\ \$\end{aligned}$$

$$\begin{aligned}\sigma^2{}_I = \Big[(10^2 * 0.1095) \quad & +(12.5^2 * 0.1095) + (15^2 * 0.1095) + (17.5^2 * 0.0958) \\ & +(20^2 * 0.082) + (22.5^2 * 0) + (25^2 * 0) + (27.5^2 * 0) + (30^2 * 0.0685) \\ & +(32.5^2 * 0.0958) + (35^2 * 0.1095) + (37.5^2 * 0.1095) \\ & +(40^2 * 0.1095)\Big] * 10^6 - (24910)^2 = 120488150\end{aligned}$$

Finally, the shape parameters of the risk impact Beta distribution were calculated, based on the minimum, maximum, mean, the variance, according to Equations (10) and (11), as follows:

$$\alpha = \frac{24910 - 10000}{40000 - 10000}\left[\frac{(24910 - 10000)(40000 - 24910)}{120488150} - 1\right] = 0.43$$

$$ss = 0.43\left[\frac{40000 - 24910}{24910 - 10000}\right] = 0.43$$

The resulting probability density function of the Beta distribution is presented in Figure 7.

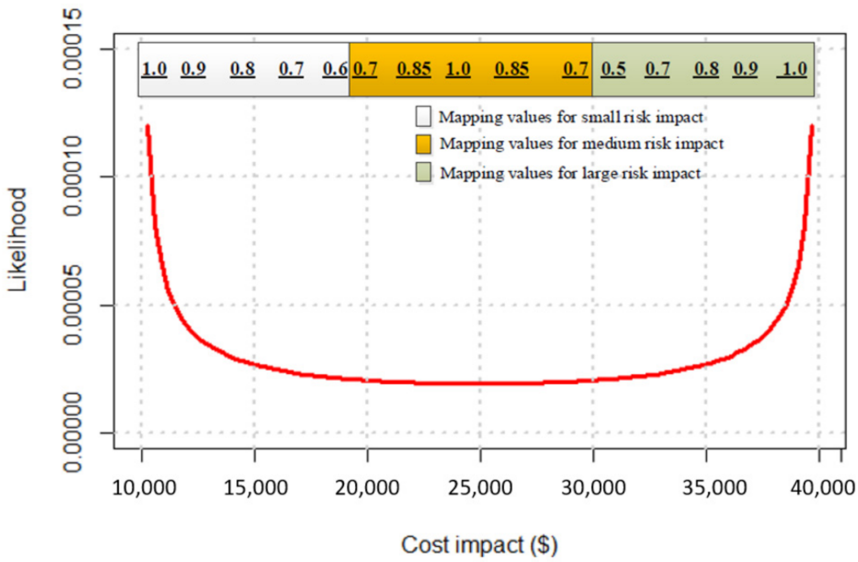


Figure 7. Marginal Beta distribution of cost-risk impact.

The marginal distribution of the schedule-risk impact was constructed in a manner similar to the marginal distribution of the cost-risk impact. The same root causes were used together with their evaluation in terms of frequency of occurrence (F) and adverse consequence (C). The mapping values were also the same as the cost-risk impact. Only the lower and upper limits of the distribution were changed. The lower limit was set to 1 day and the upper limit to 10 days. The range was divided into three subranges, namely small (1, 2, 3, 4), medium (4, 5, 6, 7), and large (7, 8, 9, 10), as shown in Figure 8, along with their mapping values. The shape parameters α = 0.412 and ß = 0.523 and the fitted Beta distribution is presented in Figure 8.

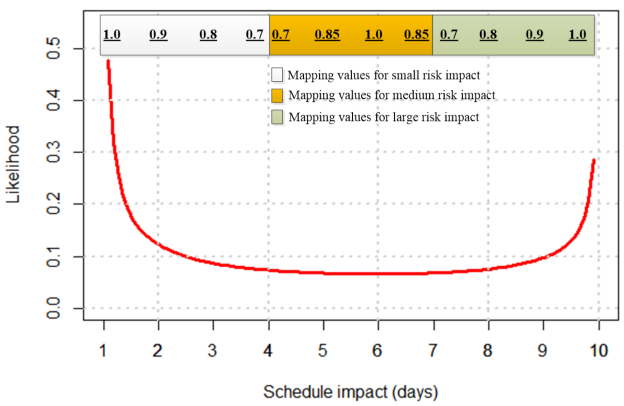


Figure 8. Marginal Beta distribution of schedule-risk impact.

4.2.2. Validation of the Marginal Distribution

Two approaches, namely sensitivity analysis and expert validation, were used to investigate the proposed methodology. In the sensitivity analysis, the influence of input parameters on the fitted distribution was investigated, while the expert validation analysis was performed to evaluate the proposed approach from the perspective of an expert.

- Sensitivity Analysis

Sensitivity analysis is a powerful technique for testing the internal consistency and reliability of models [63] by analyzing the model behavior in response to variations in input values or parameters.

The purpose of conducting a sensitivity analysis is to assess how the input variables of a model affect the resulting probability distribution, allowing analysts to identify the input variables with the greatest influence on the output of a model or system [64]. Two approaches are commonly used to conduct a sensitivity analysis: local (also known as absolute [65]) and global. The local approach allows only one variable of a model to be assessed at a time, while fixing all other variables to their original values [64]. Conversely, global sensitivity allows all input factors to vary simultaneously, where the sensitivity is evaluated over the entire range of each input factor [64].

This study applied a local approach to investigate sensitivity of three input variables: (1) selection of linguistic variables; (2) membership values of the linguistic variables; and (3) membership values for calibrating and mapping the risk impact range. Sensitivity was calculated manually by changing the values of the targeted parameters and examining the effect on the resulting marginal Beta distribution. Parameters resulting in the greatest change were deemed to have the greatest impact on outputs. Thus, experts are encouraged to remain cognizant of such parameters during the application of the proposed method. Sensitivity of these parameters was tested on the parameters of the Beta distribution. The selection of linguistic terms is at the discretion of the analyst [61]; however, the range that each linguistic term represents must be investigated. The illustrative case study was investigated again using different membership values for (*F*) and (*C*). First, the sensitivity of frequency of occurrence (*F*) was tested by introducing a new membership function, adopted from [50], where the same linguistic terms are used but the membership values were changed. Results are detailed in Table 10.

Table 10. New membership function for frequency of occurrence (*F*).

Element of Linguistic Variable	Frequency of Occurrence (*F*)				
	Very Unlikely	Unlikely	Somewhat Likely	Likely	Very Likely
0	1	0	0	0	0
0.1	1	0	0	0	0
0.2	0.5	0.5	0	0	0
0.3	0	1	0	0	0
0.4	0	0.5	0.5	0	0
0.5	0	0	1	0	0
0.6	0	0	0.5	0.5	0
0.7	0	0	0	1.0	0
0.8	0	0	0	0.5	0.5
0.9	0	0	0	0	1
1.0	0	0	0	0	1

The shape parameters $\alpha = 0.389$ and ß = 0.392 and the fitted Beta distribution is shown in Figure 9. Differences between the original distribution and the one derived following the change the in membership function of the frequency of occurrence (*F*) were minimal. The results demonstrate that small variations in membership values have little effect on the output distribution.

Next, the influence of changing the membership function of the adverse consequence (*C*) was investigated by introducing a new trapezoidal membership function, as shown in Table 11. The shape parameters $\alpha = 0.592$ and ß = 0.713 and the fitted Beta distribution are presented in Figure 10. Again, the differences between the original distribution and the one derived following the change in the membership function of the adverse consequence (*C*) were minimal. The results demonstrate that the output distribution is not sensitive to changes in the membership function of the adverse consequence (*C*).

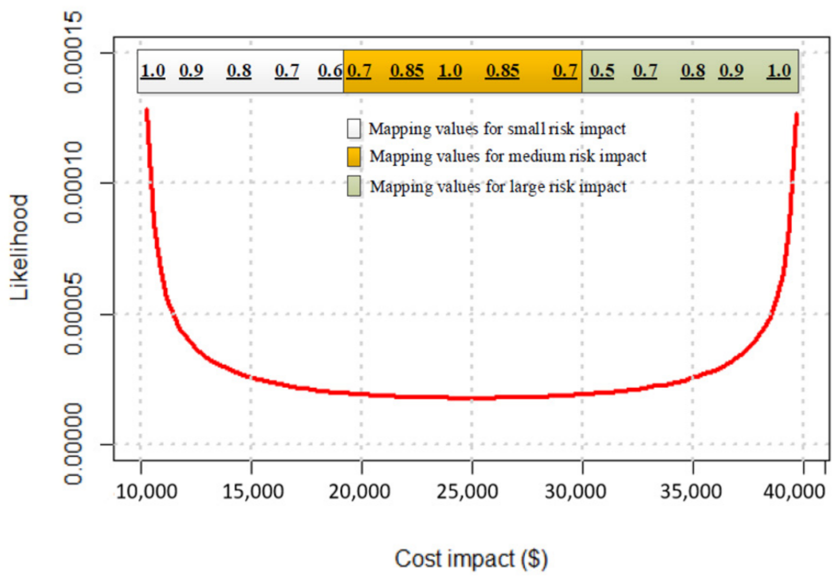


Figure 9. Marginal Beta distribution of cost-risk impact after changing the membership function of (*F*).

Table 11. New membership function for adverse consequence (*C*).

Element of Linguistic Variable	Adverse Consequence (*C*)				
	Very Small	Small	Medium	Large	Very Large
0	1	0	0	0	0
0.05	1	0	0	0	0
0.1	0.5	0	0	0	0
0.15	0	0.5	0	0	0
0.2	0	1.0	0	0	0
0.25	0	1.0	0	0	0
0.3	0	0.5	0	0	0
0.35	0	0	0.5	0	0
0.4	0	0	1	0	0
0.45	0	0	1	0	0
0.5	0	0	1	0	0
0.55	0	0	1	0	0
0.6	0	0	0.5	0	0
0.65	0	0	0	0.5	0
0.7	0	0	0	1.0	0
0.75	0	0	0	1.0	0
0.8	0	0	0	0.5	0
0.85	0	0	0	0	0.5
0.9	0	0	0	0	1
0.95	0	0	0	0	1
1.0	0	0	0	0	1

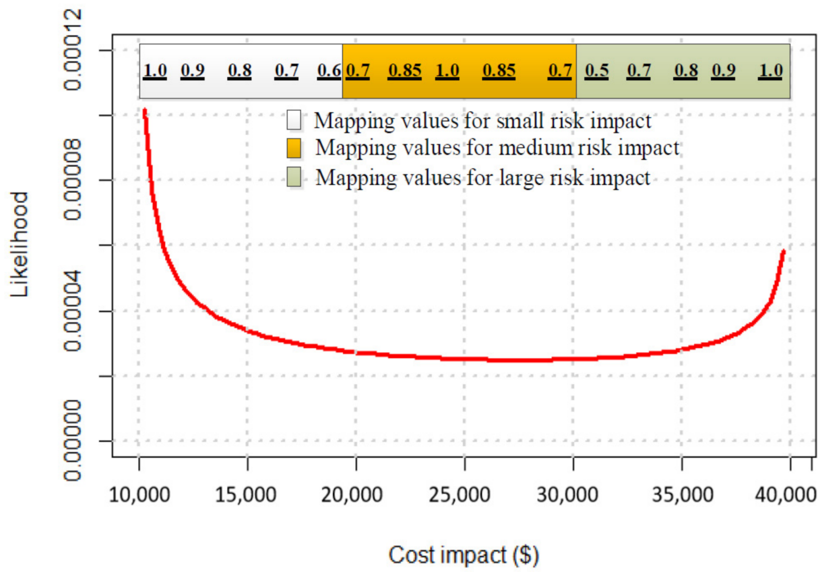


Figure 10. Marginal Beta distribution of cost-risk impact after changing membership function of (*C*).

The final sensitivity experiment investigated the influence of changing the membership values for calibrating the risk impact range. This was conducted by experimenting with different mapping values, while keeping all other values (i.e., membership functions for frequency of occurrence and adverse consequences, boundaries of the risk impact, and root causes and their evaluation) unchanged and set to their original values. The impact mapping values for each impact subset (i.e., "small impact", "medium impact", or "large impact", as in Figure 6) were changed for each trial. Specific changes are summarized in Table 12. The parameters of the resulting distributions are detailed in Table 13, and the resulting probability density functions are shown in Figure 11a–f.

Table 12. Summary of sensitivity analysis experimental parameters on mapping values.

Trial	Impact Mapping Values [1]			PDF [2]	Distribution
	Small	Medium	Large		
A	1	0	0	Figure 11a	Skews right, as values for small impact subset are equal to 1.
B	0	1	0	Figure 11b	Symmetric, with peak at middle where mapping values equal 1.
C	0	0	1	Figure 11c	Skews left, as values for large impact subset are equal to 1.
D	↓	peak = 1, other = ↓	↓	Figure 11d	Symmetric as in trial (b), but with greater variance.
E	↑ until joint point	↓ until joint point	0	Figure 11e	Skews towards small and medium joint point, where values are greatest.
F	0	↑ until joint point	↓ until joint point	Figure 11f	Skews towards medium and large joint point, where values are greatest

[1] As defined in Table 6. [2] Probability density function. Where ↓ = decreased gradually, and ↑ = increased gradually.

Table 13. Results of sensitivity analysis of mapping values.

Trial	Statistical Parameters of Beta Distribution			
	Minimum ($)	Maximum ($)	α	ß
A	10,000	40,000	1.5	7.5
B	10,000	40,000	8.5	8.5
C	10,000	40,000	7.5	1.5
D	10,000	40,000	1.3	1.3
E	10,000	40,000	12.035	32.785
F	10,000	40,000	35.175	13.44

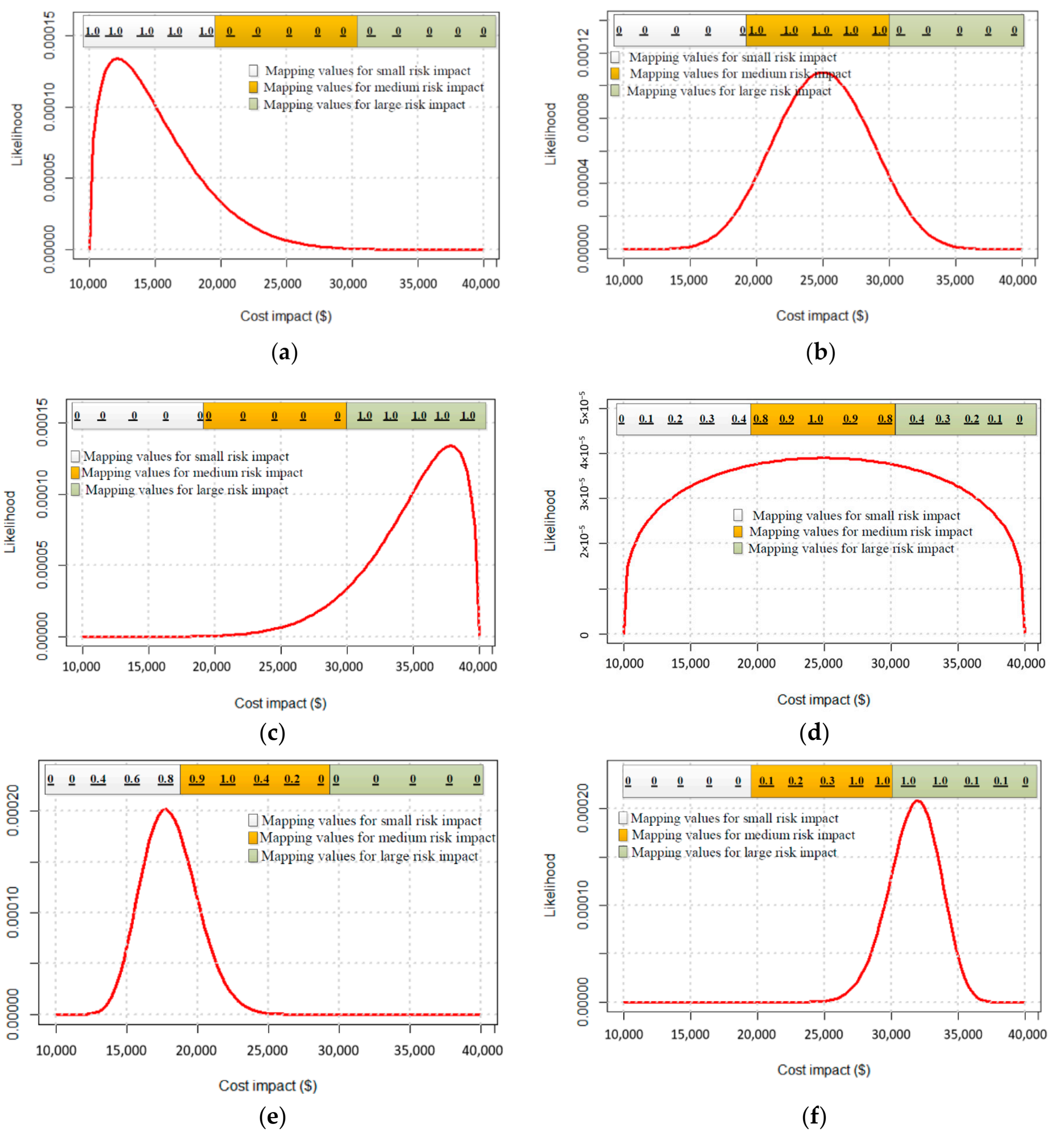


Figure 11. Marginal Beta distributions after changing the mapping value of the risk impact range in (**a**) Trial A, (**b**) Trial B, (**c**) Trial C, (**d**) Trial D, (**e**) Trial E, and (**f**) Trial F.

Results of the sensitivity analysis demonstrate that the proposed method is not particularly sensitive to changes in the membership function of the frequency of occurrence (*F*) nor the adverse consequence (*C*). In contrast, the model was shown to be sensitive to changes in the mapping values of the risk impact range. This finding indicates the importance of capturing an expert's perceived impact values as precisely as possible. As discussed previously, decomposition of (1) a risk factor into its root causes and/or (2) a risk impact range into smaller subsets can be used to reduce biases, thereby enhancing the comprehensiveness and accuracy of model results. This explains the shape of the distribution observed in the illustrative case study (Figures 9 and 10), where the expert was not confident about the mapping values for the risk impact, resulting in distribution with a uniform-like appearance.

- Expert Validation

Expert face validation is used to assess the practical applicability of a proposed method by having a subject-matter expert evaluate the results of a proposed method [63]. The illustrative case study results and sensitivity analysis were discussed with three experts (i.e., a director, a project manager, and project coordinator) from a large construction company in Alberta, Canada—each with an average of 15 years of practical experience in construction management and risk analysis. Definitions of the terms representative, comprehensive, and ease of use (described as follows) were provided to the experts. Then, experts were asked whether or not they believed that the proposed method was characterized by each definition. Results are detailed as follows.

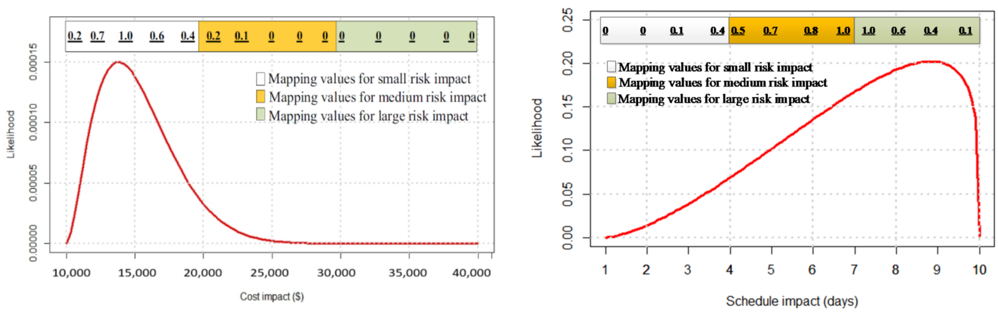
The experts agreed that the proposed method was representative, defined here as the ability of a method to representatively express the knowledge of the expert as a probability distribution function. In particular, findings that distribution shape is sensitive to the mapping values of the range impact increased their confidence that the representation was appropriate.

The experts also agreed that the proposed method was comprehensive, defined here as the ability of a method to include available information regarding the risk impact. They were satisfied with the level of information that the method allows them to incorporate while deriving the probability distribution.

In contrast, the experts indicated that the proposed method was, in its current form, not easy to use—particularly by analysts that may not be familiar with fuzzy logic computations. They agreed that full computerization of the approach would considerably facilitate its application in industry. Accordingly, this method was computerized within an in-house developed simulation engine, *SimphonyProject.NET* [47], developed for integrated assessment of risks.

4.3. Multivariate Representation

Following expert validation, a marginal distribution of the cost and schedule-risk impact based on expert knowledge was determined, as shown in Figure 12. The parameters of the final cost marginal distribution were lower limit = \$10,000, upper limit = \$40,000, shape parameter $\alpha = 2.7$, and ß = 12.745; and the parameters of the final schedule marginal distribution were lower limit = 1 day, upper limit = 10 days, shape parameter $\alpha = 2.26$, and ß = 1.13. Membership functions provided in Tables 2 and 3 were used to derive the final marginal distributions. The steps of deriving the marginal distribution of the cost-risk impact are detailed in Supplement SB. To represent the cost and schedule-risk impact of the risk factor using multivariate distribution, an expert was asked to subjectively evaluate the correlation as either weak, moderate, or strong. A strong correlation between cost and schedule impact was evaluated, and the correlation matrix in Equation (14) was consequently used in the bivariate distribution.



(a) (b)

Figure 12. Final marginal distribution of the (**a**) cost impact and (**b**) schedule impact.

Here, a multivariate normal copula [13] was chosen because it (1) does not have constraints on the marginal distributions and (2) has been successfully applied in other risk assessment studies with Beta as marginal distributions [33]. A copula package in *R* [13] was used to implement the multivariate modeling of the dependence of the cost- and schedule-risk impacts. The package allows the user to define the dependence structure and the marginal distributions separately. The dependence structure consists of the number of correlated random variables and the correlation coefficient. The marginal distributions were defined using extraDistr package [66], which allows for a generalized beta distribution that is not bound between 0 and 1. The dependence structure between the cost and schedule impact, specifically the two marginal distributions and the correlation between them, is presented in Figure 13. The output joint probability density function of the bivariate fitted distribution is presented in Figure 14 as a contour plot; a representation of the 3-dimentional surface of the joint probability distribution function is shown in Figure 15. The cumulative density function of the bivariate joint distribution is shown in Figure 16. It is this fitted distribution that will be used to model the risk factor (i.e., public obstruction) with correlated cost and schedule impact in a MCS experiment.

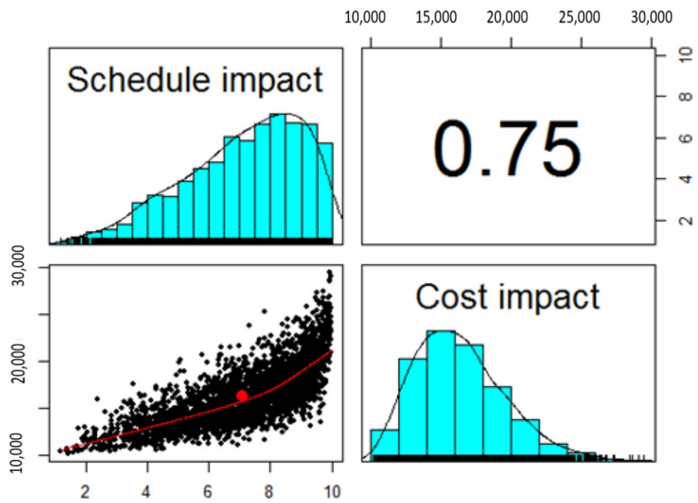


Figure 13. Dependence structure between cost and schedule-risk impact.

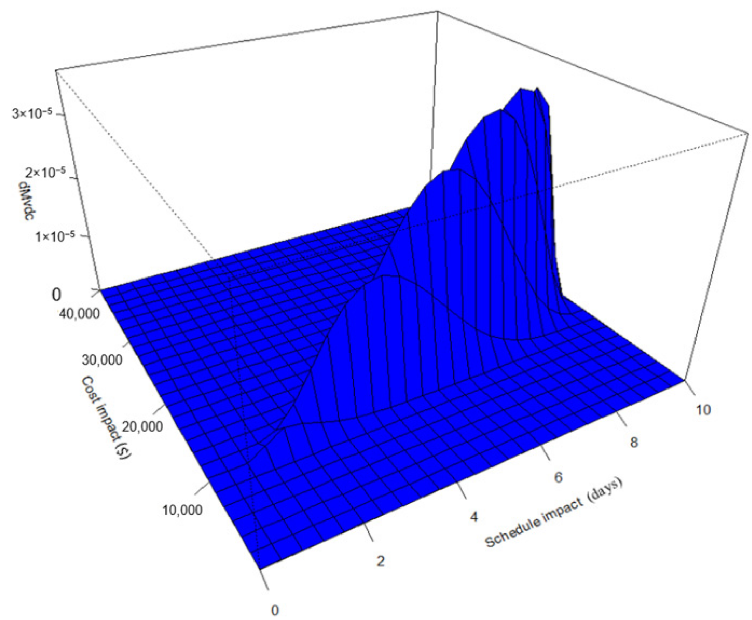


Figure 14. Probability density function of the joint distribution.

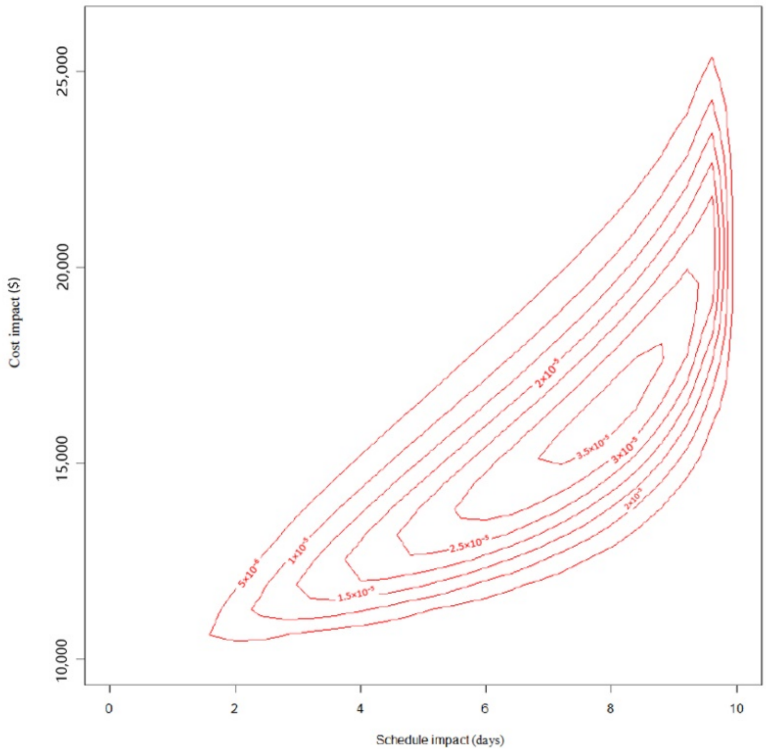


Figure 15. Contour plot of the joint probability density function.

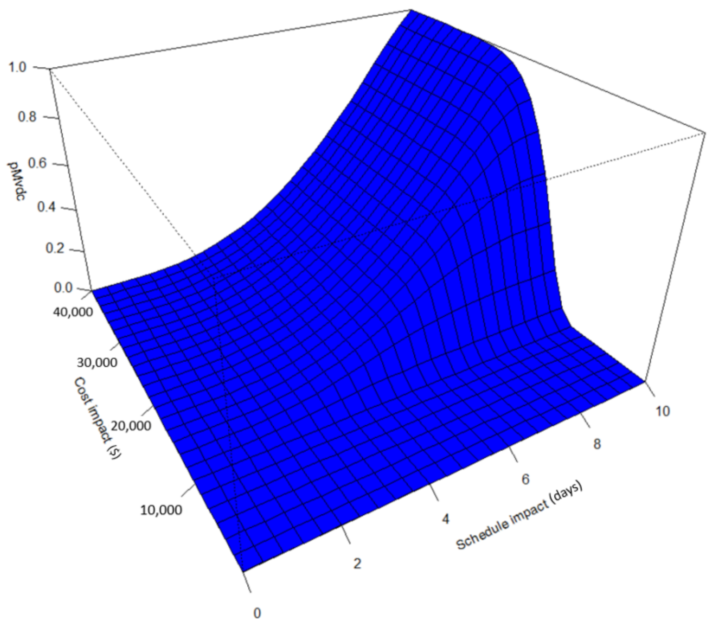


Figure 16. Cumulative density function of the joint distribution.

5. Application and Practical Benefits

Output of the model is presented in Figure 14. The process is repeated until a probability density function of the joint distribution is derived for each risk factor. These distributions can then be input into any existing MCS-based decision-support systems. Notably, for risk factors that have only schedule or risk impacts, marginal distributions will only be calculated for the cost-risk impact or schedule-risk impact (Figure 12a,b). MCS-based decision-support systems use the distributions to sample a cost-risk impact, schedule-risk impact, or joint cost-schedule risk impact value for each risk factor based on its probability of occurrence. The results of several simulation iterations are then combined to provide a number of project insights, such as (1) the expected project completion date as cumulative distribution, (2) the expected project cost as a cumulative distribution, and (3) time and cost contingencies. Readers are referred to [12,37,47] for more information on the application of MCS-based decision-support systems for risk assessment.

A primary benefit of the proposed method is its ability to incorporate detailed information into risk analysis inputs. Existing methods for deriving probability distributions from expert knowledge (e.g., elicitation of parameters for triangular, Pert, or uniform distributions) can only incorporate minimum, maximum, and most likely risk impact values. Experts are neither able to provide detailed information regarding the consideration of root causes nor perform risk impact decomposition, limiting the specificity and, in turn, accuracy, of model results. In contrast, the proposed method allows experts to consider the root causes of a risk factor while providing impact values within an impact range through value mapping. Indeed, results of the sensitivity analysis, which demonstrated that the model was particularly sensitive to changes in mapping values (Figure 11), indicate the importance of accurately capturing subjective information. Methods that allow for the decomposition of input information, such as mapping values to smaller ranges, can allow experts to more precisely express their subjective knowledge, thereby enhancing the comprehensiveness of model results.

A second benefit of the fuzzy logic approach presented here is the transformation of qualitative statements into a quantitative-like format (i.e., probability distribution function), thereby supporting the input of subjective data into quantitative methods, such as MCS. In turn, the proposed approach allows the use of one reliability-based risk assessment method throughout the life-cycle of a project. This alleviates the need for separate qualitative and quantitative methods in different phases of the project, in turn enhancing the consistency of output results.

6. Discussion

Difficulty selecting an input distribution that comprehensively represents risk impact [67] has limited the use of probabilistic simulation-based risk assessment in practice—particularly in the planning and construction phase of wind farm projects that are characterized by a lack of historical data. While various methods and recommendations to overcome this challenge have been suggested in literature [67], existing methods are not capable of incorporating a large amount of subjective knowledge in a time-sensitive, practical manner. Indeed, a flexible method that allows the expert to represent their subjective knowledge as a probability distribution has not been described in literature. The proposed approach addresses this limitation by allowing a risk analyst to (1) reliably assess the risk impact based on subjective knowledge and expertise, (2) consider the root causes of a risk factor when calculating its impact, (3) model the dependence between the cost and schedule impacts of a risk factor that has both cost and schedule impact, (4) reduce biases in expert evaluation through the decomposition of a risk factor, and (5) overcome the limitation for using MCS in practice (i.e., the need for historical data) [68].

This research study proposed a fuzzy-based multivariate analysis approach to address limitations regarding data availability for construction risk assessment of onshore wind projects. The proposed approach was used to successfully solve an illustrative case study of one risk factor common to wind projects. Benefits of proposed method, which included the ability to incorporate detailed subjective knowledge of an expert through the consideration of additional risk factor details, were demonstrated. Furthermore, mapping values provided by experts within the impact range were found to have a considerable impact on the density function of resulting distribution (Figure 11). This was in contrast to minimal impact observed following modifications to the fuzzy membership function of adverse consequence (C) and frequency of occurrence (F) for evaluating root causes (Figures 9 and 10). Although the proposed approach shares similarities with previously developed methods (i.e., determination of minimum and maximum impact values), here, the probability density function of the resulting distribution is enhanced by two notable contributions, namely (1) calculating the most probable value from the combination of the root causes and their assessment and (2) mapping values to a risk impact range. Notably, this method can be applied to any type of project characterized by access to limited data and is, therefore, not limited to wind farm construction.

While the proposed method showed an improvement over previously developed approaches, the findings of this study should be interpreted in consideration of the following limitations. First, the membership function used in this study is a linear triangular and trapezoidal membership function. Other non-linear shapes for membership functions were not investigated and may affect the results. Nevertheless, most membership functions can be accurately represented by either triangular or trapezoidal functions [69]. Second, the risk impact range was divided into three subsets; the examination of more subsets may help experts to better express their belief and confidence about a risk impact. Third, the multivariate approach proposed here will require additional data pooling efforts to ensure that data for all impact dimensions are collected. Models must be built on information that is available and may need to be adjusted in instances where data are lacking, incomplete, or in an inappropriate format for analysis. In this research, a normal copula was used based on the recommendations of previous studies. Future work will include testing multiple copulas, comparing their performance, and developing a complete MCS model for risk assessment considering the developed probability distributions by the proposed method.

7. Conclusions

Input modeling is the first step in MCS-based risk assessment of construction projects. Typically derived using historical project data, the use of MCS-based risk assessment in newer projects has been limited. Because of this, the subjective knowledge of risk analysts at a detailed level has not been properly considered when developing probability distributions for Monte Carlo risk assessment in construction. Therefore, this research tried to address the limitation of input modeling for risk assessment when historical data are lacking with only detailed subjective knowledge available. In particular, this research tried to devise a method that can make used of the experts' subjective knowledge with minimized biases when deriving a probability distribution for a risk factor impact. This research has developed a method capable of capturing and modeling expert subjective knowledge by deriving a flexible generalized Beta distribution through fuzzy logic. Distribution for risk factor impacts with either one type of impact (i.e., cost or time) or both (i.e., cost and time) can be developed with the proposed method. Risk factors that have both cost and schedule impact can be modeled using bivariate distributions by considering the correlation between cost and schedule impact of a risk factor, with the fitted generalized Beta distribution representing the marginal distributions for schedule and cost-risk impact. The method was applied to assess one risk factors common to wind farm construction. Sensitivity analysis was performed, and the model was found to be sensitive to changes in the mapping values of a risk factor impact range. In contrast, changing the shape of the membership function did not affect the resulting distribution. We found that the decomposition of a risk factor into its root causes and decomposition of the risk impact into smaller ranges allows the experts to depict their subjective knowledge more accurately, comprehensively, and with lower biases. An implication of these findings is that both expert subjective knowledge and risk impact decomposition should be taken into account when deriving input distributions for MCS risk assessment. This will enable the application of quantitative methods in the early stages of a project, thereby improving decision-making processes. The proposed approach enables the effective, representative, and comprehensive elicitation of the probability distribution for input modeling in MCS-based risk assessment when only a detailed level of subjective knowledge exists.

Supplementary Materials: The following are available online at http://www.mdpi.com/1999-4893/13/12/325/s1, Supplement SA: Detailed methodology for deriving a marginal Beta distribution from expert opinion of risk impact on cost and schedule, Supplement SB: Derivation of the marginal distribution using *SimphonyProject.NET.*

Author Contributions: Conceptualization, E.M. and S.A.; methodology, E.M. and S.A.; validation, E.M. and S.A.; formal analysis, E.M.; investigation, E.M. and P.J.; writing—original draft preparation, E.M.; writing—review and editing, E.M., P.J., and S.A.; supervision, S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This study is supported by Future Energy Systems research as part of the Canada First Research Excellent Fund (CFREF FES-T11-P01).

Acknowledgments: The authors would like to acknowledge Catherine Pretzlaw for her assistance with manuscript editing and composition.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. IRENA. *Future of Wind: Deployment, Investment, Technology, Grid Integration and Socio-Economic Aspects*; International Renewable Energy Agency: Abu Dhabi, UAE, 2019; ISBN 978-92-9260-155-3.
2. Renewable Electricity Program. Available online: https://www.alberta.ca/renewable-electricity-program.aspx (accessed on 19 June 2020).
3. Wind Energy in Alberta. Available online: https://canwea.ca/wind-energy/alberta/ (accessed on 22 October 2020).
4. Somi, S.; Gerami Seresht, N.; Fayek, A.R. Framework for risk identification of renewable energy projects using fuzzy case-based reasoning. *Sustainability* **2020**, *12*, 5231. [CrossRef]
5. Leimeister, M.; Kolios, A. A review of reliability-based methods for risk analysis and their application in the offshore wind industry. *Renew. Sustain. Energy Rev.* **2018**, *91*, 1065–1076. [CrossRef]
6. Kwak, Y.H.; Ingall, L. Exploring Monte Carlo simulation applications for project management. *Risk Manag.* **2007**, *9*, 44–57. [CrossRef]
7. Kwak, Y.; Ingall, L. Exploring Monte Carlo simulation applications for project management. *IEEE Eng. Manag. Rev.* **2009**, *37*, 83. [CrossRef]
8. Biller, B.; Gunes, C. Introduction to simulation input modeling. In Proceedings of the 2010 Winter Simulation Conference, Baltimore, MD, USA, 5–8 December 2010; pp. 49–58.
9. AbouRizk, H. Understanding and Improving Input for Quantitative Risk Analysis in the Construction Industry. Master's Thesis, University of Alberta, Edmonton, AL, Canada, 2013.
10. Smith, N.J.; Merna, T.; Jobling, P. *Managing Risk in Construction Projects*, 2nd ed.; Wiley-Blackwell: Hoboken, NJ, USA, 2009; ISBN 978-1-405-17274-5.
11. Sadeghi, N.; Fayek, A.R.; Pedrycz, W. Fuzzy Monte Carlo simulation and risk assessment in construction. *Comput. Aided Civ. Infrastruct. Eng.* **2010**, *25*, 238–252. [CrossRef]
12. Hulett, D.; Whitehead, W.; Arrow, J.; Banks, W.; Brady, D. *57R-09: Integrated Cost and Schedule Risk Analysis using Risk Drivers and Monte Carlo Simulation of a CPM Model*; AACE International: Fairmont, WV, USA, 2019.
13. Yan, J. Enjoy the joy of copulas: With a package copula. *J. Stat. Softw.* **2007**, *21*, 1–21. [CrossRef]
14. Zavadska, E.K.; Turskis, Z.; Tamošaitiene, J. Risk assessment of construction projects. *J. Civ. Eng. Manag.* **2010**, *16*, 33–46. [CrossRef]
15. Banaitiene, N.; Banaitis, A. Risk Management in Construction Projects. In *Risk Management—Current Issues and Challenges*; Banaitiene, N., Ed.; InTechOpen: London, UK, 2012; ISBN 978-953-51-0747-7.
16. Senesi, C.; Javernick-Will, A.; Molenaar, K. *RR280-11: Applying Probabilistic Risk Management in Design and Construction Projects*; Construction Industry Institute: Austin, TX, USA, 2013.
17. Duracz, A. Derivation of Probability Distributions for Risk Assessment. Master's Thesis, Stockholm University, Stockholm, Sweden, 2006.
18. Yoe, C. *Principles of Risk Analysis: Decision Making Under Uncertainty*; CRC Press: Boca Raton, FL, USA, 2016; ISBN 978-1-4398-5750-2.
19. Law, A.M. *Simulation Modeling and Analysis*, 5th ed.; McGraw-Hill Education: New York, NY, USA, 2014; ISBN 978-0-07-340132-4.
20. O'Hagan, A. Expert knowledge elicitation: Subjective but scientific. *Am. Stat.* **2019**, *73*, 69–81. [CrossRef]
21. Morris, D.E.; Oakley, J.E.; Crowe, J.A. A web-based tool for eliciting probability distributions from experts. *Environ. Model. Softw.* **2014**, *52*, 1–4. [CrossRef]

22. Cooke, R.M.; Goossens, L.H.J. Expert judgement elicitation for risk assessments of critical infrastructures. *J. Risk Res.* **2004**, *7*, 643–656. [CrossRef]
23. Galway, L. *WR-112-RC: Quantitative Risk Analysis for Project Management: A Critical Review*; Rand Corporation: Santa Monica, CA, USA, 2004.
24. Nasir, D.; McCabe, B.; Hartono, L. Evaluating Risk in Construction–Schedule Model (ERIC–S): Construction schedule risk model. *J. Constr. Eng. Manag.* **2003**, *129*, 518–527. [CrossRef]
25. van Dorp, J.R.; Duffey, M.R. Statistical dependence in risk analysis for project networks using Monte Carlo methods. *Int. J. Prod. Econ.* **1999**, *58*, 17–29. [CrossRef]
26. Galway, L.A. *Subjective Probability Distribution Elicitation in Cost Risk Analysis: A Review*; Rand Corporation: Santa Monica, CA, USA, 2007.
27. Meyer, M.; Grisar, C.; Kuhnert, F. The impact of biases on simulation-based risk aggregation: Modeling cognitive influences on risk assessment. *J. Manag. Control* **2011**, *22*, 79. [CrossRef]
28. Dikmen, I.; Budayan, C.; Talat Birgonul, M.; Hayat, E. Effects of risk attitude and controllability assumption on risk ratings: Observational study on international construction project risk assessment. *J. Manag. Eng.* **2018**, *34*, 04018037. [CrossRef]
29. Tversky, A.; Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Util. Probab. Hum. Decis. Mak.* **1975**, *11*, 141–162. [CrossRef]
30. Montibeller, G.; Winterfeldt, D. von Cognitive and motivational biases in decision and risk analysis. *Risk Anal.* **2015**, *35*, 1230–1251. [CrossRef]
31. Li, N.; Fang, D.; Sun, Y. Cognitive psychological approach for risk assessment in construction projects. *J. Manag. Eng.* **2016**, *32*, 04015037. [CrossRef]
32. Touran, A.; Wiser, E.P. Monte Carlo technique with correlated random variables. *J. Constr. Eng. Manag.* **1992**, *118*, 258–272. [CrossRef]
33. Clemen, R.T.; Reilly, T. Correlations and copulas for decision and risk analysis. *Manag. Sci.* **1999**, *45*, 208–224. [CrossRef]
34. Touran, A. Probabilistic cost estimating with subjective correlations. *J. Constr. Eng. Manag.* **1993**, *119*, 58–71. [CrossRef]
35. Ökmen, Ö.; Öztaş, A. Construction project network evaluation with correlated schedule risk analysis model. *J. Constr. Eng. Manag.* **2008**, *134*, 49–63. [CrossRef]
36. Moret, Y.; Einstein, H.H. Modeling correlations in rail line construction. *J. Constr. Eng. Manag.* **2012**, *138*, 1075–1084. [CrossRef]
37. Moret, Y.; Einstein, H.H. Construction cost and duration uncertainty model: Application to high-speed rail line project. *J. Constr. Eng. Manag.* **2016**, *142*, 05016010. [CrossRef]
38. Mawlana, M.; Hammad, A. Joint probability for evaluating the schedule and cost of stochastic simulation models. *Adv. Eng. Inform.* **2015**, *29*, 380–395. [CrossRef]
39. Analytic Method for Probabilistic Cost and Schedule Risk Analysis. Available online: https://www.nasa.gov/pdf/741989main_Analytic%20Method%20for%20Risk%20Analysis%20-649%20Final%20Report.pdf (accessed on 27 October 2020).
40. Palisade @RISK: Risk Analysis using Monte Carlo Simulation in Excel and Project. Available online: http://www.palisade.com/risk/default.asp (accessed on 22 October 2020).
41. Embrechts, P.; Lindskog, F.; Mcneil, A. Modelling dependence with copulas and applications to risk management. In *Handbook of Heavy Tailed Distributions in Finance*; Svetlozar, T.R., Ed.; Elsevier: Amsterdam, The Netherlands, 2003; pp. 329–384. ISBN 978-0-444-50896-6.
42. Rajgor, G. Building wind farms: Part five: The precarious construction phase needs careful preparation. *Renew. Energy Focus* **2011**, *12*, 28–32. [CrossRef]
43. Turner, G.; Roots, S.; Wiltshire, M.; Trueb, J.; Brown, S.; Benz, G.; Hegelbach, M. *Profiling the Risks in Solar and Wind: A Case for New Risk Management Approaches in the Renewable Energy Sector*; Swiss Reinsurance: Zurich, Switzerland, 2013.
44. Finlay-Jones, R. Putting the spin on wind energy: Risk management issues in the development of wind energy projects in Australia. *Aust. J. Multi-Discip. Eng.* **2007**, *5*, 61–68. [CrossRef]
45. Kucukali, S. Risk scorecard concept in wind energy projects: An integrated approach. *Renew. Sustain. Energy Rev.* **2016**, *56*, 975–987. [CrossRef]
46. Rolik, Y. Risk Management in implementing wind energy project. *Procedia Eng.* **2017**, *178*, 278–288. [CrossRef]

47. Mohamed, E.; Gerami Seresht, N.; Hague, S.; AbouRizk, S. Simulation-based approach for risk assessment in onshore wind farm construction projects. In Proceedings of the 9th Asia-Pacific International Symposium on Advanced Reliability and Maintenance Modelling, Vancouver, BC, Canada, 20–23 August 2020.
48. Fayek, A.R. Fuzzy logic and fuzzy hybrid techniques for construction engineering and management. *J. Constr. Eng. Manag.* **2020**, *146*, 04020064. [CrossRef]
49. Smith, G.R.; Hancher, D.E. Estimating precipitation impacts for scheduling. *J. Constr. Eng. Manag.* **1989**, *115*, 552–566. [CrossRef]
50. Budayan, C.; Dikmen, I.; Talat Birgonul, M.; Ghaziani, A. A computerized method for delay risk assessment based on fuzzy set theory using MS Project^{TM}. *KSCE J. Civ. Eng.* **2018**, *22*, 2714–2725. [CrossRef]
51. Andersen, B.; Fagerhaug, T. *Root Cause Analysis: Simplified Tools and Techniques*; ASQ Quality Press: Milwaukee, WI, USA, 2006; ISBN 978-0-87389-692-4.
52. Ayyub, B.M. *Risk Analysis in Engineering and Economics*, 2nd ed.; Chapman and Hall/CRC Press: Boca Raton, FL, USA, 2014; ISBN 9781466518254.
53. Abdelgawad, M.; Fayek, A.R. Risk management in the construction industry using combined Fuzzy FMEA and Fuzzy AHP. *J. Constr. Eng. Manag.* **2010**, *136*, 1028–1036. [CrossRef]
54. Johnson, D. Triangular approximations for continuous random variables in risk analysis. *J. Oper. Res. Soc.* **2002**, *53*, 457–467. [CrossRef]
55. AbouRizk, S.M.; Halpin, D.W.; Wilson, J.R. Fitting beta distributions based on sample data. *J. Constr. Eng. Manag.* **1994**, *120*, 288–305. [CrossRef]
56. AbouRizk, S.M.; Halpin, D.W.; Wilson, J.R. Visual interactive fitting of beta distributions. *J. Constr. Eng. Manag.* **1991**, *117*, 589–605. [CrossRef]
57. Fayek, A.R.; Lourenzutti, R. Introduction to fuzzy logic in construction engineering and management. In *Fuzzy Hybrid Computing in Construction Engineering and Management*; Emerald Publishing Limited: Melbourne, Australia, 2018; pp. 3–35. ISBN 978-1-78743-869-9.
58. AbouRizk, S.M.; Sawhney, A. Subjective and interactive duration estimation. *Can. J. Civ. Eng.* **1993**, *20*, 457–470. [CrossRef]
59. Corona-Suárez, G.A.; AbouRizk, S.M.; Karapetrovic, S. Simulation-based fuzzy logic approach to assessing the effect of project quality management on construction performance. *J. Qual. Reliab. Eng.* **2014**, *2014*, 1–18. [CrossRef]
60. Ayyub, B.M.; Haldar, A. Project scheduling using fuzzy set concepts. *J. Constr. Eng. Manag.* **1984**, *110*, 189–204. [CrossRef]
61. Oliveros, A.V.O.; Fayek, A.R. Fuzzy logic approach for activity delay analysis and schedule updating. *J. Constr. Eng. Manag.* **2005**, *131*, 42–51. [CrossRef]
62. Diógenes, J.R.F.; Claro, J.; Rodrigues, J.C. Barriers to onshore wind farm implementation in Brazil. *Energy Policy* **2019**, *128*, 253–266. [CrossRef]
63. Lucko, G.; Rojas, E.M. Research validation: Challenges and opportunities in the construction domain. *J. Constr. Eng. Manag.* **2010**, *136*, 127–135. [CrossRef]
64. Saltelli, A.; Tarantola, S.; Campolongo, F.; Ratto, M. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*; John Wiley & Sons: Hoboken, NJ, USA, 2004; ISBN 978-0-470-87093-8.
65. Ashraf, W.; Glinicki, M.A.; Olek, J. Statistical analysis and probabilistic design approach for freeze–thaw performance of ordinary Portland Cement concrete. *J. Mater. Civ. Eng.* **2018**, *30*, 04018294. [CrossRef]
66. Wolodzko, T. Additional Univariate and Multivariate Distributions, R Package. Available online: https://cran.r-project.org/web/packages/extraDistr/index.html (accessed on 27 October 2020).
67. Ospina, L.G.G.; Rueda, A.J.B.; Bessolo, A.F.M.; Tienda, J.L.P. A Methodological proposal for risk analysis in the construction of tunnels. In *Advances in Informatics and Computing in Civil and Construction Engineering*; Mutis, I., Hartmann, T., Eds.; Springer International Publishing: Basel, Switzerland, 2019; pp. 815–822. ISBN 978-3-030-00219-0.
68. Salah, A.; Moselhi, O. Contingency modelling for construction projects using fuzzy-set theory. *Eng. Constr. Archit. Manag.* **2015**, *22*, 214–241. [CrossRef]

69. Fayek, A.R.; Oduba, A. Predicting industrial construction labor productivity using fuzzy expert systems. *J. Constr. Eng. Manag.* **2005**, *131*, 938–941. [CrossRef]