

Research Article

SARIMA-Orthogonal Polynomial Curve Fitting Model for Medium-Term Load Forecasting

Herui Cui, Pengbang Wei, Yupei Mu, and Xu Peng

School of Economics and Management, North China Electric Power University, Baoding 071003, China

Correspondence should be addressed to Herui Cui; cuiherui1967@126.com

Received 21 June 2016; Accepted 4 October 2016

Academic Editor: Paolo Renna

Copyright © 2016 Herui Cui et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Seasonal component has been a key factor in time series modeling for medium-term electric load forecasting. In this paper, a seasonal-ARIMA model is developed, but the parameters of the SAR and the SMA turn out to be quite nonsignificant in most cases during the model order selection. To address this issue, the hybrid time series model based on the HP filter is utilized to extract the spectrum sequences with different frequencies and analyze interactions among various factors. Finally, an integrative forecast is made for the electricity consumption from January to November in 2014. The empirical results demonstrate that the method with HP filter could reduce the relative error caused by the interaction between the trend component and the seasonal component.

1. Introduction

To a certain extent, the medium-term power consumption is affected by the seasonal factors, the historical consumption, and consumption peaks caused by unexpected events. According to the current prediction techniques, these factors are temporarily categorized into the long-term trend (L), the seasonal fluctuation (S), the cycle volatility factors (C), and the irregular volatility factors (I). The influences of these factors superimpose on each other and thus that became a difficult problem in model construction.

In recent years, many researches have been conducted in the field of the above four fluctuation factors in power load forecasting. The neural network method is often used to make predictions on electric load [1, 2], whereas, considering the data volume, the time sequence model is more in line with the characteristics of the sequence than the neural network model [3]. Azadeh et al. [4] combined the seasonal fluctuation and the nonlinearity of forecasting with the fuzzy system and data mining techniques to analyze the monthly electricity demand in Iran. The SVM model can also be exploited to analyze the effect of the seasonal fluctuation and the long-term trend [5]. The significant trend sequence can be analyzed through the GM(1, 1) model combined with neural network method [6–8]. It follows that the characteristics such as trend and seasonal ones are the key factors which affect

the accuracy of the load forecasting in the medium-term load forecasting. The SARIMA model which could eliminate the effects of seasonal factor and irregular change factors is more suitable for the monthly electricity consumption forecasting [9, 10]. The decomposition method of sequence is often used to analyze the superimposed effect produced by seasonal change tendency and the long-term growth and decline trend [11]. Among them, the application of Hodrick-Prescott (HP) filtering method has the certain superiority in the series decomposition [12, 13].

The seasonal-ARIMA model is able to take all seasonal fluctuation of sequence into full account. However, due to the interaction of the four fluctuations, “ L ”, “ S ”, “ C ,” and “ I ”, and the interaction between the seasonal factors and the nonseasonal factors, the seasonal parameters are nonsignificant in practical applications in most cases. HP filter is based on the spectral analysis to separate the data sequences and relieve the superimposed impact of the fluctuations. In this paper, by using the HP filter we get the sequence $\{G\}$ with significant trend and the sequence $\{C\}$ with significant periodicity. With separated modeling and integrative analysis, the model can successfully relieve the mutual influence of the changing trend and improve the precision of prediction. Besides, to cope with the model order problems, the paper has conducted a long-memory test on the original data sequence. The result shows that the sequence does not meet the standard random

walk process, which puts forward new ideas with the power load forecasting.

The purpose of this paper is to design an accurate prediction model. And the remaining parts of the paper are organized as follows. Section 2 introduces the principle of the method. Section 3 describes the process of the power load forecasting by using the traditional method and the improved method and discusses the results. The last section makes some conclusions of this paper.

2. Forecasting Models

2.1. SARIMA Model. SARIMA (Seasonal Autoregressive Integrated Moving Average), which is denoted as SARIMA $(p, d, q)(P, D, Q)^S$, is based on the traditional ARIMA (p, d, q) model, and it can eliminate the periodicity influence in a prediction process and thus is a widely applied model for forecasting seasonal time series [14, 15]. The formula can be described as follows:

$$\begin{aligned} \phi_p(L) \Phi_P(L^S) (1-L)^d (1-L^S)^D y_t \\ = \theta_q(L) \Theta_Q(L^S) \varepsilon_t, \end{aligned} \quad (1)$$

where L is the backward shift operator. The integers p, q, P , and Q are the order of $\phi_p(L)$, $\theta_q(L)$, $\Phi_P(L^S)$, and $\Theta_Q(L^S)$, respectively. The integers d and D are the number of regular differences and seasonal differences, respectively, and, for a nonstationary time series y_t , $(1-L)^d y_t$ could come to a stationary series by using the difference operator $1-L$. L satisfies the formula $L^k y_t = y_{t-k}$. The formulas

$$\begin{aligned} \phi_p(L) &= 1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \dots - \phi_p L^p, \\ \theta_q(L) &= 1 - \theta_1 L - \theta_2 L^2 - \theta_3 L^3 - \dots - \theta_q L^q \end{aligned} \quad (2)$$

are polynomials in L of degrees p and q . And the formulas

$$\begin{aligned} \Phi_P(L^S) &= 1 - \Phi_1 L^S - \Phi_2 L^{2S} - \Phi_3 L^{3S} - \dots - \Phi_P L^{PS}, \\ \Theta_Q(L^S) &= 1 - \Theta_1 L^S - \Theta_2 L^{2S} - \Theta_3 L^{3S} - \dots - \Theta_Q L^{QS} \end{aligned} \quad (3)$$

are polynomials in L of degrees P and Q . And ε_t which is a current interference with variance σ^2 and mean $= 0$ is considered as the estimated residual at time t . At the same time, ε_t is an independent and identically distributed normal random variable.

In the process of the seasonal time series analysis, there are three questions that need to be analyzed.

(1) Stationary Test. The stationarity of the time series $\{y_t\}$ is the premise for building the S-ARIMA model. When it meets the condition that μ , σ^2 , and γ_l are constants in formula (4), we can define $\{y_t\}$ as weakly stationary or covariance stationary:

$$\begin{aligned} E(y_t) &= \mu, \\ \text{var}(y_t) &= \sigma^2, \\ \text{cov}(y_t, y_{t-1}) &= \gamma_l. \end{aligned} \quad (4)$$

The ADF unit root test can be used to test whether the sequence $\{y_t\}$ is stationary or not. If the sequence is nonstationary, the difference transformation would be used until the difference sequence is stationary. The stationary sequence with differential transformation is defined as $w_t = \Delta^d y_t$.

(2) Seasonal Analysis. Before we make the seasonal analysis, the autocorrelation function should be defined. It can be expressed as follows:

$$\rho_k = \frac{\sum_{t=k+1}^n (w_t - \bar{w})(w_{t-k} - \bar{w})}{\sum_{t=1}^n (w_t - \bar{w})^2}, \quad (5)$$

where w_t is a stationary sequence and \bar{w} is the average of the sequence $\{w_t\}$. By judging the autocorrelation function and the confidence interval, the periodicity and the cycle S of $\{w_t\}$ could be obtained. According to the additive model which is defined as (6), the sequence $\{w_t\}$ can be seasonally adjusted:

$$w_t = TC_t + S_t + I_t, \quad (6)$$

where TC_t means long-term trend and cycle volatility, S_t means seasonal fluctuation, and I_t means irregular volatility.

(3) Model Order Selection and Model Prediction. Firstly, the seasonally adjusted sequence is defined as $\{w_{SA}\}$. The lag intervals for endogenous function and the confidence interval of the autocorrelation function and the partial autocorrelation function should be analyzed in order to determine the order of AR(p), MA(q), SAR(P), and SMA(Q) and build the ARIMA(p, d, q)(P, D, Q)^S model. Then, according to the principle of minimum mean square error, the prediction is the conditional expectation of y_{T+1} , and it can be expressed as follows:

$$\hat{y}_T(l) = E(y_{T+l} | y_T, y_{T-1}, \dots, y_1); \quad (7)$$

when the higher-order problem exists in the ARIMA(p, d, q)(P, D, Q)^S model, we can make the long-memory test for the stationary sequence $\{w_t\}$.

2.2. R/S Method of ARFIMA Model. The long-memory analysis, which is specific to the random walk process, is put forward by H. E. Hurst in the research of the relationship between the reservoir of water flow and the storage capacity in 1951. And he puts forward the rescaled range analysis (R/S) for the long-memory analysis. Then the researchers often use this method for financial sequence analysis and build the Autoregressive Fractionally Integrated Moving Average (ARFIMA) model [16, 17]. The analysis procedure of R/S method is shown in the following paragraph.

Firstly, the sequence $\{w_t\}$ is divided into the infinite number of intervals, and the length of each interval is n . Every interval is defined as follows:

$$w_{t,n} = \sum_{u=1}^t (w_u - M_n), \quad (8)$$

where M_n is the average of the interval w_u and $w_{t,n}$ is the cumulative deviation of the interval w_u . Then the letter R can be used in denoting the difference between the maximum $w_{t,n}$ and the minimum $w_{t,n}$, and the letter S can be used in denoting the standard deviation of the sequence $\{w_u\}$, so the formula of R/S analysis can be expressed as follows:

$$K(n)^H = \frac{R}{S} = \frac{\max(w_{t,n}) - \min(w_{t,n})}{\sqrt{(1/n) \sum_{t=1}^n (w_t - \bar{w})^2}}, \quad (9)$$

where H , which is the Hurst Index, is defined as the index of n and K is a constant. The logarithm should be taken on both sides of the equation, and adjust the equation as follows:

$$H = \frac{(\log(R/S) - \log(K))}{\log(n)}. \quad (10)$$

Then the Hurst Index can be worked out by the OLS method. Finally, the long memory can be judged by the standard as follows [18]:

$$0 \leq H < 0.5 \quad \{w_t\} : \text{mean reversion process}$$

$$H = 0.5 \quad \{w_t\} : \text{standard random walk process} \quad (11)$$

$$0.5 < H < 1 \quad \{w_t\} : \text{long memory process};$$

when $0.5 < H < 1$, the original sequence is likely to have a long memory [19]; however, whether there is an ARFIMA model which is suitable for most of the medium-term load forecasting cannot be guaranteed [20].

2.3. Orthogonal Polynomial Curve Fitting. Orthogonal polynomial curve fitting is the improvement of the Ordinary Least Square (OLS). There is a premise that the independent variables must be accurate values before using the OLS method, but it is not reasonable in most cases. When the error of the independent variables reaches a certain extent, the prediction model with OLS method would produce a certain error. In view of this situation, the orthogonal polynomial curve fitting is proposed. And its basic principle is that the square sum of the orthogonal distance from all points to the fitting curve is minimum. In the OLS method the fitting polynomial can be expressed as follows:

$$\hat{y}_t = f(a, \hat{x}_t), \quad t = 1, 2, \dots, \quad (12)$$

which is fitted by the least square criterion: the distance square sum between the predicted value and actual value is minimum, and it can be expressed as follows:

$$\begin{aligned} \min \sum e_t^2 &= \min \sum (y_t - \hat{y}_t)^2 \\ &= \min \sum (y_t - \hat{a}_0 - \hat{a}_1 x_t - \dots - \hat{a}_n x_t^n)^2; \end{aligned} \quad (13)$$

then the undetermined coefficients can be got by the mean value theorem. This orthogonal polynomial curve fitting method is improved on the basis of OLS method, and the errors of the dependent variable and the independent variable

are considered to build forecasting model. And the fitting polynomial can be expressed as follows:

$$\hat{y}_t = f(a, \hat{x}_t), \quad t = 1, 2, \dots, \quad (14)$$

where \hat{x}_t is the predicted value of the independent variable x_t . The orthogonal distance error can be expressed as follows:

$$\varepsilon_t = \sqrt{\delta_t^2 + \sigma_t^2}, \quad (15)$$

where δ_t and σ_t are the random error of x_t and y_t , respectively. Then the criterion of the orthogonal polynomial curve fitting can be expressed as follows:

$$\begin{aligned} \min \sum \varepsilon_t^2 \\ = \min \sum [\delta_t^2 + (y_t - \hat{a}_0 - \hat{a}_1 x_t - \dots - \hat{a}_n x_t^n)^2]. \end{aligned} \quad (16)$$

Combining the orthogonal polynomial with the OLS method, the multinomial model can rise to the imitative effect.

The objective function can be expressed as follows:

$$D = \sum_{i=1}^n d(P_i, L)^2, \quad (17)$$

where P_i represents the real point, L represents the fitted curve, and $d(P_i, L)$ represents the orthogonal distance from the real points to the fitted curve.

The parameter equation of fitted curve L can be defined as follows:

$$\begin{aligned} x(\gamma) &= \alpha + \gamma \cos \theta, \\ y(\gamma) &= \beta + \gamma \sin \theta, \end{aligned} \quad (18)$$

where $\{\alpha, \beta\}$ is a point of fitted curve L and θ is the included angle of the tangent to the abscissa axis, so the objective function can be expressed as follows:

$$D = \sum_{i=1}^n ((x_i - \alpha) \times \sin \theta - (y_i - \beta) \cos \theta)^2. \quad (19)$$

Then we should take its partial derivative with respect to α , β , and θ in order to calculate the minimum error and the fitted curve. The equation set can be expressed as follows:

$$\begin{aligned} \frac{\partial D}{\partial \alpha} &= -2n(\bar{x} - \alpha) \sin^2 \theta + n(\bar{y} - \beta) \sin 2\theta = 0 \\ \frac{\partial D}{\partial \beta} &= n(\bar{x} - \alpha) \sin 2\theta - 2n(\bar{y} - \beta) \cos^2 \theta = 0 \\ \frac{\partial D}{\partial \theta} &= \left(\sum_{i=1}^n (x_i - \alpha)^2 - \sum_{i=1}^n (y_i - \beta)^2 \sin 2\theta \right) \\ &\quad - 2 \left(\sum_{i=1}^n (x_i - \alpha)(y_i - \beta) \right) \cos 2\theta = 0, \end{aligned} \quad (20)$$

where \bar{x} and \bar{y} are the mean values of the sequences $\{x\}$ and $\{y\}$.

2.4. Hodrick-Prescott Filter. Hodrick and Prescott first put forward Hodrick-Prescott filter (HP filter) method in the paper analyzing the economic cycle about postwar America. The method regarded the time series as the spectrum for analyzing [14, 15]. It divided the sequence into two groups, and their relationship with the original sequence is counted as

$$y_t = g_t + c_t, \quad (21)$$

where the sequence $\{G\}$ with long-term trend is denoted as $G = \{g_1, g_2, \dots, g_n\}$ and the sequence $\{C\}$ with short-term volatility is denoted as $C = \{c_1, c_2, \dots, c_n\}$. The separation process must satisfy the minimum loss function principle:

$$\min \left\{ \sum_{t=1}^n (y_t - g_t)^2 + \lambda \sum_{t=1}^n [(g_{t+1} - g_t) - (g_t - g_{t-1})]^2 \right\}, \quad (22)$$

where $\lambda = \sigma_1^2 / \sigma_2^2$, where λ is the smoothing parameter and σ_1^2 and σ_2^2 represent the standard deviation of the sequence $\{G\}$ and the sequence $\{C\}$, respectively. When λ increases, estimated total trend changes in relation to the change in the sequence which is reduced. It means that λ takes the high number, the estimated trend is more smooth, and when λ trends to infinity, estimated trend will be close to the linear function. As a general rule of thumb, when we analyze the monthly data, λ can be defined as $\lambda = 14400$.

In this paper, the HP filter is applied to the nonseasonally adjusted series, and the original sequence is divided into two sequences with the significant spectral frequency and building the model more accurately by weakening the mutual effect between the two sequences.

2.5. Error Estimation Methods. There are five basic error estimation methods; simultaneously, the model can be evaluated by relative error (RE), mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute error (MAE), which can be expressed as follows:

$$\begin{aligned} \text{RE}_i &= \frac{\hat{y}_i - y_i}{y_i} \times 100\%, \\ \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \times 100\% \right|, \\ \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \\ \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \end{aligned} \quad (23)$$

2.6. Sequence Analysis and Combination Model Building. The improved model is based on separating the original sequence by filtering analysis. Then according to the characteristics of

each sequence, the models can be established for forecasting. The detailed process is as follows.

(1) According to the HP filtering principle, the original sequence $\{y_t\}$, defined as the superposition of the waves with different frequencies, can be divided into the sequence $\{G\}$ and the sequence $\{C\}$.

(2) The sequence $\{G\}$ is defined as a function of time “ T ,” and its scatter-plot can be drawn. The error term from each point to the fitting curve is denoted as $\delta_1, \delta_2, \dots, \delta_n$, $\sigma_1, \sigma_2, \dots, \sigma_n$. Then the orthogonal polynomial with the OLS method is used for making polynomial curve fitting to minimize the sum of squared errors $\min \sum \varepsilon_t^2 = \min \sum [(\sqrt{\delta_t^2 + \sigma_t^2})^2]$.

(3) According to the polynomial fitting in the previous step, the sequences’ predictions can be got and defined as \hat{g}_t .

(4) The stationary property of the sequence $\{C\}$ is tested. If it was stationary, the correlation analysis can be used on the sequence; otherwise, the differential transform is conducted on the sequence $\{C\}$ until it is stationary. The stationary sequence is denoted as $\{w_c\}$.

(5) Through the correlation analysis of the seasonal fluctuation, the autocorrelation and moving average items can be acquired [18, 20]. According to the result, the ARIMA model can be defined as

$$y_c = \phi_p^{-1}(L) \phi_p^{-1}(L) (1-L)^{-d} (1-L^s)^{-D} \theta_q(L) \Theta_Q(L) \cdot \varepsilon_c. \quad (24)$$

(6) The rationality of the ARIMA model is tested by the residual sequence.

(7) According to the ARIMA model, the sequences’ predictions can be got and denoted as \hat{c}_t .

(8) The final prediction result can be obtained based on the principle of HP filter:

$$\hat{y}_t = \hat{g}_t + \hat{c}_t. \quad (25)$$

The improved model will produce twice prediction error in the analysis. In theory, there is the possibility of increasing the errors and reducing the prediction accuracy. But in the actual analysis, the HP filter method weakens the mutual influence of factors (including the long-term trend and the seasonal fluctuation) and utilizes integrative forecasting for models with the different characteristics of the sequence $\{G\}$ and the sequence $\{C\}$. In this way, the trend of the sequence can be effectively fitted and the influence on the seasonal trend can be reduced. Finally higher prediction accuracy can be achieved.

According to the above steps, the specific process of the improved model is shown in Figure 1.

3. Empirical Analysis

The example chooses the research data about electric power consumption from January 2004 to November 2014 in China, the data in January 2004 to December 2013 for model building, and the data in January 2014 to November 2014 for testing the prediction error.

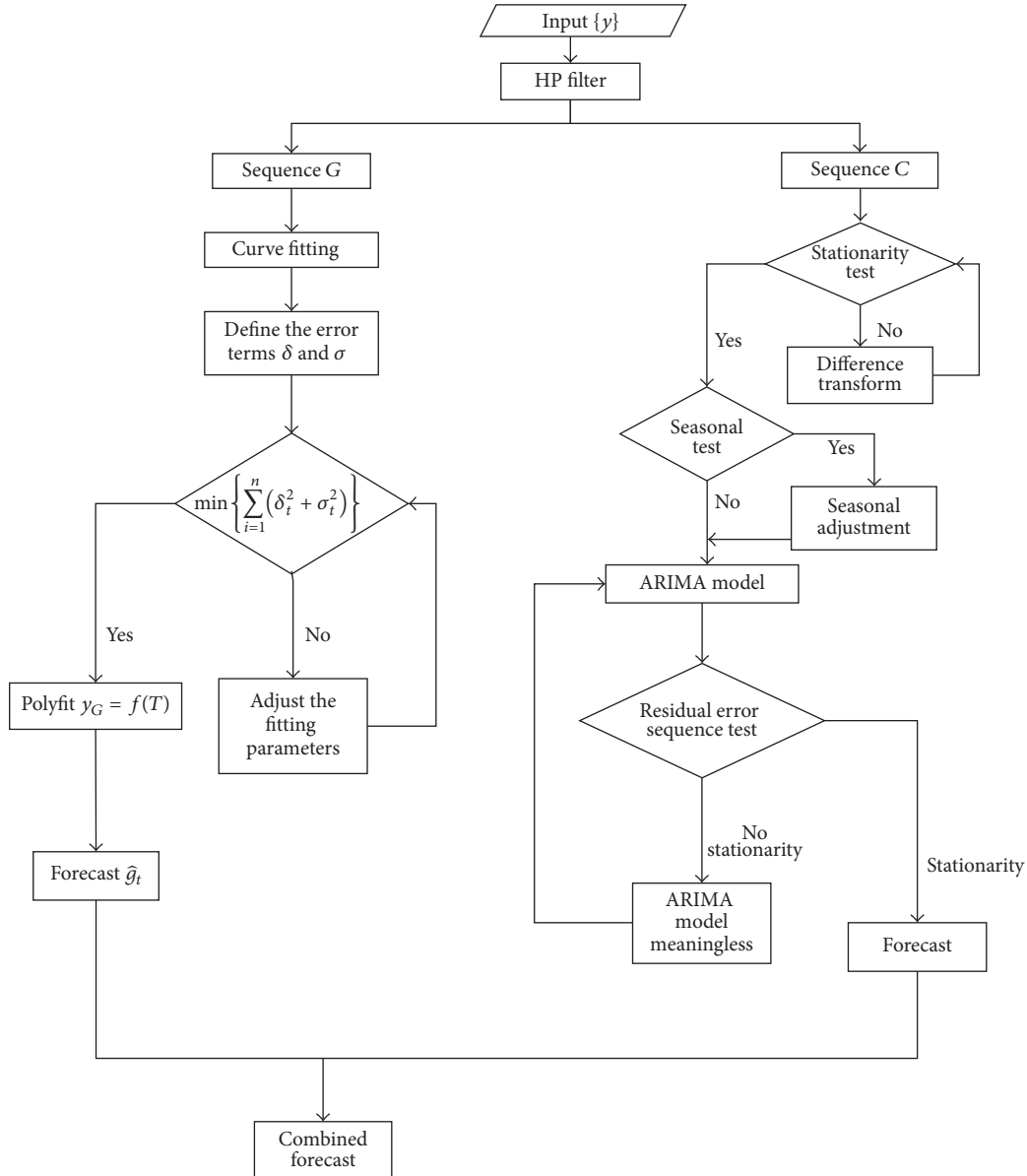


FIGURE 1: Model flow chart based on HP filter.

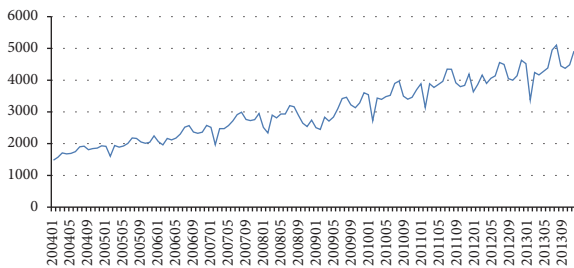


FIGURE 2: Monthly electricity consumption of China from January 2004 to December 2013.

3.1. *The Forecasting of the Seasonal-ARIMA Model.* Figure 2 plots the monthly electricity consumption data from January

2004 to December 2013; as we can see, the intercept and the trend exist in the original sequence, and the sequence is nonstationary. The result of the ADF unit root test with intercept and trend on the sequence is shown in Table 1.

The ADF unit root test demonstrates that the original sequence is nonstationary and the first-order difference of the original sequence is stationary under the 5% significant level. Through observation of the autocorrelation function of the first-order difference sequence, we discover that the sequence's seasonal cycle is 12. Therefore, the additive model is used to adjust the seasonal trend of the sequence. The analysis of the partial autocorrelation and the autocorrelation is shown in Figure 3.

On the basis of the 95% confidence level, the confidence interval of the correlation coefficient is

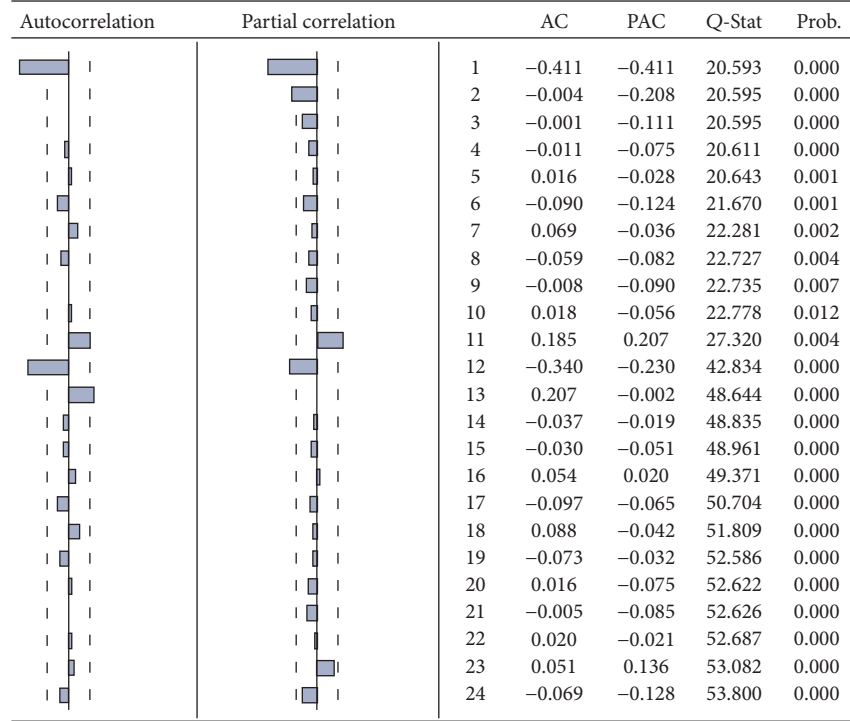


FIGURE 3: Autocorrelation figure and partial autocorrelation figure.

TABLE 1: Augmented Dickey-Fuller unit root test on d(ele).

Augmented Dickey-Fuller test statistic	<i>t</i> -statistic	Prob. *
	-3.726070	0.0248
Test critical values:		
1% level	-4.046925	
5% level	-3.452764	
10% level	-3.151911	

Prob. is the P value; the smaller the P value, the larger the significance. * indicates to reject the original hypothesis under the confidence level of 95%.

$$\left[-\frac{2}{\sqrt{n}}, \frac{2}{\sqrt{n}} \right] = [-0.183, 0.183]. \quad (26)$$

Combined with Figure 3, we can find that partial autocorrelation coefficient in 1, 2, 11 and the autocorrelation coefficient in 1, 11, 13 are not in the confidence interval. Based on the autocorrelation diagram and the partial autocorrelation diagram, OLS method is adopted to establish the seasonal-ARIMA model, and the adjustment of model parameters is based on the significance of correlation coefficient. The estimation of parameter is shown in Table 2.

After determining the model order through parameter significance testing, the time series model we obtain is $ARIMA(2, 1, 11)(1, 1, 0)^{12}$. In this model, the nonseasonal autoregressive items are $AR(1)$ and $AR(2)$, the nonseasonal moving average items are $MA(1)$ and $MA(11)$, and the seasonal autoregressive items are $SAR(1)$. According to this

model, the electricity consumption from January to November in 2014 could be forecasted and the result is shown in Table 6.

3.2. The Forecasting of the Seasonal-ARFIMA Model. In the process of model order selection, $MA(11)$ has significant influence on modeling and predicting. This phenomenon shows that the error caused by the long-term observations still influences the current monthly electricity consumption to some extent. We infer that the monthly electricity consumption may have the long-term memory characteristics, and this conjecture is confirmed by the long-term memory test using R/S method which shows the Hurst exponent:

$$H = \frac{(\log(R/S) - \log(K))}{\log(n)}, \quad (27)$$

$$H = 0.837.$$

According to the criteria, $0.5 < 0.837 < 1$, we know that the monthly electricity consumption has the long-term memory characteristic. Therefore the current forecast is influenced by the distant observations.

3.3. Integrative Model. The HP filter is applied to analyze the original sequence and the smoothing parameter λ is assigned values as 14400. And the decomposition results are shown in Figure 4. The blue curve means the original sequence. The red curve means the long-term trend, and we can find that the growth rate of the power consumption is mainly constant from 2004 to 2013. The green curve means the cyclical and

TABLE 2: The parameters estimate of the ARIMA model.

ARIMA model		AR(1)	AR(2)	AR(11)	SAR(1)
ARIMA(11, 1, 13)(1, 1, 1) ¹²	Coefficient (<i>t</i> -statistic)	-0.742 (-3.265)	-0.225 (-1.699)	0.345 (4.954)	0.543 (4.956)
ARIMA(2, 1, 11)(1, 1, 0) ¹²	Coefficient (<i>t</i> -statistic)	-0.368 (-2.353)	-0.277 (-2.488)	—	-0.660 (-3.991)
		MA(1)	MA(11)	MA(13)	SMA(1)
ARIMA(11, 1, 13)(1, 1, 1) ¹²	Coefficient (<i>t</i> -statistic)	-1.148 (-17.508)	-0.421 (-6.468)	0.536 (7.110)	0.401 (1.638)
ARIMA(2, 1, 11)(1, 1, 0) ¹²	Coefficient (<i>t</i> -statistic)	0.614 (5.127)	0.341 (3.251)	—	—
	<i>R</i> ²	AIC	SC		
ARIMA(11, 1, 13)(1, 1, 1) ¹²	0.677	12.991	13.206		
ARIMA(2, 1, 11)(1, 1, 0) ¹²	0.397	13.460	13.588		

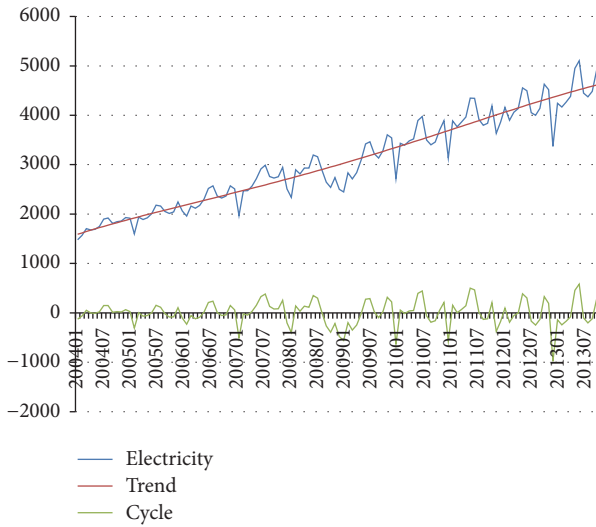


FIGURE 4: HP filter diagram of the power consumption.

irregular change. And as time goes on, the fluctuation range is more significant.

According to HP filter, the original sequence can be separated into the sequence {G} with the long-term trend and the sequence {C} with other fluctuation properties.

The data separation result using HP filter is shown in Figure 5. We can see the sequence {G} approximates a smooth curve, in which the curve of the sequence {C} fluctuates up and down around the zero.

In the perspective of statistics, the sequence {G} can be transformed into a sequence relating to T . Let the independent variable T ($T = 1, 2, \dots, 120$) represent time and let the sequence {G} represent the dependent variable of the system. A curve is fitted to describe the sequence {G} as a function of T . It turns out that the best fitting is achieved when the order of curve fitting is four and the relative error of curve fitting is shown in Figure 6.

The fitting polynomial is

$$\begin{aligned}
 G = & -0.000015186462 \times t^4 + 0.003628985114 \times t^3 \\
 & - 0.239787528844 \times t^2 + 28.15617888808 \times t \quad (28) \\
 & + 1564.292298557236.
 \end{aligned}$$

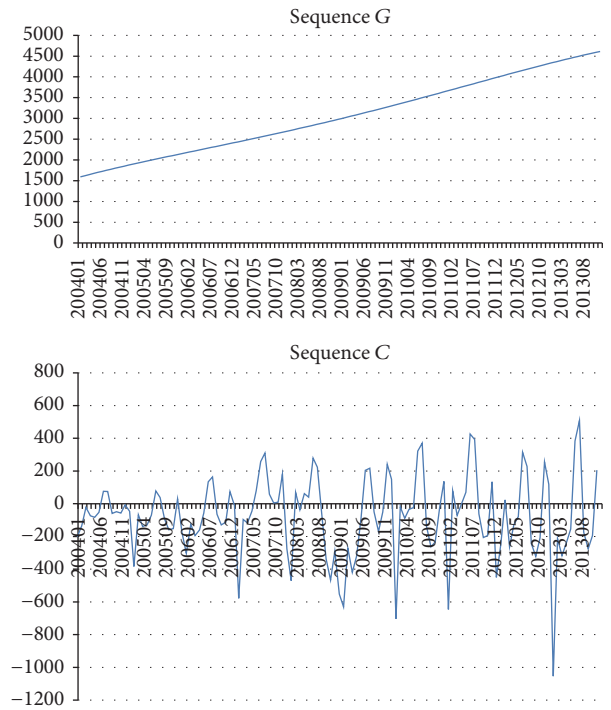


FIGURE 5: Data separation based on HP filter.

According to the fitting polynomial, the sequence from January to November 2014 can be forecasted and the results are shown in Table 3.

The sequence {C} is analyzed based on the time series model. And according to the significance we can adjust the parameters and record in Table 4. Finally, the model is ARIMA(11, 1, 1)(0, 1, 0)¹².

According to ARIMA(11, 1, 1)(0, 1, 0)¹², the sequence {C} is forecasted and the prediction results from January to November 2014 are shown in Table 5.

Finally, according to the HP filter principle, $\hat{y}_t = \hat{g}_t + \hat{c}_t$, we get the prediction of electric consumption from January to November 2014, as shown in Table 6.

3.4. Model Error Analysis. According to Table 6, we make a line chart about the absolute value of the relative error as shown in Figure 7. Except for March, May, and June, the relative errors of the integrative model are smaller than

TABLE 3: Forecasting of $\{G\}$.

Time	Predictions
Jan. 2014	4634.08
Feb. 2014	4655.72
Mar. 2014	4676.84
Apr. 2014	4697.49
May 2014	4717.36
June 2014	4736.73
Jul. 2014	4755.47
Aug. 2014	4773.55
Sep. 2014	4790.96
Oct. 2014	4807.66
Nov. 2014	4823.63

TABLE 4: The parameters estimate of the ARIMA model.

Variable	Coefficient	t -statistic
AR(1)	-0.967	-6.212
AR(2)	-0.389	-3.973
AR(11)	0.330	3.834
MA(1)	0.513	2.939
$R^2 = 0.394$	AIC = 13.527	SC = 13.570

TABLE 5: Forecasting of $\{C\}$.

Time	Predictions
Jan. 2014	-326.44
Feb. 2014	-834.04
Mar. 2014	-394.18
Apr. 2014	-378.59
May 2014	-258.20
June 2014	-208.90
Jul. 2014	418.57
Aug. 2014	402.21
Sep. 2014	-186.52
Oct. 2014	-336.92
Nov. 2014	-263.55

the SARIMA model. From the long-term forecast in July to November, the forecasting results of integrative model are superior to the forecasting results of SARIMA model. In March, May, and June, the optimization models of fitting values are not optimized. On the one hand, this is the objective result of the experimental data; on the other hand, the improved method is not perfect enough. The predicted data is in the large floating in March; this fluctuation is caused by the prediction of sequence $\{C\}$.

According to the prediction result, we make analysis for the prediction error and fill the result in Table 7. Through the comparative analysis, it can be discovered that the prediction accuracy of the improved model is significantly improved.

As shown in Table 7, the MSE of the improved model is the minimum, and it means the predictive ability of the improved model is the most stable. By observing the MAPE, the improved model is 1.817%, which is less than 2.643% of

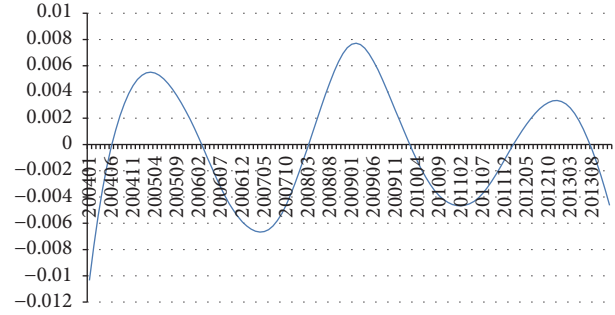
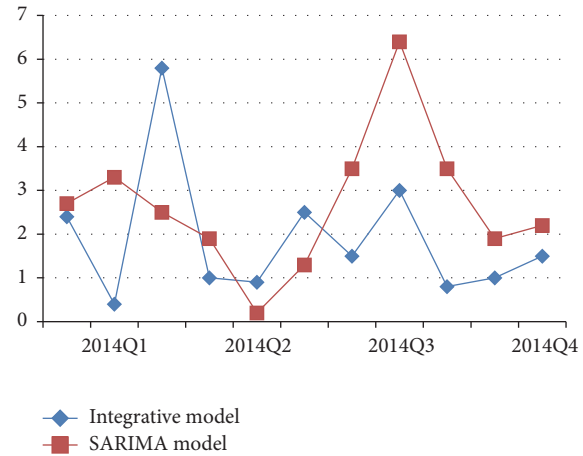
FIGURE 6: Relative error of the curve fitting about sequence $\{G\}$ in January 2004 to December 2013.

FIGURE 7: Prediction and original data of the line graph from January to November in 2014.

the SARIMA model, and it shows that the predictive results of the improved model are close to the real value. The MAE error measurement also shows the same result.

4. Conclusions

In this paper, HP filter is utilized for adjusting the time sequence data. Thus the original sequence is decomposed into the sequences with different trend, and the mutual interference between the different fluctuation items can be relieved. Moreover the relative error of load forecasting is reduced and the multistep prediction is guaranteed.

The testing result of the R/S method shows that the monthly electricity consumption has the property of long-memory process. This conclusion may be attributed to interference of the amount of the data. However, in the actual order analysis, the higher-order AR or MA indeed affects the power load forecasting. Therefore, the long memory of the time series can be considered for building the SARFIMA model in the midterm power load forecasting.

Competing Interests

The authors declared that there is no conflict of interests related to this paper.

TABLE 6: Electricity consumption forecasting results in 2014.

Month	Original data	Model			
		Seasonal-ARIMA		Combined seasonal-ARIMA	
		Prediction	RE (%)	Prediction	RE (%)
Jan. 2014	4407.06	4524.87	2.650	4307.64	-2.256
Feb. 2014	3835.00	3713.78	-3.161	3821.68	-0.347
Mar. 2014	4544.15	4439.32	-2.307	4282.66	-5.754
Apr. 2014	4356.27	4431.63	1.730	4318.90	-0.858
May 2014	4492.04	4503.10	0.246	4459.16	-0.732
June 2014	4638.50	4699.41	1.313	4527.83	-2.386
Jul. 2014	5096.96	5282.25	3.635	5174.04	1.512
Aug. 2014	5025.34	5348.25	6.426	5175.76	2.993
Sep. 2014	4569.73	4730.72	3.523	4604.44	0.760
Oct. 2014	4508.40	4591.07	1.834	4470.74	-0.834
Nov. 2014	4632.24	4736.44	2.249	4560.08	-1.558

TABLE 7: The error of forecast.

	SARIMA model	Improved model
RMSE (%)	3.052	2.340
MAPE (%)	2.643	1.817
MAE	122.386	84.289

Acknowledgments

The authors sincerely acknowledge the financial support from the National Natural Science Foundation of China (no. 71471061).

References

- [1] M. Meng, D. Niu, and W. Sun, "Forecasting monthly electric energy consumption using feature extraction," *Energies*, vol. 4, no. 10, pp. 1495–1507, 2011.
- [2] B. Yang and Y. Sun, "An improved neural network prediction model for load demand in day-ahead electricity market," in *Proceedings of the 7th World Congress on Intelligent Control and Automation (WCICA '08)*, pp. 4419–4424, Chongqing, China, June 2008.
- [3] S. S. Pappas, L. Ekonomou, D. C. Karamousantas, G. E. Chatzarakis, S. K. Katsikas, and P. Liatsis, "Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models," *Energy*, vol. 33, no. 9, pp. 1353–1360, 2008.
- [4] A. Azadeh, M. Saberi, S. F. Ghaderi, A. Gitiforouz, and V. Ebrahimipour, "Improved estimation of electricity demand function by integration of fuzzy system and data mining approach," *Energy Conversion and Management*, vol. 49, no. 8, pp. 2165–2177, 2008.
- [5] M. De Felice, A. Alessandri, and F. Catalano, "Seasonal climate forecasts for medium-term electricity demand forecasting," *Applied Energy*, vol. 137, pp. 435–444, 2015.
- [6] C. Hamzacebi and H. A. Es, "Forecasting the annual electricity consumption of Turkey using an optimized grey model," *Energy*, vol. 70, pp. 165–171, 2014.
- [7] D. Niu and M. Meng, "Research on seasonal increasing electric energy demand forecasting: a case in China," *Chinese Journal of Management Science*, vol. 8, no. 2, pp. 108–112, 2010.
- [8] N. Dongxiao, C. Zhiye, X. Mian, and X. Hong, "Combined optimum gray neural network model of the seasonal power load forecasting with the double trends," *Proceedings of the CSEE*, vol. 22, pp. 29–32, 2002.
- [9] Y. Wang, J. Wang, G. Zhao, and Y. Dong, "Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: a case study of China," *Energy Policy*, vol. 48, pp. 284–294, 2012.
- [10] Q. Zhanjun, "Medium and long-term load forecasting based on census X12-SARIMA model," in *Proceedings of the CSU-EPSA*, vol. 28, pp. 34–38, 2014.
- [11] Z. Ma and Q. Shu, "Short term load forecasting based on ESPRIT integrated algorithm," *Power System Protection and Control*, vol. 43, no. 7, pp. 90–96, 2015.
- [12] X. Zhu, W. Luan, and Y.-S. Zhu, "Analysis of macroeconomic time series by the HP filter," in *Proceedings of the 2nd International Conference on E-Business and E-Government (ICEE '11)*, pp. 3003–3006, Shanghai, China, May 2011.
- [13] Y. He, B. Wang, J. Wang, W. Xiong, and T. Xia, "Correlation between Chinese and international energy prices based on a HP filter and time difference analysis," *Energy Policy*, vol. 62, pp. 898–909, 2013.
- [14] R. S. Pindyck and D. L. Rubinfeld, *Econometric Models and Economic Forecasts*, China Machine Press, 4th edition, 1999.
- [15] F. Huanhuan and Z. Lingyun, *EViews Statistical Analysis and Application*, China Machine Press, Beijing, China, 2009.
- [16] C. W. Granger and R. Joyeux, "An introduction to long-memory time series models and fractional differencing," *Journal of Time Series Analysis*, vol. 1, no. 1, pp. 15–29, 1980.
- [17] Z. Huiming and L. Jing, *Bayesian Econometric Model*, Science Press, Beijing, China, 2009.
- [18] A. Lahiani and O. Scaillet, "Testing for threshold effect in ARFIMA models: application to US unemployment rate data," *International Journal of Forecasting*, vol. 25, no. 2, pp. 418–428, 2009.
- [19] N. Chaâbane, "A hybrid ARFIMA and neural network model for electricity price prediction," *International Journal of Electrical Power & Energy Systems*, vol. 55, pp. 187–194, 2014.
- [20] R. S. Tsay, *Analysis of Financial Time Series*, John Wiley & Sons, New York, NY, USA, 3rd edition, 2010.

