PCA-NARX Time Series Prediction Model of Surface Settlement during Excavation of Deep Foundation Pit

Qingwen Ma, Sihan Liu and Xianchang Zhao
1School of Water Conservancy Science and Engineering, Zhengzhou University, Zhengzhou City, Henan Province 450001, China
2Henan Water Conservancy Survey Co., Ltd., Zhengzhou City, Henan Province 450001, China
*Corresponding author’s e-mail: mqw2008@zzu.edu.cn

Abstract. In order to ensure that the excavation of the foundation pit can proceed smoothly in complex environment, the deformation prediction is particularly important. Because the traditional model can not accurately predict the surface settlement when the external environment changes, so the PCA-NARX foundation pit settlement prediction model is proposed. The principal component analysis (PCA) method is used to analyze the main components that affect the surface settlement and use them as parts of the external input to dynamically predict the surface settlement. Combined with Beilong Lake Foundation Pit Project in Zhengzhou, the applicability and accuracy of the model are evaluated. The results show that: (1) Comparing with the prediction results of BPNN model, NAR model and NARX model, PCA-NARX model has higher accuracy, the mean square error is only 0.1677. (2) Comparing with NARX model without PCA method, PCA-NARX model can achieve higher accuracy in the same training time, and the accuracy is improved by 12.89%, so it is more effective to predict the surface settlement during excavation. (3) The prediction results of PCA-NARX at different monitoring points also performed well, which means that it can be applied to the prediction of other monitoring contents of the foundation pit.

1. Introduction

Deformation of deep foundation pits is an important parameter in the construction process of the project. Therefore, it is of great significance to analyze the evolution of deformation based on existing surface settlement monitoring data for the project safety. But the settlement prediction is a complex and dynamic project, and the complexity and dynamic of the construction process bring many difficulties to the deformation prediction[1]. The time series prediction method is widely used in the settlement prediction because of the continuity of the settlement process. This method is based on the data of the past certain time to carry on the statistical analysis and calculate the subsequent development trend of things, but it ignores the role of external influencing factors and has certain limitations[2-5]. So the best foundation settlement prediction model should not only consider the time series of changes in the settlement, but also consider a variety of environmental factors on the settlement.

The relevant literature at home and abroad shows that the settlement changes during excavation are affected by many factors, which directly or indirectly affect the overall safety of foundation pit. Therefore, it is of great practical significance to analyze and extract the main factors affecting the
deformation of foundation pit, and to predict the trend of foundation pit settlement in the future excavation process with the historical real-time monitoring data[6-7]. NARX neural network model is a kind of dynamic neural network which can overcome the limitation that the traditional time series can not consider the external influence factors, and also consider the influence of the external input on the target[8-11]; The PCA method summarizes complex external factors to reduce dimensionality, reduces the calculation time of the NARX model and improves the calculation accuracy[12-14].

On the basis of the above contents, this paper selects PCA-NARX model as the foundation pit settlement prediction model, combines the foundation pit settlement monitoring data of Jinrong Island in Beilong Lake, Zhengzhou, Henan Province to predict the trend of foundation pit settlement change, and compares the prediction results of BP model, NAR model, NARX model and PCA-NARX model to evaluate the accuracy of the model.

2. Establishment of PCA-NARX time series prediction model
NARX time series prediction model trains the network by introducing appropriate external input, after many training, the value of delay parameter d is determined, and finally the surface settlement prediction is carried out by using the network with good training effect, the process of which is shown in Figure 1.[15].

\[
y(t+1) = f[X(t), y(t)]
\]

In the formula: \( y(t+1) \) is the settlement amount predicted by the model at \( t+1 \) moment; \( X_n(t) \) is the collection of \( n \) related factors that affect the settlement amount before the \( t \) time, which is \( X_n(t) = (x_{1n}, x_{2n}, x_{3n}, \ldots, x_{nn}) \); \( y(t) \) is the historical settlement sequence before the \( t \) time.

![Figure 1. NARX network structure diagram.](image)

2.1. Initial selection of external input variables \( X_n(t) \)
In view of the change law of foundation pit settlement, scholars at home and abroad have carried on a lot of research with engineering examples, and think that there are many main factors affecting foundation pit settlement, which can be summarized and classified as follows[16-18]:

- Geological, hydrological and climatic conditions of foundation pit;
- Various physical and mechanical indexes of soil;
- Space dimensions of foundation pit;
- Support type, support structure type and structure stiffness;
- Construction method, construction process and construction schedule;
- The internal stress characteristics of foundation pit.

But for a specific foundation pit engineering, many influencing factors have been determined in the design stage, and can not be selected as the set of influencing factors in this paper, such as plane size, insertion ratio of retaining wall (pile), thickness and stiffness, etc. In addition, if all factors are taken as external input variables, the training difficulty will be increased and the training accuracy will be reduced. Therefore, seven factors that may affect the change of foundation pit settlement are selected: internal friction angle, cohesion force, gravity, groundwater level, permeability coefficient, excavation depth of deep foundation pit, and axial force of internal support.
2.2. Correlation analysis of input variables

Because the soil above the excavated surface may contain multiple soil layers and the physical parameters of these soil layers are different, so this paper decides to select the weighted average of the above parameters as the input variables, and then construct a scientific prediction model. For evaluating the correlation degree between the above seven variables and the surface settlement of foundation pit, this paper introduces Pearson correlation coefficient to carry on the correlation analysis, and calculation formula such as formula (2).

\[
 r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}
\]

(2)

The range of correlation coefficient is \( r \in [-1,1] \). If the absolute value of the \( r \) is larger, the correlation between the variable and foundation pit settlement is stronger and vice versa. The analysis results of correlation are shown in Table 1.

| Variables                      | Correlation coefficient |
|-------------------------------|-------------------------|
| Internal friction angle       | 0.607                   |
| Cohesion force                | 0.900                   |
| Gravity                       | 0.912                   |
| Groundwater level             | 0.713                   |
| Permeability coefficient      | 0.876                   |
| Excavation depth              | 0.989                   |
| Axial force of internal support | 0.967                  |

Table 1. Variance correlation coefficient.

After analyzing Table 1, it is found that the variables of internal friction angle, cohesion, gravity, groundwater level, permeability coefficient, deep foundation pit excavation depth, axial support force and foundation pit settlement have high correlation coefficients, all greater than 0.5, which has a good positive correlation with foundation pit surface settlement. So it is considered that these seven variables can be used as the set of relevant variables used in the prediction model, and the set of relevant variables of the model can be expressed as: \( \mathbf{X}(t) = [x_1, x_2, x_3, \ldots, x_7] \).

2.3. Calculation process of PCA-NARX prediction model

Section 2.2 has preliminarily identified 7 variables related to foundation pit settlement, but there may be data redundancy between these 7 variables, that is, data duplication or surplus. Data redundancy will undoubtedly increase the amount of calculation, lead to inconsistent data, and reduce the reliability of the model. PCA method first standardizes the variable matrix \( \mathbf{X}(t) \) to eliminate the influence caused by different dimensions, then calculates the correlation coefficient, eigenvalue and eigenvector of the matrix \( \mathbf{X}(t) \), and then obtains the eigenvector and the corresponding eigenvector \( \mathbf{R} \) the matrix through the eigen equation. Finally, the first \( m \) principal components are selected according to the cumulative contribution rate of each principal component \( Q_m \). Generally, if \( Q_m \geq 85\% \), it can be guaranteed that the new variable contains the main information of the original data.

Using PCA to reduce the dimensionality of the relevant variable set \( \mathbf{X}(t) \) to obtain a new comprehensive variable \( \mathbf{X}_n(t) \). \( \mathbf{X}_n(t) \) and the historical time series of sedimentation \( y(t) \) constitute the input of the NARX network, and the neural network is trained to obtain the prediction model, then the prediction value at the next moment \( y^{'}(t+1) \) is obtained. The calculation process is shown in Figure 2:
2.4. Model performance evaluation indicators
Predictive model performance evaluation is the key to judging the predictive performance of the model. The main indicators for evaluating the performance of the NARX model are mean square error (MSE), mean absolute error (MAE), mean relative error (MRE), etc. The smaller the values of MSE, MAE, MRE and other indicators, the higher the prediction accuracy of the model. The formulas are shown in formula (3), formula (4) and formula (5) respectively.

\[
MSE = \frac{1}{N} \sum_{i=1}^{N} (y(i) - y'(i))^2 \tag{3}
\]

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |y(i) - y'(i)| \tag{4}
\]

\[
MRE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y(i) - y'(i)}{y(i)} \right| \tag{5}
\]

3. Example Application Verification
3.1. Engineering Profile
The main research object of this paper is No. 12 foundation pit located in the outer ring of Beilonghu Financial Island, Zhengdong New District, Zhengzhou City, as shown in figure 3, figure 4. The total area is about 18366.79 m², the underground is 4 stories, the foundation pit is generally excavated 12.5 m, the deepest area reaches 16 m, which belongs to the large-scale deep foundation pit engineering. The supporting type is reinforced concrete supported by bored cast-in-place pile, and an equal thickness cement-soil mixing wall waterproof curtain is arranged on the outside. According to its genetic type, lithology and engineering geological characteristics, the 16.0 m depth strata in the site are divided into 4 engineering geological unit layers and 2 engineering geological unit sublayers, 7.3 m is plain fill and 7.3 m-16.0 m is fine sand.

The buildings and roads around the proposed site are dense, the water level line near the site is high, the foundation construction is difficult, and the influence on the surrounding environment is also great, therefore, it is very important to do well the prediction and control of foundation pit surface settlement during construction.
3.2. PCA Analysis of relevant variables

This chapter selects two representative monitoring points KD25 and KD62 as targets, which are located in the middle of the inner and outer ring sides of the foundation pit. Choose continuous 90-period on-site monitoring data, and the date range of the applied data is from September 5, 2018 to November 23, 2018. The settling time curves of the two points are shown in Figure 5. Data from 1 to 50 periods are selected as training samples, and data from 51 to 80 periods are selected as test samples. Taking KD25 as an example, part of the processed sedimentation sample data from stage 1 to stage 50 is shown in Table 2.

![Figure 3. Beilong Lake Jinrong Island.](image)

![Figure 4. Schematic diagram of foundation pit shape and surrounding environment.](image)

![Figure 5. Time-lapse curves of surface settlement at monitoring points KD25 and KD62.](image)

| Serial number | $x_1(t)$ | $x_2(t)$ | $x_3(t)$ | $x_4(t)$ | $x_5(t)$ | $x_6(t)$ | $x_7(t)$ | $y(t)$ |
|---------------|----------|----------|----------|----------|----------|----------|----------|--------|
| 1             | 17.76    | 16.14    | 19.50    | 16.33    | 0.00     | 0.50     | -779.57  | -1.52  |
| 2             | 17.76    | 16.14    | 19.50    | 16.37    | 0.00     | 0.60     | -801.12  | -1.76  |
| 3             | 17.76    | 16.14    | 19.50    | 16.45    | 0.00     | 0.80     | -820.92  | -2.03  |
| 4             | 17.76    | 16.14    | 19.50    | 16.49    | 0.00     | 0.91     | -828.13  | -2.58  |
| 5             | 17.76    | 16.14    | 19.50    | 16.53    | 0.00     | 1.13     | -832.33  | -2.85  |
| 6             | 17.76    | 16.14    | 19.50    | 16.57    | 0.00     | 1.25     | -840.43  | -3.14  |

Table 2. KD25- Processing input and output variables.
According to the PCA analysis steps, the contribution rate of variables can be obtained as shown in Table 3. If the cumulative contribution rate of principal component \( Q_c \geq 85\% \), then \( m=2 \), indicating that the main information of the data can be covered only by extracting principal component 1 and principal component 2, and thus the component score coefficient matrix can be obtained, as shown in Table 4.

### Table 3. Principal component contribution statistics.

| Composition | 1    | 2    | 3    | 4    | 5    | 6    | 7    |
|-------------|------|------|------|------|------|------|------|
| Initial eigenvalues Total | 5.355 | 0.739 | 0.505 | 0.232 | 0.127 | 0.034 | 0.009 |
| Contribution rate | 76.501 | 10.554 | 7.218 | 3.311 | 1.808 | 0.481 | 0.127 |
| Cumulative contribution | 76.501 | 87.055 | 94.273 | 97.584 | 99.392 | 99.873 | 100  |

### Table 4. Component score coefficient matrix.

| Composition | V1    | V2    | V3    | V4    | V5    | V6    | V7   |
|-------------|-------|-------|-------|-------|-------|-------|------|
| 1           | -0.740 | 0.846 | -0.954 | 0.740 | 0.907 | 0.986 | -0.916 |
| 2           | 0.609  | -0.418 | -0.122 | 0.227 | 0.250 | 0.073 | -0.243 |

According to Table 4, the coefficient matrix of principal component analysis can be obtained \( A = \begin{bmatrix} -0.740 & 0.846 & -0.954 & 0.740 & 0.907 & 0.986 & -0.916 \\ 0.609 & -0.418 & -0.122 & 0.227 & 0.250 & 0.073 & -0.243 \end{bmatrix} \), and then the new comprehensive variables \( x_1, x_2 \) can be obtained.

### 3.3. Training and evaluation of predictive models

The relevant variable \( x_1, x_2 \) after dimension reduction and the historical settlement \( y \) of the foundation pit are used as the input data set of the NARX network. In order to judge the training effect of the network, 35 sets of data are used as the training set for model, 8 sets as the test set and 7 sets as the prediction set. After many times of model training, this paper chooses NARX network delay order \( d \) to be 4 and hidden layer \( m \) to be 10.

After training, a high-precision NARX dynamic neural network model can be obtained, and its overall fitting goodness factor \( R=0.99948 \), as shown in Figure 6. The index indicates that the prediction model has excellent performance and can meet the requirements.

To verify the prediction effect of the NARX neural network model, this paper predicts the settlement of the foundation pit by using the BP static neural network model, the NAR dynamic neural network model and the NARX dynamic neural network model. The comparison of the predicted effects of the four networks is shown in Table 5. The graphs of the KD25 foundation pit settlement curves predicted by the four models are summarized in Figure 7, and the graphs of KD62 are summarized in Figure 8.
The analysis of figures 7 and 8 show that: (1) the predicted value curves of the four models have good similarity with the actual value curve, but the PCA-NARX model is more consistent with the actual value curve; (2) the prediction results of either the prediction point KD25 or the point KD62, PCA-NARX model are in line with expectations.

Analysis of the data in Table 5 shows: (1) Analysis of the convergence steps and training time of the four models shows that although the error of the BP neural network is low and meets the accuracy of engineering requirements, the training speed is significantly slower than other time series models; (2) After analyzing the MSE, MAE, and MRE of the four models, it is considered that the MSE, MAE,
and MRE of the PCA-NARX model proposed in this paper are smaller than other models, and the root mean square error is 0.1135, which is 62.17% lower than the BP model, 35.29% lower than the NAR model and 12.89% lower than the NARX model, indicating that the model has higher accuracy. This is because the PCA-NARX model optimizes the network input variables relative to the traditional BP model and the NARX model, and has better generalization ability. At the same time, it takes into account the advantages of the NAR model to predict the time series data, which provides a higher precision and faster operation model for the prediction of surface settlement during foundation pit excavation.

Due to space limitations, this paper only lists the surface settlement prediction results of two representative monitoring points on the inner and outer rings of the financial island. In fact, the model still performs well on other monitoring points and maintains a high prediction accuracy. Therefore, it is believed that this model can be applied to other monitoring content, for example, horizontal displacement prediction, so that the construction unit can fully and quickly understand the safety status of the entire foundation pit project during the excavation process.

4. Conclusions

- Compared with the traditional time series prediction model and BP neural network model, the NARX time series prediction model introduces the dynamic changes of environmental factors during the construction process as an external input variable set. In this way, it not only considers the internal correlation between the subsidence time curve data, but also considers the impact of changes in the external environment on the change trend of the settlement, which can effectively improve the accuracy of prediction.

- Compared with the NARX model, the combination of the PCA method and the NARX method optimizes the network input variable set. Taking point KD25 as an example, the prediction accuracy increases by 12.89% during the same training time, and the training efficiency is also increased, which proves that PCA-NARX model is superior to the NARX model, and verifies the validity and reliability of the PCA-NARX settlement prediction model.

- The PCA-NARX model performs well in predicting points KD25, KD62 and other monitoring points, indicating that the model has good applicability to the foundation pit and can also be applied to the prediction of other monitoring contents.

References

[1] Jiang Xiaofeng, Liu Guobin, Zhang Weili, et al.(2010) A Study on Deformation Characteristics of Ultra Deep Foundation Pit in Shanghai Area Based on Measured Data. Journal of Geotechnical Engineering, 32(S2):570–573.

[2] Yin Guangzhi, Yue Shun, Zhong Tao, Li Dequan.(2009) Time series analysis of tunnel displacement based on ARMA model. Geotechnical Mechanics, 30(09):2727-2732.

[3] Xu Qing, Guo Wei, He Song.(2017) Research on Prediction of Ground Settlement in Excavation Process Based on Logistic-ARMA Combination Model. Safety and Environmental Engineering, 24(03):160-163.

[4] Zhao Zixin.(2018) Analysis and prediction of deep foundation pit deformation based on time series metabolism method. Shanghai Jiaotong University.

[5] Cao Jing, Ding Wenyun, Zhao Changshu, Song Zhigang, Liu Haiming.(2014) Time series prediction of foundation pit deformation based on LSSVM-ARMA model. Geotechnical Mechanics, 35(S2):579-586.

[6] Xu Ning.(2019) Research and Application of LSTM Prediction Model of Deep Foundation Pit Deformation. Nanchang: Jiangxi University of Technology.

[7] Ren Qingguo.(2015) Analysis of Support Optimization and Yangjiao Stress Deformation of a Special-shaped Deep Foundation Pit in Wuhan. Wuhan: Wuhan University of Technology.
[8] Civilization, Zhang Dingli, Fang Qian, Zhang Liangyi.(2015) NARXNN time series prediction model of ground subsidence during subway station construction. Journal of Rock Mechanics and Engineering,34(S1):3306-3312.

[9] Civilization, Zhang Dingli, Fang Qian, Qi Jun, Fang Huangcheng, Chen Wenbo.(2017) A Study on Nonlinear Autoregressive Time Series Prediction Method for Tunnel Surrounding Rock Deformation. Journal of Beijing Jiaotong University,41(04):1-7.

[10] L INT G,HORNE B G ,TINO P ,et al.(1996) Learning long-term dependencies in NARX recurrent neural networks. IEEE Transactions on Neural Networks,7(6):1 329-1 338.

[11] BEYHAN S ,ALCI M.(2010) Fuzzy functions based ARX model and new fuzzy basis function models for nonlinear system identification. Applied Soft Computing,10(2):439-444.

[12] Huang Zhongshan, Tian Ling, Xiangdong, Wei Yaozhong.(2018) Prediction of oil temperature trend of wind turbine gearbox based on PCA and SPC- dynamic neural network. Journal of Tsinghua University (Natural Science Edition),58(06):539-546.

[13] Yuan Hongchun, Zhao Yantao, Liu Jinsheng.(2018) Prediction of ammonia nitrogen based on PCA-NARX neural network. Journal of Dalian Ocean University,33(06):808-813.

[14] Yuan Hongchun, Huang Junhao, Zhao Yantao.(2019) Prediction of dissolved oxygen based on PCA-NARX neural network. Journal of Shandong Agricultural University (Natural Science Edition),50(05):902-907.

[15] Wong C X,Worden K.(2007) Generalised NARX shunting neuralnet-work modeling of friction. Mechanloal Systems and Signal Processing,21(1):553-572.

[16] Ge Ming.(2016) Deformation Law and Influencing Factors of Subway Foundation Pit Excavation in Tianjin Soft Soil Area. Beijing Jiaotong University.

[17] Wu Chaoyang.(2015) A Study on the Influence of Foundation Pit Construction in Subway Station on Adjacent Buildings. Hunan University.

[18] Huang Chuansheng.(2011) A Study on Prediction Method and Engineering Application of Excavation Deformation of Deep Foundation Pit of Subway. Central South University.