RANGE ESTIMATING FOR RISK MANAGEMENT USING ARTIFICIAL NEURAL NETWORKS

Annie R. Pearce[1], Rita A. Gregory[2], and Laura Williams[3]

ABSTRACT: This research developed a technique for generating range estimates to evaluate the risk of cost escalation in building construction projects, using an artificial neural network model of construction project costs. By identifying the risk of cost escalation during the planning and conceptual design of built facilities, facility owners and project managers can better focus their efforts to control total project cost in these times of decreasing capital investment budgets. The specific focus of this research was to develop a methodology for analyzing cost data from existing facilities to generate cost-probability curves and range estimates for facilities at the conceptual stage of design, that could be used in turn to identify critical variables to be managed for controlling project costs.

KEYWORDS: range estimating, cost-probability curves, cost escalation, risk, project cost control

THE PROBLEM OF MANAGING COST ESCALATION RISK IN CONSTRUCTION PROJECTS

As budgets for the construction of new facilities and the maintenance, decommissioning, or rehabilitation of existing facilities become more limited, stakeholders responsible for funding these activities

---

[1] Ph.D. Candidate, Civil & Environmental Engineering, Georgia Institute of Technology; Research Engineer, Electro-Optics, Environment, & Materials Laboratory, Safety, Health, & Environmental Technologies Division, Georgia Tech Research Institute, Atlanta, GA 30332-0837 USA (404) 894-8089 email: annie.pearce@gtri.gatech.edu
[2] Assistant Professor, School of Civil and Environmental Engineering, Construction Engineering and Management, Georgia Institute of Technology, Atlanta, GA 30332-0355 USA (404) 894-2242 email: rgregory@ce.gatech.edu
[3] M.S. Student, School of Civil and Environmental Engineering, Construction Engineering and Management, Georgia Institute of Technology, Atlanta, GA 30332-0355 USA

are seeking better ways to increase the accuracy of their project cost estimates, especially in the early phases of project planning and design. Being able to predict how much a project will cost is important not only to set sufficient funds aside to complete a project, but also to ensure that the project is not over-budgeted, potentially encouraging cost overruns or preventing other deserving projects from receiving adequate funding.

Alternative methods for cost prediction are especially important in the early planning or conceptual design phases of projects, before enough detail is known to allow a traditional quantity takeoff estimate to be performed. Since much of the allocation and/or budgeting processes often takes place during these early phases of a project, an accurate estimate is essential to ensure that projects are allocated a sufficient budget so that functional requirements are met. At the same time, an accurate project estimate can help to avoid over-budgeting that may encourage gold plating during design, or remove efficiency and performance constraints during construction. Accurate project cost estimates early in the planning and design processes can also serve as a cost-control measure to assist in managing the design process.

In this paper, we examine the potential of Artificial Neural Networks (ANNs) as a tool to support the tasks of cost prediction and risk management during the planning and design phases of the project life cycle. ANNs are a modeling tool based on the computational paradigm of the human brain, and have proven to be robust and reliable for tasks of prediction, ranking, classification, and interpretation or processing of data. We begin by examining the problem of project cost prediction in more detail, followed by the objectives of the research and a description of the methodology followed. The results of the research and a discussion of their implications are the primary contribution of this work. The paper concludes with a look at future research and applications that can stem from this proof of concept.

BACKGROUND

The problem of managing project costs is not new. The whole procurement paradigm of Project Management was created as a response to the need to ensure that projects are completed on time and within budget, to an acceptable standard of quality. However, this paradigm is typically most influential during the construction phase of a project life, whereas the greatest impact on final project cost can be achieved in the planning and design phases (Figure 1). In fact, most of the critical decisions affecting the total cost of a project are made during the project planning phase, before designers, project managers, and contractors join the project team (Hendrickson & Au 1989).

In today's cost-conscious project environment, project owners and planners need a way to predict how their early decisions will ultimately impact the final cost of a project. While this need has traditionally been addressed by heuristic or "rule-of-thumb" knowledge (e.g., "the larger the building perimeter, the greater the cost of exterior enclosure"), no quantitative method currently exists for understanding how planning choices affect final project costs. This lack of quantitative method is due primarily to the complexity of analyzing the multiple factors influencing final cost. With the multitude of interacting variables that potentially affect cost in the early planning stages of projects, performing a rigorous multivariate nonlinear regression to determine the relative importance of those variables is a nontrivial computational task.

Artificial Neural Networks (ANNs) offer an alternative to traditional methods of cost prediction based on parametric or quantity takeoff techniques. With their capacity to learn from examples and to generalize that knowledge to novel cases, ANNs provide the capability to undertake rapid modeling of systems in which the interaction between input and output variables is unknown but where representative examples of inputs and outputs exist (Wasserman 1989). ANN modeling of the process of project cost prediction provides potentially important clues to the relationships between initial planning-phase project variables and final cost. ANN models were used in this research as a quantitative approach to generating range estimates that can be used to manage project planning and design (Pearce 1997).

PROBLEM STATEMENT AND OBJECTIVES

Facility owners and construction project managers need a way to focus and prioritize their efforts to control project costs during the early phases of design and construction, when their efforts can have maximal impact on the total cost of the project. In the earliest phases of planning and design, only the most basic programming and functional decisions about the facility have been made, and the data available for predicting project costs is sparse and highly subject to change. Under these conditions, traditional methods for predicting the risk of project cost escalation, such as parametric, cost build-up, or expert system estimating, become inaccurate or impossible to implement. Stakeholders responsible for controlling project costs need an alternative to traditional cost-based risk prediction methods to help them predict how the costs of their projects may change as design and construction progresses, using the limited available data in the early phases of the project.

Based on these needs, the primary objective of the research was to develop and validate a prototype system to generate range cost estimates for projects at the conceptual design phase. The range estimating ANN model enables project planners to identify the potential for cost variation by the end of the project, thus facilitating the budgeting process. Given this objective, the following section presents the methodology used to implement the research.

RESEARCH METHODOLOGY AND OUTCOMES

To build a model for estimating the risk of cost escalation for construction projects during conceptual design, we elected to use an artificial neural network (ANN) to approximate the relationship between controllable design variables and anticipated range of final costs. ANNs are a modeling tool based loosely on the computational paradigm of the human brain (Wasserman 1989). We chose to use an ANN model to perform conceptual range estimating because of their proven track record in tasks of prediction, ranking, classification, and interpretation or processing of data (see Pearce 1997). The research focused on developing an ANN model to generate cost-probability curves and subsequent range estimates for a well-understood type of building, light commercial built facilities, as described further in the next section. The following sections describe how we scoped the range of possible facilities, generated training and test sets of data to develop the ANN model, selected ANN model parameters, and validated the resulting model.

Scoping the Range of Applicability for the Prototype Model

One of the first challenges of the research was designing a data set that was sufficiently well understood to develop and intuitively validate an ANN model of project cost. To address this challenge, we elected to develop a “clean” data set by simulating multivariate cost data using the Parametric Automated Cost Engineering System (PACES) for vertical construction, developed for the United States Air Force

(Williams 1997). The rationale behind this choice was to "prove" the accuracy and effectiveness of using ANNs to generate range estimates on a data set that was thoroughly understood (light commercial vertical construction) and "clean" (simulated). Having proved the concept on a well-understood data set, we can subsequently have confidence in applying the technique to more complex and less understood facilities. The approximation capabilities of the ANN model help to provide reasonably accurate model response for these situations, unlike other traditional conceptual cost estimating techniques (Gregory et al. in press).

The first phase of ANN model development involved scoping the work to narrow the population of construction projects to be considered in developing the ANN model. In order to demonstrate proof of concept, we sought to minimize the number of input variables considered, while still maintaining a sufficient degree of problem complexity to keep the relationships between inputs and outputs unpredictable using traditional regression techniques. After deciding to use simulated data to develop the ANN model, the next major challenge was to scope the set of variables considered in order to prove the model concept, while maintaining a sufficiently diverse population of simulated cases to test the performance of the ANN model.

We elected to limit our attention to vertical construction, since we had the Parametric Automated Cost Engineering System (PACES) model to use for scenario simulation. This cost estimating tool provides the capability to model total project costs for a variety of typical commercial building types, ranging from communication centers to medical facilities to living quarters (Burns 1997). Within each category of building types in PACES, building models based on functional space areas are included to estimate costs for specific building configurations and functional capabilities. We selected a specific facility type and functional model based on the availability of an existing database of real project data to be used for validation (ibid.). The type of facility we selected was dormitories, and the specific model we chose to simulate was for enlisted personnel dormitories in a "1+1" configuration, i.e., suites of two rooms, each housing one enlisted person, with shared bath and living areas. This type of dormitory comprises the majority of new military enlisted dormitory construction (ibid.), and thus has strong potential for initial research results to have an impact on future construction practice. The military enlisted "1+1" dormitory model provided in PACES is comparable to many dormitory designs also used in civilian commercial construction, such as the dormitories constructed for university campuses.

To further scope the problem, we chose to investigate various facility configurations for a constant number of inhabitants. In the PACES model, number of inhabitants drives the total floor area of the facility. By keeping the number of inhabitants constant, we limited the scope of variability to two main classes of variables (geometric variables and architectural variables), while keeping factors constant that are typically fixed early in the planning process. The parameters that were fixed over the sample set are shown in Table 1.

Generating a "Training Set" of Examples for Model Development

After the parameters discussed in the previous section were fixed, the next task of the research was to generate a set of samples by varying the remaining input variables in a way that is representative of the range of possible design and construction practices. All data sets for this research were simulated using the PACES cost estimating tool. Despite the relative savings in labor due to using simulated data instead of empirical data, developing cases using PACES is still quite labor-intensive. Thus, finding a modeling technique that works with representative rather than comprehensive data sets was an important consideration in selecting ANNs to model project costs. This consideration is also a very real concern for future application of the model to other types of projects using empirical data. Due to the massively interconnected and parallel nature of ANNs, a well-trained ANN model has the ability to generalize, i.e., provide reasonable outputs given a set of inputs to which it has not previously been exposed. This quality implies inversely that ANN models can be trained using sample sets which are not comprehensive but are instead representative, thus saving significant time and effort in simulating training cases.

Determining what cases should be simulated to generate a representative sample set was itself a nontrivial task, given the relatively large number of input variables, the large number of potential values for each variable, and the nonlinear interactions among input configurations. Based on informal interviews with cost engineering experts (Burns 1997, Gregory 1997, Rast 1997), we selected factors such as building perimeter length and floor-to-floor height as our input variables, which are significant controlling factors for the quantity of exterior enclosure required in a facility. Based on their field engineering experience, these experts believed that the type of cladding specified for a building has a great degree of influence on its final cost.

Table 2 shows the input variables that were considered as candidates to generate the sample set used in this research, along with possible values for each variable. While many of the input variables in isolation had a linear relationship to total project cost, in combination the variables interacted to result in a complex nonlinear cost function. Thus, significant effort was required to determine what values should be used for each variable to generate a representative sample set, while at the same time minimizing the number of scenarios that had to be generated using the PACES model.

As can be seen by inspecting Table 2, the number of possible permutations in a comprehensive data set is quite large, especially depending on the increments of sampling for continuous variables such as building perimeter or floor to floor height. Thus, determining a sampling strategy that would minimize the number of cases to be simulated was an extremely important task.

The strategy used to select configurations of input variables to represent the population of potential projects was based on empirical determination of the mathematical relationship each variable had on the total project costs. Using the PACES model, we used an incremental strategy to simulate cases at the maximum, minimum, and midpoint expected values for each variable, and plotted the relationship of these values against total project cost. If the relationship was linear, we selected a minimal number of cases across values for that variable to represent the data set. If the relationship was nonlinear, we continued to simulate cases at the midpoint of each range of values until we could pinpoint critical thresholds or discontinuities in the relationship between the variable and total cost. Additional training cases were simulated for these nonlinear variables to ensure that the ANN model could “understand” the discontinuities in these relationships.

Figures 2 and 3 show samples of plots we generated using the PACES model to determine the relationships between final cost and input parameter variation. Figure 2 shows the nonlinear effect on cost by Uniformat category of varying the number of stories with a fixed gross square footage of the building.

Figure 3 shows a composite analysis of total project cost while varying both floor-to-floor height and the number of stories, again with a fixed gross square footage for the building. As with Figure 2, the effects of changing these variables are nonlinear with respect to total project cost, and thus more than two samples are required for each parameter to create a representative sample set.

The final set of parameters used to generate the representative sample set for model development is shown in Table 3. Varying the input parameters shown in Table 3 resulted in a total of 46 training cases as indicated by X’s in the matrix. Five additional randomly selected cases were developed to serve as test cases for assessing the model’s performance with novel inputs, indicated by T's in Table 3.

Selecting ANN Model Parameters

One of the first tasks in developing an ANN model was to determine an acceptable threshold for error in output. An ANN model can be manipulated in many ways to improve its performance, including varying its internal architecture, learning paradigm or parameters, or modifying the data set used to “train” it. Given the large number of possible network configurations, selecting an acceptable level of error is important to scope the process of network experimentation. Upon developing a network that performs within an acceptable level of error, further experimentation to improve accuracy is unnecessary. To select a threshold of acceptable error for the cost estimation problem, we used a range of acceptability of {+25%, -10%}. The upper limit was based on the cost variation authorized by the U.S. Congress for military construction projects (USC 1995), while the lower limit was based on standard industry practice for vertical construction projects (Rast 1997, Gregory 1997). Thus, an ANN model which could predict direct project costs within {+25%, -10%} was considered acceptable for the purposes of this research.

The next steps in developing an ANN model to predict project costs were to select a network paradigm, and to transform the data from the PACES simulations into a form that could be fed to the network. The back-propagation paradigm of ANNs was selected, due to its demonstrated accuracy in problems of prediction and the transparency of logic underlying the theory of back-propagation neural networks (NeuralWare 1996). Transformation of data from the PACES simulations involved “squashing” each value for input and output variables to lie between {0, 1}. Squashing functions are mathematical transforms used to scale or “squash” one value between some minimum and maximum into another value between different, more convenient minimum and maximum limits. The mathematical role of the squashing function is analogous to an analog electronic amplifier, where the strength of an electronic signal is boosted if it is weak or reduced if it is strong (Eberhart & Dobbins 1990). Each value was squashed using a linear compression formula:

$$X_{squashed} = (X_{original} - X_{min}) / (X_{max} - X_{min})$$

where $X_{squashed}$ = squashed value for variable X

$X_{original}$ = original value for variable X

X_{min} = minimum value over all instances of variable X

X_{max} = maximum value over all instances of variable X

In addition to the input variables shown in Table 3, the PACES simulator automatically varied other quantitative parameters based on initial input values. The total set of twelve variables used as input to the model is shown in Table 4.

Two different configurations of outputs were considered: a single total direct cost output, and a direct cost split by Uniformat categories. The second configuration was selected, resulting in an initial set of data with fifteen outputs per case – one for each Uniformat category (see Table 5). Eight of these initial variables were excluded from the transformed data, since they did not vary substantially as a result of manipulating input variables. The seven output variables used to train the network are indicated by "yes" in Table 5.

Training, Testing, and Validating the ANN Model

After configuring the 46 training cases to serve as network input, the next step was to begin experimentation with back-propagation ANN models to obtain the required level of accuracy in predicting project direct costs. The aforementioned error thresholds were used to evaluate network performance by summing the outputs of all Uniformat subsystems and comparing the total to the PACES predicted total cost for these subsystems.

Despite many advances in the theory of ANNs, choosing an appropriate network paradigm and architecture is still largely art rather than science. Various configurations of numbers of processing elements and hidden layers were tried in the general class of back-propagation ANNs. Back-propagation is one paradigm for network learning that involves changing connection weights between hidden units based on the contribution each has made toward generating an erroneous output during training (see Wasserman 1989, NeuralWare 1996 for more explanation). The Delta learning rule (ibid.) was selected to govern the ANN training, since it resulted in the best performance during experimentation. In addition to experimentation by trial and error with various network architectures, the cascade correlation algorithm (Feldman & Lebiere 1993) for developing ANN architectures was used to guide the trial and error, using the Predict expert system ANN shell by NeuralWorks.

Various ANN architectures were tried, ranging from one hidden layer consisting of 0 processing elements to three hidden layers consisting of a total of 21 processing elements. The tested architectures, along with various error metrics, are shown in Table 6 and illustrated graphically in Figure 4.

Networks were trained for 50,000 passes of the training set, and used a learning coefficient of 0.3 with a sigmoid transfer function. The best performing network had an architecture of 12 input units, 7 output units, and one hidden layer with five processing elements. This network had an average error of –3.79% and a maximum error of –7.22% of in predicting total direct project costs over the test set.

While all ANN models seemed inclined to underestimate total direct costs on average, in fact many of the test cases resulted in slight overestimation by the models. The negative average error resulted in many cases from the fact that the models had difficulty in extrapolating to outlying cases, of which there was one in the test set. Despite this difficulty, even the worst-performing network had a maximum error of –13.74%, which is not far from the maximum industry-accepted threshold of –10%. Additional training and manipulation of ANN parameters is likely to improve even the worst-performing network to acceptable levels of performance.

Generating Project Range Estimates and Cost/Probability Functions Using the ANN Model

After an acceptably trained ANN model was developed, the next step was to use the model to generate cost/probability functions over a comprehensive set of input variable permutations. While the PACES model or some other cost estimating tool could be used to perform this task, the lack of an automated scenario generation capacity made impossible the task of running a comprehensive set of input variables to generate corresponding outputs. Given an acceptably performing ANN model with batch capabilities, running a large set of neural net cases was a simple computational task.

The outcome of this phase of the research was a cost/probability function to illustrate the potential range of project costs over all permutations of the input variables. After the shape of the cost/probability function is known, variations in final cost can be predicted based on the shape of the curve and the desired level of confidence required for the cost estimate. To generate the cost/probability function, a comprehensive set of input permutations was generated, with resolution of input variables increasing about their expected value as delineated in the PACES model. Resolution of input variables was increased over three independent variables: number of stories, building perimeter, and floor-to-floor height. By increasing the resolution of these variables, a larger number of cases was generated about the expected value of the variables, resulting in a distribution more closely resembling the existing population of dormitories. Table 7 shows the input values used to generate the permutations comprising the comprehensive data set. Expected values for each parameter are indicated by a double-bar outline. The remaining input variables (footprint, roof area, exterior wall area, exterior window area, exterior doors, number of stairwells, number of elevators, heating load, and cooling load) were calculated using the PACES equations that depend on the three independent parameters.

After the comprehensive data set was constructed and squashed, the best-performing ANN model was used to generate output values for each of the comprehensive data set cases. The resulting outputs were unsquashed and plotted as a histogram to generate an approximation of a cost/probability curve. The histogram is illustrated in Figure 5. A cumulative frequency distribution of the ANN outputs is shown in Figure 6.

The total direct cost as shown in these plots represents the sum of the ANN model's prediction for the seven Uniformat categories exhibiting variation as a result of the manipulation of input variables. Total direct cost for all Uniformat system categories can be obtained by adding $2,225,831 to each value, to account for the missing eight Uniformat categories which remained relatively constant over all samples (see Table 5).

APPLYING THE ANN MODEL TO GENERATE RANGE ESTIMATES FOR CONSTRUCTION PROJECTS

In applying the ANN model to the problem of generating range estimates, we propose four ways to represent the outputs of the model to increase its utility. In the previous section, we demonstrated one of these four ways (plotting cost-probability curves) to generate a range estimate. The following subsections describe this representation in comparison to the other three representation strategies. Having demonstrated the viability of ANN modeling to generate estimates in cost-probability form, we hypothesize that the ANN approach could work equally well for representation in the other methods.

Representing Cost-Probability Functions Using Histograms

The representation of predicted ranges of cost for a given project type used in this research was a combination of two types of curves: a histogram of costs over the whole range of possible combinations of controlling variables, and the associated curve describing probability of cost exceeding a given threshold. Figures 5 and 6 show these curves; specific range estimates can be derived from the curves based on the confidence with which the user wishes to specify the range.

Expected Value Plus Two-Point Range Representation

A second way to represent range estimates using the ANN model would be to train the ANN to respond with a single-point estimate of total project cost and a two-point range estimate of total project cost. Single-point estimates could be used directly (after normalization) from the PACES output of total project cost, or the corresponding real total project cost from a similar historical project, if such data were available. Two-point range estimates could be generated by varying the single-point PACES estimate by +50% and –25% and normalizing the resulting figures. The +50% and –25% ranges are typical of variability found in early conceptual cost estimating (Gregory 1992). Since the data set used to train the ANN model would be developed as a linear function of the point estimate used in our model, our proof-of-concept described in this paper is also a proof for the expected value plus two-point range estimate.

Representing Cost-Probability Functions Using Curve Function Parameters

A third possible configuration of outputs would be to represent potential risk and project cost curves in terms of probability function curve parameters. One method, approximation of curves using equation parameters, requires the ANN model to output curve coefficients and intersects according to the equation:

$$y = c_1x^n + c_2x^{n-1} + \ldots + c_nx + b$$

A second method, approximation of output curves using probability distribution parameters, requires assuming a probability distribution of the population of potential outputs, then using the descriptive parameters of the assumed distribution as outputs for the I/O pairs. The normal distribution and the Beta distribution are examples of commonly used probability distributions that could fit the cost-probability functions found in construction estimating.

Representation of One-of-N Classes of Possible Costs

A fourth method involves representing the desired output curve as a discrete number of ranges, each of which is represented in the ANN model by one output unit. ANN models could be trained by converting PACES single-point cost outputs into 1-of-N vectors and m-of-n vectors, where a 1-of-N vector consisted of a value of one for the output range containing the point estimate and zeros for all other vector values. The m-of-n vectors use a curve approximation to generate a normalized range of values for the vector components, representing a histogram of values centered about the single-point cost generated by the PACES model. As with the curve parameters approach, both normal and Beta distributions could be used to simulate the histograms.

IMPACTS, APPLICABILITY, AND VALIDITY OF THE RESEARCH

The benefits of being able to predict final project costs early in the project life cycle are significant. Accurate project cost predictions or estimates early in the planning and design processes can be used as a cost-control measure to assist in managing the design process. With active management during design, conflicts or cost overruns can be identified as they occur, and changed if necessary before design progresses. Significant areas of cost-escalation risk can be identified and addressed proactively during design, rather than after construction begins and budgeted dollars begin to fall short of requirements.

The research results presented in the previous sections lead to two application-related questions. First, what are the potential impacts of providing knowledge to project stakeholders about cost drivers and cost-related risk? Second, how applicable is the methodology developed in this research to other types of facilities besides 1+1 dormitories, and what are the issues relating to validation of the methodology? The following sections address each of these questions in turn.

Potential Impacts of Cost Driver and Risk Information on the A/E/C Industry

Knowledge of the cost/probability function will enable project stakeholders to estimate the risk of cost variance from the initial conceptual estimate, facilitating the budgeting process and helping to encourage "tighter" design control when the risk of cost variance is unacceptably high. Project cost control tools such as the PANNCET approach can help to improve project cost performance by providing quantitative data previously available to project stakeholders only through experiential heuristics.

Applicability and Validity of the Cost Driver Extraction Methodology

While longitudinal field validation of the results of this research was outside the scope of this research, initial heuristic validation of the research outcomes supports the methodology underlying the process of cost driver extraction. Although this proof of concept was limited in scope to one specific type of vertically-constructed facility within the entire range of construction projects, the methodology is applicable to other types of construction projects within the sector of vertical construction, as well as to other project sectors.

FUTURE RESEARCH

Three areas of additional work can stem from this research, including continued refinement of ANN performance beyond that achieved in this proof of concept, generalization of the range estimating methodology to other types of construction projects, and the development of a project-specific methodology for developing and utilizing the salient characteristics of range estimates.

Continued Refinement of ANN Performance

The first area of future research is to continue to experiment with additional configurations of ANN parameters and architecture, along with additional research into developing representative sample sets using clean, simulated data. While the purpose of this research was to demonstrate the concept of range estimating using ANNs, additional refinement and development of theory could lend substantial benefit to this work. Future research will also include testing of ANN models and range estimating techniques using real data, and experimentation with other representations of input and output data.

Generalization of Approach to Other Project Types

More research is needed to test the ANN approach to range estimation in other project types besides dormitories. Although significant differences are not anticipated for similar commercial type buildings, examples from other fields of construction such as industrial plants may pose a different challenge due to the uniqueness of each facility.

Development of Project-Specific Range Estimating Tool

One promising application of the ANN concept to project risk management is the potential to generate specific range estimates for projects at the pre-design phase of work. While many project managers rely on years of expert experience to know how to calculate a range estimate, the development of knowledge-based systems for these applications has been previously unexplored. Using ANNs to undertake the calculation of project-specific range estimates is a promising application for ANNs, extending the capabilities of generalized data set characterization and modeling for generic construction projects exploited in this research.

ACKNOWLEDGMENTS

This research was sponsored by the National Science Foundation (Grant 970-1923), the Air Force Office of Scientific Research, Bolling AFB, DC, and the Air Force Wright Laboratory Air Base Technology Branch (WL/FIVC-OL Tyndall AFB, FL) as part of the Summer Faculty and Graduate Student Research Programs. Grateful acknowledgment for technical and facilities support is given to Delta Technologies Group, Inc., owner of the commercial license for PACES software. Thanks also to the Georgia Tech Research Institute and the Georgia Tech Colleges of Engineering and Architecture for their financial support of this research, and to NeuralWare, Inc., owner of the commercial license for the NeuralWorks Professional II software used in this research.

APPENDIX. REFERENCES

Burns, T. J. (1997). Deputy Director, Construction Cost Management, HQ AFCESA/DC, Tyndall AFB. Informal interview. July 16.

Eberhart, R.C. and Dobbins, R.W. (1990). *Neural Network PC Tools: A Practical Guide*. Academic Press, San Diego.

Feldman, S.E., and Lebiere, C. (1988). "The Cascade-Correlation Learning Architecture," *Advances in Neural Information Processing Systems, v. 2*, Morgan Kaufmann, New York.

Gregory, R. A. (1992). "Development of a knowledge-based system approach for decision making in construction projects," Unpublished doctoral dissertation, University of Florida, Gainesville, FL.

Gregory, R.A. (1997). Assistant Professor of Civil & Environmental Engineering, Georgia Institute of Technology. Informal interview. July 31.

Gregory, R.A., Pearce, A.R. and Williams, L. (in press). "Cost Effective Allocation of Research Funds for the Development of Emerging Technologies," *Journal of Professional Issues in Engineering Education and Practice*.

Hendrickson, C.T. and Au, T. (1989). *Project Management for Construction*. McGraw-Hill, New York.

NeuralWare. (1996). *Neural Computing*. NeuralWare Technical Publications Group, Pittsburgh, PA.

Pearce, A. 1997. "Cost-Based Risk Predictions and Identification of Project Cost Drivers Using Artificial Neural Networks," AF/OSR Summer Research Program Technical Report (97-0415).

Rast, R.R. (1997). CEO, Delta Technologies Group, Inc. Informal interview, July 31.

USC. (1995). 104th Congress, 1st Session, No. 2, Title 10, USCode Armed Forces as amended through 12/31/94. Committee on National Security, HoR, March 1995. Page 1047, Section 2853, Authorized cost variations. Ch. 169 – Military Construction and Military Family Housing.

Wasserman, P.D. (1989). *Neural Computing: Theory and Practice*. Van Nostrand Reinhold, New York.

Williams, L. 1997. "Data Simulation Supporting Range Estimating for Research and Development Alternatives," AF/OSR Summer Research Program Technical Report (97-0416).

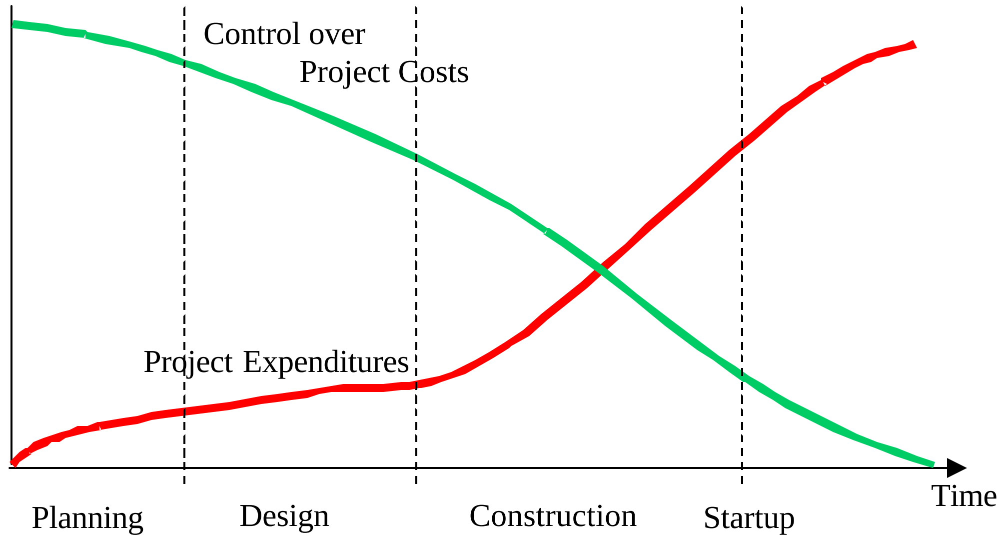

Figure 1: Cost and Control vs. Time for a Construction Project (Hendrickson & Au 1989)

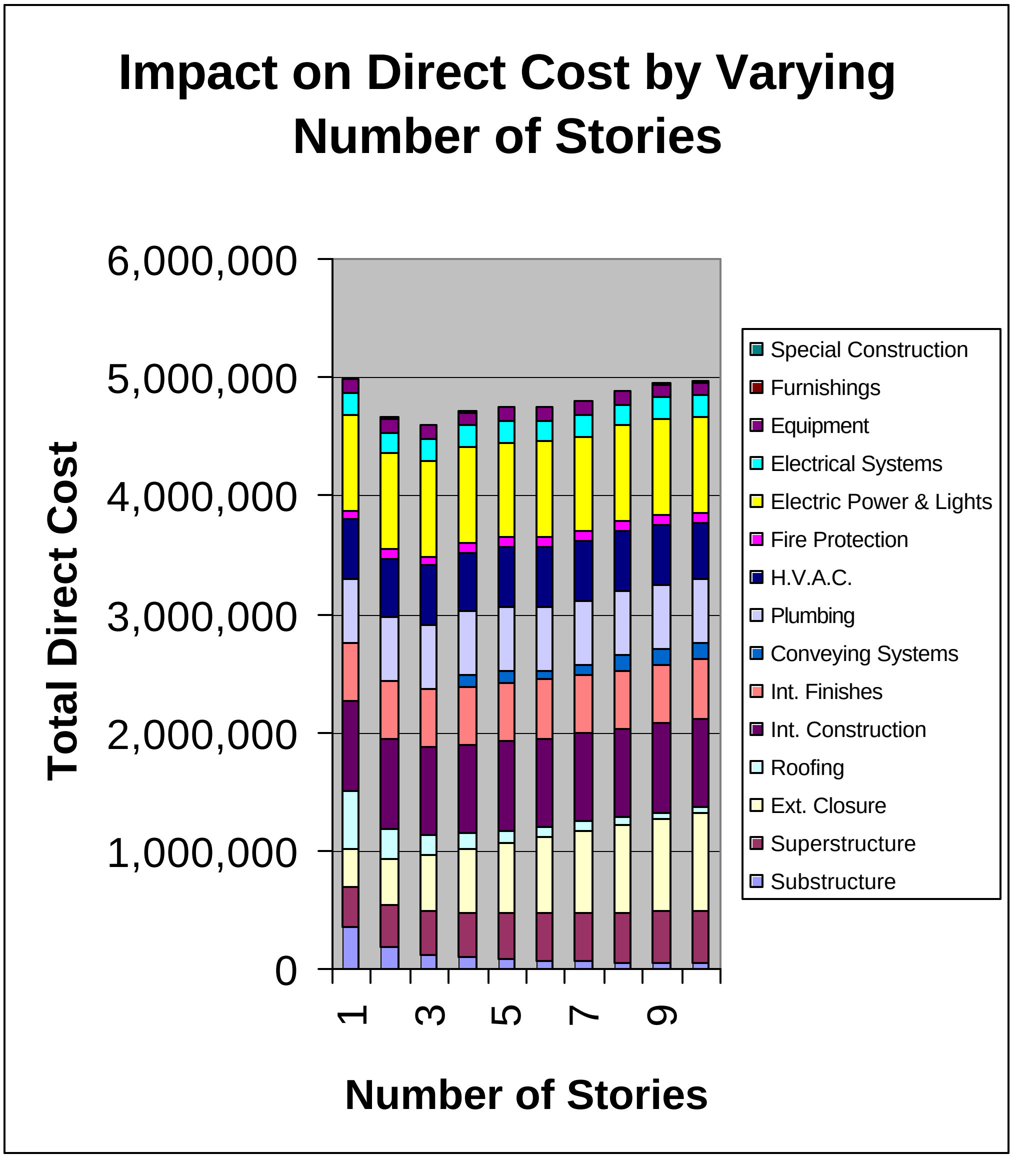

Figure 2: Dormitory Cost Variation for Various Numbers of Stories

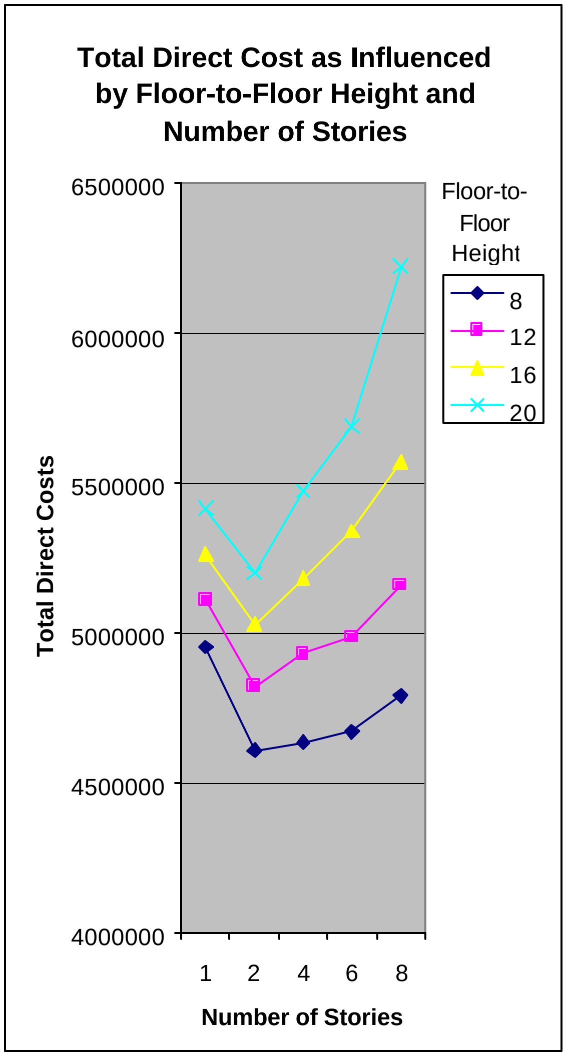

Figure 3: Cost Variations with Change in Floor-to-Floor Height + Number of Stories

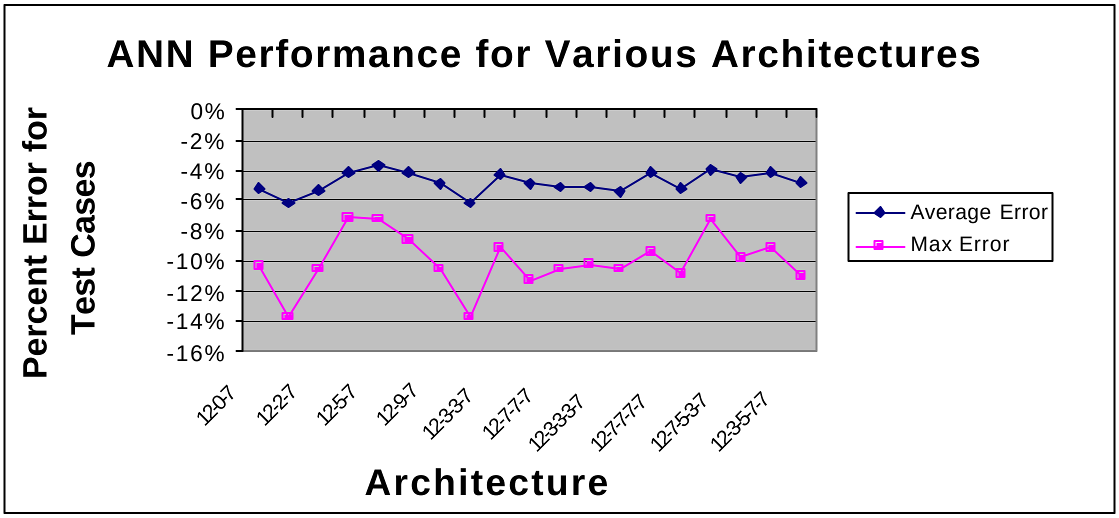

Figure 4: ANN Architectures vs. Error

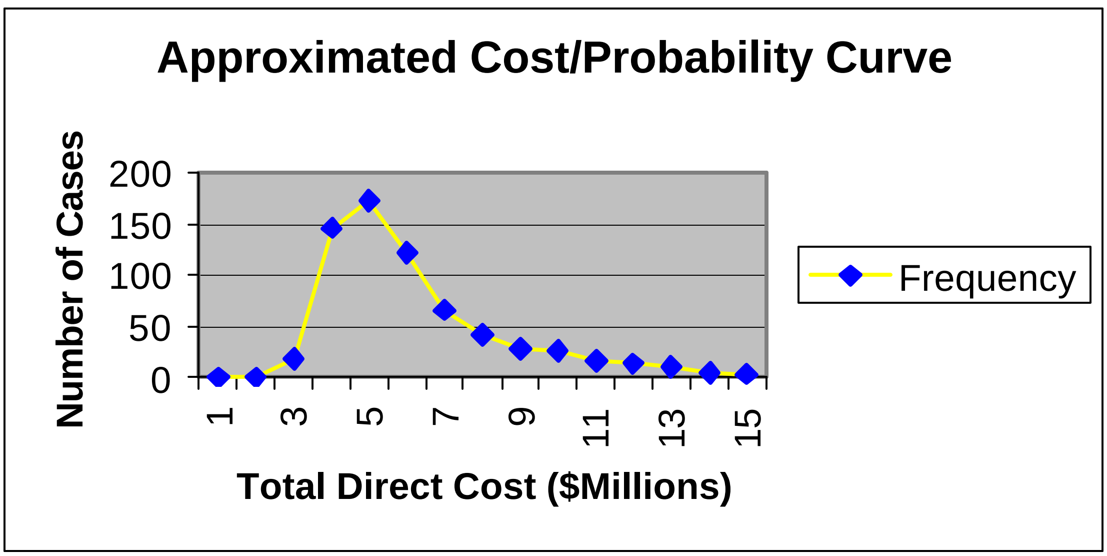

Figure 5: Frequency Distribution of Total Project Costs

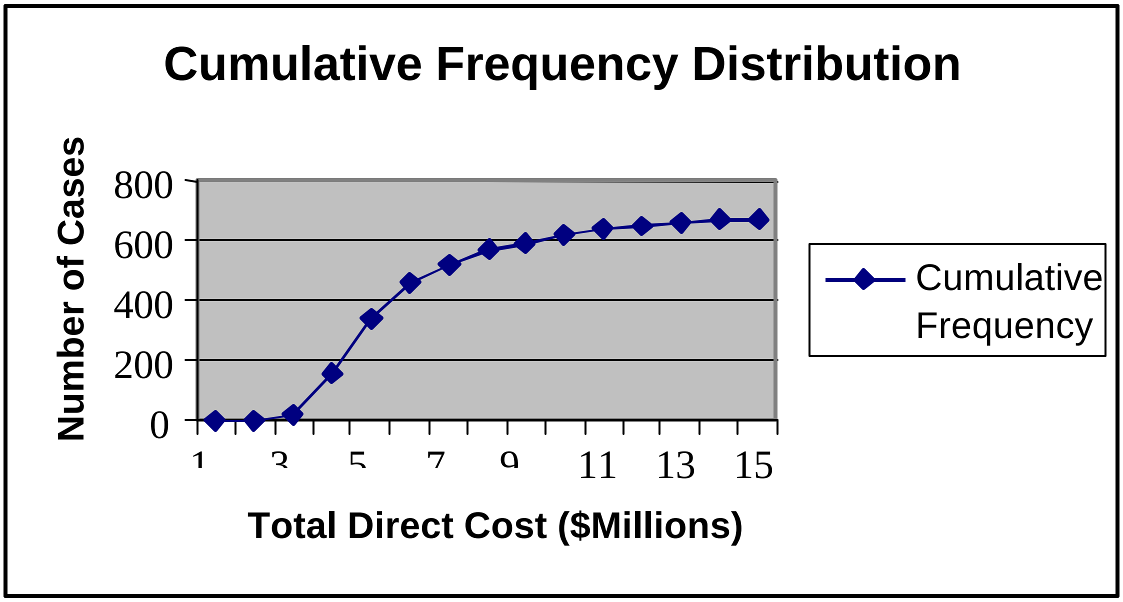

Figure 6: Cumulative Frequency Distribution of Total Project Costs

Table 1: Parameters Fixed for Problem Scoping

Parameter	Value	Comments
Facility Type	Dormitory	Selected due to availability of real cases for validation.
Building Model	Enlisted “1+1”	Greatest number of empirical cases; generalizability to civilian construction.
Location	Atlanta, GA	The index city for the cost database. Fixed since location is typically known early in the planning phase. Affects seismic and foundation conditions, and heating and cooling loads.
Number of Persons	156	Fixed to maintain constant building square footage; typically known early in the planning phase. This is the default number of persons for the PACES Dormitory model.

Table 2: Selected PACES Parameters and Possible Values

Parameter	Possible Values
Stories Above Grade	0 – 10
Perimeter	50 – 5,000 Sq. Ft.
Floor to Floor Height	0 – 50 Ft.
Floor to Ceiling Height	0 – 50 Ft.
Soil Bearing Capacity	Low, Average, High
Floor Structure Type	Concrete Frame Steel Frame with Reinforced Concrete Deck Steel Frame with Metal Joists/Steel Deck/Concrete Fill Load Bearing Walls with Metal Joists/Steel Deck/Concrete Fill Load Bearing Walls with Wood Joists/Wood Deck Load Bearing Walls with Precast/Prestressed Floors
Bay Size/Span Length	Small, Average, Large
Roofing Type	Single Membrane Built-Up Shingle Standing Seam Metal Clay Tile Metal (Typical Metal Building)
Exterior Wall Type	Brick Veneer 4” Split Rib Masonry Veneer 8” Split Rib Masonry 8” Masonry Block Tilt-up Concrete Exposed Aggregate Precast 12” CIP Concrete with Exposed Aggregate Finish Metal Sandwich Panel Stucco E.I.F.S. (Dryvit)

Table 3: Parameter Values Used in Sample Set

Floors Above Grade		1	2	3	4	5	6	7	8	9	10
Perimeter	Default	X	X	X	X	X	X	X	X	X	X
	2000 LF	X	X	X							
	3000 LF	X	X	X			T				
	4000 LF	X	X	X							
Floor to floor height	8 FT	X	X	X	X		X		X		
	10 FT			X				T			
	12 FT	X	X	X	X		X		X		
	14 FT	T		X							
	16 FT	X	X	X	X		X		X		T
	18 FT			X		T					
	20 FT	X	X	X	X		X		X		

Table 4: Input Variables for the ANN Model

Input Variable	Unit	Minimum Value	Maximum Value
Floors Above Grade	EA	1	10
Perimeter	LF	604	4000
Footprint	SF	6313	62316
Roof Area	SF	6682	65961
Floor to Floor Height	LF	8	20
Exterior Wall Area	SF	16308	129744
Exterior Window Area	SF	1201	9557
Exterior Doors	EA	4	41
Number of Stairwells	EA	0	4
Number of Elevators	EA	0	2
Heating Load	MBH	649	1137
Cooling Load	Tons	77.62	123.91

Table 5: Output Variables for the ANN Model

Uniformat Category	Used in Model?
Substructure	yes
Superstructure	yes
Exterior Closure	yes
Roofing	yes
Interior Construction	yes
Interior Finishes	no
Conveying Systems	yes
Plumbing	no
H.V.A.C.	yes
Fire Protection Systems	no
Electric Power and Light	no
Electrical Systems	no
Equipment	no
Furnishings	no
Special Construction	no

Table 6: Network Architectures and Errors

Network Architecture (input-hidden-output)	**Maximum % Error** (% of predicted total cost)	**Average Error** (% of predicted total cost)
12-0-7	-10.31%	-5.21%
12-1-7	-13.72%	-6.18%
12-2-7	-10.60%	-5.31%
12-3-7	-7.16%	-4.15%
12-5-7	-7.22%	-3.79%
12-7-7	-8.67%	-4.11%
12-9-7	-10.56%	-4.90%
12-1-1-7	-13.74%	-6.21%
12-3-3-7	-9.19%	-4.27%
12-5-5-7	-11.26%	-4.94%
12-7-7-7	-10.59%	-5.17%
12-1-1-1-7	-10.19%	-5.17%
12-3-3-3-7	-10.55%	-5.49%
12-5-5-5-7	-9.41%	-4.20%
12-7-7-7-7	-10.86%	-5.18%
12-5-3-1-7	-7.28%	-3.96%
12-7-5-3-7	-9.78%	-4.43%
12-1-3-5-7	-9.13%	-4.20%
12-3-5-7-7	-11.00%	-4.85%

Table 7: Values for Independent Variables Used to Generate Comp rehensive Data Set

Floors Above Grade		1	2	3	4	5	6	8	10
Perimeter	Default	X	X	X	X	X	X	X	X
	1000 LF	X	X	X	X	X	X	X	X
	1250 LF	X	X	X	X	X	X	X	X
	1500 LF	X	X	X	X	X	X	X	X
	2000 LF	X	X	X	X	X	X	X	X
	3000 LF	X	X	X	X	X	X	X	X
	4000 LF	X	X	X	X	X	X	X	X
Floor to floor height	8 FT	X	X	X	X	X	X	X	X
	8.5 FT	X	X	X	X	X	X	X	X
	9 FT	X	X	X	X	X	X	X	X
	9.5 FT	X	X	X	X	X	X	X	X
	10 FT	X	X	X	X	X	X	X	X
	10.5 FT	X	X	X	X	X	X	X	X
	11 FT	X	X	X	X	X	X	X	X
	12 FT	X	X	X	X	X	X	X	X
	14 FT	X	X	X	X	X	X	X	X
	16 FT	X	X	X	X	X	X	X	X
	18 FT	X	X	X	X	X	X	X	X
	20 FT	X	X	X	X	X	X	X	X