A NEURO-FUZZY HYBRID MODEL FOR PREDICTING FINAL COST OF WATER INFRASTRUCTURE PROJECTS

Dominic D. Ahiaga-Dagbui[1], Olubukola Tokede[2], Simon D. Smith[1], Sam Wamuziri[3]

[1] *School of Engineering, University of Edinburgh, EH9 3JL, UK*

[2]*School of Engineering and the Built Environment, Edinburgh Napier University, Edinburgh, EH10 5DT, UK*

[3] *Centre for Management Research, Glyndŵr University, Wrexham, LL11 2AW, UK*

Nine out of ten infrastructure projects exceed their initial cost estimates. Accuracy of construction cost estimates remains a contentious area of debate within both academia and industry. Explanations for this have ranged from scope changes, risk and uncertainty, optimism bias, technical and managerial difficulties, suspicions of corruption, lying and insufficient required information for accurate estimation. The capacity for tolerance and imprecise knowledge representation of fuzzy set theory is combined with the learning and generalising capabilities of neural networks to develop neuro-fuzzy hybrid cost models in this paper to predict likely final cost of water infrastructure projects. The will help to increase reliability, flexibility and accuracy of initial cost estimates. Neural networks is first used to develop relative numerical weightings of cost predictors extracted from primary data collected on 98 completed projects. These were then standardised into fuzzy sets to establish a consistent framework for combining the effect of each variable on the overall final cost. A three-point fuzzy lower, upper and mean estimate of likely final cost is generated to provide a tolerance range for final cost rather than the traditional single point estimate. The performance of the final models ranged from 3.3% underestimation to 1.6 % overestimation. The best models however averaged an error of 0.6% underestimation and 0.8% overestimation of final cost of the project. The results are now being extended to a larger database of about 4500 projects in collaboration with an industry partner.

Keywords: artificial neural network, cost estimation, cost modelling, cost overrun, fuzzy set theory.

INTRODUCTION

Infrastructure projects have an 86% likelihood of exceeding the initial cost estimates and 9 out of 10 of them exceed their budgets (Flyvbjerg *et al.* 2002). A key example

[1] D.Ahiaga-Dagbui@ed.ac.uk

[2] O.Tokede@napier.ac.uk

Ahiaga-Dagbui DD, Tokede O, Smith SD and Wamuziri S (2013) A neuro-fuzzy hybrid model for predicting final cost of water infrastructure projects *In:* Smith, S.D and Ahiaga-Dagbui, D.D (Eds) *Procs 29th Annual ARCOM Conference,* 2-4 September 2013, Reading, UK, Association of Researchers in Construction Management, 181-190.

is the case of the stadiums built for the 2010 FIFA World Cup games in South Africa. With overruns ranging between 5 to 94% of original cost, none of the 10 stadiums were completed within budget (Baloyi and Bekker 2011). There is overwhelming evidence in literature, and practice, which support the conclusion that cost overrun is endemic within the construction industry, irrespective of size, type, sector or geographical location of the project (see Jackson 2002; Flyvbjerg *et al.* 2004; Odeck 2004; Baloyi and Bekker 2011). Cost remains arguably one of the most important key performance indicators on most projects (Chan and Chan 2004; Yeung *et al.* 2008) so that statistics, such as the ones above, leaves most clients grossly dissatisfied, giving the industry a poor reputation regarding budget reliability (Agyakwa-Baah 2009).

Despite its importance, cost estimation is undeniably not simple, nor straightforward, largely due to the dearth of information required for detailed estimation. It is even made worse by the cloud of uncertainty that shrouds cost drivers in the early stages of the project (Hegazy 2002) and the changes that occur in scope and design of the project once construction actually begins (Love *et al.* 2011; Gil and Lundrigan 2012). It is an inexact science and estimators have to make decisions within an environment of uncertainty. Moreover, even though it is accepted that factors such as tendering method, type of client, location of project, procurement method, size of project etc. have an effect on final cost of a project, it is difficult to establish their measured financial impact (Ahiaga-Dagbui and Smith 2012). This complex web of cost influencing variables would make it seem that the decision-to-build, for most projects, is based on a somewhat unrealistic cost estimate that will inevitably be exceeded.

Against this backdrop, debates have not waned on causes and measures of cost overruns. A recent discussion on the Construction Network of Building Researchers (CNBR) left a number of unresolved questions. How accurate can estimates be? Is there an acceptable way to compare final cost of project to cost estimates? What is the most acceptable measure of cost performance on a construction project? Is it even possible to achieve certainty of cost estimates, when the very estimates are made in an environment of uncertainty? (see the Nov 2012 CNBR archive online).While the answers to these can be varied; even sometimes strongly opposing; it is difficult to disagree that clients and project financiers still require some form of reasonably accurate estimate of their likely financial commitment for a project before the project begins.

In this paper, the authors attempt to model the final cost of water infrastructure projects using gathered cost data and other project details such as location, procurement method, size of project, type of client, etc of 98 water infrastructure projects. This paper, a sequel to a previous that uses only neural networks for modelling final cost (see Ahiaga-Dagbui and Smith 2012) employs Neuro-Fuzzy (NF) hybrid models - a combination of neural networks and fuzzy set theory, drawing on synergies from the two techniques in an attempt to develop more accurate, reliable and consistent final cost models. The next section of the paper provides an overview of the two modelling techniques used in the paper- neural networks and fuzzy set theory, and then proceeds to develop a neuro-fuzzy cost estimation hybrid model before concluding with results achieved and potential extensions of this research.

NEURAL NETWORKS

Work on artificial neural networks stemmed from the curiosity to understand how the brain processes information. Haykin (1994) described the brain as a highly complex and parallel information processing system, capable of performing very complex computations many times faster than many types of computer processors. Artificial neural network (ANN) is thus just a simplistic abstraction of the biological neural networks of the brain, endowed with the capability to learn from experience (or examples) and then generalise for new cases using the acquired knowledge even within sparse or incomplete data (Anderson 1995). They are able to adapt to changing environments (or datasets) and are often referred to as universal approximators because of their ability to closely map input to output spaces in different types of problem domains (Fausett 1994). They essentially seek underlying relationships between variables and are particularly suited for complex, hard-to-learn problems, where no formal underlying theories or classical mathematical and traditional procedures exist (Adeli 2001). Neural networks are very sophisticated modelling techniques capable of modelling extremely complex functions. In particular, neural networks are non-linear (Denton and Hung 1996). For many years linear modelling (Regression), has been the commonly used technique in most modelling domains since they have well-known optimization strategies. Where the linear approximation was not valid, which was frequently the case (Boussabaine and Kirkham 2008), the models suffered accordingly.

Arguably, the strongest argument against the use of ANN is its supposed ‘black-boxness’(Olden and Jackson 2002)- it is difficult to extract knowledge from the neural network model or fully understand how it reaches its conclusions. In regression, for example, an equation with explainable physical properties is produced. This is not the case in ANN modelling - no equation results out of the model and the network weights and connections make little sense. How the inputs interact to produce the output is at best, only known to the model. In a previous model using the same data, only neural network is used to model final cost projects (Ahiaga-Dagbui and Smith 2012). In an attempt to illuminate the black-box of ANNs, the authors combine the learning and generalisation abilities of neural networks with the capacity for tolerance and imprecise knowledge representation of fuzzy set theory to develop a hybrid neuro-fuzzy cost model for cost prediction.

FUZZY SET THEORY

Fuzzy set theory is an aspect of contemporary mathematics which focuses on the ambiguities in describing events or classes. It is an attempt to formalise human abilities of conversation, reasoning, and decision-making in an environment of imprecision, uncertainty as well as conflicting and/or incomplete information (Zadeh 2008). It incorporates ‘matter of degree’ rather than crisp boundaries into decision variables (Tokede and Wamuziri 2012). Fuzzy set theory allows an approximate interpolation between observed inputs and output situations (Ross 2009) and provides a means for modelling human vagueness in judgment. It basically requires encoding certain decision parameters as fuzzy sets (Zadeh 2008).

The defining characteristic of a fuzzy set is embodied in its membership function (MF). According to Kim *et al.* (2006), an MF provides an effective way to translate subjective terms into mathematical measures. A variable in fuzzy logic could have a set of values, characterised in linguistic terms, such as short, medium or long duration of project, or poor, moderate and good ground conditions. MFs can be generated in a number of ways either using intuition or some other algorithmic or logical operations (see Ross (2009) on how to use genetic algorithm, neural networks, rank ordering or inductive reasoning in developing MFs).

Ross (2009) stipulates that fuzzy relations are analogous to classical mathematical functions and basically represent mappings for sets. Fuzzy relations share the mapping potentials exhibited by neural networks and hence provide a compatible interphase in problem solving. Relations exhibit mathematical properties such as reflexivity, transitivity and symmetry which ultimately helps in interpreting attributes in fuzzy systems (Zadeh 1994). Chen and Huang (2007) used fuzzy relations in estimating the possibility-of-meeting the completion time of a construction project.

Fuzzy relations could be also employed in establishing the strength and possible association between different pairs. This can be achieved through the composition operator - a mathematical operation that seeks to establish the relationship between similar elements in different universe of discourse (Zimmermann 2001). Two common variants of the composition operator are the max-product and max-min. According to Zimmermann (2001), the most frequently used composition operator is the max-min; though both procedures produce comparable results in many instances. The max-min composition operation basically implements the strength of one chain as equal to the strength of its weakest link; the maximum of this then represents the overall chain strength in the fuzzy system (Ross 2009). Applications in civil engineering and construction research have been reported in Ayyub (1997). For cost and risk evaluation, fuzzy sets helps in quantification of variables, whose nature could be considered as complex and fit for description within a range of options (Tokede and Wamuziri 2012). An overview of fuzzy logic applications in construction management is provided by Chan *et al.* (2009)

NEURO-FUZZY

Neural networks solves problems by identifying the underlying patterns between the variables in the data it receives (Ross 2009) and then makes predictions based on the knowledge acquired (Adya and Collopy 1998). They are powerful, easy to use (StatSoft Inc. 2011) and can deal with large number of variables and non-linear relationships (Denton and Hung 1996). Yet, they are limited by their 'black-box' nature (Patterson 1996; Olden and Jackson 2002). They also perform best when using numerical or continuous data (StatSoft Inc. 2011). The majority of the data used in this research happen to be categorical in nature - location, type of client, procurement method, etc. Fuzzy sets represent composition of graded categories using mathematics based on logical reasoning (Belohlavek *et al.* 2009). It attempts to formalise decision making in an environment of uncertainty and incomplete information (Zadeh 2008), the kind that aptly describes cost estimation of construction projects.

Tokede and Wamuziri (2012) suggest that fuzzy set theory may not function at its optimal best as a stand-alone mathematical framework. Its practicality and utility is enhanced by combining its logic with pre-existent mathematical formulations. NF hybrid models thus have the potential to effectively represent modes of reasoning and decision making that are approximate rather than exact (Zadeh 1994), the case of construction cost estimation. Yu and Lin (2006) present an NF model for mining information from incomplete construction databases whilst Bilgehan (2010) uses NF models predict concrete compressive strength. Boussabaine (2001) similarly presents NF models for modelling the likely duration of construction projects

MODEL DEVELOPMENT

The NF models reported in this paper have been developed in three main stages - the first using statistical methods to pre-process the collected data, the second using neural networks to develop relative final cost weightings of predictors and lastly using fuzzy sets to predict final cost. These stages are detailed below.

Stage One: Data and Data Pre-processing

Details on 98 water infrastructure projects completed in Scotland between 2007 and 2011were collected. The nature of the projects ranged from construction of water mains, water treatment plants, Combined Sewer Overflows (CSOs), installation of manholes or water pumps and upgrades and repairs to sewers. All the projects were target cost contracts with values between £9,000-£14 million and durations from 1-22 months.

The collected data is processed so as to structure and present the data to the model in the most suitable way. For this research, extreme values and outliers were either re-coded or deleted from the sample set and missing values replaced with the mean or mode. Input errors were corrected and all cost values were normalised to 2010 with the base year 1995 using the infrastructure resources cost indices by the Building Cost Information Services (BCIS 2012). Screening of variables to the smallest number is desirable because simpler models are easier to deploy - a model with 15 variables means information has to be known about all these variables before the model can be used for prediction. Redundant predictors - variables that do not add new information to the model because they basically contain the same information at another level with other variables were detected using spearman ranking, bi-variate histograms or cross-tabulation. Further variable screening using scree test, mean plots and optimal binning in Statistica 10 software, suggested the optimal number of variables for predicting final cost to be between 5-7 predictors.

Stage Two: Neural Network Modelling

The neural network stage of the model developed was to determine a consistent numerical weighting for all the predictors depending on their relative contribution to determining the final cost of the project. Ten initial predictors[2] were used as inputs in a 3-layered feed-forward back-propagation neural network architecture with Final Target Cost as output of the model. The 98 project cases were split in a 75:15:10% ratio for training, testing and validation respectively. The best model was developed through an iterative procedure of continually tweaking the neural network parameters i.e. hidden nodes and activation functions, to produce improved model performance. Model performance was measured using the correlation coefficient between predicted and output values as well as the Sum of Squares (SOS) of errors below:

$$SOS = \sum (T_i - O_i)^2 \qquad \text{Eqn. 1}$$

Where Oi is the prediction (network outputs)
Ti is the target (actual value) of the ith data case.

The ten best networks were retained and further tested using the validation set to produce *Figure 2*. The validation set was not used in the training of the model so can be considered as an independent verification of the model's ability to generalise on new data. This gave a quick indication of the average error level of each of the models.

---

[2] Initial list of predictors for the neural network model: Type of Soil, Site Access, Type of Location, Contractor's Need for the Project, Frequency of Project, Type of Deadline, Awarded Target cost (transformed as logTC), Type of project, Tendering Strategy, Duration (transformed as logD)

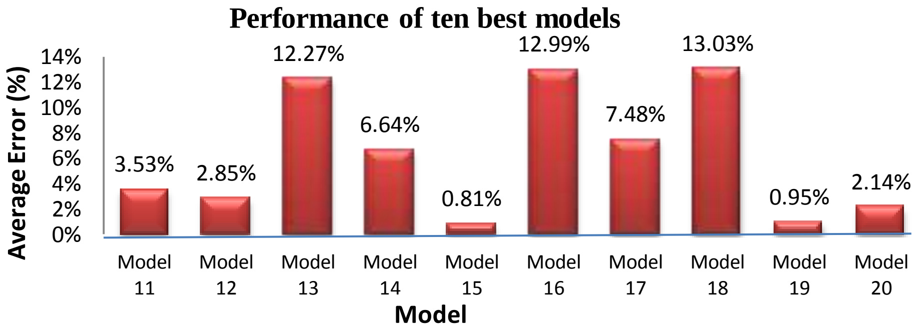

Figure 2: Performance of the ten best models

A sensitivity analysis was then carried out using the three best validated models in order to determine the contribution of each predictor to the model's performance. This was partly based on a test for parsimony using Ockham's Razor principle - one should not increase, beyond what is necessary, the number of entities required to explain anything and that all things being equal, preference should be given to the simplest hypothesis (Chase *et al.* 1996). This principle of simplicity is used to prune down the number of variables required in the model to predict the final cost, thus reducing inconsistencies, ambiguities and potential redundancies in the model. An initial ranking of all the predictors was generated based on their contribution to the model's performance. Then starting from the least important, one predictor was removed from the model at a time whilst measuring the performance of the model without that predictor. This was done until the model showed no further improvement or began to decay. The best set of predictors of final target cost after this stage are tendering strategy, site access, location, type of project, contractor's need for the project, type of soil, as well as estimated initial cost and duration (the common log of these were used in the model)

Table 1: Sensitivity analysis to determine relative ranking of predictors

Model	logTC	Tendering Strategy	Site Access	Type of Location	Project Type	Contractor's Need	Soil Type	logD
15. MLP 18-5-1	4.80	2.22	8.44	2.04	1.50	3.80	1.22	1.09
19. MLP 18-3-1	7.71	9.08	8.91	11.82	7.93	4.77	7.07	0.68
20. MLP 18-3-1	8.21	9.18	2.64	3.24	1.89	2.55	2.56	1.21
Average Weighting	6.90	6.83	6.66	5.70	3.77	3.71	3.61	0.99

Stage Three: Fuzzy Sets Modelling

Fuzzy set theory is applied at this stage of the modelling exercise to evaluate the subjective measures for each of the cost predictors in order to predict final cost. Using $\sum \ Normalized\ ranking = \frac{w_i}{\sum \ W}$=1 Eqn. 2. 2, the average weighted ranking for each of the variables from *Table* was normalized to unity in order to generate a standardised index for the subsequent fuzzy set computations (see *Table 4*)

$$\sum Normalized\ ranking = \frac{w_i}{\sum W} = 1 \qquad \text{Eqn. 2}$$

Where w*i* is the average relative weighting of the ith predictor
∑W is the sum of relative weighting of all predictors

Table 4: Normalized weighted values of the cost predictors from the neural network analysis

Factors	Tendering strategy	Site Access	Type of Location	Project Type	Contractor Need	Soil Type	Log Duration
Normalized ranking	0.22	0.21	0.18	0.12	0.12	0.11	0.04

With mean target cost to predictor plots, all predictors were fuzzified using the range set below:

$x \geq\ 5.8$,	Influence is Rather High
$5.6\ \geq\ x\ \geq 5.8$	Influence is High
$5.4\ \geq\ x\ \geq 5.6$	Influence is Medium
$x\ \ \leq 5.4$,	Influence is Low

The next stage of the fuzzy modelling involved developing membership functions. In developing these, the tolerance index is particularly relevant in evaluating and constraining the range of possibilities subject to a complex set of influencing variables, quantitatively and/or qualitatively defined. The tolerance index is vital in order to model the uncertainty in the cost values within a realistic continuum as opposed to a single figure-of-merit. For this study, the tolerances, β, were adapted to follow those indicated by Ayyub (1997) and reported in the table below.

Table 5: Values of tolerance. Source: adapted from Ayyub (1997)

β	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Poor/Low	1.0	0.9	0.7	0.4	0	0	0	0	0	0	0
Median	0	0	0.4	0.7	0.9	1.0	0.9	0.7	0.4	0	0
High	0	0	0	0	0	0	0	0.4	0.7	0.9	0
Rather High	0	0	0	0	0	0.4	0.7	0.9	1.0	0.9	0.7

Each of the project variables in the validation set was converted into fuzzy set variables using *Table 5*. According to Ross (2009), the fuzzy relation, $\tilde{T}$ of two sets, $\tilde{R}$ and $\tilde{S}$ can be defined by the set-theoretic and membership function-theoretic, mathematically expressed as:

$$\tilde{T} = \tilde{R} \circ \tilde{S} \qquad \text{Eqn. 3}$$

$$\tilde{\mu}_{T(x,z)} = \vee_{y \in Y}[\tilde{\mu}_{R(x,y)} \wedge \tilde{\mu}_{S(y,z)}] \qquad \text{Eqn. 4}$$

In Eqn. 3 above, R is a fuzzy relation on the Cartesian space X x Y. S is a fuzzy relation on Y x Z, and T is fuzzy relation on X x Z. In this cost estimation problem, R represents the set of cost predictors and S refers to the set of standard values of tolerance for linguistic descriptors of project attributes. The max-min composition operator is employed to deduce the strength and degree of relationship between specific relational pairs, which in this case, depicts the overall project cost as a fuzzy relationship of the normalised cost predictor weightings in *Table 4*, and based on the associated fuzzified project attributes deducible from *Table 5*.

The tolerance of each of the cost values in the validation set was computed, using Eqn.4 and defuzzified to obtain a 3-point estimate representing the fuzzy mean, fuzzy upper and fuzzy lower values as shown in *Table 6*. These three values provided a range of likely final cost rather than the customary single value estimate. *Table 6* shows the performance of the NF hybrid models in predicting the final cost of 10 different projects used in the validation set. This is summarised in

Table 7 along with the average model performance of the neural network model only.

The Fuzzy Upper best predicts the final cost and have the smallest percentage errors, ranging from 0.6% average underestimation to 0.8% overestimation of the likely final cost of the project. This represents an appreciable improvement in the results achieved using the neural network models only, also shown in

Table 7. The best three models at the neural network stage averaged a 1.2% under-estimation and 4.6% over-estimation of the actual final cost of the projects in the validation dataset. These results show significant promise in using neuro-fuzzy hybrid models to learn the underlying relationships between variables such as tendering strategy, site access, project location, type of soil or type of project and final cost of construction project.

Table 6: Neuro-fuzzy model validation results

		Model Prediction (log)					
Validation Cases	Actual Final Cost (log)	Fuzzy Lower (FL)	% error (FL)	Fuzzy Mean (FM)	% error (FM)	Fuzzy Upper (FU)	% error (FU)
1	5.78	5.65	2.4%	5.68	1.8%	5.75	0.5%
2	6.90	6.75	2.2%	6.77	1.9%	6.86	0.7%
3	5.41	5.35	1.1%	5.39	0.5%	5.46	-0.9%
4	5.22	5.09	2.6%	5.12	1.9%	5.20	0.5%
5	6.51	6.38	2.0%	6.41	1.6%	6.48	0.4%
6	5.95	5.85	1.7%	5.87	1.4%	5.95	-0.1%
7	6.91	6.78	1.9%	6.80	1.6%	6.89	0.4%
8	4.67	4.58	1.8%	4.62	1.1%	4.69	-0.5%
9	5.00	4.97	0.6%	4.99	0.1%	5.07	-1.6%
10	4.49	4.34	3.3%	4.36	2.9%	4.45	0.9%

Table 7: Summary of results from neuro-fuzzy model validation

	Summary of results			
	Neuro-fuzzy Lower (FL)	Neuro-fuzzy Mean (FM)	Neuro-fuzzy Upper (FU)	Neural Network Only
Average % Under-estimation	2%	1.50%	0.60%	1.2%
Average % Over-estimation	N/A	N/A	0.80%	4.6%

As already stated, even though it is agreeable that these factors affect the final cost on a project, it is difficult to assign cost measures to them as their relationship to cost are not thoroughly understood. The neuro-fuzzy hybrid models are possibly a step in the right direction in producing more accurate and realistic cost estimates at the initial stages of a construction project in an attempt to alleviate the problem of cost overruns

CONCLUSION

The research reported in this paper combines the learning and generalisation capabilities of artificial neural networks with fuzzy logic's ability to formalise human reasoning and decision making within an environment of uncertainty and incomplete information to develop neuro-fuzzy hybrid cost models for predicting the final cost of small water infrastructure projects. In particular, the research attempts to use some non-traditional cost predictors such as site access, location, tendering strategy, project and soil type to estimate likely final cost. The authors present a three-point range of possible likely final cost outcomes instead of the classical single point estimate. This might allow estimators and clients to more accurately estimate likely contingency needs for their projects. In their extended form, these models can readily be converted into stand-alone desktop applications that can allow quick simulation of what-if scenarios and also allow the easy generation of different cost estimates should project parameters change. As a sequel to a previous paper that used only neural networks, the results here show an improvement in the predictive performance and thus the results are now being extended to a database of 4500 projects with an industry partner.

REFERENCES

Adeli, H (2001) Neural networks in civil engineering: 1989-2000. *Computer-Aided Civil and Infrastructure Engineering*, **16**(2), 126-42.

Adya, M and Collopy, F (1998) How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting*, **17**(5-6), 481-95.

Agyakwa-Baah, A B (2009) *Risk Management in the Ghanaian Construction Industry* Master thesis Unpublished Thesis, Sheffield Hallam University, UK.

Ahiaga-Dagbui, D D and Smith, S D (2012) Neural networks for modelling the final target cost of water projects. *In:* Smith, S D (Ed.), *Procs 28th Annual ARCOM Conference*, 3-5 September 2012, Edinburgh, UK. Association of Researchers in Construction Management, 307-16.

Anderson, J A (1995) *An Introduction to Neural Networks*. Cambridge, Massachusetts: MIT Press.

Ayyub, B M (1997) *Uncertainty modeling and analysis in civil engineering*. CRC.

Baloyi, L and Bekker, M (2011) Causes of construction cost and time overruns: The 2010 FIFA World Cup stadia in South Africa. *Acta Structilia*, **18**(1).

BCIS (2012) BIS Construction Price and Cost Indices. In, http://www.bcis.co.uk: Building Cost Information Services, UK.

Belohlavek, R, Klir, G J, Lewis III, H W and Way, E C (2009) Concepts and fuzzy sets: Misunderstandings, misconceptions, and oversights. *International journal of approximate reasoning*, **51**(1), 23-34.

Bilgehan, M (2010) A comparative study for the concrete compressive strength estimation using neural network and neuro-fuzzy modelling approaches. *Nondestructive Testing and Evaluation*, **26**(1), 35-55.

Boussabaine, A H (2001) Neurofuzzy modelling of construction projects' duration I: principles. *Engineering Construction and Architectural Management*, **8**(2), 104-13.

Boussabaine, A H and Kirkham, R (2008) Artificial Neural Network Modeling Techniques for Applied Civil and Construction Engineering Research. *In:* Knight, A and Ruddock, L (Eds.), *Advanced research Methods in the Built Environment*. London: Wiley-Blackwell.

Chan, A P and Chan, A P (2004) Key performance indicators for measuring construction success. *Benchmarking: An International Journal*, **11**(2), 203-21.

Chan, A P, Chan, D W and Yeung, J F (2009) Overview of the application of "fuzzy techniques" in construction management research. *Journal of Construction Engineering and Management*, **135**(11), 1241-52.

Chase, S, Weiss, M, Gibbs, P, Hillman, C and Urban, N. *The Physics and Relativity FAQ* 1996 [cited 19th November, 2012. Available from http://math.ucr.edu/home/baez/physics/General/occam.html.

Denton, J W and Hung, M S (1996) A comparison of nonlinear optimization methods for supervised learning in multilayer feedforward neural networks. *European Journal of Operational Research*, **93**(2), 358-68.

Fausett, L V (1994) *Fundamentals of neural networks: architectures, algorithms, and applications*. Prentice-Hall Englewood Cliffs, NJ.

Flyvbjerg, B, Holm, M K and Buhl, S L (2002) Understanding costs in public works projects: Error or lie? *Journal of the American Planning Association*, **68**(279-295).

Flyvbjerg, B, Holm, M K S and Buhl, S (2004) What Causes Cost Overrun in Transport Infrastructure Projects? *Transport Reviews*, **24**(1), 3-18.

Gil, N and Lundrigan, C (2012) The Leadership and Governance of Megaprojects. In: *CID Technical Report No. 3/2012*: Centre for Infrastructure Development (CID), Manchester Business School, The University of Manchester, 18.

Haykin, S (1994) *Neural networks: a comprehensive foundation*. Prentice Hall PTR Upper Saddle River, NJ, USA.

Hegazy, T (2002) *Computer-based construction project management*. Upper Saddle River, NJ: Prentice Hall Inc.

Jackson, S (2002) Project cost overruns and risk management. *In:* Greenwood, D (Ed.), *Proceedings 18th Annual ARCOM Conference*, 2-4 September 2002 Newcastle, Northumbria University, UK. Association of Researchers in Construction Management, (Vol. 1) 99–108.

Kim, J, Lee, S, Hong, T and Han, S (2006) Activity vulnerability index for delay risk forecasting. *Canadian Journal of Civil Engineering*, **33**(10), 1261-70.

Love, P E D, Edwards, D J and Irani, Z (2011) Moving beyond optimism bias and strategic misrepresentation: An explanation for social infrastructure project cost overruns.

Odeck, J (2004) Cost overruns in road construction—what are their sizes and determinants? *Transport Policy*, **11**(1), 43-53.

Olden, J D and Jackson, D A (2002) Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, **154**(1-2), 135-50.

Patterson, D W (1996) *Artificial Neural Networks: Theory and Applications*. Singapore: Prentice Hall.

Ross, T J (2009) *Fuzzy logic with engineering applications*. 3ed. Chichester: John Wiley & Sons.

StatSoft Inc. (2011) Electronic Statistics Textbook. In, OK Tulsa: StatSoft, .

Tokede, O and Wamuziri, S (2012) Perceptions of fuzzy set theory in construction risk analysis. *In:* Simon, S (Ed.), *Procs 28th Annual ARCOM Conference*, 3-5 September 2012, Edinburgh, UK. Association of Researchers in Construction Management, 1197-207.

Yeung, J F, Chan, A P and Chan, D W (2008) Establishing quantitative indicators for measuring the partnering performance of construction projects in Hong Kong. *Construction Management and Economics*, **26**(3), 277-301.

Yu, W-d and Lin, H-w (2006) A VaFALCON neuro-fuzzy system for mining of incomplete construction databases. *Automation in Construction*, **15**(1), 20-32.

Zadeh, L A (1994) Fuzzy Logic, Neural Networks, and Soft Computing. *Communications of the ACM*, **37**(3), 77-84.

Zadeh, L A (2008) Is there a need for fuzzy logic? *Information Sciences*, **178**(13), 2751-79.

Zimmermann, H-J (2001) *Fuzzy set theory and its applications*. Springer.