

## CONCEPTUAL COST ESTIMATIONS USING NEURO-FUZZY AND MULTI-FACTOR EVALUATION METHODS FOR BUILDING PROJECTS

Wei-Chih WANG, Tymur BILOZEROV, Ren-Jye DZENG, Fan-Yi HSIAO, Kun-Chi WANG

*Department of Civil Engineering, National Chiao Tung University, 1001, University Road, Hsin-Chu, Taiwan*

Received 20 Oct 2013; accepted 30 May 2014

**Abstract.** During the conceptual phase of a construction project, numerous uncertainties make accurate cost estimation challenging. This work develops a new model to calculate conceptual costs of building projects for effective cost control. The proposed model integrates four mathematical techniques (sub-models), namely, (1) the component ratios sub-model, (2) fuzzy adaptive learning control network (FALCON) and fast messy genetic algorithm (fmGA) based sub-model, (3) regression sub-model, and (4) multi-factor evaluation sub-model. While the FALCON- and fmGA-based sub-model trains the historical cost data, three other sub-models assess the inputs systematically to estimate the cost of a new project. This study also closely examines the behavior of the proposed model by evaluating two modified models without considering fmGA and undertaking sensitivity analysis. Evaluation results indicate that, with the ability to more thoroughly respond to the project characteristics, the proposed model has a high probability of increasing estimation accuracies more than the three conventional methods, i.e., average unit cost, component ratios, and linear regression methods.

**Keywords:** conceptual cost estimation, component ratios method, fuzzy adaptive learning control network, fast messy genetic algorithm, regression method, multi-factor evaluations, building project.

### Introduction

Accurate cost estimation is a challenge for the project estimator during the conceptual phase of a construction project. Early cost estimates are necessary to compare design alternatives and select the most economical technical solution (Petroutsatou *et al.* 2012). An excessively low estimate can result in project overruns, while an excessively high estimate results in insufficient budget to address other critical user needs (Oberlender, Trost 2001; Wen 2010; Asmar *et al.* 2011).

This study defines the conceptual phase as the phase at which approximately 30% of the design is completed, and at which cost estimate is generally considered a budget estimate or baseline cost for a construction project (Lai *et al.* 2008; Cheng *et al.* 2010; Petroutsatou *et al.* 2012). Furthermore, this estimation challenge is difficult because only conceptual design drawings and specifications are available, and the estimations involve numerous assumptions (Asmar *et al.* 2011).

The construction industry has used several conventional conceptual cost estimation methods, such as average unit cost, cost indices, cost-capacity factors, and parametric estimation, to rapidly calculate total project cost (Barrie, Paulson 1992; Hendrickson, Au 2003; Hong *et al.* 2011). These early cost estimations at the conceptual phase are reasonably precise  $\pm 25\%$  (Petroutsatou *et al.*

2012). However, their accuracy must be increased to support an improved budget control (Wen 2010). Many project estimators thus would rather use a complex semi-detailed estimation method (i.e., a method for breaking down project cost into cost items as much as possible according to the available data) to enhance the estimation accuracy instead of applying the above conventional methods. Unfortunately, using the semi-detailed method is comparatively time-consuming, and may be ineffective in the early project phase, during which design changes frequently arise.

Since an easy-and-quick conventional method cannot provide reliable cost estimations, numerous more complex models for conceptual cost estimation have been designed, such as artificial intelligence techniques, statistical based analysis techniques (Lowe *et al.* 2006; Trost, Oberlender 2003), and case-based reasoning techniques (Chou 2009; Koo *et al.* 2011; Hong *et al.* 2011; Jin *et al.* 2012; Kim 2013). Because of the uniqueness of each construction project, a generalized conceptual cost estimation method is not readily available for all construction projects. Additionally, a proposed conceptual cost estimation method should respond to the characteristics of a new project to gain management confidences (Clark, Lorenzoni 1997). Related research thus strives continuously to develop feasible models to improve both

the evaluations of the cost effects of project characteristics and the early estimation accuracies for particular project types (Wen 2010).

Precise estimation of the project costs in the early phase must deal with three obstacles (Hsiao *et al.* 2012): (1) limited available data, (2) difficult to define relationships between the available data and total project costs, and (3) the need to capture as much as possible of the unique characteristics of a new project. To systematically deal with the three cost-estimation obstacles, this work proposes a new model that estimates building project conceptual costs using the component ratios, fuzzy adaptive learning control network (FALCON), fast messy genetic algorithm (fmGA), regression, and multi-factor evaluation methods.

## 1. Literature review

### 1.1. Textbook models for conceptual cost estimations

This section reviews the conventional conceptual cost estimation methods, including the average (project-level) unit cost, cost indices, cost-capacity factors, and parametric estimation methods. When the average (project-level) unit cost method is applied, the total cost of a project is the product of the average project-level unit cost and the total floor area of the project (Hendrickson, Au 2003). Cost indices focus on cost changes over time, while cost-capacity factors apply to changes in project size, scope, or the capacity of similar projects (Barrie, Paulson 1992).

The parametric estimation method adopts certain cost-relevant parameters (such as floor area, cubic volume, electricity generating capacity, steel production capacity, etc.) to describe a cost function in the screening estimate of a new facility (Hegazy, Ayed 1998; Hendrickson, Au 2003; PEH 2008). A cost function is developed according to one or more cost estimating relationships between the project cost (the dependent variables) and the cost-governing parameters (the independent variables) (Hegazy, Ayed 1998; PEH 2008). Several parametric estimation methods based on regression analysis and neural networks have also been suggested to increase the accuracy of conceptual cost estimates (Sonmez 2008; Gunduz *et al.* 2011). Notably, cost indices can also be incorporated into the above estimation method (Barrie, Paulson 1992).

Current textbook methods focus on total project cost, and usually do not examine cost divisions or cost item details. Although these methods can quickly obtain cost estimates by considering specific cost parameters, they often do not comprehensively deal with the three aforementioned obstacles to cost estimation.

### 1.2. Current research models for conceptual cost estimations

Regarding past research on cost estimation models, this study reviews artificial intelligence (AI) techniques. The AI-based models include Neural Networks (NNs) (Creese, Li 1995; Adeli, Wu 1998; Cheng *et al.* 2010; Petroutsa-

tu *et al.* 2012), Fuzzy Logic (FL) (Boussabaine, Elhag 1997), and Genetic Algorithms (GAs) (Kim *et al.* 2005).

For instance, Creese and Li (1995) designed a back-propagation NN application for timber bridge cost estimation. Moreover, Boussabaine and Elhag (1997) devised a neuro-fuzzy system related to FL for predicting construction project cost and duration. Meanwhile, Kim *et al.* (2005) established a cost approximation model for residential construction projects using GAs to optimize the parameters and weights of the back-propagation NN.

Yu and Lin (2006) combined NN and FL to develop a Variable Attribute Fuzzy Adaptive Learning Control Network (VaFALCON) able to handle cost estimation problems involving missing attributes. Cheng *et al.* (2009) integrated GAs, FL and NN technologies to establish a construction cost estimation model with high predictive power. Generally, NNs enable learning from past data and generalization of solutions for future applications; FL allows tolerance for real world imprecision and uncertainty; and GAs can be applied to globally optimize certain parameters (Cheng *et al.* 2009).

Based on the advantages of AI techniques, Hsiao *et al.* (2012) developed a neuro-fuzzy cost estimation model for semiconductor hookup construction. Their work combined the FALCON and fast messy genetic algorithm (fmGA) to build a training algorithm to deal with complicated relations among cost parameters in historical cost data, and applied a three-point cost estimation method to assess the uncertainties related to the inputs (for executing the training algorithm) in predicting new project cost. However, their study indicated that three-point estimation method they proposed to deal with uncertainties involved in semiconductor hookup construction projects may not be suitable for capturing the characteristics of building construction projects.

### 1.3. Related techniques used in the proposed model

This section reviews the techniques related to the proposed model, including the component ratios method, FALCON, and fmGA.

#### 1.3.1. Component ratios method

The component ratios method (also called the equipment installation cost ratios, plant cost ratios, or ratio estimating methods) assumes the existence of a ratio between major division costs and the total project cost (Barrie, Paulson 1992). Hence, when the major division costs and the ratio (= major division costs divided by total project cost based on historical data) are known, the total project cost can be calculated by dividing the major division costs by the ratio (smaller than 1.0). Focusing on certain major division costs can not only produce acceptably accurate estimates, but can also save estimation effort and time (Yu 2006).

#### 1.3.2. FALCON

FALCON is a neuro-fuzzy system that uses a learning algorithm derived from neural network theory to deter-

mine its parameters by processing data samples (Lin, Lee 1991). Initially, FALCON was designed to solve system control problems in electronics and manufacturing engineering (Lin, Lee 1991). FALCON is currently utilized to obtain construction knowledge owing to its diverse features, including the ability to handle uncertainties and its trace-back functions in problem solving (Yu, Lin 2006; Yu 2007).

Although FALCON has an excellent ability to learn from historical data, it also suffers a local optimality problem, namely, the solutions it obtains are optimal (either maximal or minimal) only within a neighboring set of solutions (Yu, Skibniewski 1999). A global optimum, which is the optimal solution among all possible solutions, is often preferred. Yu (2007) and Cheng *et al.* (2009) suggested adopting the messy GA (mGA) or fmGA to overcome such local optimal problems. Hsiao *et al.* (2012) applied the fmGA mutation and cut-splice operators to revise the fuzzy membership functions and fuzzy logic rules of FALCON to improve the cost estimation accuracy in semiconductor hookup construction projects. This investigation adopts the approach developed by Hsiao *et al.* (2012) to train historical cost data.

### 1.3.3. fmGA

GAs are search algorithms and search decision spaces for optimal solutions using methods based on natural selection and genetics (Holland 1975). Unlike simple GAs which use fixed length strings to represent possible solutions, Goldberg *et al.* (1993) developed the fmGA to apply messy chromosomes to form strings of various lengths. That fmGA allows variations in chromosome lengths benefits FALCON structural revision because the optimum precondition and consequence links structure for the fuzzy rule base may be obtained through fmGA evolution (Hsiao *et al.* 2012).

## 2. Costs of building construction projects

### 2.1. Levels of project costs

The costs of a building construction project are generally organized based on different levels of detail (Hendrickson, Au 2003). The highest level is the total project cost, while the second level is the cost division (or cost category) level, which summarizes the various cost divisions (for example, foundation and structure). The total cost of a second-level cost division itself is the sum of the costs of several third-level cost items. The cost of an item equals its unit cost (called “item-level” unit cost) multiplied by the item quantity, where the item-level unit cost is the fourth level cost. “Cost division” here means a second-level cost division.

### 2.2. Cost divisions for building projects

This work focuses on residential building projects involving reinforced concrete (RC). The construction cost of a residential housing project is represented by ten cost divisions: (1) foundation (represented by  $C_1$ ); (2) structure

( $C_2$ ); (3) external finishes ( $C_3$ ); (4) internal finishes ( $C_4$ ); (5) doors and windows ( $C_5$ ); (6) elevators ( $C_6$ ); (7) mechanical/ electrical/ plumbing (MEP) ( $C_7$ ); (8) temporary facilities ( $C_8$ ); (9) landscaping ( $C_9$ ); and (10) markup ( $C_{10}$ ). Notably, in this study, the cost division of temporary facilities ( $C_8$ ) includes several indirect costs, such as installing temporary water and electricity supply, field and home office overheads, insurance, inspection fees, and air-pollution control fees.

### 2.3. Total project cost

In this investigation, total project cost,  $C_{Tot}$  is derived by

$$C_{Tot} = (\text{construction cost}) \times (1 + \text{tax}) = (C_1 + \dots + C_i + C_8 + C_9 + C_{10}) \times (1 + t), \quad (1)$$

where  $C_1$ – $C_{10}$  are the costs of cost divisions (1)–(10), respectively. The value  $t$  denotes tax, which is a percentage (constant value, usually 5% in Taiwan) of the construction cost.

The cost of a cost division equals its “division-level” unit cost (that is, the cost required to complete a unit of work associated with a cost division) multiplied by the total floor area. That is:

$$C_i = U_i \times Q \quad \text{for } i = 1 \sim 10, \quad (2)$$

where  $U_i$  is the division-level unit cost of cost division  $i$ , and  $Q$  represents total floor area.

By integrating Eqn (2) into Eqn (1), Eqn (1) can be rewritten as:

$$C_{Tot} = (U_1Q + \dots + U_{10}Q) \times (1 + t). \quad (3)$$

Finally, Eqn (3) is rewritten as follows to reflect the effect of inflation on total construction costs:

$$C_{Tot} = (U_1Q + \dots + U_{10}Q) \left( \frac{CCI_{year}}{CCI_{base}} \right) \times (1 + t), \quad (4)$$

where  $CCI_{year}$  represents the construction cost index of the analysis year for a new project, and  $CCI_{base}$  denotes the construction cost index of the base year. In this study,  $CCI_{base}$  is set to 100 for the year 2006 based on current practice in Taiwan. Notably, Eqn (4) considers cost indices because it reflects cost changes over time.

Alternatively, if project-level unit cost ( $U_{total}$ ) of the total project (i.e., the cost required to complete a unit of work associated with a project) is available, then Eqn (4) can be rewritten as follows:

$$C_{Tot} = (U_{total} \times Q) \left( \frac{CCI_{year}}{CCI_{base}} \right) \times (1 + t). \quad (5)$$

### 2.4. Cost database of historical projects

Forty-six residential building projects located in northern Taiwan provide a historical database. All projects were

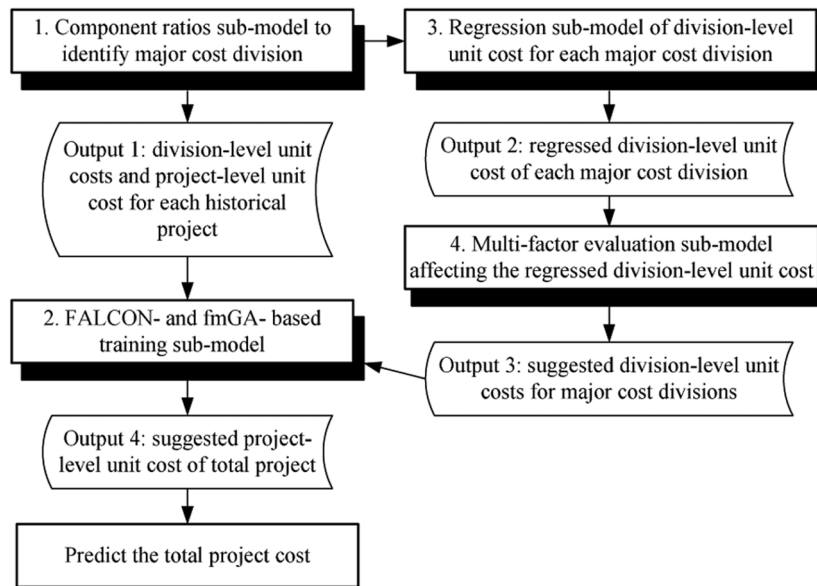


Fig. 1. Framework of the proposed model

completed by a single general contractor during 1991–2004. The projects had the following major characteristics: (1) all were RC structures; (2) average total cost (including markup) was about NT \$363,728,004 (roughly US \$12.12 million; US \$1 New Taiwan \$30); and (3) average total floor area was 12,615 m<sup>2</sup>.

### 3. Framework of the proposed model

Figure 1 displays the framework of the proposed model, which integrates four main sub-models. While the data of historical projects are trained using the FALCON- and fmGA-based sub-model, three other sub-models (i.e. component ratios, regression, and multi-factor evaluation) are designed to systematically guide cost estimators in order to assess the cost effects of the characteristics on estimating new project costs. The modelling steps are described as follows:

1. Step 1: During the conceptual phase of a building project, usually only limited data are available for new project estimation. Thus, using the component ratios method to focus on certain major cost divisions is suggested. This study uses 46 historical projects to indicate the major cost divisions of building projects. Furthermore, the division-level unit costs of the major cost divisions ( $U_i$ s) and the project-level unit cost ( $U_{total}$ ) for each historical project are identified in this step (i.e., output-1 in Fig. 1).
2. Step 2: The relationships between  $U_i$ s and  $U_{total}$  are complex. It is recommended that the FALCON-based training sub-model learn these relationships from historical projects. Since the FALCON operations may generate only a local optimum solution, the fmGA-based sub-model is used to optimize the FALCON solution to enhance the cost estimation accuracy. Following this step the training process for the proposed model is finished.

3. Step 3: New project estimation should reflect the characteristics of a new building project. A regression sub-model (built on the 46 historical projects) is proposed to determine the regressed  $U_i$  ( $RU_i$ ; output-2 in Fig. 1) of each major cost division of a new project where the total project floor area ( $Q$ ) has been predicted.
4. Step 4: A multi-factor evaluation sub-model is utilized to assess the effects of the project characteristics on the  $RU_i$  of each major cost division. A suggested  $U_i$  ( $SU_i$ ; output 3 in Fig. 1) thus is generated for each major division.
5. The  $SU_i$ s of the major cost divisions are then treated as inputs of the trained FALCON- and fmGA-based sub-model to forecast  $U_{total}$  of the total project (namely, output 4 in Fig. 1). Finally, the total project cost ( $C_{Tot}$ ) can be obtained using Eqn (5) since  $U_{total}$  is available.

The proposed model significantly differs from the neuro-fuzzy cost estimation model of Hsiao *et al.* (2012) in that the former applies the regression method and multi-factor evaluation method (rather than three-point cost estimation method used in Hsiao *et al.* 2012) to evaluate the cost effects of project characteristics for a new project. Using the multi-factor evaluation method facilitates the detailed assessment of project features to gain additional management confidence.

## 4. Details of the proposed model

### 4.1. Component ratios sub-model to identify major cost divisions

Table 1 lists the average cost and percentage contributions of individual cost divisions to total cost for 46 historical building projects. Notably, the costs are given in New Taiwan dollars. Based on the component ratios method,

Table 1. Average and percentage costs of the cost divisions in 46 historical projects

Cost divisions	Average costs (NT \$)	Percentage of cost
1. Foundation	32,269,583	9.32%
2. Structure	99,293,322	28.66%
3. External finishes	17,449,837	5.04%
4. Internal finishes	66,417,878	19.17%
5. Windows	18,962,967	5.47%
6. MEP	48,048,617	13.87%
7. Elevator	8,605,255	2.48%
8. Temporary facilities	22,327,687	6.45%
9. Landscaping	9,456,743	2.73%
10. Markup	23,575,735	6.81%

four cost divisions are identified to have the highest cost percentages, namely 28.66%, 19.17%, 13.87% and 9.32% for the structure, internal finishes, MEP and foundation divisions, respectively. Since the sum of these percentages of cost account for a high portion (about 71.02%) of the total project cost, these four divisions are termed the major cost divisions. The division-level unit costs of the four major cost divisions (inputs) and the corresponding project-level unit cost (output) of each historical project serve as the training data in the FALCON- and fmGA-based training sub-model.

### 4.2. Step 2-1: Applying FALCON

A FALCON network structure comprises five layers of nodes and two links (Lin, Lee 1991). This study applies FALCON as follows (see the right part of Fig. 2):

1. Layer 1 (input linguistic nodes): The nodes in this layer merely transmit the input values (i.e., division-level unit cost data or  $U_i$ s) to the next layer directly. That is, the  $U_i$ s of the four main cost divisions for each historical project are transmitted directly into the network.
2. Layer 2 (Input term nodes): The nodes in this layer calculate the membership functions. That is, this layer fuzzifies the input values (namely,  $U_i$ s) from Layer 1. Fuzzy partitions are determined based on the clustering relationships of both  $U_i$ s and  $U_{total}$  (i.e., the major division-level and project-level unit costs).
3. Link 1 (IF-part condition links): The connections between Layers 2 and 3 represent the fuzzy IF-THEN rule preconditions.
4. Layer 3 (rule nodes): The rule nodes in this layer perform the “fuzzy AND” operation to obtain the fired strength of various fuzzy logic rules.
5. Link 2 (THEN-part consequence links): The links between Layers 3 and 4 show the consequences of fuzzy logic rules. Only one consequence should exist for each rule node in a single output network. The

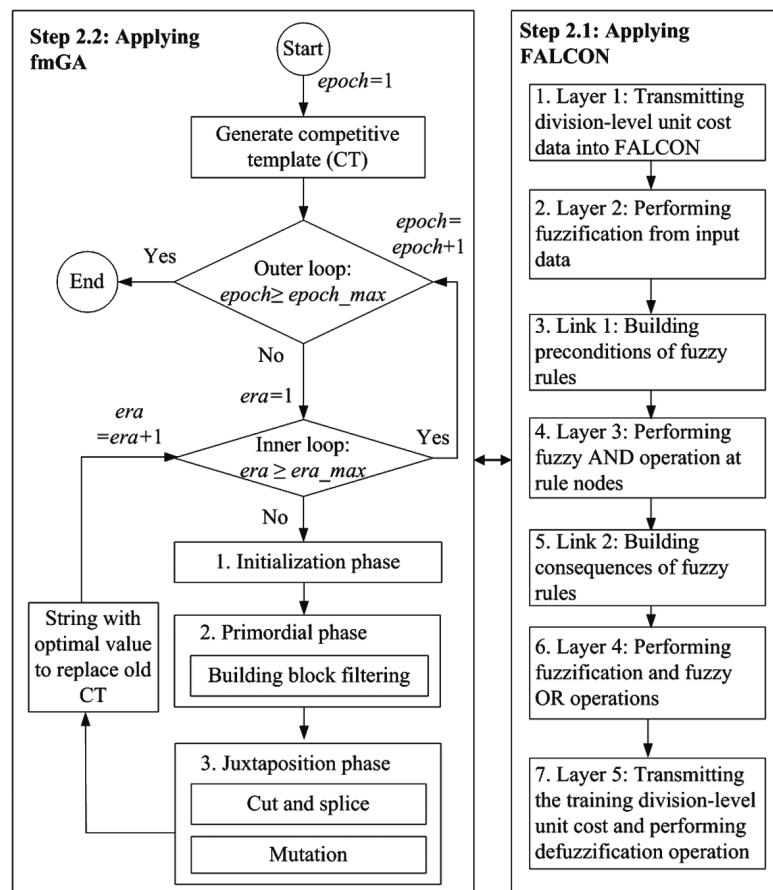


Fig. 2. Training the relationships between the division- and project-level unit costs

links are represented as numeric values 0 (disconnected) or 1 (connected).

6. Layer 4 (output term nodes): The nodes in this layer perform two functions, right-left (only performed in the training stage) and left-right (performed in both the training and usage stages) transmissions.
7. Layer 5 (output linguistic nodes): The nodes in Layer 5 also perform *right-left* (only in the training stage) and *left-right* (performed in both the training and usage stages) transmissions. In *right-left* transmission the nodes in Layer 5 perform identically to those in Layer 1; that is, they feed training data (i.e., actual  $U_{total}$ ) into the network. In *left-right* transmission the nodes at Layer 5 defuzzify fuzzy sets to provide a definite output value (i.e., estimated  $U_{total}$ ).

### 4.3. Step 2-2: Applying fmGA

Following the FALCON operations, the fmGA mutation and cut-splice operators were used to optimize the FALCON parameters, including fuzzy membership functions and fuzzy logic rules, to improve cost estimation accuracy. This requires using the fmGA variable-length chromosome to revise the fuzzy partitions (namely, the number of input term nodes in Layer 2) and fuzzy decision rules (namely, the consequence links of Link 2) of FALCON. The fmGA global search capability optimizes the parameters (means and spreads) of the membership functions in the FALCON input and output term nodes (Hsiao *et al.* 2012).

As presented in the left part of Figure 2, fmGA includes two operation loops, namely, an outer loop and inner loop. The completion of each outer loop marks an epoch, while each inner loop marks an era. As suggested by Feng and Wu (2006), this study defines the maximum number of eras (*era\_max*) as four. Additionally, the maximum number of epochs (*epoch\_max*) is defined as a preset criterion for terminating the fmGA evolution process, and is set to thirteen in this study. Furthermore, an inner loop consists of three phases (Goldberg *et al.* 1993): (1) the initialization phase – a population with sufficient strings is created to contain all possible building blocks (BBs) of the order  $k$ , where BBs refer to partial solutions to a problem; (2) primordial phase – bad genes are filtered out to maintain only chromosomes with good fitness (i.e., those containing only “good” alleles fitting to BBs); and (3) juxtapositional phase – good alleles (BBs) are rebuilt using cut-splice and mutation operations to form a high quality generation that generates optimal solutions.

Figure 2 shows that the fmGA starts with the outer loop and generates a competitive template (CT). After completing one era, the CT is replaced by a new CT (with new alleles) that has the best fitness (i.e., the highest estimation accuracy) found in that era. The operational details for the three phases can be found in Hsiao *et al.* (2012).

After evolution, the chromosome with highest fitness is fed back to FALCON to calculate the cost estimates of the new input data. Best-fit chromosomes (with opti-

num fitness) are also maintained via fmGA to provide the population and CT of the next epoch. Steps 2.1 and 2.2 (FALCON and fmGA operations) are then repeated iteratively until the fitness value converges or has reached a preset maximum era number. The total evolution process runs 52 generations (= 4 eras  $\times$  13 epochs). Finally, the fmGA operations stop and the final optimal solution is obtained.

### 4.4. Regression sub-model

The regression technique permits simple analysis to determine the influence of parameters on project cost (Lowe *et al.* 2006). A regression sub-model using the aforementioned 46 historical projects is proposed to consider the impact of total floor area during estimation. This sub-model generates the regressed division-level unit cost ( $RU_i$ ) of each major cost division for a new project. Taking the example of the foundation cost division, Figure 3 displays the regression sub-model of the  $RU_i$  for the foundation cost division.

Table 2 lists the regression equations developed to calculate the  $RU_i$ s of the four main cost divisions. In the equations,  $Q$  denotes the total floor area of a new project. Since the data of  $U_i$  for the 46 historical projects spread

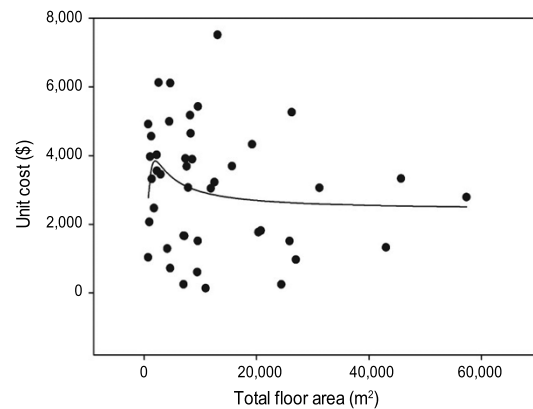


Fig. 3. Regression sub-model of division-level unit cost for foundation cost division

Table 2. Regression equations of division-level unit costs of the four major cost divisions

Cost divisions	Regression equation of division-level unit cost	R square
1. Foundation	$2,395 + (6,424,691/Q) - (8,540,036,612 / Q^2) + (2.9569 \times 10^{12} / Q^3)$	0.0495
2. Structure	$169 \times Q / (-636 + Q) + 7,129 \times Q / (174 + Q)$	0.0147
4. Internal finishes	$4,775 - 0.0126 \times Q - 1.9 \times 10^{-7} \times Q^2 + 1.16 \times 10^{-11} \times Q^3$	0.0299
6. MEP	$4,354 - 0.0269 \times Q + 2.97 \times 10^{-6} \times Q^2 - 4.6 \times 10^{-11} \times Q^3$	0.0427

Note:  $Q$  is the estimated total floor area ( $m^2$ ) of a new project.

widely, the value of the R square of each regression equation is very low; for example, R square for the foundation cost division is only 0.0495. Nevertheless, the  $RU_i$  derived from regression analysis can capture an important project characteristic, total floor area of a new project.

#### 4.5. Multi-factor evaluation sub-model

Many factors influence the of a cost division. Based on the suggestions of Wang *et al.* (2012), Table 3 lists the factors that affect the  $RU_i$ s of the four major cost divisions. Take the foundation cost division as an example. The  $RU_i$  of this division is dominated by four factors – ground improvement (F1.1), retaining wall (F1.2), excavation method (F1.3), and soil type (F1.4). Each factor is classified into different factor conditions. For instance, the ground improvement factor (F1.1) indicates whether a project site requires ground improvement. Five factor conditions are identified: no ground improvement (no cost effect); improvement via compaction; improvement by well-point dewatering; improvement via consolidation; and improvement by soil replacement (high cost effect).

This sub-model assumes the factors in a cost division  $i$  to be independent. The importance of each factor  $j$  is pair-wisely compared with other factors to obtain the weight ( $W_{i(j)}$ ) of each factor  $j$ . The evaluation result of a factor  $j$  for a given cost division  $i$  is a qualitative or quantitative value (e.g., factor condition) that is mapped to a corresponding effect value ( $E_{i(j)}$ ) to represent the effect of a factor on the  $RU_i$  of a cost division  $i$ . Multiplying the effect value ( $E_{i(j)}$ ) by its weight ( $W_{i(j)}$ ) yields a weighted effect value ( $S_{i(j)} = W_{i(j)} \times E_{i(j)}$ ) of a cost division  $i$ . Summing the weighted effect values of factors obtains the expected effect value ( $S_i = \sum S_{i(j)}$ ) of a cost division  $i$ . This process is repeated for each major cost division.

Next, considering the factor effects on the  $RU_i$  of a cost division  $i$ , a suggested division-level unit price ( $SU_i$ ) is calculated, as follows:

$$SU_i = RU_i \times [1 + (S_i - 0.5)], \quad (6)$$

Table 3. Factors affecting division-level unit costs in the four cost divisions (Wang *et al.* 2012)

Cost divisions	Factors
1. Foundation	F1.1 ground improvement; F1.2 retaining wall; F1.3 excavation method; F1.4 soil types
2. Structure	F2.1 concrete strength; F2.2 form types; F2.3 building complexity; F2.4 number of floors; F2.5 floor height; F2.6 earthquake location
4. Internal finishes	F4.1 form types; F4.2 material types of floors; F4.3 material types of internal walls; F4.4 floor height; F4.5 room area; F4.6 number of rooms in an unit
6. MEP	F6.1 functions of equipment; F6.2 quality class of MEP; F6.3 room area; F6.4 number of rooms in an unit

where  $S_i - 0.5 = 0$  indicates that factors have an average effect on the  $RU_i$ , and therefore suggests the  $SU_i$  is the regressed unit cost (namely  $SU_i = RU_i$ ). Similarly,  $S_i - 0.5$  above zero leads to  $SU_i$  higher than  $RU_i$ , and vice versa.

#### 4.6. Steps involved in new project estimation

The left side of Figure 4 illustrates the detailed steps of applying the proposed model to estimate the conceptual cost of a new project, while the right side shows the related sub-models involved in each step. For instance, the algorithms of multi-factor evaluation sub-model are applied in Steps 5.1–5.3 and Step 5.5 to calculate the influence of the expected effect value on the regressed division-level unit cost ( $RU_i$ ).

#### 4.7. Computer implementation

The operations of Step 2 (FALCON and fmGA) are performed with Matlab® version 7.5. The unit cost data are read in .dat format and the operations are run under genuine Intel 1.6 GHz CPU, 896 MB SRAM, and Windows XP computer operating systems. Training the FALCON- and fmGA-based sub-model of the 46 historical projects takes approximately 20 minutes. The operations of other steps (Steps 1, 3 and 4) of the proposed model are performed using Microsoft Excel.

### 5. Case studies

#### 5.1. Description of case projects

The proposed method is applied to three residential housing projects (namely projects I, II, and III) that were also used in the study of Wang *et al.* (2012). These three projects and the aforementioned 46 historical projects shared a common contractor. Table 4 lists the main characteristics of the three case study projects. For example, project I, made of RC, has 14 floors and three underground floors, and a total floor area of roughly 7,363 m<sup>2</sup>. Completion of Project I, achieved in mid-1999, took 20 months. The following subsections examine the evaluation results.

Table 4. General descriptions of three case projects

Characteristics	Project I	Project II	Project III
Completion time	Aug. 1999	Nov. 2000	Dec. 2004
Cost index	77.06	76.69	92.60
Construction duration (months)	20	20	23
Total floor area (m <sup>2</sup> )	7,363	4,490	14,518
Number of floors	14	13	14
Floor height (m)	3.6	3.2	3.6
Room area (m <sup>2</sup> )	163.6	171.0	272.7
Number of rooms in an unit	4	4	4
Concrete strength	4,000 psi	4,000 psi	5,000 psi
Retaining wall	Slurry wall	Slurry wall	Slurry wall
Excavation method	Bottom up	Bottom up	Bottom up
Soil type	Silt	Silt	Silt

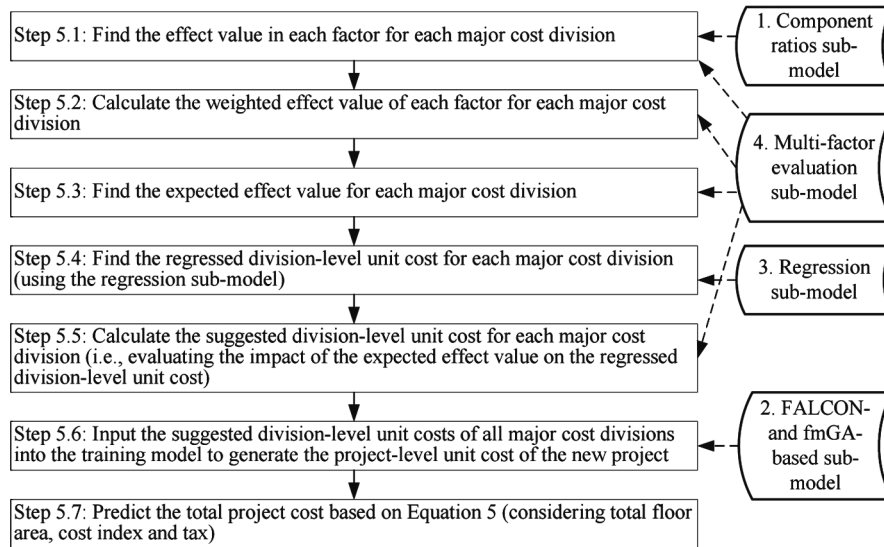


Fig. 4. Steps involved in new project estimation using the proposed model

## 5.2. Evaluation results for project I

First, Steps 5.1–5.3 are conducted, as shown in Figure 4. Table 5 lists the calculated weights, effect values, weighted effect values and expected effect values for each major cost division by applying the multi-factor evaluation sub-model. For instance, the table shows that the expected effect value ( $S_i$ ) for the foundation cost division is 0.8002 for project I.

Next, Step 5.4 is performed. Restated, given the total floor area (7,363 m<sup>2</sup>) of project I, the regressed  $U_i$  ( $RU_i$ ) of each major cost division  $i$  is generated using the regression sub-model. See the left part of Table 6. For example,  $RU_i$  of the foundation cost division is \$3,118/m<sup>2</sup>, as listed in Table 6.

Step 5.5 then calculates the suggested  $U_i$  ( $SU_i$ ) for each major cost division. See the right part of Table 6. For

Table 5.  $W_{i(j)}$ ,  $E_{i(j)}$ ,  $S_{i(j)}$  and  $S_i$  for the four major cost divisions in case project I

Cost division	Factor	Weight ( $W_{i(j)}$ )	Effect value ( $E_{i(j)}$ )	Weighted effect value ( $S_{i(j)} = E_{i(j)} \times W_{i(j)}$ )	Expected effect value ( $S_i = \sum S_{i(j)}$ )
1. Foundation	F1.1	6.32%	0.0	0.0000	0.8002
	F1.2	60.05%	1.0	0.6005	
	F1.3	27.31%	0.5	0.1365	
	F1.4	6.32%	1.0	0.0632	
2. Structure	F2.1	6.91%	0.5	0.0346	0.6618
	F2.2	6.91%	1.0	0.0691	
	F2.3	3.08%	0.5	0.0154	
	F2.4	15.84%	0.6	0.0885	
	F2.5	23.59%	1.0	0.2359	
	F2.6	43.67%	0.5	0.2183	
4. Internal finishes	F4.1	9.60%	0.0	0	0.5956
	F4.2	44.89%	0.6	0.2693	
	F4.3	24.99%	0.6	0.1499	
	F4.4	14.06%	1.0	0.1406	
	F4.5	2.85%	0.3	0.0087	
	F4.6	3.62%	0.8	0.0271	
6. MEP	F6.1	59.78%	0.5	0.2989	0.5142
	F6.2	26.04%	0.5	0.1302	
	F6.3	4.79%	0.3	0.0146	
	F6.4	9.40%	0.8	0.0705	



Table 6. Regressed and suggested division-level unit costs for the four major cost divisions in project I

Major cost division	Regressed division-level unit cost ( $RU_i$ )	Expected effect value ( $S_i = \sum S_{i(j)}$ )	Suggested division-level unit cost ( $SU_i$ )
1. Foundation	\$3,118	0.8002	\$4,054
2. Structure	\$7,113	0.6618	\$8,264
4. Internal finishes	\$4,676	0.5956	\$5,123
6. MEP	\$4,299	0.5142	\$4,360

instance, based on the  $RU_i$  (\$3,118/m<sup>2</sup>) and the expected effect value ( $S_i = 0.8002$ ), the value of  $SU_i$  of the foundation cost division is \$4,054/m<sup>2</sup> ( $= 3,118 \times (1 + 0.8002 - 0.5) = 3,118 \times 1.3002$ ) based on Eqn (6).

Next, based on Step 5.6, these four values of  $SU_i$  for the foundation, structure, internal finishes, and MEP cost divisions for project I provide inputs to the training sub-model to predict the  $U_{total}$ , which is \$34,492 /m<sup>2</sup>.

Finally, using Eqn (5) (Step 5.7), the estimated total cost of project I is calculated as follows:

$$C_{Tot} = (U_{total} \times Q) \left( \frac{CCI_{year}}{CCI_{base}} \right) \times (1 + t) =$$

$$(34,492 \times 7,363) \times (77.06 / 100) \times (1 + 0.05) =$$

$$\text{\$205,485,071,}$$

where:  $U_{total} = \$34,492/m^2$ ;  $Q = 7,363 \text{ m}^2$ ;  $CCI_{year} = 77.06$ ;  $CCI_{base} = 100$ ;  $t = 0.05$ .

Notably, the actual cost for project I was \$186,492,943. Thus, the total cost approximated by the proposed model is 10.18% ( $= (205,485,071 - 186,493,943) / 186,493,943$ ) higher than the actual project cost. Eqn (8) defines the estimation accuracy. That is, the proposed model achieves estimation accuracy of approximately 89.82% ( $= 1 - 10.18\%$ ):

$$\text{Accuracy (\%)} =$$

$$\left( 1 - \frac{\text{ABS (Estimated cost - Actual cost)}}{\text{Actual cost}} \right) \times 100\%. \quad (8)$$

Table 7. Evaluation results of three case studies using the proposed model

	Project I	Project II	Project III
Actual cost (a)	186,493,993	112,570,700	455,501,682
Suggested cost (b)	205,485,071	104,328,861	487,585,339
Difference (b - a)	18,991,078	-8,241,839	32,083,657
Difference in % ((b - a)/a)	10.18%	-7.32%	7.04%
Accuracy in % ( $1 - \text{ABS}(b - a)/a$ )	89.82%	92.68%	92.96%

### 5.3. Evaluation results for projects II and III

Similarly, the evaluation steps are also applied to case projects II and III. Table 7 summarizes the evaluation results. Restated, the proposed model achieved estimation accuracies of 89.82%, 92.68%, and 92.96% for projects I, II, and III, respectively.

### 5.4. Sensitivity analysis

To more thoroughly elucidate the behavior of the proposed model, sensitivity analysis (Park 2011) is conducted to evaluate how much the estimation accuracy changes in response to a given change in an input variable (i.e. expected effect value;  $S_i$ ) of the proposed model in the above case projects. Notably,  $S_i$  value represents the combined cost effects of all factors (i.e. project characteristics) on each major cost division  $i$ .

Figure 5 plots the sensitivity graphs based on the calculated estimation accuracies of various scenarios for the three case projects. The sensitivity analysis begins with a base-case scenario, which is developed using the initially-estimated values of  $S_i$  of four cost divisions. For instance, Table 5 lists the  $S_i$  values of the base-case scenario for project I, in which the resulting estimation accuracy is 89.82%. The value of  $S_i$  is then changed by several specified percentage points (ranging from -40% to 40%) above and below the initially-estimated value for each of the four cost divisions; meanwhile, the other variables remain constant. Then, project cost and estimation accuracy based on the changed  $S_i$  values of four cost divisions are calculated.

On average, when the  $S_i$  value is underestimated, the accuracy diminishes (as shown in the left portion of Fig. 5). Generally, the slope of each line in Figure 5 indicates that the estimation accuracy is sensitive to changes in the  $S_i$  value, implying that considering the effects of project characteristics (i.e. multi-factor evaluation sub-model) is vital to the estimation accuracy of the proposed model.

### 5.5. Analysis of the benefits of using fmGA

As mentioned earlier, despite its learning capability (based on neural network theory) from historical cost data to pre-

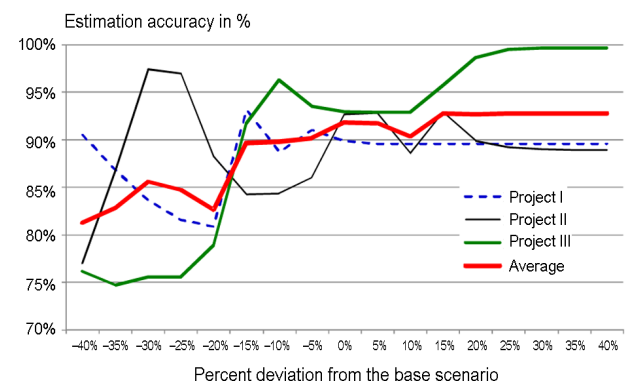


Fig. 5. Sensitivity graphs for the three case projects

dict  $U_{total}$ , FALCON fails to optimize its learning parameters (including fuzzy membership functions and fuzzy logic rules) in order to improve  $U_{total}$ . Thus, the proposed model incorporates fmGA (with its ability to optimize the learning parameters) into FALCON in order to resolve a local optimality problem of FALCON.

Notably, the proposed FALCON- and fmGA-based sub-model forecasts  $U_{total}$  (not the total project cost), which is then used to estimate the total project cost ( $C_{Tot}$ ) by using Eqn (7). Hence, the merits of using fmGA in this study are validated from the perspectives of examining the accuracy of the resulting  $C_{Tot}$  and assessing the training error rate (i.e. error rate during the training process) of the predicted  $U_{total}$ .

### 5.5.1. Examining the accuracy of resulting $C_{Tot}$

Two modified models (without fmGA) are designed. The first modified model uses only FALCON, while the second modified model incorporates simple GAs (i.e. with the training ability of sGA) into FALCON. Notably, the proposed and two modified models have the same three sub-models: component ratios, regression, and multi-factor evaluation.

Table 8 compares the proposed model and the modified models in the three case projects with respect to the evaluation results of  $C_{Tot}$ . Based on those results, the following observations are made:

Table 8. Comparisons of evaluation results of  $C_{Tot}$  using the proposed and modified models in the three case projects

	Project I	Project II	Project III
Actual cost	\$186,493,993	112,570,700	455,501,682
1. Proposed model using FALCON+fmGA (52 generations):			
Estimated $U_{total}$	34,492	28,856	34,547
Suggested cost	\$205,485,071	104,328,861	487,585,339
Accuracy in %	89.82%	92.68%	92.96%
Average accuracy of $C_{Tot}$ : 91.82%			
2. 1 <sup>st</sup> modified model using only FALCON (no training):			
Estimated $U_{total}$	33,132	24,914	31,198
Suggested cost	\$197,382,911	90,076,561	440,318,621
Accuracy in %	94.16%	80.02%	96.67%
Average accuracy of $C_{Tot}$ : 90.28%			
3. Proposed model using FALCON+fmGA (200 generations):			
Estimated $U_{total}$	32,718	26,722	33,730
Suggested cost	\$194,916,518	96,613,385	476,054,461
Accuracy in %	95.48%	85.82%	95.49%
Average accuracy of $C_{Tot}$ : 92.27%			
4. 2 <sup>nd</sup> modified model using FALCON+sGA (200 generations):			
Estimated $U_{total}$	34,707	29,080	35,149
Suggested cost	\$206,765,927	105,138,733	496,081,775
Accuracy in %	89.13%	93.40%	91.09%
Average accuracy of $C_{Tot}$ : 91.21%			

1. Average accuracy of the proposed model (for 52 generations) slightly improves the results by 1.54% (= 91.82% – 90.28%) over that of the first modified model (using only FALCON).
2. If the training process of fmGA continues to run for 200 generations (= 4 eras × 50 epochs), then the average accuracy of the proposed model increases by 0.45% (from 91.82% to 92.27%) in the three projects over that running for only 52 generations.
3. Average accuracy of the proposed model (for 200 generations) also slightly improves the results by 1.06% (= 92.27% – 91.21%) over that of the second modified model (using FALCON + sGA for 200 generations).

In sum, accuracy of the resulting total project costs for incorporating fmGA into FALCON slightly excels that of using only FALCON and FALCON+sGA in the three case projects.

### 5.5.2. Assessing the error rate of the predicted $U_{total}$

FALCON in the first modified model lacks training ability, explaining why an estimated  $U_{total}$  of new project estimation is fixed and cannot be further improved. Thus, the advantage of the training ability of fmGA is illustrated by comparing the proposed model with the second modified model (with the training ability of sGA).

Figure 6 displays the average error rates of estimated  $U_{total}$  using the proposed model and the second modified model for 46 historical projects during 200 training generations. Notably, the error rate equals 1 minus the estimation accuracy of  $U_{total}$ , and it is represented as follows:

Error rate (%) =

$$\frac{\text{ABS}(\text{Estimated } U_{total} - \text{Actual } U_{total})}{\text{Actual } U_{total}} \times 100\%. \quad (9)$$

In each generation, the value of  $U_{total}$  for each historical project is estimated and an error rate is derived using Eqn (9). The average error rate is the average value of the error rates for all 46 historical projects.

Figure 6 indicates that the average error rates of the resulting  $U_{total}$  for both models (either using fmGA or sGA) decrease as the models are trained for more generations. However, the proposed model using fmGA performs better than the second modified model (using sGA). Namely, its error rate decreases by 0.32% (= 3.53% – 3.21%) over that of the second modified model after 200 training generations.

Moreover, according to Figure 6, training the model for 33 generations can achieve a stable average error rate. Restated, conducting additional training generations negligibly affects the ability to further diminish the average error rates.

While Figure 6 examines the training error rate of  $U_{total}$  for 46 historical projects, Table 9 compares the training error rates of  $U_{total}$  for the new projects (i.e. the three case projects). Restated, according to Table 9, the average

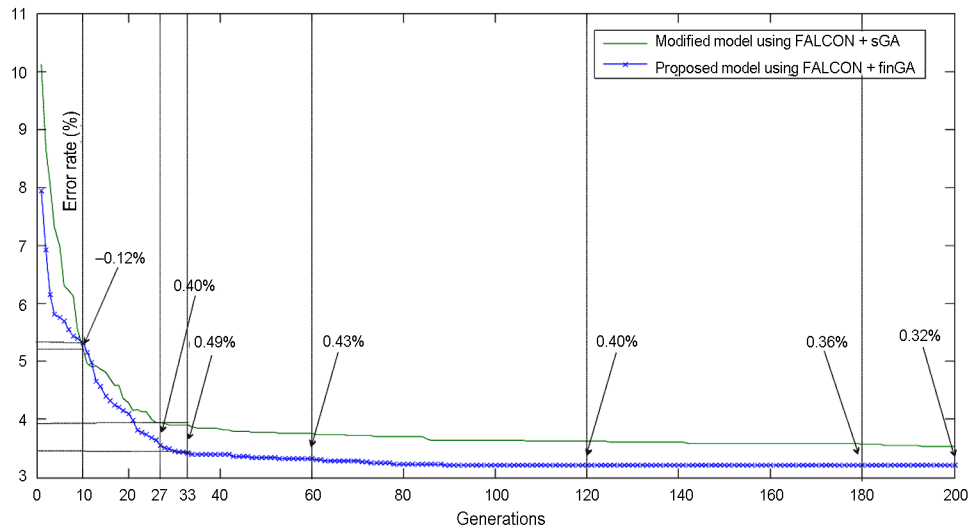


Fig. 6. Error rates of estimated  $U_{total}$  using the proposed and modified models during the training generations

error rate of  $U_{total}$  for the three case projects for the proposed model decreases by 1.99% (= 9.72% – 7.73%) and 1.24% (= 8.97% – 7.73%) over that of the first modified model and the second modified model, respectively. In summary, estimation accuracy of the resulting  $U_{total}$  for incorporating fmGA into FALCON exceeds that of using only FALCON and using sGA+ FALCON in the three case projects.

**5.6. Comparisons with three conventional methods**

The conventional average (project-level) unit cost method, the component ratios method, and the linear regression method are applied to the case projects for comparing the proposed model. Project I also illustrates the workings of these three conventional methods.

Table 9. Comparisons of error rates of  $U_{total}$  using various estimation methods for three case projects

	Project I	Project II	Project III
Actual $U_{total}$	\$31,304	31,136	32,274
1. Proposed model using FALCON+fmGA (200 generations):			
Estimated $U_{total}$	32,718	26,722	33,730
Error rate of $U_{total}$	4.52%	14.18%	4.51%
Average error rate of $U_{total}$ : 7.73%			
2. 1 <sup>st</sup> modified model using only FALCON (no training):			
Estimated $U_{total}$	33,132	24,914	31,198
Error rate of $U_{total}$	5.84%	19.98%	3.33%
Average error rate of $U_{total}$ : 9.72%			
4. 2 <sup>nd</sup> modified model using FALCON+sGA (200 generations):			
Estimated $U_{total}$	34,707	29,080	35,149
Error rate of $U_{total}$	10.87%	6.60%	8.91%
Average error rate of $U_{total}$ : 8.97%			

The average unit cost method obtains the estimated average unit cost of the total project of \$28,588/m<sup>2</sup> (calculated according to the 46 historical projects). Given that the total floor area of project I equals 7,363 m<sup>2</sup>, the estimated project cost (before considering the tax and cost index) equals \$210,493,444 (= 28,588 × 7363). The estimated project cost based on Eqn (5) then equals \$170,315,060. Since the actual project cost was \$186,492,943, the estimation accuracy of the average unit cost method is 90.58% using Eqn (8).

The component ratios method (Barrie, Paulson 1992) identifies the foundation, structure, internal finishes, and MEP as the major cost divisions. The averaged ratio between the sum of the costs of these major divisions and the whole project cost is 71.02% for the 46 historical projects. The average  $U_i$  of each major cost division can then be used to calculate the estimated cost of each division, as shown in Table 10.

For instance, the average  $U_i$  and estimated cost of foundation cost division are \$2,868 and \$21,117,084 (= 2,868 × 7363), respectively. Next, since the sum of the estimated costs of these four cost divisions is \$142,893,741, the total cost (before considering the tax and cost index)

Table 10. Estimated costs of four major cost divisions using the component ratios method for project I

Major cost division	Average division-level unit cost ( $AU_i$ )	Total floor area ( $Q$ ; m <sup>2</sup> )	Estimated cost ( $AU_i \times Q$ )
1. Foundation	\$2,868	7,363	\$21,117,084
2. Structure	\$7,463	7,363	\$54,950,069
4. Internal finishes	\$5,043	7,363	\$37,131,609
6. MEP	\$4,033	7,363	\$29,694,979
Sum:			\$142,893,741

of project I is \$201,202,113 (= 142,893,741 ÷ 71.02%). The estimated project cost based on Eqn (5) then equals \$162,783,880. Thus, the estimation accuracy using the component ratios method is 87.29% using Eqn (8).

The linear regression method develops a regression equation according to the 46 historical projects. The total project cost before considering the tax and cost index, denoted as  $Y$ , is derived as follows:

$$Y = 25,181 \times Q + 2 \times 10^7, \quad (10)$$

where the value of  $R$  square of the derived regression line is 0.8538. Following consideration of the tax and cost index for project I, the total project cost is calculated as follows:

$$C_{Tot} = Y \times \left( \frac{CCI_{year}}{CCI_{base}} \right) \times (1+t). \quad (11)$$

Since the total floor area ( $Q$ ) of project I equals 7,363 m<sup>2</sup>, the estimated project cost equals \$166,201,535 according to Eqns (10) and (11), resulting in an estimation accuracy of 89.12% in which Eqn (8) is used.

Table 11 summarizes the evaluation results using the proposed model and the three conventional methods in the three case projects. Based on those results, the following observations are made:

1. The proposed model achieves an average estimation accuracy of 91.82% for three projects.
2. The average estimation accuracies of the three case projects are 90.58%, 86.57%, and 88.87% using the average unit cost, component cost ratios, and linear regression methods, respectively.

Table 11. Comparisons of evaluation results using various estimation methods for three test projects

	Project I	Project II	Project III
Actual cost	\$186,493,993	112,570,700	455,501,682
1. Proposed model			
Accuracy in %	89.82%	92.68%	92.96%
Average accuracy	91.82%		
2. Average unit cost			
Suggested cost	\$170,315,060	103,361,664	403,488,771
Difference in %	-8.68%	-8.18%	-11.42%
Accuracy in %	91.32%	91.82%	88.58%
Average accuracy	90.58%		
3. Component ratios			
Suggested cost	\$162,783,880	98,791,104	385,646,858
Difference in %	-12.71%	-12.24%	-15.34%
Accuracy in %	87.29%	87.76%	84.66%
Average accuracy	86.57%		
4. Linear regression			
Suggested cost	\$166,201,535	107,148,066	374,897,254
Difference in %	-10.88%	-4.82%	-17.70%
Accuracy in %	89.12%	95.18%	82.30%
Average accuracy	88.87%		

3. The proposed model increases the average estimation accuracy by approximately 1.24% (= 91.82% – 90.58%), 5.25% (= 91.82% – 86.57%), and 2.95% (= 91.82% – 88.87%), over that of the average unit cost method, component ratios method, and linear regression method in these three case projects, respectively.
4. The three conventional methods tend to underestimate project costs because their estimated project costs are lower than the actual project costs in all three case studies.

To further verify the proposed model, a 16-fold cross-validation is performed. Restated, the 49 projects (the abovementioned 46 historical projects plus three case projects) are divided into 16 groups. Of the 16 groups, a single group (consisting of three projects) is retained as the validation projects for testing a particular method, and the remaining 15 groups (46 projects) are used as training projects. The cross-validation process is repeated 16 times (i.e., 3 × 16 = 48 testing projects). Evaluation results of the 48 testing projects are then averaged.

However, only the average unit cost method, component ratios method, and linear regression method are implemented in this cross-validation process, owing to the availability of the data required for incorporating these methods. The proposed model is not adopted in this cross-validation process because detailed data of the other 46 historical projects are unavailable in order to apply the multi-factor evaluation sub-model of the proposed model.

The lower portion of Table 12 summarizes the evaluation results using the three conventional methods in this cross-validation process (i.e. 48 testing projects). Based on those results, we conclude the following:

1. In the 48 testing projects, average estimation accuracies of the three case projects decrease by 15.38% (= 90.58% – 75.20%), 13.81% (= 86.57% – 72.76%), and 15.18% (= 88.87% – 73.69%) when using the average unit cost, component cost ratios, and linear regression methods, respectively.
2. The three conventional methods yield an average estimation precision of around 73.88%, which is similar to that (±25%) of Petroutsatou *et al.* (2012).
3. Again, the three conventional methods tend to underestimate project costs in this cross-validation process.

Table 12. Comparisons of evaluation results using various estimation methods

Models	No. of test projects	Average accuracy	Min. accuracy	Max. accuracy
1. Proposed model	3	91.82%	89.82%	92.96%
2. Average unit cost	3	90.58%	88.58%	91.82%
3. Component ratios	3	86.57%	84.66%	87.76%
4. Linear regression	3	88.87%	82.30%	95.18%
2. Average unit cost	48	75.20%	44.09%	99.84%
3. Component ratios	48	72.76%	42.27%	94.82%
4. Linear regression	48	73.69%	4.61%	99.26%

For instance, only four out of 48 projects generate an estimated project cost that exceeds the actual project cost when using the component ratios method.

4. Estimation accuracies of the three conventional methods diverge widely. For example, the minimum and maximum accuracies using the linear regression method are 4.61% and 99.26%, respectively.

## Conclusions

This study has attempted to solve conceptual cost estimation problems in a systematic manner by integrating four sub-models (i.e. the component ratios method, FALCON/fmGA, regression method, and multi-factor evaluation method) into the proposed model. Overall, in terms of its significance, the proposed model can more accurately reflect project characteristics of new project estimations than those of conventional methods by using the multi-factor evaluation sub-model in order to increase management confidence.

As for estimation accuracy, the proposed model should be extended to additional new projects in order to further verify its credibility. This is owing to that the estimation results of the three case projects (Table 12) only confirm that the proposed model retains a high likelihood of increasing estimation accuracy.

Three additional comments are identified after presenting these application results to the cost manager involved in the case studies. First, the conceptual phase is full of uncertainty despite 30% of the design being complete. However, confirming the total floor area could be essential to enhancing the estimation accuracy since the proposed model and the other three conventional methods all require the total floor area as inputs and also have high accuracy in the three case projects (82.30%~95.18%; see Table 11).

Second, the cost manager expressed that the proposed model should receive higher management confidence than the other three conventional methods because it can systematically reflect certain features of new projects using the multi-factor evaluation sub-model. Finally, the cost manager suggests using fuzzy-based theories to evaluate the factor conditions, which are likely uncertain during the conceptual phase.

Future research can also examine the following directions. First, computerizing the proposed model would help expedite the evaluation. Second, collecting additional historical projects should support the model training process and thus enhance estimation accuracy. Third, other neuro-fuzzy systems or optimization methods can substitute for FALCON and fmGA to find enhanced solutions.

## Acknowledgements

The authors would like to thank the Ministry of Science and Technology of Taiwan for financially supporting this research (under Contract No. NSC101-2221-E-009-137-MY3). The assistance of practitioners who provided the valuable cost data and helped implement the case studies is also greatly appreciated.

## References

- Adeli, H.; Wu, M. 1998. Regularization neural network for construction cost estimation, *Journal of Construction Engineering and Management* 124(1): 18–24.  
[http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(1998\)124:1\(18\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(1998)124:1(18))
- Asmar, M. E.; Hanna, A. S.; Whited, G. C. 2011. New approach to developing conceptual cost estimates for highway projects, *Journal of Construction Engineering and Management* 137(11): 942–949.  
[http://dx.doi.org/10.1061/\(ASCE\)CO.1943-7862.0000355](http://dx.doi.org/10.1061/(ASCE)CO.1943-7862.0000355)
- Barrie, D. S.; Paulson, B. C. 1992. *Professional construction management*. 3<sup>rd</sup> ed. New York: McGraw-Hill. 577 p.
- Boussabaine, A. H.; Elhag, T. M. S. 1997. *A neurofuzzy model for predicting cost and duration of construction projects*. RICS Research. The Royal Institution of Chartered Surveyors. London. 9 p.
- Cheng, M. Y.; Tsai, H. C.; Hsieh, W. S. 2009. Web-based conceptual cost estimates for construction projects using evolutionary fuzzy neural inference model, *Automation in Construction* 18(2): 164–172.  
<http://dx.doi.org/10.1016/j.autcon.2008.07.001>
- Cheng, M. Y.; Tsai, H. C.; Sudjono, E. 2010. Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry, *Expert Systems with Applications* 37(6): 4224–4231.  
<http://dx.doi.org/10.1016/j.eswa.2009.11.080>
- Chou, J. S. 2009. Web-based CBR system applied to early cost budgeting for pavement maintenance project, *Expert Systems with Applications* 36(2): 2947–2960.  
<http://dx.doi.org/10.1016/j.eswa.2008.01.025>
- Clark, F. D.; Lorenzoni, A. B. 1997. *Applied cost engineering*. 3<sup>rd</sup> ed. New York: Taylor & Francis Group. 419 p.
- Creese, R. C.; Li, L. 1995. Cost estimation of timber bridges using neural networks, *Cost Engineering* 37(5): 17–22.
- Feng, C. W.; Wu, H. T. 2006. Integrating fmGA and CYCLONE to optimize the schedule of dispatching RMC trucks, *Automation in Construction* 15(2): 186–199.  
<http://dx.doi.org/10.1016/j.autcon.2005.04.001>
- Goldberg, D. E.; Deb, K.; Kaegupta, H.; Harik, G. 1993. *Rapid, accurate optimization of difficult problems using fast messy genetic algorithms*. IlliGAL Report No. 93004. University of Illinois at Urbana-Champaign, IL. 36 p.
- Gunduz, M.; Ugur, L.O.; Ozturk, E. 2011. Parametric cost estimation system for light rail transit and metro trackworks, *Expert Systems with Applications* 38(3): 2873–2877.  
<http://dx.doi.org/10.1016/j.eswa.2010.08.080>
- Hegazy, T.; Ayed, A. 1998. Neural network model for parametric cost estimation of highway projects, *Journal of Construction Engineering and Management* 124(3): 210–218.  
[http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(1998\)124:3\(210\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(1998)124:3(210))
- Hendrickson, C.; Au, T. 2003. *Project management for construction – fundamental concepts for owners, engineers, architects and builders* [online], [cited 01 September 2013]. Available from Internet: <http://pmbook.ce.cmu.edu/>
- Holland, J. 1975. *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor, MI. 211 p.
- Hong, T.; Hyun, C. T.; Moon, H. S. 2011. CBR-based cost prediction model-II of the design phase for multi-family housing projects, *Expert Systems with Applications* 38(3): 2797–2808. <http://dx.doi.org/10.1016/j.eswa.2010.08.071>
- Hsiao, F. Y.; Wang, S. H.; Wang, W. C.; Wen, C. P.; Yu, W. D. 2012. Neuro-fuzzy cost estimation model enhanced by fast messy genetic algorithms for semiconductor hookup construction, *Computer-Aided Civil and Infrastructure Engineering* 27(10): 764–781.  
<http://dx.doi.org/10.1111/j.1467-8667.2012.00786.x>
- Jin, R.; Cho, K.; Hyun, C.; Son, M. 2012. MRA-based revised CBR model for cost prediction in the early stage of con-

- struction projects, *Expert Systems with Applications* 39(5): 5214–5222. <http://dx.doi.org/10.1016/j.eswa.2011.11.018>
- Kim, G. H.; Seo, D. S.; Kang, K. I. 2005. Hybrid models of neural networks and genetic algorithms for predicting preliminary cost estimates, *Journal of Computing in Civil Engineering* 19(2): 208–211. [http://dx.doi.org/10.1061/\(ASCE\)0887-3801\(2005\)19:2\(208\)](http://dx.doi.org/10.1061/(ASCE)0887-3801(2005)19:2(208))
- Kim, S. 2013. Hybrid forecasting system based on case-based reasoning and analytic hierarchy process for cost estimation, *Journal of Civil Engineering and Management* 19(1): 86–96. <http://dx.doi.org/10.3846/13923730.2012.737829>
- Koo, C. W.; Hong, T. H.; Hyun, C. T. 2011. The development of a construction cost prediction model with improved prediction capacity using the advanced CBR approach, *Expert Systems with Applications* 38(7): 8597–8606. <http://dx.doi.org/10.1016/j.eswa.2011.01.063>
- Lai, Y. T.; Wang, W. C.; Wang, H. H. 2008. AHP- and simulation-based budget determination procedure for public building construction projects, *Automation in Construction* 17(5): 623–632. <http://dx.doi.org/10.1016/j.autcon.2007.10.007>
- Lin, C. T.; Lee, C. S. G. 1991. Neural-network-based fuzzy logic control and decision system, *IEEE Transactions on Computers* 40(12): 1320–1336. <http://dx.doi.org/10.1109/12.106218>
- Lowe, D. J.; Emsley, M. W.; Harding, A. 2006. Predicting construction cost using multiple regression techniques, *Journal of Construction Engineering and Management* 132(7): 750–758. [http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(2006\)132:7\(750\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(2006)132:7(750))
- Oberlender, G. D.; Trost, S. M. 2001. Predicting accuracy of early cost estimates based on estimate quality, *Journal of Construction Engineering and Management* 127(3): 173–182. [http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(2001\)127:3\(173\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(2001)127:3(173))
- Park, C. S. 2011. *Contemporary engineering economics*. 5<sup>th</sup> ed. New Jersey: Pearson Education, Inc. 960 p.
- PEH. 2008. *Parametric estimating handbook*. 4<sup>th</sup> ed. International Society of Parametric Analysts [online], [cited 01 September 2013]. Available from Internet: [http://www.galorath.com/images/uploads/ISPA\\_PEH\\_4th\\_ed\\_Final.pdf](http://www.galorath.com/images/uploads/ISPA_PEH_4th_ed_Final.pdf)
- Petroutsatou, K.; Georgopoulos, E.; Lambropoulos, S.; Pantouvakis, J. P. 2012. Early cost estimating of road tunnel construction using neural networks, *Journal of Construction Engineering and Management* 138(6): 679–687. [http://dx.doi.org/10.1061/\(ASCE\)CO.1943-7862.0000479](http://dx.doi.org/10.1061/(ASCE)CO.1943-7862.0000479)
- Sonmez, R. 2008. Parametric range estimating of building costs using regression models and bootstrap, *Journal of Construction Engineering and Management* 134(12): 1011–1016. [http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(2008\)134:12\(1011\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(2008)134:12(1011))
- Trost, S. M.; Oberlender, G. D. 2003. Prediction accuracy of early cost estimates using factor analysis and multivariate regression, *Journal of Construction Engineering and Management* 129(2): 198–204. [http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(2003\)129:2\(198\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(2003)129:2(198))
- Wang, W. C.; Wang, S. H.; Tsui, Y. K.; Hsu, C. H. 2012. A factor-based probabilistic cost model to support bid-price estimation, *Expert Systems with Applications* 39(5): 5358–5366. <http://dx.doi.org/10.1016/j.eswa.2011.11.049>
- Wen, C. P. 2010. *Investigation of cost Estimates on semiconductor hookup construction*: Master thesis. Program of Engineering Technology and Management, National Chiao Tung University, Taiwan (in Chinese).
- Yu, W. D. 2006. PIREM: a new model for conceptual cost estimation, *Construction Management and Economics* 24(3): 259–270. <http://dx.doi.org/10.1080/01446190500183735>
- Yu, W. D. 2007. Hybrid soft computing approach for knowledge discovery in construction, *Journal of Computing in Civil Engineering* 21(5): 343–352. [http://dx.doi.org/10.1061/\(ASCE\)0887-3801\(2007\)21:5\(343\)](http://dx.doi.org/10.1061/(ASCE)0887-3801(2007)21:5(343))
- Yu, W. D.; Skibniewski, M. J. 1999. A neuro-fuzzy computational approach to constructability knowledge acquisition for construction technology evaluation, *Automation in Construction* 8(5): 539–552. [http://dx.doi.org/10.1016/S0926-5805\(98\)00104-6](http://dx.doi.org/10.1016/S0926-5805(98)00104-6)
- Yu, W. D.; Lin, H. W. 2006. A VaFALCON neuro fuzzy system for mining of incomplete construction databases, *Automation in Construction* 15(1): 33–46. <http://dx.doi.org/10.1016/j.autcon.2005.01.007>

**Wei-Chih WANG.** Professor in the Department of Civil Engineering at National Chiao Tung University, Taiwan. He is a member of the Construction Management Association of Taiwan. His research interests include construction procurement, engineering management, building information modeling, high-tech facility management, and construction simulation.

**Tymur BILOZEROV.** PhD candidate in the Department of Civil Engineering at National Chiao Tung University, Taiwan. His research interests include construction bidding and site management.

**Ren-Jye DZENG.** Professor in the Department of Civil Engineering at National Chiao Tung University, Taiwan. He is also the CEO of the Smart Living Technologies Integration Center. He received his PhD degree from the University of Michigan, Ann Arbor in 1995. His research interests include construction automation, artificial intelligence, and the accelerator and brain EEG applications for construction safety.

**Fan-Yi HSIAO.** PhD from Department of Civil Engineering, National Chiao-Tung University, Taiwan. Her research interests include construction engineering & management, artificial intelligence, and the activity simulation and RFID tracking applications for function-space assignment in buildings.

**Kun-Chi WANG.** PhD student in the Department of Civil Engineering at National Chiao Tung University, Taiwan. His research interests include building information modeling and construction management.