Prediction of Seawall Settlement Based on a Combined LS-ARIMA Model

Peng Qin¹ and Chunmei Cheng²

¹Institute of Hydraulic and Environment Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, China
²Institute of Geomatics and Municipal Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, China

Correspondence should be addressed to Peng Qin; qp021625@163.com

Received 7 April 2017; Revised 27 April 2017; Accepted 4 May 2017; Published 31 May 2017

The analysis and prediction of seawall settlement are important for seawall engineering maintenance and disaster prevention. Based on the measured seawall settlement time series data, a combined LS-ARIMA forecasting model that fits the trend item by the least-square (LS) method and the season item by the differential self-regression moving average (ARIMA) model was proposed in this study. The monitoring data of one seawall project in Zhejiang, China, is taken as an example to verify the model efficiency and prediction ability. The results show that the prediction accuracy of the new combined LS-ARIMA model was high, with the average relative error (ARE) of 0.23%, much better than that of the traditional ARIMA model (ARE = 0.70%) and the GM (1, 1) model (ARE = 33.43%). This new model has clear physical conception and can effectively improve the prediction accuracy, implying that it can fully tap the dynamic information of monitoring data. The proposed model in this study provides a new research idea for data analysis and prediction of the seawall settlement.

1. Introduction

The seawall that is made of local soil and stones is an important water conservancy facility used for typhoon defense and tide resistance in coastal areas. The embankment damage caused by the uneven seawall settlement poses a huge security risk for the economic development in the coastal areas and the personal safety of local residents [1, 2]. Therefore, the effective prediction of the seawall settlement is of great significance for the early warning of the possible deformation as well as the project safety of seawall projects.

The monitoring data of seawall settlement is different from that of reinforced concrete or steel structure engineering such as dams, bridges, or other projects. On one hand, the settlement monitoring data follows a certain law. The majority of Chinese seawall projects are filled by the soil and built on the soft-soil foundation composed of dredged silt, and so the embankment and its foundation soil will consolidate and settle under external force or its own gravity during or after a certain period of construction. This deformation is determined by the constitutive relationship of the soil, leading to the strong regularity and trend of the settlement monitoring data [2, 3]. On the other hand, the settlement monitoring data covers only a short period, with small data amount [3, 4]. Due to the lack of effective supervision, the monitoring works of the seawall settlement in the construction or post-construction period were few in the past few decades, bringing difficulty for the seawall management and maintenance. Only recently has some systematic monitoring system of seawall settlement been gradually established in some areas, with the rapid development of the monitoring technology and much more attention from relevant departments.

The seawall settlement is influenced by many factors such as construction load, pore ratio, water content, internal friction coefficient, cohesion, structure, construction period disturbance, precipitation, and wave erosion, and therefore there was no universal prediction model for the seawall settlement monitoring data [5]. At present, there
are mainly two methods for the analysis and forecasting of the seawall settlement monitoring data: (1) The consolidation coefficient inversion method: The consolidation coefficient of the collected foundation soil is first determined through the laboratory geotechnical experiment, and then the seawall settlement process is calculated by the Terzaghi’s one-dimensional consolidation theory [6]. However, since the condition of laboratory experiment is quite different from the condition and load distribution of site soil, and the laboratory geotechnical experiment is effected by many factors such as soil quality, sample representation, test technique, and calculation method, the consolidation coefficient obtained from the laboratory experiment is quite different from the real situation, which will affect the credibility of the seawall settlement prediction results. (2) The fitting and predicting algorithm: The seawall settlement process is fitted and predicted according to the variation trend of the settlement monitoring data time series, and several methods have been used for model calculation such as the empirical formula method [7], Asaoka method [8], and gray theory [9]. These methods composed of one single prediction algorithm cannot fully explore the information in the seawall settlement monitoring data. For example, the empirical formula method does not consider the influence of external environmental changes on the data during the fitting process; the Asaoka method often causes increasing prediction error due to its sensitive fitting parameters; and the gray theory shows poor predicting performance for the data without strong trend.

Both the above methods have the following disadvantages: the first inversion method is greatly influenced by the environment since its results are based on the collected foundation soil and the human experience on the laboratory geotechnical test; and the second fitting and predicting algorithm composed of one single method cannot fully explore the information of the seawall settlement monitoring data. Therefore, the establishment of a prediction model reflecting the relationship between the seawall settlement mechanism and its influencing factors can be a reasonable way to improve the prediction accuracy. Taking the monitoring data time series of one seawall project in Zhejiang, China, as an example, after analyzing the seawall settlement variation and its inducing factors, a combined LS-ARIMA forecasting model that is composed of the least-square method (LS) and the autoregressive integrated moving average (ARIMA) model was proposed in this study. This new LS-ARIMA model has great advantages because its season item has considered the environmental influence and the model combination method can fully explore the data information compared to traditional methods; therefore it can provide a reasonable and convenient calculation method for the seawall settlement prediction.

2. Basic Theory

2.1. Response Analysis of the Time Series Data. The displacement of seawall settlement time series data controlled by its own geological conditions is of approximately monotonous growth at a large time scale and that controlled by external factors is fluctuated at a small time scale. The curve of the overall displacement of the seawall settlement deformation time series data shows regular stepped change, so the time series data of seawall settlement deformation can be expressed by the additive model as follows:

\[ y_t = f_t + s_t + c_t + e_t, \]  

where \( f_t \) is the trend item reflecting the long-term trend of the time series; \( s_t \) is the season item representing the seasonal variation of the time series data; \( c_t \) is a cycle item reflecting cyclical changes of the time series, which is the nonseasonal periodic fluctuations; and \( e_t \) is a random item reflecting the random changes of the time series, which is the influence of various random factors on the sequence.

The deformation curve of the trend item \( t_t \) is similar to “S” because the displacement variation controlled by the seawall’s own soil condition can be divided into three parts: the accelerated deformation stage, the deformation development stage, and the deformation convergence stage [10, 11]. The season item \( s_t \) means the response of the monitoring data to the regular periodic fluctuations with the change of different seasons, induced by the tide level fluctuation, human activities, and so on [11]. This seasonal change can be simulated using the attribute variable generally, with the value of 0 or 1. That is, 1 is set at the time points affected by different seasons and 0 for those not affected. The following model is used to simulate the season item that is affected by \( i \) influencing factors:

\[ s_t = \sum e_i X_{i_t}, \quad i > 0, \]

where \( e_i \) is the parameter and \( X_{i_t} \) is the attribute variable, and

\[ X_{i_t} = \begin{cases} 1, & \text{If } t \text{ is within the season} \\ 0, & \text{If } t \text{ is not in the season.} \end{cases} \]

The above seasonal model has more universal significance than previous periodic variation models, because the seasonal model can simulate a purely periodic change if sufficient attribute variables are used [12].

2.2. Model Establishment Using the Addition Model. The seawall settlement is mainly affected by the seasonal rainfall and tidal level variation, and so the cycle item of the seawall monitoring data can be added to the season item, making the displacement of total seawall settlement composed of only the trend item and the season item [13]. Assuming that there is no impact of random items, so the outliers were removed and the noise of the settlement monitoring data was filtered. Thus the addition model of the time series can be generalized as follows:

\[ Y(t) = \alpha(t) + \beta(t), \]

where \( Y(t) \) is the displacement time series, \( \alpha(t) \) is the trend function, and \( \beta(t) \) is the season function.

The trend item that shifted by time can be fitted and predicted by the LS method. The original season item time series data can then be obtained after the removal of the fitted
trend item from the total displacement time series based on the addition model. The displacement of the season item is a complex nonlinear sequence influenced by multiple factors and can be predicted by the ARIMA regression prediction method. Therefore the total prediction value of the seawall settlement is obtained by superimposing the predicted values of both trend item and the season item.

3. The Combined LS-ARIMA Model Establishment

3.1. Fitting the Trend Item by the LS Method. The basic idea of the least-square curve fitting is to calculate the square sums of the difference between all the data points and their fitting points, and it is the vertical distance between the calculated point and the estimated point for the case of the two-dimensional figure error. The resultant least-square fitting curve will be obtained once the fitted curve makes the square sums of errors minimum [14].

As to the seawall settlement monitoring data \{\{(t_i, y_i) \mid i = 1, 2, \ldots, m\}\}, where \(t_i\) is the time series and \(y_i\) is the monitoring data. The fitting curve is set as \(y = \alpha(t)\), then the error distance of the \(i\)th point \(\alpha(t_i) - y_i\) and the square sum of all points is \(\sum (\alpha(t_i) - y_i)^2\). The fitting curve of the trend item \(y = \alpha(t)\) can be obtained after solving the corresponding function parameter when \(\sum (\alpha(t_i) - y_i)^2\) achieving minimum value.

The key to the LS method is to solve the minimum value of the square sums of errors. Suppose there are some known data points \(\{(t_i, y_i) \mid i = 1, 2, \ldots, m\}\) and \(\Phi\) is the function class composed of the polynomials with number less than \(m\). Now seeking \(\alpha_k(t_i) = \sum_{k=0}^{m} a_k t_i^k \in \Phi\), (5) exists.

\[
I = \sum_{i=0}^{m} (\alpha(t_i) - y_i)^2 = \sum_{i=0}^{m} (\sum_{k=0}^{m} a_k t_i^k - y_i)^2 = \min (5)
\]

Equation (5) is a multivariate function, and only when its corresponding derivative is equal to zero will its minimum exist; that is,

\[
\frac{\partial I}{\partial a_j} = \frac{2}{\sum_{i=0}^{m} (\sum_{k=0}^{m} a_k t_i^k - y_i)} t_i^j = 0, \quad j = 1, 2, \ldots, n, (6)
\]

\[
\sum_{k=0}^{m} (\sum_{i=0}^{m} a_k t_i^k) a_k = \sum_{i=0}^{m} t_i^j y_i, \quad j = 1, 2, \ldots, n. \quad (7)
\]

Equation (7) is called the normal equation, through which the unique solution \(a_k \mid k = 1, 2, \ldots, n\) can be obtained. Then the fitting function \(\alpha_k(t_i) = \sum_{k=0}^{m} a_k t_i^k\) is determined and the polynomial will be fitted when \(k > 1\) [14, 15].

Once determining the predictive equation \(\alpha(t)\) of the trend item of the seawall settlement time series, the trend item value at some future time can then be calculated.

3.2. Fitting the Season Item by the ARIMA Model. The season item of seawall settlement time series is nonstationary and not oscillated at a fixed value [16]. Although it is not a random sequence, some relevant information of the nonstationary time series can still be extracted, but the prediction accuracy by the autoregressive moving average (ARMA) model is low in this case. Usually the time series is transformed into a stationary sequence by the first-order or multiple differential method before the ARMA model is established, which will effectively improve the prediction accuracy.

The season item of the seawall settlement time series is denoted as \(\{(t_i, s_i) \mid i = 1, 2, \ldots, m\}\), where \(t\) is the time interval and \(s\) is the value of the season item time series. Record \(V\) as the differential operator, and \(s_i\) can be called a \(d\)-order homogeneous nonstable sequence if there exists a non-negative integer making \(V^d s_i = \omega_i\) a stationary time series.

If the new sequence fulfills the conditions \(\theta(B) a_i = \theta(B^d) a_i\), then the summed autoregressive moving average sequence ARIMA \((p, d, q)\) can be solved, where

\[
\begin{align*}
\theta(B) a_i &= \theta(B^d) a_i = \omega_i, \quad (8) \\
\end{align*}
\]

\([a_i]\) is the white noise sequence, and the parameters to be estimated include \(B\), the hysteresis operator, \(\varphi_1, \varphi_2, \ldots, \varphi_h\), the autoregressive coefficient, and \(\theta_1, \theta_2, \ldots, \theta_p\), the average slip coefficient.

Based on the above model building ideas, the modeling steps for the ARIMA model are concluded as follows.

Step 1 (stationary test of time series). The season item of the seawall settlement time series is first obtained after removing the trend item from the total seawall settlement time series based on the addition model (1). Then the sequence stability is determined by the correlation graph test and the ADF unit root test.

Step 2 (determine the model orders). The ARIMA model orders \(p\) and \(q\) are initially identified based on the truncation and tailing of both the autocorrelation and partial autocorrelation coefficients of the stationary sequence. Then the values of \(p\) and \(q\) are determined by means of the order criterion [16], and the Akaike information criterion (AIC) was used in this study.

Step 3 (parameters estimation and model test). The model parameters are estimated by the maximum likelihood estimation method. The model is tested after the parameter estimation, which is to test whether \([a_i]\) is smooth white noise. The model establishment will be completed if the test is passed; otherwise, the model parameters must be reestimated [16].

Step 4 (model prediction). After determining the model prediction equation \(\beta(t)\), the season item of seawall settlement monitoring data at some future time can then be estimated.

3.3. Seawater Settlement Prediction by the Combined LS-ARIMA Model. Firstly, the trend item value and the season
4 Mathematical Problems in Engineering

Seawall settlement monitoring data time series

Stationary test

Pattern recognition of the ARIMA model

Determination of ARIMA model orders \(p\) and \(q\) by AIC

White noise test

Total displacement prediction of the seawall settlement

Figure 1: Flow chart of the combined LS-ARIMA model calculation.

The least squares fitting data

Trend item time series

Season item time series

No

Yes

item value of the seawall settlement monitoring data at some time are calculated according to the prediction equation \(\alpha(t)\) and \(\beta(t)\), respectively. Secondly, the total displacement of the seawall settlement is predicted by superimposing the trend item and the season item time series. The combined LS-ARIMA model is carried out by MATLAB [17] and the flow chart of the model is shown in Figure 1.

4. Engineering Example

The seawall settlement monitoring data during the construction period of one seawall project in Zhejiang, China, was used as an example to verify the reliability of the combined LS-ARIMA model. This embankment was constructed on a silt and silty clay foundation with an earth-rock structure and a drainage procedure. 60 monitoring data points of this seawall settlement project at the 0 +166 measuring point during the construction period from July 2010 to October 2010 were used. The seawall settlement is in the deformation development stage at the chosen time and the interval of the settlement monitoring data is 3 days. The first 50 points were used as the calculation data and the last 10 points as the validation data. The monitoring data points were numbered by their measuring time for calculation convenience, and the seawall settlement time series was shown in Figure 2.

Firstly, the original seawall settlement time series data was fitted by the LS method, resulting in the fitting function

\[
\alpha(t) = -0.0144x^2 + 1.7635x + 2.723,
\]

which is also the fitted trend item time series of the seawall settlement. Secondly, the original season item time series was obtained after removing the trend item from the total seawall settlement time series based on the addition model and was also fitted by the ARIMA model. The total seawall settlement time series, the trend item time series, and the season item time series were all shown in Figure 3. It can be seen from Figure 3 that the total time series reflect the natural variation characteristics of the seawall settlement monitoring data (Figure 3(a)). The trend item time series fitted by the least squares method show smooth growth, well reflecting the trend characteristics of the total time series (Figure 3(b)). The season item time series after removal of the trend item exhibit a strong volatility due to the complex effects such as seasonal rainfall, tidal level change, and human activities (Figure 3(c)). Therefore, the season item time series will be fitted by the ARIMA model next.

The stability of the season item time series of the seawall settlement was tested using the autocorrelation function (AC) and partial autocorrelation (PAC) method (Figure 4). It can be seen that the AC and PAC values of most data points are beyond the allowable range, implying the nonsmoothness of the time series and the requirement of differential computation.

The first-order differential sequence of the cumulative displacement of the seawall settlement season item was shown in Figure 5, and the AC and PAC coefficients after the sequence test were shown in Figure 6. The results show that the first-order differential sequence of the season item time series tends to be stabilized, with the AC and PAC coefficients within the allowable range. Therefore the coefficient \(d\) in the ARIMA \((p, d, q)\) model was set to 1.

The order of \(p\) and \(q\) of the ARIMA model was then determined using the AIC method. It was found that the AIC value was minimum when \(p = 4\) and \(q = 2\), with the value of 0.1213. Therefore the coefficient \(p\) was set to 4 and \(q\) was set to 2, and the ARIMA model can then be written as

\[
\varphi(B) \omega_i = \theta(B) a_i,
\]

where

\[
\omega = (1 - B),
\]

\[
\varphi(B) = 1 - 0.154B + 0.7107B^2 - 0.1773B^3 + 0.09145B^4,
\]

\[
\theta(B) = 1 + 0.1213B + 0.09145B^2 - 0.0144B^3 + 0.0123B^4.
\]

\[
\begin{array}{c}
\text{Cumulative displacement (mm)} \\
5 & 0 & 10 & 20 & 30 & 40 & 50 \\
\end{array}
\]

\(B\) is the delays operator. Therefore, the ARIMA \((4, 1, 2)\) model can be written as

\[
\omega_i = \varphi(B) \omega_i - \theta(B) a_i = 2.723 - 0.0144x^2 + 1.7635x,
\]

Figure 2: The seawall settlement time series.
Mathematical Problems in Engineering

(a) Total time series

(b) Trend item time series

(c) Season item time series

Figure 3: The total, trend item, and season item time series of the seawall settlement.

Figure 4: The AC and PAC coefficients of the season item time series.

\[
\theta(B) = 1 + 0.3711B + 0.8503B^2 - 1.032B^3 - 0.448B^4 - 0.8085B^5,
\]

and \(a_t\) is the white noise time series.

The residuals of the season item time series fitting by the ARIMA \((4, 1, 2)\) model were calculated and shown in Figure 7, indicating the high model fitting accuracy (residual < 2 mm). Figure 7 shows that the fitting residual distribution of the model has only one maximum value at about 4 mm and four points with residuals between \(-2\) and \(+2\), which are also within a reasonable error range though. Except that, the fitting error of all samples is between the range of \(-1\) and \(+1\), mainly concentrated near 0. All the fitting results fulfill the prediction requirements, implying that the season item time series with complex fluctuations can be precisely fitted by the ARIMA model.

The total predicted values by the 10 validation data points of the seawall settlement displacement by the combined LS-ARIMA model were calculated by adding the validated trend item and season item prediction results, which were also compared with the measured values, predicted values of single ARIMA model, and GM \((1, 1)\) model from previous researches \([5, 9]\) (Table 1).
Figure 6: The AC and PAC coefficients of the first-order differential sequence.

Table 1: Comparison of the LS-ARIMA, ARIMA, and GM (1, 1) model results based on validation data.

<table>
<thead>
<tr>
<th>Point number</th>
<th>Measured value (mm)</th>
<th>LS-ARIMA model (mm)</th>
<th>Relative error (%)</th>
<th>ARIMA model (mm)</th>
<th>Relative error (%)</th>
<th>GM (1, 1) model (mm)</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>51</td>
<td>54.42</td>
<td>54.51</td>
<td>0.17</td>
<td>54.12</td>
<td>0.55</td>
<td>65.95</td>
<td>21.19</td>
</tr>
<tr>
<td>52</td>
<td>54.44</td>
<td>54.49</td>
<td><strong>0.09</strong></td>
<td>54.63</td>
<td><strong>0.35</strong></td>
<td>67.65</td>
<td>24.27</td>
</tr>
<tr>
<td>53</td>
<td>54.55</td>
<td>54.68</td>
<td>0.24</td>
<td>54.78</td>
<td>0.42</td>
<td>69.40</td>
<td>27.22</td>
</tr>
<tr>
<td>54</td>
<td>54.81</td>
<td>54.95</td>
<td>0.26</td>
<td>55.25</td>
<td>0.80</td>
<td>71.19</td>
<td>29.88</td>
</tr>
<tr>
<td>55</td>
<td>55.20</td>
<td>55.28</td>
<td>0.14</td>
<td>55.91</td>
<td>1.28</td>
<td>73.02</td>
<td>32.29</td>
</tr>
<tr>
<td>56</td>
<td>55.61</td>
<td>55.72</td>
<td>0.20</td>
<td>55.81</td>
<td>0.36</td>
<td>74.91</td>
<td>34.70</td>
</tr>
<tr>
<td>57</td>
<td>56.06</td>
<td>56.13</td>
<td>0.13</td>
<td>56.33</td>
<td>0.48</td>
<td>76.84</td>
<td>37.07</td>
</tr>
<tr>
<td>58</td>
<td>56.50</td>
<td>56.74</td>
<td>0.42</td>
<td>56.83</td>
<td>0.58</td>
<td>78.82</td>
<td>39.51</td>
</tr>
<tr>
<td>59</td>
<td>56.75</td>
<td>56.87</td>
<td>0.22</td>
<td>57.34</td>
<td>1.04</td>
<td>80.86</td>
<td>42.48</td>
</tr>
<tr>
<td>60</td>
<td>56.92</td>
<td>57.19</td>
<td><strong>0.47</strong></td>
<td>57.58</td>
<td><strong>1.16</strong></td>
<td>82.94</td>
<td>45.72</td>
</tr>
</tbody>
</table>

Note: the relative error is calculated by $\Delta_k = |y - y_1|/y$, wherein $y$ is the measured value, and $y_1$ is the predicted value.

Figure 7: The fitting residual of the season item time series by ARIMA model.

It can be seen that the largest relative error of the LS-ARIMA model is 0.47%, much lower than that of traditional ARIMA model (1.16%), while the majority of relative errors from GM (1, 1) model is beyond 20%. The average relative error $\Delta$ of the three models was calculated using $\Delta = (1/n) \sum_{k=1}^{n} \Delta_k$, wherein $\Delta_k$ is the relative error, and $n$ is the prediction length. Results show that the average relative error (ARE) of the LS-ARIMA model is 0.23%, much lower compared to the ARIMA model (0.70%) and the GM (1, 1) model (33.43%). The combined LS-ARIMA model has significantly improved the prediction accuracy by decreasing the ARE of 0.47% and 33.2% compared to the traditional ARIMA model and GM (1, 1) model. The GM (1, 1) model is more suitable for predicting the time series with strong trend items, but not ideal for the seawall settlement data prediction in this study.

5. Conclusion

(1) A combined LS-ARIMA prediction model that is combining the least-square (LS) method and the autoregressive moving average (ARIMA) model was proposed in this study, which shows good prediction
results when applied in one example of seawall settlement project in Zhejiang, China. The LS-ARIMA model that is fitting the trend item of seawall settlement time series by the LS method and fitting the season item by ARIMA model has a robust antinoise performance and can achieve better prediction accuracy than the single ARIMA and GM (1, 1) model. The proposed model can fully excavate the nonlinear dynamic information of the seawall settlement time series data, thus providing a new research idea for the prediction and analysis of the seawall settlement monitoring data.

(2) For a specific project, through the method in this study we can analyze and predict the measured settlement data at different measuring points and then draw the s–t relationship curve between the settlement displacement and time, and even the n–s relationship curve between the settlement displacement and multiple points at the same time. The large displacement difference between adjacent points means the probability of seawall crack at that time, judging from which the hidden dangers of the seawall project can be predicted, preprocessed, and even eliminated.

(3) The prediction accuracy of the LS-ARIMA model is greatly affected by the chosen the least-square order number n, and the ARIMA model orders p and q, and therefore further study needs to be carried out concerning how to optimize these parameters.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The study was in part supported by the research fund of Zhejiang Province water resources department (no. RC1534), China, the university visiting scholar “teacher professional development project” of Zhejiang Province in 2016 (no. FX2016059, no. FX2016060), and the higher education teaching reform project of Zhejiang Province in 2016 (no. kg20160504). Thanks are due to the Reclamation Technology Center of Zhejiang Province (RTCZP) for providing the data. Special thanks are due to Professor Qin Zhihai from ZJWEU for his valuable suggestions.

References
