Prediction of Operating Abnormality Rate of Charging Pile Based on Generalized AR(q) Combined Regression

Xu Xin¹, Fu Jun¹, Sun Zhijie¹, Li Xuemei¹, Zhou Guopeng¹
¹ NCEPRI(Huadian Electric Power Research Institute Co., Ltd.,) Beijing 100045, China

Abstract. The stable operation of charging pile is related to the entire operation efficiency of the charging network of electric vehicles so the prediction of charging pile operation abnormality rate can help the operational department to make operational decisions in advance. This paper uses the electric vehicle charging network operating date in the north of Hebei province, based on the feature of the anomalies records of charging pile, to combine the generalized AR(q) model and the regression model and to predict the abnormality rate of electric vehicle charging network in the north of Hebei province. It is predicted that the average absolute error is 0.0044 and the acceptable prediction effect can be obtained.

1 Introduction

Under the guidance of China's "13th Five-Year Plan" on that the electric vehicle charging infrastructure needs to be constructed in advance and the smart and efficient charging network system with vehicle piles, the coverage area of electric vehicle charging piles is continuously expanding, and the rapidly developing electric vehicle industry needs the stable running charging network. The research and prediction of the charging pile abnormality rate are of great significance on the operation of charging networks and the development of the industry. This article will carry out study on the stability of the charging network system, predict and analyze the abnormality rate of the charging pile operation. At the same time, this paper notices the limit of the time series tools and proposes the abnormality rate prediction model applicable for the charging pile system with the combination of the regression analysis.

2 Charging pile abnormal characteristics

Set the duration of the i-th charging pile has the abnormality on the t-th day as \( t \), then the abnormality rate \( r_t \) on the t-th day of this pile will be:

\[
g_t = \frac{\sum t \cdot c_t}{24}
\]

\( c_t \) includes the sum of fault duration, off-line duration and outage duration for this pile per day. The average abnormality rate \( g_t \) of the charging network system is:

\[
g_t = \frac{1}{24} \sum t \cdot c_t
\]

In which \( \sum c_t \) is the total duration of failed piles on the tth day, and \( N \) is the total number of piles. In the following analysis, we mainly focused on duration of the abnormality \( c_t \). The average abnormal duration \( g_t \) of one pile per day:

\[
g_t = \frac{\sum t \cdot c_t}{N}
\]

And the average daily abnormality rate of the system:

\[
g_t = \frac{1}{24} \sum t \cdot c_t
\]

It is easy to calculate.

Charging pile abnormality duration data has its own characteristics. (1) The length of abnormality duration is more significant than whether the abnormality duration is zero.

From the point of view of charging network operation and maintenance, most charging piles are subject to abnormalities due to various failures, offline or outage. Therefore, the problem of abnormality is actually an issue of abnormality duration, and the research focuses on those charging piles that are abnormal. (2) The data platform only responds to abnormal charging piles. If an abnormality occurs in the charging pile, the data platform will automatically display the abnormality start and recovery time; otherwise, the platform will not display any abnormal information. Hence, the data recorded by the platform is the data of the pile where the abnormality has occurred and these data can be observed. (3) The data amount of the abnormal charging piles is relatively less than that of all charging piles so it is relatively easy to handle abnormal charging pile data than all charging pile data. (4) There may be more than one charging pile will have abnormality on the same day, that is, on the t day, it may be observed that \( g_t, g_{t+1}, \ldots, g_{t+S} \), \( S > 1 \). Therefore, observations of S times of abnormality duration (S charging piles) can be obtained on day t. (5) The occurrence of abnormalities depends on the history of abnormalities. From the experience of operations, the
abnormality of charging piles is highly correlated with the past abnormalities. So the abnormality duration of the charging pile in the past has great influence on the current abnormality duration length. These characteristics showed that, $G_{it}$ depends on $G_{it-1}$, $G_{it-2}$, ... relatively strongly, where $G_{it-1}$, $G_{it-2}$, ... are not necessarily corresponding to the abnormality duration on one day or two day before t-th day but ate the duration when the first or first two abnormalities of the i-th pile occurred on t-th day. $G_{it-1}$, $G_{it-2}$, ... are the observations of the first or first two abnormalities of the i-th pile and the time intervals for these observations are not necessarily equal.

In addition to the abnormality history, there are other factors that affect the length of the abnormality duration of the charging pile. First of all, the operation and maintenance capabilities of the charging pile operation and maintenance department will have an impact on the abnormality of the charging pile. If the operation and maintenance department has a relatively high level of operation and maintenance business or can deal with the abnormality on site timely ($D_{it}$), then the abnormality duration of the charging pile within their scope of jurisdiction should be shorter. Secondly, the operation and maintenance capabilities of the charging pile manufacturers also affect the abnormality of the charging pile. If the hardware quality of the equipment supplied is relatively high or the spare parts can be replaced timely or they can deal with the abnormality on site timely ($CW_{it}$), then the length of abnormality duration of the charging pile produced by this manufacturer shall be shorter. Third, special hours. In the holidays, the use frequency of the charging piles may be different from that in the usual; During the national important meeting or other special period, the operation and maintenance department will take a major measure of power protection; Sometimes the power grid will also be outage for maintenance in a certain period. Therefore, holidays, special hours, wind direction and force, temperature ($TD_{it}$, representing the minimum temperature), PM2.5 (represented by $PM_{it}$), etc. may affect the charging pile hardware and further affect the abnormality duration.

### 3 Charging pile abnormality prediction models

According to the factors affecting the abnormality of the charging pile, we construct the prediction models as follows.

#### 3.1 Generalized AR(q) Model

According to the characteristics of the charging pile's abnormality data, the duration of the abnormality of i-th pile occurred on t-th day has the unbalanced panel data characteristics in form, but we know from long-term operation knowledge that the abnormality duration mainly depends on the length of the abnormality duration of the pile in the past. The spatial effect showed by different piles is not very obvious. We deal with the data of several piles observed in the same day as several observation values but we ignore their spatial effects. Therefore, the model can be considered as:

$$G_{it} = \epsilon_0 + \epsilon_1 G_{it-1} + \cdots + \epsilon_q G_{it-q} + \epsilon_t$$  \hspace{0.5cm} (1)

Where, $\epsilon_0, \epsilon_1, \cdots, \epsilon_q$ are parameters and $\epsilon_t$ is error term. The subscript i corresponds to the i-th pile and can be any one among N piles. But not all the N piles on the same day t can be observed and only the plies with abnormalities can be observed. At the same time, as in fact, it is impossible that all the piles are in normal state so at least one pile per day can be observed. If the spatial effect of i is ignored then only the number of observations will increases. Formula (1) is formally the AR(q) model.

In model (1), t corresponds to day t; but t-1,...,t-q in model (1) represent the days that the previous first...q-th abnormalities of i-th pile. The time interval of day t and t-1 in the model may be more than one day and for simple, it still be represented by that in (1). So the time intervals in the time series in the model (1) are generally unequal and are different from the traditional equal-interval time series. Due to formal similarities and differences in intervals, model (1) is a generalized AR(q) model and the processing needs to consider three points:

1. AR (q) order selection. The size of q in the model (1) is roughly determined, that is, it is to consider which previous abnormality is more appropriate for the charging pile.
2. Coefficient t-test. We use the least square estimation to estimate the model (1) and remove the insignificant coefficient in the model (1).
3. R2 changes. By focusing on changes in R, the goodness of fit of the model is analyzed.

#### 3.2 Operation and weather impact index

Model (1) is similar to AR(q) and has a disadvantage in application. When the data cannot provide the abnormality duration data for the first q times, so it cannot be calculated and cannot be predicted by model (1). In order to be able to predict the abnormality duration of each day, other prediction methods must be introduced to make up for the shortcomings of AR(q).

In addition to the historical abnormality factors of charging piles, there are factors such as the operation and maintenance capabilities of the operating department and manufacturers, special hours, weather, and others affect abnormality duration. Among them, weather factors are natural factors, and other factors are non-natural factors. Therefore, their impact on the abnormality duration of charging piles is divided into two categories which are analyzed with the regression model, respectively.

#### 3.2.1 Operation and maintenance impact index
The operation and maintenance impact index mainly reflects the impact of the operation and maintenance management capabilities of the operation department and the manufacturer on the abnormality duration of the charging pile and is described with the regression model as follows:

\[
G_{it} = c_0 + c_1 Y_{it,1} + \cdots + c_k Y_{it,k} + e_{it} \quad (2)
\]

Among them, \(Y_{ij}, j = 1, 2, \ldots, k\) represent the \(j\)-th operation and maintenance factor affecting the \(i\)-th pile on the \(t\)-th day. There are \(k\) factors in total, including the timeliness ratio of on-site treatment of the operation and maintenance department \(D_{it}\), the hardware quality of the equipment \(Q_{it}\), the timeliness ratio of on-site maintenance \(C_{it}\), holidays \(D_{it}\), major electricity protection days \(D_{it}\), and power outages duration \(W_{it}\) and so on. The first half of model (2) is defined as the operation and maintenance impact index \(Y_{it}\):

\[
Y_{it} = a_0 + a_1 Y_{it,1} + \cdots + a_k Y_{it,k} \quad (3)
\]

\(Y_{it}\) is not only used for operation department management, quality supervision of manufacturers, but also for abnormality duration prediction.

### 3.2.2 Weather impact index

The weather impact index mainly reflects the impact of the weather on the abnormality duration of the charging pile and is described with the regression model as follows:

\[
G_{it} = b_0 + b_1 T_{it,1} + \cdots + b_p T_{it,p} + e_{it} \quad (4)
\]

Among them, \(T_{it,j}, j = 1, 2, \ldots, p\) represent \(j\)-th weather factor affecting the \(i\)-th pile on the \(t\)-th day and there are a total of \(p\) factors, including rainy and snowy, wind direction and force, temperature \(T_{it}\), and PM2.5 and so on. The first half of model (3) is defined as the weather impact index \(T_{it}\):

\[
T_{it} = b_0 + b_1 T_{it,1} + \cdots + b_p T_{it,p} \quad (5)
\]

\(T_{it}\) can be used for abnormality duration predictions.

### 3.3 Comprehensive impact index

The operation and maintenance impact index \(Y_{it}\) combines with the weather impact index \(T_{it}\) to form a comprehensive impact index to reflect the total effect of operation and maintenance and weather on the abnormality duration of the charging pile and it is shown as follows with the regression model:

\[
G_{it} = d_0 + d_1 Y_{it} + d_2 T_{it} + e_{it} \quad (6)
\]

The sum of the first three terms of model (6) is called the comprehensive index \(F_{it}\):

\[
F_{it} = d_0 + d_1 Y_{it} + d_2 T_{it} \quad (7)
\]

The operation and maintenance impact index, the weather impact index, and the comprehensive impact index will not use the data of the \(i\)-th pile before the \(t\)-th day, and the operation and maintenance impact index and the weather impact index of each pile every day can be obtained. Therefore, the use of the comprehensive impact index the prediction of the abnormality duration of the charging pile will not be affected by the time delay and at the same time, due to the combination of the two index, it will have better prediction effect than that of the single operation and maintenance index or the weather impact index.

### 3.4 Generalized AR(q) combined regression prediction model

From the analysis of the generalized AR(q) model, it is known that there will be an abnormality in the charging pile every day. If there are \(q\) anomalies of the \(i\)-th charging pile before the \(t\)-th day, then the generalized AR(q) model (1) can be used to predict its abnormality duration; If the number of abnormality occurred before \(t\) days is less than \(q\), then it is unable to use (1) for predictions. However, the comprehensive prediction model (6) or the comprehensive prediction index (7) can be used to predict the abnormality duration. In fields related to energy prediction, the use of combined prediction techniques based on time series prediction is a relatively practical and effective choice. We will use the generalized AR(q) model (1) as the basis and combine the comprehensive prediction index model (6) to predict the abnormality duration. The prediction model is constructed as follows:

\[
G_{it} = \begin{cases} 
\varepsilon_{it} = c_0 + c_1 G_{it-1} + \cdots + c_q G_{it-q} + e_{it} \\
\varepsilon_{it} = d_0 + d_1 Y_{it} + d_2 T_{it} + e_{it} 
\end{cases} \quad (8)
\]

Once the coefficient estimates of (1) and (6) are obtained, the prediction formula can be written as:

\[
G_{it} = \begin{cases} 
\varepsilon_{it} = c_0 + c_1 G_{it-1} + \cdots + c_q G_{it-q} \\
\varepsilon_{it} = d_0 + d_1 Y_{it} + d_2 T_{it} 
\end{cases} \quad (9)
\]

Based on the prediction model (8) or (9), the predicted value of the abnormality duration of the charging pile per day can be obtained.

### 4 Estimation and prediction results

In order to obtain the prediction model, we use the operation data of the charging piles from January 2016 to August 2017 in the north of Hebei province to estimate the model (8), and then use the data of September and October of 2017 to carry out predictive analysis.

#### 4.1 Model estimation

The estimation of the model (8) includes the estimation of the generalized AR(q) model (1), and the estimation of the operation and maintenance impact index (2), the weather impact index (4), and the comprehensive impact index (6).
Here mainly consider to determine the appropriate order, choose significant historical abnormality variables, and refer to changes in R2. AIC is decreasing and R2 is increasing with the increase of q, as shown in Table 1. But in terms of application, the increase of q will also cause more missing values and more charging piles cannot be predicted. The decreasing speed of AIC and the increasing speed of R2 begin to slow down from q=30, so we choose q=35 as the reference order.

The corresponding estimation formula is

\[ q=30, \text{ so we choose } q=35 \text{ as the reference order.} \]

In order to ensure the validity of the least squares estimation as much as possible, we remove the insignificant lagged variable in the generalized AR(35) by the t-test, and finally obtain the estimation result AR(28) in Table 2, and R2=0.5126.

### 4.1.2 Operation and maintenance impact index estimation

The operation and maintenance data of the operation and maintenance departments of five cities and ten charging pile manufacturers in the north of Hebei province are used to estimate the results of the operation and maintenance impact index as shown in Table 3 and R2=0.0782.

### 4.1.3 Weather impact index estimation

The weather data from five cities in the north of Hebei province is used to estimate the results of the weather impact index as shown in Table 4 and R2 = 0.0973.

### 4.1.4 Comprehensive impact index estimate

The estimated operation and maintenance impact index (11) and weather impact index (12) are used for the further estimation of the comprehensive impact index. The results are shown in Table 5, and R2 = 0.1850.

### 4.1.5 Prediction model estimation

The comprehensive impact index corresponds to R2=0.1850 which is greater than the operation and maintenance impact index as R2=0.0782 and the weather impact index as R2=0.0973.

### 4.2 Prediction

The abnormalities duration of the charging pile can be predicted with the use of generalized AR(28) combined regression model (14). In order to compare the prediction results, we use the data of the charging piles from the September and October of 2017 in the north of Hebei province to estimate the results of the weather impact index as shown in Table 4 and R2 = 0.0973.

### Table 1 Generalized AR(q) Model Order Changes

| q  | 5   | 10  | 15  | 20  | 25  | 30  | 35  | 40  |
|----|-----|-----|-----|-----|-----|-----|-----|-----|
| R² | 0.39| 0.425| 0.435| 0.449| 0.516| 0.535| 0.546|     |
| AIC| 129331| 92450| 96628| 46663| 33794| 24925| 18956| 15160|

### Table 2 Generalized AR(q) Estimation Results

| Term | Estimate | Std Error | t Ratio | Prob>|t| |
|------|----------|-----------|---------|------|
| Intercept | 0.9671 | 0.2602 | 3.7171 | 0.0002 |
| \(G_{it-1}\) | 0.3818 | 0.0152 | 25.0431 | 0.0000 |
| \(G_{it-2}\) | 0.1317 | 0.0164 | 8.0289 | 0.0000 |
| \(G_{it-3}\) | 0.1184 | 0.0166 | 7.1461 | 0.0000 |
| \(G_{it-4}\) | 0.0797 | 0.0158 | 5.0417 | 0.0000 |
| \(G_{it-5}\) | 0.0889 | 0.0147 | 6.0294 | 0.0000 |
| \(G_{it-6}\) | 0.0511 | 0.0142 | 3.6028 | 0.0003 |
| \(G_{it-7}\) | 0.0364 | 0.0132 | 2.7501 | 0.0060 |

The corresponding estimation formula is

\[ \text{AR}(28)_{it} = 0.96 + 0.38G_{it-1} + 0.13G_{it-2} + 0.11G_{it-3} + 0.07G_{it-4} + 0.08G_{it-5} + 0.05G_{it-6} + 0.03G_{it-7} + \] (10)

### Table 3 Operation and Maintenance Index Estimation Results

| Estimate | Std Error | t Ratio | Prob>|t| |
|----------|-----------|---------|------|
| 16.5615  | 0.4424    | 37.4369 | 0.0000 |
| -0.2340  | 0.0286    | -8.1904 | 0.0000 |
| -0.2343  | 0.0557    | -4.2078 | 0.0000 |
| -2.4181  | 0.4638    | -5.2132 | 0.0000 |
| 0.4174   | 0.0091    | 45.9097 | 0.0000 |

The corresponding estimation formula is

\[ Y_{it} = 16.56 - 0.23D_{it} + 0.2343G_{it} + 2.41C_{it} + 0.41W_{it} \] (11)

### Table 4 Weather Impact Index Estimation

| Term     | Estimate | Std Error | t Ratio | Prob>|t| |
|----------|----------|-----------|---------|------|
| Intercept| 16.56    | 0.3818    | 43.6346 | 0.0000 |
| PM_{it}  | 0.1000   | 0.0042    | 24.1821 | 0.0000 |
| TD_{it}  | -0.2012  | 0.0042    | -47.8228| 0.0000 |

The corresponding estimation formula is

\[ T_{it} = 13.6346 + 0.0099PM_{it} - 0.2012TD_{it} \] (12)

### Table 5 Comprehensive Impact Index Estimation

| Term     | Estimate | Std Error | t Ratio | Prob>|t| |
|----------|----------|-----------|---------|------|
| Intercept| 14.9623  | 0.3755    | 40.0602 | 0.0000 |
| \(T_{it}\) | 1.0599   | 0.0199    | 53.1409 | 0.0000 |
| \(Y_{it}\) | 1.0500   | 0.0179    | 58.7176 | 0.0000 |

The corresponding estimation formula is

\[ Y_{it} = -14.96 + 1.06T_{it} + 0.049Y_{it} \] (13)

The comprehensive impact index corresponds to R2=0.1850 which is greater than the operation and maintenance impact index as R2=0.0782 and the weather impact index as R2=0.0973.

### 4.2 Prediction

The abnormalities duration of the charging pile can be predicted with the use of generalized AR(28) combined regression model (14). In order to compare the prediction results, we use the data of the charging piles from the September and October of 2017 in the north of Hebei province.
Hebei province for the analysis. The actual abnormality rate and the predicted abnormality rate are shown as Figure 1 and the predicted abnormality rate is basically consistent with the actual abnormality rate. For the predicted error of abnormality rate, we focus on the absolute error and the predicted absolute error is shown in Figure 2. The average absolute error is 0.0044, and the median is 0.0025, which can meet the needs of the actual prediction of the operation department.

![Figure 1 Actual abnormality rate and predicted abnormality rate](image1)

**Figure 1** Actual abnormality rate and predicted abnormality rate

![Figure 2 Predicted absolute error](image2)

**Figure 2** Predicted absolute error

5 Conclusion

Based on the operation data of the charging piles from January 2016 to August 2017 in the north of Hebei province, we used the AR(q) model's relatively accurate advantages in prediction and considered the generalized AR(q) model according to the characteristics of the abnormality duration data of the charging piles, avoiding the defect that the AR(q) model depends on lagged variables and combining the regression models. This generalized AR(q) combined regression prediction model is used for actual prediction and can be basically consistent with the actual abnormality rate. The absolute error effect generated can be accepted by the operation department.

REFERENCES

1. Wang Feilong, Sun Kaihang, Li Yanqing. Demand prediction of electric vehicle charging stations based on user classification [J]. Heilongjiang Electric Power, 2013, 35(2): 132-134.
2. Huang Qixin, Lu Zhuwei, Yang Zhengli. Research on prediction of electric vehicles charging loading

[J]. Mechanical and Electrical Information, 2016, 27(489): 140-141.
3. Chen Lidan, Nie Yongquan, Zhongqing. Prediction model of electric vehicle charging load based on trip chain [J]. Transactions of China Electrotechnical Society, 2015, 30(4): 216-225.
4. Zhang Yuyu, Zhang Yuxin. ARMA time series-based electric vehicle charging station operation state prediction and charging power demand model [J]. Electrotechnical Application, 2016, 35(21): 52-58.
5. Arias, M.B., Bae, S. Electric vehicle charging demand forecasting model based on big data technologies[J]. Applied Energy, 2016, 183(12): 327-339.
6. Majidpour, M., Qiu, C., Chu, P., Pota, HR, Gadh, R. Forecasting the EV charging load based on customer profile or station measurement[J]. Applied Energy, 2016, 163(2): 134-141.
7. Peng Zhixing, Bao Changjun, Zhao Yang, Yi Honggang, Tang Shaowen, Yu Hao, Chen N Feng. ARIMA product seasonal model and its application in the prediction of infectious diseases[J]. Application of Statistics and Management, 2008, 27(2): 362-368.
8. Bin, E. Econometrics of Panel Data Methods and Applications [M]. Oxford University Press, 2017.
9. Box, E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M. Time series analysis: forecasting and control[M]. John Wiley & Sons, Inc., 2016.
10. Deb, C., Zhang, F., Yang, J., Lee, SE, Shah, KW A review on time series forecasting techniques for building energy consumption[J]. Renewable and Sustainable Energy Reviews, 2017, 74(7): 902-924.
11. Chen Yan, Shi Zhihui. Mixed financial time series prediction model based on Adaboost and regularized ELM and Its Application [J]. Application of Statistics and Management, 2017, 36(1): 113-125.
12. Lu Lin, Xu Wei, Xiangyue, Zhang Yi, Xiong Jun. Research on optimal configuration of multi-regional charging piles based on Markov chain charging load prediction[J]. Engineering Science and Technology, 2017, 49(3): 170-178.