

Short-term Ultraviolet Index Forecasting Using ARIMA Model

Shuangyue Xiao¹, Shengchi Liu², Li Liu^(✉)
{ 1546536983@qq.com¹, 1830997169@qq.com², link_liuli@hotmail.com^(✉)}

Department of Information Science and Engineering Dalian Polytechnic University Dalian, P. R. China

Abstract. Solar energy is recognized as an ideal renewable energy. Solar photovoltaic power generation is an important way to use solar energy. Photovoltaic power generation can alleviate the existing energy crisis and alleviate various environmental problems. As a clean and renewable energy, photovoltaic power generation is playing a prominent role in smart micro-grid. The intensity of solar radiation is fluctuating, so the successful grid connection of solar power stations requires accurate power prediction. This paper focuses on the prediction of ultraviolet power index to solve the problem of power prediction of photovoltaic power station. Based on the analysis of cumulative autoregressive moving average model, the stationary of ultraviolet index time series was detected, the order of ultraviolet index model was estimated, and the ARIMA model of ultraviolet index was determined. The prediction accuracy of the model is determined by the root mean square error (RMSE) and mean absolute error (MAE).

Keywords: ARIMA model; ultraviolet index forecasting; the smart micro-grid.

1 Introduction

Smart microgrid is a kind of localized smart grid system [1]. With the rise of distributed generation technology, it is committed to solving the two major problems of distributed generation power quality control and grid reliability control. Microgrid system integrates power generation, energy storage, load and control devices [2], aiming to convert distributed energy into electric energy locally, and then supply local load nearby [3]. Microgrid itself is either independent of large power grid, or interactive with large power grid, or both, which has the characteristics of flexible distribution and complex control [4].

Based on solar radiation, photovoltaic power generation, which realizes the direct conversion of light energy and electric energy, is an important distributed generation technology [5]. In terms of decentralized energy supply, it has obvious advantages over wind power, hydropower and biomass power generation technologies [6]. Photovoltaic power generation module is an indispensable power source category for microgrid system which emphasizes local clean energy utilization [7]. Since the concept of "microgrid" was proposed at the beginning of this century, the microgrid systems developed by various countries or

✉ Corresponding author: Li Liu

Fund Project: Dalian Science and Technology Innovation Fund (2020JJ26GX029)

The Natural Science Foundation of Educational Department of Liaoning Province (Grant: J2020053)

organizations all contain photovoltaic modules [8]. However, due to the natural intermittent characteristics of photovoltaic power generation, microgrid control is more complex than traditional grid control. As the main force of distributed generation, photovoltaic power generation can accurately predict photovoltaic power, which has important practical significance for the smooth operation of smart microgrid [9].

The common statistical methods include multiple linear regression algorithm, artificial neural network (ANN) algorithm, support vector machine (SVM) algorithm and grey theory algorithm. When the multiple linear regression algorithm is used to predict the photovoltaic power generation system, only the influence of radiation and temperature on the system is considered, and the temperature model is relatively simple, and the influence of wind speed on the temperature of photovoltaic modules is not reflected in the model, so the prediction result error of the model is large; the artificial neural network algorithm is used to predict the photovoltaic system [10], and the error is minimum, and the artificial neural network model can be used the accuracy of test data can be improved continuously, and new samples can be changed at any time. But the artificial neural network algorithm is easy to fall into the local minimum problem, so that the problem can not get the optimal solution, and the generalization ability is not strong, which has a certain impact on the prediction accuracy; the least squares support vector machine algorithm, Vapnik et al [11,12], proposed the standard support vector machine (SVM) algorithm, which has been successfully applied in many fields, but the standard SVM algorithm has the disadvantages of slow calculation speed and robustness The gray theory algorithm is used to predict the power of photovoltaic power generation system [13], which only considers the data fitting of the output power and re fitting on the residual to reduce the error of power prediction. When the weather is unstable, the prediction error is large.

Generally speaking, multiple linear regression and grey theory are used in photovoltaic output prediction. Although the method is relatively simple, the prediction error is also large, and the prediction ability is weak when the weather changes. The neural network model has strong randomness and accuracy in predicting the output of photovoltaic system, but the process is relatively complicated.

In this paper, a method based on ARIMA is proposed the ultraviolet index prediction of the model, according to the historical ultraviolet index data, samples and analyzes the ultraviolet index according to the time series, and uses the historical data to model and predict the future changes. First, test the stationarity of the historical ultraviolet index data time series. Second, estimate the order of the model on the basis of the stationarity of the ultraviolet index, and get the ARIMA model, it can simplify the complexity of modeling, reduce the complexity of system design procedures, simplify the cumbersome operation, improve the system operation speed and improve the prediction accuracy.

2 ARIMA Modeling Principles and Steps

2.1 Introduction to the ARIMA Model

In this paper, ARIMA model is used in R software for predicting the wind speed [14].

AR in the ARIMA (p, d, q) model is called the autoregressive component; p is the order of the autoregressive component; MA is the moving average component; q is the order of the moving average component; I is Difference; d is the number of differences in the time series that makes the time series stationary [15].

To see that if the Time Series is Stationary or not we take help of ACF (Auto Correlation Function) and PACF (Partial Auto Correlation Function). Time series is a set of interpretations ordered (sequentially) according to the time they were observed.

2.2 Modeling Steps

ARIMA modeling generally has the following 4 steps:

(1) Inspection and processing of data series.

Before modeling, the data series must be tested, analyzed, and preprocessed to meet the requirements of normal and stable zero mean. After processing, the data sequence meets the modeling requirements, and then the next step is analyzed.

(2) Model ordering.

Determine p and q according to the AIC criterion, and calculate the AIC value corresponding to each order. The smaller the AIC value, the better. AIC is known as the Akaike's information criteria (AIC) is outlined as [16]:

$$AIC = -2\ln(L) + 2(p + q + k + 1). \quad (1)$$

where L is the likelihood function, k is the number of parameters.

(3) The estimation of parameters $\varphi_1, \varphi_2, \dots, \varphi_p$ and $\theta_1, \theta_2, \dots, \theta_q$.

This study uses Marquardt nonlinear least squares method.

(4) Model diagnosis.

Check whether the residual sequence is a white noise sequence. If it is, it means that the residual is purely caused by random interference, then the model is available, otherwise it needs to be re-modeled.

3 Experimental Procedure and Result

This paper takes the actual ultraviolet index data of Shanghai, China in the dark sky website of the United States as an example, selects the three-day ultraviolet index data from January 1, 2020 to January 3, 2020, and establishes ARIMA (p, d, q) model. The ultraviolet index in the first 9 hours before January 4, 2020 is predicted, as shown in **Figure 1**.

From the **Figure 1**, we can find that although there are some outliers in the sudden decrease of the value, the time series of the overall ultraviolet index is basically stable, but we still need to check whether it is really stable.

3.1 Data Preprocessing

The establishment of the ARIMA model requires that the time series be a stationary random process. Therefore, before modeling, the basic statistical features such as independence and stationarity of time series data need to be tested to prepare for modeling.

- Data Preprocessing

The establishment of the ARIMA model requires that the time series be a stationary random process. Therefore, before modeling, the basic statistical features such as independence and stationarity of time series data need to be tested to prepare for modeling.

- Independence Test

The overall test method proposed by Box and Pierce is used to test whether the sequence is a white noise sequence. After testing, $p < 0.05$, the data meets the independence requirements, so it can be processed with time series.

- Stationarity Test

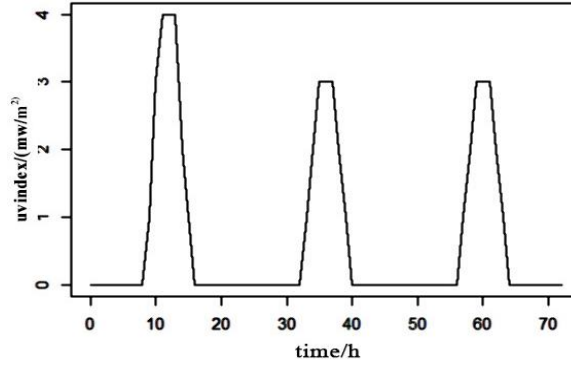


Fig. 1. The actual ultraviolet index in Shanghai from January 11th to 31st

The overall test method proposed by Box and Pierce is used to test whether the sequence is a white noise sequence. After testing, $p < 0.05$, the data meets the independence requirements, so it can be processed with time series.

The nonparametric Daniel test is used to test whether the data series have potential trends. The test method is to calculate the Spearman rank correlation coefficient σ of the time series, the formula is as follows:

$$\sigma = 1 - 6 \sum_{i=1}^n \frac{dt^2}{n(n^2-1)} \quad (2)$$

The rank of $dt = t - x_i$ is $t - R(x_i)$ for large samples, statistics $Z = (\rho - \mu_\rho) / \sigma_\rho$. In the formula $\mu_\rho = 0$; $\sigma_\rho = 1/\sqrt{n-1}$.

Then determine the significance level α , and calculate the significance level p according to the statistics. If $p < \alpha$, it means that the time series has an upward or downward trend; $p > \alpha$, means that there is no trend in the series, here $\alpha = 0.05$. According to Daniel test $p = 0.09825 > 0.05$, the time series is not stable. In order to achieve the requirements of zero equalization and the best fitting effect, the ultraviolet index series needs to be differentiated.

3.2 Model Ordering and Model Determination

First, make a first-order difference to the ultraviolet index data. The difference results are shown in **Figure 2**.

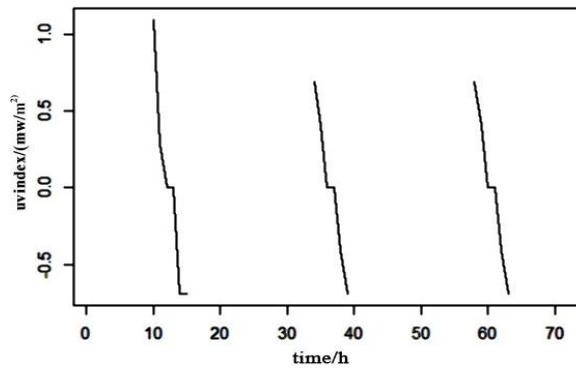


Fig. 2. Ultraviolet index data after first-order difference

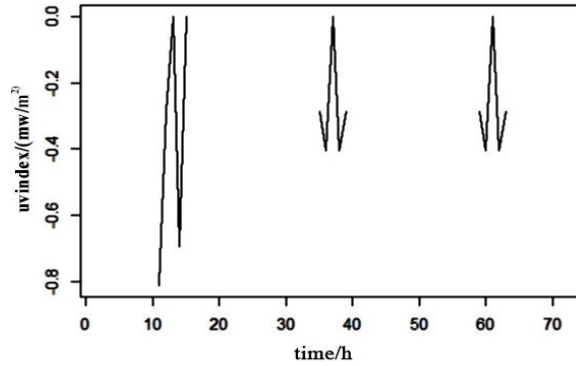


Fig. 3. Ultraviolet index data after second-order difference

Figure 2 looks much more stable than the **Figure 1**. The stationarity test of the ultraviolet index data after the first-order difference is $p=0.3628 > 0.05$, the test result does not meet the standard, the data shall be second-order differenced and the stationarity test shall be performed on the differenced result.

The second-order difference is shown in the **Figure 3**.

The stationarity test of the ultraviolet index data after the second-order difference is $p=0.6504 > 0.05$, it still does not meet the standard, so we need to carry out the third-order difference.

The third-order difference is shown in the **Figure 4**.

The stationarity test of the ultraviolet index data after the third-order difference is $p=0.01 < 0.05$, which meets the standard, the sequence is stable, and autocorrelation analysis can be performed.

After the time series data is stabilized, first calculate the original series autocorrelation function (ACF), partial autocorrelation function (PACF), and draw the autocorrelation and partial correlation function graphs. According to the autocorrelation analysis method, make a preliminary series model type and model order preliminary identify, refer to Table 1, and then judge by the AIC criterion, find the p, q corresponding to the minimum value of the model's one-step error, and select the optimal model.

For stationary time series y_t , there is

Self-covariance:

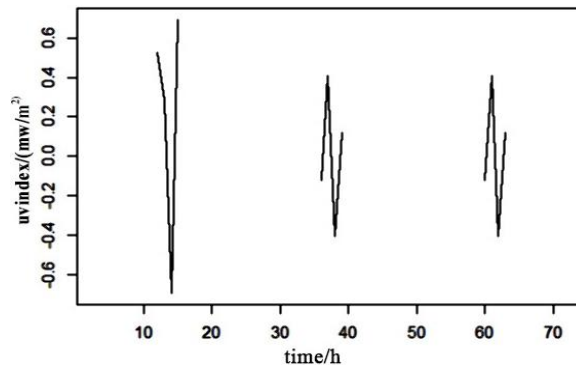


Fig. 4. Ultraviolet index data after third-order difference

Table 1. Stationary time series model structure identification rule

Category	Model recognition		
	AR(p)	MA(q)	ARMA(p,q)
Autocorrelation function(ACF)	Tailing	Censored	Tailing
Partial autocorrelation function Model recognition(PACF)	Censored	Tailing	Tailing

$$\hat{\gamma}_k = \frac{1}{N} \sum_{j=1}^{N-k} y_j y_{j+k} . \quad (3)$$

Autocorrelation function:

$$\hat{\rho} = \frac{\gamma_k}{\gamma_0} . \quad (4)$$

The partial correlation function :

$$\begin{cases} \hat{\alpha}_{11} = \hat{\rho}_1 \\ \hat{\alpha}_{k+1,k+1} = \left(\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\rho}_{k+1-j} \hat{\alpha}_{kj} \right) \times \left(1 - \sum_{j=1}^k \hat{\rho}_j \hat{\alpha}_{kj} \right) \\ \hat{\alpha}_{k+1,j} = \hat{\alpha}_{kj} - \hat{\alpha}_{k+1,k+1} \times \hat{\alpha}_{k,k-j+1} \end{cases} . \quad (5)$$

The order of the model can be preliminarily determined by judging the truncation of $\hat{\rho}_k$, $\hat{\alpha}_k$.

Make autocorrelation and partial correlation function graphs based on ultraviolet index data. The autocorrelation function graph of the ultraviolet index sequence is shown in **Figure 5**.

According to the **Figure 5**, it can be seen that the autocorrelation coefficients after the second order fluctuate around 0, that is, they are all within the confidence boundary.

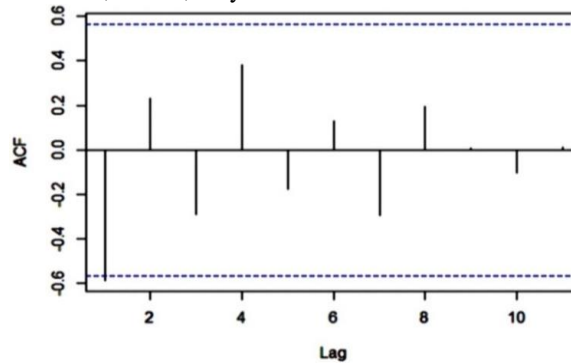


Fig. 5. The autocorrelation function graph of the ultraviolet index sequence

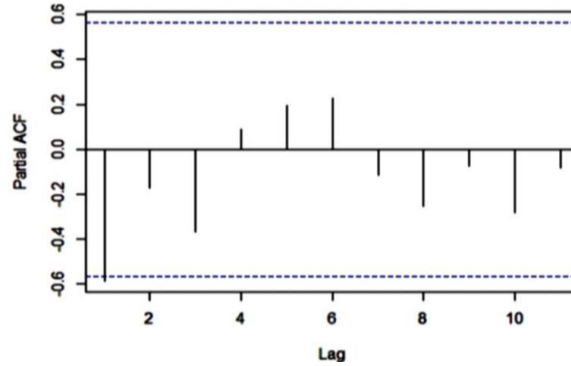


Fig. 6. The partial autocorrelation function of ultraviolet index sequence

Carry out root test ($p=0.01 < 0.05$) and white noise test ($p=0.02168 < 0.05$) on the data after autocorrelation function calculation. The test results meet the requirements for partial autocorrelation function calculation.

Make a graph of partial autocorrelation function of ultraviolet index sequence, the result is shown in **Figure 6**.

According to the partial autocorrelation graph, it can be seen that the partial autocorrelation coefficients are all within the confidence boundary after order 2.

According to the autocorrelation function graph and part of the autocorrelation function graph, the following models can be known:

- ARMA(0,2) model: that is, the autocorrelation graph is reduced to 0 after lagging by 2 order, and the partial autocorrelation is reduced to 0, it is a moving average model (MA model) with order $q=2$;
- ARMA(2,0) model: that is, the partial autocorrelation graph is reduced to 0 after the lag 2 order, and the autocorrelation is reduced to 0, it is an autoregressive model (AR model) with level $p=2$;
- ARMA(2,2) model: Even the autocorrelation and partial autocorrelation are reduced to zero. It is a mixed model (ARMA model).

It is highly subjective to directly determine the order of the model through tailing and censoring. For the estimation of model parameters, we need to perform a weighted evaluation on the loss and regularization terms. When selecting parameters, we need to balance the prediction error and model complexity. So we need to determine the order of the model according to the information criterion function method (AIC).

Finally, according to the AIC criterion and the AIC results of each model in Table 2, the optimal model is determined to be ARIMA (2,3,2), AIC=112.69.

3.3 Model Verification and Model Application

Table 2. The AIC results of each model

Model	AIC
ARIMA(0,3,2)	113.76
ARIMA(2,3,0)	130.30
ARIMA(2,3,2)	112.69

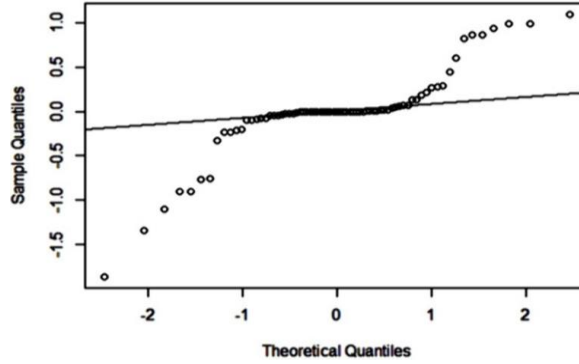


Fig. 7. Residual plot of ARIMA model

See if the model's residuals satisfy an independent normal distribution with a mean of zero.

From the **Figure 7**, it can be seen that the scatter is basically at the two ends of the straight line, so the residuals meet the normal distribution.

Significance test on the model:

$p=0.6404 > 0.05$, reflecting that the residual is a white noise sequence, so the residual white noise test passes.

Significance test on the parameters:

$$P_1 = 0.06897761 > 0.05,$$

$$P_2 = 0.8891828 > 0.05,$$

$$P_3 = 0.09903509 > 0.05,$$

$$P_4 = 0.06141894 > 0.05, \text{ parameter significance test also passed.}$$

Calculate model accuracy:

Bring in the parameters and after fitting, calculate the model accuracy. Root mean square error (RMSE) and mean absolute error (MAE) are used to evaluate the prediction accuracy of the method [17].

MAE is a common measure of forecast error in time series analysis. It measures the average magnitude of a set of forecast errors:

$$MAE = \frac{1}{N} \sum_{t=1}^n |y_t - f_t|. \quad (6)$$

where n is the number of observations during the total evaluation period, y_t is the observed value at time t , and f_t is the predicted value.

Equation (6) shows that MAE is the average value of the absolute deviation between the predicted value and the corresponding observation value. RMSE is a secondary scoring rule used to measure the average magnitude of error.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - f_t)^2}. \quad (7)$$

The difference between the predicted value and the corresponding observed value is squared, summed, and the average number of samples is averaged. Finally, take the square root of the average. Since the error is squared before averaging, RMSE gives higher weight to larger errors. This means that RMSE is most useful when big mistakes are not needed.

Table 3. Forecast results

time	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
0:00	0	-0.642477	0.6577462	-0.9866253	1.001895
1:00	0	-1.457045	1.4794547	-2.2342892	2.256699
2:00	0	-2.557124	2.5944630	-3.9206667	3.958006
3:00	0	-3.825297	3.8719525	-5.8626362	5.909291
4:00	0	-5.290319	5.3512132	-8.1069630	8.167857
5:00	0	-6.901521	6.9731178	-10.5739160	10.645513
6:00	0	-8.673376	8.7588861	-13.2874170	13.372927
7:00	0	-10.579488	10.6767307	-16.2056696	16.302912
8:00	5.048e-13	-12.626980	12.7380799	-19.3407059	19.451806

In this model, $MAE=0.2604444$, $RMSE=0.4790116$, indicating that the sequence is successfully modeled.

Use the model to predict the value of the next 9 hour:

As the value of ultraviolet index can not be negative, the final accurate result is obtained by processing the data of prediction result as shown in Table 3.

Draw a prediction map, the result is shown in **Figure 8**.

The thin dashed line represents the original data, and the solid line represents the fitted value. It can be seen from the figure that the fitting effect is very good. The prediction result is displayed as a thick dotted line, dark gray indicates a confidence interval with 80% confidence, and light gray indicates a confidence interval with 95% confidence.

3.4 Conclusion

It is feasible to predict the ultraviolet index by applying ARIMA(2,3,2) model. After verification, the model can better fit the trend of the ultraviolet index in Shanghai, China. The research results have very high value for the control of solar energy in the smart micro grid.

The implemented technique is effective and precise for ultraviolet index prediction but the ultraviolet index is affected by altitude, weather conditions and Ozone layer thickness, etc. The impact and the constraints of the model itself can only be short-term forecasts. If long-term prediction is required, the model can be optimized or combined with other models to establish a hybrid model. In the future, we plan to develop our own fuzzy logic techniques for ultraviolet index forecasting. Finally, we need to compare those techniques along with ARIMA model to find the best one.

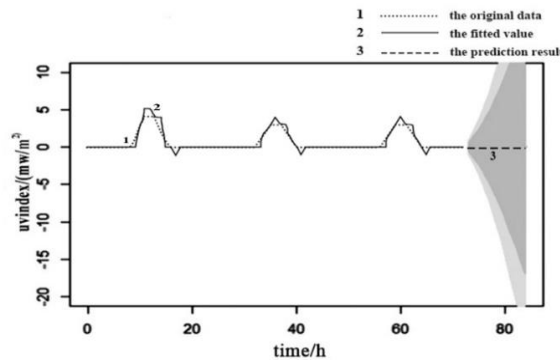


Fig. 8. Predicted value through ARIMA model

Acknowledgment

The authors acknowledge financial support from the Natural Science Foundation of Educational Department of Liaoning Province (Grant: J2020053), and Technology Innovation Fund (Grant: 2020JJ26GX029) and would like to express many thanks to the support of Dalian Key Laboratory of Smart Micro-grid and Green Recycling Industry.

References

- [1] Rana, M., Li, L.: An Overview of Distributed Microgrid State Estimation and Control for Smart Grids. *Sensors*. pp. 4302-4325 (2015)
- [2] Pavan Kumar, Y. V., Bhimasingu, R.: Key Aspects of Smart Grid Design for Distribution System Automation: Architecture and Responsibilities. *Procedia Technology*. pp. 352-359 (2015)
- [3] Venkataramanan, G., Mamay, C.: A larger role for microgrids. *IEEE Power & Energy Magazine*. pp. 78-82 (2008)
- [4] Vaikund, H., Srivani, S, G.: Trends in Energy Management System for Smart Microgrid-An Overview. *Advances in Signal and Data Processing*. pp. 15-28 (2021)
- [5] Fan, R., Xu, N., Wang, W., Li, Q., Meng, T.: Research on Photovoltaic Power Generation Efficiency Detection & Case Analysis. Vol. 6, pp. 244-249. *Proceedings of the 2017 2nd International Conference on Machinery, Electronics and Control Simulation (MECS 2017)*, China (2016)
- [6] Qt, A., Peng, W, A., Wei, T, A., et al.: Benefit allocation model of distributed photovoltaic power generation vehicle shed and energy storage charging pile based on integrated weighting-Shapley method. *Global Energy Interconnection*. pp. 375-384 (2020)
- [7] Zhao, Y., She, S., Chen, Y., Lin, P., Zhao, J., Zhang, M., Wang, L., Zhang, W., Liu, Y.: Research on Photovoltaic Power Generation Output Forecasting Model Based on Wavelet SVM and Micrometeorological Influencing Factor Analysis. *Proceedings of the 2020 the 4th International Conference on Innovation in Artificial Intelligence*. pp. 250-256 (2020)
- [8] Nikkhajoei, H., Lasseter, R, H.: Distributed Generation Interface to the CERTS Microgrid. *IEEE transactions on power delivery* pwr. pp. 1598-1608 (2009)
- [9] Yona, A., Senjyu, T., Funabashi, T.: Application of recurrent neural network to short-term-ahead generating power forecasting for photovoltaic system. *IEEE Power Engineering Society General Meeting*. pp. 24-28 (2007)
- [10] Vapnik, V.: An overview of statistical learning theory. *IEEE Transactions on Neural Networks*. pp. 988-999 (1999)
- [11] Farhangi, H.: Intelligent Micro Grid Research at BCIT. *IEEE Electrical Power and Energy Conference*. pp. 18-28 (2008)
- [12] Platt, J.: Fast training of support vector machines using sequential minimum optimization. *Advance in Kernel Methods-support Vector Learning*. pp. 185-208 (1999)
- [13] Xu, J.: Application of Grey Theory in the Load Prediction of Power System. *JiangXi Hydraulic Science and Technology*. pp. 216-219 (2005)
- [14] Wang, H., Huang, J., Zhou, H., Zhao, L., Yuan, Y.: An Integrated Variational Mode Decomposition and ARIMA Model to Forecast Air Temperature. *Sustainability*. pp. 1-11 (2019)
- [15] Zheng, P.: Telecommunication Data Forecasting based on ARIMA Model. *International Journal of Electrical and Electronics Engineers*. pp. 451-458 (2016)
- [16] Han, P., Wang, P, X., Shu, Y, Z., et al.: Drought forecasting based on the remote sensing data using ARIMA models. *Mathematical and Computer Modelling*. pp. 1398-1403 (2010)

[17] Wang, W., Lu, Y.: Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. IOP Conference Series: Materials Science and Engineering. pp. 12-49 (2018)