The Application of ARIMA Model in the Project of China Zun Tower
Heng WANG, Xu-qing ZHANG* and Guo-dong YANG
College of Geo-exploration Science and Technology, Jilin University, Changchun 130026, China
*Corresponding author

Keyword: Time series, Bolt axial force, ARIMA, Forecasting.

Abstract. This paper utilizes the ARIMA (Autoregressive Integrated Moving Average) model to analyze the axial bolt force data of foundation pit in the project of China Zun Tower and fit the specific change process of bolt axial of the anchor to predict the trend of change. At last, taking use of the smooth R squared to evaluate the accuracy and significant of the prediction results which can give the prediction of the deformation monitoring in the process of construction and operation. By means of the practical verification, the model obtains good effect to analyze and forecast the anchor axial force data in the foundation pit.

Introduction
In recent years, with the rapid development of new national urbanization, the space of cities become further tightening. In order to extend the urban space, people pay more attention to the development and supporting technology of urban underground space for impelling the development of the deep foundation pits[1]. Owing to monitor the foundation pit deformation, it requires to apply the former monitoring data to analysis and forecast the next phase of deformation of foundation pit precisely in the process of development and construction of deep foundation pit. The model of time series has the ability to analyze and predict the tendency of the complex system development. Its key idea is that using the historical information data to predict the future information.

Time Series Model
ARMA Model Principle
The autoregressive integrated moving average (ARMA) method takes the time series data as basic data obtained from systematic observation, which adopts the method of parameter and curve fitting for effective forecast. It is characterized by steady, normal and zero mean[2]. The result is not only associated with the first n observation values, but also with the first m interference terms. According to the multiple linear regression, the general form of ARMA (Auto-Regressive Moving Average Model, ARMA) model is given. The general form of the ARMA model is as follows:

\[ X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_n X_{t-n} - \theta_1 \mu_{t-1} - \theta_2 \mu_{t-2} - \ldots - \theta_m \mu_{t-m} + \mu_t \]

where \( X_t \) is the observation values, \( \varphi_i (i=1, 2, \ldots, n) \) is the auto-regressive parameters, \( \theta_j (j=1, 2, \ldots, m) \) is the moving average parameters, \( \mu_t \) is the white noise series[3].

The Improved ARMA Model – ARIMA Model
In general, the time series is not smooth, but all kinds of trend and cyclical and it cannot be described by the ARMA model directly. It can use the improved model of ARIMA to describe them for such time series.

If the sequence just has the tendency, when it can smoothly after d order difference, then we can establish the improved ARMA model-ARIMA (p, d, q) model.

Let \( \{X_t\} \) be the stationary series from the non-stationary series \( \{y_t\} \) by d-order difference. The process of d-order difference as follows:
\[ X_t = \nabla^d y_t \quad (t > d) \]  

(2)

Establish the ARMA (p, q) model from the stationary series \( \{X_t\} \):

\[ X_1 = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} \]  

(3)

where \( p \) is the order of auto regressive model (AR), \( q \) is the order of moving average (MA), \( c \) is the constant term\(^{[4]}\).

Then introduce the delay operator (denoted by \( L \)), then \( LX_t = X_{t-1} \). Thus, the k-order delay operator is defined below:

\[ L^k X_t = X_{t-k} \]

The ARMA (p, q) can be expressed as follows:

\[ (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)X_t = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)u_t + c \]  

(4)

Or

\[ \phi(L)X_t = \theta(L)u_t + c \]  

(5)

Where \( \phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p \) is autoregressive polynomial; \( \theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q \) is the moving average polynomial.

By simultaneously expressions (4) and (5) to obtain the ARIMA (p, d, q) model of non-stationary time series, which is expressed by the following expression:

\[ (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)\nabla^d y_t = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)u_t + c \]  

(6)

### Analysis of Engineering Example

#### Project Summary

The project of Beijing CBD-Z15 (China Zun Tower) is located at the core area of Beijing chaoyang. According to the national standard and requirements of technical specification, pit construction should be observed on anchor stress regularly in the process of foundation. This paper has selected the observation point of ML-03-1 which data of bolt axial force monitoring to study from the middle of the foundation pit. We choose the 50 period data of 5 days at average interval to construct the model from 2nd Jan in 2013 to 30th Dec in 2013. Then using the fitting results of former phrase 45 to predict the last phrase 5. The picture of Bolt axial force monitoring data (KN) and changes over time as follows:

![Figure 1. The raw data of bolt axial force.](image-url)
Data Analysis

Time Series Stationary Test. In order to judge whether a time series is stationary, it usually adopts augmented DF tests namely Augmented Dickey-Fuller test. This test model has three forms, respectively: (1) containing the constant term; (2) containing the constant term and linear time adjustment; (3) no constant term and linear time adjustment. This test assumes all raw series are the non-stationary time series. Firstly analyzing the time series diagram, if the constant term is significant, it will be the first test form; if the sequence has the trend and constant term, and it will be the second test form; if the time tendency and constant term is both not significant, it will be the third form.

If the t-statistic is less than the 1%, 5% and 10% critical value of the test level, it can decline the null hypothesis $H_0$ and suggest that the time series is stationary. Conversely, if the time series is not stable, it will be test again after the difference processing until meet stationary conditions \[^5\].

The Non-stationary time series was converted into the stationary series through the method of difference processing. The difference equation is a kind of recurrence relation.

The function $X_n = f(n), \quad n = \ldots, -2, -1, 0, 1, 2, \ldots$

The function $X_n = f(n)$ in the first order difference of n time is defined as:

\[
\Delta X_n = X_{n+1} - X_n = f(n + 1) - f(n)
\]

The function $X_n = f(n)$ in the first order difference of n time is defined as the difference of the first order:

\[
\Delta^2 X_n = \Delta X_{n+1} - \Delta X_n = X_{n+2} - 2X_{n+1} + X_n
\]

Simultaneously, the k-order ($\Delta^k X_n$) difference can be defined in turn \[^6\].

It can be seen from Figure 2 that the raw data is not stationary. The diagram of results under the difference processing:

![Figure 2. Anchor axial force-time after first order.](image-url)
ADF test of the data after the difference processing, the result is shown in figure:

![Figure 3. Anchor axial force-time after first order.](image)

| Table 1. First-order ADF test. |
|--------------------------------|
| t-Statistic | prob.* |
| Augmented Dickey-Fuller test statistic | -3.885403 | 0.0220 |
| 1%level | -4.205004 |
| Test critical values: | |
| 5%level | -3.526609 |
| 10%level | -3.194611 |

| Table 2. Second-order ADF test. |
|--------------------------------|
| t-Statistic | prob.* |
| Augmented Dickey-Fuller test statistic | -3.980218 | 0.0037 |
| 1%level | -3.605593 |
| Test critical values: | |
| 5%level | -2.936942 |
| 10%level | -2.606857 |

From the table 1, we can see that the t-statistic value of ADF test is -3.885403 which is greater than the 1% of the critical value of t-statistic after the first order difference. The value of the t-Statistic in the table 2 is less than the 1%, 5% and 10% of the critical value of t statistic. The results of the test decline the null hypothesis, namely the second-order series is a stationary time series [7].

**Identifying and Building Model.** In ARIMA (p, d, q) model, d is the order of the difference, p and q value is resulted from the difference of sequence of autocorrelation and partial autocorrelation figure trailing and truncated to decide. Specific principle refer to the table below:
Table 3. Model selecting principle.

| p | q | model | auto correlation function | tailing | partial autocorrelation function | truncation | tailing |
|---|---|-------|---------------------------|--------|--------------------------------|----------|--------|
| ARIMA (p,d,0) | ARIMA (0,d,q) | ARIMA (p,d,q) |

The order combination can make up different models after the preliminary selection. This paper choose the stationary R square which equivalent to the percentage of the variance of the regression equation occupy the dependent variable as the evaluation index to evaluate the model fitting\(^8\). This evaluation index can compare the difference between model of the frozen composition and simple mean model. The model is better than the simple mean model when the index value is positive\(^9\).

The equation of R square as follows:

\[
R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(y_m - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}
\]  

where SST is total sum of square, SST is the regression sum of squares, \(\bar{y}\) is the mean value, \(y_i\) is the actual value, \(y_m\) is the fitted value. When \(R^2 = 1\) denoting that all observation points fall on the regression line and it is the highest reliability.

The nature of the trailing from the autocorrelation and partial autocorrelation figure can judge out the ARIMA (p, d, q) model can be used. But the order of ACF and PACF cannot entirely determine which model is the most fitting and accurate. Therefore, we can choose the orders randomly to establish the different model on the basis of analyzing the diagram. From the fitting diagram, ARIMA (2,2,8) and ARIMA (2,2,5) can be chosen preliminarily. The index of two models as follows:

Table 4. ARIMA fitting index.

| Model      | Number of predictions | Model Fit statistics | Ljung-Box Q(18) |
|------------|-----------------------|----------------------|-----------------|
| ARIMA(2,2,8)| 0                     | 0.869 5.027          | 25.518 8 0.001  |
| ARIMA(2,2,5)| 0                     | 0.936 4.332          | 16.043 11 0.140 |

Owing to the data of stationary R-squared of ARIMA(2,2,5) higher obviously than the data of ARIMA(2,2,8),and more close to 1. We choose the ARIMA (2,2,5) model. The model formula ARIMA (2, 2, 5) as follows:

\[
Z_t = -1.091939Z_{t-1} - 0.651375Z_{t-2} - 0.4673\mu_{t-1} - 0.658813\mu_{t-2} - 0.585429\mu_{t-3} + 0.815197\mu_{t-4} - 0.058222\mu_{t-5}
\]  

White Noise Test of Residual Series Model. The white noise test results of the ARIMA(2,2,5) model as follows:
The autocorrelation value and partial correlation value of residual series model what was established in this paper basically all within the 95% confidence interval from the above diagram. The probability value of Q-stat is greater than the test level 0.05. Accordingly this residual series is white noise series.

The Model Prediction Results.

Figure 4. White noise test.

Figure 5. Fitting and forecasting results.
The prediction residual and relative error results in the following table:

| No. | Observation value | Prediction value | Residual error | Relative error |
|-----|-------------------|------------------|----------------|---------------|
| 46  | 190.25            | 191.53           | 1.28           | 0.67%         |
| 47  | 190.05            | 191.05           | 1              | 0.53%         |
| 48  | 190.11            | 190.47           | 0.36           | 0.19%         |
| 49  | 191.57            | 189.81           | -1.76          | 0.92%         |
| 50  | 193.29            | 190.51           | -2.78          | 1.43%         |

**Conclusion**

This paper utilizes the ARIMA model to analyze the bolt axial force monitoring data from foundation pit project of China Zun Tower.

The results show that:

1. Using ARIMA model can construct the model for the data of bolt axial force and have achieved the good results.
2. ARIMA model has higher precision prediction when it is used for short-term forecasting. But precision of prediction will decline over time.

**References**

[1] Lin Xun, Application of Time Series Analysis to the Building’s Deformation Monitor, Ji Lin University, (2005).

[2] Box G E P, Jenkins G M, Time Series Analysis: Forecasting and Control Holden Day, Journal of the Royal Statistical Society, 134(3), (1976).

[3] Ling Ai. Methods and Techniques on the Short-term Forecasting of Time Series. East China Normal University (2010).

[4] Yang Hao-xiong, Hu Jing. Forecast of fresh agricultural products demand based on the ARIMA model, J. Guangdong Agricultural Sciences, 11-0193-03, (2013).

[5] Zou Jiting, Application of Time Series in the Analysis of the Building Subsidence Monitor Data, Journal of Beijing University of Civil Engineering and Architecture, 25-2, (2009).

[6] Zhang Jie, Liu Xiao-ming, He Yu-long. Application of ARIMA Model in Forecasting Traffic Accidents. Journal of Beijing University of Technology, 33-12, (2007).

[7] Chi Qi-shui, Liu Xiao-xue. Study on the application of ARIMA Model in forecasting China’s coal consumption [J].Energy Research and Information, 23-2, (2007).

[8] Wu Jiang-huai. Data Analysis and Process of Building Static Deformation Monitoring. Tongji University, (2007).

[9] Ye Bing. The Analysis of Deformation Model and Deformation Forecasting. Tongji University, (2006).
### Appendix

Table 6. ML-03-1 anchor axial force raw data.

| No. Observation | Observation | Data | Data | Data | Data |
|-----------------|-------------|------|------|------|------|
| 01—05           | 196.12      | 197.92 | 194.82 | 197.38 | 192.04 |
| 06—10           | 192.29      | 198   | 193.48 | 197.61 | 194.08 |
| 11—15           | 196.62      | 195.29 | 196.85 | 194.42 | 196.22 |
| 16—20           | 197.89      | 197.82 | 200   | 192.61 | 196.85 |
| 21—25           | 199.02      | 197   | 197.9  | 196.11 | 195.99 |
| 26—30           | 194.65      | 199.55 | 196.31 | 191.52 | 195   |
| 31—35           | 193.57      | 193.1  | 191.82 | 193.28 | 195.22 |
| 36—40           | 199.65      | 194.58 | 193.04 | 194.82 | 194.01 |
| 41—45           | 193.91      | 192.44 | 191.46 | 190.43 | 192.7  |
| 46—50           | 190.25      | 190.05 | 190.11 | 191.57 | 193.29 |