emerald**insight**

Built Environment Project and Asset Management
Improving accuracy of conceptual cost estimation using MRA and ANFIS in Indonesian building projects
Dwifitra Jumas, Faizul Azli Mohd-Rahim, Nurshuhada Zainon, Wayudi P. Utama,

Article information:

Users who downloaded this article also downloaded:

For Authors
If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com
Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

Improving accuracy of conceptual cost estimation using MRA and ANFIS in Indonesian building projects

Dwifitra Jumas
*Quantity Surveying, Faculty of Built Environment, University of Malaya,
Kuala Lumpur, Malaysia and
Quantity Surveying, Universitas Bung Hatta, Padang, Indonesia*

Faizul Azli Mohd-Rahim and Nurshuhada Zainon
*Quantity Surveying, Faculty of Built Environment, University of Malaya,
Kuala Lumpur, Malaysia, and*

Wayudi P. Utama
Quantity Surveying, Universitas Bung Hatta, Padang, Indonesia

Abstract

Purpose – The purpose of this paper is to develop a conceptual cost estimation (CCE) model for building project by using a pragmatic approach, which is a mix of tools drawn from multiple regression analysis (MRA) and adaptive neuro-fuzzy inference system (ANFIS), to improve the accuracy of cost estimation at an early stage.

Design/methodology/approach – This paper presents a set of MRA and integrating MRA with ANFIS or MRANFIS. A simultaneous regression analysis was developed to determine the main cost factors from 12 variables as input variables in the ANFIS model. Cost data from 78 projects of state building in West Sumatra, Indonesia were used to indicate the advantages of the proposed model.

Findings – The result shows that the proposed model, MRANFIS, has successfully improved the mean absolute percent error (MAPE) by 2.8 percent from MRA of 10.7–7.9 percent for closeness of fit to the model data and by 3.1 percent from MRA of 9.8–6.7 percent for prediction performance to the new data.

Research limitations/implications – Because the significant variables are different for each building type, the model may be not appropriate for other buildings depending on the characteristics of building. The models can be used and analyzed based on the own historical project data for each case so that the model can be applied.

Originality/value – The study thus provides better accuracy of CCE at an early stage for state building projects in West Sumatra, Indonesia by using the integrated model of MRA and ANFIS.

Keywords Indonesia, Cost model, ANFIS, Multiple regression, Building project, Conceptual cost estimation

Paper type Research paper

Introduction

Conceptual cost estimation (CCE) is the most important preliminary process in any construction project. Due to insufficient essential information on the project at its early stages, it is very important to quickly, economically and accurately find relevant additional information (Koo *et al.*, 2011). In fact, engineers require several years to develop the necessary expertise to conduct estimating models and predict the cost of initial phase by the use of limited information (Elfaki *et al.*, 2014).

A review of studies on the accuracy of cost estimation for construction projects indicated that the estimation of accuracy level for construction projects was heavily dependent upon the availability of quality historical cost data and the level of professional expertise among other things (Moon *et al.*, 2007; Riquelme and Serpell, 2013). In theory, estimating accuracy is an indicative degree which depicts that the final price outcome of a project may vary from

the single point value used, which is being used as the estimated cost of the project. Accuracy implies closeness to the actual value, whatever it may be. It means that there are lack of errors. It also comprises two aspects in term of "bias" and "consistency" (Ashworth, 1995; Skitmore, 1991). Bias is concerned with "the average differences between actual value and forecasts," while measures of consistency are concerned with "the degree of variation around the average." Thus, accuracy is considered as an overall combination of both bias and consistency.

In addition, factors contributing to the estimation bias could be influenced by more causes and inter-correlation between many variables with different parameters (Jumas *et al.*, 2016; Azman *et al.*, 2012; Elfaki *et al.*, 2014; Shane *et al.*, 2009). Each parameter must be properly addressed to maintain an acceptable level of accuracy during the process. Therefore, the main objective of this research is to identify the main parameter as cost variables for state building, the case of West Sumatera, Indonesia in order to improve the accuracy of CCE using the integration of MRA and ANFIS.

Techniques of cost model
Many researchers have studied various methodologies for predicting cost in the initial phase by the use of limited information. Various CCE techniques have been also introduced to calculate the CCE including neural network (NN) (Adeli and Wu, 1998; Creese and Li, 1995; Hegazy and Ayed, 1998; Kim *et al.*, 2005), regression analysis (Kwak and Watson, 2005; Lowe *et al.*, 2006) and case-based reasoning (CBR) (Chou, 2009; Koo *et al.*, 2011; Marzouk and Ahmed, 2011). Some researchers have also integrated the two above methods or techniques altogether such as regression and NN (Sonmez, 2004), MRA and NN (Gunduz *et al.*, 2011), MRA and CBR (Jin *et al.*, 2012) and regression and ANFIS (Latief *et al.*, 2013).

Kim *et al.* (2004) said that regression or multiple regression analysis (MRA) is a simple and powerful statistical method that can be used as an analytical and predictive technique to examine the overall cost estimate reliability. However, Lowe *et al.* (2006) concurred that MRA led to the result of a statistical analysis, but its results were too linear to be used in a standardized model. On the other hand, it is not appropriate to describe non-linear relationships, which are multidimensional, consisting of a multiple input and output problem (Tam and Fang, 1999).

A NN as a computer system that simulates the learning process of the human brain offering offers an alternative approach for cost estimation. Ahn *et al.* (2014) said that NN can be more beneficial when it involves intuitive judgment or when its patterns of data become too irregular to identify by applying traditional techniques. Moreover, the user does not need to exert more effort to decide on the class of relations or the probability distribution of the variables (Sonmez, 2004). However, another study found the NN is a black box technique and its process is time-consuming to determine the network factors that fit the best application (Adeli and Wu, 1998; Creese and Li, 1995; Hegazy and Ayed, 1998).

Another method or technique in CCE is CBR. In CBR systems, expertise is embodied in a library of past cases which contains a description of the problem, plus a solution and/or the outcome (Marzouk and Ahmed, 2011) or expert prototype system that compares historical data at the work item-level across the case library (Chou, 2009). According to Ahn *et al.* (2014) and Marzouk and Ahmed (2011), a general CBR is able to modify, or to adapt, a retrieved solution when applied in a different problem-solving context. However, Watson (1997) in Ahn *et al.* (2014) highlighted the usefulness of CBR with structured symbolic data is more complicated than purely numeric data.

Regardless of which technique was used, the question of how many and which variable to be used for cost estimation purposed are very critical, and research on these topics is still done. Moreover, the accuracy and diversity level of cases from the historical project cost can affect the results to a certain level. Therefore, determining the significant variables is crucial in parametric cost estimation.

Methodology

This study adopted a pragmatic approach integrating MRA and adaptive neuro-fuzzy inference system (ANFIS) to improve the accuracy of cost estimation. In regression analysis, the independent and dependent variables were initially defined. Through a comprehensive literature review, 12 cost factors independent variables were identified as shown in Table I. The selection of these variables, according to Phaobunjong (2002), has fulfilled the characteristics of CCE which are well-established with a clear definition to minimize ambiguity and inconsistency, readily quantifiable or measurable value and reasonable accuracy in early project stages. On the other hand, the total cost per gross floor area (GFA) was set as independent variable instead of the project costs. The first mentioned can minimize a huge range of project cost with minimum to maximum cost (Lowe *et al.*, 2006; Cheung and Skitmore, 2006).

Before the analysis is conducted, relevant data were extracted from the historical project cost. In this research, 78 new state building projects obtained from contract documents of the successful bidders were used. All the projects are located in West Sumatra, Indonesia spanning from the period of 2005–2015. Furthermore, the data need to be normalized (Ji *et al.*, 2010; Phaobunjong, 2002) by means of adjusting the regional cost index to quantify the variability of project cost due to inflation (Sonmez, 2004). Since the index has not been available in Indonesia, the consumer price index in the following equation was used as a proxy to measure the cost indices:

$$C_j = C_i \frac{\text{cost index}_j}{\text{cost index}_i}, \tag{1}$$

where C_j = cost at location j; C_i = cost at location i.

To convert cost from one time (year) to another year, the following formula was used:

$$C_j = C_0 + ((I_{0+1} + I_{0+1} + \cdots + I_{0+n}) \times C_0), \tag{2}$$

where C_j = cost at desired year; C_O = cost at reference year; I_{0+1} = inflation at reference year +1 year; I_{0+2} = inflation at reference year+2 year; I_{0+n} = inflation at reference year +n year.

Next, MRA was employed to select the most significant variables affecting the CCE in a building project. Generally, MRA results are in a regression, as shown in the

Category	Variables	Definitions	Range
Parameter	$X1$	Gross floor area (GFA)	284–13,961 m^2
	$X2$	Building height	7.10–24.35 m
	$X3$	Number of stories	2–5
	$X4$	Average height	3.00–5.53 m
	$X5$	External wall area	383–13,540 m^2
	$X6$	Compactness (external wall area/gross external floor area)	0.15–3.42
	$X7$	Proportion of openings	0.061–0.494
	$X8$	Type of use	1 = education; 2 = office; 3 = hospital
	$X9$	Type of roof	1 = ordinary; 2 = Bagonjong
	$X10$	Type of foundation	1 = pad foundation; 2 = cyclops; 3 = Continuous footing; 4 = pile foundation; 5 = KSLL
	$X11$	The ratio of the typical floor area to the GFA	0.4–1
	$X12$	Duration	3–12 months
	Y	Cost per GFA	IDR3,362,550–8,706,432

Table I.
List of variables

following form:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n. \tag{3}$$

In addition, the intent of the model is not only for the identification and explanation of parameters affecting the construction cost, but it is also to be used in estimating the construction cost.

Finally, the integration of MRA and ANFIS was employed to develop a cost forecasting model. In this model, the variables derived from MRA were set as input variables, while the output was cost per GFA. The pair of input–output variables was then processed by using ANFIS.

Furthermore, the basic concept of ANFIS was functioned to create stipulated input–output pairs through assembling a set of fuzzy if-then rules with suitable membership functions (MFs) through implanting the fuzzy inference rule into the structure of adaptive networks (Jang, 1993). The structure of ANFIS used in this system is indicated in Figure 1.

An example of the rule base constitutes fuzzy if-then rules. The example of one rule might be "if the probability of cost per GFA occurrence is high and GFA is medium" where high and medium are the forms of fuzzy linguistic variables. FIS contains two rule bases following a linear function as described by Takagi and Sugeno (1983):

$$\text{Rule 1 : If } X_1 \text{ is } A_1 \text{ and } X_2 \text{ is } A_1 \text{ then } Y_1 = p_1 x_1 + q_1 x_2 + r_1,$$

$$\text{Rule 2 : If } X_1 \text{ is } A_2 \text{ and } X_2 \text{ is } A_2 \text{ then } Y_2 = p_2 x_1 + q_2 x_2 + r_2,$$

where "X_1 and X_2" and "Y_1 and Y_2" are numerical inputs and outputs, respectively; A and B are numerical variables; and p, q and r are the parameters determining the relation between input and output.

Layer 1: this layer shows the number of numerical inputs belonging to the different fuzzy set. Every node i in this layer is represented by the square node with the output function of the following equation:

$$\begin{cases} O_i = \mu_{A_i}(x_1) \\ O_i = \mu_{B_i}(x_2) \end{cases}, \tag{4}$$

where $\mu_{A_i}(x_1)$ and $\mu_{B_i}(x_2)$ are the MFs for fuzzy sets of A and B.

Layer 1   Layer 2   Layer 3   Layer 4   Layer 5
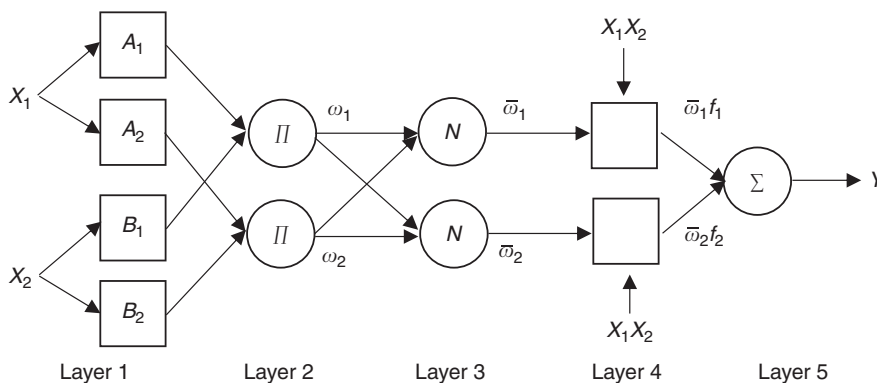
Source: Jang (1993)

Layer 2: in this layer, all incoming signals are multiplied to obtain an output, ω by which operator AND or OR are used, known as firing strength. The output is calculated using the following equation:

$$\omega_i = \mu_{A_i}(x_1) \times \mu_{B_i}(x_2). \tag{5}$$

Layer 3: every node N in this layer calculates the average ratio of previous outputs to produce a new output $\overline{\omega}$. This is obtained by the following equation:

$$\overline{\omega}_i = \frac{\omega_i}{\sum_i \omega_i}. \tag{6}$$

Layer 4: square node in this layer produces an output $\overline{\omega}_i f_i$ based on the following equation:

$$Y_i = \overline{\omega}_i f_i = \overline{\omega}_i (p_i x_1 + q_i x_2 + r_i). \tag{7}$$

Layer 5: this is an output layer in which the node calculates all outputs from Layer 4 using the following equation:

$$Y = \sum_i \overline{\omega}_i f_i = \frac{\sum_i \omega_i f_i}{\sum_i \omega_i}. \tag{8}$$

ANFIS integrates the least squares estimate (LSE) and the gradient descent method with a hybrid learning rule algorithm. This procedure is composed of a forward step in which the input signal passes forward until Layer 4, where the output parameters are then adjusted using the LSE of the error between the estimated output and the actual output. Then, on the backward step, the error rates propagate back through the system, and MFs in Layer 1 are updated by the gradient descent method (Jang, 1993). The process of these forward and backward propagations is called "epoch." The hybrid learning algorithm trains the MF parameters to mimic the training data samples.

The MRA and MRANFIS model were then evaluated to measure its closeness of fit and prediction performance. In this research, two types of error for evaluation: mean squared error (MSE) and mean absolute percent error (MAPE) were utilized. MSE and MAPE were calculated as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\text{actual}_i - \text{predicted}_i)^2, \tag{9}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|\text{actual}_i - \text{predicted}_i|}{\text{predicted}_i} \times 100\%, \tag{10}$$

where i is the project number.

Model application
MRA
The all-possible multiple regressions procedure that fits all combinations of variables was used over other variables selection procedures using "Enter method" (simultaneous regression). All variables in Table I were observed for any impact on the cost database. The "Enter method" (simultaneous regression) helped to specify the exact variables as predictors and provide a significant level based on the number of predictors (Leech *et al.*, 2011). All variables were entered/considered at the same time.

In observing the impact among variables of historical cost data, it was expected that some variables would demonstrate strong correlation among themselves. High correlations among variables indicated that there would likely be a problem with multicollinearity. In the presence of multicollinearity, the regression coefficient would be experienced with unduly large sampling variance which affects both inference and prediction (Graham, 2013; Kibria, 2003). Generally, a correlation coefficient of under −0.5 or over 0.5 indicated that the two predictors have a strong correlation (Leech *et al.*, 2011). If predictor variables were highly correlated and conceptually related to one another, Leech *et al.* (2011) suggested that they were normally aggregated to reduce not only the likelihood of multicollinearity but also the number of predictors (which typically increases power). If the predictor variables were highly correlated but conceptually different (so aggregation does not seem appropriate), the less important predictor might be eliminated before running the regression. The results of MRA models are described in Tables II and III.

A high R indicates that the model has the best fit for the data and is generally preferred. From Table II, the highest R score was produced by MRA 3, but $X3$ (number of stories) was not significant. Alternatively, MRA 2 had the next highest R value, slightly less than that of MRA 3 with both variables being significant. However, all models were significant to building cost per GFA with the coefficient of 0.000. Therefore, all independent variables in each model were used as input variables to the ANFIS.

MRA–ANFIS (MRANFIS) model
The existing data were randomly divided into two groups. The first group of 69 data was used for training system and another group of 19 data was used for testing system. Regarding the development of the model on ANFIS, the MATLAB software was used. There were three kinds of MRANFIS models that would be developed based on the input variable on MRA 1, MRA 2 and MRA 3.

Before starting the FIS training, an initial FIS model structure must be specified by choosing grid partition which generates a single-output Sugeno-type FIS by using grid partitioning on the data. In generating FIS for MRAFIS 1, as showed in Figure 2, the number of MFs of INPUT were set for [5 5] which indicated that each input had five MFs (very unlikely, unlikely, even, likely and very likely), rules (25 rules for each CCE), MFs for output (very low, low, medium, significant and high) and output (CCE). The type of MF of

Model	Model function
MRA 1	MRA 1 = 1,677,896 + 917,253 $TX1$ + 217,348 $X3$
MRA 2	MRA 2 = 1,100,503 + 1,066,364 $TX1$ + 606,401 $X9$
MRA 3	MRA 3 = 1,217,601 + 1,066,364 $TX1$ + 190,896 $X9$ + 639,407 $X3$
Note: $TX1 = \text{Log } X1$	

Table II.
Enter method
(simultaneous
regression) for CCE

Model	Independent variables	R	Variable contributes significantly to predicting CCE	p-value of the coefficient
MRA 1	$TX1, X3$	0.765	$TX1^{**}$	0.000
MRA 2	$TX1, X9$	0.794	$TX1^{**}, X9^{*}$	0.000
MRA 3	$TX1, X9, X3$	0.808	$TX1^{*}, X9^{*}$	0.000
Notes: $^{*}p < 0.05$; $^{**}p < 0.01$				

Table III.
MRA summary
predicting for CCE

input   inputmf   rule   outputmf   output

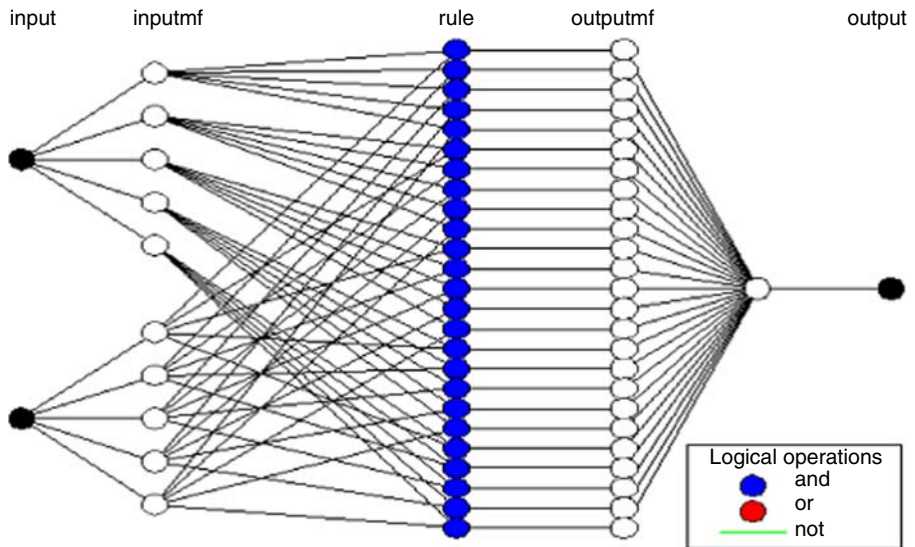

Logical operations
● and
● or
— not

Figure 2.
ANFIS structure for
MRANFIS 1

was set at "trapmf" indicating a generalized bell-shape MFs, and at OUTPUT, "linear" was chosen as the type of MF. Figure 3 shows the MFs for $X1$ and $X3$ for MRANFIS 1.

The ANFIS editor provided eight types of MFs, namely, triangular (trimf), trapezium (trapmf), generalized bell-shaped (gbellmf), Gaussian curve (gaussmf), Gaussian combination (gauss2mf), φ-shaped (pimf), difference between two sigmoidal functions (dsigmf) and product of two sigmoidal (psigmf). Regarding the selection of the best MF, different scenarios with the minimum error were chosen. Table IV shows the summary of variation in ANFIS modeling for each model.
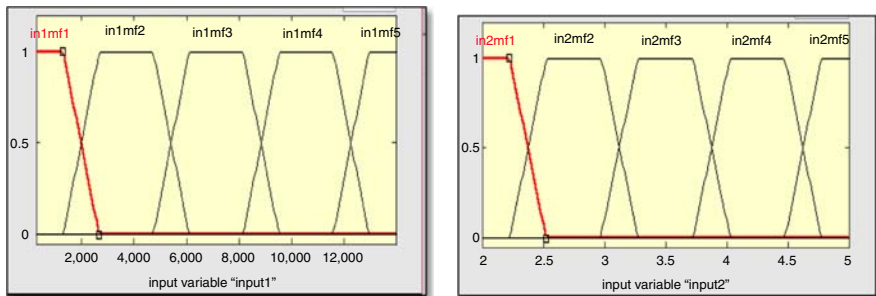

Figure 3.
Membership functions
for $X1$ and $X3$ for
MRANFIS 1

Specification		Model	
	MRANFIS 1	MRANFIS 2	MRANFIS 3
Input	2	2	3
Output neuron	1	1	1
Grid of partition	5 5	6 6	4 4 4
Rules	25	25	64
Types of MFs	Generalized bell-shaped	φ-shaped	Generalized bell-shaped
Output MFs	Linear	Linear	Linear

Table IV.
The summary of
variation in
ANFIS modeling

Model results

The measures for closeness of fit for all models were calculated by using data from all 78 projects. The prediction performances of all models were then compared by a procedure using the cross-validation technique. Three new data, not from the existing data of the 78 projects, were selected randomly as the prediction performance for the model. The results for closeness of fit and prediction performance of all models are shown in Table V and the comparison of estimate accuracy is depicted in Table VI.

The MSE and MAPE values of regression models for MRA 2 (5.43×10^{11} and 10.7, respectively) were smaller than those of MRA 1 and MRA 2, indicating that the MRA 2 provided a better fit to the data for the regression model. For similarity to closeness of fit to the data, MRA 2 also provided a good prediction performance for regression model because the values of MSE and MAPE were smaller than those of MRA 1 and MRA 2. Using the data examination and manipulation for regression model, only MRA 2 had all the significant variables (GFA, $X1$ and type of roof, $X9$) as shown in Table III. It also shows that the model of MRA 2 was selected as the best model for predicting CCE.

However, when compared to the proposed model, MRANFIS 3 with three variables of GFA ($X1$), number of stories ($X3$) and type of roof ($X9$) demonstrated a better fit to the existing data and a better prediction performance to the new data. The MSE and MAPE values of MRANFIS 3 (3.13×10^{11} and 7.90, respectively) for closeness of fit were slightly higher than those of MSE and MAPE (2.13×10^{11} and 6.7) for prediction performance. The range of estimate accuracy for MRANFIS 3 was better than MRA 2 as shown in Table V.

Conclusion

The CCE model for state building in West Sumatra, Indonesia was developed using MRA and integration of MRA with ANFIS. For the optimum model, MRA used fewer variables than ANFIS. This case was not only compared to ANFIS, but also compared to NN (Sonmez, 2004). ANFIS could identify the relationship between variables and project cost per GFA. On the contrary, regression was needed to define the class of relation (linear, quadratic, etc.) to be used in modeling. However, the regression could demonstrate the strength of relationship between two or more variables in the model.

The proposed model had successfully improved the MAPE by 2.8 percent from MRA 2 of 10.7 percent to MRANFIS 3 of 7.9 percent for closeness of fit to the data of the model by 3.1 percent from MRA 2 of 9.8 percent to MRANFIS 3 of 6.7 percent for prediction

Model	Closeness of fit		Prediction performance	
	MSE	MAPE (%)	MSE	MAPE (%)
MRA 1	5.70×10^{11}	11.5	2.94×10^{11}	11.4
MRA 2	5.43×10^{11}	10.7	2.62×10^{11}	9.8
MRA 3	1.20×10^{12}	10.9	1.51×10^{12}	26.4
MRANFIS 1	9.61×10^{12}	9.01	2.33×10^{11}	8.5
MRANFIS 2	4.01×10^{11}	9.42	7.81×10^{11}	15.1
MRANFIS 3	3.13×10^{11}	7.90	2.13×10^{11}	6.7

Table V.
Comparisons on closeness of fit and prediction performance

Model	Estimate accuracy			
	< 10%	10–20%	20–30%	30–40%
MRA 2	45	17	12	4
MRANFIS 3	54	16	8	0

Table VI.
Comparison of estimate accuracy

performance to the new data. The result was satisfactory, considering that the accuracy range of schematic design stage (for budget estimate) is ±10–30 percent (AACE, 2005).

Although efforts have been made to mitigate the errors and the fallacy in this research, the results are still subject to certain limitations. The proposed model was based on the cost data of 78 state building projects from one province only, thus, generalization of results to fully characterize the conceptual cost of Indonesia is apparently weak. Meanwhile, the use of ANFIS as a prediction instrument was limited to one output variable only. Having the limitations, this study recommends a further research to extend the boundaries of knowledge and to modify the findings. Therefore, it is encouraged to involve more data from other provinces which can represent entire Indonesia as a whole. It is also urged to employ other tools such as Feed forward NN which can modify several output variables.

References

AACE (2005), "Cost estimate classification system—as applied in engineering, procurement and construction for the process industries", TCM Framework: 7.3 – Cost estimating and budgeting, Association for the Advance of Cost Engineering International Recommended Practice No. 18R-97, VA.

Adeli, H. and Wu, M. (1998), "Regularization neural network for construction cost estimation", *Journal of Construction Engineering and Management*, Vol. 124 No. 1, pp. 18-27.

Ahn, J., Ji, S.-H., Park, M., Lee, H.-S., Kim, S. and Suh, S.-W. (2014), "The attribute impact concept: applications in case-based reasoning and parametric cost estimation", *Automation in Construction*, Vol. 43, pp. 195-203.

Ashworth, A. (1995), *Cost Studies of Buildings*, 2nd ed., Longman House, Harlow.

Azman, M.A., Samad, Z. and Ismail, S. (2012), "The accuracy of preliminary cost estimates in Public Works Department (PWD) of Peninsular Malaysia", *International Journal of Project Management*, Vol. 31, pp. 994-1005.

Cheung, F.K.T. and Skitmore, M. (2006), "Application of cross-validation techniques for modeling construction costs during the very early design stage", *Building and Environment*, Vol. 41, pp. 1973-1990.

Chou, J. (2009), "Web-based CBR system applied to early cost budgeting for pavement maintenance project", *Expert Systems with Applications*, Vol. 36 No. 2, pp. 2947-2960.

Creese, R.C. and Li, L. (1995), "Cost estimation of timber bridge using neural networks", *Cost Engineering*, Vol. 37 No. 5, pp. 17-23.

Elfaki, A.O., Alatawi, S. and Abushandi, E. (2014), "Using intelligent techniques in construction project cost estimation: 10-year survey", *Advances in Civil Engineering*, Vol. 2014, 11pp.

Graham, M.H. (2013), "Confronting multicollinearity in ecological multiple regression", *Ecology*, Vol. 84 No. 11, pp. 2809-2815.

Gunduz, M., Ugur, L. and Ozturk, E. (2011), "Parametric cost estimation system for light rail and metro trackworks", *Expert Systems with Application*, Vol. 38, pp. 2873-2877.

Hegazy, T. and Ayed, A. (1998), "Neural network model for parametric cost estimation of highway projects", *Journal of Construction Engineering and Management*, Vol. 124 No. 3, pp. 210-221.

Jang, J.S. (1993), "ANFIS: adaptive-network-based fuzzy inference system", *IEEE Transactions on System Man and Cybernetics*, Vol. 23, pp. 665-685.

Ji, S.-H., Park, M. and Lee, H.-S. (2010), "Data preprocessing-based parametric cost model for building projects: case studies of Korean construction projects", *Journal Construction Engineering Management*, Vol. 138 No. 8, pp. 844-853.

Jin, R., Cho, K., Hyun, C. and Son, M. (2012), "MRA-based revised CBR model for cost prediction in the early stage of construction projects", *Expert Systems with Applications*, Vol. 39, pp. 5214-5222.

Jumas, D., Rahim Faizul, A.M. and Zainon, N. (2016), "Influences of cost variables on the conceptual cost estimation accuracy: methods and techniques", *The Malaysia Surveyor Journal*, Vol. 51 No. 3, pp. 7-16.

Kibria, B.M.G. (2003), "Performance of some new ridge regression estimators", *Communications in Statistics*, Vol. 32 No. 2, pp. 419-435.

Kim, G.H., An, S.H. and Kang, K.I. (2004), "Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning", *Building and Environment*, Vol. 39, pp. 1235-1242.

Kim, G.H., Seo, D.S. and Kang, K.I. (2005), "Hybrid models of neural networks and genetic algorithms for predicting preliminary cost estimates", *Journal of Computing in Civil Engineering*, Vol. 19 No. 2, pp. 208-213.

Koo, C., Hong, T. and Hyun, C. (2011), "The development of a construction cost prediction model with improved prediction capacity using the advanced CBR approach", *Expert Systems with Applications*, Vol. 38 No. 7, pp. 8597-8606.

Kwak, Y.H. and Watson, R.J. (2005), "Conceptual estimating tool for technology-driven projects: exploring parametric estimating technique", *Technovation*, Vol. 25, pp. 1430-1436.

Latief, Y., Wibowo, A. and Isvara, W. (2013), "Preliminary cost estimation using regression analysis incorporated with adaptive neuro-fuzzy inference", *International Journal of Technology*, Vol. 1, pp. 63-72.

Leech, N.L., Barret, K.C. and Morgan, G.A. (2011), *IBM SPSS for Intermediate Statistics*, Routledge, Taylor & Francis Group, New York, NY.

Lowe, D.J., Emsley, M.W. and Anthony, H. (2006), "Predicting construction cost using multiple regression techniques", *Journal of Construction Engineering and Management*, Vol. 132 No. 7, pp. 750-758.

Marzouk, M.M. and Ahmed, R.M. (2011), "A case-based reasoning approach for estimating the costs of pump station projects", *Journal of Advanced Research*, Vol. 2, pp. 289-295.

Moon, S.W., Kim, J.S. and Kwon, K.N. (2007), "Effectiveness of OLAP-based cost data management", *Automation in Construction*, Vol. 16, pp. 336-344.

Phaobunjong, K. (2002), *Parametric Cost Estimation Model for Conceptual Cost Estimation of Building Construction Project*, The University of Texas at Austin, Austin, TX.

Riquelme, P. and Serpell, A. (2013), "Adding qualitative context factors to analogy estimating of construction projects", *Social and Behavioral Sciences*, Vol. 74, pp. 190-202.

Shane, J.S., Molenaar, K.R., Anderson, S.R. and Schexnayder, C. (2009), "Construction project cost escalation factors", *Journal of Management in Engineering*, Vol. 25 No. 4, pp. 221-229.

Skitmore, M. (1991), *Early Stage Construction Price Forecasting: A Review of Performance*, RICS, London.

Sonmez, R. (2004), "Conceptual cost estimation of building project with regression analysis and neural network", *Canadian Journal Civil Engineering*, Vol. 31, pp. 677-683.

Takagi, T. and Sugeno, M. (1983), "Derivation of fuzzy control rules from human operator's control actions", IFAC Symposium on Fuzzy Information Knowledge Representation and Decision Analysis, pp. 55-60.

Tam, C.M. and Fang, C.F. (1999), "Comparative cost analysis of using high-performance concrete in tall building construction by artificial neural networks", *Structural Journal*, Vol. 6, pp. 927-936.

Watson, I. (1997), *Applying Case-Based Reasoning: Techniques for Enterprise System*, Morgan Kaufmann Publishers, San Francisco, CA.

Corresponding author
Dwifitra Jumas can be contacted at: dwifitraj@gmail.com