Structural Engineering and Mechanics, Vol. 45, No. 5 (2013) 693-702

DOI: http://dx.doi.org/10.12989/sem.2013.45.5.693

A novel regression prediction model for structural engineering applications

Jeng-Wen Lin[1], Cheng-Wu Chen[*2,3] and Ting-Chang Hsu[1]

[1]*Department of Civil Engineering, Feng Chia University, Taichung 407, Taiwan, R.O.C.*
[2]*Department of Maritime Information and Technology, National Kaohsiung Marine University, Kaohsiung, Taiwan, R.O.C.*
[3]*Global Earth Observation and Data Analysis Center, National Cheng Kung University, Tainan, Taiwan 701, R.O.C.*

(Received July 1, 2012, Revised January 15, 2013, Accepted February 19, 2013)

Abstract. Recently, artificial intelligence tools are most used for structural engineering and mechanics. In order to predict reserve prices and prices of awards, this study proposed a novel regression prediction model by the intelligent Kalman filtering method. An artificial intelligent multiple regression model was established using categorized data and then a prediction model using intelligent Kalman filtering. The rather precise construction bid price model was selected for the purpose of increasing the probability to win bids in the simulation example.

Keywords: construction project and management; intelligent fuzzy regression; Kalman filtering; prediction model

1. Introduction

In recent years, many fuzzy regression models are proposed to solve many kinds of practical problems (Esfahanipour and Aghamiri 2010, Shakouri *et al*. 2009). It is believed that the tool of fuzzy regression can be used to deal with the civil engineering problem such as the prediction of tendering price on roadway construction. Therefore, a new viewpoint using fuzzy regression with Kalman filtering is developed in this paper for a prediction model in roadway construction. Being subjected to severe competition for such a long time, the construction industry is currently facing the problems of price competition and low price bidding, often leading to narrowed profit margins. Therefore, bid winners have been forced to adopt the business pattern of sub-contracting in order to reduce their risks by transferring costs and responsibilities. The sub-contracting pattern may in turn have negative effects. There are several possible situations including decrease in the quality of construction and difficulty in managing multiple tasks at the same time. It thus becomes very important to select a contractor that is in excellent financial condition. A sound financial condition is reflective of the contractor's reliability and capability of planning, organization, control and human resource management. Therefore, prior to awarding a tender, the related contractors'

∗Corresponding author, Professor, E-mail: chengwu@mail.nkmu.edu.tw

engineering experience and financial condition should be examined (Bubshalt and Al-Gobali 1996, Russell and Skibniewski 1988, Tarawneh 2004, Edum-Fotwe *et al*. 1996).

To obtain public construction contracts, construction firms normally have to participate in the tender process. For example, most road construction projects in Taiwan are submitted for public tender, with the contract being awarded to the lowest bidders. Generally, constructors, including construction companies or responsible individuals, place low bids to increase their competitiveness. However, low prices often lead to low quality and low profits can lead to the construction company closing down. When participating in the tender process, constructors should increase their chances of winning bids while still ensuring profits. Scientific methods can be used to predict the most reasonable reserve prices for the construction project and to determine whether or not to place a bid and the most appropriate bid price. Authorities often have to set reserve prices for tenders of public construction projects. Therefore, authorities responsible for construction tenders should collect reserve prices from previous construction projects which can be used as a reference to predict reserve prices for the current project.

Yan (2006) found the factors to evaluate construction projects for private school construction, using a revised Delphi questionnaire based on the fuzzy set theory. Yan determined the overall value of every evaluating factor in different combinations using game theory for the mathematical calculation process. The factors were evaluated according to the fuzzy linguistic importance level as obtained by examining the preferences of four different decision-making departments in private schools. The precise weights were then calculated through defuzzification of fuzzy weights to obtain total scores for possible proposal choices. These were used to develop a fuzzy decision making evaluation model for construction proposal evaluation. The feasibility of this method was proved with a case study.

Hong (2007) found the key factors of influence on indirect fees for construction. He developed a model, the output being the ratio of indirect to direct fees. According to data for 187 cases collected for the study and the results of statistical analyses, three input factors were determined after comparison, including project type, construction period, and project scale (represented by direct fees). That study adopted fuzzy neural networks as the prediction model. The results showed that this model could be used to predict indirect fees for construction and provide validation data for predicted indirect fees.

Lin *et al*. (2012) designed a model to assist the public sector to determine what would be a reasonable reserve price or award price. In order to ensure accurate predictions, a data classification system is established using fuzzy set theory. For each category of classified data, multiple regression analysis is applied to the linear model, the power series model, and the refined power series model. Following this article, in this study we attempt to build a public construction tender price prediction model by combining fuzzy theory with statistical refined models and Kalman filtering. This model is different from the tender price models previously built by others. It is expected that using this model it is possible to offer more accurate predictions for constructors and that authorities responsible for construction can use the results as reference to determine bid prices and set reserve prices.

2. Fuzzy regression using T-S fuzzy model

In order to establish a fuzzy regression model, the concept of the so-called Takagi-Sugeno fuzzy model is utilized in a fuzzy inference engine. According to this proposed regression methodology, data clusters are distributed into a few overlapping clusters. A T-S fuzzy model is then constructed based on the regression curves obtained for different data clusters (Lin 2012).

The fuzzy theory was originated by Zadeh (1965). He proposed the fuzzy sets to present the concept of uncertainty or fuzziness. Such uncertainty was different from the probability theory, whose uncertainty referred to the randomness of an event. For instance, Mikut *et al.* (2005) established fuzzy systems with a user-controllable trade-off between accuracy and interpretability, whose interpretability was maintained by the type of membership functions, rules, and inference mechanism; on the other hand, interpretability criteria were included in the rule evaluation. The multi-input multi-output fuzzy rules were developed by Takagi and Sugeno (1985) to establish a fuzzy polynomial regression model. According to this regression methodology, a Takagi-Sugeno (T-S) fuzzy model (Xu and Chen 2012) was built using the regression curve to process the available data with an n-th order curve fitting

$$y = \sum_{k=0}^{n} a_k x^k \tag{1}$$

where n denotes the necessary maximum order for optimal curve fitting. Once the order n is defined, it is possible to represent nonlinear relations between inputs and outputs. The parameters a_k, k = 0, 1, …, n, are determined by defining the distance (or error) between the observed and estimated data points on the polynomial being minimal so as to smoothly connect the fuzzy regression curve.

The nonlinear fuzzy regression model can be achieved by the fuzzy blending of each individual input-output realization (Hsiao *et al.* 2003). In this study, the data clusters of independent variables (e.g., the budget price, the calendar days, and the tender bond) can be described by the membership functions of fuzzy sets C_1, C_2, …, C_1 that may contain overlapping regions to represent the dependent variable (e.g., the reserve price or the price of award). The rule of fuzzy inference is then described by a set of IF-THEN rules in the following (Chen *et al.* 2009, Lin 2012)

$$\begin{aligned}
&\text{IF } x \text{ is } C_1 \quad \text{THEN } y_1 = \sum_{k=0}^{n} a_{1k} x_1^k \\
&\text{IF } x \text{ is } C_2 \quad \text{THEN } y_2 = \sum_{k=0}^{n} a_{2k} x_2^k \\
&\qquad\vdots \qquad\qquad\qquad \vdots \\
&\text{IF } x \text{ is } C_i \quad \text{THEN } y_i = \sum_{k=0}^{n} a_{ik} x_i^k
\end{aligned} \tag{2}$$

where x is the input value (e.g., the value of an independent variable) and y is the output value (e.g., the value of a dependent variable), while i = 1, 2, …, r, where r denotes the number of IF-THEN rules. By using the center of gravity defuzzification, product inference, and single fuzzifier, the final output can be inferred as (Chen *et al.* 2009, Lin 2012)

$$y = \frac{\sum_{i=1}^{r} w_i \, y_i}{\sum_{i=1}^{r} w_i} = \sum_{i=1}^{r} h_i \, y_i \tag{3}$$

assuming that $w_i \geq 0,\ i = 1, 2, \cdots, r$ and $\sum_{i=1}^{r} w_i > 0$, and thus $h_i \geq 0$ and $\sum_{i=1}^{r} h_i = 1$.

3. Kalman filter

The Kalman filter scheme provides covariance for system status prediction errors, in opposition to the adaptation gain matrix for multiple regressions, with other model parameters which need to be adjusted (Lin 2001, Lin *et al*. 2001).

In order to reduce relative modeling error, the Kalman filter is used to estimate parameters (Kim *et al*. 2011, Li *et al*. 2004, Lin 2001, Smyth *et al*. 1999)

$$y_k = \Phi_k \theta_k + e_k \tag{4}$$

$$\theta_{k+1} = \Gamma_k \theta_k \tag{5}$$

For the time step k

y_k : the q×1 measured vector;

Φ_k : the q×n observation matrix;

θ_k : the n×1 row vector, including unknown system parameters which need to be identified;

e_k : the observation error vector;

Γ_k : the n×n system transfer matrix which transfers the parameter vector at time step k to time step (k+1).

In order to estimate the parameter vector $\hat{\theta}_{k+1}$ at time step (k+1), the multiple minimum square expression of iteration can be written as

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \mathrm{W}_{k+1}\left[\bar{y}_{k+1} - \Phi_{k+1}\hat{\theta}_k\right] \tag{6}$$

$$\mathrm{W}_{k+1} = \bar{\mathrm{G}}_{k+1}^{-1}\Phi_{k+1}^T\left(\Phi_{k+1}\bar{\mathrm{G}}_{k+1}^{-1}\Phi_{k+1}^T + \mathrm{I}_q\right)^{-1} \tag{7}$$

$$\bar{\mathrm{G}}_{k+1}^{-1} = \left[\mathrm{I}_n - \mathrm{W}_k\,\Phi_k\right]\bar{\mathrm{G}}_k^{-1} \tag{8}$$

where

I_n : the identity matrix, with the order being n;

$\bar{y}_{k+1}$: the measured input/output vector at time step k+1;

W_{k+1} : the weighting matrix, updated with time;

$\bar{\mathrm{G}}_{k+1}^{-1}$: the adaptation gain matrix, updated with time;

I_q : the identity matrix, with the order being q.

The values in the adaptation gain matrix $\bar{\mathrm{G}}_{k+1}^{-1}$ directly influence estimations of system parameters. Since the hypothesis is that there is no prior information of system parameters, therefore, the initial value $\hat{\theta}_0$ is set to be close to 0 at time step k=0. The initial values in the adaptation gain matrix ($\overline{\mathrm{G}}_1^{-1} = 10^{-9}\,\mathrm{I}_n$ at time step k=0) must converge quickly. However, with more cases, more data will be used and the adaptation gain will be reduced, leading to the situation in which system parameters cannot be accurately estimated (Lin 2001, Smyth *et al*. 1999). In order to avoid this disadvantage, this study implemented a constant forgetting factor to modify the Kalman filter.

3.1 Constant forgetting factor

In order to avoid system parameters from becoming smaller and eventually reaching turnoff, it is appropriate to set smaller weights to data in previous time steps. This can be done by introducing the forgetting factor λ_k, which changes the expression of the correlation matrix to (Lin and Betti 2004, Lin and Chen 2009)

$$\mathrm{G}_{k+1} = \lambda_k\,\mathrm{G}_k + \Phi_{k+1}^T \Phi_{k+1}, \quad 0 < \lambda_k \le 1 \tag{9}$$

The adjusted $\overline{\mathrm{G}}_{k+1}^{-1}$ becomes

$$\overline{\mathrm{G}}_{k+1}^{-1} = (\frac{1}{\lambda_k})\,\mathrm{G}_k^{-1}, \quad 0 < \lambda_k \le 1 \tag{10}$$

This will lead to a new estimation for $\hat{\theta}_{k+1}$. Because of the introduction of the forgetting factor, values in the adaptation gain matrix $\overline{\mathrm{G}}_{k+1}^{-1}$ will be enhanced, further minimizing the square error function.

From the aspect of numerical implementation, the forgetting factor λ_k is set to be a constant ($\lambda_k = \lambda$, constant forgetting factor). This assumption is particularly suitable to identify system parameters with slow changes (Lin and Betti 2004, Smyth *et al*. 1999). The optimal value for the constant forgetting factor is in the range of 0.95-0.99, as determined by the number of system parameters.

3.2 Extended Kalman filter model

Kalman filtering can be considered to be a least squares regression method. Kalman filtering provides covariance of system status prediction errors and estimated parameters. This study built models to estimate system parameters using the Kalman filter.

The software MATLAB 7.1 is adopted to establish Kalman filter estimation models and system parameters to further predict reserve prices and prices of awards for road construction. In order to be able to accurately estimate system parameters caused by reduced adaptation gains when the number of time steps increases and more data are used, this study implemented the constant forgetting factor to modify the Kalman filter. Data from the power series models (4 categories) were used to build the reserve price model, while data from the linear models (3 categories) were used to build the price of award model (Lin *et al*. 2012).

4. Reserve price model

The Extended Kalman filter reserve price model was built with data from the power series models with 4 categories (Lin *et al*. 2012). The extended Kalman filter analysis was conducted with the reserve price (y) and 19 independent variables to build the reserve price model.

Category 1 data were used to illustrate the process of model establishment. Extended Kalman filter analysis was conducted with the reserve price (y) and 19 independent variables, with the implemented constant forgetting factor being 0.95 (Fig. 1).

The following regression model was obtained by substituting the final parameters from Fig. 1 (the values for the last point) into the model

$$\begin{aligned}y &= (1.38)x_1 + (3\times10^3)x_2 - (8.31)x_3 + (1.53\times10^{-7})x_4 + (1.99\times10^2)x_5 \\ &+ (1.51\times10^{-4})x_6 - (9.76\times10^{-3})x_7 - (9.77\times10^{-6})x_8 - (3.97\times10^{-2})x_9 \\ &- (1.38\times10^{-15})x_{10} - (3.42)x_{11} - (4.2\times10^{-10})x_{12} - (1.47\times10^{-9})x_{13} \\ &- (3.38\times10^{-13})x_{14} + (9.6\times10^{-5})x_{15} + (1.13\times10^{-3})x_{16} + (2.86\times10^{-11})x_{17} \\ &- (7.96\times10^{-7})x_{18} + (5.99\times10^{-8})x_{19}\end{aligned} \tag{11}$$

Formula (11) was used as the reserve price prediction model for category 1 data. The verification error was 4.92% and the prediction error was 5.47%. Models built with different forgetting factors and different categories were different. Here the example is the reserve price model built using category 1 data with the constant forgetting factor being 0.95.

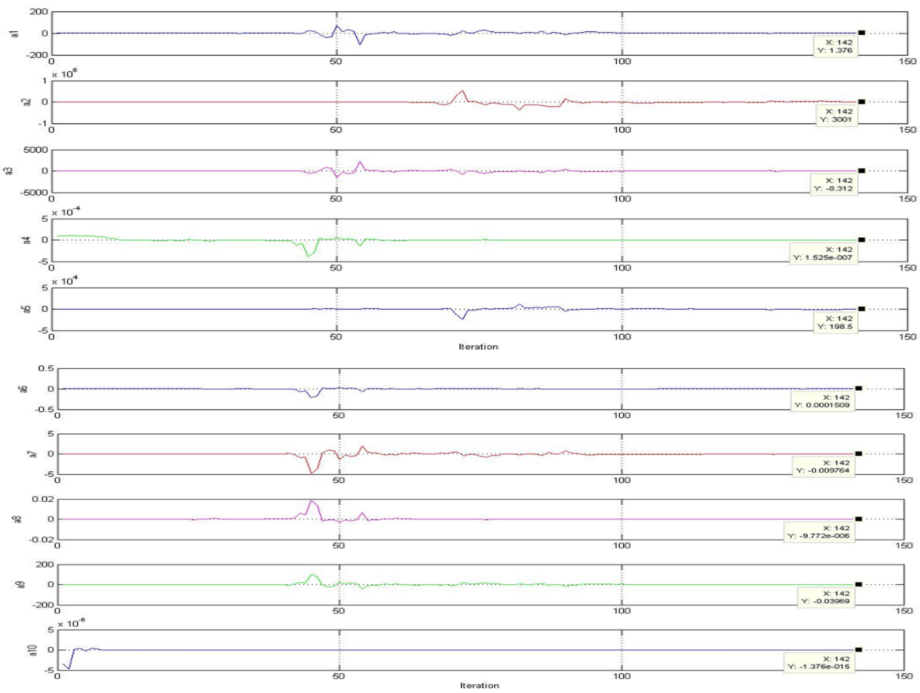

Fig. 1 System parameters for a reserve price model

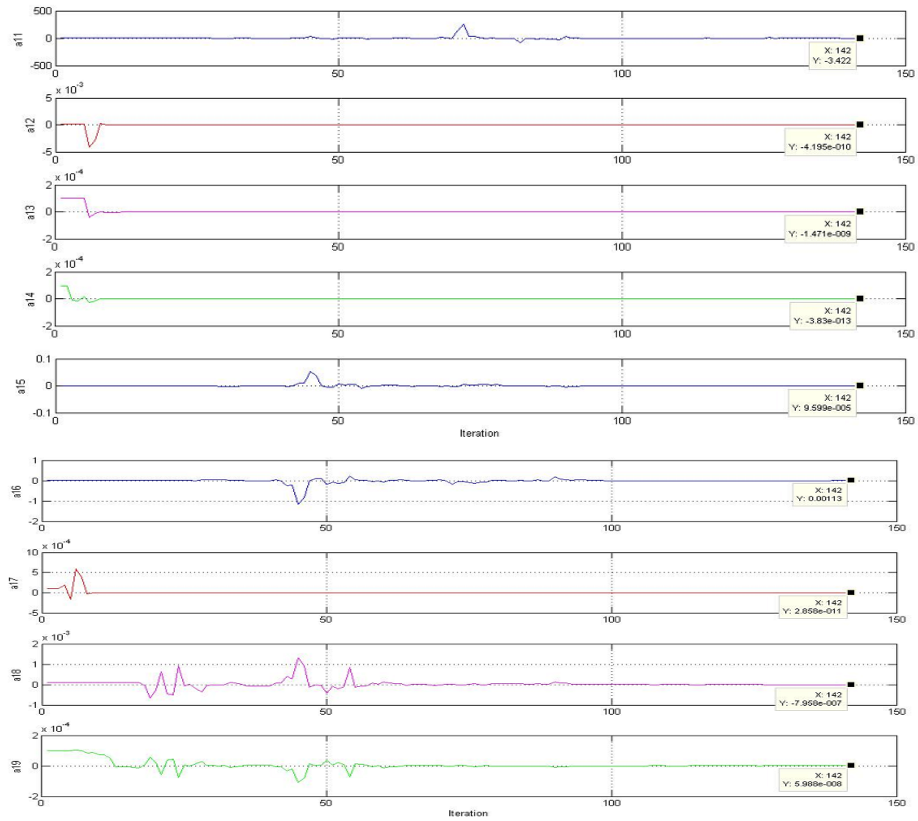

Fig. 1 Continued

Table 1 Comparison of verification error and prediction error for reserve price models built with different forgetting factors

Reserve price	Category / Forgetting factor	Category 1	Category 2	Category 3	Category 4
Verification model	0.95	4.92%	11.11%	1030.71%	2.44%
	0.96	5.95%	11.14%	458.31%	2.44%
	0.97	4.44%	11.47%	1101.19%	2.44%
	0.98	4.14%	10.28%	13931.27%	2.44%
	0.99	4.08%	10.60%	71089.63%	2.44%
Prediction model	0.95	5.47%	8.10%	1129.33%	14.91%
	0.96	8.45%	8.54%	458.29%	14.91%
	0.97	5.12%	9.18%	110026.1%	14.91%
	0.98	6.13%	8.07%	15855.46%	14.91%
	0.99	7.54%	8.85%	67150.69%	14.91%

Different constant forgetting factors were implemented on the reserve price models built using data from 4 different categories. The analysis results are shown in Table 1.

Table 2 Comparison of verification error and prediction error for price of award models built with different forgetting factors

Price of award	Category / Forgetting factor	Category 1	Category 2	Category 3
Verification model	0.95	21.15%	6.08%	5.36%
	0.96	20.71%	6.28%	5.40%
	0.97	20.1%	6.18%	5.96%
	0.98	19.32%	6.51%	5.38%
	0.99	18.31%	6.05%	5.36%
Prediction model	0.95	18.75%	8.12%	17.92%
	0.96	18.01%	8.37%	18.42%
	0.97	17.11%	8.33%	17.79%
	0.98	16.44%	8.54%	18.16%
	0.99	16.52%	8.45%	17.80%

Table 3 Results of extended Kalman filtering

Category				Category 1	Category 2	Category 3	Category 4
Reserve price verification error	4	Forgetting factor	0.95	4.92%	11.11%	1030.71%	2.44%
			0.96	5.95%	11.14%	458.31%	2.44%
			0.97	4.44%	11.47%	1101.199	2.44%
			0.98	4.14%	10.28%	13931.27%	2.44%
			0.99	4.08%	10.60%	71089.63%	2.44%
Price of award verification error	3	Forgetting factor	0.95	21.15%	6.08%	5.36%	
			0.96	20.71%	6.28%	5.40%	
			0.97	20.1%	6.18%	5.96%	
			0.98	19.32%	6.51%	5.38%	
			0.99	18.31%	6.05%	5.36%	

5. Price of award model

The extended Kalman filter price of award model was built using data from a linear model, with 3 categories (Lin *et al*. 2012). Extended Kalman filter analysis was conducted with the price of award (y: the dependent variable), the budget price (x1: the independent variable), the calendar days (x2: the independent variable), and the tender bond (x3: the independent variable) to build the price of award model. Because the method used to build the price of award model was the same as the one used to build the reserve price mode, only the price of award model is illustrated in this article. Different constant forgetting factors were implemented for the price of award models built using data from 3 different categories. The analysis results are shown in Table 2.

Table 3 shows that only the price of award model built using category 1 data with the constant forgetting factor being 0.99 did not require high adaptation values, because its verification error was decreased from 19.46% to 18.31% through Kalman filtering. Models built with other category data showed no improvement.

6. Conclusions

By referencing purchase announcement guidelines by the government and previous studies, three factors were chosen as influential for reserve price modeling and price of award modeling for road construction projects. First of all, the collected cases were categorized through a categorization system built by applying fuzzy theory. Then the categorized data were used to build the reserve price model and price of award model for road construction projects using fuzzy regression and extended Kalman filtering. This study used data for public tenders for construction projects in 2005 as an example. Data from construction projects with amounts under NT$50,000,000 were used to develop a reserve price model and price of award model for road construction projects. The conclusions are summarized below:

1. Constructors which participate in tenders may predict prices of award as a reference using information of the budget price, the calendar days, and the tender bond. In addition, proprietors may also use this model as a reference to set reserve prices.

2. Comparison of the model built with uncategorized data and the one built with categorized data using fuzzy theory proved that the categorization system developed in this study could actually reduce errors. In other words, applying fuzzy theory to bidding price models is feasible.

3. The optimal reserve price model was the power series model (4 categories) while the optimal price of award model was the linear model (3 categories). These models were selected as the reserve price model and the price of award model. The price of award model built with category 1 data through extended Kalman filtering could reduce errors (the forgetting factor used was 0.99) from 19.46% o 18.31%.

The reserve price model was more feasible. However, predictions of prices of award are probably influenced by many complex factors. Therefore, it was not practical to develop a more effective prediction model.

Acknowledgments

The authors would like to thank the National Science Council, R.O.C., for their financial support of this research under Contract Nos. NSC 97-2221-E-035-045, NSC 100-2221-E-022-013-MY2 and NSC 100-2628-E-022-002-MY2.

References

Bubshalt, A.A. and Al-Gobali, K.H. (1996), “Contractor prequalification in Saudi Arabia,” *Journal of Management in Engineering*, ASCE, **12**(2), 50-54.

Chen, C.W., Wang, M.H.L. and Lin, J.W. (2009), “Managing target the cash balance in construction firms using a fuzzy regression approach”, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **17**(5), 667-684.

Edum-Fotwe, F., Price, A. and Thorpe, A. (1996), “A review of financial ratio tools for contractor insolvency”, *Construction Management and Economics*, **14**, 189-198.

Esfahanipour, A. and Aghamiri, W. (2010), “Adapted neuro-fuzzy inference system on indirect approach TSK fuzzy rule base for stock market analysis”, *Expert Systems with Applications*, **37**, 4742-4748.

Hong, J.Y. (2007), “Fuzzy neural network model for estimating project indirect cost ratio”, Master Thesis, Department of Construction Engineering, National Kaohsiung First University of Science and Technology, Taiwan. (In Chinese)

Hsiao, F.H., Chiang, W.L. and Chen, C.W. (2003), "Application and fuzzy H^∞ control via T-S fuzzy models for nonlinear time-delay systems", *International Journal on Artificial Intelligence Tools*, **12**(2), 117-137.

Kim, C.W., Kawatani, M., Ozaki, R. and Makihata, N. (2011), "Recovering missing data transmitted from a wireless sensor node for vibration-based bridge health monitoring", *Structural Engineering and Mechanics*, **38**(4), 417-428.

Klir, G.J. and Yuan, B. (1995), *Fuzzy Sets and Fuzzy Logic Theory and Applications*, Prentic-Hall, Englewood Cliffs, NJ.

Li, S.J., Suzuki, Y. and Noori, M.N. (2004), "Improvement of parameter estimation for non-linear hysteretic systems with slip by a fast Bayesian bootstrap filter", *International Journal of Nonlinear Mechanics*, **39**(9), 1435-1445.

Lin, J.W. (2001), "Adaptive algorithms for the identification of nonlinear structural systems", Ph.D. Dissertation, Columbia University, New York.

Lin, J.W. (2012), "Fuzzy regression decision systems for assessment of the potential vulnerability of bridge to earthquakes", *Natural Hazards*, **64**(1), 211-221.

Lin, J.W., Betti, R., Smyth, A.W. and Longman, R.W. (2001), "On-line identification of non-linear hysteretic structural systems using a variable trace approach", *Earthquake Engineering and Structural Dynamics*, **30**(9), 1279-1303.

Lin, J.W. and Betti, R. (2004), "On-line identification and damage detection in non-linear structural systems using a variable forgetting factor approach", *Earthquake Engineering and Structural Dynamics*, **33**(4), 419-444.

Lin, J.W., Chen, C.W. and Hsu, T.C. (2012), "Fuzzy statistical refinement for the forecasting of tenders for roadway construction", *Journal of Marine Science and Technology*, **20**(4), 410-417.

Lin, J.W. and Chen, H.J. (2009), "Repetitive identification of structural systems using a nonlinear model parameter refinement approach", *Shock and Vibration*, **16**(3), 229-240.

Mikut, R., Jäkel, J., Gröll, L. (2005), "Interpretability issues in data-based learning of fuzzy systems", *Fuzzy Sets and Systems*, **150**(2), 179-197.

Niimura, T. and Nakashima, T. (2001), "Deregulated electricity market data representation by fuzzy regression models", *IEEE Trans. Syst., Man, Cybern.*, **31**, 320-326.

Russell, J.S. and Skibniewwski, M.J. (1988), "Decision criteria in contractor prequalification", *Journal of Management in Engineering*, ASCE, **4**(2), 148-164.

Shakouri, H., Nadimi, R., and Ghaderi, F. (2009), "A decision support tool for apparel coordination through integrating the knowledge-based attribute evaluation expert system and the T-S fuzzy neural network", *Expert Systems with Applications*, **36**, 2377-2390.

Smyth, A.W., Masri, S.F., Chassiakos, A.G. and Caughey, T.K. (1999), "On-line parametric identification of MDOF nonlinear hysteretic systems", *Journal of Engineering Mechanics*, **125**, 133-142.

Takagi, T. and Sugeno, M. (1985), "Fuzzy identification of systems and its applications to modeling and control", *IEEE Transactions on Systems, Man and Cybernetics*, **15**(1), 116-132.

Tarawneh, S.A. (2004), "Evaluation of pre-qualification criteria: client perspective; Jordan case study", *Journal of Applied Sciences*, **4**(3), 354-363.

Xu, Y. and Chen, J. (2012), "Fuzzy control for geometrically nonlinear vibration of piezoelectric flexible plates", *Structural Engineering and Mechanics*, **43**(2), 163-177.

Yan, Z.X. (2006), "Application of fuzzy set theory for evaluation of construction alternatives for private schools", Master Thesis, Department of Construction Engineering, National Kaohsiung First University of Science and Technology, Taiwan. (In Chinese)

Zadeh, L.A. (1965), "Fuzzy sets", *Information and Control*, **8**, 338-353.