CHAPTER II

LITERATURE REVIEW

Time estimating is crucial in construction planning and scheduling. Traditional methods of duration estimating are based on a deterministic approach and usually ignore the inherent variability of the real world. A probabilistic methodology is better suited (Cor, 1998). The development of a single point estimating together with the evaluation of the contingencies resulting from risk (e.g., operation risk and project risk) is necessary because the limitations of a deterministic approach are overcome.

A number of probabilistic approaches have been developed. One of the most effective probabilistic approaches is simulation. In addition, several methods are proposed to estimate parameters of a probability density function including moment matching, maximum likelihood estimation and least-squares fit of the cumulative distribution function. These are numerical methods which can be used to compute the parameters of the underlying probability density function, if sample observations are available. In an absence of data, subjective information is used to estimate the parameters of the probability density function expected to describe the construction operations. The direct fractile assessments are considered as the most reliable method in eliciting an accurate subjective estimation of a statistical characteristic such as mode, mean, variance or selected percentiles. The duration multipliers, which are subjectively estimated, are used find these statistical characteristics.

Although this method makes assessors become less prone to bias, the multiplier developing process does not explicitly evaluate effects of factors such as weather, labor skills, site conditions, and management quality which would be risk factors when they probably extend duration of project activities and project completion times. A better understanding of effects of risk factors on duration would make the schedule a more useful management tool by providing a better estimate of uncertainty in duration of project activities and helping to focus attention on significant risk factors that have the great impact on project completion times

Currently, models proposed to analyze impacts of risk factors on duration of project activities and a project duration are rarely found in the construction industry. In traditional construction scheduling methods, the duration of each activity is estimated without explicit and systematic evaluation and integration impacts of risk factors on the duration. This research is to propose a methodology for quantifying impacts of risk factors influencing duration of project activities and project completion times and providing information to support schedule risk management. The following sections are to review previous researches on construction duration estimating methods and risk analysis.

2.1 Deterministic Approach for Duration Estimating

Impacts of risk can be presented in terms of delay, cost overrun, and poor quality of work. Bordoli and Baldwin (1998) classified delays into six delay mechanisms and proposed the delay quantification method to calculate delay duration and update the operation duration and the project duration based on the basic techniques in scheduling, i.e. bar charts. Shi et al. (2001) also applied risk evaluation to the bar charts. The proposed method is visually effective in communicating a schedule. However, this effectiveness diminishes, as projects become complex.

Other delay analysis methods have been developed, for example as-planned method, as-built method, modified as-built method, modified as-planned, daily window delay analysis (Bubshait and Cunningham, 1998; Hegazy and Zhaug, 1995 and 2006). Onofrio and Stetson (1990) summarized several methods used to analyze delays including a global impact method, net impact method, adjusted as-planned CPM approach, adjusted as-built approach, collapse as-built schedule approach, impacted updated CPM approach, modification impact analysis approach, and time impact updated CPM approach. The difference between these approaches is the way delays are incorporated within the analysis and the types of the schedule affected by the delays. For instance, the global impact method determines total delay as a sum of delay durations displayed on an as-built bar chart schedule. The sum of the delay durations plus concurrence of delays is the output of the net impact approach. The as-planned CPM approach incorporates the delays into the as-planned schedule, but no factual data are included. The adjusted as-built method illustrates all delays in the as-built schedule, but it fails to meet the dynamic nature of CPM because it is applied only once to perform the computation of the critical path. However, these models do not consider impact of risk factors on duration.

Pipattanapiwong (2004) mentioned that the quantification of impacts of risk on activity duration is difficult because the dependencies among risk factors and also relationships between risk factors and activities, and correlation between durations of activities cannot be clearly defined and fully measurable. Generally, activity duration is directly and independently estimated without formally accounting for risk and the dependencies between risk and activities. For example, the critical path method (CPM) approach is deterministic in nature; a process thereby ignores the effects of uncertainty by using a single value for the estimated duration of each activity. Several researches have integrated the dependency identification into the basic techniques in scheduling (i.e. the work breakdown structure (WBS) and CPM scheduling technique).

Mulholland and Christian (1999) classified problems that have been documented with traditional scheduling processes shown in Table 2.1. The problems within each classification are divided into the following three sources of constraints, which can limit effective use of traditional scheduling processes including (1) technical problems involved with network concepts, (2) process problems in the application of network methodology, and (3) problems involved with making the process work in construction. One of the more serious problems with traditional scheduling processes is that they do not evaluate explicitly the uncertainty and risky nature of the internal and external environment of a project and an operation so that a realistic project performance time cannot be adequately estimated.

Kim (2003) and Cor (1998) conducted the research related to the prediction of duration. They also produced improvements in the scheduling technique (e.g., CPM) by considering the dynamic nature of CPM. It is found out that CPM cannot be used to demonstrate delay mechanisms and suspension of works due to the failure to link the schedule up with the information in the job record. In addition, CPM is not so effective in the demonstration of productivity-related impacts because it is not capable of addressing the productivity-related issues.

Table 2.1 Classification of problems in traditional project scheduling processes (Mulholland and Christian, 1999)

Classification of problems	Sources of problems		
	Network theory	Application of network methodology	Realities of construction process
(i) Nature of engineering and construction work			
Complex organization		x	
Activity dependencies	x		
Other techniques more appropriate		x	
Critical activities cost separation		x	
(ii) Resource allocation process			
Networks inadequacy considering resources		x	
Equal importance to each activity in resource allocation		x	
Effect on resource program for each activity with single value time estimate	x		
Resource allocation algorithms use of priority rules	x		
Limited resource types in resource allocation program	x		
Resource allocation processes problem with project scope details			x
(iii) Complexities of project operations and scheduling methods			
Schedules too complex for field			x
Choice of logic in network models	x		
Specialized scheduling operation team required			x
Interpretation of excessive computer output		x	
(iv) Contractors lack of confidence in network scheduling processes			
Creditability that construction process can be modeled			x
Many secondary business activities occur but not shown on schedule		x	

Table 2.1 Classification of problems in traditional project scheduling processes (Mulholland and Christian, 1999) (con't)

Classification of problems	Sources of problems		
	Network theory	Application of network methodology	Realities of construction process
(v) Activity time estimating			
Lack of reliable and realistic time estimates for activities			x
Nature of activity duration requires experience			x
Time estimates only marginally address schedule risk		x	

Several techniques have been developed to quantify productivity-related impacts, such as time and motion study, expert opinion, and historical productivity data. The time and motion study is used to measure the performance difference between impacted and non-impacted operations (Cushman and Carpenter, 1990). However, the reliability depends mainly on a number of observations. Although expert opinion is used to prove and quantify delay, the answer relies on the factual basis, the expert's personal creditability, the analysis method which is selected by experts. This method is usually applied to situations where limited data is available. Historical productivity data are compared to productivity data of works being carried out on an existing project. However, this method gives the overall impact rather than the segregated impacts caused by particular risk. Generally, these tools are not as effective as the tools for delay analysis. These limitations definitely justify a search for more powerful tools for quantifying the productivity-related impacts and delays resulting from risk.

2.2 Definition of Risk and Uncertainties

Risk has different meanings to different people. The concept of risk varies according to viewpoint, attitudes and experience. Engineers, designers, and contractors consider risk from the technological perspective; while lenders and developers view risk from the economic and financial side. Therefore, risk is generally considered as abstract concept whose measurement is very difficult.

A large or complex construction project is fraught with the uncertainties. Risk is exposure to consequences of uncertainty. In the other words, uncertainty may be considered as a range of events that may happen and produce risks affecting the project. Uncertainty can directly or indirectly lead to increased project duration. The project duration is uncertain until the project has been completed. Since one of the objectives of construction projects is usually stated as a target established for time, the most important risk in construction are the failure to meet this target (Ahuja and Nandakumar, 1985).

Under conditions of uncertainty, the construction project planning involves risks. A planned schedule is really an educated guess of the activities that a project will require, how these activities interrelate, and how long each activity will take. The uncertainties associated with these activities are the effects of interactions between activities and the duration of the activities. For example, one delayed activity may delay several activities or cause their duration to take longer that expected.

In a project context, risk is defined as the chance of something happening that will have an impact upon objectives. It includes the possibility of loss or gain, or variation from a desired or planned outcome, as a consequence of the uncertainty associated with following a particular course of action. Risk may be defined as the likelihood of a detrimental event occurring to the project and associated consequences if it does Chapman (1997).

Many different classifications of risk have been developed. However, most of these have considered the source criteria as the most importance. Construction project risks are broadly classified into technical, construction, legal, natural, logistic, social, economic, financial, commercial, and political. Instead of the source criteria, risks are classified based on the location of the impact of risks in the elements of the project. Typically, risks are also categorized into dynamic/static, corporate/individual, internal/external, positive/negative, acceptable/unacceptable and insurable/non-insurable.

The time goal of project cannot be achieved because of inherent risk pertaining to it. An effective project management is possible through systematic risk management along with the other tools of managing project. Laufer and Tucker (1987) suggested that for planning to become effective, methods should be changed (e.g., gathering and diffusing of information), policies should be modified (e.g., the role of planning and control), assumptions should be adjusted (e.g., attitude to uncertainty) and the overall philosophy of project management should be reexamined. Barker (1986) identified factors like size, complexity, advance technology and increasing involvement of external risk factors combined to increase the uncertainty inherent in today's projects.

By nature of construction projects, the available data are often incomplete or insufficient and invariably contain variability. The risk assessment and project scheduling must rely on predictions or estimations based on idealized models with unknown degrees of imperfections relative to reality, and thus involve additional uncertainty. It is important to recognize the presence of all major sources of uncertainty in the risk assessment and project scheduling. The sources of uncertainty may be broadly classified into two types: (i) sources that are related to natural randomness of the underlying phenomenon that is exhibited as variability in the observed information, and (ii) sources that are associated with inaccuracies in the assessment of risk and the prediction and estimation of project duration because of insufficient or imperfect knowledge. The term aleatory uncertainty is used to describe the former, while the term epistemic uncertainty is used to describe the latter. Typically, the two types of uncertainty are treated separately. However, they should be combined and analyzed as a total uncertainty.

To determine every uncertainty, the sources of the two types of uncertainty, and their effects on the risk assessment and project scheduling should be determined. The aleatory (database) uncertainty is associated with the inherent variability of basic information pertaining to the risk assessment and project scheduling that can be observed and described. Sources of the aleatory uncertainty can commonly be singled out from other contributions to uncertainty by their representation as randomly distributed quantities that can take on values in an established or known range, but for which the exact value will vary by chance from time to time. A probability distribution is a common mathematical representation of the aleatory uncertainty. The aleatory uncertainty is part of construction projects and, therefore, may not be reduced or modified. When substantial experimental data are available for estimating a distribution, there is no debate that the correct model for aleatory uncertainty is a probability distribution which is propagated through a modeling and simulation process.

The epistemic (or knowledge–based) uncertainty, on the other hand, is associated with imperfect knowledge of a construction project, and may be reduced through an increase in knowledge or information and application of better risk assessment method and project scheduling model. Examples of sources of the epistemic uncertainty are: unavailability and insufficiency of data for fixed (but unknown) parameter, limited understanding of complex construction projects, and the occurrence of fault sequences or conditions not identified for inclusion in the risk assessment and project scheduling.

Effects of these two types of uncertainty are different. The effect of the aleatory randomness leads to a calculated probability or risk, whereas the effect of the epistemic type expresses an uncertainty in the estimated probability or risk. The uncertainty (or error bounds) of a calculated probability or risk is as important as the risk itself. The National Research Council (1994) has emphasized the importance of quantifying the uncertainty in the calculated risk. Theoretically, the epistemic uncertainty includes inaccuracies in the prescribed form of probability distributions and in all parameters. However, for practical purposes, the epistemic uncertainty may be limited to the estimation of the mean or median values. The mathematical representation of epistemic uncertainty has proven to be much more of a challenge. The preeminent issue in uncertainty analysis is the representation, aggregation, and propagation of epistemic uncertainty as well as combination of epistemic and aleatory uncertainty. The quantification of the two types of uncertainty should include the appropriate concepts and methods.

2.3 Probabilistic Approach for Duration Estimating

Duration estimating is crucial in every phase of construction projects from planning to execution. An estimator has to produce an accurate estimate that is able to highlight reality. The estimate is considered accurate if it is sufficiently close to actual duration (AbouRizk and Halpin, 1992). The professional experience and judgment, and a matte of relevant historical data have a great impact on an estimate. The estimator must employ the best information available to estimate duration for providing resources and carrying out works. This information includes crew time data from previous works. Data are associated with the calculations based on detailed analysis of the construction process combining with the estimator's best guess of time. An estimator has to use data from past projects with knowledge of their similarity to the ongoing work to provide an accurate estimate. The adjustment of the historical data by subjective data is required to improve accuracy and reflect the realities of the current project.

The traditional methods of duration estimating are often oriented towards a deterministic approach which cannot deal with the variability of the real world. To overcome the limitations of the deterministic approach, the probabilistic methodology for creating a range estimate is proposed by quantifying contingencies and then adding them into a single point estimate.

When applying a range-estimating methodology, durations of project activities are assumed to be random variables instead of known parameters. Then, a probability distribution of project duration is established. The obtained probability distribution is compared to a single-point estimate provided by a deterministic approach. The resulting probability distribution assists in defining duration of project activities that are able to be associated with prescribed levels of confidence. The manager can use the resulting probability distribution to produce a useful estimate of a project completion time. Most range duration estimates are prepared using probabilistic techniques such as Monte Carlo simulation. In this technique, durations of project activities, each defined as random variables, are sampled according to the derived distribution functions. The project completion time is calculated by systematically summing up durations of project activities. This sampling process is repeated a large number of times to produce sufficient samples, or probable distribution of project duration.

Halpin (1992) stated that simulation suits to construction because it provides experiments in construction operations which can be used to evaluate potential impacts of risk factors or improvements to schedule. This research attempts to provide a better estimate of uncertainty in duration of an activity and a project by using simulation. The following section is to review simulation and applications of simulation to construction.

2.3.1 Simulation

Simulation can be considered as a process of replicating the real world on a set of assumptions and conceived models of reality (AbouRizk and Halpin, 1992). An operation is imitated over time or production cycle by the simulation. Generally, simulation is utilized to demonstrate a large complex, non – existent process as it provides simulated experiments which are less expensive than physical experiments provided by a real system. Simulation is done purposely to understand the behaviors of a process because simulation can present behaviors of a system under various alternative scenarios. The changes of any real operation after applying new components or reduced parts of the existing operations can be observed by changing the input variable values and assumptions provided during the development of the simulation model. The simulation is very useful to investigate and understand construction operations in the design stage as it can predict the performance of a system not yet built.

Simulation can be of continuous random variables or discrete events. The quantification of changes in a system over a period of time is provided in the continuous simulation. The modeling of a system as it evolves over time by a representation in which state variables change instantaneously at separate point in time is the definition of the discrete event simulation provided by Law and Kelton (1991). Simulating the processes of an order arriving at a plant or the delivery of material is an example of the discrete event simulation of which the variable values change at each point in the process, i.e., from raw materials to a finished product.

Simulation is also considered as a tool in decision making. System performance can be examined under alternative methods or environments by using the simulation. Then, an optimal design of a system can be achieved. Discrete event simulation can be applied to analysis of construction processes and activities affected by risk factors. Results provided by performing the construction risk assessment can be used to improve realization of simulation results. Risky scenarios, resource constraints, and potential bottlenecks can be observed in the construction simulation conducted based on particular designed construction scenarios. Discrete event simulation can be applied to the statistical assessment of possible outcomes of construction scheduling by using probability distributions to model uncertainties in the basic random variables. Construction sequences, construction times, utilization of facilities, materials, labor and equipment, and transportation can be modeled by a construction simulation.

A simulation model capturing construction processes is developed to make a comparison between resources estimated to be utilized in a project and those actually utilized. It can highlight critical areas and bottlenecks of the relevant construction scenarios. As noted, the simulation evolves and understands behaviors of a system under various alternative scenarios. The changes of a real system or process after applying new components or reducing parts of the existing system can be observed by changing input variable values and assumptions provided during the development of a simulation model. The simulation is particularly useful to investigate and understand the construction CPM schedules in the design phase and construction phase as it is able to predict the construction performance, although the construction is not yet carried out or performed on unforeseen conditions. To estimate the construction duration, the probability distributions of decision variables are used to represent uncertainty of duration estimate. Simulation allows users to update whenever new data are available so as to provide the better estimation performance.

2.3.2 Application of Simulation to Construction

Considerable efforts have been spent to improve on modeling construction processes using computer simulation. This section is devoted to the demonstration of the applications of simulation modeling in the construction industry. Also, combinations between simulation and other approaches together with advantages and limitations are discussed.

Although duration of project activities can be approximately estimated by using equipment manufacture's charts, the estimated duration has to be modified by considering impacts of affecting factors (e.g., working conditions, productivity factors). Zayed and Halpin (2004) stated that it is difficult for the estimators to evaluate pilling productivity as an enormous number of problems relating to, for example, subsurface obstacles, lack of contractor experience, and site planning difficulties. This problem can be also found out upon other types of construction. Therefore, it is necessary to use sophisticated techniques to analyze the problem and determine the closest optimal solution. Several tools for assessing construction process productivity using simulation technique have been proposed (Zayed and Halpin, 2004; Marzouk and Moselhi, 2001; Zhang et al., 2004). However, modeling techniques for construction operations of pile foundations based on simulation (Zayed and Halpin, 2004), earthmoving operations based on simulation (AbouRizk and Halpin, 1992) and regression (Smith, 1999) still lack an ability to provide users with an interpretable duration estimation model and accurate estimated duration.

Construction-oriented discrete-event simulation often faces the problem of defining uncertain information input, such as subjectivity in selecting and establishing probability distributions that results from insufficient or lack of site productivity data. Inappropriate inputs lead to misleading simulation outputs, and therefore error-prone or sub-optimal planning and management decisions. Difficulties in input modeling, including the lack of confidence in input information, have limited the use of discrete-event simulation as a practical tool for construction.

Input modeling for activity duration is generally accomplished through importing observed data or probability distributions derived from sample data. However, it is time-costly to collect large amount of field data, especially for a large operational system that cannot be practically imitated (Law and Kelton, 1991). In addition, definition of probability distributions might be considered to be intractable. It might be theoretically too complicated and computationally too expensive. Several researches have provided definitions of the probability distributions for activity duration through sample data in the construction field (AbouRizk and Halpin, 1992). The considerable factors affecting duration-input modeling consist of gathering data of sufficient quality, quantity, and variety and methods used to establish a probability density function of duration. In the construction industry, it is not possible to collect a sufficient amount of data statistically required for establishing probability distributions for modeling simulation based on the probability theories. This is because of the uniqueness of construction activities, inappropriate data collecting process, and unqualified respondents. The difficulties in the establishment of a probability distributions of inputs in the simulation processes are not only brought about by uncertain data, but also uncertainty factors (e.g., task uncertainty, weather uncertainty, organization uncertainty) stemming from the complexity of a construction project. These uncertainty factors have been discussed in previous section.

2.4 Distributions of Duration

The probabilistic methodology is generally employed to quantify uncertain activity duration due to factors having impact on such activity. It is also used to create the range estimate of duration of project activities. The attributes of affecting factors and activity duration are considered as random variables. The probability distribution is used to represent duration is able to be associated with the prescribed levels of confidence. A probabilistic technique such as Monte Carlo simulation is usually used to prepare the range duration estimate of a project by sampling the activity durations according to their distribution functions and then summing up the simulated activity durations.

The Monte Carlo simulation technique uses a random number generator to simulate construction processes a multiple number of times and then store and record the results for each iteration. A range-estimating methodology can help an estimator to incorporate the inherent variability in activity duration with the project duration and to assign a level of confidence to each selected value.

To utilize a range-estimating methodology, typically historical data acquired through completed projects are used to define the data set for the individual activity duration. As the historical data are unable to exactly represent existing working condition, the historical data have to be adjusted for scope, location, and time differences. After all historical data are adjusted, it is desirable to model the data as a probability distribution to enable a range-estimating approach for predicting further project duration.

The selection of probability distributions and parameter estimating techniques for modeling the available data and developing the construction schedule is an important task of developing a duration-estimating system. Normally, a triangular distribution and a least-squares (LS) technique for estimating parameters of a distribution are chosen for a construction project. However, there are several types of probability density functions including the normal, lognormal, uniform, gamma, and beta distributions. The probability distributions should have the following properties modified from the probability distribution properties mentioned in (AbouRizk and Halpin, 1992):

a. On any estimate, upper and lower limits exist, beyond which a reasonable estimator is relatively certain that no value will occur. Consequently, a closed-ended distribution is desirable.
b. The distribution should be continuous. It is illogical to assume that the probability distribution for project duration is discrete. The construction duration may have any value within reasonably defined limits.

c. It should be assumed that the probability of occurrence of an event decreases as the upper and lower limits are approached. Given this constraint, the probability distributions should have a convex shape rather concave.
d. As a result of the previous consideration, the distributions of construction duration will be unimodal. This is particularly relevant to construction duration, as it is expected that duration will have one most-probable value.
e. Since the actual durations will have a greater freedom to be higher than lower with respect to the estimate, skewness should be expected.

Each of these desired constraints are satisfied by the beta and triangular distributions. As note, the triangular distribution is usually selected for modeling the historical data. For the beta density function, it yields a bimodel, u-shaped distribution Four parameters must be specified in the general form of the beta distributions or, alternatively, the standard form of the beta distribution may be used, which requires only two shape parameters but strictly limits the possible (x) values to a range of zero to one [0,1]. The standard form of the beta distribution may be accordingly established in a simply manner. However, data samples have to be rescaled and shifted within the model to accommodate the full range of values required to create probabilistic duration estimates at various hierarchical levels of a project. This requires an estimation of the minimum and maximum values and also the two shape parameters α and β for the distribution instead of the minimum, maximum, and most likely values required for the triangular distribution. These make the beta family of distributions less desirable to be used when it is compared to the triangular distribution. The triangular distribution is usually selected for the estimating system because it has more intuitive understanding.

Weiler (1965) stated that many of the errors in an output of a simulation model can be traced to assigning wrong values of the parameters of a distribution. However, if the form of the probability density function is unknown, a further error may be introduced by assuming one distribution function when actually some other distribution would have been appropriate. Nevertheless, this type of error is likely to be small compared with errors in assigning the wrong values to the distribution parameters. AbouRizk and Sawhney (1993) mentioned that the choices of a distribution function do not affect the output results of the simulation model, if they are derived from the same original data.

Although the triangular distribution is suited for estimating construction duration, little researches have studied on how to actually model uncertainties involved in duration of project activities affected by risk factors. The integration of the risk assessment results into the establishment of a triangular distribution of activity duration is also rarely determined. Most duration-estimating methods using triangular distributions presume that estimators can easily approximate the minimum, maximum, and most likely duration. Automatic generation of probability-density functions assists users in quickly performing reliable probabilistic estimates for numerous combinations of time and location project scenarios. Therefore, it is very important to develop a methodology for modeling data as a triangular distribution with an aid of expert opinion or other subjective evaluations. The challenge is to automate a procedure that would accurately define the triangular distribution when only subjective data are available for an analysis. Subjective data representing possibility and consequence of risk factors and their impacts may be relatively few or many depending on experts' judgment.

2.5 Risk Analysis

Risk and uncertainty occur not only on large capital projects: while size can be one of the major causes of risk, other factors include complexity, speed of construction, location of the project and degree of unfamiliarity to the client. Perry (1986) suggested that a significant improvement to project management performance may result from greater attention to the whole process of risk management. Several researches and literature related to a risk analysis have been presented. There are four types of risk analysis: sensitivity analysis, probability analysis, decision three analysis, and simulation (Perry, 1986). Frank (1987) mentioned that the distinction between quantifiable risk forms the basis for risk analysis and risk assessment during the project handling stage. The quantifiable risks can be easily assessed by using the change management. Uncertainty with regard to the project duration is essentially influenced by the effects of the qualitative risks. Hull (1988) segregated the risk analysis into two phases: qualitative phases and quantitative phases in sequence. The probability approach and simulation are used to perform the risk analysis. Probability models suffer from two major limitations: (i) non–availability of detailed quantitative information in real project world, and (ii) imprecise and ill defined decision problem of project participating agencies where utility and perception of decision maker plays and important role. However, definitions of the risk analysis provided in project management, engineering, construction, and other industries are not consistent. This section focuses on the risk analysis defined for a project management aspect. An application of the risk analysis to the evaluation of impacts of risk on duration is particularly concerned.

Several definitions of risk and attributes of a risk factor in construction have been discussed (Al-Bahar and Chandall,1990; Chapman, 1997; Smith, 1990). However, risk might be an ambiguous term, if the definition of risk is not well provided. There are five attributes related to risk suggested by AbouRizk and Sawhney (1993) including likelihood, outcome, significance, causal scenario, and population. The modified definition of risk provided by AbouRizk and Sawhney (1993) is used in this research as it is the generally accepted expression for risk.

In the context of project management, risk was defined as "how likely the event is to occur, i.e. the chance of certain occurrences adversely affecting project objectives. Commonly, risk is characterized by: 1) event, i.e. precisely what might occur in a project, 2) degree of probability, i.e. how likely the event is to occur, and 3) consequence, i.e. a total loss or losses which could result.

2.5.1 Likelihood and Significant Consequence

The assessment and evaluation of uncertainties related to an event is the concept of risk. Typically, a pair of likelihood (probability of occurrence) of an event and the outcomes (consequences) related to the occurrence of the event is used to measure risk. However, the probability of occurrence is not pairing off with the corresponding consequence by a mathematical operation, a scalar or vector quantity. The probability of occurrence is just matched with the corresponding consequence. The equation often used to represent the relationship between the probability of occurrence and consequence is Eq. (2.1).

$$\text{Risk} = [(Ps(u_1),\ Cs(u_1)),\ (Ps(u_2),\ Cs(u_2)),\ldots,\ (Ps(u_i),\ Cs(u_i))] \qquad (2.1)$$

In this equation $Ps(u_i)$ is the likelihood of event u_i and $Cs(u_i)$ is the occurrence outcome of the event. Normally, the product of likelihood of occurrence and the impact of an event is used to evaluate risk as shown in Eq. (2.2).

$$RI(u_i) = Risk\left\{\frac{Consequence}{Time}\right\} = Likelihood\left\{\frac{Event}{Time}\right\} \times \mathrm{Im}\,pact\left\{\frac{Consequence}{Event}\right\} \quad (2.2)$$

A probability is used to define the likelihood and risk is evaluated in terms of an uncertainty related to a particular damage or loss.

An amount of gain or loss is usually used to represent the significance of each risk consequence of which the dimensions are, for example, days of the extended duration of an activity. This significance is used to present a utility in the decision making. The utility is the representation of risk which displays the significance of the consequence.

2.5.2 Risk Classifications

Construction risk can be classified into physical risks and capability. The physical risks are the events that obstruct the completion of a project or increase the costs and schedule. The physical risks are the things that are beyond control of a project team including acts of God, weather, impracticability or other events. The capability related risks is events associated with the work performance which can be handled by management. The capability related risks includes poor quality, safety and equipment selection. Other classification of risks is provided by the Project Management Institute (PMI) (1996) including external (uncontrollable) and internal (controllable) events.

2.5.3 Uncertainty in Risk Analysis

Uncertainty associated with the probability and consequence of a risk factor is the main concept of risk. The occurrence probability and consequence of a risk factor are also keys for distinguishing between risk and uncertainty. The uncertainty varies between certain, the case in which the probability of occurrence is 100%, and impossible, the case in which the probability of occurrence is 0%. It means that the uncertainty exists when probability of occurrence of an event is not known (Nkado, 1992 and 1995). Uncertainty also exists when consequence of an event is not known (Smith, 1999). For example, if how many days of delay of a construction operation and its probability of occurrence and consequence cannot be estimated or quantified, the risk assessment involves uncertainty.

Ayyub (1985) stated that uncertainty is related to vague and ambiguity and can be presented in terms of non-cognitive and cognitive. McCahon (1987) mentioned that the risk assessment community has begun to make a clear distinction between two types of uncertainties: aleatory (depending on chance) and epistemic (related to the knowledge of things) uncertainty. In this way, the aleatory uncertainty becomes the non-cognitive uncertainty, while the epistemic uncertainty becomes the cognitive uncertainty. These two words are interchangeable.

Non-cognitive or aleatory uncertainty is referred to in the literature as variability, irreducible uncertainty, inherent uncertainty and stochastic uncertainty. The term aleatory uncertainty is used to describe the inherent variation associated with the physical system or the environment under consideration. Sources of aleatory uncertainty can commonly be singled out from other contributors to uncertainty by their representation as randomly distributed quantities that can take on values in an established or known range, but for which the exact value will vary by chance from unit to unit or from time to time. The use of limited information and modeling uncertainty by using simplifying assumptions is the main causes of the non-cognitive uncertainty.

Cognitive or epistemic uncertainty is also termed reducible uncertainty, subjective uncertainty and model form uncertainty. Epistemic uncertainty derives from some level of ignorance or incomplete information, subjective judgement and knowledge of some characteristics in each phase or activity of the modeling process, the system or the surrounding environment. As a result, an increase in knowledge or information can lead to a reduction in the predicted uncertainty of the response of the system. Examples of sources of epistemic uncertainty are when there is little or no experimental data for a fixed (but unknown) physical parameter, limited understanding of complex physical parameter, limited understanding of complex physical processes, and the occurrence of fault sequences or environmental conditions not identified for inclusion in the analysis of a system.

Several methods are proposed to perform the risk analysis. Normally, statistical methods are used in the risk assessment and impact evaluation. Recently, various methods (i.e., the multidimensional scaling, weighted scores, index method, the analytic hierarchy process (AHP) and soft computing technology) are applied to the risk assessment. However, these methods may not appropriately address the dynamic and complex nature of construction projects and a variety of different formats of construction documents. Difficulties in data acquiring are often observed during applying these methods. Consequently, they might be processed on missing data.

The selection of risk assessment and impact evaluation methods to efficiently represent, aggregate, and propagate different types of uncertainty through computational models are clearly important. It does not only depend upon types of uncertainty involved in the risk assessment, but also the availability of data for evaluating the probability and consequence of risk factors. For example, if plenty of statistical data are available for the particular risk assessment, normally the quantitative risk assessment will be performed. The statistical and probabilistic theories are used to analyze the non-cognitive uncertainty involved in the risk assessment and impact evaluation from the quantitative data. Conversely, if the available data are incomplete or not directly applicable, a qualitative assessment will be performed. The subjective data are used in the risk assessment and impact evaluation. The use of probability analysis is limited to the qualification of qualitative aspects. The lack of know-how in measuring strategic and intangible (qualitative) parameters leads the probabilistic models to ignore their contribution to the overall analysis.

Newer mathematics, which extend or otherwise depart from the statistical probabilistic theories, are applied and are known collectively by the generalized information theory (GIT). For example, possibility theory, fuzzy set theory, and evidence theory. These alleviate the shortcoming of the statistical and probabilistic theories, where the users need only to determine a possible range, and perhaps even a most likely value for each parameter, without the input of each factor's relative frequency. Typically, a fuzzy set theory based approach is used to analyze the cognitive uncertainty involved in the subjective risk assessment and impact evaluation. The fuzzy set theory is an appropriate vehicle as it is based on the concept that all values within a certain range are possible, with the exact value being unknown. A range of values or an interval is assigned subjectively, but the individual values in the interval are not assigned a relative belief values. An expert may feel that a given parameter is within a certain range and may even have an intuitive feel for the best value within that range.

2.6 Risk Assessment

Typically, the risk assessment is divided into qualitative and quantitative risk assessments. The selection of the appropriate risk assessment method depended primarily on data available to evaluate risk and the level of comfort provided by the selected methods. Pipattanapiwong (2004) suggested that a proper way to quantify probability of risk, either subjective or objective, relies mainly on the recurring problems resulting from project risk. Typically, existing quantitative data are unavailable and unobservable. Besides, a historical data required to conduct quantitative risk analysis is not always available and even when available, it might not be applicable in the current project because some project characteristics and environment are unique. Thus, historical data may not represent every part of the current project. As a result, it is inevitable that the subjective risk assessment is adopted for providing values of the occurrence probabilities and consequences of risk based on expert's judgement and experience.

To provide a framework for providing the subjective risk assessment in association with a construction project, the following issues are determined: 1) a brief description of risk, 2) stages of a project when risk might occur, 3) elements of a project that could be affected, 4) factors that cause risk for a project, 5) relationships with other types of risk, 6) the likelihood of risk occurring, and 7) how risk could affect a project.

2.6.1 Qualitative Risk Assessment

Qualitative (or subjective) risk assessment is carried out by using the direct judgment, ranking options, comparing options, and descriptive analysis (Flanagan and Norman, 1993). Blair (2001) summarized the characteristics of qualitative risk assessment methods which included safety/review audit, checklist, what-if analysis, hazard and operability study (HAZOP), preliminary hazard analysis (PrHA), risk assessment matrix tables, analytical hierarchy process (AHP), expected monetary value(EMV) using Delphi method, and influence diagram. Based on Bender's risk assessment model, the risk assessment matrix tables were applied to the risk assessment of an offshore construction project. This method is an effective qualitative risk assessment method which can be used as a guideline for the construction risk assessment, so that the quantification of risk using this method is provided in detail.

2.6.1.1 Qualitative Risk Assessment Matrix Tables

The likelihood of occurrence and consequence of a risk factor are qualitatively described by the qualitative risk assessment matrix tables. The qualitative risk assessment is provided by assigning values in association with the likelihood of occurrence and consequence of an event. The quantified risks for various scenarios can be used to perform comparisons among scenarios. Table 2.2 and 2.3 show the descriptions of the likelihood of occurrence and consequences of an adverse scenario, respectively. Expert elicitation and actual probability data are used to provide the levels of occurrence. Table 2.3 provides categories of consequence related to schedule, other consequences (e.g., cost, technical performance) could be modified from this table.

Table 2.2 Likelihood of occurrence

Level	Description
1	Improbable, minimal, remote, can assume occurrence will not happen
2	Unlikely, small, yet possible over the life of a project
3	Occasional, likely to occur over life of project
4	Probable, highly likely, will occur at least once over the life of a project
5	Frequent, likely to occur more than once over the life of a project

Table 2.3 Consequences associated with schedule

Description	Schedule
1. Negligible	Minimal or not impact
2. Acceptable	Milestones slip, use float to recover overall schedule
3. Marginal	Some critical path items impacts that results in minor delays
4. Critical	Major and lengthy delays to a critical path item
5. Catastrophic	Multiple delays to critical path items that result in multiple and lengthy delays

Expert elicitation and actual data can be used to define the consequence described in Table 2.3 in terms of schedule delay. A better way to quantity the consequences of a risk factor is the use of extended duration to represent such consequences. The extended duration can be calculated in a simply manner by comparing the original planned schedule to the current planed schedule.

The risk matrix is formed by combining the above tables. Then, the risk assessment is provided based on the pairing of the likelihood of occurrence and consequence. Table 2.4 shows the pairing of the likelihood of occurrence and consequence. Then, Table 2.4 is considered as the risk assessment matrix.

Table 2.4 Risk assessment matrix

Likelihood level	Consequence level				
	1	2	3	4	5
1	1	2	4	8	10
2	3	4	8	13	14
3	5	6	12	16	18
4	7	8	16	20	22
5	9	10	21	24	25
Risk index	Suggested criteria				
1-5	Acceptable				
6-10	Acceptable – with review from management				
11-19	Undesirable – decision required				
20-25	Unacceptable - alternative solution required				

Risk factors can be prioritized and categorized by using the conducted list as shown in Table 2.4. Bender (2000) suggested that the risk assessment matrix is particularly suitable to the construction industry as it can be based on semi-quantitative data. As each construction project is unique, specific data may not be available and risk probability and consequence may be approximately described in relative terms. Moreover, after funding is arranged, construction is generally fast paced. A quick and effective risk quantifying method is necessary. Consequently, the risk assessment matrix is suitable for providing the risk assessment for a construction project as it could be performed quickly and effectively with good quantification performance, easy understanding and meaningful result representation.

2.6.1.2 The Analytic Hierarchy Process (AHP) and Fuzzy Analytic Hierarchy Process (FAHP)

An analytic hierarchy process (AHP) is one of the effective multicriteria decision making (MCDM) analysis tools used in modeling unstructured problems. The AHP presents the evaluation criteria in the form of a hierarchical structure. Then, the decision makers will perform the pairwise comparisons so as to represent the relative importance of each of the criteria or the degree of preference of one factor over another with respect to each criterion by using a ratio scale. The eigenvector prioritization method is applied based on three principles: decomposition, comparative judgement, and synthesis of priorities. The pairwise comparisons of the elements for a given level are evaluated regarding their parents in the next-higher level based on the requirement of the comparative judgement principle. Matrices presenting the pairwise comparisons are later used to construct ratio scores representing the local priorities of elements in the level. A composite (global) set of priorities for elements at any lower level of the hierarchy is derived by combining the calculated ratio-scale local priorities according to the requirement of the synthesis of priorities.

Typically, the discrete scale has a merit of simplicity and ease of an application. However, the uncertainty associated with the mapping of a decision maker's perception (or judgement) to a particular number (i.e., 1,2,..,9) cannot be presented by the discrete scale, but the fuzzy scale. The FAHP is one of the effective approaches used to address the uncertainty and vagueness from the subjective perception and the experience of humans in decision-making process. Büyüközkan (2007) suggests that the FAHP is an appropriate approach for overcoming difficulties in explicitly expressing the decision maker's preference involved in the comparison process.

Currently, FAHP has been applied to the selection of software development strategy, the evaluation of government websites (Büyüközkan, 2007), and the selection of global suppliers (Chan and Kumar, 2007). Assessors were asked to qualitatively estimate the level of impact that each affecting factor or a set of affecting factors has on the comparison of a pair of elements. It is believed that this approach of qualitative comparison is very practical because the impacts of uncertainty are easily expressed in linguistic terms (Chang, 1996). There is no inherent restriction on the number of levels of impacts which is used for each affecting factor or a set of affecting factors. FAHP can be applied to carry out the risk assessment by using five linguistic terms: equally significant, moderately significant, strongly significant, very strongly significant, and extremely significant of which the numerical ratings are 1, 3, 5, 7, and 9, respectively.

However, fuzzy numbers cannot be linearly ordered as shown in Figure 2.1. The decision maker might think that the fuzzy number B is larger than the fuzzy number A because the center of B is larger than the center of A. In contrast to the previous statement, the decision maker might think that the fuzzy number B might be smaller than the fuzzy number A because B spreads widely over the small numbers while A is concentrated on large numbers. Therefore, this problem is no crisp answer. Büyüközkan's extent analysis method on FAHP is one of the perfect fuzzy ranking methods. It is the integration of Chang's extent analysis method (Chang, 1996) and Zhu et al's improvement method (Zhu, 1999), which provides the computational simplicity and effectiveness.

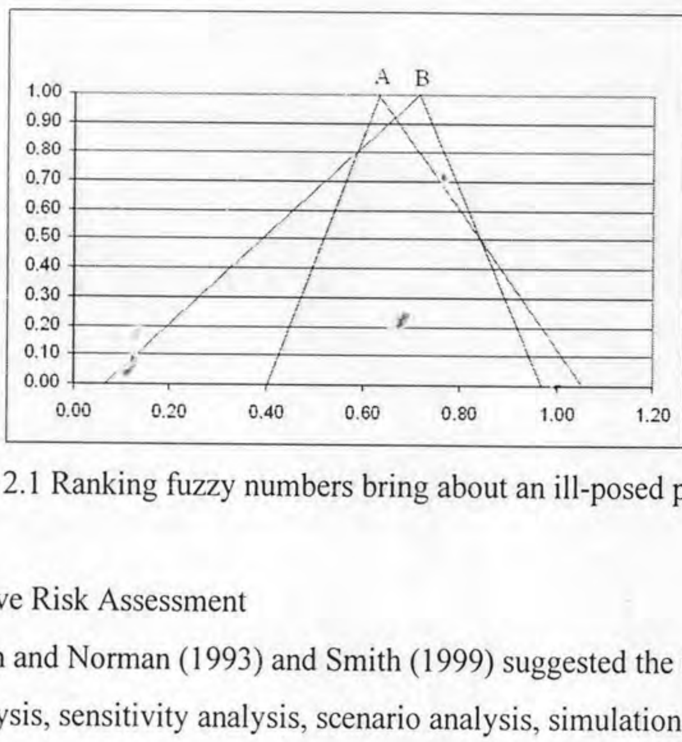

Figure 2.1 Ranking fuzzy numbers bring about an ill-posed problem

2.6.2 Quantitative Risk Assessment

Flanagan and Norman (1993) and Smith (1999) suggested the use of probability analysis, sensitivity analysis, scenario analysis, simulation analysis, correlation analysis, portfolio theory, decision tree, and fuzzy stochastic application for the risk assessment. Bender (2000) summarized the application of other methods including simulation, Failure Modes and Effects Analysis (FMEA), Fault Tree Analysis (FTA), Event Tree Analysis (ETA), expected NPV. This research concentrated on the application of simulation and fuzzy stochastic applications and the descriptions of these two approaches are accordingly provided in detail in the following section.

An objective assessment of probabilities and consequences can be exactly and precisely examined in a perfect risk scenario (Kumamoto and Henley, 1996). The exact probabilities and consequences can be developed by using the representative data together with the best method of conducting associated probabilities and consequences. However, it is totally impractical in the construction industry because of uniqueness in terms of type, function, and person involved in each project. Databases for construction projects might be acquired for in-house (company specific), governmental agencies collected from industry sources, and academic literature (Bender, 2000).

Typically, main database for construction projects is recently acquired from in-house (company specific) and academic literature. The in-house historical databases are considered as a probable best source of data used to assess the probability of an occurrence or consequence of a risk factor. Nevertheless, problems related to inadequate, unavailable, and supplemented with personal knowledge are usually found out from the in-house database (Al-Bahar and Crandell, 1990). Bender (2000) stated that these databases are company and project specific and should not be uniformly applied to new projects.

Statistics reported in the literature can be used to assess a probability of an occurrence or consequence of a risk factor. However, a special examination for their applicability to the system under observation must be required. In addition, this data is typically wide broad and generally used in understanding and developing guideline for an actual assessment.

2.7 Methods for Modeling Probability Distributions Accounting for Impacts of Risk

Risk factors can influence either a specific activity or multiple activities on a particular project. For instance, if unforeseeable site conditions are found during drilling a bored hole, it also is likely to influence other activities carried out at the same time (but perhaps in different location) on the project. The unforeseeable site conditions will increase the duration of each activity affected by the unforeseeable site condition. Moreover, the unforeseeable site condition might induce other risk factors to increasingly and frequently influence activities taking place on the project. If this correlation effect occurs for many activities contained in a network path, the variability of the path's duration may be dramatically increased. The variability in activities' durations probably leads to variability in the project' duration, especially for activities contained in the critical or near critical path. Consequently, increased variability in the project's duration increases the uncertainty of completing the project by a target date. Thus, the correlation effect will lead to an unexpected schedule overrun.

A better understanding of the effects of risk factors and correlation among activities affected by risk factors would make the schedule a more useful management tool by providing a better estimate of uncertainty in the project duration and helping to focus attention on the risk factors having the greatest impact on the activity duration and the project duration.

In the critical-path method (CPM) and probabilistic models such as program evaluation and review technique (PERT) and Monte Carlo simulation, impacts of risk factors are not explicitly measured during estimating duration. Moreover, the duration of each activity is entered or evaluated independently of the durations of other activities in the network. Although these methods are generally used in construction scheduling, these methods will not capture the impacts of risk on a particular activity and also the correlation that probably exists between durations of different activities in the schedule network.

Several methods are proposed to quantify impacts of risk on duration of construction activities and determine the correlation in a project's duration. A stochastic simulation network model with probabilistic node logic has been used to develop risk analysis models by the Accountancy Estimating and Pricing Service (AEPS) of the Ministry of Defense Procurement Executive in UK. AbouRizk and Sawhney (1993) stated that, in the developing concept of transitional correlation, the risks from project time and economic variables are better understood and quantified by an alternative analytical approach built on the PNET algorithm. The analytical approach is validated by using Monte Carlo simulation. The validation results are represented by cumulative distributions for project time variables obtained from the analytical approach which are further compared to those obtained from Monte Carlo simulation.

These methods can be broadly divided into two groups: direct and indirect elicitation. Direct elicitation models evaluate the effect of correlation with explicit specification of correlation coefficients. Direct elicitation models consist of the exact simulation, the quantile simulation, the modified second random number, and factored simulation.

2.7.1 Direct Elicitation Model

For the exact simulation (Touran and Wiser, 1992), a proper assessment of the joint probability density function for the risk factors and the correlated variables is required to conduct an exact simulation analysis incorporating the effect of risk factors and the correlation. If the evaluation of impacts of risk and the correlation coefficient is difficult to perform, the qualitative estimates are adopted.

The quantile simulation uses a facility in available Monte Carlo-simulation software to capture the effect of correlation when the correlation coefficient is known (Martinez, 1996). If the correlation coefficient is positive, the sampling procedure increases the probability of sampling the same quantiles from two probability density functions. If the correlation coefficient is negative, there will be a higher probability of sampling the n^{th} percentile and 100- n^{th} percentile from the two probability density functions. For the modified second random number simulation, the random numbers are selected by using the value of the correlation coefficient.

In a stochastic network model dealing with correlated duration which is known as the factored simulation (Woolery and Crandall, 1983), the duration of an activity consists of a time distribution for the activity duration under optimum condition and a series of time distributions for various problems (risk factors) that may extend the activity duration. This method assumes the risk factors are independent, but the impacts of the same risk factor on multiple activities are assumed to be partially or perfectly correlated. In this way, the effect of uncertainty is logically evaluated by adjusting a base duration or originally estimated duration by a series of factor-related distributions. However, because the risk factors are assumed to be independent, indirect impacts of some risk factors on other risk factors are not determined. The indirect impacts of risk factors might have significant impacts on the activity duration.

2.7.2 Indirect Elicitation Model

Indirect elicitation models do not require correlation coefficient as input. In addition, the evaluation of effect of correlation between duration of different activities might be provided with or without the evaluation of impacts of risk factors.

Several models have been proposed to evaluate impacts of risk factors on activity duration and then establish a probability function for simulation by accounting for results from the risk assessment and impact evaluation. These models are considered as the probabilistic density function developing model – based risk analysis.

The direct approaches developed for integrating the evaluated impacts of risk factors with the probabilistic density function might be difficult to be applied to construction industry. Ultimate outputs from simulation might be adjusted by adding the value of impacts of risk factors so as to reflect real world practices.

Zayed and Halpin (2001) proposed a productivity index (PI). The ultimate output obtained from the simulation particularly modeling a bored pile process is then multiplied by PI in order to overcome limitations associated with the establishment of a probabilistic density function accounting for impacts of risk factors. The evaluation of PI is provided by considering factors affecting the construction productivity. These factors include 1) soil type (e.g., sand, clay, and stiff clay), 2) drilling type (e.g., auger, bucket), 3) method of spoil removal, the size of hauling units, and space considerations at the construction site, 4) pile axis adjustment, 5) equipment operator efficiency, 6) weather conditions, 7) concrete pouring method and efficiency, 8) waiting time for other operations, 9) job and management conditions, and 10) cycle time.

Wang and Demsetz (2000) summarized the indirect-elicitation models proposed to establish a probability distribution for the activity duration by considering risk factors influencing duration of multiple activities on a particular project. Indirect-elicitation models include the simulation-based MUD (Model for Uncertainty Determination) (Carr, 1979); the DYNASTRAT (DYNAmic-STRATegy) model for dynamically allocating resources (Padilla and Carr, 1991); the PRODUF (Project Duration Forecast) (Ahuja and Nandakumar, 1985) used to create objective duration distributions of activities before processing conventional Monte Carlo-simulation procedures; the PLATFORM (AbouRizk and Sawhney, 1993) being a rule-based method that updates durations of uncompleted activities based on the durations of completed activities; and the CEV (Conditional Expected Value) model (Ranasinghe and Russell, 1992). The MUD, DYNASTART, and PRODUF require extensive inputs and historical data, while the PLATFORM depends on the performance of completed activities and treats all risk factors as having the same effects. The simulation-based model NETCOR (NETworks under CORrelated uncertainty) evaluates schedule networks when activity durations are correlated by using qualitative estimates of the sensitivity of each activity to each risk factor.

For NETCOR model, uncertainty in an activity's duration distribution, called a grandparent distribution, is a combination of a base duration and variations due to different risk factors. Uncertainty in duration of project activities affected by a particular risk factor is represented by the parent distribution. Different risk factors are assumed to cause independent effects. The durations of project activities are correlated only through the impacts of shared risk factors. The NETCOR model deals with the correlation by drawing duration samples from related portions of the parent distributions for activities that are sensitive to a given risk factor. Based on different degree of a risk factor, the parent distribution is disaggregated into child distributions. When a simulation model is executed under a particular degree of a risk factor, sample durations will be independently drawn from the corresponding child distribution.

Vuong and Watanabe (2001) introduced the risk analysis model using Monte Carlo simulation to quantify risk affecting project duration. Pipatanapong (2004) developed the duration valuation process (DVP) based on the basic set and probability theory in subjective elicitation of probability. In DVP, firstly the risk checklist is used to identify causes of risks. The relationship and transformation of causes and consequence of risks and the corresponding types of delays are demonstrated by the hierarchical structure of risk and uncertainty (HSRU). The DVP uses 1) the productivity rate of work to estimate operation duration, 2) the delay mechanisms to evaluate impacts of particular risk factors and then establish probability distributions of duration of project activities accounting for risk, and 3) the Monte Carlo simulation in spreadsheet based on the CPM scheduling method to estimate project duration.

The DVP provides a three-point estimate (i.e., optimistic, most likely, and pessimistic) which is later used to define a triangular distribution of duration. In the DVP, the optimistic duration is the original duration, the most likely duration is the expected duration, and the pessimistic duration is the original duration plus the calculated delay. The expected impacted duration is original duration plus total delay duration multiplying by the product of probabilities of delays, while the total delay duration is the added up delay durations. The probability distributions are fed to the developed simulation model processing by Monte Carlo simulation. The estimated project duration accounting for impacts of risks is a final output of the simulation model.

It can be observed that delay duration is estimated without considering effects of dependency between sets of risk factors and also the correlation between durations of project activities. In addition, probability-based methodology cannot deal with cognitive uncertainty involved in the subjective risk assessment. Furthermore, the conditional probability and multiplication rule in the probability theory are limited by the union operation, but relationships between risk factors and delays might be represented by union and intersection operations.

Most of the risk assessment and delay evaluation techniques employ the probabilistic methods and always rely on historical data (Mulholland and Christian, 1999; Hall, 1990; Vuong and Watanabe, 2001). However, by the nature of construction projects, historical data applied to the risk assessment and impact evaluation are usually fragmented, insufficient and unavailable. This problem stems from the uniqueness and large size of construction projects and rapid changes of construction operations which in turn creates a great difficulty in the data collection and interpretation (Yu and Liu, 2005). A large number of risk factors involved with construction projects also brings about uncertain data. Additionally, a consequence of missing data resulting from lack of construction projects leads to an uneven distribution of the construction duration.

As generally construction projects are complex and ill-defined, the assessors cannot provide the subjective risk assessment by giving the clear insight into the relationships between all risk factors and delays. In order to keep the interpretability and transparency, the assessors should examine the probability of a particular risk factor in conjunction with assessing risk factors all together. Moreover, the quantitative data acquired through the job records should be integrated with the subjective data so as to improve the accuracy of the risk assessment.

2.8 Fuzzy Network Scheduling and Modeling Probability Distribution Accounting for Impacts of Risk Factors

Kangari and Riggs (1989) classified risk assessment and impact evaluation models incorporated within the duration (productivity) evaluation models into two groups: 1) a classical model, i.e. probabilistic analysis, and 2) a conceptual model, i.e. fuzzy set analysis. Some probabilistic methods applied to provide the risk assessment and impact evaluation are described in the previous section. Bender (2000) also mentioned that simulation and fuzzy applications are considered as efficient quantitative risk assessment techniques. The simulation depends on the statistical and probabilistic theories, while the fuzzy applications are based on the fuzzy set and logic theories.

This section presents project – scheduling methods based on the fuzzy set and logic theory. The idea of fuzzy sets and logic is that an artificial logic system can be developed to emulate the linguistic way humans think and judge, yet achieve consistency by following accountable rules. The fuzzy set and logic theory can be used to formalize vague data and represent their fuzziness that can be entered into computation and possibility theoretic interpretation. Fuzzy logic therefore provides a mean of performing linguistic computations to quantify risk (Blair et al, 2000).

To measure the significance of risk factors affecting the activity duration and the project duration, and cognitive uncertainty, fuzzy set operations and fuzzy logic theory have been applied to model relationships between those durations and risk factors such as unforeseeable adverse site condition, weather, and labor performance (Ayyub and Haldar, 1984; AbouRizk and Sawhney, 1993). Specifically, Ayyub and Haldar (1984) employed the fuzzy set theory to model the combined effect of adverse weather conditions and labor experience to the activity duration. Fuzzy relation was used to establish relationships between the frequency of occurrence and its consequences of risk factors, and between the consequences and the durations of project activities affected by risk factors. Fuzzy relation and compositions were later used to build the crossed – relation between the frequency of occurrence and durations of project activities. Pairs of activity durations and their membership values were the output of this method. AbouRizk and Sawhney (1993) used the method proposed by Ayyub and Haldar (1984) to develop a computer system of which the input data were associated with the factors affecting duration of any activity being considered, their likelihood of occurrence and adverse consequences of affecting factors. The outputs of the developed system were the mean, variance, the shape parameters of the beta distribution from subjective data provided by the uses.

Wu and Hadipriono (1994) used a technique called the fuzzy modus ponens deduction (FMPD) to assess the impacts of affecting factors which in this study were called duration factors, on activity durations. This method provided the optimistic and pessimistic durations as required in PERT for the most likely activity durations. While the initial output of fuzzy logic behind any statement applied to the reasoning behind the risk assessment and impact evaluation is confidence levels, an ultimate output is analyzed by comparing the confidence levels of statements involved in a fuzzy logic system. The qualitative domain expert opinion can be captured by using fuzzy sets and logic for an achievable and affordable schedule. Even though this technique cannot substitute for deterministic and probabilistic scheduling methods, it does complement the set of modeling methods. Thus, this enables a better and more extensive risk assessment in cases of vague and incomplete project information. Ayyub and Haldar (1985) and Blair et al (2000) also performed risk-based decision-making by using a fuzzy set and logic theory methodology.

Fuzzy set theory has been employed to represent the uncertainty network calculations. Fuzzy arithmetic operations were used to calculate the early start and finish times along with fuzzy project durations (Chanas and Kamburowski, 1981, Dubois and Prade 1988). The typical sum and minus operation are known to be pessimistic (McCahon, 1987). Other operations are therefore proposed to avoid the inflation of the imprecision. The proposed sum and minus operations provide more optimistic results by averaging the imprecision.

The phenomenon coming from the fact that $\overline{X}+\overline{Y}-\overline{Y}\neq\overline{X}, \overline{X} and \overline{Y}$ being two fuzzy numbers that may occur at the backward calculation of the activity network has been addressed in Nasution (1999) Lorterapong and Moselhi (1996). The problem associated with the determination of critical path in a fuzzy environment which is composed of finding a critical path (and presenting its criticality) and searching for all of the critical paths of an activity network was determined in (Chanas and Kamburowski, 1981). Impacts of risk factors such as adverse site conditions, weather, and labor performance evaluated by applying fuzzy logic approaches were considered in the calculation of the early start and finish times along with fuzzy project durations

(Oliveros and Fayek, 2005).Fuzzy network scheduling enables planning experts to describe a project with approximate time data.

In addition to the above applications, many scheduling problems can be simulated and managed by fuzzy inference system as long as interactions between attributes of affecting factors and duration of project activities affected by such factors are known. The fuzzy set theory has been used to model uncertain production environments through continuous simulation (Dohnal, 1983; Fishwick, 1991; Negi and Lee, 1992; Petroive et al, 1998; Southall and Wyatt, 1988). However, examples of modeling fuzzily estimated durations of project activities through combining the fuzzy set theory with the discrete-event simulation are scarce, especially for the construction-oriented simulation. Zhang et al. (2004) presented a fuzzy discrete-event simulation approach that incorporated the fuzzy set theory to handle subjectivity, vagueness or imprecision in estimating activity duration and overcoming some problems in using probability distributions.

A risk-based schedule was developed by using the expert opinion to model a construction simulation utilizing a fuzzy stochastic technique (Blair, 1999). The steps in this technique include (1) collecting and inputting subjective information, (2) quantifying subjective information using fuzzy sets, (3) estimating various parameters of distributions, including maximum and minimum values, and the mean and variance, (4) examining graphical display of distributions and updating the estimated parameters if fit is not satisfied, and (5) inputting the stochastic estimate of duration of project activities into simulation model. The risk-based schedule technique presented that the fuzzy logic system and the scheduling method (e.g., simulation) complemented each other well. Specifically, the fuzzy set and the fuzzy logic system have been used to establish probability distributions which can be further inputted into the simulation process, while the calculation of project completion time will be provided by running a simulation model.

The use of the fuzzy set and logic theory is particularly suited to construction issues because of a lack of quantitative data and the propensity of construction experts to express uncertainty and risk factors in linguistic terms rather than mathematical expressions. However, by the nature of construction projects, logical interactions are not always known, and only a limited number of input-output data is observable. A fuzzy system constructed by expert knowledge alone will usually not perform as required because experts can be wrong about the location of the fuzzy sets and a number of rules.

Long time and a lot of efforts are required for acquiring a correct, complete and consistent set of fuzzy IF-THEN rules or the fuzzy associative memories (FAMs). Either the fuzzy IF-THEN rule or the FAMs is a context-dependent. Furthermore, in the context of construction which vagueness and subjectivity of natural linguistic statements are generally found, fuzzy rules based on qualitative knowledge of experts alone cannot adequately model a complex ill-defined operation in a construction project. In addition to time and efforts required for establishing fuzzy rules, it usually takes a lot of time to define and tune parameters which quantitatively define linguistic labels. Finally, expert knowledge in the form of linguistic statements is limited in an application to a construction project where expert knowledge is often unavailable and the complexity of construction processes to be modeled is unlimited.

To overcome these inherent limitations of the development of fuzzy systems particularly stemming from the lack of enough expert knowledge, fuzzy rules should be generated by learning from examples. Learning techniques derived from neural networks are applied to the establishment of fuzzy systems which lead accordingly to the development of neurofuzzy modeling techniques. Numerical reasoning techniques (e.g., artificial neural networks (ANN) and neurofuzzy system (NFS)) have been applied to provide the risk assessment and analysis as they are good at prediction and estimation. Several researches use these techniques to quantify impacts of risk on duration of construction operations (Marzouk and Moselhi, 2001; Zhang et al., 2004; Yu and Liu, 2005). One of the main problems related to the development and applications of these techniques is that they require sufficient information to construct their internal knowledge structure, such as the interconnection weights of networks in an ANN and the fuzzy decision rules in fuzzy logic decision making systems (FLDSs). However, in a real world problem, a limited number of input-output pairs are usually observable, so that the outputs for arbitrary inputs cannot be obtained. Consequently, the readability and performance of the resulting model cannot be achieved.

2.9 Applications of Adaptive Neurofuzzy Inference System: ANFIS

Factual data associated with an existing project are usually unavailable and unobservable, especially for the quantitative data. Expert's questionnaire, historical record, and simulation become the data acquiring approaches used to gather data for training and testing the neurofuzzy network. Most researches have concentrated on training the neurofuzzy systems on subjective data and historical data collected from expert's questionnaire and historical record, respectively. However, construction data acquired from these two approaches are always unavailable and insufficient for training the neurofuzzy models. Underestimated neurofuzzy networks resulting from having the number of training pairs fewer than the number of system parameters is usually observed and bringing about independent approximation.

This section is devoted to the development of the neurofuzzy models trained on simulation data which is rarely found in the literature on the construction risk assessment and construction scheduling. The data driven models (e.g., neural network and neurofuzzy system) trained on simulation data is an interesting choice for overcoming a problem related to insufficient training data as simulation is able to create a number of experiments related to input-output variables so that a required number of input-output pairs for training the data driven models can be collected (Kilmer, Smith, and Shuman, 1996)

To be a guideline for developing and applying the neurofuzzy models trained on simulation data for evaluating impacts of risk on construction operation, the development and application of the neurofuzzy models trained on simulation data to other areas are discussed.

An important study concerning the prediction of the system performance under various combinations of process parameters was conducted by Wang (1997). A two-dimensional axisymmetric quasic-static finite element model, in conjunction with an adaptive-network-based fuzzy inference system (ANFIS) for chemical polishing process (CMP) was developed. The data of non-uniformity on wafer surface can be obtained under different conditions of the carrier load, the pads elastic modulus and thickness by using the developed finite element model for CMP. Three input process parameters have been used in the ANFIS model to predict the non-uniformity on wafer surfaces. In this model, thee types of membership functions of analysis in ANFIS training were adopted and their difference compared the accuracy rate of prediction of non-uniformity on wafer surfaces. This prediction model was modified from a study determining the prediction of surface roughness. The ANFIS was used to accurately establish the relationship between inputs (e.g., cutting parameters and the features of surface image) and the output (e.g., surface roughness). The modification was made by considering a study examining the detection of tool-failure in a single point turning operation by using ANFIS (Lo, 2002). This model applied three types of membership functions for analysis in ANFIS training and compared their differences regarding the accuracy rate of the tool-state detection.

Considering the prediction model developed by Lo and Lin (2005), firstly the ranges of affecting factors (or inputs) were decided. Based on the system conditions under the designed ranges of affecting factors, a total of 35 sets of simulated data were obtained by using the developed two-dimensional axisymmetric quasic-static finite element model (FEM). Among them, 27 sets of data were used to train the ANFIS. The other eight sets of data were used to validate the accuracy of the prediction of non-uniformity on wafer surface. Briefly, the model consists of nine steps: (1) develop FEM simulation, (2) get training data and testing data, (3) load training data, (4) set input parameter membership function, (5) input training data into ANFIS system, (6) train data by ANFIS, (7) get results in terms of fuzzy decision rules in the ANFIS system; (8) input testing data into ANFIS system, and (9) test the trained ANFIS system.

During the training in ANFIS, 27 sets of data were used to conduct 600 times of learning. After training the ANFIS system, there were a total of 27 fuzzy rules in the architecture. Among the architecture, each input parameters membership function was divided into three regions, namely, the small, medium, and large regions. Initial and final triangular membership functions of the input variables were compared. Degree of differences is used to evaluate impacts of the input variables on the output variable. The comparison between the finite element simulated values versus predicted values provided by ANFIS was performed by changing types of membership functions including triangular membership function, Gauss membership function, and bell membership function. It can be observed that the ANFIS can obtain a higher accuracy rate of prediction of system performance by adopting the triangular membership function to conduct system training. The conclusion can be drawn that the combination of FEM and ANFIS methods can be used effectively in the prediction of system performance.

In addition to the above application, ANFIS has also been used to represent the uncertainty. Previous studies have illustrated the feasibility in utilizing ANFIS to calculate the impact of affecting factors. The outputs of this method are a knowledge network which can be presented in the form of fuzzy IF-THEN rules and a nonlinear equation. The affecting factors can be prioritized based on their significance. However, it has not been employed to address uncertainties in the network calculation.

Based on the developed scheduling methods previously explained, the fuzzy sets were used to represent uncertain duration and specified resources that are flexibly demanded. Uncertainty durations of project activities which were represented by fuzzy numbers were inputted into a simulation model. In this way, uncertainties stemming from dynamic characteristics of a construction operation under real – time conditions can be modeled by the fuzzy simulation. Theoretically, the fuzzy sets are however able to represent only a particular type of uncertainty (i.e., epistemic uncertainty). The mathematics for the fuzzy variables might inappropriately analyze every uncertainty involved in duration of project activities. The use of the mathematics for random–fuzzy variables in the network calculation could lead to misleading planning and decision making as another type of uncertainty cannot be addressed by the fuzzy set theory. The application of a highly risky project, such as fuzzy sets and their mathematics under restrictions on a particular type of uncertainty have limited the use of the fuzzy simulation as a practical problem – solving tool for construction projects (e.g., a bored piling project, a tunnel construction project).

The approaches applying both probability and possibility theory to integrate the random variables with the fuzzy variables for appropriately handling every uncertainty affecting activity durations are required to overcome the individual limitations of these two theories. However, these approaches are seldom developed. A random–fuzzy network scheduling method of which the activity duration is represented by a random–fuzzy variable is able to transcend the limitations discussed above. The fuzzy theory is employed to address systematic and unknown contributions to uncertainty, subjectivity, vagueness or imprecision in probabilistic estimating and risk analysis, while the random contributions to uncertainty are handled by applying the probability distribution. The mathematics for random–fuzzy variables (Salicone, 2007) can be used in the network calculation in order to calculate the project completion time. The credibility coefficients are used in the forward and backward path calculation and the measurement of the criticality of activities. Appendix presents the ANFIS architecture and ANFIS learning algorithm in detail.

2.10 Summary

This chapter has reviewed previous works pertaining to several relevant aspects in project scheduling, risk assessment and impact evaluation. The main concepts of these aspects are associated with the measurement of the uncertainties in attributes of a risk factor and duration of project activities which stems from random, systematic, and unknown contributions. Simulation based on probability theories has been applied to the construction risk assessment and construction scheduling in order to model non – cognitive uncertainty due to the random contribution. Fuzzy set theory, on the other hand, can be applied to produce an effective construction risk assessment model and provide the network calculation which is capable of examining systematic and unknown contributions to the cognitive uncertainty. However, both cognitive and non – cognitive uncertainties. The integration of these two theories into the project scheduling, risk assessment and impact evaluation is rarely provided to address every uncertainty.

In the construction simulation, decision variables (i.e., duration of project activities) are estimated by considering impacts of risk factors (i.e., adverse working conditions, bad weather, low productivity factors) which can be evaluated by introducing randomness into the analysis of construction processes. Input modeling for representing activity duration is generally accomplished through importing probability distributions of the random variables which are derived from sample data. However, it is time-costly to collect large amount of field data and also the historical data, especially for data related to dynamic processes typically involved in a construction project. Moreover, the selection of probability density functions, the calculation of the distribution parameters, the determination of the goodness of fit and correlations between random variables has to be provided in the scheduling methods based probability theory which in turn becomes a immense barrier for practically applying the probability theory. The examination of impact of risk factors on duration of project activities also depends mainly on assessors' attitude and judgement which cannot be appropriately examined by using the probability theory. Although fuzzy set theory appears to be an attractive solution approach in this regard, a fuzzy number is able to represent only a particular uncertainty and fuzzy rules based on qualitative knowledge alone can adequately model only very simple processes. For complex ill-defined processes, it usually takes a lot of time to define and tune the parameters which are quantitatively defined by linguistic labels. Consequently, expert knowledge in the form of linguistic statements is limited in practice.

Data driven methods (e.g., artificial neural networks (ANN) and neurofuzzy system (NFS)) have been developed to overcome the inherent limitations of the fuzzy systems due to the lack of enough expert knowledge. The NFSs can be applied to produce the risk assessment and risk analysis as they are good at prediction and estimation. As there are some problems with the development of the neurofuzzy system which the users must be aware. Therefore, a neurofuzzy modeling methodology for the risk assessment and impact evaluation must be carefully developed and applied to a construction project. Good neurofuzzy models should not be time-consuming and costly. Besides, the neurofuzzy models should have either readability or accuracy of the resulting model. Random–fuzzy variables in conjunction with their mathematics based on the contribution of the probability theory and the fuzzy set theory are introduced to overcome these limitations. To represent the attributes of risk factors and duration of project activities by the random–fuzzy number, probability–possibility transformation methods (e.g., statistical method, neurofuzzy metamodel) are used to establish membership functions of the random–fuzzy variables.