



## PREDICTING FREEWAY PAVEMENT CONSTRUCTION COST USING A BACK-PROPAGATION NEURAL NETWORK: A CASE STUDY IN HENAN, CHINA

Jie He<sup>1</sup>✉, Zhiguo Qi<sup>2</sup>, Wen Hang<sup>3</sup>, Chihang Zhao<sup>4</sup>, Mark King<sup>5</sup>

<sup>1, 2, 3, 4</sup>Transportation College, Southeast University, Sipailou 2#, Nanjing, Jiangsu Province, 210018 P.R. China

<sup>1, 5</sup>Centre for Accident Research and Road Safety, Queensland University of Technology,  
130 Victoria Park Road, Kelvin Grove, QLD, 4059 Australia

E-mails: <sup>1</sup>hejie@seu.edu.cn; <sup>2</sup>176761442@qq.com; <sup>3</sup>czghw@126.com; <sup>4</sup>chihangzhao@seu.edu.cn;  
<sup>5</sup>mark.king@qut.edu.au

**Abstract.** The objective of this research was to develop a model to estimate future freeway pavement construction costs in Henan Province, China. A comprehensive set of factors contributing to the cost of freeway pavement construction were included in the model formulation. These factors comprehensively reflect the characteristics of region and topography and altitude variation, the cost of labour, material, and equipment, and time-related variables such as index numbers of labour prices, material prices and equipment prices. An Artificial Neural Network model using the Back-Propagation learning algorithm was developed to estimate the cost of freeway pavement construction. A total of 88 valid freeway cases were obtained from freeway construction projects let by the Henan Transportation Department during the period 1994–2007. Data from a random selection of 81 freeway cases were used to train the Neural Network model and the remaining data were used to test the performance of the Neural Network model. The tested model was used to predict freeway pavement construction costs in 2010 based on predictions of input values. In addition, this paper provides a suggested correction for the prediction of the value for the future freeway pavement construction costs. Since the change in future freeway pavement construction cost is affected by many factors, the predictions obtained by the proposed method, and therefore the model, will need to be tested once actual data are obtained.

**Keywords:** pavement, construction cost, Neural Network, predicting model, Matlab.

### 1. Introduction

Henan Province is located in the centre of China. Its area is 167 000 km<sup>2</sup>, it has a population of 99 mln and contains 18 cities ranging in size from 1.5 mln to 11 mln people. In 1994, Henan's first freeway was built from Zhengzhou to Kaifeng. In 2007, the total length of the freeways in Henan Province was 4556 km, and by 2020, it is expected that the total length will reach 6280 km.

In the past 10 years, the freeway network in China has developed very quickly and the total investment has been huge. However, in some freeway projects, the final construction cost is higher than the estimated cost at the detailed design stage, which in turn is higher than the conceptual cost at the preliminary design stage. In the context of government financial accountability practices in China this presents challenges; any deviation is likely to be queried, and the Secretary of the Provincial Transportation Dept or a senior official in the department will often

have to defend the increased costs publicly or in the state legislature. As a result, the legislature and the public will have perceptions of incompetence and erosion. A more accurate cost estimation process for freeways would therefore contribute to greater public and government confidence in the operation of infrastructure planning and development agencies, as well as contributing to more efficient budget processes.

Researches have indicated that project definition in the early planning process is an important factor leading to project success (Le *et al.* 2010; Scott-Young, Samson 2008; Thomas, Fernández 2008). To prepare reliable budgets for freeway construction programs, road authorities must have accurate estimates of future funding allocations they are likely to receive, and future project costs for long term infrastructure programs. While future funding is obviously never known with a great deal of certainty, it is more often the inaccurate estimation of project

costs that causes greater disruption to the execution of construction programs.

Various critical factors must be identified to estimate construction costs effectively. Several studies have set out to identify relevant factors, ranging from generic management and financial factors through to those that are specific to the industry under consideration. Stoy *et al.* (2008) identified quantitative cost factors such as absolute size, construction duration, and compactness as influence factors for good bidding information. Liu *et al.* (2011) found uncertain factors such as meteorological factor has a great uncertainty in the construction schedule of hydropower construction.

Pinto and Mantel (1990) identified the ten critical factors such as project scope, management goals, time planning and management, communication with owner, etc. In a study conducted in Newfoundland, Hegazy and Ayed (1998) found that season, location, type of project, contract duration, and contract size had a significant impact on an individual contract cost. Wilmot and Cheng (2003) described future construction cost in terms of predicted index values based on forecasts of the price of construction labour, materials, and equipment and the expected contract characteristics and contract environments. In a building construction study conducted by Cheng *et al.* (2009b), ten key quantitative factors were identified in the planning stage of projects. Six were quantitative: floors underground, total floor area, floors aboveground, site area, the number of households and households in adjacent buildings; and four were qualitative: soil condition, seismic zone, interior decoration and electromechanical infrastructure. Thus, examination of the literature shows that a wide variety of factors have been found to influence construction costs.

Factors such as those described above have been used in models of construction costs, but the models rarely attempt to use a comprehensive set of factors. In part, this is a consequence of the methods used for estimation. Shi, Li (2008) integrated rough sets (RS) theory and Artificial Neural Network (ANN) to forecast construction project cost. To overcome cost overruns in projects, some methods such as Probabilistic Simulation (Chou *et al.* 2009) and Support Vector Machine (Cheng *et al.* 2010; Chou 2011) have been used to develop appropriate cost models for predicting the expected project cost.

On the contrary, regression analysis represents a traditional approach (Khosrowshahi, Kaka 1996), an inherent disadvantage of which is its requirement of a defined mathematical form for cost functions, i.e. the nature of the relationships between variables must be assumed at the outset. In addition, such traditional methods of estimating project costs are hampered by the large number of important variables and the interactions between them. In addition, some of the variables that influence construction costs, such as the cost of labour, equipment, and materials, are usually highly correlated with each other, resulting in multicollinearity in the model when more than one of them is included. Thus, traditional methods are limited

in their potential applicability to the estimation of construction costs.

As a comparatively new method, Neural Network (NN) models have no implicit functional form and therefore have greater freedom to fit the data than do regression models. It is therefore possible that the greater flexibility in the relationship between input and output variables in NN might translate into a better model than that achieved with regression analysis. One purpose of the research reported in this paper is to use NN to identify a better model.

Some researchers have employed NN models to estimate the construction costs of individual projects (Ji *et al.* 2009). By combining NN and fuzzy logic, Boussabaine (1999), Boussebaine and Elhag (1999) developed neuro-fuzzy systems to estimate the construction cost and project duration of individual building projects. Wilmot and Mei (2005) developed a NN model to estimate highway construction cost escalation over time. Cheng *et al.* (2009b) developed an evolutionary fuzzy neural inference model to estimate costs at the concept stage. Ma *et al.* (2012) propose to modify the existing model (a single cost for cost-sensitive neural networks), the traditional back-propagation neural networks (TNN), by extending the back-propagation error equation for multiple cost decisions. Yip *et al.* (2014) presents a comparative study on the applications of general regression neural network (GRNN) models and conventional Box – Jenkins time series models to predict the maintenance cost of construction equipment.

Furthermore, hybrid models (combining NN and other approaches) have also been developed to estimate construction costs. Hegazy and Ayed (1998) used NN to develop a parametric cost estimating model for highway projects, with optimal NN weightings optimized by genetic algorithms. Kim *et al.* (2005) applied hybrid models of NN and genetic algorithms to residential building cost estimation in order to predict preliminary cost estimates.

These studies indicate that NN and NN hybrid models have been used instead of traditional methods to estimate the cost, duration, and other features of construction project costs, including highway construction projects. However, it is also clear from the limited literature that NN models have usually been used only for individual construction projects, rather than investigating the overall cost of construction across a range of projects, and examining how their cost alters over time. This approach has an inherent limitation, i.e. that the models developed are relevant only to the case studied, and will therefore not be readily generalizable to other projects. And the models discussed also lack relevance to similar projects undertaken some time later, as the model is specific to a particular time as well, whereas some of the important variables are changing over time in ways which they are modelled. The objective of this paper is to address these issues by developing a NN model based on a range of freeway pavement construction projects and taking temporal factors into consideration. In particular, this research will apply a back-propagation (BP) NN model to predict design cost estimates for

freeway pavement construction projects, using historical data on freeway construction projects in Henan Province as a case study of the application of the approach.

## 2. Influential factors analysis

The first step in developing the model is to identify which factors influence the costs of freeway pavement construction, so that they are considered for inclusion in the model. These factors have been categorized below as location, resource or time factors, though they also incorporate other variables, e.g. location is related to altitude and topography, both of which influence pavement construction costs in Henan. While the list of potential influencing factors is quite lengthy, a balance needs to be found, such that the number of factors is sufficient to provide adequate forecasts of costs, but not too large for practical application in a management setting. There is no clear guideline as to what the ideal number of factors should be. In this study, it was judged that the nine factors described below (two location, four resource and three time-related factors) should provide a more comprehensive basis for modelling and forecasting than has previously been the case, without creating disproportionate information needs.

### 2.1. Location factors

The freeway construction projects were located across Henan, which is characterized by differences in climate, geology, and topography that might be expected to have an influence on the cost of the projects. These characteristics tended to vary together, so that it was possible to define

just three regions based on climate, geology and topography. The region factor is given in Table 1.

An important practical issue for highway construction is the amount of variation in altitude along the road, as greater variation increases costs. This is related to topography, which is taken into account in a broad sense in the regional categories above, but the degree of variation between individual projects pointed to a need to develop categories at the project level. Variation in altitude was therefore divided into five categories from “very small”, which described roads that were essentially flat, to variations of between 450 m and 800 m. The variation in altitude factor categories (B1 to B5) are listed in Table 2, along with an indication of where the projects for each category took place, the range of absolute altitudes which applied there, and the freeway contracts which fell into these categories.

### 2.2. Resource factors

Labour, material and equipment are the main resources for a construction project. For simplicity, this study randomly selected five cases as an example to illustrate pavement construction cost components as shown in Table 3. Material costs constituted nearly 85% of pavement construction costs and the equipment costs constituted nearly 12%.

A construction project usually requires more than 100 types of material. The components of pavement material costs of the five cases are shown in Table 4, with the largest four components listed separately. Taken together, the two largest components (concrete and asphalt costs and stone costs) accounted for approximately 76% of material costs.

**Table 1.** Region factor

Region list	Corresponding city	Average pavement cost, \$/km
A1 West Henan	Sanmenxia, Luoyang, Jiyuan, Jiaozuo	692 998
A2 South Henan	Xinyang, Nanyang, Zhumadian	956 815
A3 East Henan and North Henan	Anyang, Puyang, Hebi, Xinxiang, Zhengzhou, Kaifeng Xuchang, Shangqiu, Luohe, Zhoukou, Pingdingshan	980 169

**Table 2.** Altitude factor

	Location	Altitude, m	Variation in altitude	Corresponding freeways and contracts
B1	East Henan Plain	0~50	Very small	15 contracts from Shang-Zhou Freeway, Kai-Tong Freeway
B2	East Henan Plain	0~100	≤50 m	2 contracts from Shang-Ze Freeway, 6 contracts from Yong-Bo Freeway, 4 contracts from Xu-Bo Freeway, Xin-Chang Freeway, Chang-Feng Freeway, Xi-Zhou Freeway, Zhou-Jia Freeway, Hu-Xin Freeway, Pu-Fan Freeway
B3	Nanyang Basin	100~200	≤100 m	2 contracts from Nan-Deng Freeway, 11 contracts from Da-Guang Freeway, 13 contracts from Feng-Nan Freeway, 2 contracts from Luo-Ping Freeway, Mi-Nan Freeway, An-Nan Freeway, Ping-Zheng Freeway, Xin-Zhu Freeway, Ji-Jiao Freeway, Jiao-Xiu Freeway, Zhu-Mi Freeway
	Plain	0~100	≤100 m	
	West Henan Plain	50~100	≤100 m	
B4	Plain and Mountain	50~500	≤450 m	8 contracts from Da-Xin Freeway, King-Mi Freeway, Ji-Jin Freeway, Yu-Deng Freeway, Mi-Tong Freeway
B5	Mountain	200~1000	≤800 m	Ji-Feng Freeway, Zheng-Shi Freeway, 15 contracts from Feng-Nan Freeway

**Table 3.** Components of pavement cost of five cases

Cost category	Shang-Zhou 11	Xu-Bo (Fugou)	Da-Guang 6	Ming-Tong	Zhou-Jia	Average
Labour costs	3.59	2.27	1.33	2.87	2.44	2.50
Material costs	81.87	85.30	86.54	85.90	83.83	84.69
Equipment costs	13.43	11.11	10.37	11.07	12.52	11.70
Other costs	1.11	1.33	1.76	0.15	1.21	1.11

**Table 4.** Components of pavement material costs for five cases

Cost category	Shang-Zhou 11	Xu-Bo (Fugou)	Da-Guang 6	Ming-Tong	Zhou-Jia	Average
Stone costs	38.21	23.1	24.84	32.21	21.86	28.04
Concrete and asphalt costs	41.71	52.73	43.86	45.98	54.45	47.75
Water and power costs	1.38	0.59	0.42	2.43	1.79	1.32
Oil costs	6.09	5.35	2.52	3.47	4.97	4.48
Other costs	12.61	18.23	28.36	15.91	16.93	18.41

For simplicity, the authors proposed to use the costs of crushed stone (diameter 4 cm) as a proxy for stone costs. The quantity of stone category material is written in the following form:

$$N_{stone} = \sum_{i=1}^n N_i \cdot \frac{P_{io}}{P_{Crushed\ Stone(4\ cm)o}}, \quad (1)$$

where  $N_{stone}$  – the quantity of stone category material;  $N_i$  – the quantity of material  $i$ ;  $P_{io}$  – the bid price of material  $i$  at time  $o$ ;  $P_{Crushed\ Stone(4\ cm)o}$  – the crushed stone price at time  $o$ . This equation provides equivalence between all other stone materials and crushed stone. This allows the quantity of other stone materials to be included in the evaluation by linking it to an equivalent amount of additional crushed stone.

A similar process is used to quantify an equivalent amount of Calx (calcium oxide) as a proxy for concrete and asphalt material costs. The quantity of concrete and asphalt category material is written in the following form:

$$N_{Concrete\ And\ Asphalt} = \sum_{j=1}^n N_j \cdot \frac{P_{jo}}{P_{calxo}}, \quad (2)$$

where  $N_{Concrete\ And\ Asphalt}$  – the quantity of concrete and asphalt category material;  $N_j$  – the quantity of material  $j$ ;  $P_{jo}$  – the bid price of material  $j$  at time  $o$ ;  $P_{calxo}$  – the calx price at time  $o$ .

**2.3. Time-related factors**

Freeway construction costs change over time because the Index Number of Prices (INP) always changes (Levinson, Gillen 1997; Wilmot, Cheng 2003; Wilmot, Mei 2005; Tawfek *et al.* 2012). Three indices were developed to reflect how construction costs change as a result of changes in the

INP. They were Index Number of Labour Prices (INLP), Index Number of Material Prices (INMP), and Index Number of Equipment Prices (INEP).

First, INLP is written in the following form:

$$\left\{ \begin{aligned} INLP_K^L &= \sum_{i=1}^n W_{ik}^L \cdot \frac{p_{ik}^L}{p_{io}^L} \cdot 100 \\ W_{ik}^L &= \left( \frac{M_{ik}^L \cdot p_{ik}^L}{\sum_{i=1}^n M_{ik}^L \cdot p_{ik}^L} \right) \end{aligned} \right. , \quad (3)$$

where  $INLP_K^L$  – the INLP at time  $k$  based on the INLP at time  $o$ ;  $W_{ik}^L$  – the proportion of labour item  $i$  at time  $k$ , and for each time  $k$ ,  $\sum_{i=1}^n W_{ik}^L = 1$ ;  $M_{ik}^L$  – the quantity of labour item  $i$  at time  $k$ ;  $n$  – the number of main labour items;  $p_{ik}^L$  – the price of labour item  $i$  at time  $k$ ;  $p_{io}^L$  – the price of labour item  $i$  at time  $o$ .

Similarly, INMP is written as:

$$\left\{ \begin{aligned} INMP_K^M &= \sum_{i=1}^n W_{ik}^M \cdot \frac{p_{ik}^M}{p_{io}^M} \cdot 100 \\ W_{ik}^M &= \left( \frac{M_{ik}^M \cdot p_{ik}^M}{\sum_{i=1}^n M_{ik}^M \cdot p_{ik}^M} \right) \end{aligned} \right. , \quad (4)$$

where  $INMP_K^M$  – the INMP at time  $k$  based on the INMP at time  $o$ ;  $W_{ik}^M$  – the proportion of material item  $i$  at time

$k$ , and for each time  $k$ ,  $\sum_{i=1}^n W_{ik}^M = 1$ ;  $M_{ik}^M$  – the quantity of material item  $i$  at time  $k$ ;  $n$  – the number of material items;  $p_{ik}^M$  – the price of material item  $i$  at time  $k$ ;  $p_{io}^M$  – the price of material item  $i$  at time  $o$ . In practice, this study did not base INMP calculations on all material items, selecting only major items which together accounted for more than 76% of all construction material costs.

The INEP has the same form:

$$\left\{ \begin{aligned} INEP_K^E &= \sum_{i=1}^n W_{ik}^E \cdot \frac{p_{ik}^E}{p_{io}^E} \cdot 100 \\ W_{ik}^E &= \left( \frac{M_{ik}^E \cdot p_{ik}^E}{\sum_{i=1}^n M_{ik}^E \cdot p_{ik}^E} \right) \end{aligned} \right. , \quad (5)$$

where  $INEP_K^E$  – the INEP at time  $k$  based on the INEP at time  $o$ ;  $W_{ik}^E$  – the proportion of equipment item  $i$  at time  $k$ , and for each time  $k$ ,  $\sum_{i=1}^n W_{ik}^E = 1$ ;  $M_{ik}^E$  – the quantity of equipment item  $i$  at time  $k$ ;  $n$  – the number of equipment items;  $p_{ik}^E$  – the price of using equipment item  $i$  at time  $k$ ;  $p_{io}^E$  – the price of using equipment item  $i$  at time  $o$ . As with the materials costs, this study did not base INEP calculations on all equipment items, selecting only major items which together accounted for more than 80% of all equipment costs.

### 3. Data for model development

Data were obtained for freeway construction projects contracted by the Henan Transportation Department during the period 1994–2007. Some nonstandard design and construction projects were removed from the data base. The

effective data consisted of contractual information on 88 projects, all of which were four lane divided carriageway freeways with 120 km/h speed limits. Pavement construction cost factors for a sample of the projects are shown in Table 5.

Eighty one of the 88 projects in the data set were used as a training data set, which was designed to comply with the following criteria for minimum size and proportion of total data set.

The minimum training set for the NN is written as follows:

$$N = n_1 \cdot n_2 \cdot \dots \cdot n_m, \quad (6)$$

where  $m$  – the number of the factors of a BP NN;  $n$  – the possible value of each factor (Shi 1995);  $N$  – the number of combinations of all possible values of the  $m$  parameters. The training sample set is considered incomplete in terms of solving the problem without a sample equal to or greater than  $N$ .

The problem in this paper has 9 factors. The location factor has three possible values; the altitude factor has five possible values; the labour cost per km is simple and does not involve categories, so it has only one possible value; the two largest components of the cost of materials are “concrete and asphalt” and “stone”, so it has two possible values; and the cost of equipment is simple and therefore has one possible value. The influence of “other costs” on the overall cost of pavement construction is not significant compared with other resource costs; it is assigned a value of 1 in our calculation. The possible values of the Index Number of Labour Prices (INLP) and Index Number of Equipment Prices (INEP), which correspond with the labour and equipment resource factors, are both 1. As the possible value of different types of materials has been taken into account in the materials resource costs, the value of the Index Number of Material Prices (INMP) is set at 1 to avoid recalculation.

**Table 5.** Construction cost factors of the pavement construction

The case	Location factors		Resources factors				Time-related factors		
	Region	Altitude	Labor, \$/km	$N_{stone}$ , \$/km	$N_{Concrete\ and\ asphalt}$ , \$/km	Equipment, \$/km	$INMP_K^M$	$INEP_K^E$	$INLP_K^L$
1 Shang-Zhou 1	3	1	1436.332	6121.730	2227.225	174.2843	342.028	121.670	105.119
2 Shang-Zhou 2	3	1	1399.976	6024.234	2256.014	170.0689	332.085	127.850	105.119
3 Shang-Zhou 3	3	1	1409.092	6131.177	2235.037	159.245	323.589	128.547	105.119
4 Shang-Zhou 4	3	1	1161.524	5039.263	1834.701	138.6546	315.328	127.917	105.119
5 Shang-Zhou 5	3	1	1471.174	6236.763	2254.881	178.4577	316.268	127.463	105.119
6 Shang-Zhou 6	3	1	1362.988	5944.144	2170.559	168.4096	308.263	127.943	105.119
7 Shang-Zhou 7	3	1	1820.468	6135.445	2134.949	185.7811	303.403	125.389	105.119
8 Shang-Zhou 8	3	1	1551.688	6132.593	2206.586	175.6292	309.491	127.163	105.119
9 Shang-Zhou 9	3	1	1327.206	5696.227	2080.068	155.0239	311.207	128.052	105.119
10 Shang-Zhou 10	3	1	1443.558	6174.912	2248.998	174.9466	296.499	127.814	105.119
11 Shang-Zhou 11	3	1	1146.895	4467.522	3913.307	128.5997	290.243	127.066	105.119
12 Shang-Zhou 12	3	1	1054.623	4208.745	1515.270	121.2685	284.558	127.204	105.119

From the discussion above, it can safely be concluded that the minimum size of the training set for the BP NN used in this paper was 30. In theory, the more training samples, the better, but in practice there are limitations on the number of road segments available. 88 samples are gathered; most researchers will select 90% of them as training samples and use the remainder for testing. This research selected 81 of the 88 samples as training samples and used the remaining 7 for testing.

**4. Artificial neural network models for construction cost estimation**

Artificial Neural Networks (NNs) were selected to model the pavement construction cost. ANNs are versatile because of their highly distributed parallel structures and adaptive learning processes (Cheng *et al.* 2009b; Raab *et al.* 2013; Šliupas, Bazaras 2013; Wilmot, Mei 2005). Of the many structures available for NNs, the multilayer feed-forward network was chosen for this study because such networks have the ability to deal with complex systems and yet are relatively easy to construct (Hegazy, Ayed 1998; Hunter *et al.* 2012; Ji *et al.* 2009). To train the model, the back-propagation (BP) learning algorithm was used because it has strong classification and generalization capabilities (Cheng *et al.* 2009a; Li, Chen 2012; Xiaokang, Mei 2010). The form of neural network used in this study is common in civil engineering applications.

In theory, a three layer BP network consisting of an input layer,  $n$  input variables are mapped to  $m$  target output variables in a hidden layer and an output layer. Therefore, the general form of the neural network models used in this study is represented as the simple three layers shown in Fig. 1.

The number of neurons in the hidden layer is difficult to ascertain and is normally found by experiment and experience.

The number of neurons in the hidden layer is directly related to the requirements of the problem and the number of neurons in the input or output layer. If the number is too small, there will be insufficient information acquired by the network to resolve the question; if there are too many neurons, it will increase the number of iterations of the network, thus extending the training time and reducing network generalization, thus decreasing predictive power.

First, the number of neurons in the hidden layer is determined using empirical formulae during the design of the network. Second, the network is trained using different neuron numbers. Finally, the optimal number of neurons is obtained by comparing the operating results. The general empirical formula used to determine the number of neurons in the hidden layer (Hirose *et al.* 1991; Sheela, Deepa 2013) is:

$$i = \sqrt{n + m} + a, \tag{7}$$

where  $i$  – the number of hidden neurons;  $n$  – the number of input neurons;  $m$  – the number of output neurons;  $a$  – a constant and  $1 < a < 10$ .

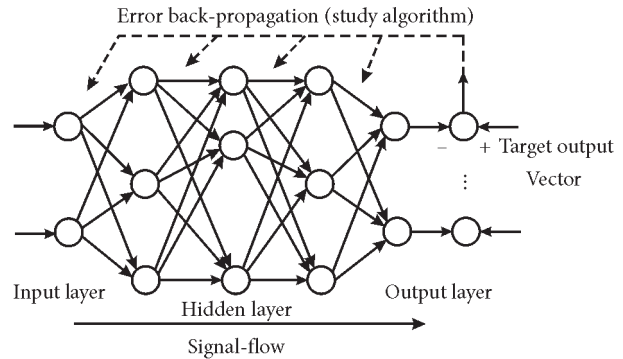


Fig. 1. Structure of three-layer NN BP network

According to Kolmogorov’s theorem, if the number of neurons in the input layer was  $n$ , then the number of neurons in the hidden layer is  $2n + 1$ .  $i$  is written as:

$$i = 2n + 1. \tag{8}$$

And  $i$  is written as:

$$i = \log_2^n, \tag{9}$$

where  $n$  – the number of input neurons;  $i$  – the number of hidden neurons.

In this study, the max and min number of hidden neurons ( $i_{max}$ ,  $i_{min}$ ) was determined by (7), (8) and (9), while training the network from the min to max increased the number of neurons by one. The optimal number of hidden neurons was selected by convergence data and training error using the operating results of different neurons number.

**4.1. Pavement construction cost model development**

Nine neurons were used in the input layer. These arose from the construction cost factors identified earlier and shown in Table 5 (region, variation in altitude, labour costs, stone costs, concrete and asphalt costs, and equipment costs, INLP, INMP and INEP).

The min and max determined by (7), (8) and (9) were 4 and 19. The training error and testing error that varied through different numbers of neurons are listed in Table 6. According to the changes of training step and training error listed in Table 6, the training error gradually decreased with the increase of the number of hidden layer neurons, but it rebounds when the number was 17 to 19. In summary, the optimal hidden neurons number was 16.

Only one neuron appeared in the output layer, representing pavement construction cost.

**4.2. MATLAB program**

The MATLAB software package was used to estimate the neural network models. The MATLAB training function for BP network has training functions `traingd`, `trainrp`, `traingdf`, `trainscg`, `trainlm`, `trainbr` and so on. Each has its own characteristics but no single function is adapted to the training process in all cases (Adeli, Wu 1998; Minli, Shanshan 2012). There are also many improved BP algorithms

**Table 6.** Model testing using different numbers of hidden neurons

Hidden neurons number	Training error	Training step	Testing error							Mean absolute
			1	2	3	4	5	6	7	
4	0.00999766	732	-0.2380	0.0054	0.1989	0.0654	0.0843	0.1078	0.1302	0.83
5	0.00999745	699	-0.1205	-0.0011	0.1798	0.0558	-0.0162	-0.0756	-0.1056	0.8336
6	0.00999479	567	-0.1730	0.0715	0.4289	0.1823	0.0850	0.1029	-0.1171	1.1607
7	0.00999827	594	-0.2297	-0.0133	0.2288	0.1299	0.0648	0.1309	-0.1327	0.9301
8	0.00995648	623	-0.1163	0.1055	0.3286	0.2708	0.0796	0.1345	-0.1178	1.1531
9	0.00998786	698	-0.2908	-0.0597	-0.0133	0.0321	0.0891	0.0812	-0.1020	0.7682
10	0.00997415	597	-0.1937	0.0537	0.3125	0.0433	0.0623	0.0367	-0.1193	0.8215
11	0.00993162	729	-0.2009	0.0288	0.0824	0.0588	0.1092	0.1357	-0.1255	0.7413
12	0.00998699	704	-0.2675	-0.0858	0.1634	-0.0681	0.1110	0.0849	-0.1086	0.8893
13	0.00998885	951	-0.1718	0.0482	0.2409	0.1448	0.0829	0.1091	-0.1380	0.8357
14	0.00999893	572	-0.1953	-0.0516	0.0472	-0.0897	0.0834	0.0494	-0.0832	0.7198
15	0.00999976	560	-0.1872	-0.0571	0.0489	-0.0033	0.1466	0.1387	-0.1021	0.6839
16	0.00998793	506	-0.1549	-0.0466	0.0747	-0.0148	0.1716	0.1404	-0.0957	0.6938
17	0.00996809	706	-0.1929	0.0917	0.3651	0.1035	0.0971	0.0788	-0.1214	1.0956
18	0.00999765	562	-0.3058	-0.0344	0.0517	-0.1001	0.1368	0.0906	-0.0939	0.8133
19	0.00997144	600	-0.2086	0.0086	0.1096	0.0094	0.0501	0.0618	-0.1136	1.0591

**Table 7.** Model testing using different training functions

Training function	Training step	Training error	Testing error						
Trainlm	8	0.0088614	-0.1085	0.0077	0.0029	-0.0481	-0.0164	0.1486	-0.1292
Trainscg	81	0.00999032	-0.0682	-0.0782	0.1663	-0.0040	0.0479	0.0161	-0.1156
Traincgf	71	0.00989892	-0.0224	0.0962	0.3613	0.0891	0.1148	0.0338	-0.0878
Trainbr	159	0.00999072	-2.5384e-4	4.1466e-5	4.9974e-4	-2.5216e-4	6.8190e-4	1.3440e-4	0.0103
Traingd	7000	0.032115	-	-	-	-	-	-	-
Trainrp	59	0.00995126	-0.0503	0.0596	0.0947	0.0144	0.1017	0.0990	-0.1365
Traingda	506	0.00999976	-0.0549	-0.0466	0.0747	-0.0148	0.1716	0.1404	-0.0957
Traingdm	7000	0.052164	-	-	-	-	-	-	-
Traingdx	716	0.00999408	-0.0715	0.0697	0.3900	0.1921	0.0883	0.1093	-0.1240

**Table 8.** Model testing using 10 runs of the trained NN

Run number	Shang-Zhou4	Shang-Zhou (SQ)02	Yong-Bo A4	An-Nan	Daguang-Xin8	Ji-Jin	Feng-Nan08
1	2.53	-0.012	-0.45	-0.069	-0.12	-0.30	-0.44
2	-2.49	-0.13	-0.40	-0.0069	-0.18	-0.0093	-0.52
3	-1.82	-0.33	-0.56	-0.039	0.19	-0.013	-2.66
4	-1.07	-0.75	-0.19	-0.015	0.69	-0.16	-0.86
5	-2.37	-1.33	-0.27	-0.034	0.57	-0.18	-0.82
6	3.33	-0.50	-0.23	0.017	-0.0051	-0.16	-2.87
7	-2.72	-0.66	-0.045	-0.052	-0.60	-0.027	-0.49
8	-1.93	0.28	0.091	-0.11	0.41	-0.22	-0.81
9	-2.25	-0.51	-0.66	-0.10	-0.077	-0.037	-0.60
10	-1.87	-0.78	0.16	-0.043	0.68	-0.13	-0.86
MAPE	2.24	0.53	0.31	0.048	0.35	0.12	1.09

such as the algorithm with adaptive study velocity and the additive momentum which is implemented using Matlab function ‘traingdx’, the gradient descent with momentum function that is implemented using ‘traingdm’, and the gradient descent adaptive function which is implemented by ‘traingda’ etc.

The training data set was used to map the input variable pattern to the target output pattern and minimize the error by adjusting the weights of the network links in an iterative process. Training was set to stop after 7000 iterations or until convergence of the root mean square error (RMSE) to a value less than 0.01.

Observing the changes of training step and training error obtained by different training functions determined the number of neurons in each layer of the network. The result showed that ‘trainbr’ was the best function, as its testing error was the minimum; even ‘trainlm’ and ‘trainrp’ had a small training step, but their testing error was relatively large; ‘trainscg’, ‘traingcf’ and ‘traingd’ showed much worse results; ‘traingdm’ and ‘traingd’ showed the worst results. In short, ‘trainbr’ is chosen as the training function for the network.

**4.3. Model testing**

A random selection of 81 freeway cases were used as a training data set for the neural network model and the remaining 7 freeway cases were used as a testing set on which the performance of the NN model was evaluated. The testing set projects were Shang-Zhou 4, Shang-Zhou(SQ)02, Yong-Bo A4, An-Nan, Daguang-Xin 8, Ji-Jin and Feng-Nan 08.

The NN model was programmed in MATLAB, with each run producing a slightly different result. The results of 10 runs on the testing set are listed in the Table 8. The statistical measure mean absolute percentage error (MAPE) was used to measure the performance of the models. The MAPE of the seven test cases varied from 0.048% to 2.24%, and the mean MAPE was 0.67%. The implications of this value for the accuracy of cost estimates are discussed below.

**5. Forecasting future costs**

**5.1. Predicting**

The model was used to forecast the change in future freeway pavement construction costs based on predictions of input values.

Input variables such as labour,  $N_{stone}$ ,  $N_{Concrete}$  and  $N_{Asphalt}$  and equipment costs utilized average values observed between 1994 and 2007. The next two variables, variation in altitude and region, were taken from Tables 1 and 2 based on the specific location and characteristics of each contract. The other three variables  $INMP_K^M$ ,  $INEP_K^E$ ,  $INLP_K^L$  and were calculated using (3), (4), and (5) and were based on forecasts of future GDP. The resulting values are listed in Table 9.

Using these values, the freeway pavement construction costs predicted by the model for 2010 are shown in Table 10.

**5.2. Accuracy of predicted costs**

As noted above, the MAPE is around 0.67%, which needs to be taken into account in the predictions made by the

**Table 9.** Input values to predict freeway construction costs in 2010

	Region	Variation in altitude	Labour, \$/km	$N_{stone}$ \$/km	$N_{Concrete}$ and $N_{Asphalt}$ \$/km	Equipment, \$/km	$INMP_K^M$	$INEP_K^E$	$INLP_K^L$
Shang-Zhou (SQ)02	3	1	1573.58	6091.74	3331.30	195.09	291.65	140.70	105.119
Daguang-Xin8	2	4	839.29	4316.35	2804.54	186.33	336.65	129.24	105.119
Ji-Jin	1	4	1008.03	4499.84	4568.21	159.78	194.47	145.55	105.119

**Table 10.** Predicted freeway pavement construction costs for 2010 (RMSE is 5.5%)

Run number	Shang-Zhou(SQ)02, \$/km	Daguang-Xin8, \$/km	Ji-Jin, \$/km
1	1 235 363	976 003	716 929
2	1 315 463	968 878	489 248
3	1 260 963	1 055 278	488 152
4	1 352 458	836 262	494 183
5	1 140 996	886 604	540 177
6	1 265 787	1 017 648	540 542
7	1 160 771	812 265	730 435
8	1 236 395	913 139	706 566
9	1 347 458	873 621	749 226
10	1 322 176	1 000 095	486 502
Average value	1 263 788	933 978	594 201



**Table 11.** Expected range of the construction costs, \$/km

The range of cost	Shang-Zhou(SQ)02	Daguang-Xin8	Ji-Jin
Forecast min cost	1 255 376	927 804	590 233
Forecast max cost	1 272 358	940 342	598 169

model. The MAPE is used to calculate an expected range within which the actual future cost is expected to fall.

MAPE is written in the following form:

$$MAPE = \frac{|C_{Forecast} - C_{Reality}|}{C_{Reality}} \quad (10)$$

Hence,  $C_{Reality}$  is expressed as

$$C_{Reality} = \frac{C_{Forecast}}{MAPE + 1} = \frac{1}{MAPE + 1} \cdot C_{Forecast} \quad (C_{Forecast} \geq C_{Reality}), \quad (11)$$

$$C_{Reality} = \frac{C_{Forecast}}{1 - MAPE} = \frac{1}{1 - MAPE} \cdot C_{Forecast} \quad (C_{Forecast} < C_{Reality}), \quad (12)$$

where  $\frac{1}{MAPE + 1}$  and  $\frac{1}{1 - MAPE}$  are the correction coefficients.

Therefore, the range of future freeway pavement construction costs is written in the following form:

$$\left[ \frac{C_{Forecast}}{1 + 0.0067}, \frac{C_{Forecast}}{1 - 0.0067} \right] \quad (13)$$

Using (13), the range of future freeway pavement construction costs in 2010 is shown in Table 11.

## 6. Discussion

The paper presents prediction of the construction cost of freeway pavement in Henan, China using an Artificial Neural Network. It seems to be informative and to provide accurate forecasts. However, the following issues need to be taken into account:

- there are more than 9 factors that influence the costs of freeway pavement construction – increasing the number of factors would give a more accurate model, but a greater sample size might also be needed;
- for NN, in theory, the more training samples the better, but in practice there are limits to numbers of road segments available;
- the price of product is changed suddenly due to the international economic situation;
- the model will need to be tested using actual data.

In short, there is still some distance to go in order to pursue this approach as an engineering application, but in the meantime it is useful for the Secretary of the Provincial

Transportation Department or a senior official in the department to adopt as a reference.

## 7. Conclusion

1. This paper has explored a new approach to the estimation of future freeway pavement construction costs by using a Neural Network trained with real data from 81 construction projects and incorporating a more comprehensive set of factors than is typically employed. Data were obtained from freeway construction projects let by the Henan Transportation Department during the period 1994–2007. The data consisted of information on 88 freeway contracts. Data from a random selection of 81 freeway cases were used to train a neural network model and the remaining data were used to test the performance of the Neural Network model. Finally, the likely range of pavement construction cost of three freeways in 2010 was predicted.

2. The factors used in the Neural Network model in this study reflect the characteristics of location (region – which incorporates differences in climate, geology and topography – and variation in altitude along the constructed road), resources costs (labour costs, and proxy costs for stone, concrete and asphalt, and equipment), and time-related changes dependent on indexation costs of labour, materials and equipment (Index Number of Prices, Index Number of Labour Prices and Index Number of Equipment Prices, respectively). Neural Network models have usually been used only for individual construction projects, rather than investigating the overall cost of construction across a range of projects, and examining how their cost alters over time, so that this approach represents a new way of addressing the problem of predicting future pavement construction costs.

3. A question which arises is the generalizability of the results, however this is relevant to the specific model derived rather than the process described. While the Neural Network was developed using data from Henan Province, the principal factors and the applicability of the Neural Network process are transferable to other locations. The nine factors used in this study will not all be applicable in another location, and this would need to be determined at the outset through consultation with the relevant agencies and experts.

4. The general form of the neural network model used in this study was three layers and for training, the Back-Propagation learning algorithm was used. In addition, the MATLAB® software package was used to estimate the neural network models, utilizing the training function 'trainbr' with characteristics of adaptive study velocity and the additive momentum method. Again, alternative approaches are tested and other software packages used.

5. One limitation which requires further testing relates to the success of the model's predictions in practice. This study has shown how to develop and train the model, and has tested how consistent its predictions were when applied to a different set of cases, but no attempt was made to test the accuracy of the predictions in practice. This requires a longer term study with greater amounts of data.

### Acknowledgments

The authors would like to thank National Natural Science Foundation of China for its financial support (Project No. 51078087), Education Department of Jiangsu Province (sponsored by Qing Lan Project), Henan Province Transportation Department for its financial support (project "Analysis of Freeway Construction Costs") and for providing essential data and support, and Professor Rod Troutbeck (Centre for Accident Research and Road Safety – Queensland (CARRS-Q), Queensland University of Technology) for grammar edit. Their assistance is gratefully acknowledged.

### References

- Adeli, H.; Wu, M. 1998. Regularization Neural Network for Construction Cost Estimation, *Journal of Construction Engineering and Management* 124(1): 18–24.  
[http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(1998\)124:1\(18\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(1998)124:1(18))
- Boussabaine, A. H. 1999. *Neurofuzzy Modeling of Projects Duration*. Research Report No. 4. School of Architecture and Building Engineering, University of Liverpool, Liverpool, UK, 1–39.
- Boussabaine, A. H.; Elhag, T. M. S. 1999. *Tender Price Estimation Using ANN Methods*. Research Report No. 3, School of Architecture and Building Engineering, University of Liverpool, Liverpool, UK, 1–57.
- Cheng, M.-Y.; Peng, H.-S.; Wu, Y.-W.; Chen, T.-L. 2010. Estimate at Completion for Construction Projects Using Evolutionary Support Vector Machine Inference Model, *Automation in Construction* 19(5): 619–629.  
<http://dx.doi.org/10.1016/j.autcon.2010.02.008>
- Cheng, M. -Y.; Tsai, H.-C.; Sudjono, E. 2009a. Evolutionary Fuzzy Hybrid Neural Network for Conceptual Cost Estimates in Construction Projects, in *Proc. of the 26<sup>th</sup> ISARC, USA*, 512–519.
- Cheng, M. -Y.; Tsai, H.-C.; Hsieh, W.-S. 2009b. Web-Based Conceptual Cost Estimates for Construction Projects Using Evolutionary Fuzzy Neural Inference Model, *Automation in Construction* 18(2): 164–172.  
<http://dx.doi.org/10.1016/j.autcon.2008.07.001>
- Chou, J. -S. 2011. Cost Simulation in an Item-Based Project Involving Construction Engineering and Management, *International Journal of Project Management* 29(6): 706–717.  
<http://dx.doi.org/10.1016/j.ijproman.2010.07.010>
- Chou, J.-S.; Yang, I.-T.; Chong, W. K. 2009. Probabilistic Simulation for Developing Likelihood Distribution of Engineering Project Cost, *Automation in Construction* 18(5): 570–577.  
<http://dx.doi.org/10.1016/j.autcon.2008.12.001>
- Hegazy, T.; Ayed, A. 1998. Neural Network Model for Parametric Cost Estimation of Highway Projects, *Journal of Construction Engineering and Management* 124(3): 210–218.  
[http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(1998\)124:3\(210\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(1998)124:3(210))
- Hirose, Y.; Yamashita, K.; Hijiya, S. 1991. Back-Propagation Algorithm which Varies the Number Of Hidden Units, *Neural Networks* 4(1): 61–66.  
[http://dx.doi.org/10.1016/0893-6080\(91\)90032-Z](http://dx.doi.org/10.1016/0893-6080(91)90032-Z)
- Hunter, D.; Hao, Y.; Pukish, M. S.; Kolbusz, J.; Wilamowski, B. M. 2012. Selection of Proper Neural Network Sizes and Architectures: a Comparative Study, *IEEE Transactions on Industrial Informatics* 8(2): 228–240.  
<http://dx.doi.org/10.1109/TII.2012.2187914>
- Ji, Z.; Li, Y. 2009. The Application of RBF Neural Network on Construction Cost Forecasting, *Knowledge Discovery and Data Mining*: 32–35. <http://dx.doi.org/10.1109/WKDD.2009.53>
- Khosrowshahi, F.; Kaka, A. P. 1996. Estimation of Project Total Cost and Duration for Housing Projects in the UK, *Building and Environment* 31(4): 373–383.  
[http://dx.doi.org/10.1016/0360-1323\(96\)00003-0](http://dx.doi.org/10.1016/0360-1323(96)00003-0)
- Kim, G. H.; Seo, D. S.; Kang, K. I. 2005. Hybrid Models of Neural Networks and Genetic Algorithms for Predicting Preliminary Cost Estimates, *Journal of Computing in Civil Engineering* 19(2): 208–211.  
[http://dx.doi.org/10.1061/\(ASCE\)0887-3801\(2005\)19:2\(208\)](http://dx.doi.org/10.1061/(ASCE)0887-3801(2005)19:2(208))
- Le, T.; Caldas, C. H.; Gibson, G. E.; Thole, M. 2010. Assessing Scope and Managing Risk in the Highway Project Development Process, *Journal of Construction Engineering and Management* 135(9): 900–910.  
[http://dx.doi.org/10.1061/\(ASCE\)CO.1943-7862.0000052](http://dx.doi.org/10.1061/(ASCE)CO.1943-7862.0000052)
- Levinson, D. M.; Gillen, D. 1997. The Full Cost of Intercity Highway Transportation, *Transportation Research – D* 3(4): 207–223. [http://dx.doi.org/10.1016/S1361-9209\(97\)00037-0](http://dx.doi.org/10.1016/S1361-9209(97)00037-0)
- Li, M.-S.; Chen, W.-C. 2012. Application of BP Neural Network Algorithm in Sustainable Development of Highway Construction Projects, *Physics Procedia* 25: 1212–1217.  
<http://dx.doi.org/10.1016/j.phpro.2012.03.222>
- Liu, X.; Wang, Z.; Yin, H. 2011. Cost Evaluation for Hydropower Construction Using Multiple Monte Carlo Simulation Method, *International Conference on Management and Service Science*: 1–4.  
<http://dx.doi.org/10.1109/ICMSS.2011.5998164>
- Ma, G.-Z.; Song, E.-M.; Hung, Ch.-Ch.; Li, S.; Huang, D.-Sh. 2012. Multiple Costs Based Decision Making with Back-Propagation Neural Networks, *Decision Support Systems* 52(3): 657–663.  
<http://dx.doi.org/10.1016/j.dss.2011.10.023>
- Minli, Z.; Shanshan, Q. 2012. Research on the Application of Artificial Neural Networks in Tender Offer for Construction Projects, *Physics Procedia* 24(C): 1781–1788.  
<http://dx.doi.org/10.1016/j.phpro.2012.02.262>
- Pinto, J. K.; Mantel, S. J.; Jr. 1990. The Causes of Project Failure, *IEEE Transactions on Engineering Management* 37(4): 269–276.  
<http://dx.doi.org/10.1109/17.62322>
- Raab, C.; Halim, A. E. A. O.; Partl, M. N. 2013. Utilisation of Artificial Neural Network for the Analysis of Interlayer Shear Properties, *The Baltic Journal of Road and Bridge Engineering* 8(2): 107–116. <http://dx.doi.org/10.3846/bjrbe.2013.14>
- Scott-Young, C.; Samson, D. 2008. Project Success and Project Team Management: Evidence from Capital Projects in the Process Industries, *Journal of Operations Management* 26(6): 749–766. <http://dx.doi.org/10.1016/j.jom.2007.10.006>

- Sheela, G. K.; Deepa, S. N. 2013. Review on Methods to Fix Number of Hidden Neurons in Neural Networks, *Mathematical Problems in Engineering* 2013, Article ID 425740, 11 p., 2013. <http://dx.doi.org/10.1155/2013/425740>
- Shi, H.; Li, W. 2008. The Integrated Methodology of Rough Set Theory and Artificial Neural-Network for Construction Project Cost Prediction, *Intelligent Information Technology Application* 60–64. <http://dx.doi.org/10.1109/IITA.2008.238>
- Shi, L. 1995. Training Sample Size of BP Neural Network, *North China Earthquake Sciences* 13(2): 26–30.
- Šliupas, T.; Bazaras, Z. 2013. Forecasting the Risk of Traffic Accidents by Using the Artificial Neural Networks, *The Baltic Journal of Road and Bridge Engineering* 8(4): –293. <http://dx.doi.org/10.3846/bjrbe.2013.37>
- Stoy, C.; Pollalis, S.; Schalcher, H.-R. 2008. Drivers for Cost Estimating in Early Design: Case Study of Residential Construction, *Journal of Construction Engineering and Management* 134(1): 32–39. [http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(2008\)134:1\(32\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(2008)134:1(32))
- Tawfek, H. S.; Mohammed, H. E. –D. H.; Razek, M. E. A. 2012. Assessment of the Expected Cost of Quality (COQ) in Construction Projects in Egypt Using Artificial Neural Network Model, *HBRC Journal* 8(2): 132–143. <http://dx.doi.org/10.1016/j.hbrcej.2012.09.009>
- Thomas, G.; Fernández, W. 2008. Success in IT Projects: a Matter of Definition?, *International Journal of Project Management* 26(7): 733–742. <http://dx.doi.org/10.1016/j.ijproman.2008.06.003>
- Wilmot, C. G.; Cheng, G. 2003. Estimating Future Highway Construction Costs, *Journal of Construction Engineering and Management* 129(3): 272–279. [http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(2003\)129:3\(272\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(2003)129:3(272))
- Wilmot, C. G.; Mei, B. 2005. Neural Network Modeling of Highway Construction Costs, *Journal of Construction Engineering and Management* 131(7): 765–771. [http://dx.doi.org/10.1061/\(ASCE\)0733-9364\(2005\)131:7\(765\)](http://dx.doi.org/10.1061/(ASCE)0733-9364(2005)131:7(765))
- Xiaokang, H.; Mei, L. 2010. Research on Construction Cost Control Based upon BP Neural Network and Theory of Constraint, *International Conference on Management and Service Science (MASS)*: 1–4. <http://dx.doi.org/10.1109/ICMSS.2010.5576622>
- Yip, H.-L.; Fan, H.-Q.; Chiang, Y.-H. 2014. Predicting the Maintenance Cost of Construction Equipment: Comparison between General Regression Neural Network and Box–Jenkins Time Series Models, *Automation in Construction* 38: 30–38. <http://dx.doi.org/10.1016/j.autcon.2013.10.024>

Received 20 December 2011; accepted 1 February 2012