ISSN 2616-5775 Vol. 4, Issue 2: 11-19, DOI: 10.25236/AJCIS.2021.040203

Prediction of Optical Power Data Based on Optimized ARIMA Model

Chaoju Hu[1, *], Yunpeng Gao[1]

[1]North China Electric Power University, Baoding, China
**Corresponding author*

Abstract: *In order to make the operation of optical fiber protection system more stable and improve the accuracy of time series prediction for a small amount of optical power data samples, this paper presents an ARIMA model prediction method based on improved wavelet decomposition. This method uses the improved wavelet decomposition is multistage discrete wavelet decomposition SWT and improved it. Different from the conventional method of decomposition and reconstruction of signals, this paper directly uses wavelet decomposition coefficient for modeling, which simplifies the process of input data construction and reduces the data loss caused by data reconstruction. ARIMA is autoregressive integrated moving Average model. Building a combination model and using the data to conduct simulation experiments. Experimental results verify that the prediction accuracy of this optimization model is higher than that of the ARIMA model alone and prove that this model is superior to a small amount of sample optical power data.*

Keywords: *the time series, wavelet decomposition, ARIMA.*

1. Introduction

Fiber optic communication has the advantages of fast transmission speed, weak information attenuation and strong resistance to external interference. Therefore, optical fiber is widely used in various fields. In this premise, the signal security in optical fiber communication has become a more critical problem. Therefore, optical fiber protection system (OLP) arises at the historic moment. At present, optical fiber protection system has been widely used in electric power, network communication, railway communication and other fields [1,2,3]. In OLP, the system automatically detects and collects the optical power data in the optical network in real time. If the decay value in the channel exceeds the normal decay value range, it will issue a warning and switch the current channel to the standby channel [1]. However, this switching mode is bound to cause communication interruption for a short time, and even when the OLP switching module fails, it will directly lead to the interruption of network communication. In order to enhance the security and stability of optical fiber protection system, a prediction method based on machine learning algorithm is proposed in this paper. A small number of optical power data samples collected in real time are used to predict the optical power in the short term. According to the prediction results, the optical fiber protection system can react in advance and improve the stability of communication network.

Which method can be used to accurately predict the result has become a problem worth studying. From the perspective of prediction, optical power data belongs to one-dimensional discrete time series. Currently, more methods used in the field of machine learning to predict time series include Differential Autoregressive Moving Average (ARIMA), Long and Short Term Memory Network (LSTM), Support Vector Machine (SVM), etc. These methods all have their own characteristics. For example, the ARIMA model requires that the historical data of optical power is in a state of steady change, and the prediction premise is that the data itself has some inertia-like characteristics, so ARIMA is often used to predict the data with strong regularity. LSTM has good nonlinear prediction ability, can consider to the various factors affecting the historical data and the prediction precision is [4], currently use LSTM effect is better, but this model for large sample data of predicted results is superior to the small sample data, some small sample data, the forecast effect is very poor, at the same time, this model takes more computing resources. The generalization performance of SVM is related to the selection of kernel function and parameters, but the prediction results will lose the influence of peak value [5]. At the same time, there is no good theoretical support for the selection of construction kernel function and parameters up to now. In view of the above methods, this paper believes that a single model can no longer meet the requirements of

ISSN 2616-5775 Vol. 4, Issue 2: 11-19, DOI: 10.25236/AJCIS.2021.040203

accurate prediction. Using the experimental results in recent years for reference, the combined model can better improve the prediction accuracy [6]. Be Reddy et al. used the combined model of ARIMA to predict the atmospheric temperature and obtained accurate experimental results [7].

Wavelet decomposition can extract multi-scale characteristics from the sequence and transform the study of complex multi-scale time series into the study of multiple time series of different frequencies. It is an effective means to capture abnormal time series data by using different targeted methods for effective analysis of different sequences obtained by decomposition [8]. However, in the conventional wavelet decomposition process, the data is decomposed first to get the decomposition coefficient, then the data components are reconstructed and finally the modeling is carried out. However, in the process of reconstruction of the data components, the data operation amount will be increased and a small amount of original data features will be lost, which has an impact on the prediction accuracy of the final result. However, in the process of optical power prediction, the characteristics of the data component does not make sense the accuracy of the final results have practical significance, thus to improve the wavelet decomposition, not refactoring data component, direct use of decomposition coefficient model, can effectively reduce data computation and reduce the loss of data in order to improve the prediction precision.

Therefore, an ARIMA model based on improved wavelet decomposition is proposed in this paper to predict and verify a small amount of optical power data.

2. Organization of the Text

Before you to forecast data, the data preprocessing operations, in addition to the stationarity of data processing (the model data stationarity method using the finite difference method), to the wavelet decomposition, the data after the wavelet decomposition, the original data can be decomposed into high frequency and low frequency data, then respectively for high frequency and low frequency data modeling using ARIMA model, and then the predictive results of the ARIMA model in wavelet reconstruction to get the final prediction results.

Combination model flow:

a) Wavelet decomposition: a group of original data is decomposed by SWT (the number of decomposition layers is 1, and the wavelet function is DB1) to obtain two sets of data, namely, the high frequency coefficient and the low frequency coefficient.

b) Data preprocessing: missing value processing and data resampling are performed on the two sets of data respectively, and then the two sets of data are transformed into one-dimensional time series.

c) ARIMA prediction, the processed high frequency and low frequency coefficients were modeled respectively, and the prediction results were obtained by differential processing (first-order difference) and selecting appropriate parameter values (p,q).

d) Wavelet reconstruction. The prediction results of the two models were respectively reconstructed by wavelet to get the final prediction results.

The model selected in this paper is based on the characteristics of optical power data. According to the characteristics of the original data, the data is preprocessed appropriately, so that it can better meet the requirements of the prediction model for input data, so as to make the prediction results more accurate. Optical power data have strong stability and regularity, accompanied by occasional outliers. Therefore, under a small amount of historical optical power data, ARIMA model is the most appropriate choice. At the same time, in order to improve the prediction accuracy, before data preprocessing, the original data is decomposed by wavelet, and then the optical power data is subdivided into high-frequency and low-frequency data, so that the features of the original data are extracted according to different scales, and then the prediction is made. Be Reddy et al. used ARIMA model based on wavelet decomposition to predict the earth atmospheric data, and the results showed that this model could make the prediction results more accurate [7].

2.1. Wavelet Decomposition

Wavelet analysis is a method of data preprocessing and data analysis. Its purpose is to express the signal more clearly and concisely, and to highlight the time-frequency characteristics of the signal more prominently. In this model, multi-stage discrete wavelet decomposition (SWT) is used to decompose the original time series into two arrays, namely, the high-frequency coefficient array and the low-frequency coefficient array. Mallat algorithm is generally used to decompose the original signal [9]. Through the Mallat algorithm, the original data is decomposed into two groups of coefficient arrays representing the features of the original signal. The calculation formulas are shown in Formula (1) and Formula (2) [8].

$$A_{j+1,k} = \sum_m h_0(m-2k)A_{j,m} \tag{1}$$

$$D_{j+1,k} = \sum_m h_1(m-2k)A_{j,m} \tag{2}$$

The j represents decomposition scale. The k, m represents the translation variable. The $A_{j,k}$ represents the approximate coefficient. The $D_{j,k}$ represents the detail coefficient. The h_1 and h_0 represents low pass and high pass filters, respectively.

2.2. Wavelet Reconstruction

The original sequence can be reconstructed by using the two groups of wavelet coefficients after decomposition. The reconstruction of wavelet coefficients is shown in Equation (3) [8].

$$A_{j-1,m} = \sum_m h_0(m-2k)A_{j,k} + \sum_m h_1(m-2k)D_{j,k} \tag{3}$$

The process of wavelet reconstruction is the process of inverse operation of two groups of coefficient data according to the standard of wavelet function.

2.3. ARIMA Model

2.3.1. Introduction to The Model

The ARMA model is a very important model for studying time series, and has been applied in many fields, including climate, medicine, power and other fields, and is mainly used to predict time series [12-16]. This model is actually a linear combination of AR model and MA model [10]. The ARIMA model used in this article is actually an ARMA model with differential operations. The purpose of difference is to make the original unstationary time series become stationary [11].

The AR model (autoregressive model) represents the relationship between the current value and the historical value, and uses the historical time data of the variable itself to predict itself. The most important parameter of this model is p, that is, the current value is related to the historical value of order p. The formula (4) of the autoregressive model is:

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t \tag{4}$$

y_t represents the current value, μ represents the constant term, p represents the order, γ_i represents the autocorrelation coefficient, and ε_t represents the error.

The MA model (moving average model) represents the accumulation of error terms in the autoregressive model, and the most important parameter is q, which is the accumulation of q-order error terms. The formula (5) of the moving average model is:

$$y_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{5}$$

y_t represents the current value, μ represents the constant term, q represents the order, θ_i represents the error term coefficient, and ε_t represents the error.

The principle of the ARIMA model has been explained in the above, that is, the linear group sum of AR and MA, that is, its formula (6) is:

ISSN 2616-5775 Vol. 4, Issue 2: 11-19, DOI: 10.25236/AJCIS.2021.040203

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{6}$$

2.3.2. Modeling Steps

a) Data preprocessing

First, the original data is processed with missing values and data resampling. The missing data is averaged and resampled at the desired time interval to form a time series that meets the requirements of the experiment.

b) Data smoothing

The ARIMA model requires the input data to maintain stability. The so-called stationarity means that the curve obtained by the sample data can still maintain the continuity of the existing form "inertia" for a period of time in the future, so the stability test must be performed before the data is input to the model. The commonly used stationarity detection method is the ADF single-root detection method. When the data does not pass the test, the difference method is used to make the data stable.

c) Model order

After the data is smoothed, the input parameters of the model should be selected, that is, the order of the model. The main parameters of the model, namely p and q, are determined by the ACF (autocorrelation coefficient) and PACF (partial autocorrelation coefficient) of the stationary data. The method is to determine the values of p and q by observing the autocorrelation coefficient graph and the partial autocorrelation coefficient graph. The criteria for determining the parameters are shown in Table 1.

Table 1: Order code determined

Model	ACF	PACF
AR(p)	attenuation tends to zero	fall to the confidence interval after order p
MA(q)	fall to the confidence interval after order q	attenuation tends to zero
ARMA(p,q)	attenuation tends to zero after order q	attenuation tends to zero after order p

However, in some original data time series, the values of p and q cannot be determined directly by observing the autocorrelation coefficient graph and the partial autocorrelation coefficient graph. It is necessary to select multiple sets of parameter combinations that meet the parameter determination criteria according to the Akaike information criterion (Akaike Information Criterion, AIC) and Bayesian Information Criterion (Bayesian Information Criterion, BIC) to find the optimal parameter values to determine the final model order.

d) Model prediction and evaluation

The data and the determined optimal parameters are input into the model, and the results are obtained and fitted with the original data. The standard deviation (MSE) and root mean square error (RMSE) are used to measure the quality of the model. The more the RMSE and MSE values are smaller, it proves that the prediction effect of the model is better [10].

3. Experiment Procedure

3.1. Data Preparation

The original data used in this experiment comes from the optical power data provided by an optical fiber network protection system in Baoding. A piece of continuous 90 data from the full-sky optical power data on July 1 is sampled as samples, of which the first 80 data are the training set, and the last 10 data are the test set. The original data is shown in Figure 1.

ISSN 2616-5775 Vol. 4, Issue 2: 11-19, DOI: 10.25236/AJCIS.2021.040203

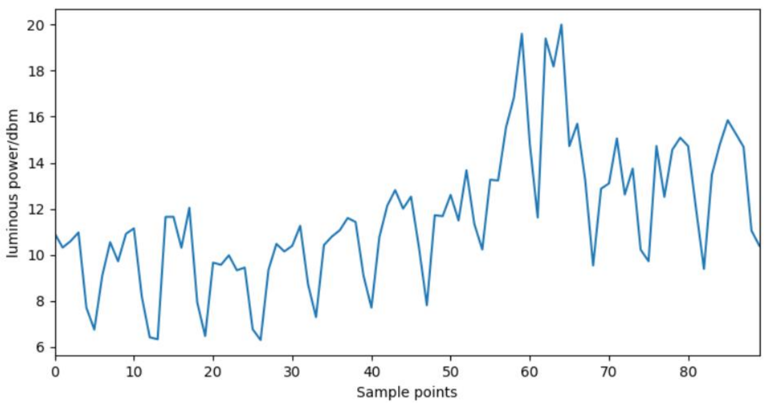

Figure 1: Original data

3.2. Wavelet Decomposition

According to the characteristics of optical power data, it needs to be wavelet decomposed into different components for subsequent analysis. The collected original data is decomposed by wavelet, the number of decomposition layers is 1, and the wavelet function is db1. After decomposition, the high-frequency coefficient components and low-frequency coefficient components of the original data are obtained, as shown in Figure 2 and Figure 3.

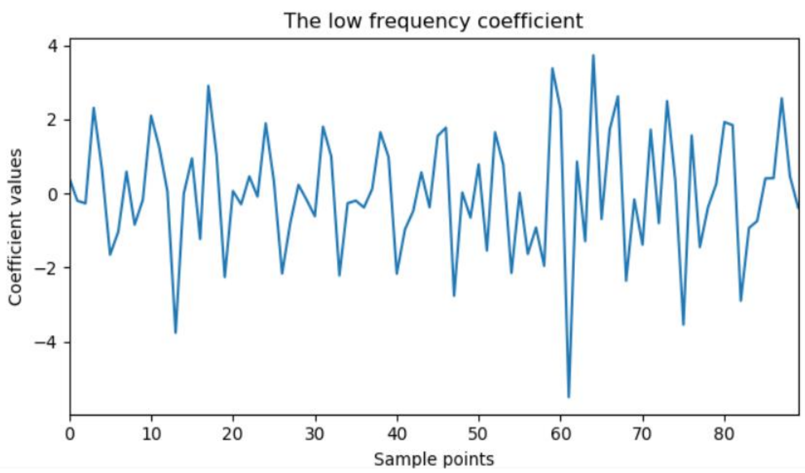

Figure 2: Low-frequency coefficient component

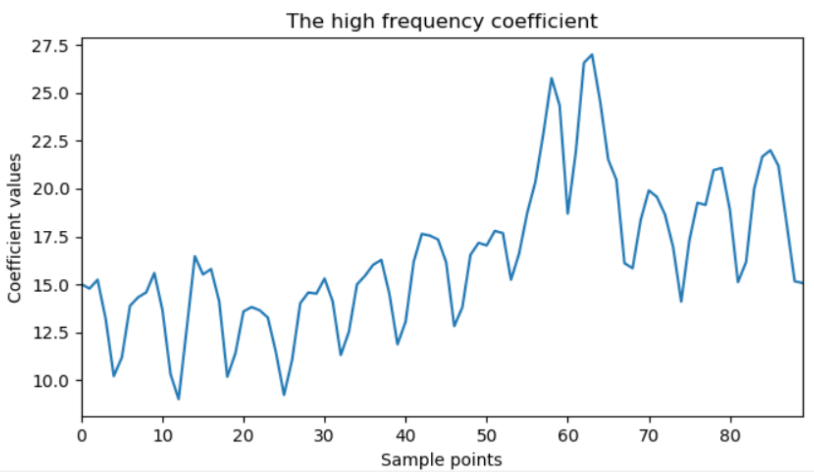

Figure 3: High-frequency coefficient component

ISSN 2616-5775 Vol. 4, Issue 2: 11-19, DOI: 10.25236/AJCIS.2021.040203

3.3. Data Smoothing

The low-frequency and high-frequency coefficient components are obtained through wavelet decomposition of the original data. These two sets of coefficient components also reflect the details and trend characteristics of the original data. Observing the characteristics of the two sets of components separately, obviously does not meet the ARIMA model's requirements for data stationarity, so we use the difference method to smooth the two sets of components.

The first-order difference is performed on the high-frequency and low-frequency components respectively, and then the ADF unit root detection method is used to detect the stability of the data. After the first-order difference, the high-frequency and low-frequency components are tested by ADF unit root, and the p-values are -4.721 and -5.087 ($P<0.001$) respectively, which are statistically significant. Both results have passed the test, and the data has been stabilized.

3.4. Select Suitable Parameter

After differential operation, both components achieve data stability. Then the two components are modeled, and the autocorrelation coefficients and partial autocorrelation coefficients of the low-frequency components are shown in Figure 4 and Figure 5.

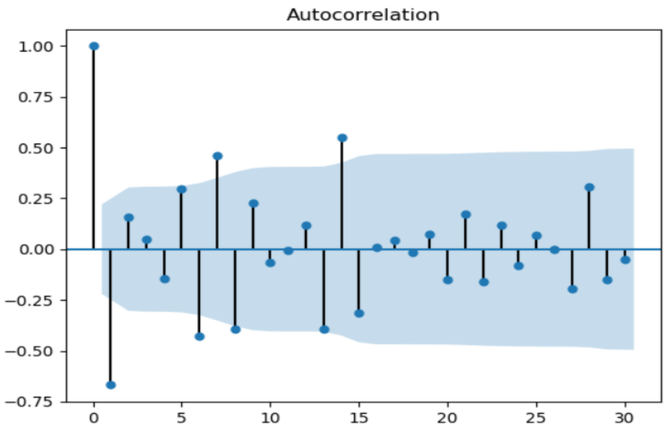

Figure 4: Low frequency autocorrelation coefficient

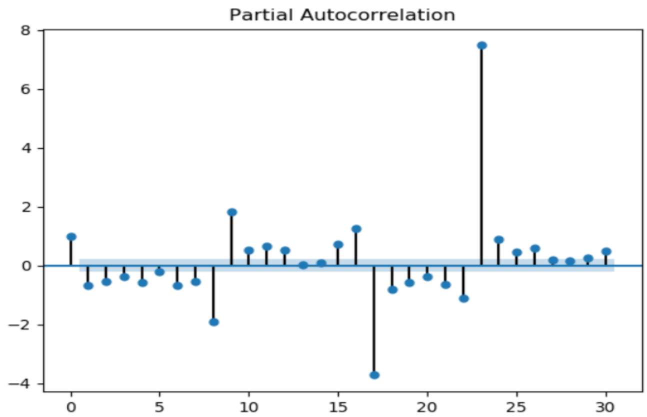

Figure 5: Low frequency partial autocorrelation coefficient

It can be seen from Figure 4 that the autocorrelation coefficients of the low-frequency coefficients begin to decay rapidly after the first order and all fall within the confidence interval. From Figure 5 you can see that the partial autocorrelation coefficients of the low-frequency coefficients start to oscillate from the seventh order. type. Therefore, for low frequency components, ARIMA (7,1,1) is used for modeling and prediction.

The autocorrelation coefficients and partial autocorrelation coefficients of high-frequency components are shown in Figure 6 and Figure 7.

ISSN 2616-5775 Vol. 4, Issue 2: 11-19, DOI: 10.25236/AJCIS.2021.040203

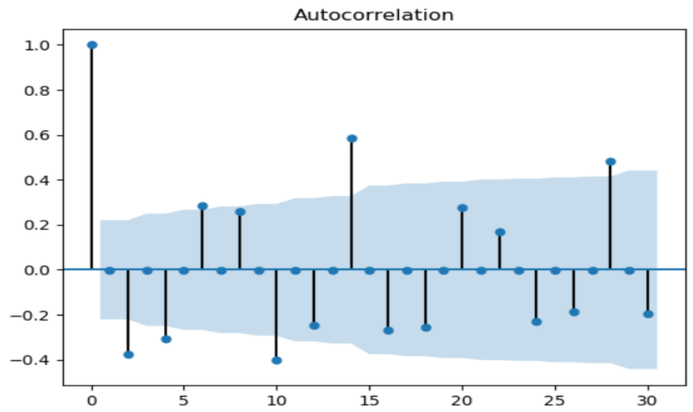

Figure 6: High frequency autocorrelation coefficient

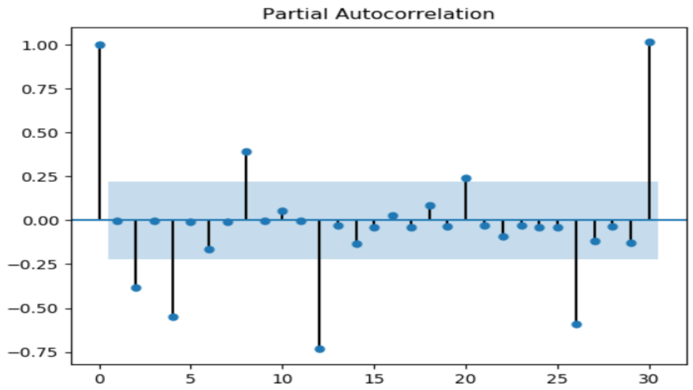

Figure 7: High frequency partial autocorrelation coefficient

From Figure 6 we can observe that the autocorrelation coefficient of the high-frequency component begins to show an oscillating type after order 0. It can be seen from Figure 7 that the partial autocorrelation coefficient shows an oscillating type after order 7 and ARIMA (7,1,0) to model high-frequency components.

It is worth mentioning that the two coefficient components in this experiment can be directly determined by the autocorrelation coefficient graph and the partial autocorrelation coefficient graph, so it is no longer necessary to use the AIC and BIC criteria to select the optimal parameters.

3.5. Fitting Result

The low-frequency and high-frequency components are modeled by ARIMA (7,1,1) and ARIMA(7,1,0) respectively, and the result obtained is then subjected to wavelet inverse operation to obtain the final predicted value. Fit the predicted results with the original data and compare them with the fitting results of the ARIMA model without wavelet decomposition, as shown in Figure 8 and Figure 9.

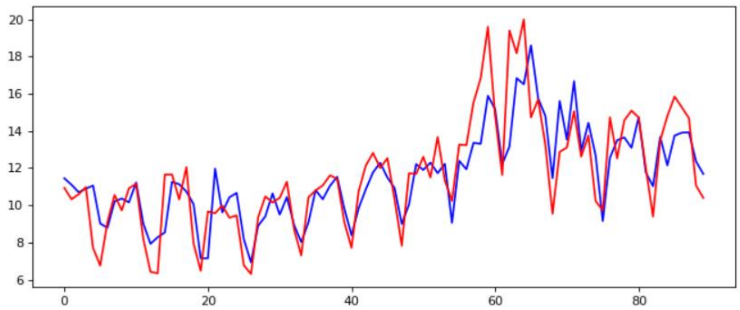

Figure 8: The result of single model fitting

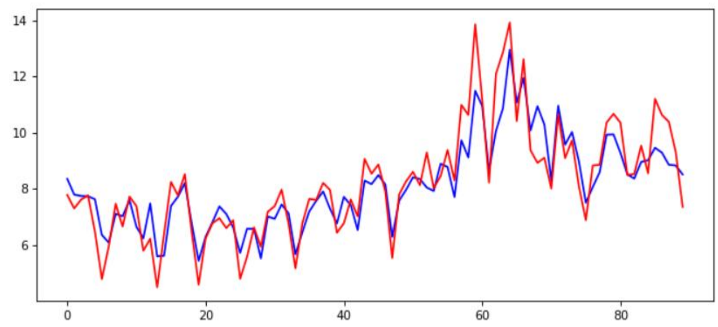

Figure 9: Based on the fitting results of wavelet decomposition model

The red in the figure represents the original data, and the blue represents the prediction result. The results show that the standard deviation (MSE) and root mean square error (RMSE) of the fitting results of the unoptimized model (2.654 and 1.629, respectively. The standard deviation (MSE) and root mean square error (RMSE) of the fitting results based on the wavelet decomposition model RMSE) are 0.661 and 0.813.

Finally, use the trained model to make predictions, set the prediction range to the last 12, and then compare the prediction results with the original data, as shown in Figure 10 and Figure 11.

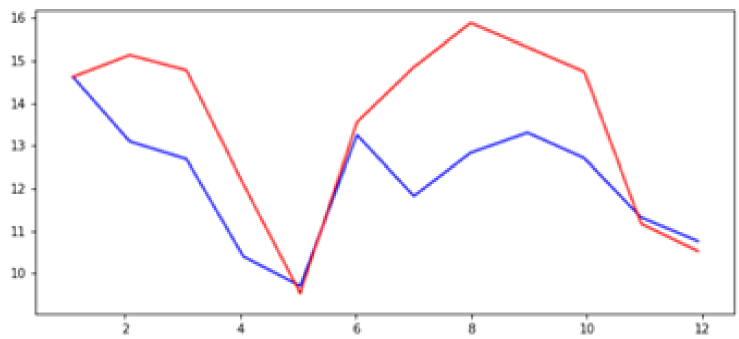

Figure 10: Single model prediction results

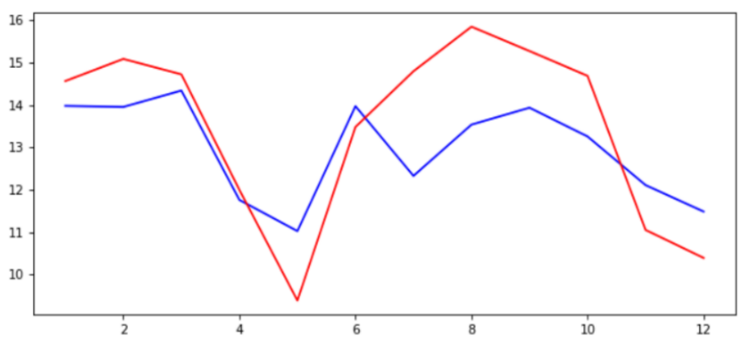

Figure 11: Based on the prediction results of wavelet decomposition

4. Conclusion

Taking the characteristics of optical power data into account, this paper proposes an ARIMA model based on improved wavelet decomposition, which aims to improve the prediction accuracy of optical power under a small amount of raw data. Through experiments, and the results of using a single ARIMA model for comparison, this combined model has higher prediction accuracy under a small amount of optical power data.

In addition, in the process of wavelet decomposition, this paper uses the improved multi-dimensional discrete wavelet decomposition method to decompose the original data, and directly uses the component coefficients for the model training process, simplifying the process of reconstructing component data from the component coefficients, and improving the efficiency of training the model. Adapting to a small amount of optical power data not only improves the performance of the prediction model, but also reduces the loss of original data characteristics caused by the reconstruction process, and improves the prediction accuracy. This model provides a new idea and method for the study of medium and short-term optical power prediction in optical protection systems. At present, the prediction of optical power in the channel is still in its infancy in the communication network, and there are many limitations. In the field of optical power prediction, there are still many problems waiting to be solved and improved, and more exploration and exploration are needed.

Finally, this model is only limited to the prediction effect of the optical power data under a small number of samples, and does not consider the effect of the model under a large sample. At the same time, the wavelet decomposition process is also limited to the case of decomposing one layer. Later, we will consider adding an LSTM model to further eliminate non-linear factors and improve model prediction accuracy.

References

[1] Hanchuan Fan. (2013)Application of Optical Fiber Line Protection System in Communication Network. Practical Electronics, 12, 146-146.

[2] Hua Qin, Gang Li, Jian Yang, et al. (2020) Application of uninterpolated Loss-type optical fiber protection system in electric Power Communication Network. Inner Mongolia Science Technology & Economy, 5, 97-99.

[3] Li Yumei. (2014)Security Protection of railway optical communication System by using intelligent optical fiber protection module. Technology Innovation and Application, 26, 72-72.

[4] Hua Qin, Gang Li, Jian Yang, et al. (2020) Application of uninterpolated Loss-type optical fiber protection system in electric Power Communication Network. Inner Mongolia Science Technology & Economy, 5, 97-99.

[5] Amin M A A, Hoque M A, et al. (2019) Comparison of ARIMA and SVM for Short-term Load Forecasting. 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON). Jaipur, India, 1-6.

[6] W. Cheng, Y. Zhou, Y. Guo, et al. (2019) Research on prediction method based on ARIMA-BP combination model. Porc of the 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE). Xiamen, China, 663-666.

[7] CN Babu, BE Reddy. (2012)Prediction data mining at global mean temperature using variants of ARIMA model. International Conference on Progress in Engineering Science and Management, 256-260.

[8] Xiang Li, Ling Peng, Jing Shao, et al. (2016)Study on air pollution prediction based on wavelet decomposition and ARMA Model. Environmental Engineering, 8, 110-113.

[9] Mallat S G. (1989) A theory for multiresolution signal decomposition: The wavelet representation. Pattern Analysis and Machine Intelligence. IEEE Transactions on, 11, 674-693.

[10] Hao Fu, Shengda Wang, Shan Jiang, et al. (2019) Optical power prediction method based on ARMA-LSTM. Optical Communication Technology, 43, 20-23.

[11] Shiguang Peng, Xianhui Geng. (2020)Based on the ARIMA and GM (1, 1) model of China's soybean imports and imports of prediction. Soybean science, 8, 1-13.

[12] Chadsuthi S, Iamsirithaworn S, Triampo W, et al. (2015)Modeling seasonal influenza transmission and its association with climate factors in Thailand using time-series and ARIMAX anal yses. Comput Math Methods Med, 436495.

[13] Liu L, Luan RS, Yin F, et al. (2016) predicting the incidence of hand, foot and mouth disease in Sichuan Province, China using the ARIMA model. EPidemiol Infect, 144, 144-151.

[14] T Wang, J Liu, Y Zhou, et al. (2016)Prevalence of hemorrhagic fever with renal syndrome in Yi yuan County, China, 2005-2014. BMC Infect Dis, 16, 69.

[15] C Narendra Babu, B Eswara Reddy. (2012)Predictive data mining of global mean temperatures using variants of the ARIMA model. Porc of IEEE International Conference on Progress in Engineering, Science and Management (ICAESM-2012). Tamil Nadu, Nagapattinam, 256-260.

[16] Z Zou, X Wu, Z Zhao, et al. (2016) Short-term Power Load forecasting based on the combination of BP Neural Network and ARIMA. Porc of the IEEE 4th Congress on Information Technology and Electromechanical Engineering (ITOEC). Chongqing, China, 1671-1674.