Journal of
Marine Science and Technology

Volume 24 | Issue 2 Article 2

USING PRINCIPAL COMPONENT ANALYSIS WITH A BACK-PROPAGATION NEURAL NETWORK TO PREDICT INDUSTRIAL BUILDING CONSTRUCTION DURATION

Sou-Sen Leu
Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, R.O.C

Chi-Min Liu
Construction Engineering division of Formosa Plastic Group, Tunhwa N. Rd., Songshan District, Taipei City, Taiwan, R.O.C., d9305101@mail.ntust.edu.tw

Recommended Citation
Leu, Sou-Sen and Liu, Chi-Min (2016) "USING PRINCIPAL COMPONENT ANALYSIS WITH A BACK-PROPAGATION NEURAL NETWORK TO PREDICT INDUSTRIAL BUILDING CONSTRUCTION DURATION," *Journal of Marine Science and Technology*: Vol. 24: Iss. 2, Article 2.
DOI: 10.6119/JMST-015-0325-2
Available at: https://jmstt.ntou.edu.tw/journal/vol24/iss2/2

Journal of Marine Science and Technology, Vol. 24, No. 2, pp. 82-90 (2016)
DOI: 10.6119/JMST-015-0325-2

USING PRINCIPAL COMPONENT ANALYSIS WITH A BACK-PROPAGATION NEURAL NETWORK TO PREDICT INDUSTRIAL BUILDING CONSTRUCTION DURATION

Sou-Sen Leu[1] and Chi-Min Liu[2]

Key words: construction duration, prediction, principal component analysis (PCA), artificial neural network (ANN).

ABSTRACT

Industrial businesses must respond efficiently to market demands; therefore, industrial construction must accurately predict the project duration at the pre-investment stage. In practice, project duration predictions rely on the experience of project managers. To provide impartial expertise and quantitative estimate the predicted duration of constructing an industrial building, an extensive history of industrial building cases were collected to form a database. Principal component analysis was applied to the database to identify key factors to serve as input data for a back-propagation neural network (BP-NN) that was used to estimate the project duration. Three prediction models were identified and developed separately based on the total cost for large, medium, and small construction projects. The derived BP-NN prediction models are applicable for estimating construction duration during the initial stages of a project.

I. INTRODUCTION

Because the industrial marketplace is subject to rapid change of new competition, an accurate and expedient forecast of the amount of time required to construct a building is critical because it enables business to remain competitive. For example, the building construction cost of 12-in fly ash brick is only 5% of the total project cost—the remaining 95% of the cost includes equipment, installation, test runs, operation, and other factors. In addition, monthly sales constitute approximately four times the cost of constructing a building, creating an even greater incentive to complete construction on schedule. Ultimately, time is the principal concern of an industrial construction project. This study proposes a methodology for predicting the duration of industrial building construction projects that involves using principal component analysis (PCA), a back-propagation neural network (BP-NN), and a database containing 50 years of records of petrochemical industrial construction in Taiwan. The research scope of this study was limited to predicting industrial building construction duration, and the time requirements of equipment purchases, installation, test runs, and operation were excluded from the analysis.

II. PREDICTION ON PROJECT DURATION

Duration prediction has been extensively studied in numerous fields including management science (Yang et al., 2003), security inspection (Ding et al., 2003), medical research (Kelly, 2002), trade analysis (Goulielmos and Siropoulou, 2006; Huang et al., 2010), natural events (Monton and Kierland, 2006), and supplier selection (Jaskowski et al., 2010; Lam et al., 2010). Prediction methods can be classified into two categories: bottom-up methods and top-down approach methods. Table 1 shows a comparison between these methods. Bottom-up methods involve considering orders, resources, and the duration of each task in a construction project. To apply bottom-up methods, a skilled engineer's experience and attention to detail regarding the design are required for an accurate schedule prediction. Bromilow (1969) indicated that only 12.5% of cases are completed on schedule, 40% are completed late, and 47.5% are completed before scheduled. Factors of uncertainty that can affect construction duration include the engineer's experience, a contractor's skill level, weather, economic conditions, price changes, and project alterations; additionally, a detailed design requires a substantial amount of time to prepare. Despite these uncertainties, accurately estimating construction duration is still crucial during the early stage of a project in professional practice.

Paper submitted 03/28/14; revised 02/04/15; accepted 03/25/15. Author for correspondence: Chi-Min Liu (e-mail: d9305101@mail.ntust.edu.tw).
[1] *Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, R.O.C.*
[2] *Construction Engineering division of Formosa Plastic Group, Tunhwa N. Rd., Songshan District, Taipei City, Taiwan, R.O.C.*

Table 1. Comparison between duration prediction methods.

Bottom-up Methods		Top-down Methods
Scheduling Methods	Empirical Methods	Statistical Methods
Scheduling methods used in Program Evaluation and Review Technique (PERT): MS_Project, P3, Gantt chart, Critical Path Method, Probability Network Evaluation Technique (PNET), etc.	Risk analysis (considering probability distribution and cumulative probability) with uncertainty (Hui, 2003; Žujoet al., 2009)	Linear regression (Lam et al., 2010)
Monte Carlo Casimulation (Song et al., 2008)	DELPHI method (Liu, 2002)	Case-based reasoning (Yau et al., 1998)
Optimization techniques such as TABU Search (Zhang, 2002), Genetic Algorithms (Lin, 2003), etc.	Expert's option (Yao, 2002)	Construction Database (Lin, 2005)
		Forecast model (Zou et al., 2004)
Comparison		
1. Based on duration of each task in a construction project. 2. Restricted by the finished time of detail design and the experience of engineers. 3. The task order could be changed at the construction location. 4. The probability obtained from simulation runs may not be able to provide solid reference for decision making. 5. External conditions cannot be considered such as weather, political issues, material prices, constructors' skill level, etc.	1. Based on duration of each task. 2. Expert's opinion or risk parameters are guessed and cannot be verified.	1. Linear relationship could over-simplify the relationship between construction duration and related factors. 2. Database is not easy to build. 3. Database could incorporate unknown factors. 4. Consideration on duration using macro-scope viewpoint. 5. Provide reference data in the early stage of a construction project.

Top-down approaches start from the case study of construction projects, decompose the relative factors, and build a reliable model based on information in a history database. These approaches can be used to estimate construction duration directly by integrating known project information into an artificial intelligence algorithm. Statistical and heuristic methods have been widely applied in top-down approaches. Hojjat Adeli (2001) thoroughly reviewed artificial neural network (ANN) applications in civil engineering, structural engineering, and engineering management. Combining PCA and ANN to forecasting models has been studied in several fields (Jan, 2003; Ran et al., 2004; Wang et al., 2009; Ma et al., 2011), but not to industrial construction.

The current study adopted a top-down approach by combining PCA with ANN to estimate the duration of constructing industrial buildings. To develop a practical model, a database containing 50 years of history data was used. Various factors were analyzed, such as location, weather, price variation, the number of design changes, and contractor skill level. The proposed method enables directly estimating construction duration at the early stage of an industrial construction project. Although the proposed method cannot entirely replace the detailed planning involved in estimating a construction period, the results can serve as a critical reference for signing con- tracts and managing operational strategies.

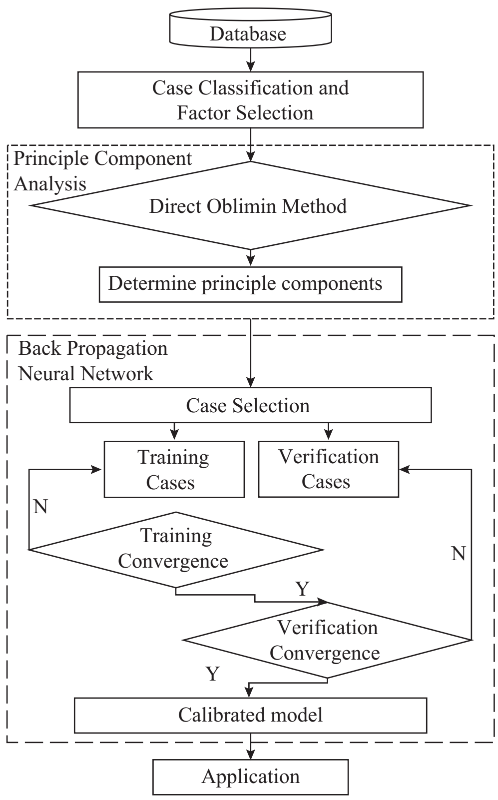


Fig. 1. Flowchart.

III. METHODOLOGY

Fig. 1 presents a flowchart depicting the methodology employed in this study. First, a set of target cases and factor selection criteria were selected, which are explained in the subsequent section. In PCA, direct oblimin rotation was used to determine the critical principle components of the selected

Table 2. Basic information of the selected cases.

Type \ Contractors	Civil engineering contractors	Level A construction company	Level B construction company	Level C construction company	Engineering material company	Miscellaneous Company	Consulting company	Engineering company	Total
AF_Decoration works style	18	4	10	1	7	1	3	14	58
AH_Steel Structure works style	62	11	146	6	20	44	12	37	338
AS_Reinforce Concrete structure works style	34	3	122	11	71	31	9	66	347
AT_Miscellaneous works style	11	0	11	2	11	4	7	19	65
AU_the utilities work style	1	12	0	1	2	1	0	1	18
CB_Harbor and bridge works style	40	47	13	31	6	19	5	13	174
CF_RC Foundation works style	1		76	4	2	2	2	89	176
CG_Ground preparation, backfilling and road pavement works style	23	60	25	14	36	24	20	31	233
HM_Machinery and Electrical construction work style	74	1	2	26	0	0	1	25	129
Total	264	138	405	96	155	126	59	295	1538

Table 3. Introduction of selected factors.

Direct Factor		
Category	Factor	Instruction
Case type	Work Type (WT) (Jan, 2003; Lam et al., 2010)	9 work types[1]
	Contract Type (CT) (Jan, 2003)	7 contract types[2]
Participant	Contractor Level (CL) (Liu, 2002; Jan, 2003; Jaskowski et al., 2010; Cheng et al., 2013)	The level is classified using company type (such as Inc., Ltd, etc.) and the skill level approved by the government [3]
	Establishment Year (EY)	The founded year of the company
	Capital (Ca) (Jan, 2003)	The founded capital (NTD)
	Number of Stuffs (NS)	Total number of people/works available for the contractor
	Recent Revenue (RR) (Jan, 2003)	The revenue of the contractor in recent years (millions of NTD/Year)
	Inspector (ID) (Jan, 2003; Lam et al., 2010)	Personal inspector which relates the construction duration to the person who can control the quality of the construction
	Supervisor (Su) (Liu, 2002)	18 organizations of supervisors
Location (site)	Location (Lo) (Jan, 2003)	The construction location[4]
	Project Effective Year (PEY)	The starting year of the construction project
Time	Project Effective Date (PED)	The starting time of the construction project (YY/MM/DD)
	Project Due Date (PDD)	The ending time of the construction project (YY/MM/DD)
	Number of Design Change (NC) (Liu, 2002)	The number of design changes required by owners (times)
Derived Factors		
Category	Factor	Instruction
Participant	Seniority (Se) (Liu et al., 2012)	The experienced year of the contractor which can be expressed as (PED-EY)
Time	Start Season (SS) (Jan, 2003)	The season when the project started. Four seasons: spring (Jan.-Mar.), summer (Apr.-Jun.), Fall (Jul.-Sep.), and Winter (Oct.-Dec.)
	Work Difficulty (WD) (Liu, 2002; Jan, 2003; Lam et al., 2010)	WD = Year (PED-FY)/3 WD is used to indicate the management rules that are getting more and more restricted.
	Price Index (PI) (Lam et al., 2010)	PI is provided by government. The PI value equals to 1.000 for the reference year 1991.
	Duration (Du)	Construction duration using 0.5 year as an interval.

1. Work Type (WT) include AF_Remodeling, AH_Steel works, AS_Structure Engineering, AT_Miscellaneous construction, AU_Utility works, CB_Harbor and bridge, CF_Foundation, CG_Ground & road, HM_Machinery.
2. Seven contract types are outsourcing-processing, outsourcing, turnkey, outsourcing appointment, outsourcing design, outsourcing design (type D), and procurement.
3. Contractor Level includes general company, grade A constructional company, grade B constructional company, grade C constructional company, pre-mix plant or material supplier, consulting company, engineer incorporated company (Inc.), limited company (Ltd.) and some other company (decoration, surveying, landscape gardening companies).
4. Locations include Taipei, Taichung, Tainan, Ilan, Linkou, Nantou, Taoyuan, Tarzan, Kaohsiung, Keelung, Mailiao, Chiayi, Changhua, Shulin, and Guanyin.

factors. The obtained principle components and construction duration records from the history database were used to train the BP-NN model. The objective function denotes the difference between the true value and estimated result. The stopping criteria of the algorithm were the number of iterations and mean square error (MSE). The construction duration is presented as the construction cost per day (NT$/day). This general representation is applicable to construction projects of various scales.

1. Database

The database comprised more than 20,000 cases of industrial building construction. The selection criteria were that the construction duration was longer than 6 months and no missing data. After filtering the construction data according to these requirements, 1,538 cases were identified. Table 2 shows the construction types and contractor classification of these cases.

Although numerous factors may affect construction duration, this study first classified a set of major categories, and then selected the corresponding representative factors from each category. All of the selected factors were quantified to facilitate conducting a scientific analysis and developing a forecasting model.

Recent studies (Lin, 2005) have indicated that critical factors include constructability, workspace acquisition, learning curve, weather, supervision efficiency, building type, contract systems, management effectiveness, district environment, and financial issues. The present study classified the factors into four categories: case type, participant, location, and time. Each category contains additional descriptive factors that could be directly obtained from the studied database. By contrast, factors requiring further calculation or were obtained from another database (e.g., economic indices) were regarded as derived factors. Table 3 lists these factors and their definitions. All of these factors were subjected to PCA and then applied to the BP-NN model for training.

2. Construction Cost

Although construction costs are strongly related to construction duration (Fig. 2), they are not considered a factor because different construction cost levels involve distinct relationships between the factors and construction duration. To account for the influence of construction costs, the factors were classified into three categories based on the project scale (large-, medium-, and small-scale construction projects), and the training was performed separately for each model.

(1) Large-scale: Construction cost over NT$50 million, with a construction duration of more than 42 months (n = 183).
(2) Medium-scale: Construction cost between NT$10 million and NT$50 million dollars, with a construction duration of 12 to 42 months (n = 399).
(3) Small-scale: Construction cost below NT$10 million, with a construction duration of 6 to 18 months (n = 956).

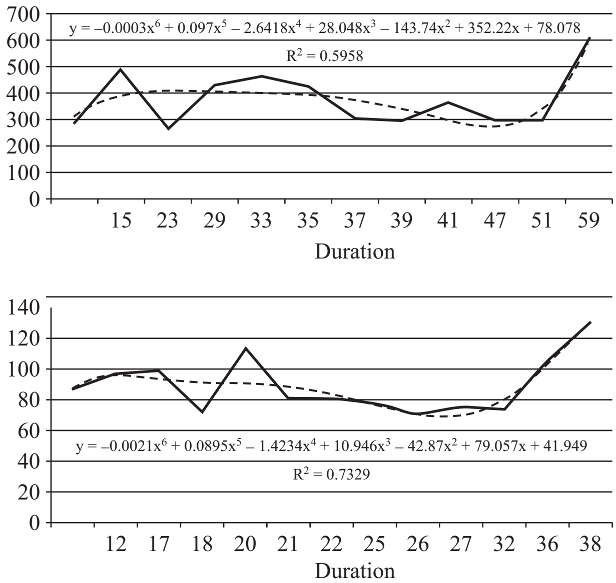
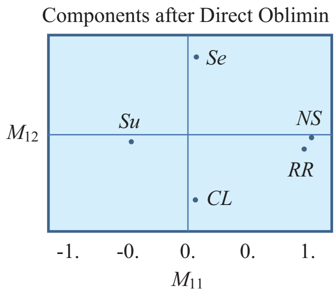


Fig. 2. Two examples of relationship between construction cost and duration.



Coefficient of components

	Compo-	
	M_{11}	M_{12}
Su	-0.227	-0.087
Se	0.067	0.594
NS	0.474	0.025
CL	-0.014	-0.591
RR	0.460	-0.064

Method: Component analys i
Rotation: Oblimin rotation with Kaiser criterion

Notes: *Su*, *Se*, *NS*, *CL* and *RR* are independent variables (x_i) of the fallowing formula (13) for a maximal value.

Fig. 3. An output example from PCA using SPSS 17.

3. Principal Component Analysis

PCA is a statistical method for converting potentially correlated variables of observation data into a set of linearly uncorrelated variables called principal components. The dimension of a principle component is equal to or less than the dimension of the original variable. The selected principle components can be used as input data for BP-NN model training. Because the original variables may be correlated, direct oblimin rotation is used for obtaining a non-orthogonal (oblique) solution, resulting in higher eigenvalues but diminished interpretability of the variables (Chen, 2005). In the present study, PCA was conducted using SPSS Version 17. Fig. 3 shows the output from PCA, wherein two principal components were obtained from analyzing five factors.

PCA was performed separately for each construction project scale.

A. Large-scale construction project: Three sets of principal components were selected.
(1) Two principal components, denoted as M_{11} and M_{12}, were

derived from the following five original variables: supervisor *Su*, seniority *Se*, number of stuffs *NS*, contractor level *CL*, and recent revenue *RR*. The following two components explain 69.155% of the variance:

$$M_{11} = -0.227Su + 0.067Se + 0.474NS - 0.014CL + 0.460RR \tag{1}$$

$$M_{12} = -0.087Su + 0.594Se + 0.025NS - 0.591CL + 0.064RR \tag{2}$$

(2) Two principal components, denoted as T_{11} and T_{12}, were derived from the following three original variables: start season *SS*, duration *Du*, and number of design changes *Nc*. The following two components explain 78.916% of variance:

$$T_{11} = 0.001SS + 0.603Du + 0.612Nc \tag{3}$$

$$T_{12} = 0.978SS - 0.127Du + 0.126Nc \tag{4}$$

(3) Other variables include location *Lo*, work difficulty *WD*, price index *PI*, and work type *WT*.

B. Medium-scale construction project: Three sets of principal components were selected.

(1) Two principal components, denoted as O_{11} and O_{12}, were derived from the following three original variables: contract type *CT*, work type *WT*, and price index *PI*. The following two components explained 57.563% of the variance:

$$O_{11} = 0.185CT + 0.614WT - 0.145PI \tag{5}$$

$$O_{12} = 0.586CT + 0.028WT + 0.737PI \tag{6}$$

(2) Two principal components, denoted as M_{21} and M_{22}, were derived from the following three original variables: supervisor *Su*, inspector *ID*, number of stuffs *NS*, and seniority *Se*. The following two components explained 63.963% of the variance:

$$M_{21} = 0.064Su - 0.069ID + 0.598NS + 0.607Se \tag{7}$$

$$M_{22} = 0.684Su + 0.606ID + 0.006NS + 0.008Se \tag{8}$$

(3) Two principal components, denoted as T_{21} and T_{22}, were derived from the following three original variables: start season *SS*, duration *Du*, and work difficulty *WD*. The following two components explained 70.174% of the variance:

$$T_{21} = 0.763SS - 0.064Du + 0.593WD \tag{9}$$

$$T_{22} = -0.237SS + 0.851Du + 0.428WD \tag{10}$$

(4) Other variables include recent revenue *RR*, capital *Ca*, number of design changes *NC*, and location *Lo*.

C. Small-scale construction project: One set of principal components was selected.

(1) Two principal components, denoted as T_{31} and T_{32}, were derived from the following four original variables: start season *SS*, price index *PI*, work difficulty *WD*, and duration *Du*. The following two components explained 88.869% of the variance.

$$T_{31} = -0.005SS + 0.024PI + 0.98WD + 0.36Du \tag{11}$$

$$T_{32} = 1.000SS - 0.001PI + 0.001Du \tag{12}$$

(2) Other variables include recent revenue *RR*, and number of stuffs *NS*.

The results obtained from PCA (Table 4) indicated that several groups of variable sets can be classified according project scale (i.e., large, medium, and small) and factor type (i.e., participant, case type, time, and location),

$$max\ var(x_1^*) = max\ var(\sum_{i=1}^{n} x_i \times cos(\theta_j)) \tag{13}$$

find the θ_j for *max* variance of x_1^*; where θ_j is the angle rotate of axis.

Then, the maximal variance for one of these variables is able to be refined as the so-called Principal Component after trials as PCA_1. The others will be replaced as PCA_2. Repeat the process to screen out all Principal Components proposed in the manuscript, which are close to independent to each other. In case of those failed to be chosen, if there is no appropriate substitute, they will be abandoned.

4. Back-Propagation Neural Network (BP-NN)

Fig. 4 illustrates the calibration of the BP-NN model from using the obtained principal components as input data. The calibrated BP-NN model was used to predict the construction duration in units of construction cost per day (NT$/day). NeuroSolutions is used for the BP-NN model development. The structure of the ANN model features a single hidden layer based on the back-propagation approach, which is a supervised learning network. In the input layer, the number of neurons was equal to the number of principal components. The activity function adopted a summation function, which was a weighted summation of the neuron output from the preceding layer. For the model training, the input data were the principal components and the construction duration from the database. The steepest descent method was used to determine the optimal solution, which is an optimal weighting matrix. The hyper tangent was selected as the transfer function.

Table 4. Classification of principal components for BPNN model training.

Scale of the construction	Principal component			
	participant	case type	time	location
Large scale 50 million plus dollars (duration is larger than 3.5 years)	(1) M_{11} and M_{12}: [Consist of *Su*, *Se*, *NS, CL*, *RR*]	(1) *WT*	(1) T_{11} and T_{12}: [Consist of *SS*, *Du*, *and CN*] (2) *PI* (3) *WD*	(1) *Lo*
Medium scale Between 10 million and 50 million dollars (duration is between 1 and 3.5 years)	(1) M_{21} and M_{22}: [Consist of *NS, Se*, *Su*, *ID*] (2) *RR* (3) *Ca*	(1) O_{11} and O_{12}: [Consist of *WT, CT, PI*]	(1) T_{21} and T_{22}: [Consist of *SS*, *Du*, *WD*] (2) *CN*	(1) *Lo*
Small scale Less than 10 million dollars (duration is between 0.5 and 1.5 years)	(1) *NS* (2) *RR*		(1) T_{31} and T_{32}: [Consist of *SS*, *Du*, *WD*, *PI*]	

Notes:

Some factors have underline as *NS*, *RR* that not produced from PCA process but include in the BP-NN model. For the reason have two:

(1) the purpose of PCA process is reduce the number of variables, but the *NS* and *RR* in the same attribute field just two variables, no need to do the process.

(2) *NS* and *RR* are affect the duration indeed base on the domain knowledge, so we must be join these variables in the model to check the influence to the duration.

We set the price index as a variable to avoid waste time of calculation in the future (different years).

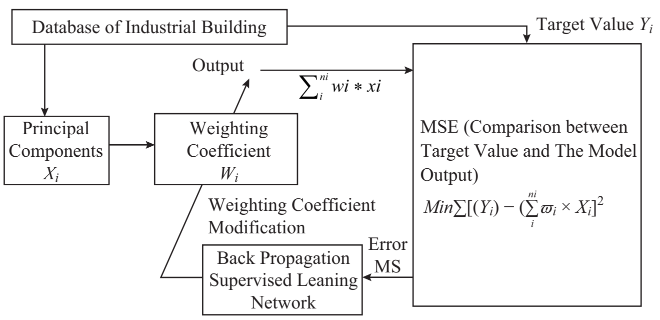


Fig. 4. Illustration of BP-NN model.

The objective function is expressed in Eq. (14):

$$Min\sum[(Y_i)-(\sum_i^{ni}\varpi_i \times X_i)]^2 \quad (14)$$

where Y_i is the target value, which is the construction duration obtained from the database; ϖ_i is the weight; and X_i is the output from the neuron output from the preceding layer.

Two stopping criteria were the number of iterations (5,000 runs) and MSE (<0.05), which is expressed in Eq. (15):

$$MSE = \frac{\sum_{j=0}^{P}\sum_{i=0}^{N}(d_{ij}-y_{ij})^2}{NP} \quad (15)$$

where d_{ij} is the output which result from the model operation; y_{ij} denotes the known construction cost per day (depend var.); and N and P denote the number of independent variables.

The model training was performed separately according to the project scale. Table 5 shows the calibrated model parameters, which are the weights of the hidden layer. The dimension of the weight matrix varies with the construction project scale. The MSE of the large-, medium-, and small-scale projects are 0.02-0.05, 0.06-0.10, and 0.08-0.11 respectively, where lower MSE values indicate more accurate calibration of the BP-NN model.

5. Results and Discussion

Fig 5 shows a comparison output between the predicted (a dotted line) and real construction duration. Although the calibrated model shows only an intangible statistical relationship between the construction duration and principal components, this study attempted to reveal the physical meaning that may exist behind the black box model. In addition to the project scale, the relationship between construction duration and (1) participant, (2) location, (3) time, and (4) case type are addressed.

(1) Participant

(a) Large-scale projects:

The major factors of the principal components are number of stuffs *NS* and recent revenue *RR* for M_{11}, and seniority *Se* and contractor level *CL* for M_{12}. All of these factors indicated that the capability of the contractors has the strongest influence on the construction duration. Moreover, M_{11} and M_{12} also have large weights in the BP-NN model, further indicating the importance of this variable. Therefore, contractor capability must be considered as a constraint in large-scale construction projects.

Table 5 The calibrated weights for hidden layer (*10^{-1}).

	Number of neuron (#)																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Larger Case (dimension 8 × 5)																			
LO	-5.3	-1.1	-4.3	1.7	-2.0	--	--	--	--	--	--	--	--	--	--	--	--	--	--
WD	***8.9***	4.7	-3.3	4.5	-1.3	--	--	--	--	--	--	--	--	--	--	--	--	--	--
WT	2.6	-15.5	***21.0***	-6.0	0.9	--	--	--	--	--	--	--	--	--	--	--	--	--	--
PI	4.1	-5.6	5.8	***6.6***	***-4.1***	--	--	--	--	--	--	--	--	--	--	--	--	--	--
M_{11}	-1.5	***-15.8***	-3.9	1.7	-2.0	--	--	--	--	--	--	--	--	--	--	--	--	--	--
M_{12}	-2.8	-7.1	9.7	-5.0	-3.1	--	--	--	--	--	--	--	--	--	--	--	--	--	--
T_{11}	1.0	9.0	-3.8	2.1	0.1	--	--	--	--	--	--	--	--	--	--	--	--	--	--
T_{12}	-3.9	8.8	8.0	1.9	0.3	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Medium Case (dimension 10 × 10)																			
LO	4.4	-3.2	-3.9	-0.9	2.6	0.0	4.7	1.5	-3.2	-0.5	--	--	--	--	--	--	--	--	--
NC	-4.6	4.2	-3.8	0.1	-3.1	0.5	3.1	3.8	-3.6	3.7	--	--	--	--	--	--	--	--	--
RR	-2.5	1.3	-0.8	-3.7	2.9	1.8	1.1	-3.3	1.3	1.2	--	--	--	--	--	--	--	--	--
CA	4.5	-1.2	1.2	-3.0	-4.3	-4.2	-4.2	0.8	-1.7	-0.9	--	--	--	--	--	--	--	--	--
O_{11}	4.3	-0.1	2.3	4.6	3.2	2.8	2.2	-2.3	4.7	4.5	--	--	--	--	--	--	--	--	--
O_{12}	-4.3	-0.6	1.5	4.7	0.5	0.5	-3.1	2.3	-0.2	-4.9	--	--	--	--	--	--	--	--	--
M_{21}	***4.2***	-0.9	1.6	2.8	2.0	-0.6	0.9	-1.5	-3.3	-3.1	--	--	--	--	--	--	--	--	--
M_{22}	0.2	2.0	-0.2	-4.4	***-4.7***	0.4	1.5	-2.9	0.5	-4.6	--	--	--	--	--	--	--	--	--
T_{21}	-4.8	3.5	0.2	2.9	-4.4	1.0	-3.7	4.1	-0.1	-0.5	--	--	--	--	--	--	--	--	--
T_{22}	4.5	4.1	-3.0	-3.3	-0.2	4.7	4.4	-3.0	2.6	-0.4	--	--	--	--	--	--	--	--	--
Small Case (dimension 4×19)																			
NS	3.1	-1.4	0.7	0.6	3.7	-0.2	-3.2	1.1	-2.6	1.4	-4.4	-4.2	4.2	2.7	3.8	1.4	-4.2	-0.1	-1.4
RR	-2.3	0.9	-2.6	3.4	-4.9	-0.6	-1.8	1.5	-5.1	-2.0	-1.4	4.1	-1.1	-4.3	2.9	1.3	1.3	-0.8	-0.1
T_{31}	4.4	-1.0	4.0	-0.4	1.6	1.5	-4.4	-0.6	-1.4	2.7	4.4	0.7	0.9	-2.6	-0.1	1.5	-3.7	1.5	2.4
T_{32}	3.1	3.2	1.7	3.5	-5.5	2.1	-0.2	2.3	0.1	1.7	0.3	0.9	-2.6	0.8	3.1	0.8	-1.7	0.3	-2.3

Note:
1. The number should times 0.1 for real value.
2. The weights are case dependent which may not be directly applied to other cases.

(b) Medium-scale projects:
M_{21} shows that contractor capability is a major influence in the number of stuffs *NS* and seniority *Se*. In addition to contractors level *CL*, the supervisor Su and inspector *ID* play crucial roles in medium-scale projects.

(c) Small-scale projects:
Although no principal factors are generated, the number of stuffs *NS* and recent revenue *RR* were critical factors.

(2) Time:
Starting season *SS* was the major factor for all large-, medium-, and small-scale projects. According to additional analysis for *SS*, the projects starting in summer have a negative impact on construction duration. This might be attributed to the typhoon season, which can prolong the construction period. In addition to *SS*, the number of design changes *NC* may be crucial for larger cases; however *NC* was non-significant for medium- and small-scale projects.

(3) Location:
The location *Lo* had a marked influence on large- and medium-scale projects. For construction projects in Taipei City, the cost per unit of time tends to be higher than in other areas because of restrictive regulations, higher risk of damaging the areas surrounding a construction site, and higher costs for labor. For projects located in suburban areas, such as in reclaimed land areas, the cost per unit of time is relatively lower, which could be attributed to less stringent regulations and easier mobility of construction equipment.

(4) Case type:
Regarding the work type *WT* and work difficulty *WD* variables, the *WT* has significant impact on large- and medium-scale projects. The *WT* may imply the complexity of a project. Ranked in descending order, the construction costs per unit of time are harbor engineering, steel works, normal constructions, and structural engineering. The *WD* variable had a marked impact for all project scales, indicating that the work difficulty directly influences the con-

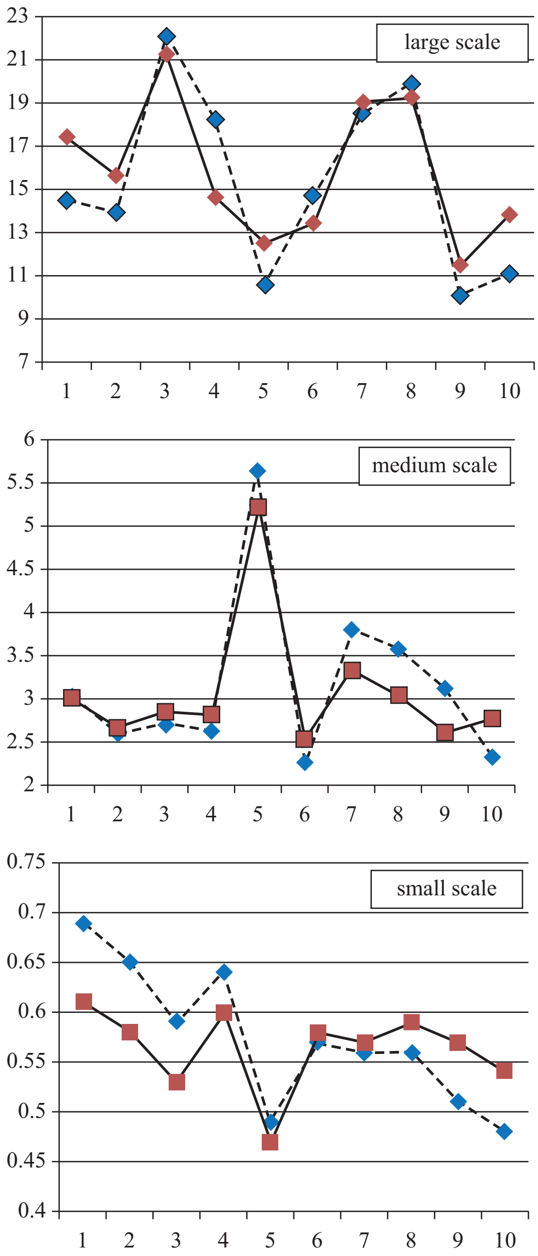


Fig. 5. The verification results of selected 10 cases.

struction cost per unit of time.

IV. CONCLUSION

This study used a history database containing 50 years of industrial building construction cases in the petrochemical industry in Taiwan. Because of the variety of cases, they were classified into three categories: large-, medium-, and small-scale projects, and three ANN models were independently trained for each category to improve the prediction results. To facilitate comparison, the prediction duration was represented as the construction cost per unit of time. The results demonstrate the considerable applicability of the proposed methodology. Although this study focused on industrial building construction, the proposed methodology may be applicable for other types of buildings.

REFERENCES

Bromilow, F. J. (1969). Contract time performance: expectations and reality. Building Forum 1, 70.

Chen, S. H. (2005). Principal component analysis class notes of power point files. Feng Chia University Transportation Technology and Management, Taichung, Taiwan.

Cheng, T., J. Teizer, G. C. Migliaccio and U. C. Gatti (2013). Automated task-level activity analysis through fusion of real time location sensors and worker's thoracic posture data. Automation in Construction 29, 24-39.

Ding, B., Y. M. Zhang, X. Y. Peng, Q. C. Li and H. Y. Tang (2013). A hybrid approach for the analysis and prediction of elevator passenger flow in an office building. Automation in Construction 35, 69-78.

Goulielmos, A. M. and E. Siropoulou (2006). Determining the duration of cycles in the market of second-hand tanker ships, 1976-2001: Is Prediction Possible?. International Journal of Bifurcation and Chaos 16, 7, 2119-2127.

Hojjat, A. (2001). Neural networks in civil engineering: 1989~2000. Computer-Aided Civil and Infrastructure Engineering 16, 126-142.

Huang, T., B. Q. Huang, Y. Huang, M. T. Kechadi and B. Buckley (2010). Using pca to predict customer churn in telecommunication dataset. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6441, 2, 326-335.

Hui, F. Z. (2003). Operation-related impact of uncertainty on the total duration of the assessment model. Master's thesis, Department of Construction Engineering, Chaoyang University of Technology, Taichung, Taiwan.

Jan, Q. C. (2003). Combination of fuzzy neural network and genetic algorithm on forecast project duration. Master's thesis, National Cheng Kung University, Tainan, Taiwan.

Jaskowski, P., S. Biruk and R. Bucon (2010). Assessing contractor selection criteria weights with fuzzy AHP method application in group decision environment. Automation in Construction 19, 120-126.

Kelly, W. E. (2002). Anxiety and the prediction of task duration: a preliminary analysis. The Journal of Psychology 136, 1, 53-58.

Khosrowshahi, F. and A. P. Kakat (1996). Estimation of project total cost and duration for housing projects in the U.K. Building and Environment 31, 4, 373-383.

Lam, K. C., R. Tao and M. C. K. Lam (2010). A material supplier selection model for property developers using fuzzy principal component analysis. Automation in Construction 19, 608-618.

Lin, F. Z. (2003). Evolutionary computation part II: genetic algorithms and three application examples. Journal of Intelligent Technology and Applied Statistics, 33-44.

Liu, W. Z. (2002). Application of DELPHI method to a reasonable period: a case study of major railways. Master's thesis, National Chiao Tung University, Department of Civil Engineering, Hsinchu, Taiwan.

Lin, C. C. (2005). Construction progress and productivity analysis on spawning gravels long tunnel. Master's thesis, Department of Civil Engineering, Feng Chia University, Taichung, Taiwan.

Liu, S. S. and C. J. Wang (2012). Optimizing linear project scheduling with multi-skilled crews. Automation in Construction 24, 16-23.

Monton, B. and B. Kierland (2006). How to predict future duration from present age. The Philosophical Quarterly 56, 222, 16-38 January.

Ma, Z., X. Luo, Du, P. J. Hou and D. Duan (2011). A new method to predict the reservoir porosity based on fuzzy-PCA and neural network. Proceedings of the 2011 Chinese Control and Decision Conference, 2587-2590.

Ran, W., J. Qiao and X. Ye (2004). Soft-Measuring approach to on-line predict BOD based on PCA time-delay neural network. Proceedings of the World Congress on Intelligent Control and Automation (WCICA) 4, 3483-3489.

Song, L. and F. Ramos (2008). A framework for real-time simulation of heavy construction operations. Proceedings of the 2008 Winter Simulation

Conference, 2385-2395.

Wang, Z., J. Xu, F. Lu and Y. Zhang (2009). Using the method combining PCA with BP neural network to predict water demand for urban development. 5th International Conference on Natural Computation 2, 621-625.

Yang, J., D. J. Edwards and E. D. L. Peter (2003). A computational intelligent fuzzy model approach for excavator cycle time simulation. Automation in Construction 12, 725-735.

Yao, M. (2002). Study on water pipes swap works: comparison on open-cut and on-site construction method of curing schedule lined. Master Thesis, Institute of Civil Engineering, National Taiwan University, Taipei, Taiwan.

Yau, N. J. and J. B. Yang (1998). Case-Based reasoning in construction management. Computer-Aided Civil and Infrastructure Engineering 13, 143-150.

Zhang, L. (2002). Tabu search method is applied to the project schedule/cost tradeoff optimization of research. Master's thesis, Department of Construction Engineering, Chaoyang University of Technology, Taichung, Taiwan.

Zou, M. C. and Z. H. Sun (2004). Prediction mode in comparison with the integration of spatial data mining: a case study of Chi-Chi earthquake caused landslides spatial database. Journal of Geographical Science 38, 93-109.

Žujo, V. and D. Car-r-Pušić (2009). Modeling of risk factor influence on construction time overrun by applying multi variant regression analysis. CIB Joint International Symposium 2009 - Construction Facing Worldwide Challenges.