UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

IDENTIFYING COST, SCHEDULE, AND PERFORMANCE RISKS IN PROJECT PLANNING AND CONTROL – A FUZZY LOGIC APPROACH

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By
WAYNE JONES
Norman, Oklahoma
2001

UMI Number: 9994067

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

IDENTIFYING COST, SCHEDULE, AND PERFORMANCE RISKS
IN PROJECT PLANNING AND CONTROL –
A FUZZY LOGIC APPROACH

A Dissertation APPROVED FOR THE
SCHOOL OF INDUSTRIAL ENGINEERING

BY

Adedeji B. Badiru

Shivakumar Raman

Theodore Trafalis

Robert Schlegel

James D. Baldwin

Acknowledgments

I feel very fortunate to have had the opportunity to work under the direction of Dr. Adedeji Badiru throughout my graduate studies. I would also like to thank my other committee members, Dr. Shivakumar Raman, Dr. Theodore Trafalis, Dr. Robert Schlegel, and Dr. James Baldwin, for the support that each of them has given to me, as well as the industrial engineering department staff.

I am especially thankful for the support my wife, Amber, and my parents, Booker and Louise Jones, provided during this endeavor. Finally, I am thankful for the opportunity and strength provided by the Lord to pursue this study.

Table of Contents

Figures

Tables

ABSTRACT

In a "real" project management environment historical cost, schedule, and performance data are often not available. The lack of historical data requires the estimation of cost, schedule, and performance parameters. The uncertainties associated with parameter estimation results in inherent project risks. The identification and quantification of project risks associated with parameter estimation requires analytical tools that are effective and usable in project planning and control.

A review of risk identification and quantification methods revealed the need for additional methods to assess cost, schedule, and performance estimation. A risk model was developed using fuzzy set theory. The risk model was tested using a sample radar development project. The results obtained from the model proved that a practical approach incorporating subject-matter expert assessment and fuzzy set theory could be used to both identify and quantify project risks. Outputs from the model had sufficient fidelity for decision-makers to determine areas for additional surveillance and/or control.

CHAPTER 1

INTRODUCTION

Project management can be defined as the planning, scheduling and controlling of project activities and resources to achieve project objectives. A project can be very simple or extremely complex. Project management techniques are widely used in many areas including construction, manufacturing, marketing, transportation, and software development.

In this better, faster, and cheaper era, project managers face many challenges in their attempts to perform effective project management. Too many projects suffer because the cost, schedule and performance goals are unachievable. Unrealistic cost, schedule, and performance estimates and a failure to quantify and communicate the uncertainty of these estimates to managers and stakeholders often results in project failure. Project risks are seldom quantified in a manner that the estimators, management hierarchy and the customer mutually understand and accept. Estimate uncertainties often result in project cost and schedule overruns and degraded performance.

1.1 Introduction to Cost Estimation

Cost estimation generally involves predicting labor, material, utilities or other costs over time. Empirically-based cost estimation models supporting project management began to appear in literature during the 1970s and 1980s. These cost models were derived from the collection and analysis of large amounts of project data. Modelers used the data to fit a curve and analyze the parameters that affected the curve. The better the project cost data and cost model, the closer the predicted cost was to the final actual cost at project completion.

Statistical models, usually in regression analysis form, have been used to predict project cost. A disadvantage of regression-based techniques is their requirement to define in mathematical form the cost function that best fits the available historical data.

When cost data are not available, non-statistical cost models are generally used. Non-statistical cost models are generally more judgmental than those based on regression analysis. The use of non-statistical cost models puts additional emphasis on the competence and credibility of the model developer and usually increases the skepticism of auditors and higher management.

1.2 Introduction to Schedule Estimation

Schedules are an important part of any project plan. In classical project scheduling, network diagramming is a technique that uses rectangles to represent each activity. Each rectangle is connected to the rectangle or rectangles representing activities that succeed it in time. The critical path is found by determining each and every path, and then determining which path is of longest duration. The critical path determines the project's duration. Research throughout the years has focused on the use of the beta distribution to model variable activity times in the Program Evaluation and Review Technique (PERT). Justification for using a weighted average of optimistic, most likely, and pessimistic times is based on the beta distribution's ability to handle skewness and its ease of use for computing the mean activity times.

Probabilistic PERT is a technique for including the risk or uncertainty inherent in the completion of every activity in a project. Uncertainty, or risk, in the completion of an activity can be expressed mathematically if the duration of each activity is considered to be a random variable. Therefore, the duration of the critical path in a network diagram can also be viewed as a random variable because it represents the sum of random variables.

1.3 Introduction to Performance Estimation

Performance in the context of this research addresses the technical attributes of product deliverables. Performance estimation is the measurement of how well the project will meet technical requirements. The difference between the plan and the actual measure represents a technical variance which is either bad, good or somewhere in between, depending upon the level of tolerance allowed. The quantification of the variance in terms of its overall impact on the technical development process is difficult to measure. Performance estimation techniques have lacked the ability to methodically collect and organize technical data. It is important to quantitatively measure performance to ensure that intended requirements are met. Performance estimation is an area that has received little documented research attention in the context of project management.

1.4 Introduction to Cost, Schedule, and Performance Parameters

Each project has specific cost, schedule, and performance parameters. Examples of some of the parameters that may be of importance to project success are shown in Table 1. In sophisticated cost accounting systems, costs can be allocated to a number of categories such as fixed, variable, direct, indirect, sunk, etc. Schedule parameters are usually denoted in terms of time, such as hours, days, weeks or months. Performance parameters represent the technical requirements of the project. It is important to measure the degree of uncertainty in the estimates of these parameters to determine possible project impacts due to inherent risks associated with estimation.

Parameter	Type	Example
Cost	Direct, Indirect	Labor, material, taxes, insurance, depreciation, utilities, packaging, shipping
Schedule	Durations, start, stop	Hours, days, weeks, months, years
Performance	Operational	Speed, weight, range, velocity, accuracy, receiver sensitivity, horsepower, etc....
	Supportability	Mean Time Between Failure (MTBF), modularity, expansion capability, Mean Time Between Maintenance (MTBM)
	Producibility	Critical material available, special manufacturing equipment available, special facility available, skilled people to build
	Engineering Processes	Staffing, design ability
	Affordability	Design to cost, life cycle cost

Table 1. Cost, Schedule, and Performance Parameters

1.5 Introduction to Project Risk Identification

Risk is a measure of the potential inability to achieve overall project objectives within defined cost, schedule, and performance constraints. Risk management is the act or practice of dealing with risk. It includes planning for risk, assessing (identifying and analyzing) risk areas, developing risk-control options, and monitoring risk. Figure 1 shows a risk management model.

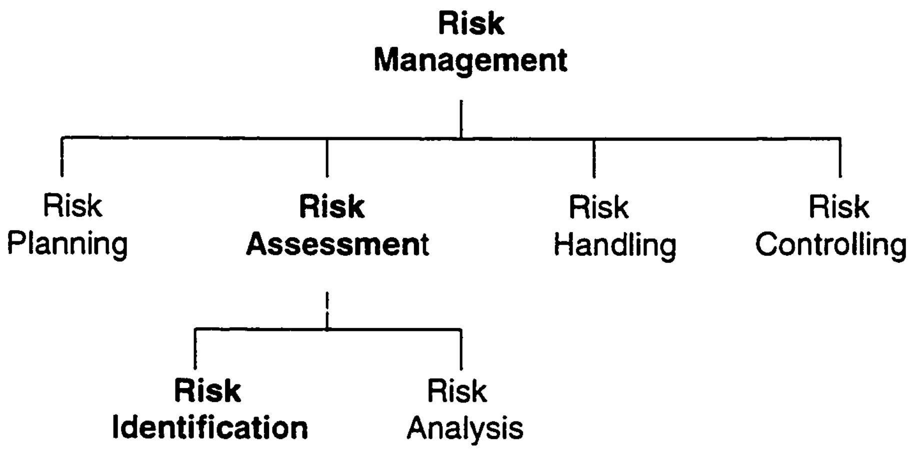


Figure 1. Risk Management

Risk planning is the process of developing and documenting an organized, comprehensive, and interactive strategy, to address project uncertainties. Risk assessment is the process of identifying and analyzing critical program areas to identify impediments to meeting cost, schedule, and performance objectives. Risk identification is the process of examining the project areas to determine the source of potential risks. Risk analysis is the process of examining each identified risk area or process to refine the description of the risk, isolating the cause, and determining the effects. Risk control is the process of identifying, evaluating, selecting, and implementing options in order to set risks at acceptable levels given project constraints and objectives. Risk monitoring is the process of systematically tracking and evaluating the effects of risk controlling actions against established metrics.

This research is focused on the identification portion of risk assessment as it relates to project cost, schedule, or performance.

1.6 Introduction to Artificial Intelligence in Project Management

Managers of very large projects have found computer-aided decision support beneficial. In addition to multidimensional capacities for data manipulation, retrieval, and display, computers are an invaluable resource for trading off design alternatives, analyzing scheduling options, tracking resource availability, and developing project cost estimates. Computers have been used to build project scenarios by modeling specific problems to be solved.

Artificial intelligence is an area that holds great promise in providing additional assistance to the project manager. Artificial intelligence is a field that has experienced rapid growth and diversity in application and practice in the last decade. The repertoire of artificial intelligence techniques has evolved and expanded to include traditional symbolic methods such as knowledge-based systems, logical reasoning, symbolic machine learning, search techniques, and natural language processing. Some of the newer fields include expert systems, neural networks, genetic algorithms, fuzzy systems, rough set theory, chaotic systems, and hybrid systems. This research focused on the use of fuzzy logic to quantify project risks due to uncertainty in cost, schedule, or performance estimates.

1.7 Research Objectives

A systematic risk identification approach using expert judgement coupled with a risk quantification method was envisioned to have utility in project planning and control. Quantification of identified risks associated with cost, schedule, or performance estimates has received little attention in previous research. Some risk assessment research has been performed for cost and schedule using Monte Carlo simulation to address estimate variability using probability distribution functions.

There were two research objectives. The first objective was to develop a mathematical risk model. The second objective was to use the model to identify and quantify project risks. The model derived from this research effort resulted in the development of a fuzzy set risk model.

CHAPTER 2

LITERATURE REVIEW OF COST ESTIMATION

Cost estimation generally involves predicting labor, material, utilities or other costs over time given a small subset of factual data. Statistical models, usually of regression form, have assisted with this projection. A neural network (NN) (see Appendix E) approach has been used on construction cost data to develop a parametric cost-estimating model for highway projects (Hegazy and Ayed, 1998). Traditionally, cost-estimating relationships (CER) are developed using regression analysis on historical data. A major disadvantage of regression-based techniques is their requirement to have a defined mathematical form for the cost function that best fits the available historical data. Another disadvantage of regression-based techniques is their unsuitability for handling large numbers of variables needed in construction projects. NNs are commonly used for difficult tasks involving intuitive judgment or requiring the detection of data patterns that elude conventional analytic techniques. Typical NNs consists of a group of processing elements organized into a sequence of layers with connections and weights between successive layers.

Hegazy and Ayed demonstrated that a NN could be developed using a spreadsheet program similar to the ones that are often used in construction cost estimation. Data from 18 highway construction projects in Newfoundland during a five year period were used to "train" the NN. The NN was developed using an Excel spreadsheet program. Excel macros were used to compare the results of successive iterations of cost inputs and the resulting budget costs. The spreadsheet implementation by Smith and Mason (1997) demonstrated the practicality of using spreadsheet programs to develop adequate NN models for use in construction project management. A comparison with traditional implementations of a NN model was also performed by Smith and Mason (1997) to determine the adequacy of the spreadsheet implementation. The spreadsheet implementation was found to track closely with the more traditional implementation.

Smith and Mason (1997) described the trade-offs of using NNs for cost estimation in a variety of simulated environments. The study also included observations on the usability, accuracy and sensitivity of NNs versus regression analysis for cost estimation. A function in two variables using a simulated data set was selected so that sampling bias, sample noise and sample sizes could be controlled. A nonlinear function ($z = 20x + y^3 + xy + 400$) with two independent cost driver variables, x and y, were used to determine the amount of output resources, z, required. The nominal range of x was 0 to 100 and of y was 0 to 50. The experiment tested four factors: (1) the modeling method used to develop the CER; (2) the sample size available for CER construction; (3) the magnitude and distribution of data imperfections (noise); and (4) the bias of the sample. For each CER method, a full factorial experiment with five levels of construction sample size, three levels of noise and three levels of bias was created resulting in a total of 45 separate prediction models for each CER. A total of 45 neural network models were built for the same experiment. Each NN consisted of two input neurons, one output neuron and two intermediate hidden layers with two neurons each. The following classical backpropagation algorithm was added to allow current weight changes to be based on past weight changes:

$$D_pW_{ij} = \eta(\alpha D_{p-1}W_{ij} + (1-\alpha)\delta_{pi}O_{pi})$$

where D_pW_{ij} is the change in weight connecting neuron j to neuron i for input vector p, O_{pi} is the output of neuron i for input vector p, δ_{pi} is the error of the output of neuron i for input vector p times the derivative of the sigmoidal transfer function, η is the training rate, and α is the smoothing factor. The results of these experiments indicated that when an a-priori CER is known, the regression model provides superior results. However, when an a-priori CER is unknown, the neural network approach is of nearly comparable precision. In addition, the NN was less dependent on the sample data used and more robust to the conditions of the problem, which was seen in lower variance across all factors. The results of this experiment suggested that an artificial NN may be an attractive substitute for regression analysis for cost estimation.

Douglis (1998) performed an analysis that focused on actual cost savings per project associated with using automated estimation and planning tools versus a manual approach. The analysis included a quantification of benefits that software executives received through the use of knowledge-based estimating tools. The knowledge-based estimation tools possessed the capability to develop a detailed task-level plan that considered project size, complexity, classification, languages, technology, process, and environment. The cost savings were in the 2% to 7% range on an individual project basis. When multiplied by the number of project managers within the organization responsible for planning and the total number of project plans, great savings can potentially be realized. The analysis took into account two factors: the number of planning days currently required to develop a manual estimate and the anticipated time saved in planning based on using knowledge-based estimation tools. The response varied by organization based upon organization size, project size, and process discipline. The analysis focused on efficiency gains and savings derived from better project planning. The results from the analysis presented in this study indicated that potential savings for an organization could be achieved using knowledge based estimation tools. The projected savings from improved project planning through knowledge-based estimation tools will improve overall project efficiency.

Cost estimation is an integral part of the procurement process of major Air Force and Navy weapon systems. Despite this essential role, the cost estimation process often provides decision-makers and analysts with limited insight due to the complex nature of the cost models that typically contain 20-30 CERS and 50-100 variables.

In an effort to provide Air Force and Navy decision-makers and analysts with additional insight into cost estimation, Campbell (1995) demonstrated a methodology that: (1) identified the critical cost drivers of a cost model; (2)

estimated the effects of these cost drivers; and (3) approximated the variance of the cost model to support confidence interval estimation.

The Navy's Tomahawk Baseline Improvement Program (TBIP) model is based on a model developed by the Program Analysis and Evaluation (PA&E) group of the Office of the Secretary of Defense. The PA&E model is a spreadsheet-based model that is used to estimate production, engineering, manufacturing, and development costs of the program. The PA&E analysts developed the model using a parametric approach based on CERs.

Using the TBIP model, a series of experiments were designed in conjunction with regression analysis to model the critical cost drivers. This model was different from the original cost model, due to the addition of confidence intervals. The estimation of the variance contained in the original cost model allowed the construction of confidence intervals for the revised model. A comparison of the intervals constructed using the revised model with those generated by the original model verified the model produced a close approximation to the original model and allowed the facilitation of "what-if" analysis.

Automation was achieved by linking the cost model to a spreadsheet containing the appropriate orthogonal experimental design. The spreadsheet contained ± 1's that represented the high and low settings of the factors. The cost model used the factor values and calculated the cost at each design point.

The cost was subsequently passed to a fourth spreadsheet that maintained the cost estimate for each run of the model.

The confidence interval estimate provided more useful information by capturing the uncertainty associated with the cost estimate. The confidence intervals related the uncertainty associated with the cost estimate through the width of the interval – the wider the interval, the more uncertainty.

Lederer and Prasad (1992) conducted a study of the cost estimating practices reported by 115 computing managers and professionals. The objective of the study was to develop a better understanding of the cost estimating process in general. The accurate prediction of information systems development costs has been a critical issue for managers. The study results were presented as prescriptions to computing managers to enable them with a method to improve the accuracy of the cost estimating process.

Much of the prior research on cost estimating has focused on the study of algorithmic technique. Lederer and Prasad (1992) identified several factors that they believed affected systems development. These factors included system size and complexity, personnel capabilities and experience, hardware constraints, the use of modem software tools and practices, users' understanding of information technology, the volatility of their requirements, and many others. To use an algorithmic technique, the estimator quantified each factor for the proposed system based on historical data about past development projects or on intuition and experience. The estimator then mathematically projected the cost of the new system based on the factors.

Lederer and Prasad (1992) suggested that to understand cost estimating, another research approach was needed. The suggested approach was to study the actual experiences of practicing computing managers and other computing professionals to obtain guidance in the cost estimating process. The approach included a questionnaire based on a previous case study of the cost estimating process. Respondents answered the questions in terms of what their organization defined as a "large project". To define a frame of reference, an arbitrary cost figure of $50,000 was chosen for a "large project".

The results of the study indicated that 84% of the respondents thought that cost estimating was very important. The respondents also reported that approximately 63% of all large projects, significantly overrun their estimates. Based on the results of the study Lederer and Prasad (1992) developed the following prescriptive guidelines:

1. Assign the initial estimating task to the final developers.
2. Delay finalizing the initial estimate until a thorough study has been accomplished.
3. Anticipate and control user changes.
4. Monitor the progress of the proposed project.
5. Evaluate project progress using independent auditors.
6. Use the cost estimate to evaluate project personnel.

7. Management should carefully study and approve the cost estimate.
8. Rely on documented facts, expert judgment, standards, and simple formulas rather than guessing, intuition, personal memory, and complex formulas.
9. Don't rely soleiy on cost estimating software for an accurate estimate.

The guidelines developed by Lederer and Prasad (1992) will likely result in improved cost estimation for computing managers that choose to adopt them.

Ting, et al (1999) developed a cost estimating methodology using the Multi-Attribute Utility Theory (MAUT) and fuzzy set theory. The use of fuzzy set theory for cost models and affordability applications was developed to address the problems of: (1) limited data for materials; (2) processes with limited empirical data; and (3) manufacturing processes with little or no previous manufacturing base. The methodology included the integration of a fuzzy set cost method and expert opinions to develop a cost model for incomplete or uncertain data. MAUT has been researched for many years. The main steps involved in applying MAUT to cost estimation are as follows:

1. Identify the objects of evaluation and the functions that the evaluation is intended to perform.
2. Identify a set of attributes that contributes to the overall product cost. Proper selection of attributes will increase the accuracy of cost estimation.

3. Construct a utility function for each level of attributes. The basic form of the utility function is: $U(x_i^\wedge) = pU(x_i^*) + (1\text{-}p)U(x_i^0)$; where $U(x_i^\wedge)$ is the utility function of attribute i at $\wedge$ level, x_i^* is the highest level giving the highest cost of attribute i; and x_i^0 is the lowest level leading to lowest cost of attribute i.

In conventional utility theory, the decision-maker's preferences are taken as utility values. The utility values are obtained through the decision maker's answers to the preference-query questions. The utility calculation is usually based on probability theory. The fuzzy set method was used to acquire utility values since the information was uncertain. In the traditional MAUT approach, the decision variables are deterministic and the utility values are crisp. Therefore, the general MAUT method was unable to handle problems with incomplete and uncertain data. In applying the utility theory to costing, experts generally provide a range of utility values for a specific cost driver. Fuzzy set theory was used to obtain a sensible result. Based on this understanding, a new method, called the Fuzzy Multi-Attribute Utility (FMAU) method, which combines MAUT and fuzzy set theory, was developed to deal with the cost-estimation issue.

The FMAU method was found to be effective for the cost estimation examples where it was applied when information about an object was incomplete or uncertain. In addition, this method was found to be more efficient than other traditional cost models because it was not necessary to collect a great amount of historical data. With FMAU, cost estimation was performed through a systematic procedure using the experts' experiences and opinions. Because of the fuzzy operations on opinions from a number of experts, the subjectivity was reduced in estimating cost.

The empirical evidence supporting the use of learning curves for planning is well documented in the literature. Learning curve models attempt to explain the phenomenon of increasing productivity with experience. The first reported use of the learning curve phenomenon was by Wright (1936), and since then an extensive number of papers have reported its use in industrial applications and research settings.

Wright's model assumed that costs decreased by a certain percentage as the number of produced units doubled. Extensions of Wright's model to account for work-in-progress and for use in project control have also been proposed to consider typical data gathering problems and scenarios in industry settings. Wright's learning curve model, $y = a\, x^{-b}$, a log-linear model, is often referred to as the "cumulative average" model because y represents the average cost of all units produced up to the xth unit. These models are discrete in unit time for cost calculation.

Badiru (1992) presented a computational survey of various univariate and multivariate learning curve models. The conventional univariate learning curve expresses a dependent variable (e.g., production cost) in terms of an independent variable (e.g., cumulative output). The log-linear model is referred to as the conventional learning curve model. The two basic forms of the log-linear model are the average cost function and the unit cost function. The average cost model specifies the relationship between the cumulative average cost per unit and cumulative production. The relationship indicates that cumulative cost per unit will decrease by a constant percentage as the cumulative production volume doubles. The model is expressed by the following equation:

$$y_x = a\ x^{-b}$$

where:

y = cumulative average cost of producing x units

a = cost of the first unit

x = cumulative production count

b = the learning curve exponent.

Figure 2 shows the graph of the log-linear learning curve that is a hyperbola.

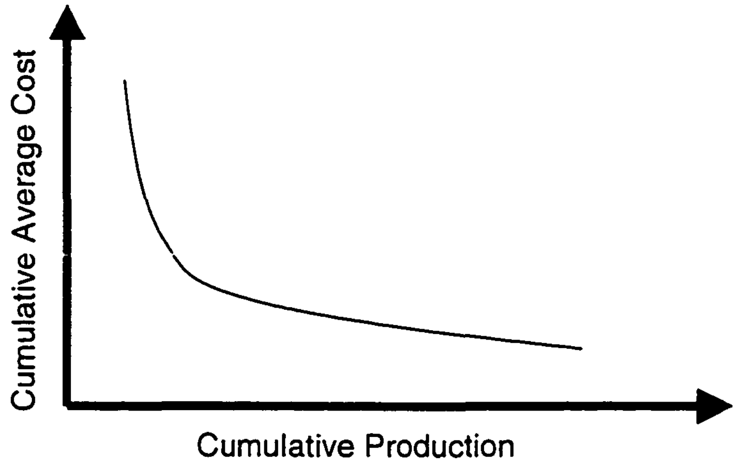


Figure 2. Log-linear Learning Curve Model

Expression for Total Cost: Using the basic cumulative average cost function, the total cost of producing x units is computed as follows:

$$y_{tc} = a\, x^{(1-b)}.$$

Expression for Unit Cost: The unit cost of producing the xth unit is given by:

$$y_{uc} = a\, x^{(1-b)} - a(x - 1)^{(1-b)}.$$

Expression for Marginal Cost: The marginal cost of producing the xth unit is given by:

$$y_{MC} = d(y_{tc}) / dx = (1 - b)(a\, x^{-b})$$

Badiru (1992) also discussed the importance of extensions and modifications of conventional leaming curves to achieve realistic analysis of productivity gain. Quantitative and qualitative factors interact to compound the productivity analysis problem. Multivariate models have been determined to be useful for detailed cost and productivity analysis in many economic and production processes. Badiru (1992) performed a comparison of the univariate model to a bivariate model. The bivariate model provided a slightly better fit than the univariate model. The bivariate model also provided more detailed information about factor interactions and better utilization of available data. Badiru (1992) suggested that the results of the computational experiment could be generalized to make a case for the appropriateness of multivariate models in many learning curve analyses.

In either the cumulative average or unit cost approach, an approximation is required to convert one type of cost to the other. This approximation can create difficulties both in empirical studies of production costs and in formulating analytical models for production planning that include learning. The use of the continuous form of the log-linear model overcomes this discrete formulation problem. By making the assumption that learning can occur continuously, learning curve projections can be made from mid-units, thus eliminating any approximation error.

Smunt (1999) examined the continuous learning approach for log-linear learning curve models and its use in analyzing productivity trends in manufacturing databases. In particular Smunt (1999) presented the derivation of the mid-unit model, a continuous form of the log-linear learning curve, which can accurately provide production cost estimates from either cumulative average costs or unit costs. The rate of increase of the total cost function was described by the first derivative of that total cost function. The fact that the rate of increase of a total cost (TC) function can be defined as the addition of unit costs over time, the unit cost function became:

$$Y_u(x) = d(TC) / d(x) = (1 - b)(a)(x^{-b}).$$

The use of the mid-unit model requires the determination of the mid-unit for most calculations and for regression analysis. Normally, the mid-unit is calculated for a production batch so that average costs can be projected from one batch to another. In essence, the average cost for a batch is the unit cost for the mid-unit of the batch. Therefore, projection of a previous batch average cost simply requires that unit costs be projected from one mid-unit to another mid-unit.

Within any given batch where x_2 is a quantity at the end of the batch and x_1 is the quantity prior to the beginning of the batch, the total cost of the batch (using the cumulative cost equation) and the unit cost were determined to be:

$$\text{Total batch cost} = y_{tc} = (a)(x_2^{1-b}) - (a)(x_1^{1-b}), \text{ and}$$

$$\text{Unit cost} = Y_{uc} = (1 - b)(a)(x^{-b}).$$

Frequently, a production process will experience a change in the learning rate. When a break in the learning curve is expected, a method to project beyond this breaking point to a different learning curve was needed. Smunt (1999) derived a formula to project on a "dog-leg" learning curve. To derive the formula to project on a "dog-leg" learning curve two separate projections were needed. First the projection from batch 1 to the breaking point (BP) and then from the BP to batch 2. The following equation projects from batch 1 to BP:

$$Y_u(BP) = \quad (BP^{-b}) / (X_{m1}) \qquad \text{(average batch cost}_1\text{)}.$$

The following equation projects from BP to batch 2:

$$Y_u(x_{m2}) = [(X_{m2}^{-b}) / (BP^{-b})] \quad (Y_u(BP)).$$

The advantages achieved by Smunt through the continuous form of the log-linear learning curve included computational speed and computational accuracy. The computational speed advantage was due to an increased ability to project all types of costs: unit; batch average; cumulative average; and total costs; from the continuous unit cost function. An additional advantage of being able to project from mid-unit to mid-unit was the ability to use regression analysis on available historical cost data. Smunt's research showed that a log-linear learning curve model provided good "fits" of empirical data for many products and processes.

The finance function at General Electric (GE) devised a method in the 1960's to figure costs caused or driven by activities, rather than the traditional method of assigning indirect and overhead costs to corporate functions such as marketing, production, or engineering, based on some measure such as labor (Johnson, 1992). This was found to be necessary because in some cases, labor costs did not vary directly with the majority of activities and, therefore, cost allocations based on that measure were grossly inaccurate. GE also traced costs upstream to the driver of the activities. Usually, this was a cross-functional analysis because activities in one department would likely generate activities in other departments (Johnson, 1992).

Company management employed these costs, derived from activities, as management accounting information. By so doing, costs could be managed by controlling activities and drivers of activities that actually caused costs. This was a different approach than the use of standard product costing methods to control costs.

The new method was not taken as far as it could have been because all the activity costs were not totaled in order to get an estimated output cost. The resurgence of Activity Based Costing (ABC) in the 1980's focused not only on the costs of activities and the drivers, but also on estimations of output costs from summing all the costs of the generators of activities. Vast improvements in computer capability made this much easier in the 1980's than in the 1960's (Johnson, 1992).

ABC attempts to assign overhead costs based on the activities that generate the costs, rather than arbitrarily assigning costs simply because the organization incurs them. Traditionally, overhead costs have been treated as having little or no causal relationship to levels of service. ABC, however, operates on the premise that many of the costs that are treated as overhead, are, in fact, variable costs. Examining overhead costs uncovered cause and effect relationships that linked activities with overhead.

The primary benefit of using ABC is to obtain more accurate costing. Advocates of ABC contend that most organizations have a poor idea of the actual costs of providing products or services. In most organizations, direct labor has declined as a major input for production and the volume of indirect costs has grown (Snyder and Davenport, 1997).

ABC can be an effective tool in an organization that produces more than one output (Rotch, 1991). If only one service or product results from work processes, then all the costs associated with that organization must be borne by the one product. In this case, budgeting is simpler because changes in activities will be passed on to the cost of the one output. To correctly assign costs to each product, the individual activity costs must be separated from the total and traced from the output back to the activity cost drivers. The ABC model provides visibility into the costs caused by activities upstream from the output.

There are two basic approaches to implement an ABC system. One is the top-down approach where the business processes are identified first, followed by the activities. The second approach is to start from the bottom by identifying activities first and then arranging them into business processes. The activities will normally be specified in detail for the departments and areas covered by the activity analysis. If the top-down approach is taken, it will ultimately require the detailed analysis at the lower level to validate, add, delete, change, and modify the initial definitions of activities and business processes.

Data must be collected in order to identify the resources the organization draws upon, the activities performed, and the products produced. There is also a need to identify the manner in which the activities consume the resources, and how products consume activities. There are many ways of gathering this type of data, all of which have advantages and disadvantages.

Activity centers are chosen on the basis of aggregating and dis-aggregating cost information. Ideally, an activity center is a discrete part of the production process; for example, in a manufacturing environment, it could be a stamping plant or paint room. Sometimes activity centers are not so distinct, such as a workstation where multiple tasks are performed. If activity centers represent different stages in a manufacturing process, then the resources that go into them should be split into separate cost pools. The cost pools will represent how much of each resource is consumed by the defining activity. Multiple cost pools associated with a single activity center are grouped together to determine the total cost of that activity. The use of ABC affects project cost estimates from the standpoint of what is included in the project cost estimate.

The Air Force Research Laboratories (AFRL) implemented an ABC system in 1997. A study was conducted by the Air Force Institute of Technology (AFIT) and documented in a Thesis written by Memminger and Wrona (1999). The purpose of the study was to examine the initial implementation of an ABC system within AFRL. The study reviewed the initial purposes for implementing ABC within AFRL and determined whether or not the goals were attained.

Exploratory research was conducted to obtain qualitative data. The primary data collection technique that was used to gather data was personal interviews. A traceability matrix that included research objectives was used to develop the research questions. The traceability matrix led to investigative interview question development.

Problems were noted with AFRL's current ABC system. First, when the system was initially designed, it was hampered by a rigid structure that was predetermined and there was insufficient training for the personnel in charge of development. The issue involved concerns with the implementation of the current ABC system. The resulting analysis showed that there were many steps that could have been taken to ensure a successful ABC system. The research concluded that ABC was a potentially beneficial tool that could have been used by AFRL if it had been developed and implemented in a different manner. ABC, as implemented, was not a beneficial tool at lower organizational levels within AFRL. In order to attain the full benefits of ABC, it was recommended that the ABC implementation be revised and include input from the lower organizational levels as necessary to achieve desired goals.

CHAPTER 3

LITERATURE REVIEW OF SCHEDULE ESTIMATION

Throughout the years, research on schedule has centered on the use of the beta distribution to model variable activity times in the Program Evaluation and Review Technique (PERT). Project managers have accepted the assumption that the mean time and variance for each activity can be determined from three time estimates (optimistic, most likely and pessimistic) based on the beta distribution.

Probabilistic PERT has the ability to represent the critical path in the precedence diagram as a random variable. The expected time t_i and the variance s_i^2 of each activity, A_i on the critical path, and the expected duration T_E of the project is:

$$T_E = \sum_{i=1}^{J} t_i$$

where there are J activities on the critical path. The variance S_E^2 of the project is:

$$S_E^2 = \sum_{i=1}^{J} s_i^2$$

If the probability density of an activity is not known a-priori but the project specialists believe that the duration of the activity will most often be between some optimistic value t_o and some pessimistic value t_p and will most likely be completed at a time t_m, then the probability density function is assumed to be a beta-distribution. The activity time for each activity is calculated using:

$$t_i = \frac{t_p + 4\,t_m + t_o}{6}$$

and

$$s_i = \frac{t_p - t_o}{6}$$

Risk and uncertainty are directly related to the differences between the pessimistic and optimistic times t_p and t_o. Probabilistic PERT shows both the critical path and the activities on the critical path that have high uncertainty or risk.

Many researchers have evaluated the use of the beta distribution for PERT analysis attempting to find theoretical justification for its use and/or to suggest other methods for calculating mean activity time. An early investigator, Grubbs (1962), criticized the use of the beta distribution, because the assumption is based primarily on empirical evidence. Farnum and Stanton (1987) showed that the empirically-based assumption had a theoretical basis. Farnum and Stanton had the belief that the activity means could be theoretically justified as (a + 4m + b) / 6 within some ranges of values for the three time estimates where a is the optimistic completion time, m is the most likely completion and b is the pessimistic completion. Farnum and Stanton recommended a refinement of the activity means if the modal value falls outside of a specific range. A similar conclusion was suggested by Littlefield and Randolph (1987), and Gallagher (1987).

Farnum and Stanton (1987) showed that the PERT estimate of the most likely or modal value, m_x, can be converted to an estimate of the mean, μ_x, using $\mu_x = (4\, m_x + 1) / 6$ where σ_x could be estimated to be 1/6, if $0.13 \leq m_x \leq 0.87$. The mean of the activity time, μ_y, could be approximately estimated from $\mu_y = a + \mu_x (b - a)$. If the estimated mode is near the upper or lower limits of the distribution, the mean and standard deviation should be calculated using the following rules: If $m_x < 0.13$ then μ_x can be estimated by $\mu_x = 2 / (2 + 1 / m_x)$ and σ_x can be estimated by $\sigma_x = [m_x^2 (1 - m_x) / (1 + m_x)]^{1/2}$.

Moitra (1990) suggested that the skewness of the data can impact the justification for accepting the beta distribution to determine mean activity times. Moitra found that when the degree of skewness is low, the beta distribution works well. However, if the degree of skewness is high, the beta assumption provides a broader distribution than the true one. Moitra suggested that a more general way of expressing the mean and standard deviation was to use:

$$\mu = (a + b + km) / (k + 2) \quad \text{and} \quad \sigma = (b - a) / c.$$

Weights of $k = 4$ and $c = 6$ result in the beta distributed mean activity times and standard deviations, which are optimal for a wide range of m values but not for extreme values of m as suggested by Farnum and Stanton (1987). Along with the more generic formula for determining the mean activity times, Moitra (1990) recommended that information be solicited from the decision-maker regarding the amount of skewness; high, medium, or low. To accomplish this the decision-maker would be asked to identify his uncertainty using a qualitative scale to determine the degree of skewness. Overall, the method attempted to obtain the level of confidence of the person making the estimates rather than arbitrarily setting the estimates based solely on the beta distribution.

Cottrell (1999) proposed a simplified variation of PERT. The proposed simplification of PERT was intended to reduce the number of time estimates were required for each task from three to two. The reduction of time estimates had dual objectives to decrease: (1) the level of effort needed to apply PERT; and (2) the required knowledge of activity durations. A symmetric duration normal distribution was a basic assumption adopted by Cottrell. The choice of the two values was made from the PERT parameters of *a, m*, or *b*. Moder (1983) reported that most time estimates are optimistic. The result is that actual project durations are generally longer than those forecasted. A more conservative estimate is to use *m* and *b*, where *m* is the mean. The simplified PERT technique reduced the level of effort and required knowledge associated with the time estimates due to the two estimates versus three. The remainder of the procedure was identical to the conventional method.

A comparison was made between the simplified PERT and the conventional PERT. The comparison involved 12 project networks. The indication was that the simplified PERT produced shorter expected project durations, but greater project duration variances. The simplified and conventional PERT means were compared. The activity duration means of the simplified PERT were shown to have a greater error than those computed using conventional PERT, especially when the distribution was highly skewed.

Shipley, et al (1997) proposed a methodology that did not rely on the beta distribution or any variation of the beta distribution to caiculate the expected project completion times. The methodology incorporated into the estimation of the activity means: (1) a measure of the decision maker's level of confidence; and (2) the decision maker's uncertainty reflected by a probability distribution. Shipley, Korvin and Omer (1997) developed the Belief in Fuzzy Probability Estimations of Time (BIFPET) methodology. The BIFPET methodology used human judgment instead of stochastic assumptions to determine project completion times. The BIFPET methodology adopted the PERT process of defining the activities and their precedence relations.

The person responsible for the completion of an activity specifies the optimistic, most likely, and pessimistic times. The person also supplies probabilities of these times being accurate.

Uretech Machinery Intemational, Inc. is a supplier of a full line of equipment for the flexible foam industry and was chosen to participate in the study to compare BIFPET to PERT. The activity supervisors on the shop floor assigned fuzzy probabilities to their estimated optimistic, most likely, and pessimistic completion times. The ranges of the time estimates were used in a fuzzy logic framework to determine expected completion times.

With PERT, the standard deviation was calculated by evenly distributing the difference between pessimistic and optimistic time estimates. In the fuzzy probability process, weights were assigned to the difference between pessimistic and optimistic time estimates on the basis of the belief that the probabilities of completing the activities on time represented the true reality of the situation. The BIFPET methodology provided a more realistic completion schedule, when contrasted with PERT generated results due to the incorporation of inputs from activity supervisors.

In order to overcome the limitations of PERT and still quantify risk, Finley and Fisher (1994) proposed a project schedule estimator using Monte Carlo methods. The method involved using random numbers in calculations of a schedule estimate. Monte Carlo methods provided a means to overcome PERT's singular focus on the critical path and also allowed for asymmetric duration distributions. Two shortcomings existed with duration estimates using the Monte Carlo methods. First, the problem of needing three duration estimates still existed. Secondly, the problem was how to simulate the correlation between activities.

Finley and Fisher (1994) used two techniques to overcome Monte Carlo shortcomings. The first technique was to run the Monte Carlo simulation only on the major milestones. This tended to lump many of the interdependent activities together and reduced the number of activities for which high and low durations needed to be estimated. The second technique was to consider only paths through the network that may have an effect on the overall duration. These techniques were found to further reduce the simulation complexity and reduce the number of activities involved.

A hypothetical construction project to build a retail space in a shopping mall was used by Finley and Fisher (1994) to test the proposed techniques. The project included 85 activities ranging from initial economic studies through detailed design and construction. The project's Work Breakdown Structure (WBS) and activity network were developed with a commercially-available project management software package. The target project duration, as determined by the length of the critical path, was 335 days.

In developing the Monte Carlo simulation, the 85 activities and over 100,000 possible paths were consolidated into 19 ranged elements and six paths. Two Monte Carlo models were used to simulate the network. The models were set up to randomly select a duration for each element and then sum the duration of each block in the abbreviated schedule. The first model only examined the critical path and simply summed the durations of the individual elements. This was repeated 1,000 times and the durations were cataloged until the probabilities of completing the projects were known for all possible durations. The second model was then created to consider all of the possible paths through the network. First, durations were randomly selected for the individual activities. Then, the longest path was found by running comparisons at points where paths diverged, ultimately picking the maximum of the two path durations. The durations were progressively summed until the longest path was determined. The second model was also run with 1,000 sets of randomly selected numbers.

The simulation that considered only the critical path indicated that there was a 42.5 percent probability that the project duration would exceed 335 days. The "all-paths" simulation probability was 70 percent. On the other hand, both projects had a 99.95 percent chance of being completed in 367 days, because that was the sum of the maximum durations for the longest possible path. At the bottom end, the critical path simulation had a 99.95 percent probability of taking longer than 295 days, while the "all-paths" method number was 314 days. The main conclusion was that the probability of finishing early was substantially reduced, when all possible paths were considered. Finley and Fisher (1994) theorized that prioritizing risks to project schedules enabled managers to reduce the risk of overrunning the desired completion date for a project.

Li and Love (1997) developed a Genetic Algorithm (GA) (see Appendix F) to optimize time and cost in construction planning. Time-cost optimization problems in construction projects are characterized by the constraints on the time and cost requirements. These problems are difficult to solve because they do not have unique solutions. Typically, if a project is running behind the scheduled plan, one option is to compress some activities on the critical path, so that the target completion time can be met. If the durations of activities are compressed it is almost inevitable that the cost of these activities will be increased. Some activities can be expedited at a lower cost than others, therefore, when a choice of activities is presented, the cheaper ones should be compressed more than expensive ones. The objective of optimization is to find the minimum total direct costs for which the reduction in time can be achieved.

Two methods have been used in the past to solve time-cost optimization problems. One method required the visual identification of several activities on the critical path and the enumeration of possible alternatives for allocating reduction time to the activities. The cost of each of these alternatives was evaluated and the one with the minimum cost was selected as the final solution. The second method was based on linear programming. Linear programming was used to establish the total cost of a project as the objective function with the associated constraints.

These two methods had a common deficiency, which was the global optimality of solutions from these methods could not be guaranteed. As combinatorial optimization problems, time-cost optimization problems, have been solved using GAs. Recent research has shown that GAs are robust and have the capacity to efficiently search complex solution spaces. The robustness of GAs is attributed to their capacity to locate the global optimum in a multi-modal landscape. Therefore, GAs are less likely to restrict the search to a local optimum when compared with point to point movement, or gradient descent optimization techniques (Goldberg, 1989).

GAs are a set of tools based on natural selection and mechanisms of population genetics. GAs employ a random yet directed search for locating the globally optimal solution. Typically, a set of GAs requires a representation scheme to encode feasible solutions to the optimization problem. Usually a solution is represented as a linear string called a chromosome whose length varies with each application. Some measure of fitness is applied to the solutions to construct better solutions. There are three basic operators in the GA system: reproduction (or selection), crossover, and mutation. Reproduction is a process in which strings are duplicated according to their fitness magnitude. Crossover is a process in which the newly reproduced strings are randomly coupled, and each couple of a string partially exchanges information. Mutation is the occasional random alteration of the value of one of the bits in the string.

Initial experiments with the GA system enabled Li and Love (1997) to determine that the length of time required for convergence was mainly caused by mutation and crossover. Crossover was improved by not allowing two identical strings to exchange information, thus eliminating wasted computational time. To improve the performance of the mutation operation, the strings were evaluated against the objective function and the strings that had the lower costs were kept as better solutions.

A construction project was used to evaluate the performance of the improved GA and the basic GA. The building of a single residential house was the selected construction project. The project required the reduction of total project duration from 64 to 57 days. Initially, eight feasible solutions were manually prepared and encoded as strings for initial inputs to the basic GA and the improved GA systems. The best manual solution had a total cost of $18,550. After approximately 400 generations of genetic operations, the improved GA converged at a value of $15,670, while the basic GA converged at the same value after about 8,000 generations. The improved GA enhanced the viability of using GAs in cost and schedule estimation.

Lorterapong and Moselhi (1996) recognized that the estimation of construction project activity durations required expert knowledge. Lorterapong and Moselhi also recognized that statements made by these experts usually contained some level of imprecision. Previous studies have demonstrated the use of fuzzy set theory for quantifying the imprecision associated with the durations of project activities. These studies in large part did not address the processing of this information for the purpose of generating a complete schedule. Lorterapong and Moselhi proposed a network scheduling method based on fuzzy set theory. The proposed method incorporated a number of features that facilitated: (1) the representation of imprecise activity durations; (2) the calculation of scheduling parameters; and (3) the interpretation of the fuzzy results generated.

The Lorterapong and Moselhi (1996) method used traditional fuzzy set operations on an example schedule network to calculate the imprecise activity durations. A Monte Carlo simulation was also performed on the example schedule network. The results indicated that the proposed method was capable of providing schedules that could appropriately account for the nature, as well as the type of uncertainties normally encountered in construction projects. The results were in close agreement with those obtained using Monte Carlo simulation. A new method, called fuzzy network scheduling (FNET), was developed as a result of the experiment to model schedule uncertainties.

CHAPTER 4

LITERATURE REVIEW OF PERFORMANCE ESTIMATION

The concept of a technical parameter hierarchy has been proposed as a method to systematically establish a framework for technical performance baselining (Kulick, 1999). The foundation of the technical performance baseline is the hierarchy that identifies all the measurable key technical elements and establishes their relative relationships and importance. The hierarchy is a comprehensive representation of the technical risk factors associated with the project. Typically medium and high risk areas are covered, however, low risks can also be included. Once the critical technical elements are identified in the form of measurable parameters, they are organized into a network similar to an organizational chart, with some parameters detailed further into "children". Typically, the highest level of the hierarchy represents the system level requirements for the project and the parameters underneath are "children". The highest level of the hierarchy can also represent operational requirements.

The technical parameter hierarchy provides a foundation for the technical performance baseline by identifying key technical parameters and establishing their relative importance. The hierarchy offers a structure to quantify the complex relationships between technical parameters.

As an example, a particular system might have a deployment life requirement of six hours. Key to this requirement is both the electrical and mechanical life of the system to be deployed. Integral to the electrical life ,might be the power consumption of the system, and the battery capacity. The parameter relationships are illustrated in Figure 3.

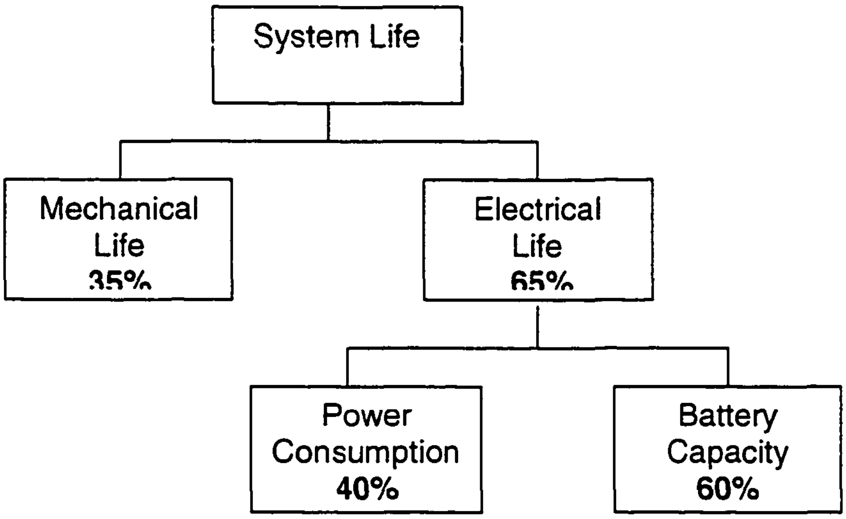


Figure 3. Technical Parameter Hierarchy

In addition to representing the proper relationship between the parameters in the hierarchy, the relative importance of each parameter must be established. The relative importance is established through a scheme of weightings. This relative importance can also be used to express risk.

In the above example, electrical life is almost twice as important in meeting the system life requirement. This means that the mechanical system is not as critical to system operations or that the mechanical systems are not expected to fail as often. Weightings associated with "children" must always add up to 100 percent.

Weightings facilitate two important functions. First, technical scores are "rolled-up" through the hierarchy to achieve a summary technical score at any level. Second, the process of calculating a "composite" technical score for associated technical parameters can be used to facilitate performance quantification.

Kulick (1998) proposed using the technical parameter hierarchy to assess technical risk. For each technical parameter in the hierarchy, a set of profiles was built to characterize the relative confidence of the estimator. The framework provided a method to organize the estimator's uncertainty into confidence levels. Kulick (1998) believed that this approach provided an enhanced technique to address technical uncertainty.

Kulick (1998) also investigated the use of Bayesian probabilistic reasoning to measure uncertainty. Bayes' Theorem (Appendix G) is a process of drawing inferences about objects or events where uncertainty exists. Bayesian techniques and networks have been used for years in diagnostic expert systems. A Bayesian network is constructed with one node used for each variable. For technical performance a Bayesian approach was proposed by Kulick (1998) to provide a framework to measure technical performance. The nodes are connected by directional links. A link pointing from node A to node B indicated that A caused B partially or in total, that A and B are functionally related, or they are statistically correlated.

Once constructed, the network was used to determine the value of each parameter from the known information that was entered. Figure 4 shows an example Bayesian network with associated probabilities to achieve an optimal life system design.

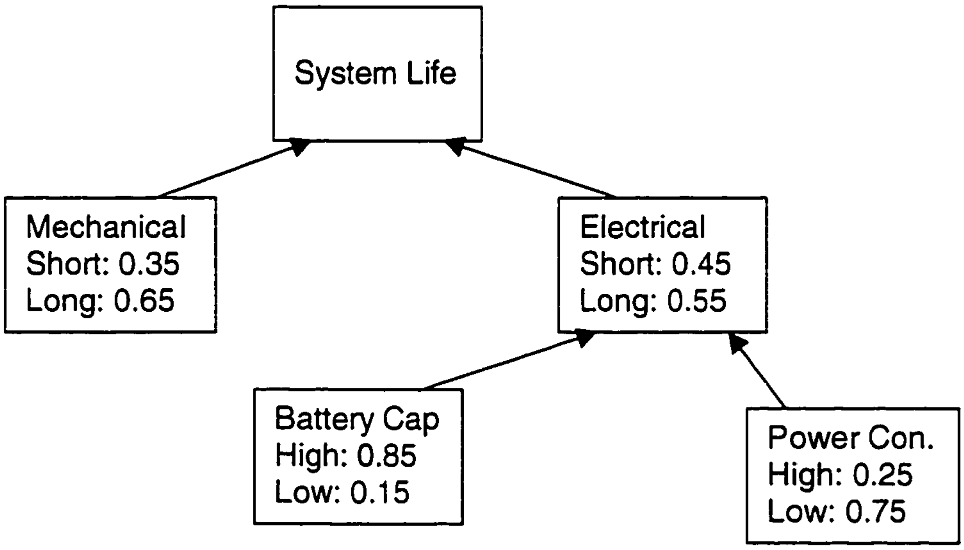


Figure 4. Bayesian Network

Kulick's (1998) Bayesian approach used the relationship data to propagate probabilities of success to defined nodes of the network which were correlated to WBS elements. The relationships between the parameters and the

WBS elements were used to determine the probability of success for each WBS element.

Anderson and Mason (1996) conducted research on the tradeoff between control system design risk and aircraft performance risk. The objective of the research was to determine the best overall design or the design that minimized design risks across several disciplines. The research was conducted on three different aircraft horizontal tail sizes ranging from a large tail to no tail at all. From fundamental aircraft design theory, the largest tail size will be the easiest to control. On the other hand, considering overall aircraft performance, a smaller tail must be considered. Since control system design and aircraft performance estimation are traditionally two separate aerospace engineering disciplines the optimization problem falls into the category of Multidisciplinary Design Optimization (MDO).

In the past, the structures discipline of aerospace engineering has used vehicle weight as a good indicator of design risk. The aircraft performance discipline uses drag counts or range and endurance as appropriate design risk metrics. Anderson and Mason (1996) used a fuzzy logic inference engine to assign a numerical value representing control system design risk. The fuzzy logic system consisted of rules that were developed from specifications that the control system had to meet. Anderson and Mason (1996) considered flying qualities, ride comfort, turbulence rejection, and stability margins in the control system design.

A model was designed to study horizontal tail sizing. Using the trimmed drag coefficient as a performance risk metric, a cost or objective function of the form $J = C_D + \mu R$ was chosen, where C_D is the trimmed drag coefficient and R is the control system design risk. The weighting factor μ was applied to the control risk value so that the relative influence of the two different metrics could be examined. The results of the analysis showed that the tailless configuration led to the lowest value for performance risk but the control design risk had the highest value. The MDO approach provided a numerical basis to perform tradeoff analysis of aircraft performance. The chosen design included a downsized horizontal tail that contributed to aircraft performance and the control system goals.

Simulation modeling and design of experiments (DOE) are two tools that industrial engineers use to improve their management of production systems. The simulation model emulates an actual system to help determine the effects of changing parameters on performance. The DOE, or experimental design, is used to plan an experiment to allow data analysis by statistical methods. Porcaro (1996) proposed the use of these two tools jointly to further improve performance analysis when it is impractical to run an actual system. DOE was proposed to generate the guidelines for the experiment, and as a method to analyze findings, while the simulation offered a way to implement the experiment. The study by Porcaro (1996) tested the effectiveness of these two tools to confirm their applicability.

The scope of the project was limited to the chemical blackening line and the adjacent strip line for Snap on-Tools. The first steps were to define the scope and objectives of the project and to develop a functional description of the existing system. Performance measures and assumptions were then documented. The baseline model was developed to provide a basis for comparison. The process consisted primarily of a series of tanks. The product was loaded in cage-like barrels and sequentially dipped in each tank, where it was processed for a period of time. One operator was responsible for loading the barrels, moving them between tanks, timing the dip time at each tank, and unloading and packaging the product at the end of the process. The primary performance measures were the barrels processed (directly related to pounds/hour the plant measure); and operator idle time. The plant wanted to evaluate several different alternatives to the current process, to determine the effect on the amount of product that could be processed. Eight possible configurations were examined by the plant project team for implementation. The results of the above scenarios were tabulated and graphed.

The baseline model indicated that the operator had sufficient time to complete all tasks involved. The reasons for performance problems were not due to basic system design. Significant performance improvements were determined to be possible through minor reconfiguration.

The DOE study brought out two points that were not evident in the initial study. The automated hoist had very little effect on production output and the extra locations per tank had almost no effect. Both of these points represented significant savings of $100,000, or more, in capital investment. The use of statistical experimental design highlighted the advantages of computer simulation. Simulation allowed the analyst to experiment with systems that were impractical to experiment with in real life. The dynamic models incorporated the random happenings that disrupted the real system. Furthermore, the information was much more representative of the real system than what could have been extracted from "static" analyses. The DOE techniques provided a front-end planning and analysis tool for the simulation project. The DOE helped define the project because it forced the consideration of performance measures, as well as the factors that affected them.

The Weibull distribution has been widely used as a failure model, particularly for mechanical components. This distribution has a varied shape and requires a fairly large sample size to produce accurate statistical estimators. In practice, sample sizes are almost always small and subjective judgement is applied, aided by a Weibull plot of the test data to determine the adequacy of the component design to meet reliability goals. The Weibull distribution is as follows:

$$f(\theta) = (P / \theta)\,(X / \theta)^{P-1}\ \exp(-(X / \theta)^{P} \quad \text{for } X > 0,\ P > 0,\ \theta > 0$$

Souza and Lamberson (1995) proposed a Bayesian procedure that incorporated past experience with design and development programs and actual field performance data. The Bayesian approach was expected to provide a structure for the application of subjective judgement.

Most of the difficulties associated with performing a Bayesian reliability analysis concerned the identification, selection, and justification of the prior distribution. As a basis for this procedure Souza and Lamberson (1995) assumed that one source of input would be a subjective percentile estimate of the Weibull distribution. The input was to be obtained from the practitioner based on past experience from a particular testing scheme. Questions to the practitioner would be of the form, "When do you expect to start seeing failures?". The response could be, "At about 6000 cycles.". Another source of input was proposed to be obtained directly from the results of past tests. As past tests results are accumulated and documented they could be combined in some fashion to yield a percentile estimate.

Prior knowledge or belief about a reliable life (X_R) for an arbitrary and fixed reliability, R, was incorporated into the estimation process used by Erto (1982). The Weibull distribution, used by Erto (1982) and by Erto and Rapone (1984) was used to incorporate X_R into the estimation process. The assumption was that an engineer or practitioner can express his knowledge or experience in terms of the reliable life (X_R) of a certain design item.

Souza and Lamberson (1995) applied the Bayesian procedure using an example taken from Nelson (1982), where a more traditional statistical procedure was used to estimate the values and confidence intervals for the shape and scale parameters of a Weibull sampling distribution. The example was on the breakdown time of an insulating fluid. A purpose of the test was to assess whether the distribution was exponential. If the distribution was exponential the fluid has a constant failure rate consistent with engineering opinion that such fluids do not age. The traditionai Weibull approach was compared to the Bayesian approach and the results indicated that the Bayesian procedure validated the engineering opinion about the constant failure rate and the fact that such fluids do not age. The results also showed that the input voltage had an effect on the failure rates. The estimate of the expected value of θ using the traditional approach tended to be closer to the results obtained by the Bayesian procedure.

The results obtained by Souza and Lamberson (1995) using the Bayesian approach to estimate the parameters of the Weibull sampling distribution, in the insulating fluid example, indicated that the use of the Bayesian statistic in reliability is not a well-defined process. The results also indicated that a certain level of knowledge about quantification of subjective information is required to use this method. Finally, how this approach is used in the future will depend on the formulation of the reliability hypothesis.

CHAPTER 5

LITERATURE REVIEW OF RISK IDENTIFICATION

Project risk is defined as an undesirable event that diminishes the chance of achieving cost, schedule, and technical performance objectives. The project risks can be internal or external to the organization.

All projects have degrees of risks. The key to successful project management is not to wish risks away, be frightened by them, or to be too optimistic about them. Rather, risks must be identified and dealt with to achieve some level of success.

Failure to keep within the estimated cost, to achieve the required completion date, and to achieve the required technical requirements, are the key problems that all project managers attempt to avoid. Therefore, project managers should identify and assess potential risks, and develop response actions to control and manage the identified risks, in addressing these problems.

Bent and Humphreys (1996) described risk analysis as a tool or method for quantifying uncertainties and their inherent risk. Risk analysis is a formalized structured approach defining the uncertainties and assessing the probability of risk associated with each uncertain item and/or event. Risk analysis allows the project manager to qualify and quantify the sensitivity of risk to the major facets of the project, namely cost, schedule, and performance.

The tools of risk analysis vary from intuition or “gut feeling” and judgment to simple manual models, to computerized simulation models. The identification and proper management of risk is a vital ingredient of successful project execution. Many studies that have been performed, concluded that today's project managers quite often have inadequate business skills, poor decision-making capabilities, and inadequate risk management capabilities (Bent and Humphreys, 1996).

The decision-making process inherently contains varying degrees of certainty, uncertainty, and risk. Certainty only exists when the exact conditions and circumstances can be specified during the period of time covered by the decision. This is rare in the project management business. Risk occurs when it is possible to specify a degree of probability or possibility for a number of likely outcomes. It is common for probability or possibility estimates to be made with historical data, or, in its absence, by personal experience. Uncertainty is present when it is not possible to specify the relative likelihood of any outcome. This occurs most often in situations where there is no historical data available, or when the task is outside the experience of project personnel.

Several authors have formulated different risk management approaches. Cooper and Chapman (1987) identified a risk management approach that included a multiphased risk analysis covering identification, evaluation, control and management of risks. Hertz and Thomas (1983) proposed a logical sequence of steps consisting of risk identification, risk measurement, and risk evaluation. They linked risk management with strategic planning and management. Charette (1989) treated risk analysis and risk management as two separate concepts, and defined risk engineering as a process consisting of both risk analysis and risk management.

Risk identification, risk measurement, and risk assessment constitutes the basic set of tools required to identify potential risk factors and assess the impact of the consequences of identified risk factors. They are also used to assess the likelihood of occurrence of these consequences and to develop the corresponding risk profiles that are necessary for the accomplishment of project objectives. In risk evaluation, project managers are required to evaluate several decision alternatives based on the risk profiles generated during risk identification, risk measurement, and risk assessment. A course of action is chosen to contain and control risks. The final phase – risk control and monitoring – is a method to periodically review project progress and to provide status to senior management and other personnel involved with project execution.

Project cost estimates generally contain a degree of uncertainty. To assess the uncertainty in a project's costs, the total costs needs to be broken down into parts. The uncertainty of each part can be identified and then the parts can be put together to get a picture of the project's risks associated with cost estimates. The standard way to breakdown a project is a Work Breakdown Structure (WBS) (Tumer, 1993). A WBS can be drawn up at any level of detail, from a simple top-level view down to the lowest level at which an individual can describe his or her work. The appropriate level of detail will vary, depending on the goals of the assessment. The appropriate level of detail for a cost risk assessment depends on several factors including the amount of time available to carry out the assessment, a specific area of concern, or an area that is critical to the success of the project.

Traditionally, cost estimates are point estimates. These point estimates may or may not accurately indicate the possible range of values that the "true" value may assume (Toakley, 1995). When estimating, the most common method of allowing for uncertainty is to add a percentage amount to the most likely estimate of the final cost. The amount added is usually called a contingency (Thompson and Perry, 1992).

Contingencies are often allowed in cost estimates. The objective of contingency allocations is to ensure that the estimated project cost is realistic and sufficient to meet any cost incurred due to risks and uncertainties. Thompson and Perry (1992) pointed out several weaknesses in using a contingency amount. Some of the problems revolve around the fact that the percentage amount is, most likely, arbitrarily derived and not necessarily appropriate for the specific project. There is also a tendency to double count risks because some estimators are inclined to include contingencies in their best estimates. A percentage addition is a single-figure prediction of estimated cost, which can imply a degree of certainty that is not justified. The percentage allows for cost risks in terms of contingency, and tends to ignore schedule and performance risks.

To alleviate the usually inflated contingency estimates, Mak and Picken (2000) described the Estimating Risk Analysis (ERA) technique that has been adopted by the Hong Kong Government for all public works projects. ERA has been used to estimate the contingency of a project by identifying and costing risk events associated with a project. The starting point for ERA was the base estimate, which was an estimate of the known scope of the project. The contingencies determined by the ERA process are then added to the base estimate.

The first step in the ERA process is to identify project risk items. The risk items are then categorized as either (1) fixed; or (2) variable. Fixed risk events are those that either happen in total or not at all. If the event happens, the maximum cost will be incurred; if not, then no cost will be incurred. Variable risk events are events that will occur, but the extent that they will occur is uncertain. The cost incurred will be uncertain and variable.

The relationship between risk type and risk allowance is shown in Table 2.

Type of risk	Average risk allowance	Maximum risk allowance
Fixed risk	(Probability) X (max.cost)	Maximum cost
Variable risk	Chance (percentage) of being exceeded (50% for example)	Chance (percentage) of being exceeded (10% for example)

Table 2. Relationship Between Risk Allowance and Risk Type

For fixed risk, maximum cost is incurred if the event occurs. For variable risk, the probability of the event occurring is multiplied by the maximum risk allowance. After the risk events are identified the fixed and variable risks allowances are calculated.

In December 1997, a summary of completed projects were received by Mak and Picken from the Hong Kong Government. The summary included the contract sum, original contingency, amount of additions, amount of omissions, final account amount, and start date of 332 building projects. Forty-five of these building projects used the ERA process to determine contingencies and 287 were performed using the traditional method. The means and standard deviations were compared between the projects with ERA and those without ERA. An F-statistical analysis showed that the F value (9.5412) was much higher than the critical value of F (1.7998) at a 1% significance level ($p < 0.00001$). Thus, the hypothesis that the variances of the two populations were equal, was rejected. The variances of the two populations indicated that the variability of contingency allowance for ERA projects was much lower with actual project costs.

Burchett (1994) studied the need for risk management models to assess risks in capital investments in the construction of an extra-high voltage (EHV) transmission line project. Mak (1995) formulated a risk management model and examined its ability to control and manage risks associated with a transmission system project to improve operation and maintenance activities. Mok (1994) investigated the feasibility of applying risk management to cost estimates for building services installations, to improve the quality of estimating total building service installation costs. Leung (1994) developed a risk management model to evaluate and select project proposals that satisfied predetermined safety and reliability objectives, for projects at Mass Transit Railway Corporation of Hong Kong. Lo (1995) investigated the feasibility of applying a risk management model for improving electricity supply reliability in a distribution system. Yu (1996) developed a knowledge-based expert system to identify, evaluate and manage project schedule risks associated with an EHV substation construction project. Ntuen and Mallik (1987) designed a knowledge-based system (KBS) for project cost estimating that was used to assist cost engineers in choosing appropriate models.

Leung, et al (1998) developed an integrated KBS to assist project managers in identifying potential risk factors and the corresponding project risks. The risk identification KBS was established by incorporating a project risk identification model. The specific expert knowledge was represented using a rule-based forward chaining search process. The rule-based structure took the form of an IF/THEN structure that logically related information contained in the IF part to other information contained in the THEN part. The working memory contained facts about the problem that were derived during the consultation. The inference engine was developed to work with the facts contained in the working memory and the domain knowledge was contained in the knowledge-base.

In order to evaluate how the proposed system could be applied in a real engineering project environment, a prototype of a risk identification KBS was configured for an EHV transmission line construction project. The experience of senior project engineers was used to develop the KBS. The KBS was tested to identify potential risk factors. The results showed that the system could be used to provide project managers with useful information on risk factors that could enable the manager to take corrective actions to control and manage identified risk factors.

Riggs, et al (1994) addressed integrating cost, schedule, and technical risk for project management. The methodology they developed was based on using the Analytic Hierarchy Process (AHP) as a basis to elicit utility functions that represent the project manager's relative preference for cost, schedule, or performance success. The process of formulating cost, schedule, and technical utility functions consisted of the following four steps:

(1) Quantifying technical (T), cost (C), and schedule (S) objectives for the utility function (TCS), using the AHP.

(2) Constructing the decision nodes and chance nodes for the logic tree associated with the decision process that eventually supported alternative selections.

(3) Assigning probabilities to the decision tree using AHP.

(4) Determining the series of decisions (path) that maximized the expected value for the decision tree.

The logic tree was composed of chance event nodes, decision nodes, and terminal branches. Probabilities were assigned to each branch emanating from a chance node, and the branches emanating from each decision node represented different decisions that could be made. The terminal branches did not have any forks emanating from them and had either dollar values, x, or utility values, u(x), assigned to them. Once the logic tree was structured and the probabilities associated with each chance node assigned, the decision tree was "rolled-up". For chance events the expected value of that node was calculated based on mathematical expectation. For decision nodes, the branch that yielded the maximum expected value was the decision of choice. The final solution to the logic tree was the series of decisions that maximized the expected value for the tree.

In many project management situations, the intent is not to maximize profit, but rather to maximize some subjective preference for achieving technical/performance criteria, staying within cost/budget limitations, and meeting schedule milestones (Riggs, et al 1994). Technical (T), cost (C), and schedule (S) outcomes were treated as either succeeding or failing, and were treated as binary variables:

T = technical success and T' = technical failure,

C = budgetary success and C' = budgetary failure, and

S = schedule success and S' = schedule failure.

The following eight possible outcomes collectively constituted the TCS utility functions:

1. TCS: u(TCS) = 1
2. T'CS: u(T'CS) < 1
3. TC'S: u(TC'S) < 1
4. TCS': u(TCS') < 1
5. T'C'S: u(T'C'S) < 1
6. T'CS': u(T'CS') < 1
7. TC'S': u(TC'S') < 1
8. T'C'S': u((T'C'S') = 0

The utility for the most favorable situation (complete success) is defined as u(TCS) = 1, and the utility for the least favorable situation (complete failure) is u(T'C'S') = 0. The remaining six utilities for those situations fall between complete success and complete failure. A TCS questionnaire was designed to elicit the relative weights for the objectives of technical, cost, and schedule. The results from the TCS questionnaire were used with the AHP process to calculate the importance (weight) of technical success (W_T), schedule success (W_S) and cost success (W_C). The resulting utility functions were defined as follows:

1. $u(TCS) = W_T + Wc + W_S = 1$
2. $u(T'CS) = Wc + W_S$
3. $u(TC'S) = W_T + W_S$
4. $u(TCS') = W_T + W_C$

5. $u((T'C'S) = W_S$

6. $u(T'CS') = W_C$

7. $u(TC'S') = W_T$

8. $u(T'C'S') = 0$

The utility values were used with the decision tree to obtain the sequence of decisions that maximized the TCS utility function. The resulting software implementation provided a mathematically based technique that addressed the shortfall of risk identification tools.

Schedule uncertainty is derived from many of the same issues associated with cost uncertainty. Conventional schedule planning is based on activity networks that are analyzed to find the critical or longest possible path from start to finish. Schedule risk analysis operates in much the same way, but allows for uncertainty in the definition of the network, its durations, and its logical structure. A schedule risk model can be built entirely independently of the cost risk model, or the performance risk model, for the same project. However, especially where the main cost is labor, the two are usually closely related. Simple relationships between cost, schedule, and performance do not always hold. It is important to look at each estimate carefully and decide if there is a link between a cost, schedule, or performance estimate.

Cooper (1994) reported from a 1992 worldwide survey, that a majority of construction projects failed to achieve schedule objectives. A survey by Laufer and Stukhart (1992) of 40 U.S. construction managers and owners indicted that for scope and design objectives only 35% of the projects considered had low uncertainty and the remaining 65% had medium to very high uncertainty at the beginning of construction. Laufer and Howell (1993) also confirmed the finding that a majority of construction projects failed to achieve schedule objectives. They concluded that approximately 80% of projects at the beginning of construction possessed a high level of uncertainty. The amount of uncertainty in the internal and external environments of a project is an important factor in determining whether there will be a schedule overrun (Mulholland and Christian, 1999).

Typically, the PERT technique has been used to develop project schedules and to compute the probability that a project will be completed on or before a scheduled time. The information from the PERT calculation can be used to determine the lower (5%) and the upper (95%) confidence limits of the schedule distribution of a project. An alternative to the probability statement and a descriptive method to convey the uncertainty in the project's schedule is the cumulative density function (CDF).

To address the uncertainties associated with estimating project schedules Mulholland and Christian (1999) developed a computer-based system for the assessment of construction schedule risk. The system consisted of three key features: (1) a hypertext information system for schedule risk identification; (2) a spreadsheet to describe and evaluate project uncertainty; and (3) direct pictorial information to assist the decision makers in selecting a realistic project completion time.

The computer-based system consisted of a risk factor identification module that contained information acquired from experts and previous construction projects. Statistical techniques were embedded in an Excel spreadsheet. The output of the system included schedule confidence limits and the risk profiles of the critical-path activities. The knowledge base consisted of four risk dimensions, namely, engineering design, procurement, construction, and project management. The Excel spreadsheet was used to model schedule risk based on inputs for optimistic, likely, and pessimistic activity times. The system also used the relative importance of risks and expected activity time as inputs. The Excel spreadsheet model also provided a means for sensitivity analyses for the different outcomes. Sensitivity analyses were performed by varying one uncertain element at a time and examining the effect of the change in that element on the total project schedule. The system provided a structured approach to identify the sources of risk in a project. The results of the risk analyses were used to make assessments and to adjust the overall project schedule.

Technical performance risk assessment is the process of determining the likelihood that the estimated technical parameters can be achieved. The technical performance parameters are related to system hardware, software, human factors, and logistics.

Software projects have been difficult to manage and too many of them end in failure. In 1995, annual U.S. spending on software projects reached approximately $250 billion and encompassed an estimated 175,000 projects (Johnson, 1995). In 1995, U.S. companies spent an estimated $59 billion in cost overruns on information system projects and another $81 billion on canceled software projects (Johnson, 1995). One explanation for the high failure rate was that managers were not taking prudent measures to assess and manage the risks involved in these projects.

Using a systematic approach, Keil, et al (1998) tapped the experience of more than 40 software project managers from around the world to identify a universal set of risk factors. The three most important risk factors were judged to be a lack of top management commitment to the project, a failure to gain user commitment, and a misunderstanding of the requirements. These and other identified risk factors were proposed for a checklist to be used to conduct future risk assessments of software projects.

One of the most interesting findings from this study was the fact that the risks perceived to be most important often lie outside the direct control of the project manager. Most of the participants in the study indicated their perceptions of risk were higher for those items over which they had little or no control. Based on this observation, Keil, et al (1998) developed and presented a typology of project risk factors and used this to suggest possible risk mitigation strategies.

One of the strengths of this approach was the focus on a higher-level risk framework by thinking about four distinct types of software project risk.

A comparison of the results with Boehm's (1991), revealed that some of the most important risks identified by the Keil, et al (1998) were missing entirely from Boehm's (1991) top 10 risk list. Boehm's (1991) list focused on execution risks while the study of Keil, Cule, Lyytinen, and Schmidt (1998) was not restricted to project execution risks. The framework presented in this study encouraged managers to explore a broader set of factors in performing risk assessments. Looking to the future, the effectiveness of different strategies for managing each type of risk needs to be carefully assessed.

In construction, alternative types of equipment or material can accomplish the tasks and functions required for an operation. Therefore, in most project scenarios, a new technology or method has to compete with traditional technological alternatives. Because of changes in the construction environment, the productivity and cost of a technology often exhibits great variability. This is especially true for unproven methods and products. Therefore, any effort to introduce a state-of-the-art technology to replace existing ones must address the inherent element of technical risk.

To address risk concerns of builders, Chao and Skibniewski (1995) applied utility theory under uncertainty to evaluate a new construction technology. The approach was based on establishing a nonlinear utility function of cost to translate the uncertain outcomes of an alternative into utility measurements. The utility measurements were weighted by their corresponding probabilities and added together to produce a total utility score for the alternative. Although the utility function was assumed to model the value scale of a decision-maker in a situation that involved risk, it was difficult to get required information to develop the form and parameters of the function.

Chao and Skibniewski (1998) developed a fuzzy decision support system that incorporated a risk factor for evaluating a new construction technology. The fuzzy logic approach, like other quantitative methods, was intended to streamline the decision analysis process and produce an evaluation according to the decision-maker's value system and judgements. Using the concepts of fuzzy sets and fuzzy logic, the methodology presented by Chao and Skibniewski (1998) produced an evaluation guided by decision rules that reflected the builder's priorities and concerns. The decision rules also served as a company record for reference in future projects.

An advantage of applying fuzzy-logic-based decision analysis, rather than the utility theory approach, to new construction technology implementation was the use of verbal descriptions in the rules. The amount of calculation involved was as great in the fuzzy logic system as the utility theory approach, the use of computers made the fuzzy-logic-based decision system easier to use.

CHAPTER 6

LITERATURE REVIEW OF FUZZY LOGIC

The primary issues in the development of a realistic cost, schedule, and performance risk identification methodology involves the management of uncertainty. A proven means of handling uncertainty is through the concept of fuzzy logic. In 1965 Dr. Lofti Zadeh developed fuzzy set theory. The type of uncertainty that this theory was meant to handle has as its roots the type of imprecision and ambiguity, which is so prevalent in human communication and thought. In particular, fuzzy set theory frees us from the so-called law of contradiction and allows us to entertain conflicting propositions. Zadeh's original paper sparked the interest of many researchers worldwide, and this has resulted in the rapid development of the field.

The concept of fuzzy set theory differs from that of conventional crisp sets mainly in the degree by which an object belongs to a set. In crisp set theory, objects are either included or they are excluded from a set. In fuzzy set theory, objects are described in such a way to permit a gradual transition from being a member of a set to a nonmember. Each object contains a degree of membership ranging from zero to one, where zero signifies non-membership, one indicates full-membership, and values in between describe the degrees of partial membership.

Theoretically, fuzzy numbers can take various shapes. In modeling real-life problems, linear approximations such as the trapezoidal and triangular fuzzy numbers are frequently used.

Crisp values can be regarded as a single value or set of values where the degrees of membership in the set assume a unit value and zero otherwise. Figure 5 shows an example of a crisp value for a single value, t_1. In practice, project managers might not have, or in some cases, might not want to specify a sharp (crisp) boundary between the possible and impossible values. A smooth transition between these two values may be preferred. Fuzzy values can be generated by imposing fuzzy boundaries on each situation. The triangular distribution shown in Figure 6 represents the situation where the value, t_3, is the most likely value between t_2 and t_4. The trapezoidal distribution shown in Figure 7 represents a more imprecise situation, where the most plausible value falls between t_6 and t_7.

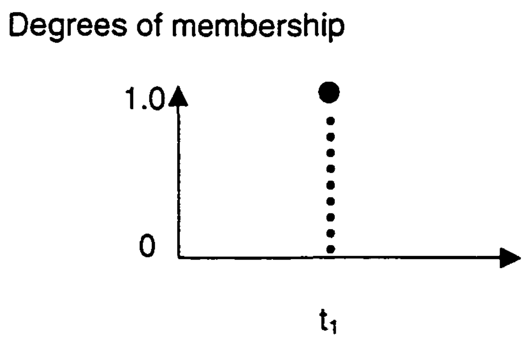


Figure 5. Crisp Value

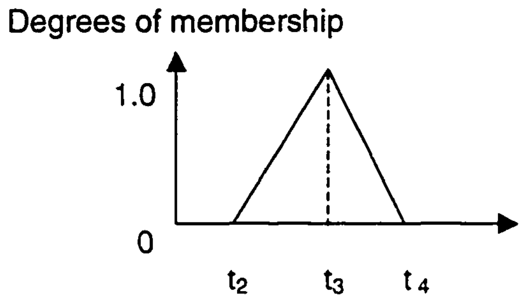


Figure 6. Fuzzy Triangle Approximation Value

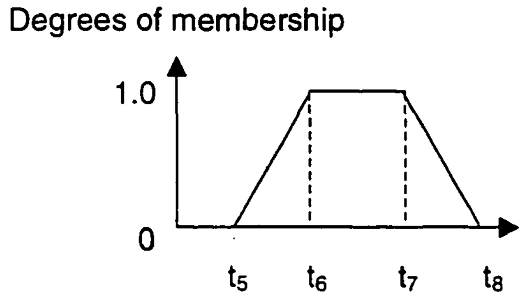


Figure 7. Fuzzy Trapezoidal Approximation Value

The triangular membership functions are of the form:

$$A(x) = \begin{cases} 0 & \text{if } x < a - b \text{ or } x > a + b \\ 1 + \dfrac{x - a}{b} & \text{if } a - b < x \leq a \\ 1 - \dfrac{x - a}{b} & \text{if } a < x \leq a + b \end{cases}$$

where a and b are any real numbers.

The trapezoidal membership functions are of the form:

$$A(x) = \begin{cases} 0 & \text{if } x < a \text{ or } x > d \\ \dfrac{x-a}{b-a} & \text{if } a \le x < b \\ 1 & \text{if } b \le x < c \\ \dfrac{x-d}{c-d} & \text{if } c \le x \le d \end{cases}$$

where $a < b < c < d$.

Badiru (1992) defined a fuzzy set A as a set of objects specified over a sample space X. For the finite set X defined as:

$$X = x_1, x_2, \ldots, x_n$$

The fuzzy set A was represented by the linear combination:

$$A = u_1(x_1), u_2(x_2), \ldots, u_n(x_n),$$

where u is the grade of membership of x in A. In general, for a sample space of objects defined as $X = \{x\}$, the fuzzy set A in X is a set of ordered pairs defined as:

$$A = \{x, u_A(x)\}, x \in X.$$

A value of $u_A(x) = 0$ implies that x is not a member of A, while $u_A(x) = 1$ implies that x is completely contained in A. For values of $u_A(x)$ between 0 and 1, x is a partial member of A.

Fuzzy sets deal with the type of uncertainty that arises when the boundaries of a class of objects are not sharply defined. Membership in such classes is a matter of degree rather than certainty one way or another, and is specified mathematically by fuzzy sets. A fuzzy value is typically associated with a degree of belief of some expert. An increasingly prevalent view is that this method of encoding information is inadequate. Assigning an exact number to the expert's opinion is too restrictive. The use of intervals is appropriate in some situations involving impreciseness. Kandel (1992) defined a fuzzy interval with a trapezoidal form. Kandel (1992) believed that an interval of values was more realistic.

Similar to the ordinary fuzzy case where values are in the unit interval, an algorithm is needed to check the equality of the expressions. The use of intervals is the simplest method for propagating uncertainty through mathematical calculations. Interval mathematics is an extension of real mathematics in which the standard operators (+, -, *, /) are applied to intervals of real numbers rather than to the real numbers themselves. The intervals are carried out through all the calculations and result in an interval that bounds the answer.

Kandel (1992) extended the application of mathematical operations to intervals of real numbers, the following example for two intervals $X = [X', X'']$ and $Y = [Y', Y'']$ yield the following rules:

Addition:

$$[X + Y] = [\min(A), \max(A)]$$

where the set A is given by

$$A = (X' + Y', X' + Y'', X'' + Y', X'' + Y'')$$

Subtraction:

$$[X - Y] = [\min(B), \max(B)]$$

where the set B is given by

$$B = (X' - Y', X' - Y'', X'' - Y', X'' - Y'')$$

Multiplication:

$$[X * Y] = [\min(C), \max(C)]$$

where the set C is given by

$$C = (X' * Y', X' * Y'', X'' * Y', X'' * Y'')$$

Division:

$$[X / Y] = [\min(D), \max(D)]$$

where the set D is given by

$$D = (X' / Y', X' / Y'', X'' / Y', X'' / Y'').$$

Yager (2000) proposed the use of fuzzy modeling as a tool for constructing customized (intelligent) decision making functions. These functions were constructed to provide a method to evaluate alternative courses of action in a way that reflected as much as possible the preferences of the responsible decision-maker. The fuzzy model was expected to provide a bridge between natural language expressions and formal mathematical representations that are useful for incorporating the kinds of concepts required for an intelligent decision valuation function.

The fuzzy system model used to construct the decision valuation functions incorporated both decision attitude and probabilistic information about the payoffs. A_i corresponded to a collection of actions available to a decision-maker. E_i was used to denote the fuzzy subsets, where $E_i(j)$ indicated the possibility of state of nature j occurring if alternative i was selected. The object state of nature j under the selection of alternative i was characterized by both its probability of occurrence, P_j, and its payoff C_{ij}. In general terms $E_i(j)$ = Function $((P_j, C_{ij}, A_i))$. Thus, the subjective possibility was a function of the probability and payoff, as well as the choice of the alternative selected.

In decision environments where no information was available regarding the probability of the states of nature, this method has been used to compare alternatives. Using this method, the decision-maker was required to provide a value $\alpha \in [0, 1]$ indicating their degree of optimism. For a given alternative A_i:

$$\text{Best}(A_i) = \text{Max}_j[C_{ij}]$$

$$\text{Worst}(A_i) = \text{Min}_j[C_{ij}]$$

The valuation of this alternative was obtained as a weighted average of the two extremes:

$$V_i = \alpha \text{ Best}(A_i) + (1 - \alpha) \text{ Worst}(A_i).$$

The weighting was determined by the measure of optimism. An extension of this approach allowed the use of probabilistic information. In Yager's approach the following two rule knowledge base described the decision-maker's valuation function.

> $Rule_1$: For the alternative under consideration, if there existed a state of nature with a feasible chance and with a payoff close to maximal for that alternative then the valuation was that payoff.
>
> $Rule_2$: For the alternative under consideration, if there existed a state of nature with a feasible chance and with a payoff close to minimal for that alternative then the valuation was that payoff.

Using these fuzzy rules and applying fuzzy systems modeling, the valuation of an alternative (A_i) was possible. The fuzzy subset M_i was close to the maximal payoff for this alternative, and the fuzzy subset N_i, was close to the minimal payoff for this alternative. The following is the valuation model for alternative A_i:

$$V = \frac{\alpha uq + (1 - \alpha)uq}{\alpha u + (1 - \alpha)u}$$

where $u = Max_j[M_i(C_{ij})\ F(P_j)]$, the degree of which there exists a feasible state of nature that has a payoff close to the maximal payoff of A_i, and q is the value of C_{ij}, where the value of u is achieved.

The primary purpose of Yager's work was to show the potential for using fuzzy systems modeling, to allow the responsible decision-maker to specify the desired performance of a valuation function. The purpose of the work was not empirical but rather to introduce an additional tool for decision-makers.

Project selection is a very complex decision-making process and involves many factors. The problem of project selection is of significant interest in engineering management. Many of the project selection problems are associated with uncertainty. The uncertainty of subjective judgement is present during the selection process. Project selection can involve a high level of risk due to uncertainties in human judgment. To address the risk associated with project selection Machacha and Bhattacharya (2000) proposed the use of fuzzy logic to select projects. To solve the problem using fuzzy logic the variables were defined and a fuzzy graph was developed. The fuzzy graph contained the range of the variable on the X-axis and the degree that the variable was contained in the fuzzy set on the Y-axis.

Machacha and Bhattacharya (2000) performed a case study using a fuzzy logic system to aid in the selection of a software product. The fuzzy logic system had as its main objective, the selection of the optimal software product. While making a decision to buy software, a consumer generally is uncertain about the criteria to use to evaluate various alternatives. An optimal decision corresponds to a selection that comes closest to meeting the consumer's desired benefits and satisfaction. The decision mechanism is constrained by the uncertainty inherent in the determination of the relative importance of each attribute. The classification of alternatives was addressed through an expert evaluation of the degree that each attribute was contained in each available option. The degree that the expert thinks a particular computer product was undesirable, acceptable, or desirable was reflected by the assignment of weights to the attributes. The attributes were the inputs to the fuzzy system.

The experts were asked to estimate the percentage of the attributes they considered to be best, good, or worst. Fuzzy values were assigned based on the percentages given by the experts. The fuzzy values and their linguistic parameters were set as follows: 70% to 100% were the "best" rating; 40% to 85% corresponded to a "good" rating; and 0% to 50% corresponded to the worst rating.

To manipulate the fuzzy numbers, a Qbasic program was written to handle the matrix that contained the fuzzy values. The implementation aided in the software selection process, but was very limited in that it only addressed linear relationships. The developers have plans to continue the development and include more than a one dimensional matrix. The example implementation provided reasonable results based on the intended purpose of the fuzzy logic system.

CHAPTER 7
CONCEPTUAL MODEL

7.1 Model Concept

The uncertainty associated with project cost, schedule, and performance estimates can be caused by a number of factors such as linguistic imprecision, unpredictability, lack of knowledge, and random errors, just to name a few. Uncertainty is inherent in most decision-making processes and the availability of data for deterministic and probabilistic models quite often do not exist. The identification of the sources of estimate uncertainty and the resulting effect are vital to project risk assessment. A model that considers expert judgement to address risk identification and quantification has utility in a "real" project decision environment.

The conceptual model for this research assumes that basic project management techniques have been used to organize the project. A Work Breakdown Structure (WBS) is typically the starting point to define the components of the project. It is also assumed that project estimates for cost, schedule, and performance have been made using a combination of the techniques described in the previous literature reviews.

The model uses project component (cost, schedule, or performance) estimates and expert judgments. The model must have the flexibility to handle the range of values associated with each of the cost, schedule, or performance parameters.

Output from a model that considers expert judgement and previously developed risk decision levels must present the decision-maker with enough information to determine risk for each of the project parameters. The model requires enough output fidelity to specifically identify the risk associated with each of the project components. The steps contained in Figure 8 will be used to develop and demonstrate the model.

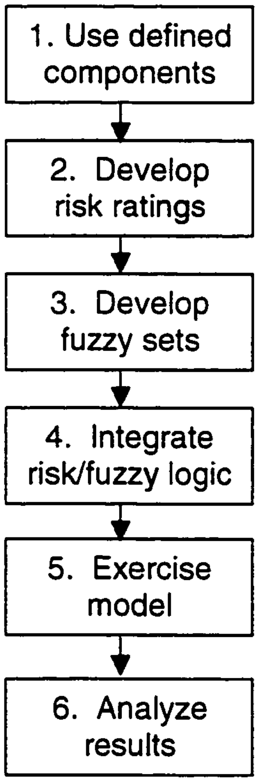


Figure 8. Implementation Steps

An example project was used to provide a logical framework from which a mathematical model could be developed. The chosen project example is the development of an aircraft radar system. Cost, schedule, and performance estimates were used to demonstrate the concept. Project activities include system design, software development, hardware procurement, integration and testing, and production.

7.2 System Components

The WBS shown in Figure 9 contains the basic project components for the radar development example. Level zero is the overall radar development project. Level 1 shows the basic project components. Typically, software development is one of the most difficult to manage project components and, therefore, it is decomposed into greater detail.

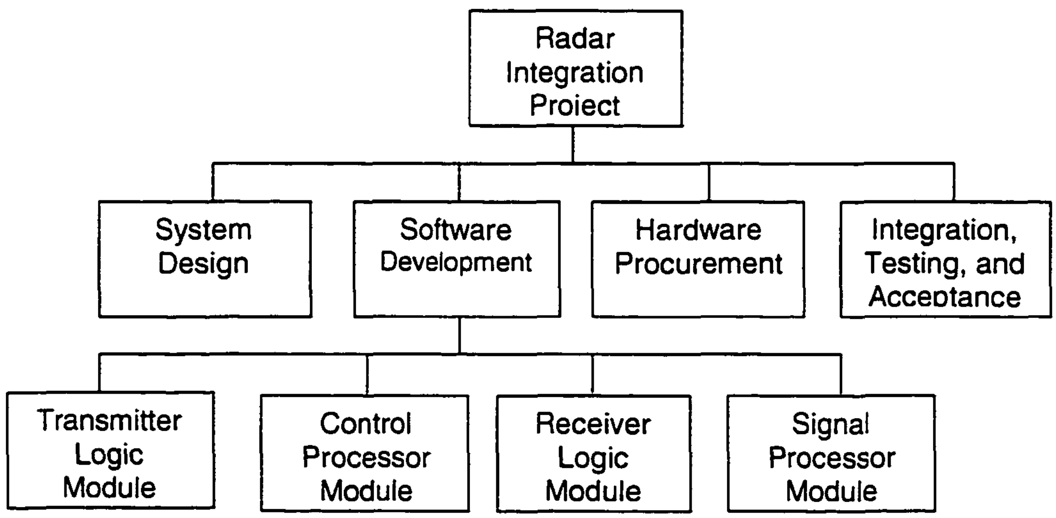


Figure 9. Radar Integration Project WBS

Project costs have been estimated for each major activity of the project. The costs for the project are summarized in Table 3. The costs for system design includes the man-hours, design support equipment such as computer aided design, circuit design, and simulation tools needed to accomplish the design. The costs associated with software development includes the man-hours, computers, etc. Commercially available hardware components are purchased. The hardware and software are subsequently integrated and tested.

ACTIVITY	COMPONENT	ESTIMATE
System Design	Antenna	160,000
	Transmitter	352,000
	Power supply	79,000
	Control data processor	400,000
	Receiver	352,000
	Signal processor	245,000
	Displays	160,000
Software Development	Transmitter	270,000
	Control data processor	536,000
	Receiver	300,000
	Signal processor	700,000
Hardware Procurement	Circuit boards	250,000
	Chassis	170,000
	Power supplies	100,000
	Signal processor	910,000
	Antennas	825,000
	Displays	200,000
Integration and test	Hardware and Software	1,710,000
Production	Fabricate 100 units	3,000,000

Table 3. Radar System Cost Estimates

A high level schedule for the radar system example is shown in Figure 10. The schedule estimates represent required time in days for each of the major project components. System design precedes software development and hardware procurement. Integration and testing are required for the hardware and software. The production of 100 radar units begins after the system has been successfully integrated and tested.

The bold connector lines show the critical path for the project. The critical path for the project indicates that the project will take a total of 731 days to complete.

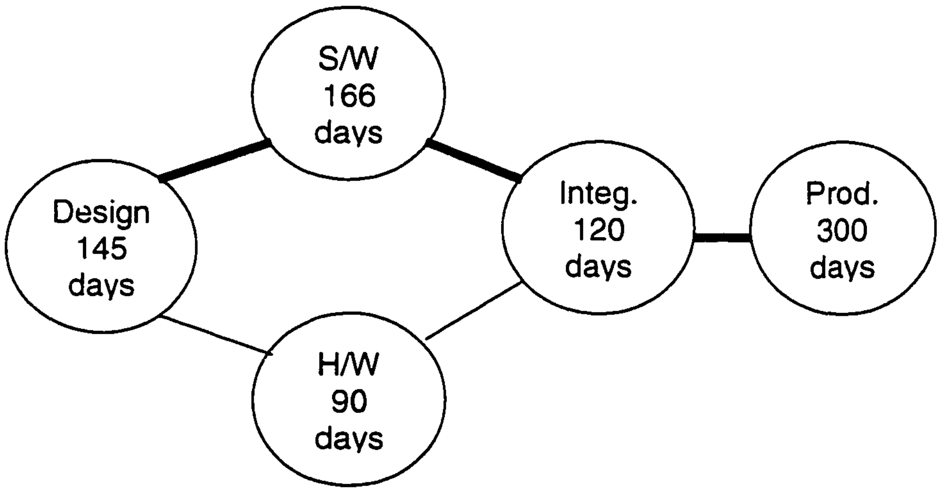


Figure 10. Radar System Schedule Estimates

There are several critical system performance requirements associated with a radar system. The critical performance requirements that are of concern for the radar development project are shown in Table 4.

PARAMETERS	REQUIREMENTS
Transmitter range	35 nautical miles
Receiver range	35 nautical miles
Identification time	3 seconds
Location accuracy	10 feet
Weight	< 45 lbs.
Size	< 24" X 12" X 36"
MTBF	3000 hours

Table 4. Radar System Performance Attributes

Some of the areas to consider during risk identification are shown in Table 5. Based on the areas shown in Table 5, expert risk assessments of cost, schedule, or performance estimates can be performed.

Risk Area	Risk Consideration
Cost	Requirements are too constrictive or identify specific solutions that force high cost. Design is not cost effective.
Schedule	Schedule is optimistic, "best case", rather than realistic. Product is larger than estimated. Contractor does not deliver components when promised. Facilities are not available on time. Proper personnel are not available.
Performance	Design implications are not sufficiently considered in concept exploration. System will not satisfy user requirements. Design relies on immature technologies or "exotic" materials to achieve performance objectives. Technology relies on complex hardware, software, or integration. Operational requirements are not properly established or are vaguely stated. Required operating environment is not sufficiently described.

Table 5. Risk Identification Considerations

7.3 Risk Rating

A critical aspect of risk identification and quantification is data collection. The two primary sources of data collection are interviews of subject-matter experts and analogy comparisons with similar systems. The framework for this research is based on risk assessments performed by subject-matter experts. Risk assessment requires the identification of both the likelihood and consequences of an event. The likelihood, as well as the consequences of risk events, are best defined as a range of possibilities. Generally, subject-matter experts relate to risks in linguistic terms. Table 6 shows risk likelihood criteria. The criteria provide a range of linguistic values to represent the likelihood of risks in the project estimates. The range of associated numerical values is shown in terms of percentages where: (1) the first number is the lower value of the range; (2) the second and third are the most likely range of percentages; and (3) the fourth value is the upper value of the range.

Level	What is the Likelihood the Estimate Is Underestimated?
a	Remote (0, 1, 3, 5) %
b	Unlikely (3, 7, 11, 15)%
c	Likely (10, 17, 26, 35)%
d	Highly likely (30, 38, 46, 55)%
e	Near certainty (> 50)%

Table 6. Likelihood Criteria

Each project estimate that has a likelihood of occurrence also has an associated consequence. For the cost, schedule, and performance estimates, the magnitude of the impact is based on the subject-matter expert's assessment. The values contained in the risk criteria should be derived from corporate policies and expert-judgments. The consequence criteria are contained in Table 7. Cost and performance are assumed to have the same project consequences for this example. The schedule for this project is not as critical as cost and performance.

Level	Given the Risk is Realized, What is the Magnitude of the Impact		
	Cost	**Schedule**	**Performance**
1	No impact (0, 1, 3, 5)%	No impact (0, 4, 6, 10)%	No impact (0, 1, 3, 5)%
2	Small (Contingency funding needed to complete) (3, 5, 7, 9)%	Small (Additional time required to complete) (6, 11, 13, 18)%	Small (Acceptable with some reduction in margin) (3, 5, 7, 9)%
3	Medium (Funding beyond Contingency required) (7, 10, 12 15)%	Medium (Not able to meet schedule minor slip in schedule required) (14, 20, 24, 30)%	Medium (Acceptable with significant reduction in margin) (7, 10, 12, 15)%
4	Large (Significant funding needed) (10, 17, 23, 30)%	Large (Key milestone or critical path impacted major slip needed) (20, 35, 45, 60)%	Large (Acceptable - no remaining margin) (10, 17, 23, 30)%
5	Catastrophic (Cannot complete due to funding shortfall) (25, 100)%	Catastrophic (Cannot meet schedule need) (50, 100)%	Catastrophic (Unacceptable performance) (25, 100)%

Table 7. Consequences Criteria

7.4 Fuzzy Membership Sets

Fuzzy sets provide a methodology to identify and quantify identified risk. An assessment of both the likelihood and consequence of risks associated with project cost, schedule, or performance form the basis for the development of the fuzzy sets. The trapezoidal fuzzy membership function has a structure that can represent the range of values solicited from subject-matter experts. The range of values associated with cost, schedule and performance estimates can be easily represented by trapezoidal fuzzy sets. An advantage of the fuzzy approach is that each possible level does not have to be crisply defined. Intermediate levels can be accounted for through membership in more than one fuzzy set.

For the aircraft radar system project example, a determination by a subject-matter expert of the likelihood that the cost, schedule, or performance estimate is understated is required. Based on risk considerations similar to the ones contained in Table 5, the subject-matter expert makes an assessment of risk likelihood and consequence for each activity.

The overall risk rating criteria considers both likelihood and consequences. Using these risk ratings, activities that have high, moderate, or low risk can be identified. Risk ratings shown in Table 8 were adopted from the Risk Management Guide for DoD Acquisition, Jan 2000.

HIGH – Unacceptable. Major disruption likely. Different approach required. Priority management attention required.
MODERATE – Some disruption. Different approach may be required. Additional management attention may be needed.
LOW – Minimum impact. Minimum oversight needed to ensure risk remains low.

Table 8. Overall Risk Rating Criteria

The risk rating matrix (Table 9) shows the resultant relationship between likelihood and consequences. The high (H), medium (M), and low (L) risk ratings are determined by the values of "likelihood" and "consequences" shown in Tables 6 and 7. Table 9 also corresponds to the fuzzy rules developed earlier in the chapter. "Likelihood" and "consequences" are the inputs to the fuzzy logic system and the outputs are high (H), medium (M), or low (L).

Likelihood		1	2	3	4	5
	e	L	M	H	H	H
	d	L	M	M	H	H
	c	L	M	M	M	H
	b	L	L	L	M	M
	a	L	L	L	L	M
		Consequences				

Table 9. Risk Rating Matrix

Figure 11 shows a conceptual fuzzy risk model that can "reason" or map inputs to outputs. The fuzzy rules for the model are of the form "*IF A AND B, THEN C*". Inputs to the fuzzy system are "A" and "B". "A" is the input for the likelihood fuzzy set and "B" is the input for the consequence fuzzy set. Once "A" and "B" are input, the degree of membership is determined for all of the combinations of fuzzy sets. A fuzzy operation is required to combine each of the subsequent fuzzy set outputs and obtain a crisp number. The process of determining this crisp number is called defuzzification. "C" is the consequence of the fuzzy rule "*IF A AND B, THEN C*".

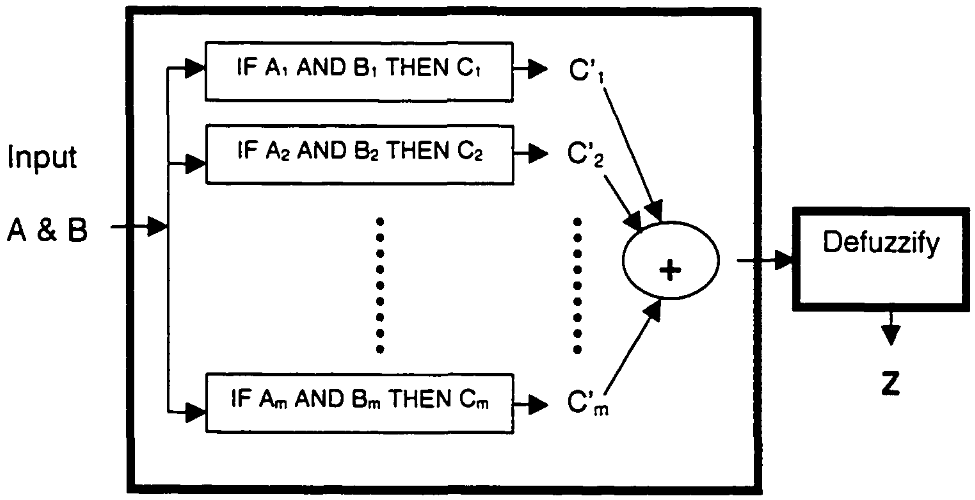


Figure 11. Fuzzy System Model

Generic fuzzy membership functions for likelihood of occurrence are shown in Figure 12. Likelihood of occurrence has five fuzzy sets (Figure 12), which are remote (R), unlikely (UL), likely (L), highly likely (UL), and near certainty (NC). The cost, schedule, and performance consequences membership functions are shown in Figures 13, 14, and 15. The consequences fuzzy sets for cost, schedule and performance are also divided into five fuzzy sets, which are no impact (NI), small (S), medium (M), large (L), and catastrophic (C).

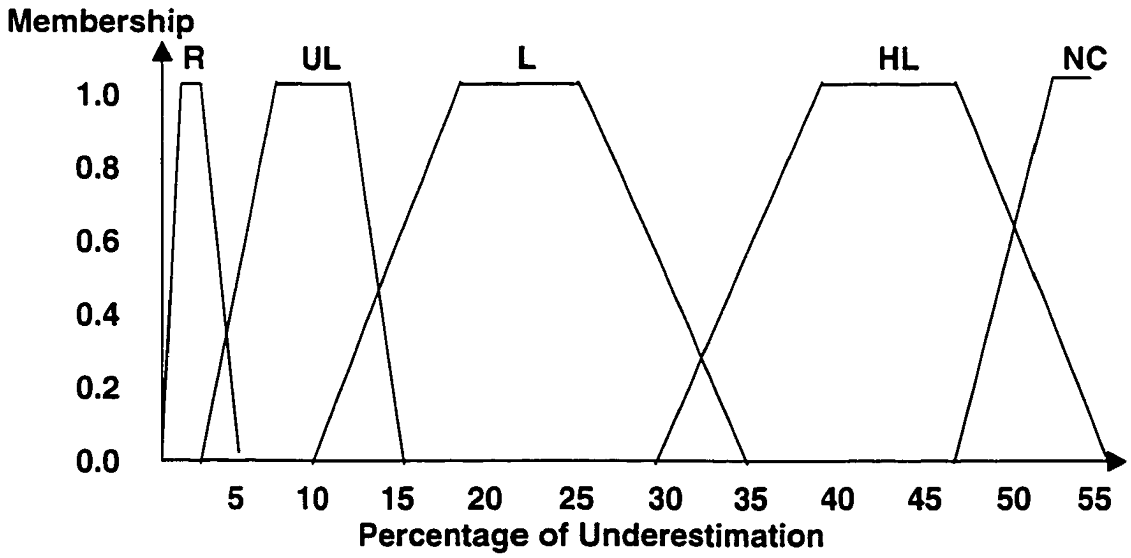


Figure 12. Likelihood of Occurrence Membership Functions

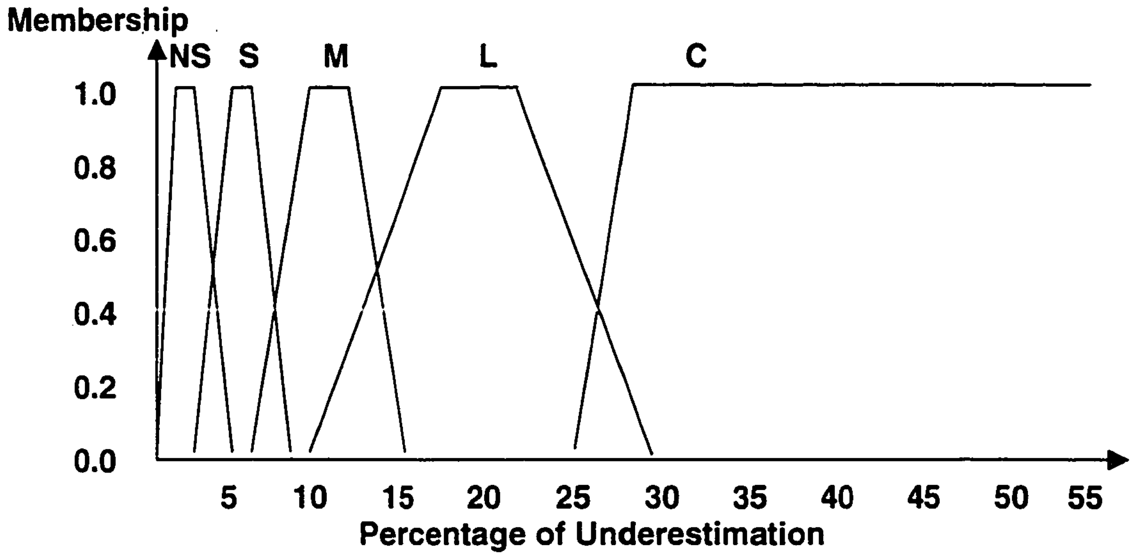


Figure 13. Cost Consequence Membership Functions

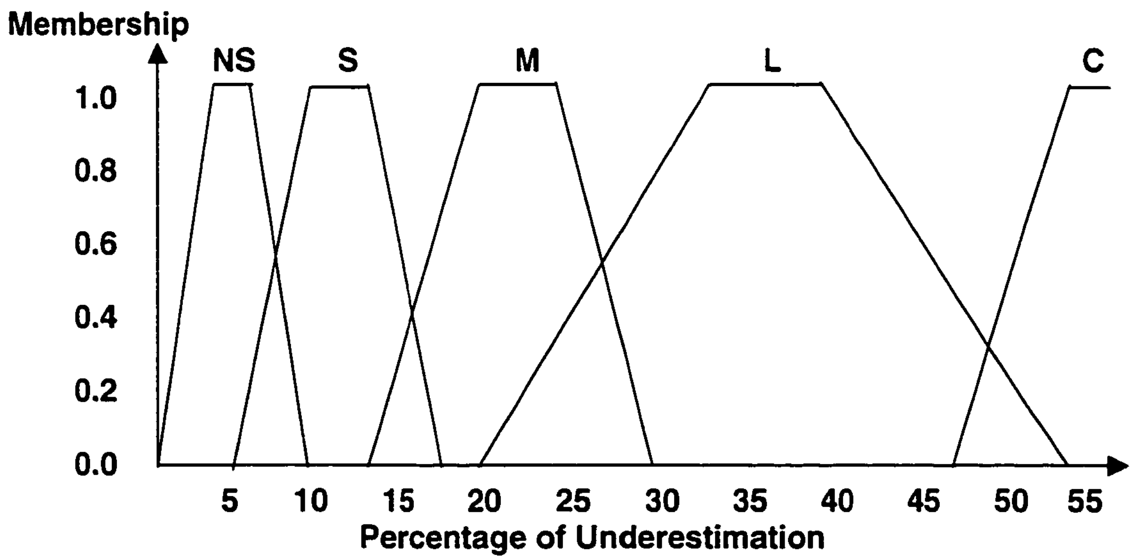


Figure 14. Schedule Consequence Membership Functions

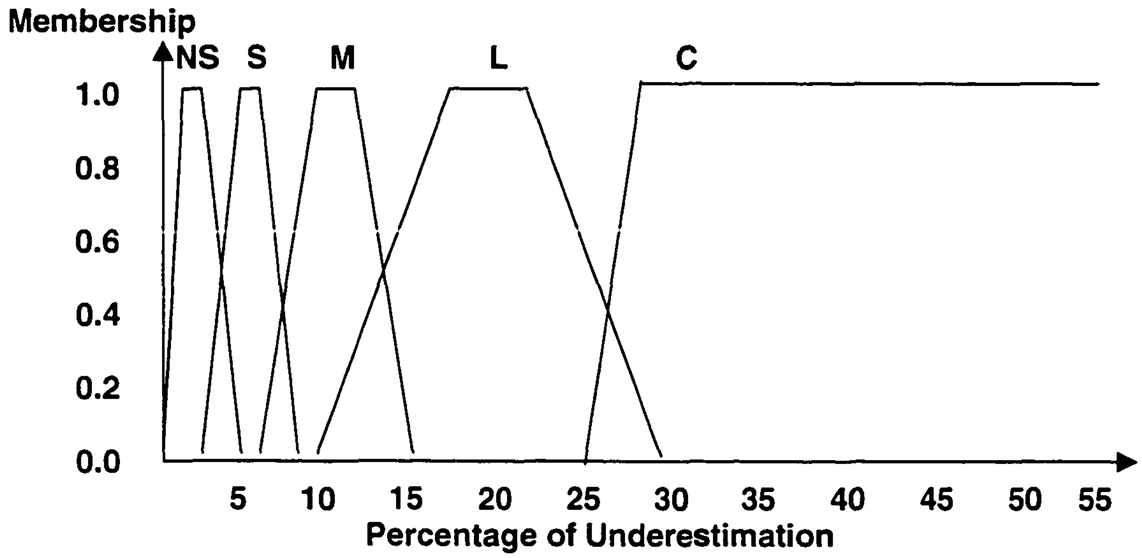


Figure 15. Performance Consequence Membership Functions

7.5 Fuzzification and Rules

Inputs are solicited from the subject-matter expert for both the likelihood that the estimate is understated and the resulting consequence. The degree of membership within the likelihood and consequence fuzzy sets are determined for subsequent application within the fuzzy rules.

Fuzzy rules are defined in simple language terms. Inherent to fuzzy rules, no decisions are required for breakpoints. There are also no decisions to be made about the functional form of the relationships. The subject-matter expert's inputs for both likelihood and consequence are associated with the risk level through the fuzzy rules. The result of a rule (low, medium, or high) is based on the Risk Rating Matrix shown in Table 9. The rules for the radar system example are divided into three categories for cost, schedule, and performance. Example rules for cost, schedule, and performance are shown below:

Cost rules for likelihood and consequence of underestimation

Rule 1: IF remote AND no impact THEN project risk is low.

Rule 2: IF remote AND small THEN project risk is low.

Rule 3: IF remote AND medium THEN project risk is low.

Rule 4: IF remote AND large THEN project risk is low.

Rule 5: IF remote AND catastrophic THEN project risk is medium.

Rule 6: IF unlikely AND no impact THEN project risk is low.

Rule 7: IF unlikely AND small THEN project risk is low.

Rule 8: IF unlikely AND medium THEN project risk is low.

Rule 9: IF unlikely AND large THEN project risk is medium.

Rule 10: IF unlikely AND catastrophic THEN project risk is medium.

Rule 11: IF likely AND no impact THEN project risk is low.

Rule 12: IF likely AND small THEN project risk is medium.

Rule 13: IF likely AND medium THEN project risk is medium.

Rule 14: IF likely AND large THEN project risk is medium.

Rule 15: IF likely AND catastrophic THEN project risk is high.

Rule 16: IF highly likely AND no impact THEN project risk is low.

Rule 17: IF highly likely AND small THEN project risk is medium.

Rule 18: IF highly likely AND medium THEN project risk is medium.

Rule 19: IF highly likely AND large THEN project risk is high.

Rule 20: IF highly likely AND catastrophic THEN project risk is high.

Rule 21: IF near certainty AND no impact THEN project risk is low.

Rule 22: IF near certainty AND small THEN project risk is medium.

Rule 23: IF near certainty AND medium THEN project risk is high.

Rule 24: IF near certainty AND large THEN project risk is high.

Rule 25: IF near certainty AND catastrophic THEN project risk is high.

Schedule rules for likelihood and consequence of underestimation

Rule 26: IF remote AND no impact THEN project risk is low.

Rule 27: IF remote AND small THEN project risk is low.

Rule 28: IF remote AND medium THEN project risk is low.

Rule 29: IF remote AND large THEN project risk is low.

Rule 30: IF remote AND catastrophic THEN project risk is medium.

Rule 31: IF unlikely AND no impact THEN project risk is low.

Rule 32: IF unlikely AND small THEN project risk is low.

Rule 33: IF unlikely AND medium THEN project risk is low.

Rule 34: IF unlikely AND large THEN project risk is medium.

Rule 35: IF unlikely AND catastrophic THEN project risk is medium.

Rule 36: IF likely AND no impact THEN project risk is low.

Rule 37: IF likely AND small THEN project risk is medium.

Rule 38: IF likely AND medium THEN project risk is medium.

Rule 39: IF likely AND large THEN project risk is medium.

Rule 40: IF likely AND catastrophic THEN project risk is high.

Rule 41: IF highly likely AND no impact THEN project risk is low.

Rule 42: IF highly likely AND small THEN project risk is medium.

Rule 43: IF highly likely AND medium THEN project risk is medium.

Rule 44: IF highly likely AND large THEN project risk is high.

Rule 45: IF highly likely AND catastrophic THEN project risk is high.

Rule 46: IF near certainty AND no impact THEN project risk is low.

Rule 47: IF near certainty AND small THEN project risk is medium.

Rule 48: IF near certainty AND medium THEN project risk is high.

Rule 49: IF near certainty AND large THEN project risk is high.

Rule 50: IF near certainty AND catastrophic THEN project risk is high.

Performance rules for likelihood and consequence of underestimation

Rule 51: IF remote AND no impact THEN project risk is low.

Rule 52: IF remote AND small THEN project risk is low.

Rule 53: IF remote AND medium THEN project risk is low.

Rule 54: IF remote AND large THEN project risk is low.

Rule 55: IF remote AND catastrophic THEN project risk is medium.

Rule 56: IF unlikely AND no impact THEN project risk is low.

Rule 57: IF unlikely AND small THEN project risk is low.

Rule 58: IF unlikely AND medium THEN project risk is low.

Rule 59: IF unlikely AND large THEN project risk is medium.

Rule 60: IF unlikely AND catastrophic THEN project risk is medium.

Rule 61: IF likely AND no impact THEN project risk is low.

Rule 62: IF likely AND small THEN project risk is medium.

Rule 63: IF likely AND medium THEN project risk is medium.

Rule 64: IF likely AND large THEN project risk is medium.

Rule 65: IF likely AND catastrophic THEN project risk is high.

Rule 66: IF highly likely AND no impact THEN project risk is low.

Rule 67: IF highly likely AND small THEN project risk is medium.

Rule 68: IF highly likely AND medium THEN project risk is medium.

Rule 69: IF highly likely AND large THEN project risk is high.

Rule 70: IF highly likely AND catastrophic THEN project risk is high.

Rule 71: IF near certainty AND no impact THEN project risk is low.

Rule 72: IF near certainty AND small THEN project risk is medium.

Rule 73: IF near certainty AND medium THEN project risk is high.

Rule 74: IF near certainty AND large THEN project risk is high.

Rule 75: IF near certainty AND catastrophic THEN project risk is high.

From the example radar system development project, a cost parameter will be used to demonstrate how the fuzzy model can be used to logically calculate project risks for a single parameter. The following parameter has been chosen to logically demonstrate the fuzzy model.

Cost: Antenna cost (from Table 3) = $160,000

Subject-matter expert inputs are needed for the fuzzy model. The inputs based on the subject-matter expert's assessment of the likelihood of the estimate being understated and the associated consequence are:

Cost: Likelihood (Table 5) = 40% likely underestimated

Cost: Consequence (Table 6) = 27% more funding needed

The likelihood that the cost is underestimated can be determined from the likelihood membership functions shown in Figure 12. The degree of membership that the input (A = 40) is contained in each of the fuzzy sets is 0 with the exception of the highly likely set, where it has a membership value of 1. The consequence of the cost being underestimated is determined from the subject-matter expert's input (B = 27). Using the cost membership functions shown in Figure 13 the degree of membership is approximately 0.25 in the likely set, 0.6 in the catastrophic set, and 0 in all the other sets. Using the max and min fuzzy operators, C = max(1, min(0.25,1) = 1.0. From the previously developed fuzzy rules: *If cost is "highly likely" understated AND consequence is "catastrophic" THEN project risk is "high"*

Each of the project parameters for cost, schedule, and performance can be calculated using a fuzzy model similar to the one shown in Figure 11. The model provides the logical framework from which a generalized mathematical model was developed.

CHAPTER 8

METHODOLOGY

8.1 Methodology Goals

The research methodology contained in this chapter is a formalization of the project risk conceptual model presented in Figure 11, Chapter 7. The risk model includes both "likelihood" of occurrence and "consequence". Figure 16 shows the generic risk rating classification levels based on "likelihood" and "consequence". The risk associated with each project estimate (cost, schedule, or performance) is categorized as low (L), medium (M), or high (H). The actual risk ratings are shown in Table 9 in Chapter 7.

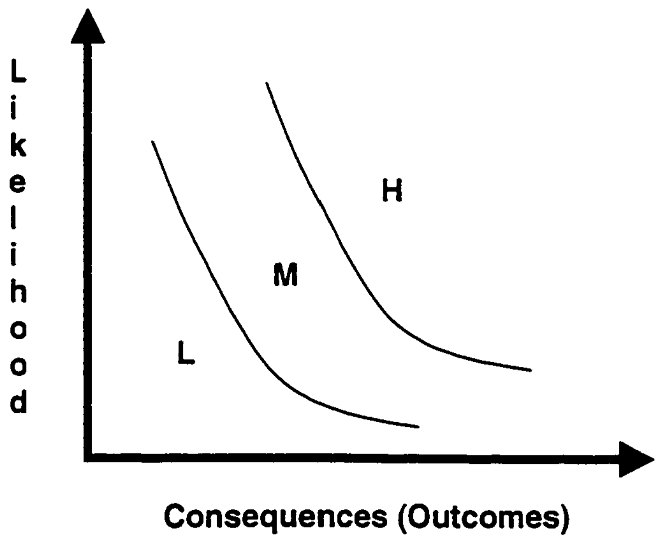


Figure 16. Risk Classification

Identifying the sources and nature of risks and the uncertainty associated with project activities is considered the first step in the risk assessment process. Risk identification requires a complete description of the risk events that might occur and attempts to answer the questions: "What can go wrong?" and "What are the consequences?". The input-output relationships with respect to the project variables, and the effect of a myriad of consequences constitutes the heart of the risk identification model.

The classification of risk is based on both risk "likelihood" and "consequences". The transition between low, medium, and high risk is gradual and the boundaries are generally not crisply defined for an actual project. Fuzzy sets offer a smooth and seamless transition between different levels of information granularity or levels of specificity. Increasing the number of fuzzy sets for the system's variables provides greater granularity. Increased granularity moves closer to numerical models. Decreased granularity moves closer to qualitative models. The fuzzy model uses collected data and facts from subject-matter experts as the basis for fuzzy set formulation.

The fuzzy logic model is an aggregate of the linguistic labels for risk "likelihood" and "consequence". The crux of the model is that the essential relationships between system variables are described in terms of fuzzy sets rather than numerical quantities.

A trapezoidal fuzzy set (TFS) can be represented by a quadruple (a, b, c, d), where a and d are the lower and upper bounds, b and c are the lower and upper modal values, respectively, and x is an element between a and d. The generic membership function for the TFS can be expressed as:

$$A(x) = \begin{cases} 1 & \text{if } x < a \text{ or } x > d \\ \dfrac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } b \leq x \leq c \\ \dfrac{x-d}{c-d} & \text{if } c < x \leq d \end{cases}$$

where a < b < c < d.

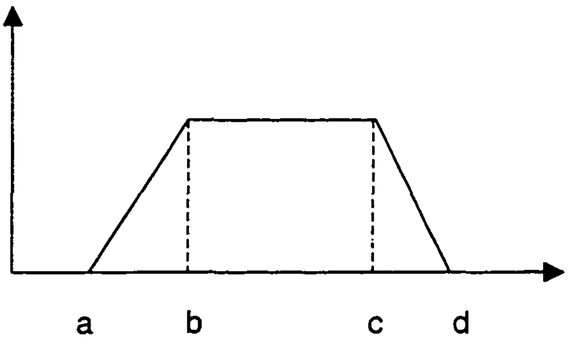


Figure 17. Trapezoidal Values

The magnitudes of the left spread, the middle plateau, and the right spread signify, collectively, the degrees of uncertainties associated with each TFS.

There are a number of operations that can be performed on TFSs. The risk model was based on the following operations:

If $M = (a_1, b_1, c_1, d_1)$ and $N = (a_2, b_2, c_2, d_2)$

$M \subseteq N$ iff $A_M(x) \leq A_N(x)$

$M \oplus N = (a_1 + a_2, b_1 + b_2, c_1 + c_2, d_1 + d_2)$

$M \Theta N = (a_1 - a_2, b_1 - b_2, c_1 - c_2, d_1 - d_2)$

$\max(M, N) = [\vee(a_1, a_2), \vee(b_1, b_2), \vee(c_1, c_2), \vee(d_1, d_2)]$

$\min(M, N) = [\wedge(a_1, a_2), \wedge(b_1, b_2), \wedge(c_1, c_2), \wedge(d_1, d_2)]$

$M \cap N = \{x, A_M(x) \wedge A_N(x)]\}$

where $\oplus$ = fuzzy addition; Θ = fuzzy subtraction;

$\vee$ = maximum; and $\wedge$ = minimum.

To develop the model, the following steps were required:

1. Define inputs to the model in terms of likelihood that estimates for cost, schedule, or performance are understated and the resulting consequence of the estimate.
2. Fuzzify the inputs by fetching their membership values from the fuzzy sets for the preconditions of the decision rules presented in Chapter 7.
3. Determine the firing strength of each fuzzy rule based on the fetched membership values.
4. Aggregate the firing strengths of all rules.
5. Defuzzify and present the resulting risk.

8.2 Model Inputs/Outputs

The collected data, facts, and expert judgements about the project are transformed from a numerical estimate into the framework formed by fuzzy sets. This phase required the development of an input interface. The fuzzy sets (linguistic labels) were used to build the input interface and played a crucial role in anchoring all the available pieces of information into a processing context. The input for the likelihood fuzzy sets is an expert judgement percentage (x) that the estimate (cost, schedule, or performance) is underestimated. The input for the consequence fuzzy sets for cost, schedule, or performance estimates is also a percentage (y).

Due to overlapping of the membership functions the crisp input may have membership in more than one fuzzy set. The rule base with preconditions that match the inputs to a non-zero membership value can also have more than one rule that meets the conditions.

The outputs from the model are the risk rating level crisp value (z), which gives the firing strength of the fuzzy rule. This value gives the subject-matter expert the risk level associated with the two input percentages for likelihood and consequences.

8.3 Membership Sets

As presented in Chapter 7, risk likelihood is represented by the five fuzzy sets remote, unlikely, likely, highly likely, and near certainty using the trapezoidal distribution. The risks associated with cost, schedule, and performance estimates have possible membership in the five fuzzy sets. Figure 18 shows the five likelihood membership sets. Figures 19, 20, and 21 shows the membership sets for cost, schedule, and performance, respectively.

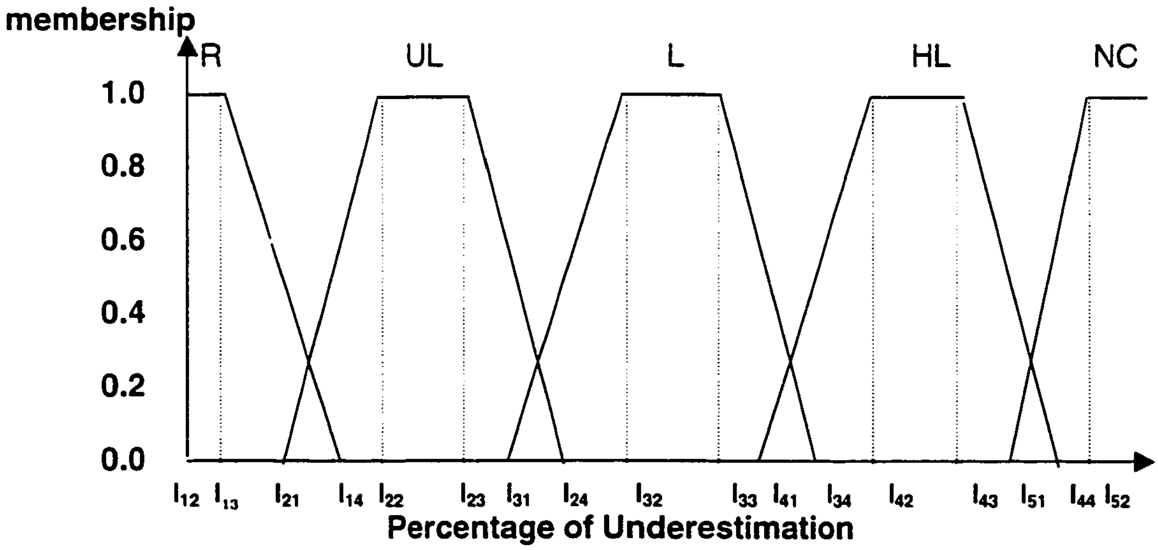


Figure 18. Likelihood of Occurrence Membership Functions

In this research, likelihood of occurrence linguistic mappings were defined as follows:

Remote (R) – $l_{11}, l_{12}, l_{13}, l_{14}$

Unlikely (UL) – $l_{21}, l_{22}, l_{23}, l_{24}$

Likely (L) – $l_{31}, l_{32}, l_{33}, l_{34}$

Highly Likely (HL) – $l_{41}, l_{42}, l_{43}, l_{44}$

Near Certainty (NC) – $l_{51}. l_{52}. l_{53}. l_{54}$

Remote (R)

For $x > l_{14}$; $f(x) = 0$

For $l_{12} \leq x \leq l_{13}$; $f(x) = 1$

For $l_{13} < x \leq l_{14}$; $f(x) = (x - l_{14}) / (l_{14} - l_{13})$

Unlikely (UL)

For $x < l_{21}$ or $x > l_{24}$; $f(x) = 0$

For $l_{21} \leq x < l_{22}$; $f(x) = (x - l_{21}) / (l_{22} - l_{21})$

For $l_{22} \leq x \leq l_{23}$; $f(x) = 1$

For $l_{23} < x \leq l_{24}$; $f(x) = (x - l_{24}) / (l_{23} - l_{24})$

Likely (L)

For $x < l_{31}$ or $x > l_{34}$; $f(x) = 0$

For $l_{31} \leq x < l_{32}$; $f(x) = (x - l_{31}) / (l_{32} - l_{31})$

For $l_{32} \leq x \leq l_{33}$; $f(x) = 1$

For $l_{33} < x \leq l_{34}$; $f(x) = (x - l_{34}) / (l_{33} - l_{34})$

Highly Likely (HL)

For $x < l_{41}$ or $x > l_{44}$; $f(x) = 0$

For $l_{41} \leq x < l_{42}$; $f(x) = (x - l_{41}) / (l_{42} - l_{41})$

For $l_{42} \leq x \leq l_{43}$; $f(x) = 1$

For $l_{43} < x \leq l_{44}$; $f(x) = (x - l_{44}) / (l_{43} - l_{44})$

Near Certainty (NC)

For $x < l_{51}$ or $x > l_{54}$; $f(x) = 0$

For $l_{51} \leq x < l_{52}$; $f(x) = (x - l_{51}) / (l_{52} - l_{51})$

For $l_{52} \leq x \leq l_{53}$; $f(x) = 1$

For $l_{53} < x \leq l_{54}$; $f(x) = (x - l_{54}) / (l_{53} - l_{54})$

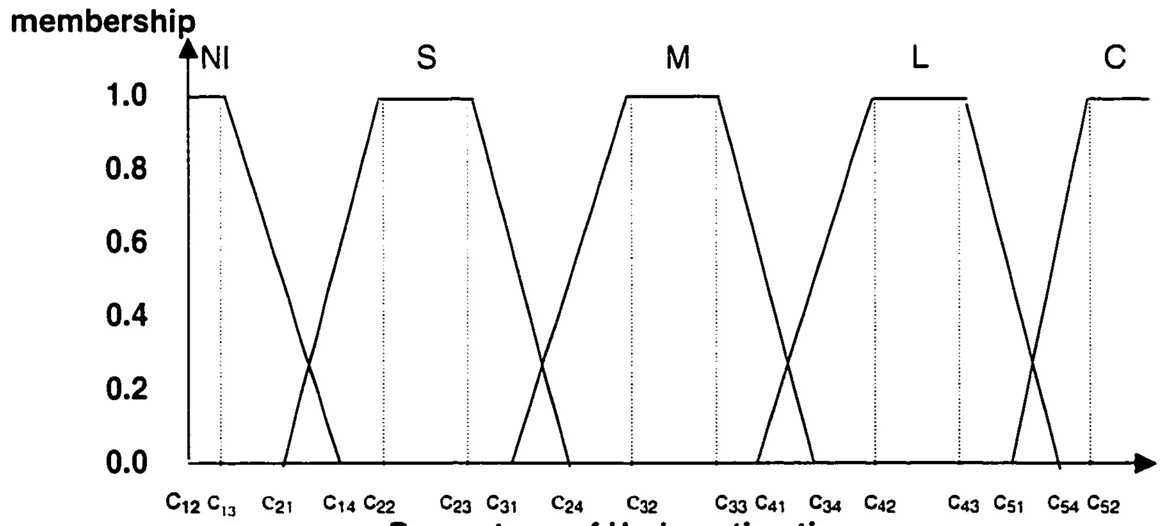


Figure 19. Cost Consequence Membership Functions

The cost consequence linguistic mappings for this research were defined as follows:

No Impact (NI) – c_{11}, c_{12}, c_{13}, c_{14}

Small (S) – c_{21}, c_{22}, c_{23}, c_{24}

Medium (M) – c_{31}, c_{32}, c_{33}, c_{34}

Large (L) – c_{41}, c_{42}, c_{43}, c_{44}

Catastrophic (C) – c_{51}, c_{52}, c_{53}, c_{54}

No Impact (NI)

For $y > c_{14}$; $f(y) = 0$

For $c_{12} \le y \le c_{13}$; $f(y) = 1$

For $c_{13} < y \le c_{14}$; $f(y) = (y - c_{14}) / (c_{14} - c_{13})$

Small (S)

For $y < c_{21}$ or $y > c_{24}$; $f(y) = 0$

For $c_{21} \le y < c_{22}$; $f(y) = (y - c_{21}) / (c_{22} - c_{21})$

For $c_{22} \le y \le c_{23}$; $f(y) = 1$

For $c_{23} < y \le c_{24}$; $f(y) = (y - c_{24}) / (c_{23} - c_{24})$

Medium (M)

For $y < c_{31}$ or $y > c_{34}$; $f(y) = 0$

For $c_{31} \le y < c_{32}$; $f(y) = (y - c_{31}) / (c_{32} - c_{31})$

For $c_{32} \le y \le c_{33}$; $f(y) = 1$

For $c_{33} < y \le c_{34}$; $f(y) = (y - c_{34}) / (c_{33} - c_{34})$

Large (L)

For $y < c_{41}$ or $y > c_{44}$; $f(y) = 0$

For $c_{41} \le y < c_{42}$; $f(y) = (y - c_{41}) / (c_{42} - c_{41})$

For $c_{42} \le y \le c_{43}$; $f(y) = 1$

For $c_{43} < y \le c_{44}$; $f(y) = (y - c_{44}) / (c_{43} - c_{44})$

Significant (S)

For $y < c_{51}$ or $y > c_{54}$; $f(y) = 0$

For $c_{51} \le y < c_{52}$; $f(y) = (y - c_{51}) / (c_{52} - c_{51})$

For $c_{52} \le y \le c_{53}$; $f(y) = 1$

For $c_{53} < y \le c_{54}$; $f(y) = (y - c_{54}) / (c_{53} - c_{54})$

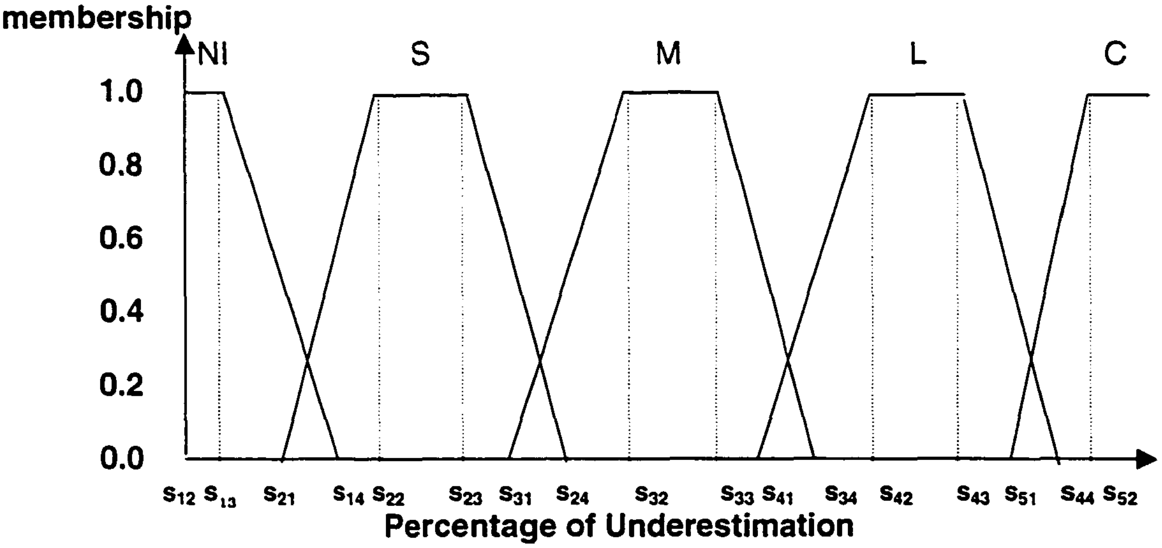


Figure 20. Schedule Consequence Membership Functions

The schedule consequence linguistic mappings for this research were defined as follows:

No Impact (NI) – s_{11}, s_{12}, s_{13}, s_{14}

Small (S) – s_{21}, s_{22}, s_{23}, s_{24}

Medium (M) – s_{31}, s_{32}, s_{33}, s_{34}

Large (L) – s_{41}, s_{42}, s_{43}, s_{44}

Catastrophic (C) – s_{51}, s_{52}, s_{53}, s_{54}

No Impact (NI)

For $y > s_{14}$; $f(y) = 0$

For $s_{12} \le y \le s_{13}$; $f(y) = 1$

For $s_{13} < y \le s_{14}$; $f(y) = (y - s_{14}) / (s_{14} - s_{13})$

Small (S)

For $y < s_{21}$ or $y > s_{24}$; $f(y) = 0$

For $s_{21} \le y < s_{22}$; $f(y) = (y - s_{21}) / (s_{22} - s_{21})$

For $s_{22} \le y \le s_{23}$; $f(y) = 1$

For $s_{23} < y \le s_{24}$; $f(y) = (y - s_{24}) / (s_{23} - s_{24})$

Medium (M)

For $y < s_{31}$ or $y > s_{34}$; $f(y) = 0$

For $s_{31} \le y < s_{32}$; $f(y) = (y - s_{31}) / (s_{32} - s_{31})$

For $s_{32} \le y \le s_{33}$; $f(y) = 1$

For $s_{33} < y \le s_{34}$; $f(y) = (y - s_{34}) / (s_{33} - s_{34})$

Large (L)

For $y < s_{41}$ or $y > s_{44}$; $f(y) = 0$

For $s_{41} \le y < s_{42}$; $f(y) = (y - s_{41}) / (s_{42} - s_{41})$

For $s_{42} \leq y \leq s_{43}$; $f(y) = 1$

For $s_{43} < y \leq s_{44}$; $f(y) = (y - s_{44}) / (s_{43} - s_{44})$

Catastrophic (C)

For $y < s_{51}$ or $y > s_{54}$; $f(y) = 0$

For $s_{51} \leq y < s_{52}$; $f(y) = (y - s_{51}) / (s_{52} - s_{51})$

For $s_{52} \leq y \leq s_{53}$; $f(y) = 1$

For $s_{53} < y \leq s_{54}$; $f(y) = (y - s_{54}) / (s_{53} - s_{54})$

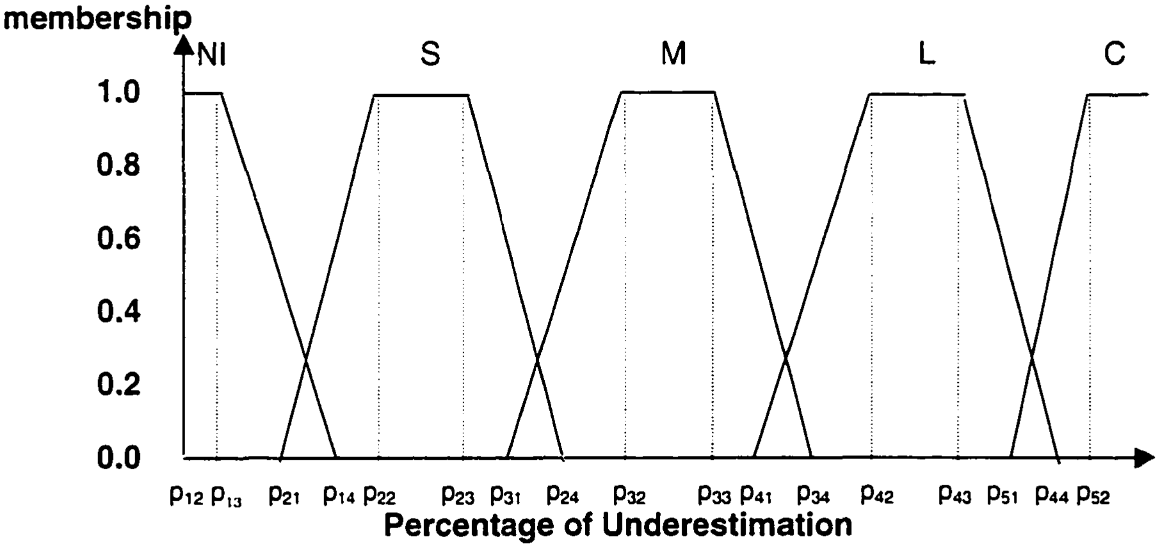


Figure 21. Performance Consequence Membership Functions

The performance linguistic mappings for this research were defined as follows:

No Impact (NI) – p_{11}, p_{12}, p_{13}, p_{14}

Small (S) – p_{21}, p_{22}, p_{23}, p_{24}

Medium (M) – $p_{31}, p_{32}, p_{33}, p_{34}$

Large (L) – $p_{41}, p_{42}, p_{43}, p_{44}$

Catastrophic (C) – $p_{51}, p_{52}, p_{53}, p_{54}$

No Impact (NI)

For $y > p_{14}$; $f(y) = 0$

For $p_{12} \le y \le p_{13}$; $f(y) = 1$

For $p_{13} < y \le p_{14}$; $f(y) = (y - p_{14}) / (p_{14} - p_{13})$

Small (S)

For $y < p_{21}$ or $y > p_{24}$; $f(y) = 0$

For $p_{21} \le y < p_{22}$; $f(y) = (y - p_{21}) / (p_{22} - p_{21})$

For $p_{22} \le y \le p_{23}$; $f(y) = 1$

For $p_{23} < y \le p_{24}$; $f(y) = (y - p_{24}) / (p_{23} - p_{24})$

Medium (M)

For $y < p_{31}$ or $y > p_{34}$; $f(y) = 0$

For $p_{31} \le y < p_{32}$; $f(y) = (y - p_{31}) / (p_{32} - p_{31})$

For $p_{32} \le y \le p_{33}$; $f(y) = 1$

For $p_{33} < y \le p_{34}$; $f(y) = (y - p_{34}) / (p_{33} - p_{34})$

Large (L)

For $y < p_{41}$ or $y > p_{44}$; $f(y) = 0$

For $p_{41} \le y < p_{42}$; $f(y) = (y - p_{41}) / (p_{42} - p_{41})$

For $p_{42} \le y \le p_{43}$; $f(y) = 1$

For $p_{43} < y \le p_{44}$; $f(y) = (y - p_{44}) / (p_{43} - p_{44})$

Catastrophic (C)

For $y < p_{51}$ or $y > p_{54}$; $f(y) = 0$

For $p_{51} \le y < p_{52}$; $f(y) = (y - p_{51}) / (p_{52} - p_{51})$

For $p_{52} \le y \le p_{53}$; $f(y) = 1$

For $p_{53} < y \le p_{54}$; $f(y) = (y - p_{54}) / (p_{53} - p_{54})$

Fuzzy inference requires the association of a fuzzy membership function with each implication. For the rule "IF A THEN B", a possible membership function is:

$$\mu_{A\to B}(u, v) = \min[\mu_A(u), \mu_B(v)]$$

where u and v span the universes of discourse corresponding to the terms A and B. It is important to note that $\mu_{A\to B}(u, v)$ is a two-dimensional set.

Zadeh (1983) showed that fuzzy evidence can be propagated through a rule by using a compositional rule of inference (CRI). The goal of CRI is to infer a new membership function for the term B in the consequent of the rule. If the available data is A_{data} the CRI commonly found in the literature is:

$$\mu_B(v) = \sup\{\min[\mu_{Adata}(u), \mu_{A\to B}(u, v)]\}$$

where u spans the universe of discourse for the term A and v is the same for the term B. In practice a faster way to implement CRI is to use the following:

$$\sup \min[\mu_{Adata}, \mu_A]$$

and use the number that was obtained to truncate the membership function for the consequent term B. The number obtained from the sup-min operation above is the degree-of-match between A_{data} and A.

To achieve the goals of the risk model, fuzzy evidence must be propagated through multiple rules simultaneously and the results aggregated. There are two important questions that need to be resolved to develop the fuzzy risk identification model.

1) Is the order in which the rules are invoked important?
2) How can conclusions be drawn from all the rules that are fired from a given set of facts?

Lee (1990) developed three Lemmas to answer the questions of rule order and how conclusions can be drawn from multiple rules.

Lemma 1: If A can match the antecedents of rules R_1, R_2, ..., R_n, each rule being of the form $A_i \rightarrow B_i$, the overall conclusion from the rules satisfies the following distributive relationship:

$$B = A \bigcup_{i=1}^{n} R_i = \bigcup_{i=1}^{n} A \; R_i.$$

A set of rules that can be fired by a given fact is represented by the union operator. B is the fuzzy set obtained by aggregating the appropriately modified conclusion fuzzy sets B_1, B_2, ..., B_n. The representation of a set of rules by a union operator is justified because the different rules represent alternative ways of asserting the conclusion. This lemma provides the framework to propagate a fuzzy fact through two different rules. The propagation should be performed separately and then take the fuzzy union of the conclusion fuzzy sets. This lemma is limited to multiple rules with only one term in each rule antecedent.

Lemma 2: Given two facts A and B corresponding to two different linguistic variables that can match the two facts in the antecedent of a single rule of the form "IF A_i AND B_i THEN C_i", the conclusion fuzzy set for C_i can be obtained by the following formula:

$$C_i = [A \ \ (A_i \rightarrow C_i)] \cap [B \ \ (B_i \rightarrow C_i)].$$

Lemma 3: Given two facts A and B corresponding to different linguistic variables that can match the two facts in each of rules R_1, R_2, ..., R_n, each rule being of the form "IF A_i AND B_i THEN C_i", an overall conclusion can be drawn from the rules due to the following distributive relationship:

$$C = \bigcup_{i=1}^{n} [A \ (A_i \rightarrow C_i)] \cap [B \ \ (B_i \rightarrow C_i)].$$

This lemma provides a basis for a final conclusion that is made by propagating the fuzzy evidence through each rule separately and then taking a fuzzy union of the resulting conclusions.

8.4 Fuzzy Decision Rules and Firing Strengths

Project risks associated with cost, schedule, and performance are determined by mathematical operations on the formulated fuzzy sets. A fuzzy decision rule has certain preconditions, which are to be matched with given facts, as well as certain consequences that result when the preconditions are met.

Each precondition or consequence in a rule includes an instance of the fuzzy variable involved. The fuzzy rules defined in Chapter 7 have the general form: IF precondition-1 AND precondition-2 THEN consequence-1.

The parameter mapping framework for the fuzzy sets was presented earlier in this chapter. For the "likelihood" and "consequence" fuzzy sets the firing strength of rule i determines how much the fuzzy set will contribute to the risk event. Figure 22 shows the basic components of multi-rule firings. The likelihood (l), consequence (m), and the firing strength (f) for the rule are the preconditions and consequence of the fuzzy rule. The composite (F) is the max of the rules that have a positive firing strength.

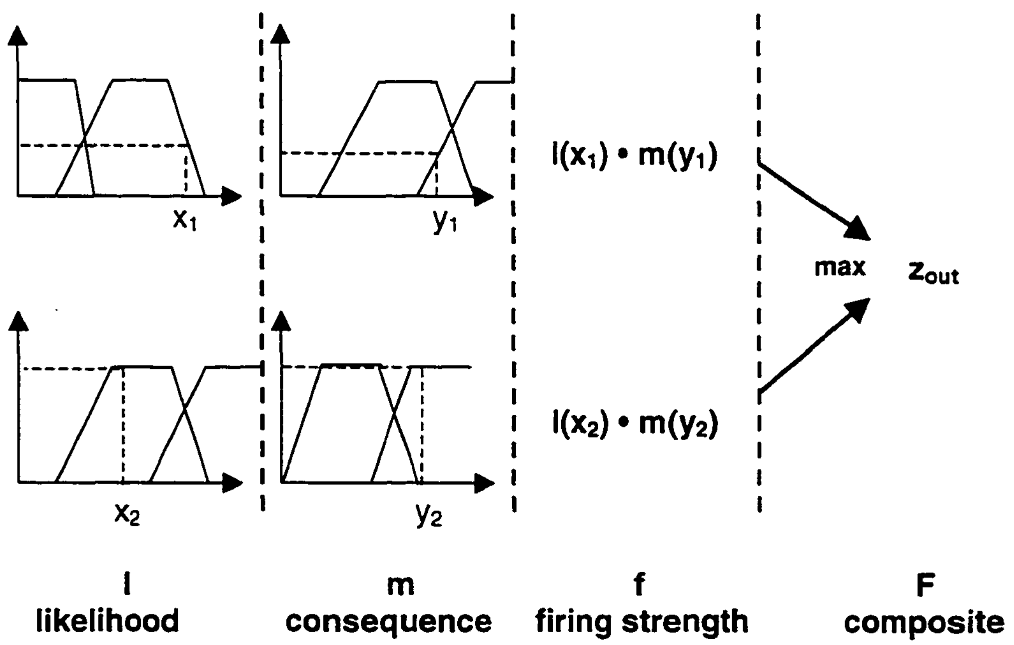


Figure 22. Composite Membership Functions

After the likelihood (x) and consequence (y) inputs are determined by the subject-matter expert, the next step is to determine the firing strength of the fuzzy rules. The firing strength is determined by the fetched membership values from the fuzzy sets based on the input percentages (x and y). From the input value x, the fetched membership value is determined using the trapezoidal likelihood equations presented earlier. The fetched membership value is l_c for the likelihood cost function. Similarly, the fetched membership values are l_s and l_p for schedule and performance, respectively. Risk likelihood (l) is determined by the following equations:

$$l = (x - a) / (b - a), \quad \text{for } a \le x < b$$

$$l = 1, \quad \text{for } b \le x \le c$$

$$l = (x - d) / (c - d), \quad \text{for } c < x \le d$$

where a and d are the lower and upper bounds; and b and c are the lower and upper modal values of the trapezoid.

Input (y) to the risk consequence membership functions is used to determine the fetched membership values of m_c, m_s, and m_p for cost, schedule, and performance, respectively. Risk consequences (m) are determined from the input value (y) using the following equations:

$$m = (y - a) / (b - a), \quad \text{for } a \le y < b$$

$$m = 1, \quad \text{for } b \le y \le c$$

$$m = (y - d) / (c - d), \quad \text{for } c < y \le d$$

where a and d are the lower and upper bounds; and b and c are the lower and upper modal values of the trapezoid.

Since the two preconditions of a rule are connected by the AND relation, the product of multiplication of the corresponding memberships is used to represent the firing strengths shown below:

$f_c = l_c \bullet m_c$, for cost likelihood and consequence

$f_s = l_s \bullet m_s$, for schedule likelihood and consequence

$f_p = l_p \bullet m_p$, for performance likelihood and consequence

where l_c, l_s, and l_p are the fetched likelihood membership values for cost, schedule, and performance, respectively; m_c, m_s, and m_p are the fetched consequence membership values for cost, schedule, and performance, respectively. f_c, f_s, and f_p are the firing strengths of the likelihood and consequence rules; $0 \leq f \leq 1$.

8.5 Aggregated Rule Consequences

The basic structure of fuzzy logic sets makes overlap of the rules inevitable. A crisp conclusion is required rather than a fuzzy set. To arrive at a definite conclusion, defuzzification is required for the aggregate consequence. The union operation is used to aggregate the overall consequence of executing all n rules. As shown in Figure 22 the union of the firing strengths (f) form the composite function. The fuzzy logic max operation is shown below:

$F_i = \max(f_1, f_2, \ldots, f_n)$

where F_i is the maximum firing strength of n rules, $0 \leq F_i \leq 1$.

F_i represents a single risk output for a cost, schedule, or performance estimate. To aggregate total project risk for all the project estimates the project state for each estimate needs to be maintained. The risk state of the project is represented by a two dimensional vector that contains the risk level and firing strength for each of the project estimates. The risk state vector is:

$$R = \{(F_1, K_1), (F_2, K_2), \ldots (F_n, K_n)\}$$

where F_i is the identified risk firing strength; and K_i is the risk rating for estimate i. The values contained in the risk state vector R represent the composite of risks for all project estimates.

Specifically, for cost, schedule, and performance project estimates, a state vector is needed for each. Three state vectors contain the information from which total project risk quantification can be obtained. The state vectors for cost, schedule, and performance are:

$$R_c = \{(F_{c1}, K_{c1}), (F_{c2}, K_{c2}), \ldots (F_{cn}, K_{cn})\}$$

$$R_s = \{(F_{s1}, K_{s1}), (F_{s2}, K_{s2}), \ldots (F_{sn}, K_{sn})\}$$

$$R_p = \{(F_{p1}, K_{p1}), (F_{p2}, K_{p2}), \ldots (F_{pn}, K_{pn})\}$$

The risk rating (K) has a value of L (low), M (medium), or H (high). The quantitative value (F) is a value in the range $0 \leq F \leq 1$. The values of K and F provide the subject-matter expert with a risk rating and a quantitative measure of the inputs for likelihood and consequence. Overall project risks for cost, schedule, or performance are assumed at the highest risk rating level.

CHAPTER 9

MODEL IMPLEMENTATION

9.1 Risk Model Flowchart

Implementation of the risk model was based on the mathematical framework presented in Chapter 8. A flowchart of the model is shown in Figure 23. Risk likelihood (x) and consequence (y) are inputs to the model. Both inputs are percentages. The firing strength of the likelihood (l) is determined based on the input x. Consequence firing strength is determined from the input y. The overall rule fuzzy strength (f) is determined by the product of the likelihood (l) and the consequence (m). The aggregate (F) of all the fuzzy rules is determined by the max fuzzy operator. Finally the crisp output z_{out} is determined from the max operation.

Each iteration of the model produces firing strengths for likelihood (l), consequence (m), and a defuzzied crisp output (z_{out}). Likelihood (l) and consequence (m) form the basis to determine the fuzzy rules that are "activated" based on the inputs (x and y).

Twenty-five rules were required for the combinations of 5 likelihood and 5 consequence fuzzy membership sets for cost, schedule, or performance. A total of 75 fuzzy rules were needed to represent the likelihood and consequences of cost, schedule, and performance.

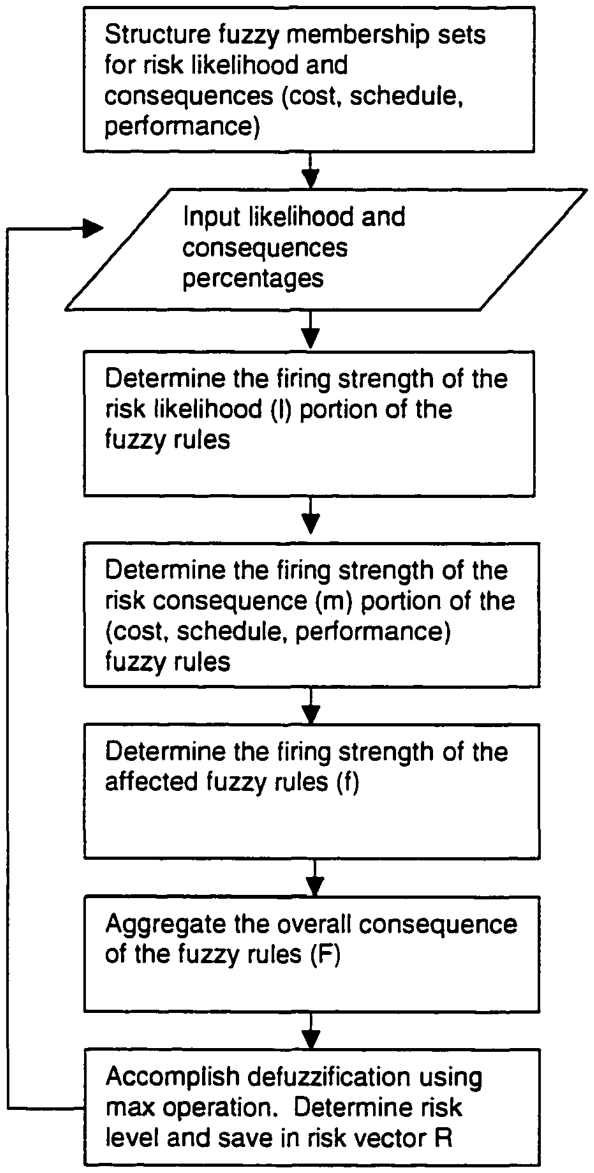


Figure 23. Risk Model Flowchart

9.2 MATLAB Program

The risk model was implemented using MATLAB. The MATLAB program that implemented the model is shown below.

```
%  Fuzzy Risk Quantification Program
%  This program was developed to quantify likelihood and
consequence
%
% Likelihood
L = [0 0 3 5; 3 7 11 15; 10 17 26 35; 30 38 46 55; 50 60
100 100];
% cost
C = [0 0 3 5; 3 5 7 9; 7 10 12 15; 10 17 23 30; 25 100 100
100];
% Rules strengths
R = [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];
RRC = [0 0 0 0 1 0 0 0 1 1 0 1 1 1 2 0 1 1 2 2 0 1 2 2 2];
Rcf = [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];
% initialize cost risk vector
Rcl = blanks(25); % initialize cost risk vector
n = 1; % initialize output counter
%
%  Get # of estimates to evaluate
%
nest = input('Enter the number of input estimates >  '); %
# estimates
%
% Get inputs for the program
%
while n < nest+1
X = input('Enter Likelihood Percentage >  ');  % input
likelihood percentage
Y = input('Enter Cost Percentage >  ');  % input cost
percentage
FL = [0 0 0 0 0];  % firings matrix for likelihood
FC = [0 0 0 0 0];  % firings matrix for cost
FS = [0 0 0 0 0];  % firings matrix for schedule
FP = [0 0 0 0 0];  % firings matrix for performance
Fmax = 0;  % fuzzy maximum of aggregated rules
Indx = 0;  % fuzzy maximum rule index
```

```
Zout = 0;  % crisp output value
%
% Likelihood Firings
%
for i = 1:5
  if X > L(i,1) & X < L(i,2)
    FL(i) = (X - L(i,1)) / (L(i,2) - L(i,1));
   end
   if X >= L(i,2) & X <= L(i,3);
    FL(i) = 1;
   end
   if X > L(i,3) & X < L(i,4);
    FL(i) = (X - L(i,4)) / (L(i,3) - L(i,4));
   end
end
%
% Cost Firings
%
for i = 1:5
  if Y > C(i,1) & Y < C(i,2)
    FC(i) = (Y - C(i,1)) / (C(i,2) - C(i,1));
   end
   if Y >= C(i,2) & Y <= C(i,3);
    FC(i) = 1;
   end
   if Y > C(i,3) & Y < C(i,4);
    FC(i) = (Y - C(i,4)) / (C(i,3) - C(i,4));
   end
end
%
%  Determine the firing strength of the rules based on the
strength of the risk
%    likelihood and the cost consequence.
%
k = 0;
for i = 1:5
   for j = 1:5
    k = k + 1;
    R(k)= FL(i) * FC(j);
  end
end
%
% Determine the aggregated rule consequences
%
for i = 1:25
```

```
   if Fmax < R(i)
      Fmax = R(i);
      Indx = i;
   end
end
%
% Determine crisp output value
%
Zout = Fmax;

%
% Determine risk output from rules
%
i = Indx;
if R(i) > 0 & RRC(i) == 0
   rule = i;
   risk = 'L';
end
if R(i) > 0 & RRC(i) == 1
   rule = i;
   risk = 'M';
end
if R(i) > 0 & RRC(i) == 2
   rule = i;
   risk = 'H';
end
%
% Save risk level and firing strength
%
Rcl(n) = risk;
Rcf(n) = Zout;
n = n + 1;
end
%
% Print Report
%
disp(Rcl)
disp(Rcf)
```

9.3 Risk Model Input

Data contained in the example presented in Chapter 7 were used as the inputs to the risk model. The trapezoidal fuzzy membership values for cost likelihood were presented previously in Table 6. Cost consequence membership values were taken from Table 7.

For the radar system development example, 19 cost estimates were identified. A likelihood input (x) and consequence input (y) were required for each estimate. As shown by data in Table 10, software development and integration and test are historically higher risk activities. Although the data are for a radar system development project, the method is applicable for other applications where cost, schedule, or performance estimates are available.

Component	Estimate	Likelihood X	Consequence Y
System design			
Antenna	160,000	15	22
Transmitter	352,000	10	25
Power supply	79,000	3	5
Control data processor	400,000	5	15
Receiver	352,000	8	15
Signal processor	245,000	6	10
Displays	160,000	5	10
Software development			
Transmitter	270,000	25	30
Control data processor	536,000	33	40
Receiver	300,000	25	35
Signal processor	700,000	30	35
Hardware Procurement			
Circuit boards	250,000	2	5
Chassis	170,000	2	7
Power supplies	100,000	3	5
Signal processor	910,000	5	10
Antennas	825,000	6	15
Displays	200,000	2	5
Integration and test			
Hardware and software	1,710,000	10	25
Production			
Fabricate 100 units	3,000,000	5	10

Table 10. Model Input Cost Data

Inputs for schedule likelihood (x) and consequence (y) for the radar example are shown in Table 11. Although a high level schedule was presented for the example consisting of only 5 milestones, the model can handle greater fidelity.

Activity	Timeline	Likelihood X	Consequence Y
Design	145 days	5	10
Software	166 days	25	45
Hardware	90 days	5	7
Integration	120 days	10	14
Production	300 days	7	15

Table 11. Model Input Schedule Data

Project performance inputs for likelihood (x) and consequence (y) are shown in Table 12.

Parameter	Attribute	Likelihood X	Consequence Y
Transmitter range	35 nm	5	10
Receiver range	35 nm	5	10
Identification time	3 Sec.	2	7
Location accuracy	10 ft.	5	15
Weight	<45 lbs.	5	5
Size	24X12X36	5	5
MTBF	3000 hrs	10	15

Table 12. Model Input Performance Data

9.4 Risk Model Output

Model outputs for a likelihood input value (x) of 15% and a cost consequence (y) of 22% for the system design of the radar antenna are shown in Table 13.

Likelihood Input (x)	Consequence Input (y)	Rule	l(x)	m(y)	f	F
15	22	1	0	0	0	
		2	0	0	0	
		3	0	0	0	
		4	0	1	0	
		5	0	0	0	
		6	0	0	0	
		7	0	0	0	
		8	0	0	0	
		9	0	1	0	
		10	0	0	0	
		11	0.7143	0	0	
		12	0.7143	0	0	
		13	0.7143	0	0	
		14	0.7143	1	0.7143	0.7143
		15	0.7143	0	0	
		16	0	0	0	
		17	0	0	0	
		18	0	0	0	
		19	0	1	0	
		20	0	0	0	
		21	0	0	0	
		22	0	0	0	
		23	0	0	0	
		24	0	1	0	
		25	0	0	0	

Table 13. Risk Model Output

Table 13 shows that for a likelihood input value (x) of 15% and a cost consequence of input value (y) of 22%, the firing strengths, and the resulting maximum value of 0.7143 for Rule 14. The risk rating associated with Rule 14 is medium (M).

The risk rating is saved in the cost risk state vector R_c parameters (K_{c1} = M and F_{c1} = 0.7143). This process was repeated for each of the remaining cost estimates. Schedule and performance estimates are also computed using the same process. The resulting vectors R_c, R_s, and R_p contain the risk rating levels (K) and quantitative strengths (F) for cost, schedule, and performance estimates, respectively.

Cost outputs for the 19 estimates are:

1. $R_{c1} = (K_{c1} = M; F_{c1} = 0.7143)$
2. $R_{c2} = (K_{c2} = M; F_{c2} = 0.7143)$
3. $R_{c3} = (K_{c3} = L; F_{c1} = 1.0000)$
4. $R_{c4} = (K_{c1} = M; F_{c1} = 0.3571)$
5. $R_{c5} = (K_{c5} = M; F_{c5} = 0.7143)$
6. $R_{c6} = (K_{c6} = L; F_{c1} = 0.7500)$
7. $R_{c7} = (K_{c7} = L; F_{c7} = 0.5000)$
8. $R_{c8} = (K_{c8} = H; F_{c8} = 0.0667)$
9. $R_{c9} = (K_{c9} = H; F_{c9} = 0.0750)$
10. $R_{c10} = (K_{c10} = H; F_{c10} = 0.1333)$
11. $R_{c11} = (K_{c11} = H; F_{c11} = 0.0741)$
12. $R_{c12} = (K_{c12} = L; F_{c12} = 1.0000)$
13. $R_{c13} = (K_{c13} = L; F_{c13} = 1.0000)$
14. $R_{c14} = (K_{c14} = L; F_{c14} = 1.0000)$
15. $R_{c15} = (K_{c15} = L; F_{c15} = 0.5000)$

16. $R_{c16} = (K_{c16} = M; F_{c16} = 0.5357)$

17. $R_{c17} = (K_{c17} = L; F_{c17} = 1.0000)$

18. $R_{c18} = (K_{c18} = M; F_{c18} = 0.7143)$

19. $R_{c19} = (K_{c19} = L; F_{c19} = 0.5000)$

Schedule outputs for the 5 estimates are:

1. $R_{s1} = (K_{s5} = L; F_{s1} = 0.4000)$
2. $R_{s2} = (K_{s5} = M; F_{s2} = 1.0000)$
3. $R_{s3} = (K_{s5} = L; F_{s3} = 0.3750)$
4. $R_{s4} = (K_{s5} = L; F_{s4} = 0.8000)$
5. $R_{s5} = (K_{s5} = L; F_{s5} = 0.6000)$

Performance outputs for the 7 estimates are:

1. $R_{p1} = (K_{p1} = L; F_{p1} = 0.5000)$
2. $R_{p2} = (K_{p2} = L; F_{p2} = 0.5000)$
3. $R_{p3} = (K_{p3} = L; F_{p3} = 1.0000)$
4. $R_{p4} = (K_{p5} = M; F_{p4} = 0.3571)$
5. $R_{p5} = (K_{p5} = L; F_{p5} = 0.5000)$
6. $R_{p6} = (K_{p6} = L; F_{p6} = 0.5000)$
7. $R_{p7} = (K_{p7} = M; F_{p7} = 0.7143)$

The risks associated with each of the cost, schedule, and performance estimates provide a detailed framework to assess project risks. Overall risk for cost, schedule, or performance corresponds to the highest level of risk identified for the estimates. Therefore, project cost risk is high (H), schedule risk is medium (M), and performance risk is medium (M).

CHAPTER 10

MODEL EVALUATION

10.1 Discussion

The fuzzy logic risk model developed in Chapter 8 and implemented in Chapter 9 offered a methodology for identifying and quantifying project risks associated with uncertainties in cost, schedule, and performance estimates. For the risk model developed in this research, analyses were performed to determine how the model would respond to independent inputs for likelihood (x) and consequence (y).

Inputs to the model can vary greatly depending on the application and the estimating ability of the subject-matter expert. Three situations were evaluated to determine the model's ability to delineate project risks based on varied input estimates. The first situation is where the input values vary between the lower and upper range of possible input values. This situation was referred to as mixed. The second situation was where the estimates were good. Finally, the third situation was where the estimates were understated (bad). A uniformly distributed Monte Carlo simulation was used to determine the ability of the model to address these three situations. A graphical representation of the uniform distributions shown in Figures 24, 25, and 26 represents mixed, good, and bad estimates, respectively.

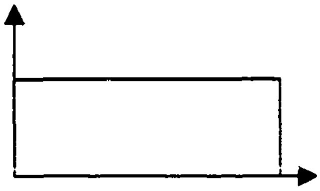


Figure 24. Mixed

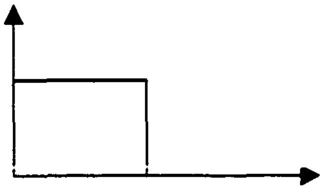


Figure 25. Good

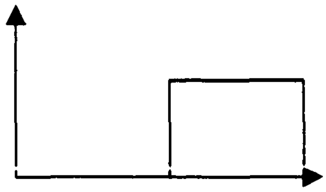


Figure 26. Bad

The uniform distribution is:

$$f(x) = \begin{cases} 1/(b-a) & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

$$\mu = (a+b)/2$$

$$\sigma^2 = (b-a)^2 / 12$$

There are three combinations of inputs for likelihood and consequence. A total of nine combinations were considered as shown below:

	Likelihood	Consequence
1.	mixed	mixed
2.	good	good
3.	bad	bad
4.	mixed	good
5.	mixed	bad
6.	good	mixed
7.	bad	mixed

8. good bad

9. bad good

10.2 Mixed Likelihood and Mixed Consequence Estimates

The simulation generated, 1000 likelihood (x) and consequence (y) inputs were normalized to the full range of possible values specified by the fuzzy membership sets in Chapter 8. The inputs for likelihood and consequences are shown in Figures 27, and 28, respectively. The mean (μ) and standard deviation (σ) for the 1000 randomly generated cost inputs were:

$l(x)$: $\mu = 30.4208$; $\sigma = 17.2759$

$m(y)$: $\mu = 14.7845$; $\sigma = 8.8215$

Risk levels determined by the fuzzy rules for the 1000 x and y inputs are shown in Figure 29. The risk levels were compared to the trapezoidal fuzzy set (TFS) values shown in Table 14. The top values and bottom values in each cell correspond to the consequence and likelihood, respectively. With $\mu = 30.4208$ for x and $\mu = 14.7845$ for y a comparison with the fuzzy membership set values shown in Table 14 – it is intuitively apparent that a preponderance of the risk will fall in the medium range (center of table).

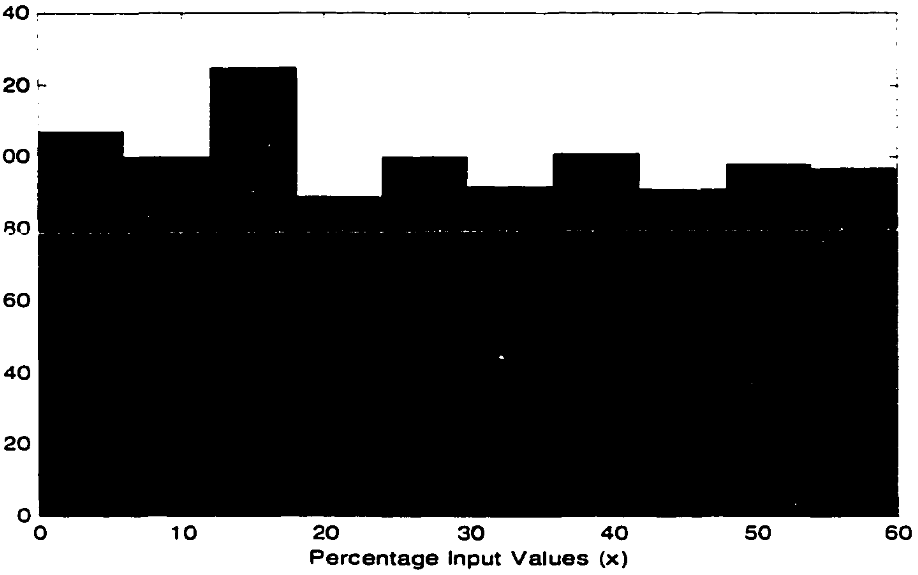


Figure 27. Mixed Cost Likelihood Input Values

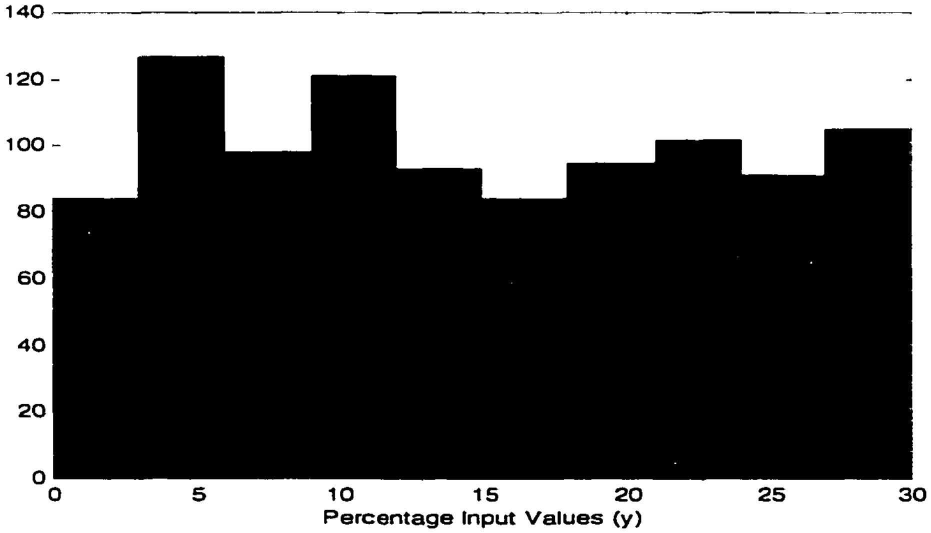


Figure 28. Mixed Cost Consequence Input Values

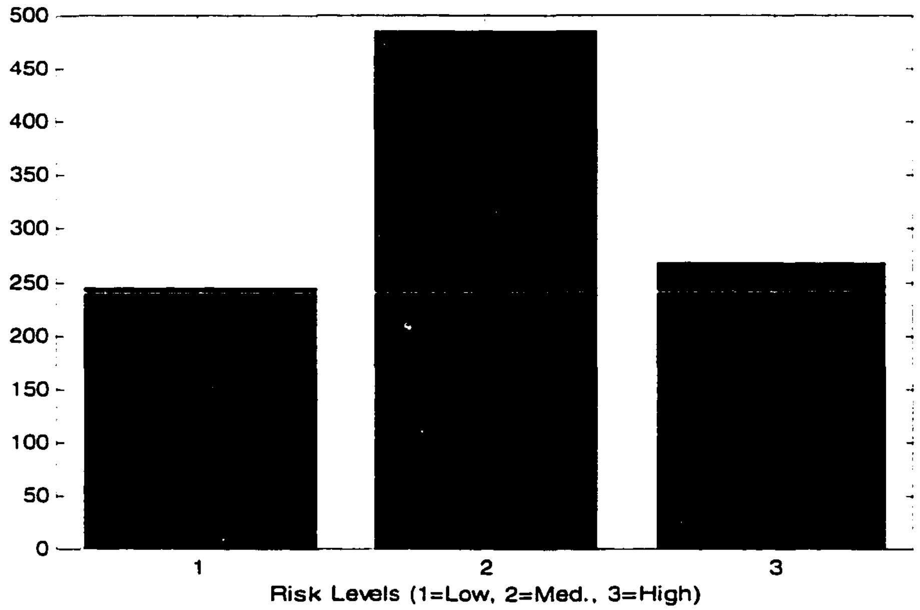


Figure 29. Mixed/Mixed Cost Risk Levels

Likelihood	1	2	3	4	5
e	0, 1, 3, 5 50, 60, ∞	3, 5, 7, 9 50, 60, ∞	7,10,12,15 50, 60, ∞	10,17,23,30 50, 60, ∞	25, ∞ 50,60, ∞
d	0, 1, 3, 5 30,38,46,55	3, 5, 7, 9 30,38,46,55	7,10,12,15 30,38,46,55	10,17,23,30 30,38,46,55	25, ∞ 30,38,46,55
c	0, 1, 3, 5 10,17,26,35	3, 5, 7, 9 10,17,26,35	7,10,12,15 10,17,26,35	10,17,23,30 10,17,26,35	25, ∞ 10,17,26,35
b	0, 1, 3, 5 3,7,11,15	3, 5, 7, 9 3,7,11,15	7,10,12,15 3,7,11,15	10,17,23,30 3,7,11,15	25, ∞ 3,7,11,15
a	0, 1, 3, 5 0, 1, 3, 5	3, 5, 7, 9 0, 1, 3, 5	7,10,12,15 0, 1, 3, 5	10,17,23,30 0, 1, 3, 5	25, ∞ 0, 1, 3, 5

Consequences

Table 14. Cost Fuzzy Trapezoidal Membership Values

Figures 30 and 31shows the inputs for likelihood and consequences, respectively. Figure 32 shows the schedule risk levels. The schedule likelihood and consequence values for 1000 iterations were:

l(x): $\mu = 30.2912$; $\sigma = 17.4886$

m(y): $\mu = 29.5689$; $\sigma = 17.6430$

The risk levels compared with the schedule trapezoidal fuzzy sets shown in Table 15 indicate that about half of the risk were in the medium range as expected.

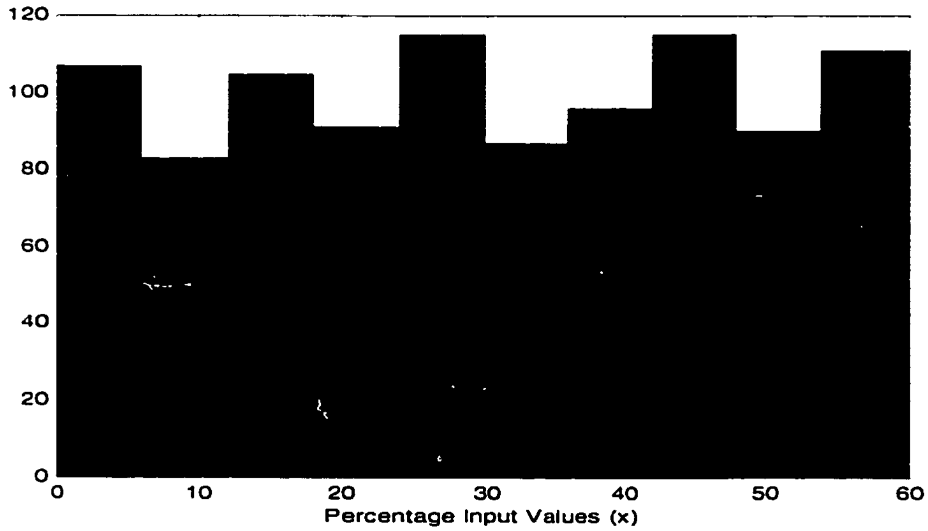


Figure 30. Mixed Schedule Likelihood Input Values

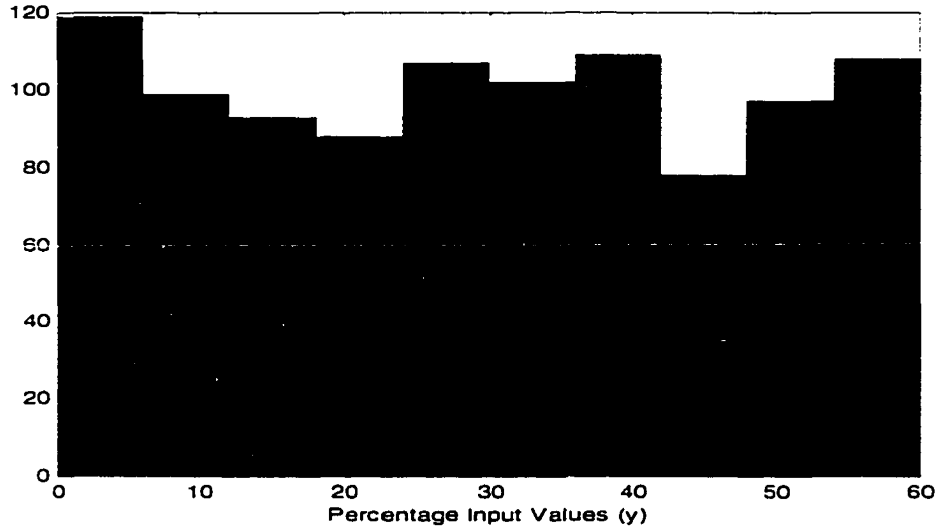


Figure 31. Mixed Schedule Consequence Input Values

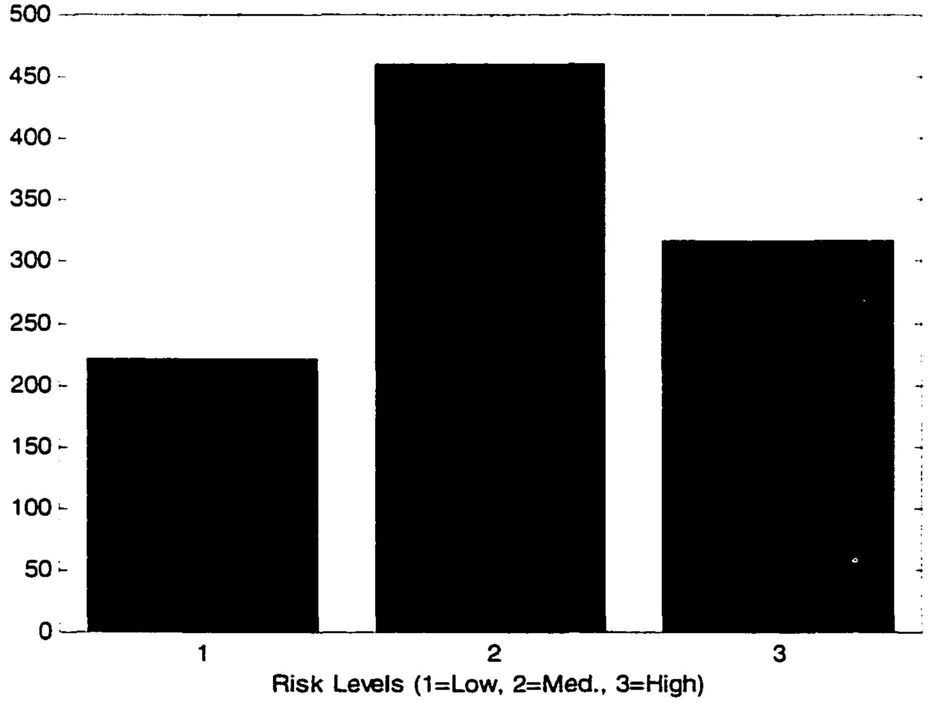


Figure 32. Mixed/Mixed Schedule Risk Levels

Likelihood					
e	0, 4, 6, 10 50, 60, ∞	6,11, 13,18 50, 60, ∞	14,20,24,30 50, 60, ∞	20,35,45,60 50, 60, ∞	50, ∞ 50,60, ∞
d	0, 4, 6, 10 30,38,46,55	6,11,13, 18 30,38,46,55	14,20,24,30 30,38,46,55	20,35,45,60 30,38,46,55	50, ∞ 30,38,46,55
c	0, 4, 6, 10 10,17,26,35	6,11,13,18 10,17,26,35	14,20,24,30 10,17,26,35	20,35,45,60 10,17,26,35	50, ∞ 10,17,26,35
b	0, 4, 6, 10 3,7,11,15	6,11,13,18 3,7,11,15	14,20,24,30 3,7,11,15	20,35,45,60 3,7,11,15	50, ∞ 3,7,11,15
a	0, 4, 6, 10 0, 1, 3, 5	6,11,13,18 0, 1, 3, 5	14,20,24,30 0, 1, 3, 5	20,35,45,60 0, 1, 3, 5	50, ∞ 0, 1, 3, 5
	1	2	3	4	5
			Consequences		

Table 15. Schedule Fuzzy Trapezoidal Membership Values

The performance simulation results were essentially the same as those from the cost simulation due to the same fuzzy trapezoidal membership values used in the example.

The risk levels when compared to the performance trapezoidal fuzzy sets also fell primarily in the medium range for the values shown in Table 16.

Likelihood					
e	0, 1, 3, 5 50, 60, ∞	3, 5, 7, 9 50, 60, ∞	7,10,12,15 50, 60, ∞	10,17,23,30 50, 60, ∞	25, ∞ 50,60, ∞
d	0, 1, 3, 5 30,38,46,55	3, 5, 7, 9 30,38,46,55	7,10,12,15 30,38,46,55	10,17,23,30 30,38,46,55	25, ∞ 30,38,46,55
c	0, 1, 3, 5 10,17,26,35	3, 5, 7, 9 10,17,26,35	7,10,12,15 10,17,26,35	10,17,23,30 10,17,26,35	25, ∞ 10,17,26,35
b	0, 1, 3, 5 3,7,11,15	3, 5, 7, 9 3,7,11,15	7,10,12,15 3,7,11,15	10,17,23,30 3,7,11,15	25, ∞ 3,7,11,15
a	0, 1, 3, 5 0, 1, 3, 5	3, 5, 7, 9 0, 1, 3, 5	7,10,12,15 0, 1, 3, 5	10,17,23,30 0, 1, 3, 5	25, ∞ 0, 1, 3, 5
Consequences	1	2	3	4	5

Table 16. Performance Fuzzy Trapezoidal Membership Values

10.3 Good Likelihood and Good Consequence Estimates

The remainder of the analysis was focused on measuring the ability of the model to determine risk levels based on varied inputs. To reduce redundancy, cost was chosen as the parameter for comparison. The uniform distribution was normalized in the Monte Carlo simulation to represent the situation where the project has relatively good estimates. The uniform distribution for "good" estimates has predominant values towards the lower end of the fuzzy membership sets. Both risk likelihood and consequences are shown in Figures 33 and 34, respectively. The associated values are:

$l(x)$: $\mu = 10.1403$; $\sigma = 5.7586$

$m(y)$: $\mu = 5.5349$; $\sigma = 3.1455$

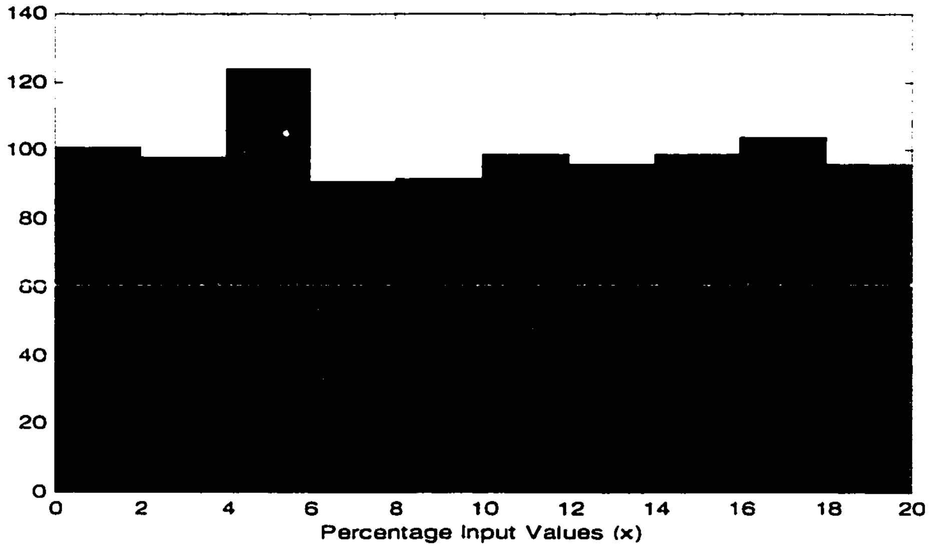


Figure 33. Good Cost Likelihood Input Values

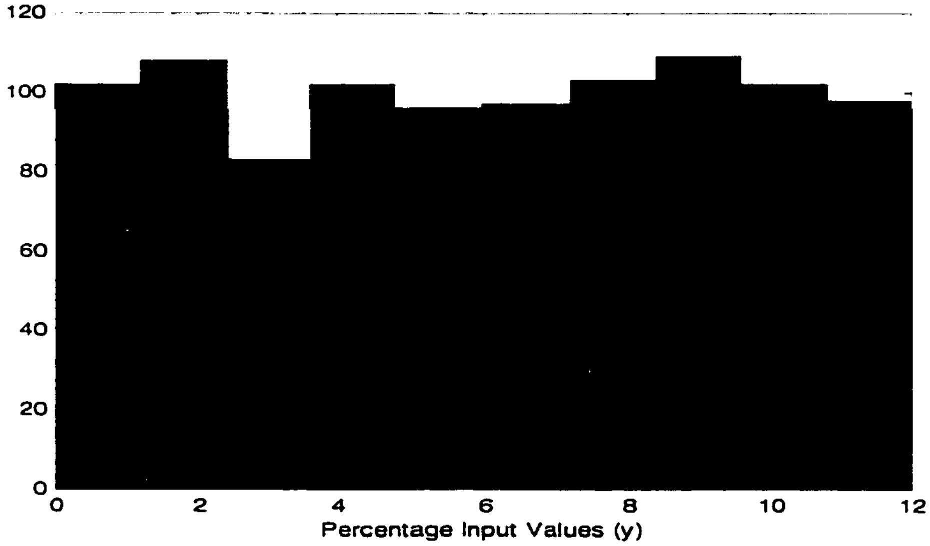


Figure 34. Good Cost Consequence Input Values

The risk levels associated with likelihood and consequences are shown in Figure 35. The likelihood and consequence values result in risk levels primarily in the low range with a few in the medium range.

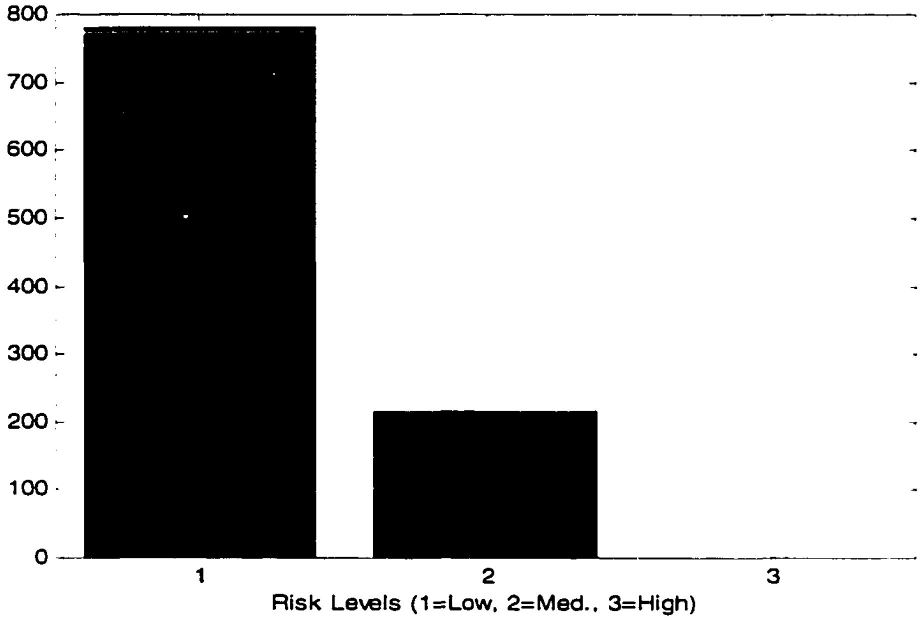


Figure 35. Good/Good Cost Risk Levels

10.4 Bad Likelihood and Bad Consequence Estimates

The situation where the project estimates are understated and the resulting consequences are "bad" will result in higher project risk. Figures 36 and 37 shows likelihood and consequences, respectively. The associated values are:

l(x): μ = 40.0354; σ = 11.6533

m(y): μ = 20.3956; σ = 5.4583

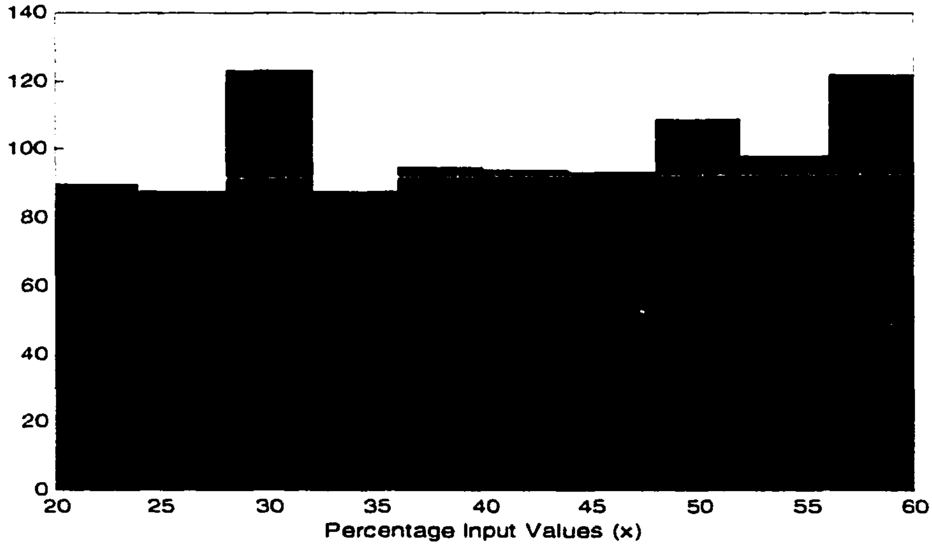


Figure 36. Bad Cost Likelihood Input Values

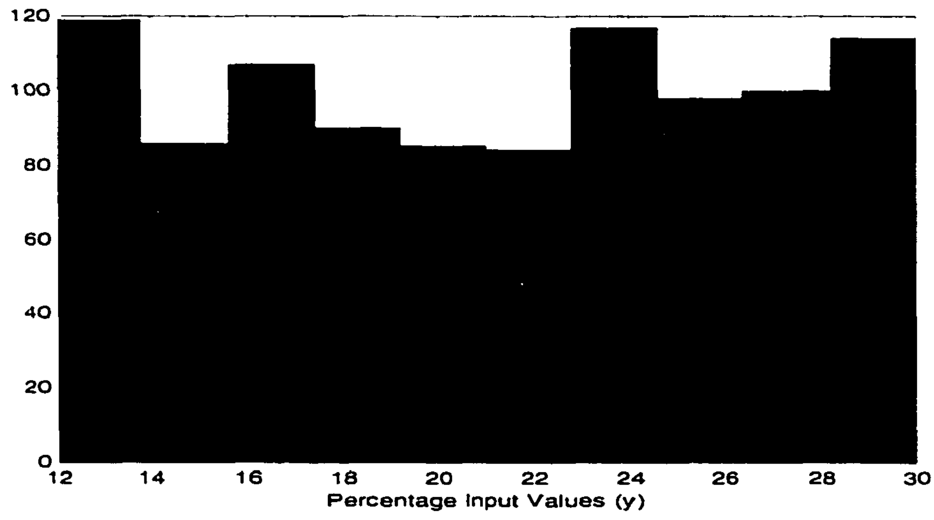


Figure 37. Bad Cost Consequence Input Values

Figure 38 shows the risk levels associated with bad likelihood and bad consequence. As expected the predominant risk levels are in the high range.

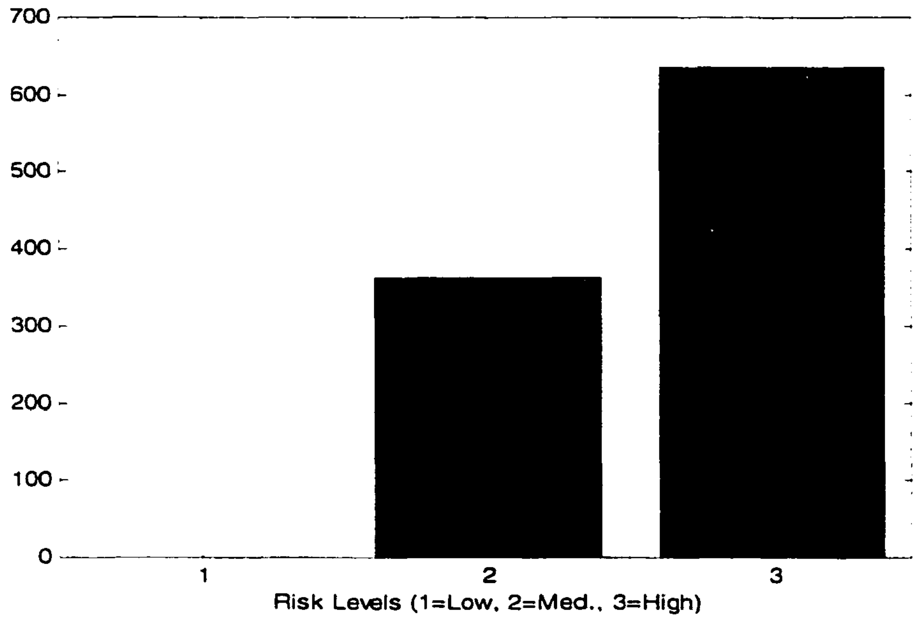


Figure 38. Bad/Bad Cost Risk Levels

10.5 Remaining Likelihood and Consequence Estimates Combinations

Figures 39 through 44 show the resulting risk levels derived from various combinations of mixed, good, and bad likelihood and consequence.

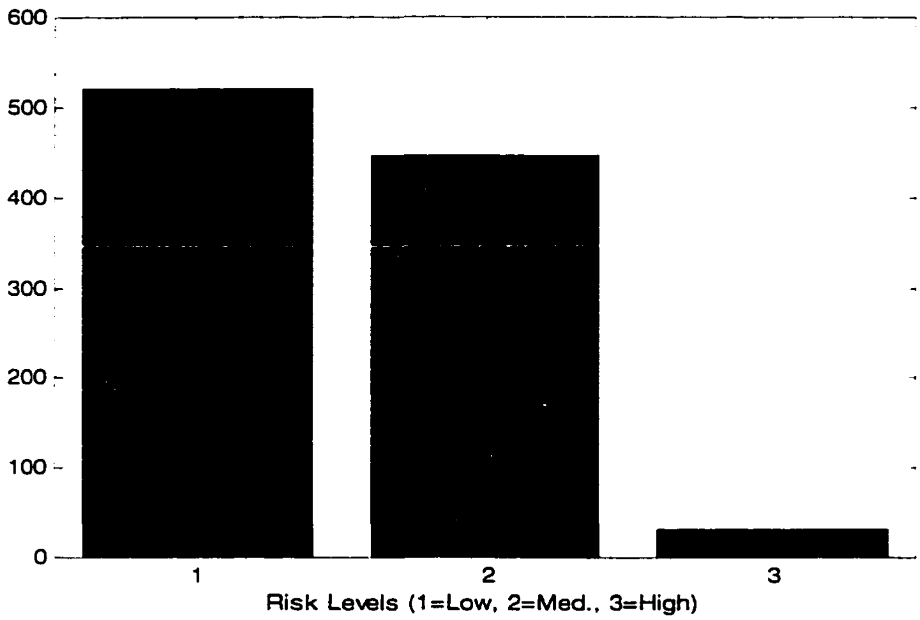


Figure 39. Mixed/Good Cost Risk Levels

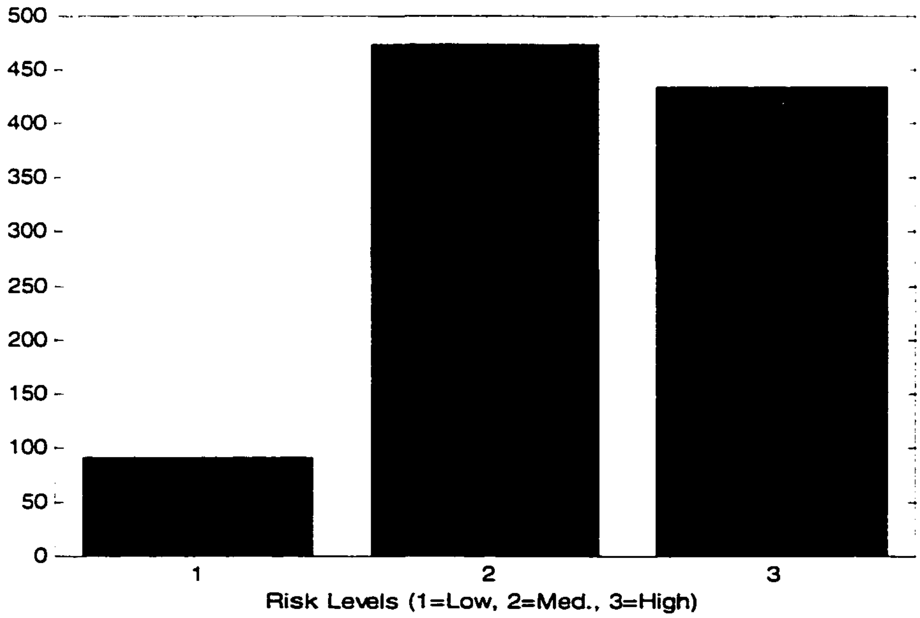


Figure 40. Mixed/Bad Cost Risk Levels

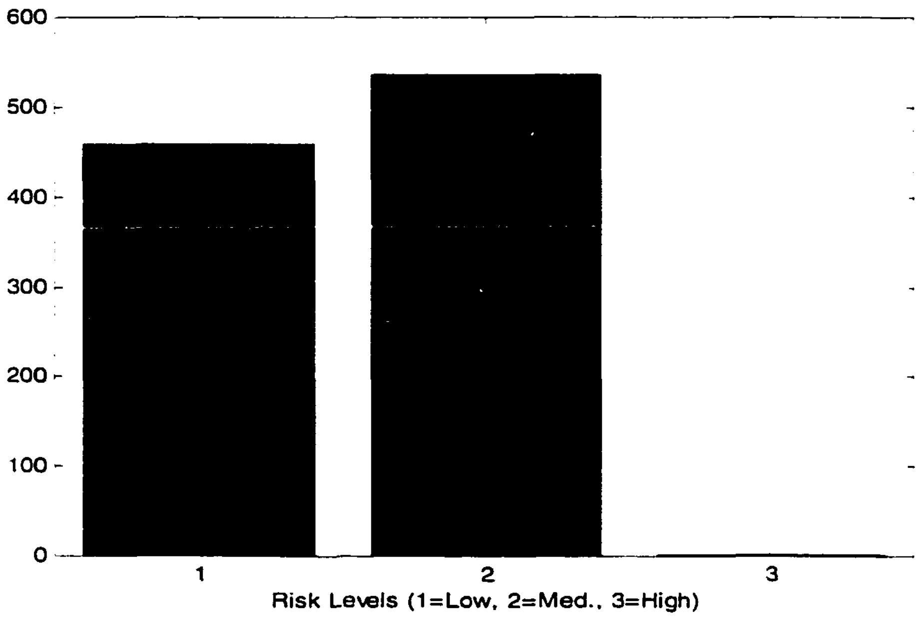


Figure 41. Good/Mixed Cost Risk Levels

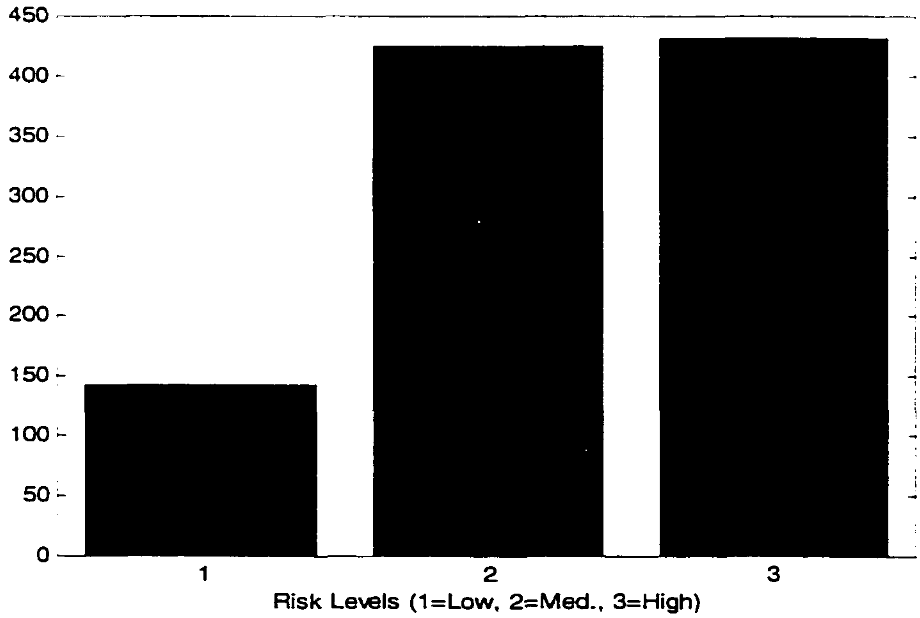


Figure 42. Bad/Mixed Cost Risk Levels

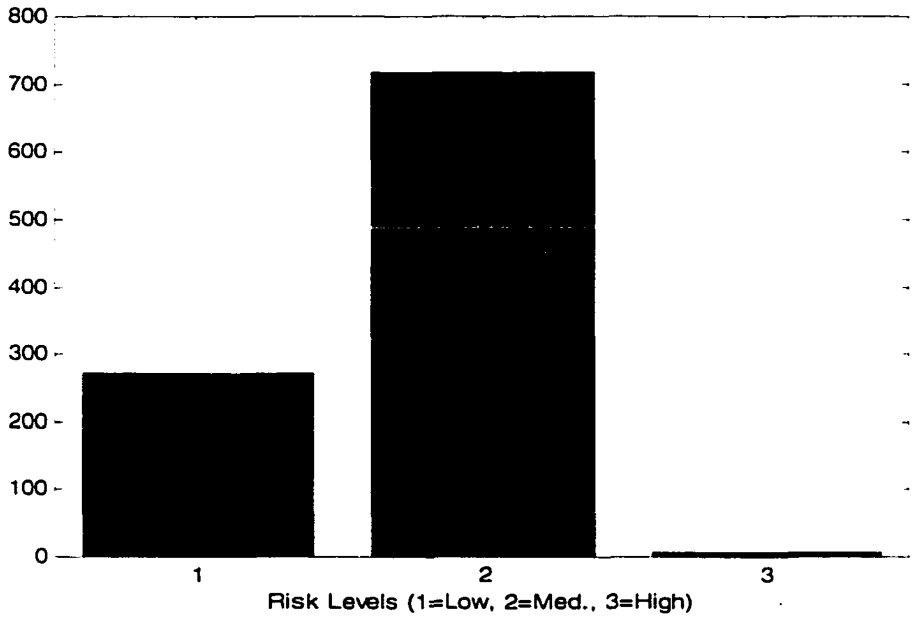


Figure 43. Good/Bad Cost Risk Levels

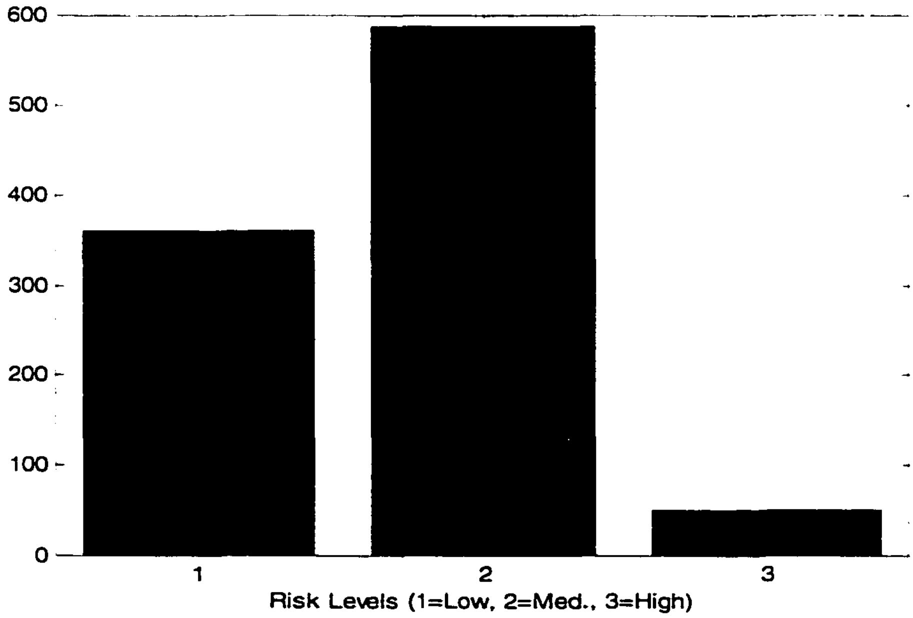


Figure 44. Bad/Good Cost Risk Levels

Model results were consistent with the likelihood and consequence inputs. The corresponding statistical values and graphical representations revealed that the pragmatic approach of using expert judgment and fuzzy sets could be used to identify and quantify project risks associated with cost, schedule, and performance estimates. The five likelihood and consequence membership sets provided a reasonable and practical level of granularity. The values of the trapezoidal membership sets were easily modified to reflect the range of values used in the example radar development project.

CHAPTER 11

CONCLUSIONS

Cost, schedule, and performance are regarded as the “magical” combination that is continuously reviewed by project managers throughout the life cycle of their projects. These three factors form the basis for project control. Therefore, to achieve project success, each of these factors must be understood and properly estimated. This research effort was focused on the identification and quantification of project risk associated with the estimation of cost, schedule, and performance. The risk model developed using fuzzy membership sets identified and quantified project risks associated with all three of the project parameters. Although the literature indicated that a few researchers have addressed cost or schedule, their efforts have not addressed the combination of cost, schedule, and performance. The key to the success of the development of the model in this work was largely attributed to the implementation methodology that included the combination of fuzzy logic and project risk analysis. Specifically, the model used fuzzy logic to address risk likelihood and the resulting potential consequences for the project estimates.

Since the risk model developed in this research effort is the only known tool that can capture total project risks due to cost, schedule, or performance estimates, it provides a unique practical method to assess project risk. The model also provides a tool to capture risk management corporate policies from subject-matter experts using fuzzy membership sets.

The fuzzy logic risk model implemented using MATLAB was demonstrated and evaluated using Monte Carlo simulations. The evaluation showed that the model consistently identified project risks based on a variety of likelihood and consequence inputs.

11.1 Contributions

The model developed in this research contributed to project risk identification and quantification by providing:

1. A structured methodology to capture corporate risk policies using fuzzy membership sets.
2. A methodology to represent imprecise linguistic variables and an interpretation of the results of the fuzzy sets.
3. A practical seamless method to transition between varying levels of project risk.
4. A framework that can be easily adapted to achieve required risk identification and quantification granularity.
5. A consistent tool to identify project risks based on individual cost, schedule, or performance estimates.

11.2 Further Research

This research effort concentrated on the identification and quantification of project risks due to cost, schedule, or performance estimation. In the future the research should be conducted on actual projects and collected data. The collected data can be correlated with actual project results to further demonstrate the utility of the model.

The research treated cost, schedule, and performance independently and did not address any of the dependent project parameters. Further research efforts should also focus on the correlation between project parameters.

REFERENCES

Badiru, A. (1992). "Computational Survey of Univariate and Multivariate Learning Curve Models", IEEE Transactions on Engineering management, Vol. 39, No. 2, pp. 176-188.

Badiru. A. and Pulat, P. (1995). Comprehensive Project Management, Prentice Hall, Englewood Cliffs, NJ.

Badiru, A. (1992). Expert Systems Applications in Engineering and Manufacturing, Prentice Hall, Englewood Cliffs, NJ.

Belgum, S. (1995). Thesis: "Activity Based Costing at the Naval Postgraduate School", Naval Postgraduate School, Monterey, CA.

Bent, J. and Humphreys, K. (1996). Effective Project Management Through Applied Cost and Schedule Control, Marcel Dekker, Inc., New York, NY.

Boehm, B.W. (1991). "Software Risk Management: Principles and Practices", IEEE Software, Vol. 8, No. 1, pp. 32-41.

Burdick, R. and Graybill, F. (1992). Confidence Intervals on Variance Components, Marcel Dekker, Inc., New York, NY.

Campbell, P. (1995). Thesis: "The Development of a Metamodel for a Major Weapon System Cost Model", Air Force Institute of Technology, Wright Patterson AFB, OH.

Chao, L. and Skibniewski, M. (1995). "Decision Analysis for New Construction Technology Implementation", Civil Engineering Systems, Vol 12, pp. 67-82.

Chao, L. and Skibniewski, M. (1998). "Fuzzy Logic for Evaluating Alternative Construction Technology", Journal of Construction Engineering and Management, Vol. 124, No. 4, pp. 297-304.

Charette, R. (1989). Software Engineering Risk Analysis and Management, Intertex Publications, McGraw-Hill, New York, NY.

Chevallier, N. and Russell, A. (1998). "Automated Schedule Generation", Canadian Journal of Civil Engineering, Vol. 25, pp. 1059-1077.

Chiu, C. and Park, C. (1998). "Capital Budgeting Decisions with Fuzzy Projects", The Engineering Economist, Vol. 43, No. 2, pp. 125-150.

Chua, D., Kog, Y. and Loh, P. (1997). "Model for Construction Budget Performance – Neural Network Approach", Journal of Construction Engineering and Management, Vol. 123, pp. 214-222.

Cooper, D. and Chapman, C. (1987). "Risk Analysis for Large Projects: Models, Methods and Cases, Wiley, Chichester.

Cooper, K. (1994). "The $2,000 Hour: How Managers Influence Project Performance Through the Rework Cycle", Project Management Journal, Vol. XXV(1), pp. 11-24.

Cottrell, W. (1999). "Simplified Program Evaluation and Review Technique (PERT)", Journal of Construction Engineering and Management, Vol. 125, No. 1, pp. 16-22.

Deshotels, R. and Zimmerman, R. (1995). Cost Effective Risk Assessment for Process Design, McGraw-Hill Inc., New York, NY.

DoD Defense Systems Management College (2000). Risk Management Guide for DoD Acquisition, Third Addition, Defense Systems Management College Process, Fort Belvoir, VA.

Douglis, C. (1998). "Cost Benefit Discussion for Knowledge-Based Estimation Tools", Project Management Journal, pp. 8-10.

Dubois, D. and Prade, H. (1980). Fuzzy Sets and Systems: Theory and Application, Academic Press, New York, NY.

Elazouni, A., Nosair, I., Mohieldin, Y., and Mohamed, A. (1997). "Estimating Resource Requirements at Conceptual Design Stage Using Neural Networks", Journal of Computing in Civil Engineering, October, pp. 217-223.

Emani, M., Goldenberg, A. and Turken, I. (1998). "Development of a Systematic Methodology of Fuzzy Logic Modeling", IEEE Transactions on Fuzzy Systems, Vol. 6, pp 346-361.

Erto, P. (1982). "New Practical Bayes Estimators for the 2-Parameter Weibull Distribution", IEEE Transactions on Reliability, R-31(2), pp. 194-197.

Erto, P. and Rapone, M. (1984). "Non-Informative and Practical Bayesian Confidence Bounds for Reliable Life in the Weibull Model", Reliability Engineering, Vol. 7, pp. 181-191.

Finley, E. and Fisher, D. (1994). "Project Scheduling Risk Assessment Using Monte Carlo Methods", Cost Engineering, Vol. 36, No. 10, pp. 24-28.

Goldberg, D. (1989). Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley Publishing Co., Reading, MA.

Golenko-Ginsberg, D. (1988). "On the Distribution of Activity Times in PERT", Journal Operations Research Society, Vol. 39, pp. 767-771.

Gray, S. (1995). Practical Risk Assessment for Project Management, John Wiley & Sons, New York, NY.

Grupe, F.H., Von Sadovszky, T., and Owrango, M. (1995). "An Executive Guide to Artificial Intelligence", Information Strategy.

Haimes, Y. (1998). Risk Modeling, Assessment, and Management, John Wiley & Sons Inc., New York.

Hegazy, T. and Ayed, A. (1998). "Neural Network Model for Parametric Cost Estimation of Highway Project", Journal of Construction Engineering and Management, Vol. 124, No. 3, pp. 210-218.

Hertz, D. and Thomas, H. (1983). Risk Analysis and Its Application, Wiley, Chicester.

Jain, L. and Martin, N. (1999). Fusion of Neural Networks, Fuzzy Sets, and Genetic Algorithms, CRC Press, Boca Raton, FL.

Johnson, H. (1992). Relevance Regained: From Top-Down Control to Bottom-Up Empowerment, The Free Press, NY, pp. 133-139.

Johnson, J. (1995). "Chaos: The Dollar Drain of IT Project Failures", Application Development Trends, pp 41-47.

Kandel, A. (1992). Fuzzy Expert Systems, CRC Press, Boca Raton, FL.

Kandel, A. (1986). Fuzzy Mathematical Techniques with Applications, Addison-Wesley Publishing Co., Reading, MA.

Kartam, N. and Tongthong, T. (1997). "Potential of Artificial Neural Networks for Resource Scheduling", Artificial Intelligence for Engineering Design, Analysis and Manufacturing, Vol. 11, pp. 171-185.

Kaufmann, A. and Gupta, M. (1988). Fuzzy Mathematical Model in Engineering and Management Science, North-Holland, Amsterdam.

Kehoe, J., Dodson, W., Reeve, R., and Gustav, P. (1995). Activity Based Management in Govemment, Coopers & Lybrand L.L.P., Wash D.C..

Keil, M., Cule, P., Lyytinen, K., and Schmidt, R. (1998). "A Framwork for Identifying Software Project Risks", Communications of ACM, pp. 76-88.

Kolluru, R., Bartell, S., Pitblado, R., and Stricoff, S. (1996). Risk Assessment and Management Handbook for Environmental, Health, and Safety Professionals, McGraw Hill Inc., New York.

Kosko, B. (1993). Fuzzy Thinking, Hyperion, New York.

Kulick, K. (1997). "Technical Performance Measurement – A Systematic Approach to Planning, Integration and Assessment", The Measurable News, p. 15.

Laufer, A. and Howell, G. (1993). "Construction Planning: Revising the Paradigm", Project Management Journal, Vol. XXIV(3), pp. 23-33.

Laufer, A. and Stukhart, G. (1998). "Incentive Programs in Construction Projects: The Contingency Approach", Project Management Joumal, Vol. XXII(2), 23-30.

Lederer, A. and Prasad, J. (1992). "Nine Management Guidelines for Better Cost Estimating", Communications of the ACM, Vol. 35, No. 2, pp. 50-60.

Lee, K., Kim, H. and Lee, J. (1998). "Case and Constraint-based Project Planning for Apartment Construction", AI Magazine, Vol. 19, pp. 13-24.

Leondes, C. (1998). Fuzzy Logic and Expert Systems Applications, Academic Press, San Diego.

Leung, H., Chuah, K., and Tummala, R. (1998). "A Knowledge-based System for Identifying Potential Project Risks", Omega Intemational Journal of Management Science, Vol. 26, pp. 623-638.

Li, H. and Love, P. (1997). "Using Improved Genetic Algorithms to Facilitate Time-Cost Optimization", Joumal of Construction and Engineering Management, Vol. 123, No. 3, pp. 233-237.

Littlefield, T. and Randolph, P. (1987). "An Answer to Saieni's Question on PERT Times", Management Science, Vol. 33, pp. 1357-1359.

Lorterapong, P. and Moselhi, O. (1996). "Project Network Analysis Using Fuzzy Sets Theory", Journal of Construction Engineering and Management, Vol. 122, No. 4, pp. 308-318.

Lui, L. (1997). "A Synergistic Approach to Project Management in Information Systems Development", International Journal of Project Management, pp 173-179.

Machacha, L. and Bhattacharya, P. (2000). "A Fuzzy Logic Based Approach to Project Selection", IEEE Transactions on Engineering Management, Vol. 47, No. 1, pp. 65-73.

Mak, C. (1995). Dissertation: "Risk Management for Improving Transmission Operations and Maintenance Activities", Department of Manufacturing Engineering, City University of Hong Kong, Kowloon, Hong Kong.

Mak, S. and Picken, D. (2000). "Using Risk Analysis to Determine Construction Project Contingencies", Journal of Construction Engineering and Management, pp. 130-136.

McNeill, D. and Freiberger, P. (1993). Fuzzy Logic, Simon & Schuster, New York.

Mok, C. (1994). Dissertation: "The Application of Risk Management Process in Building Services Cost Estimation", Department of Manufacturing Engineering, City Polytechnic of Hong Kong, Kowloon, Hong Kong.

Maus, R., and Keyes, J. (1991). Handbook of Expert Systems in Manufacturing, McGraw-Hill, St. Louis, Mo.

Melchers, R. and Stewart, M. (1995). Integrated Risk Assessment, Proceedings of the Conference on Integrated Risk Assessment, New Castle, Australia.

Memminger, M. and Wrona, J. (1999). Thesis: "Implementation of an Activity Based Costing System in a Air Force Laboratory Environment", Air Force Institute of Technology, Wright Patterson AFB OH.

Mishkoff, H. (1998). Understanding Artificial Intelligence, Howards W. Sams & Co., Indianapolis, IN.

Moder, J. (1983). Project Management with CPM, PERT and Precedence Diagramming, 3rd Edition, Van Nostrand Reinold, New York, NY.

Mulholland, B. and Christian, J. (1999). "Risk Assessment in Construction Schedules", Journal of Construction Engineering and Management, pp. 8-15.

Munakata, T. (1998). Fundamentals of the New Artificial Intelligence, Springer, New York, NY.

Negrotti, M. (1991). Understanding the Artificial: On the Future Shape of Artificial Intelligence, Springer-Verlag, Berlin-Heidelberg, Germany.

Nelson, W. (1982). Applied Life Data Analysis, John Wiley & Sons, Inc.

Nguyen, H. and Walker, E. (1999). Fuzzy Logic, Chapman & Hall/CRC, Wash DC.

Ntuen, C. and Mallik, A. (1987). "Applying Artificial Intelligence to Project Cost Estimating", Cost Engineering, Vol. 29, pp. 8-13.

Pedrycz, W. (1995). Fuzzy Sets Engineering, CRC Press Inc., Boca Raton, Florida.

Porcaro, D. (1996). "Simulation Modeling and DOE", IIE Solutions, Vol. 28, No.9, pp. 24-30.

Raymond, F. (1999). "Quantify Risk to Manage Cost and Schedule", Acquisition Review, Vol. 6, No. 2., pp. 147-155.

Riggs, J., Brown, S., and Trueblood, R. 1994). "Integration of Technical, Cost, and Schedule Risk in Project Management", Computers Operations Research, Vol. 21, No. 5, pp. 521-533.

Rogers, J. (1999). "Tools and Techniques for Decomposing and Managing Complex Design", Vol. 36, pp. 266-274.

Rotch, W. (1991). Implementing Activity-Based Costing, Executive Enterprises Publications Co., Inc., New York, NY, pp. 71-76.

Schneider, M., Kandel, A., Langholz, G. and Chew, G. (1996). Fuzzy Expert Systems Tools, John Wiley & Sons, New York, NY.

Sebastian, H. and Antonsson, E. (1996). Fuzzy Sets in Engineering Design and Configuration, Kluwer Academic Publishers, Boston.

Shipley, M., Korvin, A. and Omer, K. (1997). "BIFPET Methodology versus PERT in Project Management: Fuzzy Probability Instead of the Beta Distribution", Journal of Engineering and Technology Management, Vol. 14, pp. 49-65.

Smith, A. and Mason, A. (1997). "Cost Estimation Predictive Modeling: Regression Versus Neural Network", The Engineering Economist, Vol. 42, No. 2, pp. 137-159.

Smunt, R. (1999). "Log-linear and Non-Log-linear Learning Curve Models for Production Research and Cost Estimation", International Journal Production Research, Vol. 37, No. 17, pp. 3901-3911.

Snyder, H. and Davenport, E. (1997). "What does it really cost: Allocating Direct Cost", The Bottom Line: Managing Library Finances, Vol. 10, No. 4, pp. 158-164.

Souza, D and Lamberson, L. (1995). "Bayesian Wiebull Reliability Estimation", IIE Transactions, Vol. 27, No. 3, pp. 311-321.

Tabtabi, H., Kartam, N., Flood, I., and Alex, A. (1997). "Construction Project Control Using Artificial Neural Networks", Artificial Intelligence for Engineering Design, Analysis and Manufacturing, Vol. 11, pp. 45-57.

Tavares, V., (1999). Advanced Models for Project Management, Kluwer, Boston, MA.

Terano, T., Asai, K. and Sugeno, M. (1992). Fuzzy Systems Theory, Academic Press Inc., San Diego, CA.

Thompson, A. and Perry, J. (1992). Engineering Construction Risks: A Guide to Project Risk Analysis and Risk Management, Thomas Telford, London.

Ting, P., Zhang, C., Wang, B., Deshmukh, A., and Dubrosky, B. (1999). "Product and Process Cost Estimation with Fuzzy Multi-Attribute utility Theory", Engineering Economist, Vol. 44, pp. 303-317.

Toaklley, A. (1995). "Risk Management Applications – a Review", Australian Institute of Builders Papers, Vol. 6, pp. 77-85.

Weber, S. (1999). "Model for Construction Budget Performance – Neural Network", Journal of Construction Engineering and Management, Vol. 125, May/June, pp. 213-214.

Wright, T. (1936). "Factors Affecting the Cost of Airplanes", Journal of Aeronautical Sciences, Vol. 3, pp. 122-128.

Yager, R. (2000). "Fuzzy Modeling for Intelligent Decision Making Under Uncertainty", IEEE Transactions of Systems, Man, and Cybernetics, Vol. 30, No. 1, February, pp. 60-70.

Zadeh, L. A. (1983). "The Role of Fuzzy Logic in the Management of Uncertainty in Expert Systems", Fuzzy Sets and Systems II Amsterdam, The Netherlands, pp.199-227.

APPENDIX A

REGRESSION ANALYSIS

The purpose of regression analysis is to improve our ability to predict the next "real world" occurrence based on historical data. Regression analysis is defined as the mathematical nature of the association between two variables. The association is determined in the form of a mathematical equation. The equation provides the ability to predict one variable on the basis of the knowledge of the other variable. The variable whose value is to be predicted is called the dependent variable. The variable about which knowledge is available or can be obtained is called the independent variable. The functional relationship can be described graphically (on a common X-Y coordinate system) by a straight line and mathematically by the common form:

$$y = a + bx$$

where y = (represents) the calculated value of y - the dependent variable

x = the independent variable

b = the slope of the line, the change in y divided by the corresponding change in x

a and b are constants for any value of x and y

A.1 Curve Fitting

There are two standard methods of curve fitting. In one method the analyst plots the data and fits a smooth curve to the data. This is known as the graphical method. The other method uses formulas or a "best-fit" approach where an appropriate theoretical curve is assumed and mathematical procedures are used to determine the "best-fit" curve. This is known as the Least Squares Best Fit (LSBF) method. The simplest model to handle is the straight line.

The LSBF method specifies the one line that best fits the data set. The method minimizes the sum of the squared deviations of the observed values of Y and calculated values of Y. For example, if the distances: $(Y_1 - Y_{C1})$, $(Y_2 - Y_{C2})$, $(Y_3 - Y_{C3})$, $(Y_4 - Y_{C4})$, etc., parallel to the Y-axis, are measured from the observed data points to the curve, then the LSBF line is the one that minimizes the following equation (see Figure A.1):

$$(Y_1 - Y_{C1})^2 + (Y_2 - Y_{C2})^2 + (Y_3 - Y_{C3})^2 + \ldots + (Y_n - Y_{Cn})^2$$

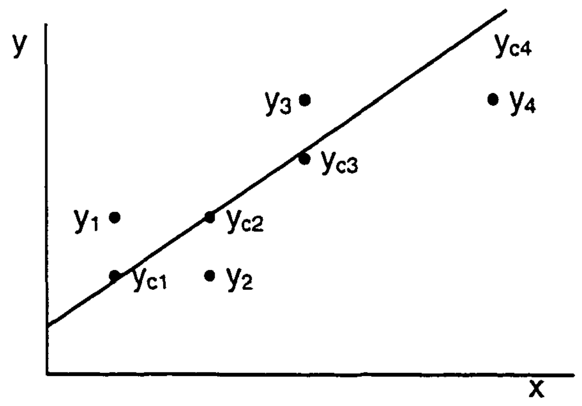


Figure A.1. LSBF Line.

The sum of the deviations from the observed value of Y, and the calculated value of $Y - Y_c$ squared, is a minimum; i.e., $(Y - Y_C)^2$ is a minimum. This same distance, $(Y - Y_C)$ is the error term or residual. Therefore, the LSBF line is one that can be defined as:

ΣE^2 is a minimum because $\Sigma (Y - Y_C)^2 = \Sigma E^2$

For a straight line,

$Y = a + bx$

and, with N points,

$(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \ldots (Y_n, Y_n)$

The sum of the squares of the distances is a minimum if,

$$\Sigma Y = AN + B\Sigma X \text{ and}$$

$$\Sigma XY = A\Sigma X + B\Sigma X^2$$

The arithmetic mean is the sum of the values of the independent variable divided by the number of observations or $\Sigma X/n = \bar{x}$ and the sum of the "Ys" divided by the number of observations or $\Sigma Y/n = \bar{y}$. The parameters, a and b, define a unique line with a Y-intercept of a and a slope of b. The values needed to solve for a and b, are shown in the spreadsheet below.

X	Y	X*Y	X^2	Y^2
X_1	Y_1	$X_1 * Y_1$	X_1^2	Y_1^2
X_2	Y_2	$X_2 * Y_2$	X_2^2	Y_2^2
X_3	Y_3	$X_3 * Y_3$	X_3^2	Y_3^2
-	-	-	-	-
-	-	-	-	-
ΣX_n	ΣY_n	$\Sigma(X_n * Y_n)$	ΣX_n^2	ΣY_n^2

Table A.1. Sums, Squares, and Cross Products

A.2 Correlation Analysis

The LSBF regression equations are used to determine goodness of fit. In order to make this determination, a check is made for the goodness of fit, the coefficient of correlation (R), and the related coefficient of determination (R^2).

One indicator of the "goodness" of fit of a LSBF regression equation is correlation analysis. Correlation analysis considers how closely the observed points fall to the LSBF regression equation. The assumption is that the more closely the observed values are to the regression equation, the better the fit.

The coefficient of determination (R^2) represents the proportion of variation in the dependent variable that has been explained or accounted for by the regression line. The value of the coefficient of determination may vary from zero to one. A coefficient of determination of zero indicates that none of the variation in Y is explained by the regression equation; whereas a coefficient of determination of one indicates that 100 percent of the variation of Y has been explained by the regression equation.

In order to calculate R^2 the following equation is used:

$$R^2 = \frac{(\sum xy - n\overline{xy})^2}{(\sum x^2 - \overline{x}\sum x) \bullet (\sum y^2 - \overline{y}\sum y)}$$

APPENDIX B

WORK BREAKDOWN STRUCTURE (WBS)

A key strategy of effective planning is to partition a project into manageable chunks that can be individually planned estimated and controlled. The WBS is a powerful tool for expressing the scope or extent of a project in simple graphic terms. It represents the project in terms of the hierarchy of deliverables and services it will produce. Tasks that are contained in the WBS collectively describe the total project. The tasks may involve physical objects, services and data. The WBS provides the link between the end objective and the operations required to achieve that objective. The WBS starts with a single box at the top, which represents the whole project. The project is then partitioned into its components with lower level boxes.

Although the WBS is typically used for component partitioning, it is proposed to be used to structure the project so that associated cost, schedule and performance estimates can be organized. By assigning a code to each WBS element, tracking can be achieved throughout the project life cycle.

The WBS is a graphical tool that displays the project's statement of work making it easier to understand and communicate. The WBS is employed from the earliest stages of project planning. The WBS supports the principle of management through deliverables thus providing a map of what is to be produced.

The WBS can be used to partition the major project deliverables into smaller components to improve the accuracy of cost estimates. It has been used to provide a mechanism for collecting and organizing actual costs. Figure B.1 shows a high level WBS. The box at the top represents the total system and is referred to as WBS level 0. Lower levels describe the project components in increasing detail and are numbered 1, 2, 3, and so on. The concept of WBS is important as it allows the designation of the level of detail from which cost estimates and other project information can be organized. The lowest levels of a WBS represent discrete deliverable items against which project parameters can be measured.

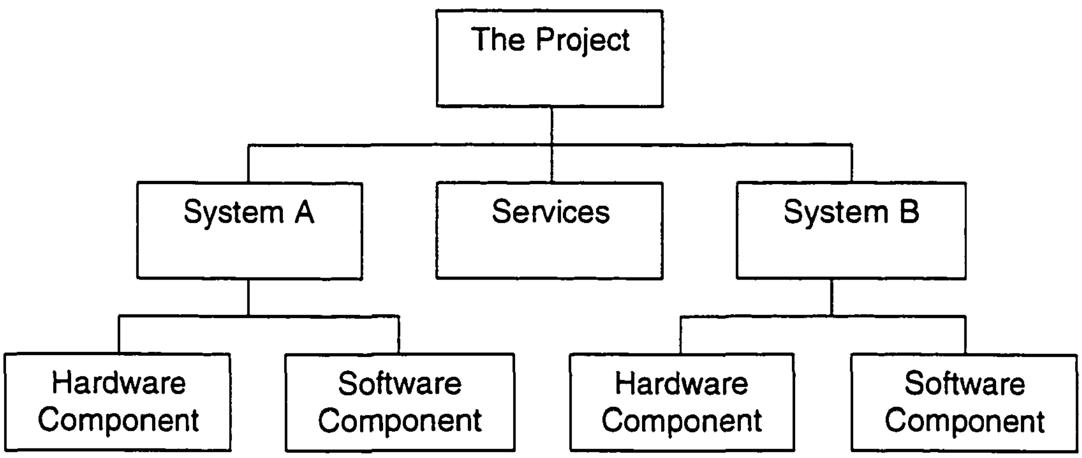


Figure B.1. Work Breakdown Structure

APPENDIX C

ACTIVITY BASED COSTING (ABC)

Allocating costs in an ABC system is no different, in principle, than any allocation process. The steps involved are:

1. Select the products/services about which to gather costs;
2. Assign the direct costs;
3. Examine each overhead cost associated with the product/service to determine if some cost driver exists that predicts the overhead cost;
4. Apply any remaining overhead using some standard basis.

Activity based costing differs from traditional costing to the extent that it relies on multiple activity-related bases to allocate overhead instead of allocating overhead based on a simple algorithmic basis.

Traditional cost systems use a two-stage procedure to assign an organization's indirect and support expenses to outputs. Operating expenses are assigned first to cost pools and second, to the outputs of the production process. These traditional two-stage assignment procedures distort reported costs. The traditional systems assign costs from cost pools to outputs using volume drivers such as labor, machine hours, material purchases and units produced. Since many indirect and support resources are not used in proportion to the number of output units produced, these traditional systems provide inaccurate measures of the cost of support activities for individual outputs. The traditional two stage approach is shown in Figure C.1.

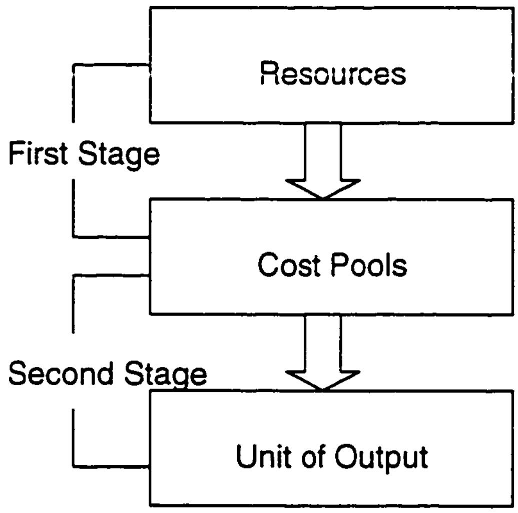


Figure C.1. Traditional Two Stage Cost System

Activity based cost systems differ from traditional cost systems by assigning the usage of all organizational resources to the activities performed by these resources and then linking the cost of these activities to outputs such as products, services, customers and projects. In particular, activity-based systems measure more accurately the cost of activities not proportional to the volume of outputs produced. In manufacturing processes, four categories of activities can be identified: unit, batch, product and facility. The ABC approach is shown in Figure C.2.

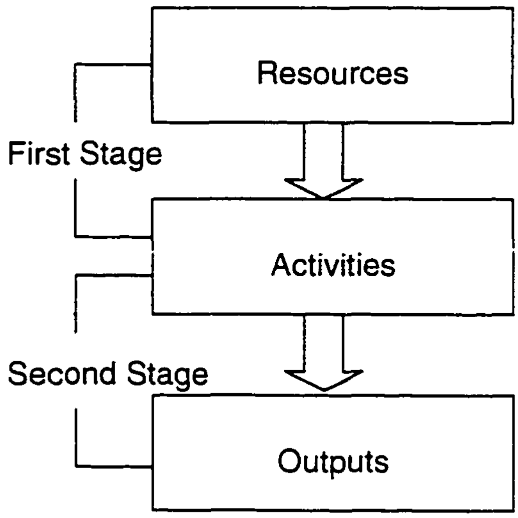


Figure C.2. Activity Based Cost System

In ABC there is a hierarchy of operations. Each level is an aggregation of the ones below it. Executives make decisions about the highest level, such as what goals to set and where to make high-cost improvement investments. Managers and employees working in the lower levels contribute information and recommendations to guide decisions at the higher levels. They also determine how best to implement goals within their processes or activities.

An activity is a unit of work usually done by one or more persons belonging to the same office, branch or other small group. Within an activity there are discrete tasks. Tasks are made up of even smaller units of work called steps.

Every activity has inputs. Inside the activity, the inputs are transformed or converted into identifiable outputs: a product or service, and sometimes information on the output. The external components of the transformation include the people who do the work in the activity; the equipment, methods, and supplies they use; and the physical environment where the activity exists.

Activities have identifiable boundaries, or starting and ending points. They start when people inside the transformation component gain control over inputs, and stop when these same people hand over control of outputs to another activity, or to external customers.

All activities have customers and suppliers. An activity's internal customer is another activity inside the organization that receives output or information from the first activity. External suppliers are people and organizations outside an agency that provide it with materials, information, or services.

APPENDIX D

FUZZY LOGIC OPERATIONS

D.1 Operations on Fuzzy Sets

Operations on fuzzy sets are similar to operations on conventional sets. The following relationships hold for fuzzy sets A and B (Badiru, 1992):

Equality:

$$A = B \text{ if and only if } u_A(x) = u_B(x),\ \forall\, x \in X$$

Containment:

$$A \subseteq B \text{ if and only if } u_A(x) \leq u_B(x),\ \forall\, x \in X$$

Intersection:

$$u_{A \cap B}(x) = \min\{u_A(x), u_B(x)\}$$

Union:

$$u_{A \cup B}(x) = \max\{u_A(x), u_B(x)\}$$

Complement:

$$u_{A'}(x) = 1 - u_A(x)$$

The intersection of two fuzzy sets A and B creates the largest fuzzy set membership values that is a subset of both A and B. While the union of two fuzzy sets A and B creates the smallest fuzzy membership values that is a subset of A and B.

D.2 Operational Properties

Operational Properties for fuzzy sets A, B and C are shown below (Badiru, 1992):

Distributive property:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Associative property:

$$(A \cup B) \cup C = A \cup (B \cup C)$$

$$(A \cap B) \cap C = A \cap (B \cap C)$$

Commutative property:

$$A \cap B = B \cap A$$

$$A \cup B = B \cup A$$

Idempotence property:

$$A \cap A = A$$

$$A \cup A = A$$

DeMorgan's law:

$$u_{(A \cap B)'}(x) = u_{A' \cup B'}(x)$$

$$u_{(A \cup B)'}(x) = u_{A' \cap B'}(x)$$

Concentration:

$$CON(A) = A^2$$

Dilation:

$$DIL(A) = A^{0.5}$$

D.3 Mapping Functions

In order to graphically represent fuzzy values, the actual characteristics of the membership functions must be simulated over a mapping function. A mapping function is a graphical representation of an element as it passes through the continuum of membership values. The goal of the mapping function is to describe subjective and ambiguous estimates in a membership domain in the range [0, 1]. The mapping function provides a means to view the progression of changes in the state of given variables over a membership function.

APPENDIX E

NEURAL NETWORKS

A neural network is an abstract computer model of the human brain. Similar to the brain, a neural network is composed of artificial neurons called units and interconnections. A neural network is viewed as a graph with neurons represented as nodes and interconnections as edges. Neural networks are quite different than expert systems and case-based reasoning. Neural networks are used to create new knowledge. These systems attempt to employ the same learning-through-repetition technique that humans use. Given a series of examples, the neural network learns by inducing patterns that distinguish the examples from one another. Every neuron model consists of a processing element with synaptic input connections and a single output. The signal flow of neuron inputs, x_i, is considered to be unidirectional as indicated by arrows, as is a neuron's output signal flow. A general neuron symbol is shown in Figure E.1. This figure shows a set of weights and the neuron's processing unit, or node.

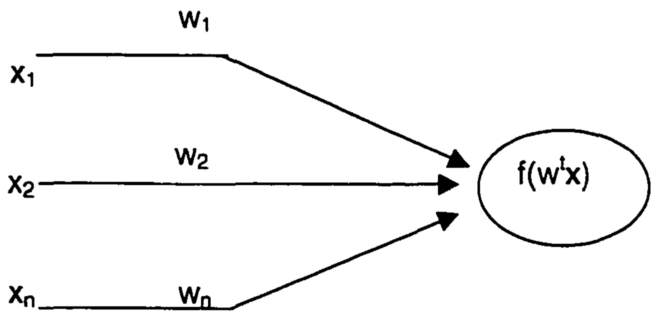


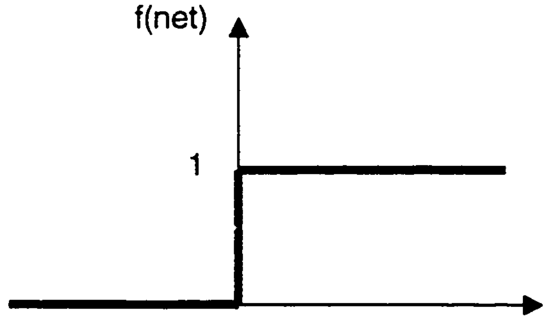
Figure E.1. Neuron with Processing Node and Synaptic Connections

There are two kinds of artificial neurons from which most networks are built. The two models differ mainly in the type of data they can handle, which is either binary or continuous. In the first case several discrete numbers – x_1, x_2, ... x_n – enter the artificial neuron as inputs. Using the biological nerve cell as a neuron from other neurons. In the mathematical model, the aggregation of inputs is represented by the weighted sum:

$$net = w_1\,x_1 + w_2\,x_2 + \ldots\,w_n\,x_n$$

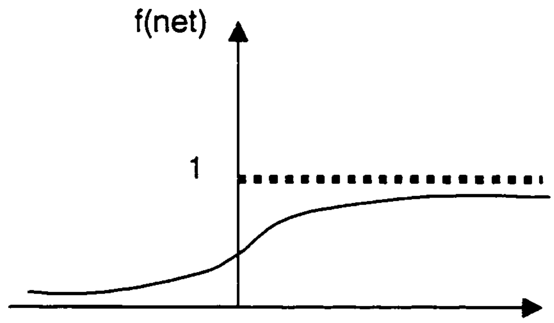
where the "weights" w_1, w_2, ... w_3, which can be any real numbers, measure the strength of the connection of each input to the neuron body. These numbers are also known as synaptic weights. In an artificial neuron, a typical threshold value is zero, and when it "fires" the combined inputs produce an output value f(net) = 1, if net is greater than zero; otherwise, it does not "fire" (f(net) = 0). The function f is called the neuron's activation function.

A continuous neuron accepts any real numbers as inputs and it can respond with any number between 0 and 1 as output – a continuous range of values, or graded response. The activation f(net) of the continuous neuron is also a function of the weighted sum net. Figure E.2 shows a graph of the discrete neural activation function. Figure E.3 shows a graph of the continuous activation function.



$$f(net) = \begin{cases} 1, \text{ if } net > 0 \\ 0, \text{ if } net \leq 0 \end{cases}$$

Figure E.2. Discrete Activation Function



$$f(net) = \frac{1}{1 + \exp(-\lambda net)}$$

E.3. Continuous Activation Function

The neuron output signal is given by the following relationship:

$o = f (w^t x)$ or $o = f (\Sigma w_i x_i)$

where w is the weight vector defined as:

$$w = [\ w_1\ w_2\ ....w_n\ ]^t$$

and x is the input vector defined as:

$$x = [\ x_1\ x_2\ ....x_n\ ]^t$$

The function f (w^t x) is called an activation function and its domain is the set of activation values, net, of the neuron model. The variable net is defined as a scalar product of the weigh and input vector:

$$net = w^t x$$

The activation function is an analog of the biological neuron's membrane potential. The neuron as a processing node performs the operation of summation of its weighted inputs, or the scalar product computation to obtain net. The activation function is used to perform the non-linear operation f(net). Typical discrete activation functions that are used are as follows:

$$f(net) \equiv \frac{2}{1 + \exp(-\lambda net)} - 1$$

$$f(net) \equiv sgn(net) = \begin{cases} +1, & net > 0 \\ -1, & net < 0 \end{cases}$$

The following activation functions are used for the continuous situation:

$$f(net) \equiv \frac{1}{1 + \exp(-\lambda net)}$$

$$f(net) \equiv sgn(net) = \begin{array}{l} +1, \ net > 0 \\ 0, \ net < 0 \end{array}$$

The back-propagation algorithm is designed to train feed-forward networks composed of two or more layers of neurons, and connected so that the outputs from one layer become the inputs to the next one. In addition, the activation functions of the neurons must be continuous (to allow for the use of differential calculus). The algorithm derives its name from the fact that the weight adjustments dictated by the learning rules propagate "backwards", from the output layer towards the input layer. Figure E.4 shows a network consisting of two layers.

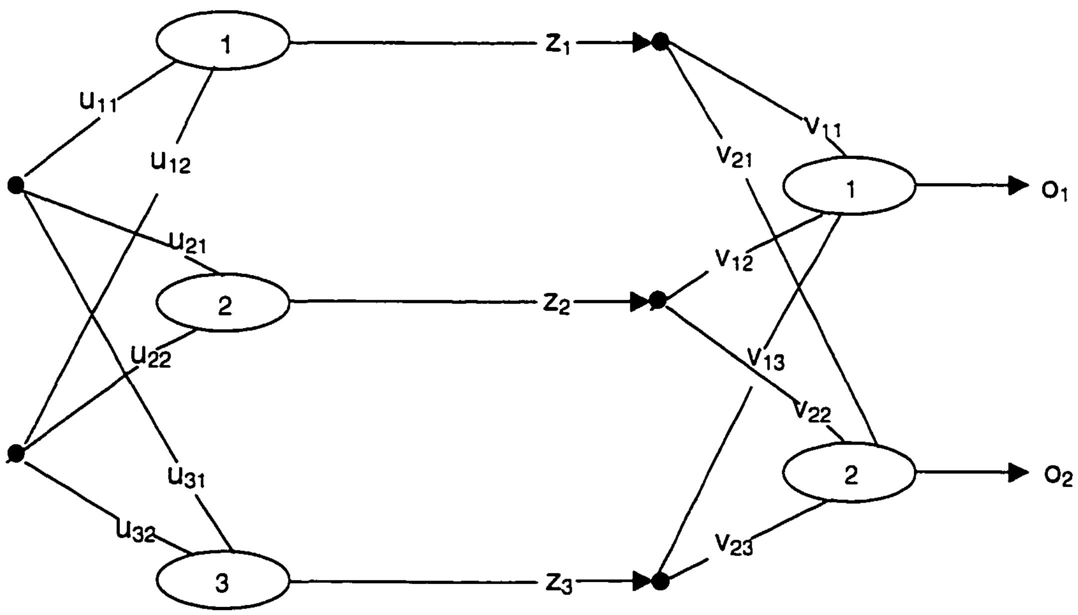


Figure E.4. Two Layer Network

The three neurons in the first (or hidden) layer receive inputs y_1 and y_2, and respond with intermediate, or hidden, outputs z_1, z_2, and z_3. These are then passed on to the two neurons in the second (output) layer, which transform the z_j into the final outputs o_1 and o_2 according to the formulas:

$$o_i = f(s_i) \quad \text{for } i = 1, 2$$

where f if the neuron's activation function; s_i is the weighted sum:

$$s_i = v_{i1} z_1 + v_{i2} z_2 + v_{i3} z_3$$

and v_{ik} is the weight, or strength, of the connection joining the k-th input to the i-th neuron. The z_j are themselves computed in a similar way, by applying f to the weighted sum of the inputs y_1 and y_2 :

$$z_j = f(u_{j1} y_1 + u_{j2} y_2), \; j = 1, 2, 3$$

After a given input pattern $y = [y_1, y_2]$ has been processed, the network responds with an output vector $o = [o_1, o_2]$. The response error E is calculated by comparing o with the desired response, which is another vector $d = [d_1, d_2]$. E is defined by:

$$E = (d_1 - o_1)^2 + (d_2 - o_2)^2$$

that is, the sum of the squares of the local errors $d_k - o_k$ at each output neuron.

In the simplest case, when E is a function of only two variables, w_1 and w_2, the error can be visualized as surface in 3-dimensional space hanging above the 2-dimensional weight plane. Starting at any given point on this surface, there is one direction that corresponds to the steepest climb, or, equivalently, to the fastest rate of increase in error. Mathematically, the direction defined by the gradient of the error is a 2-dimensional vector ∇E whose components are calculated using differential calculus (partial derivatives $\delta E / \delta w_1$ and $\delta E / \delta w_2$). The error decreases most rapidly in the direction opposite that of ∇E, which is that of the vector $-\nabla E$. This is the direction of steepest descent. For the initial weight vector w, E(w) is the error for these particular weights.

The new weight vector $w^{'}$ is dictated by the imperative to reduce the error as rapidly as possible. Thus, a weight increment Δw that moves the current weight w along the direction of steepest descent is computed. Starting with the output layer, the algorithm calculates the components of Δw layer by layer. The "error signal" travels backwards, from output to input, enters into the calculation of the weight increments, from which the algorithm's name "back-propagation" is derived.

The back-propagation is an efficient technique for calculating the gradient error in one sweep through the network, working with only one input pattern at a time. By reducing the error the weights are adjusted (trained) to the best solution.

APPENDIX F

GENETIC ALGORITHMS

A genetic algorithm is an iterative procedure that consists of a constant-size population of individuals, each one represented by a finite string of symbols, known as the genome. Adaptation in natural populations aims at improving the fitness, and therefore the chances for survival, of the group as a whole. Genetic algorithms, on the other hand, are mostly concerned with "breeding" one exceptional individual, whose "genetic code" would represent the optimal or near-optimal solution to a problem. In order to increase the likelihood of such a desirable event, the algorithm gradually improves the average quality of the entire "generation" of potential solutions, just as in the biological case.

An initial population of individuals is generated at random or heuristically. Every evolutionary step, known as a generation is decoded and evaluated according to some predefined quality criterion, referred to as the fitness function. To form a new population (the next generation), individuals are selected according to their fitness. Selection along cannot introduce new individuals into the population. New individuals are generated by genetically-inspired operators. The most popular ones are the crossover and mutation. Crossover is performed with probability p between two selected individuals, called parents, by exchanging parts of their genomes to form two new individuals called offspring.

The strategy of a genetic algorithm is based on the mechanisms of natural selection and evolution. Its implementation normally requires considerable computing resources to store the data, automate the operations, and speed up the evolutionary clock. After some preliminary steps (coding, selecting the initial solution pool, etc.), the evolutionary plan proceeds in a cyclic fashion, producing a new "generation" of potential solutions after each cycle. A combination of chance and controlled "reproduction" will favor the development of first-rate solutions in the long run, perhaps after thousands of generations. There is no guarantee that the model will actually reach an optimal or near-optimal solution. The case for the plan's eventual success rests on statistical arguments rather than on exact mathematical proof.

Genetic algorithms are stochastic iterative processes that are not guaranteed to converge. The termination condition may be specified as some fixed, maximal number of generations or as the attainment of an acceptable fitness level.

APPENDIX G

PROBABILITY DISTRIBUTIONS

G.1 Beta Distribution

PERT analysis is based on the probabilistic properties of activity duration being modeled by the beta probability density function. The beta distribution is defined by two end points and two shape parameters. The beta distribution was originally chosen because it is a reasonable distribution to model activity times due to the end points and shape definition. The general beta probability density function is as follows:

$$f(t) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \cdot \frac{1}{(b-a)^{\alpha+\beta-1}} \cdot (t-a)^{\alpha-1}(b-t)^{\beta-1}$$

$$\text{for } a \leq t \leq b \quad \text{and} \quad \alpha > 0,\ \beta > 0$$

where a = lower end point of distribution

b = upper end point of distribution

α, β are the shape parameters for the distribution

The mean, variance and mode of the beta distribution are shown in the following equations:

$$\mu = a + (b-a)\frac{\alpha}{\alpha+\beta}$$

$$\sigma^2 = (b-a)^2 \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

$$m = \frac{\alpha(\beta - 1) + b(\alpha - 1)}{\alpha + \beta - 2}$$

G.2 Triangular Distribution

The triangular probability distribution has been used as an alternative to the beta distribution in estimating activity times. The three parameters associated with the triangular distribution are minimum value (a), mode (m) and maximum (b). The triangular function is as shown:

$$f(t) = \frac{2(t-a)}{(m-a)(b-a)} \qquad \text{for } a \le t \le m$$

$$f(t) = \frac{2(b-t)}{(b-m)(b-a)} \qquad \text{for } m \le t \le b$$

where the mean and variance are defined as:

$$\mu = \frac{a+m+b}{3}$$

$$\sigma^2 = \frac{a(a-m) + b(b-a) + m(m-b)}{18}$$

G.3 Uniform Distribution

The uniform distribution is used when the extremes of an activity duration can be estimated. In this situation the assumption is that the intermediate values are equally likely to occur. The uniform distribution is defined as follows:

$$f(t) = \frac{1}{b - a} \qquad \text{for } a \le t \le b$$

$$f(t) = 0 \qquad \text{otherwise}$$

where the mean and variance are:

$$\mu = \frac{a + b}{2}$$

$$\sigma^2 = \frac{(b-a)^2}{12}$$

G.4 Bayes Theorem

$B_1, B_2, \ldots, B_n$ is the set of events forming a partition of the sample space S, where $P(B_n) \neq 0$, for $i = 1, 2, \ldots, n$, and if A is any event of S such that $P(A_k) \neq 0$; then for $k = 1, 2, \ldots, n$:

$$P(B_k \mid A) = \frac{P(B_k \cap A)}{\sum_{i=1}^{n} P(B_i \cap A)}$$

$$= \frac{P(B_k)\, P(A \mid B_k)}{\sum_{i=1}^{n} P(B_i)\, P(A \mid B_i)}$$