Computer-Aided Civil and Infrastructure Engineering **27** (2012) 764–781

Neuro-Fuzzy Cost Estimation Model Enhanced by Fast Messy Genetic Algorithms for Semiconductor Hookup Construction

Fan-Yi Hsiao

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

Shih-Hsu Wang

Department of Civil Engineering, R.O.C Military Academy, Kaohsiung, Taiwan

Wei-Chih Wang*

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

Chao-Pao Wen

Program of Engineering Technology and Management, College of Engineering, National Chiao Tung University, Hsin-Chu, Taiwan

&

Wen-Der Yu

Department of Construction Management, Chung Hua University, Hsin-Chu, Taiwan

Abstract: *Semiconductor hookup construction (i.e., constructing process tool piping systems) is critical to semiconductor fabrication plant completion. During the conceptual project phase, it is difficult to conduct an accurate cost estimate due to the great amount of uncertain cost items. This study proposes a new model for estimating semiconductor hookup construction project costs. The developed model, called FALCON-COST, integrates the component ratios method, fuzzy adaptive learning control network (FALCON), fast messy genetic algorithm (fmGA), and three-point cost estimation method to systematically deal with a cost-estimating environment involving limited and uncertain data. In addition, the proposed model improves the current FALCON by devising a new algorithm to conduct building block selection and random gene deletion so that fmGA operations can be implemented in FALCON. The results of 54 case studies demonstrate that the proposed model has estimation accuracy of 83.82%, meaning it is approximately 22.74%, 23.08%, and 21.95% more accurate than the conventional average cost method, component ratios method, and modified FALCON-COST method, respectively. Providing project managers with reliable cost estimates is essential for effectively controlling project costs.*

*To whom correspondence should be addressed. E-mail: *weichih@mail.nctu.edu.tw*.

DOI: 10.1111/j.1467-8667.2012.00786.x

1 INTRODUCTION

High technology semiconductor fabrication has been an essential part of Taiwan's national economic growth in past decades. Numerous semiconductor fabrication plants (or fabs) have supported this development (Liu et al., 2010). Semiconductor hookup construction plays a critical role in fab completion (Hong, 2001; Beer et al., 2008). During the conceptual phase of a semiconductor hookup construction project (i.e., constructing process tool piping systems for a process module; see Section 2 for further illustration), it is preferable to conduct an accurate cost estimate for effective cost control. In practice, due to the requirement for rapid semiconductor plant construction, the semiconductor plant management team may not be able to ascertain the semiconductor process and/or the detailed requirements (such as the needed diameters and quantities of pipes) for a process module at that time. Even if the process is confirmed, the originally planned process may be altered after the plant is completed because of the rapid advancements and changes in semiconductor processing technologies. Estimating the costs of a semiconductor hookup construction project occurs in an uncertain estimation data environment.

The conventional methods used to estimate semiconductor hookup construction project costs either take the average costs from historical data or rely heavily on experienced estimators' intuition. Both existing methods perform poorly, that is, resulting in a huge gap between the estimated cost and the final project cost. Underestimated costs result in poorly allocated budget execution, whereas over-estimated costs push aside other jobs required for completing a fab (Wu, 2007; Wen, 2010). Therefore, it is necessary to develop an improved cost estimation method for semiconductor hookup construction projects to effectively control the total fab cost.

Artificial intelligence technologies, such as neural networks (NNs) (Ahmadlou and Adeli, 2010; Sedano et al., 2010), fuzzy logic (FL) (Jiang and Adeli, 2003; Lee and Pinheiro dos Santos, 2011; Ma et al., 2011; Reuter, 2011; Ross et al., 2011), and genetic algorithms (GAs) (Kim and Adeli, 2001; Carro-Calvo et al., 2010; Martínez-Ballesteros et al., 2010; Baraldi et al., 2011) have been widely used in construction engineering and project cost estimations (Creese and Li, 1995; Kim et al., 2005; Yu and Lin, 2006; Cheng et al., 2009). For instance, Creese and Li (1995) developed a back-propagation NN (Hung and Adeli, 1993; Koprinkova-Hriatova, 2010) application for the timber bridge parametric cost estimation. Boussabaine and Elhag (1997) designed a neuro-fuzzy system (NFS) related to FL for predicting the cost and duration of a construction project. Karim and Adeli (1999a,b) developed CONSCOM model for construction scheduling, cost optimization, and change-order management using neurocomputing and object technologies (Hung and Adeli, 1994; Adeli and Yu, 1995; Jiang and Adeli, 2004; Zhang et al., 2011). Senouci and Adeli (2001) presented a resource scheduling model using the neural dynamics model of Adeli and Park.

Kim et al. (2005) established a cost approximation model for residential 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) that was able to handle cost estimation missing attribute problems. Cheng et al. (2009) integrated GAs, FL, and NN technologies to establish a cost estimation model with an extremely high predictive power for construction costs. Rokni and Fayek (2010) proposed a multicriteria optimization approach for industrial shop scheduling using fuzzy set theory. In general, NNs provide the ability to learn from past data and generalize solutions for future applications. The FL allows tolerance for real world imprecision and uncertainty. GAs can be applied toward the global optimization of parameters (Cheng et al., 2009).

In a seminal book, Adeli and Hung (1995) advocated and presented the synergistic integration of the three areas of computation intelligence: NN, FL, and GA and showed how such a multiparadigm approach can help solve complicated pattern recognition problems such as face recognition and engineering design. Since then many authors have followed their multiparadigm approach, but the great majority have focused on integration of just two, such as FL and NNs (Gonzalez-Olvera et al., 2010; Li et al., 2010; Scherer, 2010; Theodoridis et al., 2010; Wang et al., 2010; Freitag et al., 2011) or FL and evolutionary computing (Iglesias et al., 2010; Patrinos et al., 2010). This work proposes an innovative model, called FALCON-COST, for estimating semiconductor hookup construction project costs (Wang et al., 2012) using NN, FL, and GA. To systematically deal with a cost-estimating environment involving limited and uncertain data, this proposed model integrates the component ratios method, fuzzy adaptive learning control network (FALCON), fast messy GA (fmGA), and three-point cost estimation method.

The remainder of this article is organized as follows. Section 2 introduces semiconductor hookup construction and three cost-estimating characteristics. Section 3 reviews some of the existing cost estimation models. Section 4 elucidates the techniques adopted by the proposed model. Section 5 presents a general description of the proposed model. Section 6 presents the details of the proposed model. Section 7 discusses the application results from 54 case studies. The results using the proposed model and three other methods are compared. Section 8 presents our conclusions and offers recommendations for future research.

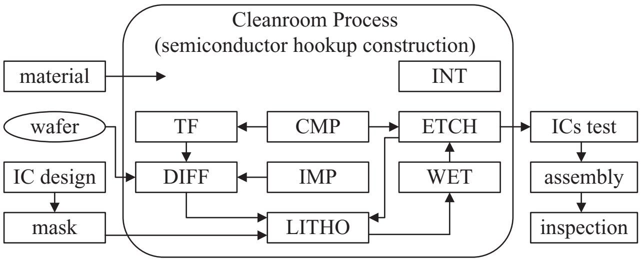

Fig. 1. Workflow of main process modules for a semiconductor fab.

2 SEMICONDUCTOR HOOKUP CONSTRUCTION

In an intensely competitive market environment, completing a fab construction project, from ground breaking to first wafer production, takes about 12 months (Chasey and Merchant, 2000). Under normal market conditions, project completion takes about 18 months, and hookup construction usually begins in the 11th month and lasts for 3 months (Wu, 2007; Wen, 2010). Figure 1 displays the main production process modules used to manufacture microchips in a semiconductor fab (Hong, 2001; Wu, 2007; Wen, 2010). There are eight process modules related to semiconductor hookup construction, including chemical mechanical planarization (CMP), diffusion (DIFF), ETCH (dry etching), integration (INT), LITHO, implant (IMP), thin film (TF), and wet etching (WET) (Hong, 2001; Wu, 2007; Wen, 2010). The construction of piping systems for a process tool related to a particular process module (or simply called a process module tool) is called a semiconductor hookup construction project (or simply called a project hereafter). Constructing the piping systems required for hookup is the last job in finishing a semiconductor fab, and is the first job to require the installation of process module tools for production (Hong, 2001). A semiconductor fab construction project requires conducting several hookup construction projects to install numerous process module tools.

A hookup construction project requires constructing some or all of the 11 types of piping systems (corresponding to 11 cost items), including (Wen, 2010): (1) bulk gas, (2) special gas, (3) chemical, (4) pumping line, (5) process cooling water (PCW), (6) ultra pure water (UPW), (7) drain, (8) power, (9) exhaust, (10) process vacuum (PV), and (11) foundation. For example, the CMP project (i.e., a project that is related to the CMP process module tool) is mainly involved in bulk gas, UPW, exhaust, and foundation piping systems. The construction for a DIFF project is related mainly to bulk gas, special gas, pumping line, and exhaust piping systems.

During the early phase of a hookup construction project, cost estimators are required to provide a best estimate to facilitate effective budget allocation for the project (Wu, 2007; Wen, 2010). In this stage, the type of process module tool is often known. However, the required quantities and specifications (such as the pipe diameters) for the piping systems are frequently unknown, primarily because tool demands/needs (such as installation locations and manufacturing brands) are still being planned (Wen, 2010). Thus, the project cost cannot be estimated using the well-known unit cost method (that is, the total cost is the summation of the products of the quantities multiplied by the corresponding unit costs) (Hendrickson and Au, 2003). The conventional approach uses an average cost method according to historical project cost data. That is, the estimated total cost for a new project is the summation of the average costs of the 11 cost items in historical projects. In addition, cost estimates are sometimes simply based on the experience of estimators to directly generate total project cost. The accuracy of estimates made with both the average cost method and the experience-based method are often unacceptable, leading to poor budget planning (Wen, 2010).

Accurately estimating hookup construction project cost during the conceptual phase is difficult because cost estimators must base their calculations on limited and uncertain cost data. To be effective, cost estimation models must deal with three cost-estimating characteristics. The first characteristic is that only limited data is available for estimating new projects. As indicated earlier, typically only some piping systems dominate the work invested in each project. The second characteristic is that the relationships (reasoning rules) between the costs of the dominant piping systems and total project cost are complicated. The third characteristic is that the details (e.g., quantities and specifications) of the piping systems are uncertain in evaluating a new project cost. Therefore, estimator experience is still required.

3 REVIEW OF COST ESTIMATION MODELS

Accurately estimating costs is an essential task in effectively managing construction projects. Several cost estimation models have been proposed to account for the effects of uncertainties. These recent cost models involve FL (Sarma and Adeli, 2000, 2002), NNs (Adeli and Wu, 1998; Adeli and Karim, 1997, 2001), simulation (Wang, 2002; Wang et al., 2007; Wang et al., 2008; Sadeghi et al., 2010; Chou, 2011), and other systematic approaches (Diekmann and Featherman, 1998; Oberlender and Trost, 2001).

Numerous conceptual cost estimation methods, such as unit cost, cost indices, cost-capacity factors, component ratios, and parametric estimation, have been designed to quickly compute a total project cost in the construction industry (Sarma and Adeli, 1998; Barrie and Paulson, 1992; Hendrickson and Au, 2003; Hong et al., 2011). Cost indices focus on cost changes over time, whereas cost-capacity factors apply to changes in size, scope, or the capacity of similar projects (Barrie and Paulson, 1992). They reflect the increase in cost with size, as a result of economics of scale.

The parametric estimation method has been widely applied. The parametric estimation method takes a single parameter (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 (Hendrickson and Au, 2003; PEH, 2008). A parametric cost estimation method includes one or several cost-estimating relationships between the cost (the dependent variables) and the cost-governing parameters (the independent variables) (Hegazy and Ayed, 1998; PEH, 2008). Cost indices can also be incorporated into the parametric estimation method for reflecting the cost changes over time (Oberlender, 2000; Barrie and Paulson, 1992).

Parametric cost estimation methods are often used by both contractors and government bodies in the project planning and budgeting stages (Hegazy and Ayed, 1998). Several parametric estimation methods based on regression analysis and NNs have been suggested to improve the accuracy of conceptual cost estimates (Sonmez, 2008; Gunduz et al., 2011). The regression technique allows a relatively simple analysis to sort out the impact of the parameters on the cost of a project (Lowe et al., 2006). NNs based on artificial intelligence offer an alternative approach to estimate the costs of building projects (Kim et al., 2005) and highway projects (Hegazy and Ayed, 1998).

Generally, the current conceptual cost-estimating methods focus on the level of total project cost (i.e., they usually do not examine any cost divisions or cost item details) and generate estimates that can vary widely in terms of accuracy. In addition, current methods have been developed for various projects, such as building projects, highway projects, and oil refinery projects. However, no methods have been developed for capturing the aforementioned three cost-estimating characteristics encountered in semiconductor hookup cost estimations (Wen, 2010).

4 REVIEW OF RELATED TECHNIQUES

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

4.1 Component ratios method

In the component ratios method (also called equipment installation cost ratios, plant cost ratios, or ratio estimating method), it is assumed that a ratio (or factor) exists between the total project cost and the cost of a major cost item (Barrie and Paulson, 1992). Hence, when the cost of the major cost item and the ratio (=total project cost divided by major item cost based on historical data) are known, the total project cost can be calculated by multiplying the major item cost by the ratio (greater than 1.0). A variation on this component ratios method takes the cost of each major item separately, multiplies each by its own ratio, then takes the sum of the factored items (Barrie and Paulson, 1992).

Following the component ratios method concept, Yu (2006) further developed a principal-item ratio estimation method (PIREM) by considering only the selected 20% cost items (called principal items) and their associated principal item ratios to calculate the overall cost. This "20%" number is determined according to the Pareto Optimum Criterion (named "80/20 principle") which implies that 80% of the overall project cost is determined by 20% of the cost items (Koch, 1997). Yu (2006) discussed public civil construction projects and building projects (unlike the present investigation). However, his study encouraged the belief that focusing on certain principal costs could not only produce acceptably accurate estimates, but it could also save estimation effort and time. Hence, the 80/20 principle is used to identify major cost items (to support the component ratios method adopted in this study) with the difference that the sum of the principal item costs does not exactly equal 80%. See the model for further illustrations.

4.2 FALCON

FALCON, one of the NFSs, is a fuzzy system that uses a learning algorithm derived from NN theory to determine its parameters by processing data samples (Lin and Lee, 1991). Lin and Lee (1991) developed FALCON to solve system control problems in electronics and manufacturing engineering. However, FALCON has been utilized to acquire construction knowledge due to its numerous features, such as the ability to handle uncertainties and trace-back functions for problem solving (Yu and Skibniewski, 1999). Furthermore, the FALCON network structure graphically shows how it captures the complex IF-THEN reasoning rules (Lin and Lee, 1991). Most importantly, FALCON has been modified to support conceptual cost estimation in building projects (Yu and Lin, 2006; Yu, 2007).

FALCON's learning ability is based on the Kohonen learning rule and supervised learning algorithm. In the traditional FALCON methodology there is no mechanism for rule refinement after Kohonen learning. As indicated by Yu and Skibniewski (1999), it has been found that after many computational experiments the two learning algorithms (the competitive learning for initial rule connection and the back-propagation for fine-tuning of membership functions) in the FALCON methodology encounter severe local optimal problems. A local optimum of a combinatorial optimization problem is a solution that is optimal (either maximal or minimal) within a neighboring set of solutions. This contrasts with a global optimum, which is the optimal solution among all possible solutions.

First, as the FALCON FL rules are determined by the fuzzy partitions of the linguistic terms defined for input attributes by the decision maker, it results in enormous FL rules that contain redundant precondition links and unnecessary consequence links. Such a problem is essentially due to using two algorithms in the traditional FALCON method. Using the Kohonen learning rule first to roughly determine the fuzzy membership functions of fuzzy linguistic terms may have imposed an erroneous precondition structure for the FL rules. As a result, the consequence links obtained by competitive learning rule (based on the precondition structure previously determined) in the unsupervised learning phase in the traditional FALCON may be erroneous. Second, back-propagation is adopted for supervised parameter learning in fuzzy membership functions (for both input and output layers) on the primitive fuzzy rules determined in the structure learning. Because the primitive fuzzy rules are erroneous, the supervised parameter learning results may be easily captured in local optimum (Lin and Lee, 1996; Yu and Skibniewski, 1999).

Because there is no mechanism to revise the FL rules using back-propagation once they are determined, the local optimum problem cannot be improved in the traditional FALCON method (Lin and Lee, 1996). Yu (2007) and Cheng et al., (2009) suggested adopting the messy GA (mGA) and fmGA, respectively, for structure revision and parameter learning in a NFS, in which FALCON is one kind of NFS. This study, thus, applies the fmGA mutation and cut-splice operators to revise the fuzzy membership functions and FL rules of FALCON to improve the cost estimation accuracy.

4.3 fmGA

GAs, originally proposed by Holland (1975), are search algorithms and they search through a decision space for optimal solutions based on the mechanics of natural selection and genetics. Using GAs for civil engineering problem solutions may go as far back as 1993 (Adeli and Cheng, 1993, 1994a,b; Adeli and Kumar, 1995a,b). GAs have also been applied in other disciplines such as construction engineering (Cheng and Yan, 2009; Al-Bazi and Dawood, 2010), transportation engineering (Lee and Wei, 2010; Putha et al., 2012), highway engineering (Kang et al., 2009), and structural engineering (Marano et al., 2011).

To explore an individual gene's contribution to the fitness value during the evolution process, mGA was developed (Goldberg et al., 1989). Unlike the simple GAs which use fixed length strings to represent possible solutions, Goldberg et al. (1993) further developed the fast mGA (fmGA) to apply messy chromosomes to form strings of various lengths.

The fmGA chromosome is divided into two parts: allele locus and allele value. The allele locus represents the allele serial number. The allele value is the value of the allele serial number. The major difference between the fmGA and traditional GA lies in the fact that the fmGA allows for variations in the chromosome lengths and the allele locus and allele value evolution may happen simultaneously. This variable chromosome length characteristic provides a desirable capability for FALCON structural revision because the optimum precondition and consequence links structure for the fuzzy rule base may be obtained via the fmGA evolution process.

Four distinct features differentiate the fmGA from the traditional GA (Feng and Wu, 2006): (1) variable length chromosomes can be adopted in fmGA; (2) simple cut and splice is used to replace the GA operator mechanism; (3) the optimization process contains a primordial phase and a juxtapositional phase; and (4) competitive templates (CTs) are adopted to retain the most outstanding gene building blocks (BBs) in each generation.

After applying the cut-splice operator to the chromosome, the problems of chromosomes being over- or under-specified may result (Feng and Wu, 2006). If the chromosomes are over-specified the fmGA will screen out repeated genes from left to right on a first-come-first-served basis. If the chromosomes are under-specified the fmGA will make the chromosome with the optimum fitness in the previous generation be the CT and make up for the missing genes.

5 GENERAL DESCRIPTION OF THE PROPOSED MODEL

This section provides an overview of the proposed cost-estimating model, called FALCON-COST. The modeling steps are displayed in the left part of Figure 2. Restated, the component ratios method, FALCON, and fmGA are integrated to generate the original FALCON-COST that will be trained (or learned) from

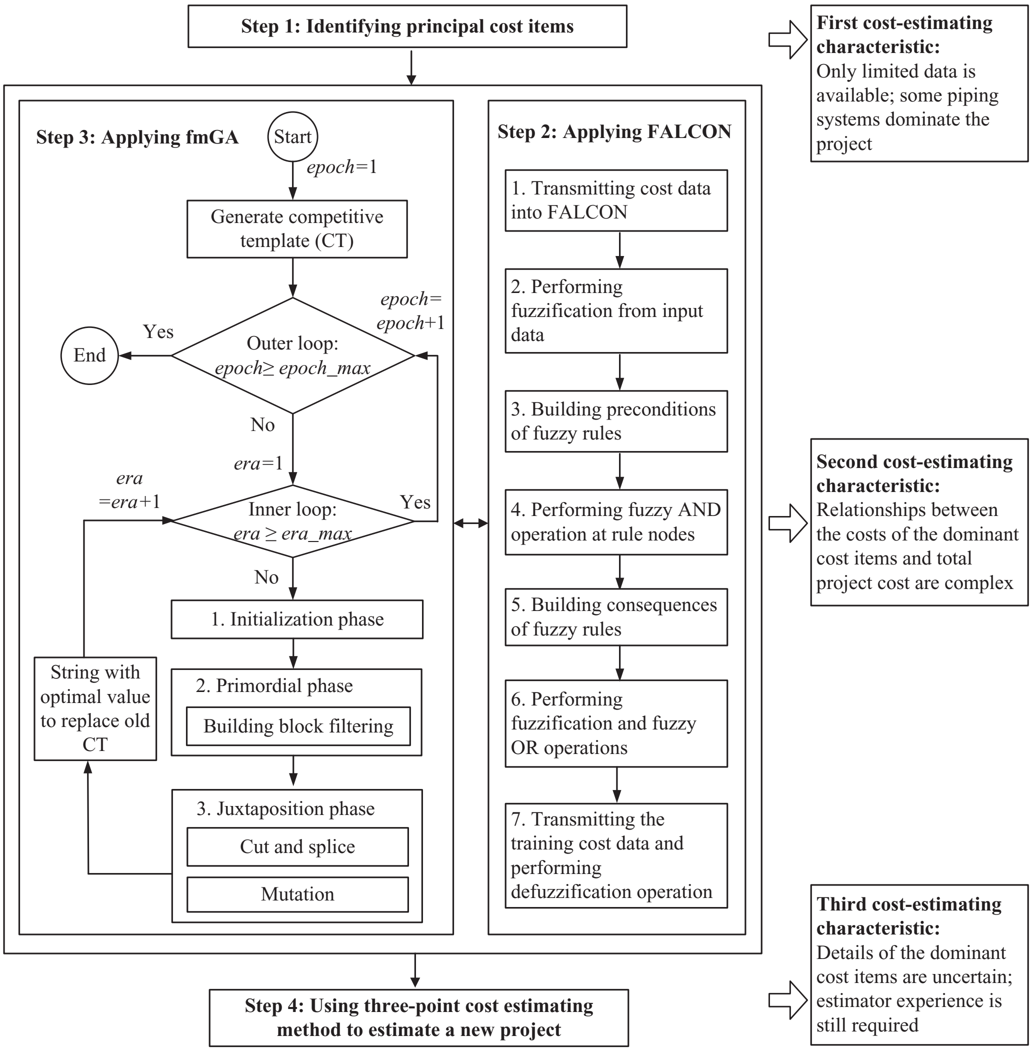

Fig. 2. Proposed model to meet three cost-estimating characteristics.

historical projects. See Steps 1–3. The three-point estimation method is then applied to support the trained FALCON-COST to predict the total cost of a new project. See Step 4.

This proposed model aims to systematically guide cost estimators to conduct their estimations for dealing with the above three cost-estimating characteristics. See the right part of Figure 2. Namely, the component ratios method reflects the first characteristic and focuses exclusively on the cost items of those dominant piping systems. The FALCON and fmGA methods are used to solve the complex relationships between those dominant cost items and total project costs (second characteristic). The three-point estimation method deals with the third characteristic in assessing the uncertainties of the dominant cost items in a new project. The major modeling steps of the FALCON-COST are described as follows.

Step 1—Unlike many existing conceptual cost estimation models that center on the level of total cost, the proposed model predicts the total costs by analyzing the item-level costs. To reflect the environment lacking sufficient and certain data for cost estimating, the component ratios method is adopted to identify the major (or principal) cost items to forecast the total cost of a project. In this study, 241 historical projects are used to indicate the principal cost items of the projects for the eight types of process modules.

Step 2—The FALCON is used to learn the relationships between the principal item costs (inputs) and the corresponding total cost (output) of each historical project for each type of process module. The main FALCON operations include (Lin and Lee, 1991): calculating membership functions from network input, performing the "*fuzzy AND*" operation to determine the fired rules, and performing the "*fuzzy OR*" operation

Table 1
Training and test project amounts in each process module

Module	*Training project amounts*	*First set of test project amounts*	*Second set of test project amounts*
CMP	15	2	2
DIFF	41	4	4
ETCH	36	4	4
INT	29	3	3
LITHO	16	3	3
IMP	30	3	3
TF	27	3	3
WET	47	5	5
Total	241	27	27

to aggregate the linguistic fuzzy cost estimation term memberships and finally carrying out defuzzification to derive a total project cost estimate (output value).

Step 3—To overcome the local optimum problem caused by FALCON, FALCON's fuzzy membership functions and FL rules are optimized through the fmGA mutation and cut-splice operators to enhance the cost estimation accuracy. After completing this step the training process for the original FALCON-COST is finished.

Step 4—To compute the cost of a new project, the three-point cost estimation method is applied to determine the expected cost of each principal cost item. These expected principal item costs are then treated as the inputs for the trained FALCON-COST to generate the total cost of the project.

6 DETAILS OF THE PROPOSED MODEL

This section illustrates the detailed FALCON-COST development. Semiconductor hookup construction consists of eight process modules. A cost estimation model must be established for each individual module because the piping systems related to each module vary.

6.1 Historical projects

Two hundred forty-one historical projects related to the same semiconductor plant were used to develop and train the FALCON-COST. The left and right parts of Table 1 present the training and test project amounts, respectively, for the eight types of process modules.

6.2 Step 1: Identifying principal cost items

Based on the component ratios method, three to four principal cost items of each project for each process module are indicated. Table 2 lists the average cost and the percentage of each principal cost item identified in each module. Notably, the costs are in New Taiwan dollars ($1 U.S. dollar ≅ $30 New Taiwan dollars). For instance, four cost items from 15 historical CMP projects are identified to have the highest cost percentages. Restated, the cost percentages are 8.7%, 15.7%, 33.7%, and 16.5% for the bulk gas, UPW, exhaust, and foundation piping systems, respectively. Because the sum of these percentages of cost account for a high portion (about 74.9%) of the total cost, these four items are called the principal cost items. The principal cost items for the projects for seven other process modules were also determined using a similar process. The principal item costs (inputs) and the corresponding total cost (output) for each historical project are used as the training data in the proposed model.

6.3 Step 2: Applying FALCON

A FALCON network structure consists of five layers of nodes and two links, including: the input linguistic nodes (Layer 1), input term nodes (Layer 2), IF-part condition links (Link 1), rule nodes (Layer 3), THEN-part consequence links (Link 2), output term nodes (Layer 4) and output linguistic nodes (Layer 5) (Lin and Lee, 1991). Layer 2 and Layer 3 are connected by Link 1, whereas Layer 3 and Layer 4 are connected by Link 2. The computation results for each node will be passed on to the next layer of nodes through the neuron synaptic weights and become the input value for the next layer. A description of these FALCON layers and links is further illustrated below (Lin and Lee, 1991).

1. Layer 1 (input linguistic nodes): The nodes in this layer just transmit the input values (i.e., cost data) to the next layer directly. For example, the actual costs of the four principal cost items (i.e., bulk gas, UPW, exhaust, and foundation) for a CMP project are transmitted directly into the network.
2. Layer 2 (input term nodes): The nodes in this layer are responsible for calculating the membership functions. That is, this layer conducts fuzzification on the input values (i.e., cost data) from Layer 1. Fuzzy partitions are determined based on the clustering relationships of the principal item costs and the total costs. For instance, Figure 3 depicts the clusters identified in each principal item after conducting the fuzzy partitions for the CMP module. Based on the cost data graphic distribution, two clusters (high and low) are identified for

each of the bulk gas, UPW, and exhaust principal items, whereas three clusters (high, medium, and low) are for the foundation principal item. As a result, the input parameters are partitioned according to the number of identified clusters.

3. Link 1 (IF-part condition links): The connections between Layer 2 and Layer 3 represent the fuzzy IF-THEN rule preconditions. Take the CMP module for example. Based on Layer 2 operations there are 2, 2, 2, and 3 clusters identified for the bulk gas, UPW, exhaust, and foundation items, respectively. Therefore, 24 (=2×2×2×3) IF-part FL rules are generated.
4. Layer 3 (rule nodes): The rule nodes in this layer perform the "*fuzzy AND*" operation to derive the fired strength of various FL rules. For example, in the CMP module, four input cost parameters are involved. These input parameters are fuzzified in Layer 2 (input term nodes) with corresponding fuzzy partitions (i.e., 2, 2, 2, and 3). A membership value (ranging from 0 to 1) is then given to each input term node. Every rule node is connected to five output term nodes in Layer 4 for the CMP module.
5. Link 2 (THEN-part consequence links): The links between Layer 3 and Layer 4 present the consequences of FL rules. There should be no more than one consequence for each rule node in a single output network. The links are represented as numeric value 0 (disconnected) or 1 (connected).
6. Layer 4 (output term nodes): The nodes at this layer perform two functions, *right–left* (only performed in training stage) and *left–right* (for both training and usage stages) transmissions. In *right–left* transmission the training data (i.e., actual total project cost) from the output layer (i.e., Layer 5) are transmitted into Layer 4. Thus, the fuzzy operation of this layer is exactly the same as Layer 2. That is, the output cost data is mapped through the membership functions of the output fuzzy linguistic terms. In *left–right* transmission the output term nodes carry out the "*fuzzy OR*" operation to sum up the membership functions of the fired rules obtained from Link 2.
7. Layer 5 (output linguistic nodes): The nodes in Layer 5 also perform *right–left* (only for training stage) and *left–right* (for both training and usage stages) transmissions. In *right–left* transmission the nodes at Layer 5 act precisely the same as Layer 1, that is, feeding the training data (i.e., actual total project cost) into the network. In *left–right* transmission the nodes at Layer 5 perform the

Table 2

Average costs and percentages of the principal cost items in each process module

Module / *Cost item*	*CMP*	*DIFF*	*ETCH*	*INT*	*LITHO*	*IMP*	*TF*	*WET*
Bulk gas	$143,459 8.7%	$282,234 16.4%	$768,333 19.0%	$67,950 14.1%	$215,554 12.8%	$151,807 15.1%	$857,853 24.0%	$282,712 13.8%
Specialty gas		$245,510 14.3%	$819,391 20.2%				$787,540 22.0%	
Pumping line		$412,006 24.0%					$507,120 14.2%	
PCW					$186,562 11.1%			
UPW	$258,308 15.7%							$430,736 21.0%
Drain								$267,435 13.1%
Power				$62,002 12.8%		$99,932 9.9%		
Exhaust	$554,114 33.7%	$293,638 17.1%	$1,342,478 33.1%		$379,487 22.6%	$278,059 27.6%	$551,778 15.4%	$465,667 22.7%
Foundation	$271,050 16.5%			$242,613 50.2%	$367,176 21.9%	$190,682 18.9%		
Chemical								
PV								
Sum of percentages	74.9%	71.3%	68.9%	80.9%	69.0%	76.3%	70.4%	68.7%
Amount of projects	15	41	36	29	16	30	27	47

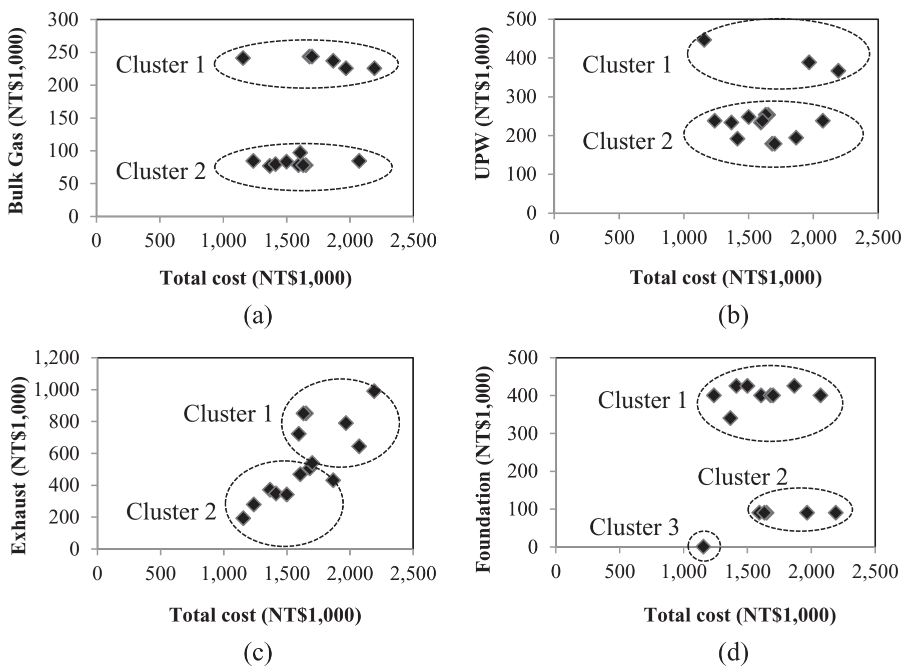

Fig. 3. Clusters identified in each principal item for the CMP module. (a) Bulk gas, (b) UPW, (c) exhaust, and (d) foundation.

defuzzification of fuzzy set to provide a definite output value (i.e., estimated total project cost).

6.4 Step 3: Applying fmGA

After completing the FALCON operations the fmGA's mutation and cut-splice operators are then used to optimize the FALCON's parameters including fuzzy membership functions and FL rules for improving the cost estimation accuracy. To do so, the fmGA variable-length chromosome is utilized to revise the fuzzy partitions (i.e., the number of input term nodes in Layer 2) and fuzzy decision rules (i.e., the consequence links of Link 2) of FALCON. The fmGA global search capability is employed to optimize the parameters (means and spreads) of the membership functions in FALCON input and output term nodes.

To perform the above-mentioned functions, the fmGA chromosome models a set of FALCON solutions, where the parameters (e.g., input membership functions, FL rules, and output membership functions) of FALCON are represented as the chromosome gene values. For example, Figure 4 presents the composition of a sample chromosome for the CMP module. Every membership function contains a mean and a spread. The numbers of fuzzy partitions for the input and output linguistic nodes are [2, 2, 2, 3] and [5], respectively. Thus, there are 52 alleles in an fmGA chromosome. Restated, 52 alleles $= 18$ ($=2\times2 + 2\times2 + 2\times2 + 3\times2$ input membership parameters) $+ 24$ ($=2\times2\times2\times3$ rules) $+ 10$ ($=5\times2$ output membership parameters).

As also presented in the left part of Figure 2, fmGA includes two operation loops, an outer loop and inner loop. Finishing an outer loop is called an epoch, whereas conducting an inner loop is called an era. As suggested by Feng and Wu (2006), this study defines the maximum number of eras (era_max) as 4. In addition, the maximum number of epochs (epoch_max) is defined as a preset criterion for terminating the fmGA evolution process. In this study, epoch_max is 10. 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 BBs of the order k, where BBs refer to partial solutions of a problem; (2) primordial phase—bad genes are filtered out to maintain only the chromosomes with good fitness (i.e., containing only "good" alleles fitting to BBs); and (3) juxtapositional phase—those good alleles (BBs) are rebuilt using cut-splice and mutation operations to form a high quality generation that tends to generate an optimal solution.

As depicted in Figure 2 the fmGA starts with the outer loop and generates a CT. For example, the CT for the CMP module is an fmGA chromosome representing a set of FALCON solutions (refer to Figure 4). After completing one era, the CT will be replaced by a new CT (with new alleles) with the best fitness (i.e.,

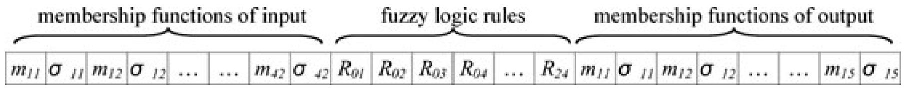

Fig. 4. Chromosome composition in fmGA for the CMP module.

the highest estimation accuracy) found in that era. The operational details for the three phases are further described as follows.

6.4.1 Initialization phase. As suggested by Feng and Wu (2006), the chromosome population size (n) in this study is determined by Equation (1) to ensure a sufficient quantity of chromosomes:

$$n = \frac{\binom{l}{\lambda}}{\binom{l-k}{\lambda-k}} 2c(\alpha)\beta^2 (M-1) 2^k \qquad (1)$$

where

- l is the chromosome length. For example, the value of l will be 52 (=18 + 24 + 10) for the CMP module. See Figure 4.
- k is the number of fuzzy rules. For example, the value of k should be set as 24 (=2×2×2×3) for the CMP module. However, computations will be overwhelming if $k = 24$ is applied to Equation (1) for deriving n. Hence, $k = 4$ is used here as suggested by Goldberg et al. (1993).
- λ is a random value generally set to be l–k, $k<\lambda\leq l$. For example, the value of λ will be 48 (=l–k = 52–4) for the CMP module.
- $c(\alpha)$ where α is probability square of a normal distribution, which is set to be 1.
- β is the ratio of a chromosome with the optimum fitness to those with the second best fitness in the same era; it is set as 1.
- M is the BBs' coefficient, it is set as 2.

The fitness of a chromosome is evaluated based on the estimation accuracy, defined in Equation (2). This estimation accuracy, in terms of percentage indicates the difference between the estimated total cost and the actual total project cost.

$$\text{Accuracy}(\%) = \left(1 - \frac{\text{ABS(Estimated cost} - \text{Actual cost)}}{\text{Actual cost}}\right) \times 100\% \qquad (2)$$

6.4.2 Primordial phase. This phase performs two operations, namely building-block filtering and threshold selection (Goldberg et al., 1993). The building-block filtering includes BBs selection and random allele deletion. A BB is a set of alleles, which are a subset of strings that are short, low-order, and high performance. The key to building-block filtering is to pump enough copies of the good BBs so that even after random allele deletion eliminates a number of copies, one or more copies remain for subsequent processing (Goldberg, 2002). In addition, in threshold selection, a genetic threshold mechanism (also called tournament selection) is applied to restrict competition between BBs that have little in common (Goldberg et al., 1991).

The BBs in this study are built to represent the FALCON parameters, including the means (m_{ij}) and spreads (σ_{ij}) of input and output term nodes and the fuzzy rule links (R_{ij}) of Link 2. The building-block filtering process details are illustrated as follows:

1. In BBs selection, a chromosome with the best fitness in the previous era is picked to be the CT. The alleles of FL rule nodes for this CT are selected as the BBs that are used to replace the alleles of FL rule nodes for the reproduced chromosomes in the next era.
2. In random allele deletion these BBs will replace the genes of FL rules for the other 80% of next-era chromosomes with worse fitnesses. The alleles of the means and spreads for those chromosomes will be randomly deleted by 5% (in the CMP module, 5%≒3 alleles) in each era. The deleted alleles are replaced by the alleles stored in the CT. The minimum number of alleles in the chromosomes is kept the same as the number of BBs (i.e., 24 for the CMP module) after the deletion, as suggested by Goldberg (2002).

In addition to the manipulation of BBs that enrich chromosome diversity, this study adopts a new algorithm that combines Kohonen, competitive learning rules, and the fmGA operations to enhance the traditional FALCON learning rules for escaping from the aforementioned local optimum problem.

6.4.3 Juxtapositional phase. In each era, two outcomes can result, the fitness value of a specific chromosome is higher or lower than (or equal to) that of the CT in the previous era:

1. If the fitness value is higher than that in the previous era, it means a better fitness estimation result is generated through the cut-splice operator.

The cut and splice operation is then performed. The one-cut point of chromosomes is randomly selected. Both the splice rate and the cut rate in this study are set to 1. All chromosomes except the one selected for the CT are evolved.

2. If the fitness value is lower than (or equal to) that in the previous era, it implies that no-better-estimation result can be generated using the cut-splice operation alone. The mutation operator is therefore employed. In performing the mutation operation the mutation probability (P_m) is set to 5% (as suggested by Goldberg et al., 1993). All parameters (such as the input membership functions, fuzzy rules, and output membership functions) of the chromosomes are mutated and the allele locus to be mutated is selected randomly.

After evolution the chromosome with highest fitness is fed back to FALCON for calculating the cost estimations of the new input data. Best-fit chromosomes (with optimum fitness) will also be maintained by fmGA to provide the population and the CT of the next epoch. Then Steps 2–3 (FALCON and fmGA operations) are repeated iteratively until the fitness value converges, or it has reached a preset maximum era number (set to be 10 in this study). Finally, the fmGA operations stop and the final optimal FALCON-COST structure is derived. Notably, three types of data are required to train the original FALCON-COST: (1) principal item costs and the total cost for each historical project, (2) FALCON's fuzzy partitions, and (3) fmGA's maximum numbers of evolution era and epoch.

6.5 Step 4: Using three-point cost estimation method to estimate a new project

Although the required quantities and specifications of the piping systems for a new project are uncertain, the costs of the principal items (for the piping systems) must be provided to run the proposed model for meeting the third cost-estimating characteristic mentioned above. For instance, in estimating the total cost of a CMP project, the costs of the four principal items (i.e., bulk gas, UPW, exhaust, and foundation) should be derived. The estimation guess for each principal cost item is performed by asking the question: how much cost will be higher and/or lower (in terms of percentage) than the average historical cost, based on his knowledge on this principal item of the project? As indicated by several cost-estimating mangers specialized in fab construction, an experienced manager should be able to make a reasonable guess of the cost of each principal item for a project based on the available cost information.

To increase the objectivity of input evaluations, the widely-used three-point estimation approach is adopted to systematically guide a cost-estimating manager to assess the uncertainties surrounding the costs of principal items (Moder et al., 1983; Oberlender, 2000; Peurifoy and Oberlender, 2002). By introducing an expected percentage variable, the expected cost (denoted as $C_{i(j)}$) of each principle item j ($j = 1, 2, 3$, and/or 4) for a project related to a process module i ($i = 1, 2, \ldots, 8$) is derived as,

$$\begin{aligned} C_{i(j)} &= EP_{i(j)} \times C_{i(j)(\mathrm{ave})} \\ &= \frac{a_{i(j)} + 4m_{i(j)} + b_{i(j)}}{6} \times C_{i(j)(\mathrm{ave})} \end{aligned} \tag{3}$$

in which $EP_{i(j)}$ is an expected percentage variable of $C_{i(j)}$. Furthermore, $a_{i(j)}$, $m_{i(j)}$, and $b_{i(j)}$ are the optimistic, most likely, and pessimistic values (expressed as percentages) of $EP_{i(j)}$, respectively. $C_{i(j)(\mathrm{ave})}$ is the average historical cost of $C_{i(j)}$. Notably, in estimating a new project, the calculated expected costs ($C_{i(j)}$) of the three (or four) key items serve as inputs to the FALCON-COST model for predicting total project cost.

6.6 Computer implementation

The operations of Steps 2 and 3 (FALCON and fmGA) are built with Matlab® version 7.5. The cost data are read in .dat format and the operations are run under the Genuine Intel 1.6GHz CPU, 896MB SRAM, and Windows XP computer operating systems. Training the FALCON-COST for the 15 historical projects of the CMP module will take approximately 20 minutes. The Steps 1 and 4 operations for FALCON-COST are performed in Microsoft Excel.

6.7 Training of the FALCON-COST

The actual costs of 241 historical projects are used to train the proposed models (i.e., the algorithms related to the FALCON and fmGA steps**)** with respect to the eight types of process modules. For example, 15 historical projects are used to train the CMP model. The training helps develop the model to reduce its error rate (=1−estimation accuracy) by running for several evolutions (=40 eras). Figure 5 depicts the model training error rate in each epoch of the CMP model after the evolutions reached 10 epochs. At that time the estimation accuracy is about 97.85%, and thus the error rate has been reduced to only around 2.15% (=1–0.9785). The same processes are used to train the models for the other seven process modules.

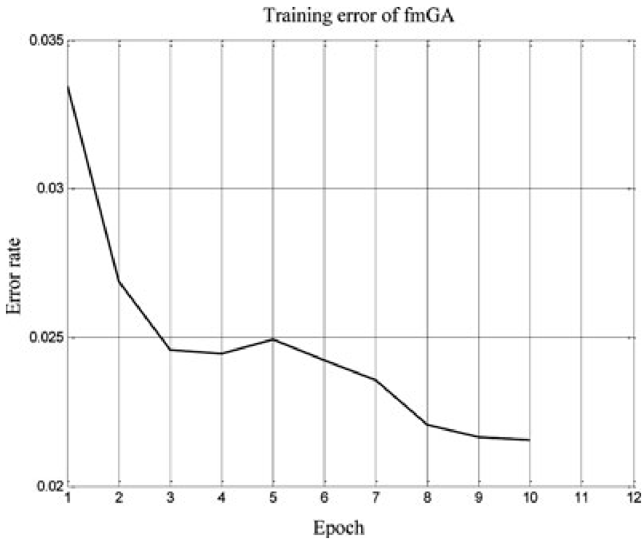

Fig. 5. Model training error rate in each epoch for the CMP module.

7 CASE STUDIES

A two-step cross-fold validation process is conducted to test the proposed model, and is described as follows. First, a set (first set) of 27 additional hookup construction projects (project numbers 1–27) for the same semiconductor plant is used to test the proposed model that is trained using the aforementioned 241 projects. After completing the first set of 27 case studies, the validation process proceeds to the next step. Namely, a second set of 27 test projects (project numbers 28–54) are randomly selected from the original 241 training projects. Notably, 241 projects are still involved in the training because the first set of 27 test projects are entered into the training pool. A total of 54 projects are tested. The middle and right columns of Table 1 list the first and second sets of test project amounts for each process module, respectively.

Section 7.1 illustrates how the FALCON-COST is applied to estimate the cost of a new project. Section 7.2 compares the results of the 54 case studies with those using the conventional average cost method, the component ratios method, and a modified FALCON-COST method. Section 7.3 discusses the FALCON-COST improvements over the conventional FALCON.

7.1 Case project application

As indicated earlier, a cost-estimating manger must provide three-point estimations for the cost of each principal item for a project. The expected cost, $C_{i(j)}$, for each principle item in the project can then be derived. These derived expected costs for the principle items for the project are used as inputs to calculate the total project cost using FALCON-COST Steps 2 and 3. Notably, all the training projects and test projects are related to a single company; additionally, a cost-estimating manager involved in the 54 test projects was asked to provide the inputs realistically and consistently.

Take the No. 1 test project (a CMP project) shown in Table 3 for example. To estimate the expected cost ($C_{i(j)}$) of the bulk gas (a principal item) for the new project the manager inputs the $a_{i(j)}$, $m_{i(j)}$, and $b_{i(j)}$ values. That is, based on his understanding of the requirements for this new project, he provides the optimistic % (or lowest %), most likely %, and pessimistic % (or highest %) of the average costs from historical projects. In this example, the $a_{i(j)}$, $m_{i(j)}$, and $b_{i(j)}$ values are 92%, 102%, and 105% of the average cost, respectively. Thus, the expected percentage of the average cost for the bulk gas for this new project, $EP_{i(j)}$, equals 100.8333% (=(92%+4×102%+105%)/6). The average historical cost of $C_{i(j)}$ for CMP projects, $C_{i(j)(ave)}$, is \$143,459. Hence, the expected cost ($C_{i(j)}$) of the bulk gas is \$144,645 (=100.8333%×\$143,459) according to Equation (3).

Similarly, in the No. 1 test project, the expected costs ($C_{i(j)}$) of the other three principal piping items (i.e., UPW, exhaust, and foundation) are calculated as \$260,461, \$577,202, and \$266,532, respectively. Table 3 summarizes the calculated expected costs of the four principal items for this test project. These four $C_{i(j)}$ values are used as inputs for generating the estimated total cost of this new project. The estimated cost of this project using the proposed model is \$2,223,200.

The actual cost of this project was \$2,022,555. Thus, the estimation accuracy of the proposed model is 96.10% (=1–ABS(1–2,223,200/2,022,555)) according to Equation (2). Similarly, the evaluation steps are also applied to the other 53 test projects. Table 4 lists the evaluation results for the first set of 27 test projects.

7.2 Comparisons with three methods

The conventional average cost method, the component ratios method, and a modified FALCON-COST are applied to the 54 test projects. The No. 1 test project is also used to illustrate how these three methods work. When the conventional average cost method is utilized, the estimated cost of the test project equals \$1,642,981 (=total costs of 15 CMP historical projects divided by 15). Because the actual cost of this project was \$2,022,555, the estimation accuracy of the average cost method is 81.23% using Equation (2).

In the component ratios method (Barrie and Paulson, 1992), the four cost items: bulk gas, UPW, exhaust, and foundation for the CMP project are identified as the principal items. The averaged ratio between the whole

Table 3
Calculated expected costs of the principal items for No. 1 test project

Principal item	*Average cost $C_{i(j)(\mathrm{ave})}$*	*Optimistic %, $a_{i(j)}$*	*Most likely %, $m_{i(j)}$*	*Pessimistic %, $b_{i(j)}$*	*Expected cost, $C_{i(j)}$*
Bulk gas	$143,459	92%	102%	105%	$144,654
UPW	$258,308	75%	95%	150%	$260,461
Exhaust	$554,114	60%	105%	145%	$577,202
Foundation	$271,050	30%	105%	140%	$266,532

Table 4
Estimated costs and estimation accuracies using the proposed model for the first set of 27 test projects

Module	*Test project number*	*Estimated cost ($)*	*Actual cost ($)*	*Accuracy (%)*	*Average of accuracy (%)*	*Standard deviation (%)*
CMP	1	2,101,500	2,022,555	96.10	97.02	1.30
	2	2,141,600	2,186,601	97.94		
DIFF	3	2,194,600	2,466,930	88.96	84.28	11.61
	4	2,128,100	2,176,044	97.80		
	5	2,009,800	2,534,026	79.31		
	6	2,067,600	2,910,376	71.04		
ETCH	7	2,849,700	2,745,135	96.19	79.93	12.07
	8	3,707,200	4,922,227	75.32		
	9	3,929,700	5,809,517	67.64		
	10	3,492,600	2,924,495	80.57		
INT	11	570,090	784,678	72.65	89.19	14.37
	12	570,090	578,209	98.60		
	13	570,090	549,890	96.33		
LITHO	14	2,518,400	2,689,975	93.62	91.44	9.66
	15	2,519,000	2,523,427	99.82		
	16	2,477,900	2,079,944	80.87		
IMP	17	1,359,500	1,225,856	89.10	83.88	7.42
	18	1,787,700	1,584,328	87.16		
	19	811,970	1,076,994	75.39		
TF	20	2,261,300	3,983,913	56.76	78.06	20.87
	21	2,144,600	2,716,373	78.95		
	22	2,130,600	2,163,491	98.48		
WET	23	2,519,100	2,860,966	88.05	90.05	9.05
	24	2,815,400	2,830,734	99.46		
	25	3,183,100	3,433,748	92.70		
	26	2,512,600	2,660,107	94.45		
	27	2,507,400	3,316,608	75.60		

project cost and the sum of the costs of these principal items is 1.335 (=100%/74.9%; see Table 2) for the 15 historical CMP projects. Because the sum of the averaged costs of these principal items is $1,226,931 for the same historical projects, the total cost of a new project is $1,637,732 (=1,226,931×1.335). Table 5 summarizes the calculations using the component ratios method for the No. 1 test project. Because the actual cost of this project was $2,022,555, the estimation accuracy using the component ratios method is 80.97%.

The modified FALCON-COST revises the details of Step 4 in Figure 2. Restated, the historical average cost ($C_{i(j)(ave)}$) rather than the expected costs ($C_{i(j)}$) (obtained using three-point estimations) of the principal items is used to estimate the project cost. For example, in the No. 1 test project, the average costs of the four cost items listed on the left of Table 3 are directly used as inputs for the FALCON-COST. When the modified FALCON-COST is applied, the estimated cost of the test project equals $2,080,100. Because the actual cost of this project was $2,022,555, the estimation accuracy using Equation (2) is 97.15%.

Similar evaluation steps are also applied to the other 53 test projects. Table 6 summarizes the estimation accuracies of the evaluation results using four methods. In the 54 test projects, the proposed

Table 5
Estimated cost using the component ratios method for No. 1 test project

Principal item	*Averaged cost ($)*	*Ratio*	*Estimated cost ($)*
Bulk gas	143,459	1.335	1,637,732
UPW	258,308		
Exhaust	554,114		
Foundation	271,050		
Subtotal cost	1,226,931		

model achieved average estimation accuracy of 83.82%. This represented an improvement of about 22.74% (=83.82%−61.08%) compared with the average cost method, around 23.08% (=83.82%−60.74%) compared with the component ratios method, and approximately 21.95% (=83.82%−61.87%) compared with the modified FALCON-COST method. Moreover, the proposed model has smaller standard deviation of estimation accuracy than alternative models (just 13.46%), meaning it can provide more consistent estimations than the other three methods.

These 54 case studies yield two additional observations. First, these case studies confirm the poor performance of the conventional average cost method. Specifically, the average accuracies of the projects related to IMP and WET process modules using the average cost method are only 4.52% (118.44% for standard deviation) and 37.47% (78.07% for standard deviation), respectively. Analyzing the historical project data reveals that these high inaccuracies likely resulted from the large cost deviation among historical projects even within the same type of process model. Consequently, clustering data (for example, high and low cost clusters) for analysis is crucial for capturing this high cost deviation. In the proposed model, this clustering capability can be found in Layer 2 of FALCON and in the example shown in Figure 3.

Second, compared with the other three methods, only the proposed model reflected certain features of a new project using the three-point cost estimation method, potentially significantly contributing to the improved estimation accuracy. The modified FALCON-COST using historical average costs rather than expert opinions (obtained using three-point estimations) as modeling inputs achieved average accuracy of just 61.87% (with standard deviation of 35.45%). Restated, the model outputs are sensitive to inputs, and thus the proposed model is particularly recommended for experienced cost estimators familiar with semiconductor hookup construction.

7.3 Discussions on the FALCON-COST improvements

As indicated earlier, FALCON-COST applies the fmGA mutation and cut-splice operators to optimize the conventional FALCON fuzzy membership functions and FL rules. To verify this improvement, this study uses the CMP module (including 15 historical projects) as an example.

Table 7 compares the FALCON parameters (i.e., m_{ij} and σ_{ij} of input membership functions; R_{ij} of FL rules;

Table 6
Comparisons of results using various estimation methods for 54 test projects

		Estimation accuracy (%)							
		Average cost method		*Component ratios method*		*Modified FALCON-COST model*		*Proposed model*	
Module	*Amount of test projects*	*Average*	*Std Dev*	*Average*	*Std Dev*	*Average*	*Std Dev*	*Average*	*Std Dev*
CMP	4	86.13	9.56	86.11	9.82	83.94	14.16	93.87	5.57
DIFF	8	65.57	7.39	65.92	7.39	72.32	7.11	79.85	10.58
ETCH	8	66.57	15.54	64.96	20.63	43.86	50.55	78.33	14.43
INT	6	74.51	15.96	71.38	15.26	72.26	22.90	92.69	10.16
LITHO	6	79.70	14.10	78.85	13.82	78.00	18.03	83.88	13.98
IMP	6	4.52	118.44	7.41	107.25	42.08	71.60	76.98	11.32
TF	6	74.12	21.60	74.67	29.68	64.40	21.02	76.81	16.88
WET	10	37.47	78.07	36.62	82.58	38.12	7.22	88.14	13.82
Average accuracy (%)		61.08	55.54	60.74	54.67	61.87	35.45	83.82	13.46

Note: "Std Dev" is an abbreviation of "standard deviation."

Table 7
Comparisons of FALCON parameters for the CMP module

(a) Input membership functions

	m_{11}	σ_{11}	m_{12}	σ_{12}	m_{21}	σ_{21}	m_{22}	σ_{22}	m_{31}	σ_{31}	m_{32}	σ_{32}	m_{41}	σ_{41}	m_{42}	σ_{42}	m_{43}	σ_{43}
FALCON-COST	0.1	0.2	1.0	0.1	0.2	0.1	1.0	0.2	0.4	0.3	1.0	0.2	0.2	1.2	0.9	0.1	1.0	0.1
Traditional FALCON	0.1	1.0	0.2	1.0	0.4	1.0	0.2	0.9	1.0	0.0	0.1	0.1	0.2	0.3	0.2	0.0	0.1	0.0

(b) Fuzzy logic rules

	R_{01}	R_{02}	R_{03}	R_{04}	R_{05}	R_{06}	R_{07}	R_{08}	R_{09}	R_{10}	R_{11}	R_{12}	R_{13}	R_{14}	R_{15}	R_{16}	R_{17}	R_{18}	R_{19}	R_{20}	R_{21}	R_{22}	R_{23}	R_{24}
FALCON-COST	2	4	5	2	5	2	1	4	4	1	4	4	2	5	1	2	4	4	2	5	5	2	5	5
Traditional FALCON	2	4	5	2	4	4	2	4	4	2	4	4	2	5	5	2	4	4	2	5	5	2	5	5

(c) Output membership functions

	m_{11}	σ_{11}	m_{12}	σ_{12}	m_{13}	σ_{13}	m_{14}	σ_{14}	m_{15}	σ_{15}
FALCON-COST	0.0	0.0	0.2	0.0	0.7	0.1	0.9	0.0	1.0	0.1
Traditional FALCON	0.0	0.2	0.6	0.9	1.0	0.0	0.0	0.1	0.0	0.1

m_{ij} and σ_{ij} of output membership functions) trained by the traditional FALCON and FALCON-COST for the CMP module. From this table some of the parameter values produced by both models vary greatly. For instance, in the input membership functions (see Table 7a), great differences in the parameter values are found for σ_{11}, m_{12}, σ_{12}, σ_{21}, m_{22}, σ_{22}, m_{31}, σ_{31}, m_{32}, σ_{41}, m_{42}, and m_{43}. In other words, the proposed FALCON-COST model is able to revise the parameters for generating a new FALCON structure to achieve an improved solution. The traditional FALCON model does not provide this capability.

As presented in Section 6.7, the estimation accuracy of the trained FALCON-COST is around 97.85% (error rate = 2.15%) after training using 15 CMP historical projects. When the same historical projects were applied to the conventional FALCON, the error rate increased to about 5.25% (=1–0.9475). This comparison verifies the FALCON-COST improvement.

8 CONCLUSION

This investigation has contributed to several aspects. First, the proposed model has enhanced the estimation accuracy of the cost of hookup construction projects. The 54 test projects achieved increases in estimation accuracy of approximately 22.74%, 23.08%, and 21.95% over the conventional average cost method, the component ratios method, and a modified FALCON-COST method, respectively. In addition, the proposed model is currently being implemented by the case-study company to facilitate cost estimations for new projects specifically related to INT and WET process modules tools.

Second, integrating the four techniques (that is, the component ratios method, FALCON, fmGA, and the three-point estimation method) into the proposed model is innovative, and can systematically deal with real world cost estimation problems.

Third, the proposed model improves the current FALCON by applying the fmGA mutation and cut-splice operators. That is, in the primordial phase of fmGA, a new algorithm is developed to conduct BB selection and random gene deletion, so that fmGA operations can be implemented in FALCON.

Future research may include the following directions. First, computerizing the proposed model will help expedite the evaluation. Second, collecting additional historical projects should support the model training process for enhancing the estimation accuracy. Third, applying the proposed model to new projects for conducting a before-the-fact analysis can further verify the practicality of the model. Fourth, other attribute ranking algorithms, such as Analytical Hierarchy Process (Saaty, 1978), or analysis of the observation frequencies of cost items may help identify major cost items. Fifth, FALCON may be substituted by another type of NFS, such as adaptive network-based fuzzy inference system (ANFIS) (Jang, 1993). Other optimization methods, such as Tabu search (Fan and Machemehl, 2008), simulated annealing (Paya et al., 2008; Zeferino et al., 2009; Oliveira and Petraglia, 2011), and ant colony (Vitins and Axhausen, 2009; Putha et al., 2012) may also be substituted for fmGA in the proposed model to find enhanced solutions.

Sixth, although the proposed model is devised specifically for semiconductor hookup construction projects, it can be modified to apply to other similar decision-making problems with similar cost-estimating characteristics, including conceptual cost estimation problems in building projects (Yu, 2006; Cheng et al., 2009), bid-price determination under limited bid preparation time (Wang et al., 2007), and project success prediction problems (Cheng et al., 2010). For instance, during the conceptual phase of a building project, project management often needs to calculate the project cost given a conceptual design situation involving unavailable and uncertain cost data. At such times, by treating the major cost categories (or cost divisions) in the building project as the main cost items discussed in this study, and using the relevant historical building project data for training, the FALCON-COST model can easily be refined and applied.

ACKNOWLEDGMENTS

The authors would like to thank the editor and the reviewers for their careful evaluation and thoughtful suggestions. The authors also thank the National Science Council of Taiwan (Contract No. NSC97–2221-E-009–134) and the Ministry of Education of Taiwan via the Aim for the Top University (MOU-ATU) program for financially supporting this research.

REFERENCES

Adeli, H. & Cheng, N.-T. (1993), Integrated genetic algorithm for optimization of space structures, *Journal of Aerospace Engineering*, **6**(4), 315–28.

Adeli, H. & Cheng, N.-T. (1994a), Augmented Lagrangian genetic algorithm for structural optimization, *Journal of Aerospace Engineering*, **7**(1), 104–18.

Adeli, H. & Cheng, N.-T. (1994b), Concurrent genetic algorithms for optimization of large structures, *Journal of Aerospace Engineering*, **7**(3), 276–96.

Adeli, H. & Hung, S. L. (1995), *Machine Learning—Neural Networks, Genetic Algorithms, and Fuzzy Systems*, John Wiley and Sons, New York.

Adeli, H. & Karim, A. (1997), Scheduling/cost optimization and neural dynamics model for construction, *Journal of Construction Management and Engineering*, **123**(4), 450–8.

Adeli, H. & Karim, A. (2001), *Construction Scheduling, Cost Optimization, and Management—A New Model Based on Neurocomputing and Object Technologies*, Spon Press, London.

Adeli, H. & Kumar, S. (1995a), Distributed genetic algorithms for structural optimization, *Journal of Aerospace Engineering*, **8**(3), 156–63.

Adeli, H. & Kumar, S. (1995b), Concurrent structural optimization on a massively parallel supercomputer, *Journal of Structural Engineering*, **121**(11), 1588–97.

Adeli, H. & Wu, M. (1998), Regularization neural network for construction cost estimation, *Journal of Construction Engineering and Management*, **124**(1), 18–24.

Adeli, H. & Yu, G. (1995), An integrated computing environment for solution of complex engineering problems using the object-oriented programming paradigm and a blackboard architecture, *Computers and Structures*, **54**(2), 255–65.

Ahmadlou, M. & Adeli, H. (2010), Enhanced probabilistic neural network with local decision circles: a robust classifier, *Integrated Computer-Aided Engineering*, **17**(3), 197–210.

Al-Bazi, A. & Dawood, N. (2010), Developing crew allocation system for precast industry using genetic algorithms, *Computer-Aided Civil and Infrastructure Engineering*, **25**(8), 581–95.

Baraldi, P., Canesi, R., Zio, E., Seraoui, R. & Chevalier, R. (2011), Genetic algorithm-based wrapper approach for grouping condition monitoring signals of nuclear power plant components, *Integrated Computer-Aided Engineering*, **18**(3), 221–34.

Barrie, D. S. & Paulson, B. C. (1992), *Professional Construction Management*, 3rd edn, McGraw-Hill, New York.

Beers, A. J., Williams, D. & Tarr, A. (2008), Tool hookup: a paradigm shift to modularization, *Solid State Technology*, September, 36–7.

Boussabaine, A. H. & Elhag, T. M. S. (1997), *A neuro-fuzzy model for predicting cost and duration of construction projects*, RICS Research (9p), The Royal Institution of Chartered Surveyors, London.

Carro-Calvo, L., Salcedo-Sanz, S., Ortiz-Garc, G. & Portilla-Figueras, A. (2010), An incremental-encoding evolutionary algorithm for color reduction in images, *Integrated Computer-Aided Engineering*, **17**(3), 261–9.

Chasey, A. D. & Merchant, S. (2000), Issues for construction of 300-mm fab, *Journal of Construction Engineering and Management*, **126**(6), 451–7.

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), 164–72.

Cheng, M. Y., Wu, Y. W. & Wu, C. F. (2010), Project success prediction using an evolutionary support vector machine inference model, *Automation in Construction*, **19**(3), 619–29.

Cheng, T. M. & Yan, R. Z. (2009), Integrating messy genetic algorithms and simulation to optimize resource utilization, *Computer-Aided Civil and Infrastructure Engineering*, **24**(6), 401–15.

Chou, J. S. (2011), Cost simulation in an item-based project involving construction engineering and management, *International Journal of Project Management*, **29**, 706–17.

Creese, R. C. & Li, L. (1995), Cost estimation of timber bridges using neural networks, *Cost Engineering*, **37**(5), 17–22.

Diekmann, J. E. & Featherman, W. D. (1998), Assessing cost uncertainty: lessons from environmental restoration projects, *Journal of Construction Engineering and Management*, **124**(6), 445–51.

Fan, W. & Machemehl, R. B. (2008), Tabu search strategies for the public transportation network optimizations with variable transit demand, *Computer-Aided Civil and Infrastructure Engineering*, **23**(7), 502–20.

Feng, C. W. & Wu, H. T. (2006), Integrating fmGA and CYCLONE to optimize the schedule of dispatching RMC trucks, *Automation in Construction*, **15**(2), 186–99.

Freitag, S., Graf, W. & Kaliske, M. (2011), Recurrent neural networks for fuzzy data, *Integrated Computer-Aided Engineering*, **18**(3), 265–80.

Goldberg, D. E. (2002), *The Design of Innovation: Lessons from and for Competent Genetic Algorithms*, 1st edn, Kluwer Academic Publishers, Boston.

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.

Goldberg, D. E., Deb, K. & Korb, B. (1991), Don't worry, be messy, in *Proceedings of the Fourth International Conference on Genetic Algorithms*, La Jolla, CA, 24–30.

Goldberg, D. E., Korb, B. & Deb, K. (1989), Messy genetic algorithms: motivation, analysis, and first results, *Complex System*, **3**, 493–530.

Gonzalez-Olvera, M. A., Gallardo-Hernandez, A. G., Tang, Y., Revilla-Monsalve, M. C. & Islas-Andrade, S. (2010), A discrete-time recurrent neurofuzzy network for black-box modeling of insulin dynamics in diabetic type-1 patients, *International Journal of Neural Systems*, **20**(2), 149–58.

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), 2873–7.

Hegazy, T. & Ayed, A. (1998), Neural network model for parametric cost estimation of highway projects, *Journal of Construction Engineering and Management*, **124**(3), 210–8.

Hendrickson, C. & Au, T. (2003), Project Management for Construction—Fundamental Concepts for Owners, Engineers, Architects and Builders, Available at: http://www.ce.cmu.edu/pmbook/, accessed March 5, 2012.

Holland, J. (1975), *Adaptation in Natural and Artificial System*, University of Michigan Press, Ann Arbor, MI.

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), 2797–808.

Hong, X. (2001), *Introduction to Semiconductor Manufacturing Technology*, Prentice Hall, NJ.

Hung, S. L. & Adeli, H. (1993), Parallel backpropagation learning algorithms on Cray Y-MP8/864 supercomputer, *Neurocomputing*, **5**(6), 287–302.

Hung, S. L. & Adeli, H. (1994), Object-oriented backpropagation and its application to structural design, *Neurocomputing*, **6**(1), 45–55.

Iglesias, J. A., Angelov, P., Ledezma, A. & Sanchis, A. (2010), Human activity recognition based on evolving fuzzy systems, *International Journal of Neural Systems*, **20**(5), 355–64.

Jang, J. S. (1993), ANFIS: adaptive-network-based fuzzy inference system, *IEEE Transactions on Systems, Man, and Cybernetics*, **23**(3), 665–85.

Jiang, X. & Adeli, H. (2003), Fuzzy clustering approach for accurate embedding dimension identification in chaotic time series, *Integrated Computer-Aided Engineering*, **10**(3), 287–302.

Jiang, X. & Adeli, H. (2004), Object-oriented model for freeway work zone capacity and queue delay estimation, *Computer-Aided Civil and Infrastructure Engineering*, **19**(2), 144–56.

Kang, M. W., Schonfeld, P. & Yang, N. (2009), Prescreening and repairing in a genetic algorithm for highway alignment optimization, *Computer-Aided Civil and Infrastructure Engineering*, **24**(2), 109–19.

Karim, A. & Adeli, H. (1999a), Object-oriented information model for construction project management, *Journal of Construction Engineering and Management*, *ASCE*, **125**(5), 361–7.

Karim, A. & Adeli, H. (1999b), CONSCOM: an OO construction scheduling and change management system, *Journal of Construction Engineering and Management*, *ASCE*, **125**(5), 368–76.

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), 208–11.

Kim, H. & Adeli, H. (2001), Discrete cost optimization of composite floors using a floating point genetic algorithm, *Engineering Optimization*, **33**(4), 485–501.

Koch, R. (1997), *The 80/20 Principle—The Secret of Achieving More Less*, Nicholas Brealey Publishing Ltd, London.

Koprinkova-Hriatova, P. (2010), Backpropagation through time training of a neuro-fuzzy controller, *International Journal of Neural Systems*, **20**(5), 421–8.

Lee, J. & Pinheiro dos Santos, W. (2011), An adaptive fuzzy-based system to simulate, quantify, and compensate color blindness, *Integrated Computer-Aided Engineering*, **18**(1), 29–40.

Lee, Y. & Wei, C. H. (2010), A computerized feature selection using genetic algorithms to forecast freeway accident duration times, *Computer-Aided Civil and Infrastructure Engineering*, **25**(2), 132–48.

Li, X., Er, M. J., Lim, B. S., Zhou, J. H., Gan, O. P. & Rutkowski, L. (2010), Fuzzy regression modeling for tool performance prediction and degradation detection, *International Journal of Neural Systems*, **20**(5), 405–19.

Lin, C. T. & Lee, C. S. G. (1991), Neural-network-based fuzzy logic control and decision system, *IEEE Transactions on Computers*, **40**(12), 1320–36.

Lin, C. T. & Lee, C. S. G. (1996), *Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems*, Prentice Hall, New Jersey.

Liu, C. H., Lin, S. J. & Lewis, C. (2010), Life cycle assessment of DRAM in Taiwan's semiconductor industry, *Journal of Cleaner Production*, **18**, 419–25.

Lowe, D. J., Emsley, M. W. & Harding, A. (2006), Predicting construction cost using multiple regression techniques, *Journal of Construction Engineering and Management*, **132**(7), 750–8.

Ma, Z. M., Zhang, F., Yan, L. & Cheng, J. (2011), Extracting knowledge from fuzzy relational databases with description logic, *Integrated Computer-Aided Engineering*, **18**(2), 181–200.

Marano, G. C., Quaranta, G. & Monti, G. (2011), Modified genetic algorithm for the dynamic identification of structural systems using incomplete measurements, *Computer-Aided Civil and Infrastructure Engineering*, **26**, 92–110.

Martínez-Ballesteros, M., Troncoso, A., Martínez-Álvarez, F. & Riquelm, J. C. (2010), Mining quantitative association rules based on evolutionary computation and its application to atmospheric pollution, *Integrated Computer-Aided Engineering*, **17**(3), 227–42.

Moder, J. J., Philips, C. R. & Davis, E. W. (1983), *Project Management with CPM, PERT and Precedence Diagramming*, 3rd edn, Van Nostrand Reinhold, New York.

Oberlender, G. D. (2000), *Project Management for Engineering and Construction*, 2nd edn, McGraw-Hill, New York.

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), 173–82.

Oliveira, H. A. & Petraglia, A. (2011), Global optimization using dimensional jumping and fuzzy adaptive simulated annealing, *Applied Soft Computing*, **11**, 4175–82.

Patrinos, P., Alexandridis, A., Ninos, K. & Sarimveis, H. (2010), Variable selection in nonlinear modeling based on RBF networks and evolutionary computation, *International Journal of Neural Systems*, **20**(5), 365–79.

Paya, I., Yepes, V., González-Vidosa, F. & Hospitaler, A. (2008), Multiobjective optimization of concrete frames by simulated annealing, *Computer-Aided Civil and Infrastructure Engineering*, **23**(8), 596–610.

PEH, (2008), *Parametric Estimating Handbook*, 4th edn. International Society of Parametric Analysts. Available at: http://www.galorath.com/images/uploads/ISPA_PEH_4th_ed_Final. pdf, accessed February 17, 2012.

Peurifoy, R. L. & Oberlender, G. D. (2002), *Estimating Construction Costs*, 5th edn, McGraw-Hill, New York.

Putha, R., Quadrifoglio, L. & Zechman, E. (2012), Comparing ant colony optimization and genetic algorithm approaches for solving traffic signal coordination under oversaturation conditions, *Computer-Aided Civil and Infrastructure Engineering*, **27**, 14–28.

Reuter, U. (2011), A fuzzy approach for modeling nonstochastic heterogeneous data in engineering based on cluster analysis, *Integrated Computer-Aided Engineering*, **18**(3), 281–9.

Rokni, S. & Fayek, A. R. (2010), A multi-criteria optimization framework for industrial shop scheduling using fuzzy set theory, *Integrated Computer-Aided Engineering*, **17**(3), 175–96.

Ross, T., Reda Taha, M., Kim, J. J. & Gilfeather, F. (2011), Logical models for the propagation of disparate information and uncertainty across effectivity trees, *Integrated Computer-Aided Engineering*, **18**(3), 251–64.

Saaty, T. L. (1978), Exploring the interface between the hierarchies, multiple objectives and the fuzzy sets, *Fuzzy Sets and Systems*, **1**, 57–68.

Sadeghi, N., Fayek, A. R. & Pedrycz, W. (2010), Fuzzy Monte Carlo simulation and risk assessment in construction, *Computer-Aided Civil and Infrastructure Engineering*, **25**(4), 238–52.

Sarma, K. & Adeli, H. (1998), Cost optimization of concrete structures, *Journal of Structural Engineering*, **124**(5), 570–8.

Sarma, K. C. & Adeli, H. (2000), Fuzzy discrete multicriteria cost optimization of steel structures, *Journal of Structural Engineering*, **126**(11), 1339–47.

Sarma, K. C. & Adeli, H. (2002), Life-cycle cost optimization of steel structures, *International Journal for Numerical Methods in Engineering*, **55**(12), 1451–62.

Scherer, R. (2010), Designing boosting ensemble of relational fuzzy systems, *International Journal of Neural Systems*, **20**(5), 381–8.

Sedano, J., Curiel, L., Corchado, E., de la Cal, E. & Villar, J. R. (2010), A soft computing method for detecting lifetime building thermal insulation failures, *Integrated Computer-Aided Engineering*, **17**(2), 103–15.

Senouci, A. B. & Adeli, H. (2001), Resource scheduling using neural dynamics model of Adeli and Park, *Journal of Construction Engineering and Management, ASCE*, **127**(1), 28–34.

Sonmez, R. (2008), Parametric range estimating of building costs using regression models and bootstrap, *Journal of Construction Engineering and Management*, **134**(12), 1011–6.

Theodoridis, D. C., Boutalis, Y. S. & Christodoulou, M. A. (2010), Indirect adaptive control of unknown multi variable nonlinear systems with parametric and dynamic uncertainties using a new neuro-fuzzy system description, *International Journal of Neural Systems*, **20**(2), 129–48.

Vitins, B. J. & Axhausen, K. W. (2009), Optimization of large transport networks using the ant colony heuristic, *Computer-Aided Civil and Infrastructure Engineering*, **24**(1), 1–14.

Wang, C. H., Lee, W. H. & Huang, Y. C. (2008), Re: influence of rate stability on project cost simulation, *Computer-Aided Civil and Infrastructure Engineering*, **23**, 45–58.

Wang, W. C. (2002), Simulation-facilitated model for assessing cost correlations, *Computer-Aided Civil and Infrastructure Engineering*, **17**(5), 368–80.

Wang, W. C., Dzeng, R. J. & Lu, Y. H. (2007), Integration of simulation-based cost model and multi-criteria evaluation model for bid price decisions, *Computer-Aided Civil and Infrastructure Engineering*, **22**, 223–35.

Wang, N., Joo, M., Meng, X. Y. & Li, X. (2010), An online self-organizing scheme for parsimonious and accurate fuzzy neural networks, *International Journal of Neural Systems*, **20**(5), 389–403.

Wang, W. C., Wang, S. S., Wen, C. P., Hsiao, F. Y. & Yu, W. D. (2012), Application of FALCON and fmGA to cost estimation for semiconductor hookup construction in *Proceedings of the 14th International Conference on Computing in Civil and Building Engineering (14th ICCCBE)*, Moscow.

Wen, C. P. (2010), Investigation of cost estimates on semiconductor hookup construction, (in Chinese). Master thesis, Program of Engineering Technology and Management, National Chiao Tung University, Taiwan.

Wu, C. J. (2007), Investigation on hookup engineering management of semiconductor equipments—case study of a 12-inch wafer fab (in Chinese). Master thesis, Engineering Management Graduate Program, National Cheng Kung University, Taiwan.

Yu, W. D. (2006), PIREM: a new model for conceptual cost estimation, *Construction Management and Economics*, **24**(3), 259–70.

Yu, W. D. (2007), Hybrid soft computing approach for knowledge discovery in construction, *Journal of Computing in Civil Engineering*, **21**(5), 343–52.

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.

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), 539–52.

Zeferino, J. A., Antunes, A. P. & Cunha, M. C. (2009), An efficient simulated annealing algorithm for regional wastewater systems planning, *Computer-Aided Civil and Infrastructure Engineering*, **24**(5), 359–70.

Zhang, F., Ma, Z. M. & Yan, L. (2011), Construction of ontology from object-oriented database model, *Integrated Computer-Aided Engineering*, **18**(4), 327–47.