IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 9, 2017, accepted July 29, 2017, date of publication August 21, 2017, date of current version September 6, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2738029

Short-Term Load-Forecasting Method Based on Wavelet Decomposition With Second-Order Gray Neural Network Model Combined With ADF Test

BOWEN LI, JING ZHANG, YU HE, AND YANG WANG
School of Electrical Engineering, Guizhou University, Guiyang 550025, China

Corresponding author: Jing Zhang (zhangjing@gzu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51007009, in part by the Guizhou Science and Technology Foundation, China, under Grant [2016]1036, and in part by the Graduate Innovation Fund of Guizhou University, China, under Grant 2017065.

ABSTRACT Improving the accuracy of power system load forecasting is important for economic dispatch. However, a load sequence is highly nonstationary and hence makes accurate forecasting difficult. In this paper, a method based on wavelet decomposition (WD) and a second-order gray neural network combined with an augmented Dickey–Fuller (ADF) test is proposed to improve the accuracy of load forecasting. First, the load sequence is decomposed by WD to reduce the nonstationary load sequence. Then, the ADF test is adopted as the test method for the stationary load sequence of each decomposed component after WD in which the test results determine the best WD level. Finally, because forecasting the wavelet details characterized by high frequencies is difficult owing to its fluctuation, a second-order gray forecasting model is used to forecast each component after WD. Furthermore, to obtain the optimum parameters of the second-order gray forecasting model, the neural network mapping approach is used to build the second-order gray neural network forecasting model. The simulation result of a real load sequence verifies that the method proposed in this paper can effectively improve the load-forecasting accuracy.

INDEX TERMS Augmented Dickey–Fuller test, load forecasting, neural network mapping, second-order gray neural network forecasting, stationary load sequence.

I. INTRODUCTION

Load forecasting is one of the key challenges in the development of a power-supply plan and the balance between supply and demand of a power grid. It is also the foundation of power market operation as well as an important link in power planning. Improving the accuracy of load forecasting is helpful in improving the utilization ratio of power equipment and reducing energy consumption. Load-forecasting methods can be broadly divided into two categories. The first type of load-forecasting model is the statistical model such as the time-series method, which is developed based on the analysis of the inherent characteristics of historical data. The second type of load-forecasting model is based on related factors such as meteorological factors or price. However, quantifying the complicated interactions of different climate and load conditions needs more consideration, which, for the second model, is the auxiliary input. Therefore, we use the first model in the present study.

Until the present, many load-forecasting methods such as the continuation method, artificial neural networks [1]–[3], time-series method [4], Kalman filters [5], [6], support vector machine [7], and gray forecasting model [8], [9] are commonly used. In [1], load-forecasting models based on different neural networks were compared. The results showed that load-forecasting models based on the Elman neural network load were more accurate. The effect of climate change on load forecasting was discussed in [10]. In [11], a new short-term load-forecasting method was proposed, and the fitting method was discussed. The short-term load forecasting method based on root locus was presented, and the error convergence was guaranteed [12]. In [13], by considering the India electric power market as the research object, short-term load forecasting based on an extreme learning machine was proposed. A short-term load-forecasting model based on the feedback network and principal component decomposition was proposed in [14]. Big data theory was widely used in a

load-forecasting model in [15] and [16]. A load-forecasting model based on wavelet network was proposed in [17]. In [18], the spatial load forecasting of an urban distribution network based on the cloud and the cellular automata theories was presented. As far as the recent studies are concerned, most of the models were directly modeled and predicted. The nonstationary characteristics of the load are important factors that affecting the accuracy of forecasting.

The wavelet decomposition (WD) method has been widely investigated [19]–[22], as it can reduce the nonstationary characteristics of a series and improve forecast accuracy. In [19], the selection of a mother wavelet was studied. Typically, the load series is decomposed by a wavelet, and the components are then separately modeled [20]–[22]. With regard to the recent research works, two problems arise in the use of WD to process the time series. First, no theoretical basis is available for the determination of the WD level, and secondly, it lacks the ability to predict high-frequency components.

To solve the first problem, a method of selecting the WD series based on an augmented Dickey–Fuller (ADF) test is proposed in this paper. For the second problem, a load-forecasting model based on a second-order gray forecasting model is adopted. To obtain the optimum parameters of the second-order gray forecasting model, a neural network mapping approach is used to build the second-order gray neural network forecasting model [GNNM (2, 1)]. The simulation result verifies that the method effectively improves the load-forecasting accuracy.

II. WD BASED ON ADF TEST

A. WD

WD is a powerful tool for analyzing nonstationary and nonlinear signals. WD is used to process the load series in this work, and can reduce the nonstationarity of the load series and use it as the basis to improve the forecast accuracy. WD is expressed by the following equation:

$$\mathrm{W}T_x(b,a)=\frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty}x(t)\psi^*\left(\frac{t-b}{a}\right)dt, \tag{1}$$

where $\mathrm{W}T_x(b,a)$ is the coefficient that represents the extent to which load sequence $x\ (t)$ and the scaled mother wavelet match. Thus, for all a and b associated with a signal, the set of all wavelet coefficients $\mathrm{W}T_x(b,a)$ is related to the mother wavelet, where a and b are real and * denotes the complex conjugation. Scale parameter a controls the spread of the wavelet, and translation parameter b determines its central position.

We let $a=\frac{1}{2^s}$ and $b=\frac{k}{2^s}$, where s and k are integers. The discrete wavelet transform is described by the equation:

$$\omega_{k,s}=\int_{-\infty}^{+\infty}x(t)\psi_{k,s}(t)dt. \tag{2}$$

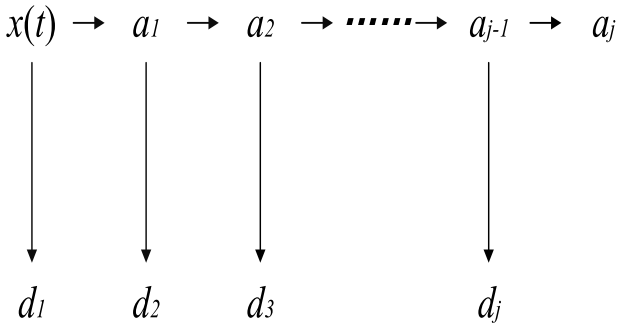

FIGURE 1. Mallat decomposition algorithm.

The inverse wavelet transform is described by the following equation:

$$x(t)=C_\varphi^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dadb}{a^2}\mathrm{W}T_x(b,a)\psi^*\left(\frac{t-b}{a}\right). \tag{3}$$

The Mallat algorithm [23] is a fast algorithm for computing the wavelet transform based on multiresolution analysis. The load series is projected into the scale space and wavelet subspace to obtain the approximate and detailed signals. The Mallat algorithm flow is shown in Fig. 1.

The decomposition process can be expressed as follows:

$$\begin{cases}a_{j+1}=H(a_j)\\ d_{j+1}=G(d_j),\end{cases} \tag{4}$$

where j is the decomposition level of the Mallat algorithm. $H(\bullet)$ is the low-frequency decomposition function, and $G(\bullet)$ is the high-frequency decomposition function, which are similar to low-pass filters and high-pass filters, respectively. During this process, the reconstruction only needs the coefficient vectors, which are generated by down sampling the length of the series into half. Therefore, before the reconstruction, the coefficient should be modified by allocating zeros between the samples.

$$\begin{cases}A_j=(H^*)^j a_j\\ D_j=(H^*)^{j-1}G^* d_j,\end{cases} \tag{5}$$

where H^* is the dual operator of H and G^* is the dual operator of G.

After this process is completed, load sequence x can be smoothened by eliminating the high frequencies or by separating it into high and low frequencies as follows:

$$x(t)=\sum_{\mathrm{i}=1}^{j}D_i(t)+A_j(t), \tag{6}$$

where $A_j(t)$ is an approximation signal or a fundamental component and $D_j(t)$ is denoted as a detailed signal or a high-frequency component.

To obtain the optimum parameters of the second-order gray forecasting model, a neural network mapping approach is used to build the GNNM (2, 1). The respective outputs are reconstructed to obtain the final forecasted load.

B. DETERMINATION OF THE BEST WD LEVEL BASED ON ADF TEST

The selection of the number of decomposition layers and the mother wavelet affects the forecast results. The selection method of the mother wavelet was studied in [19]. In the present work, the optimal WD level is studied. When the decomposition level is increased, the low-frequency components become more stable, and the forecasting accuracy of the fundamental component increases. However, the number of high-frequency components increases in addition. Thus, the forecast accuracy decreases as the number of components increases. Therefore, we need to find a balance between the WD level and the stability of the wavelet components.

To determine the balance, a model that evaluates the stability of the components after WD and that determines the optimal decomposition level based on the ADF test is proposed in this paper. The ADF test is an improved method of the unit root test. Its principle is to check whether a unit root is present in a sequence. If no unit root is present, the sequence is stationary; otherwise, it is nonstationary. The ADF test is usually used to test the stationarity of an economic time series. The determination of the optimal WD level by the ADF test is described as follows:

Step 1: We set $j = 1$.

Step 2: The load sequence is transacted by the j-level wavelet transformation to obtain the different components.

Step 3: The ADF test is used to evaluate the stability of the components after WD.

Step 4: If each component is stable, j is the best WD level. Otherwise, $j = j + 1$, and Step 2 is repeated.

At the same time, the ADF test can not only ensure the stability of the sequence and improve the forecast accuracy but also eliminate the many WD levels. Using this method, we derive the optimal layer of WD.

III. GRAY MODEL

Gray system theory was put forward as early as in the 1980s. It is characterized by using gray mathematics to deal with uncertainty and makes full use of known data to find the law of a system. A forecast model based on the gray-system theory has the advantages of being a simple model, requires less historical data, has high forecast accuracy, can be conveniently calculated, and does not need to consider the distribution.

GM (N, M) is a gray model whose order is N, and the number of variables is M. Until the present, GM (1, 1) has been widely used owing to its simplicity. However, the GM (1, 1) model contains only one exponential component and only one characteristic root; thus, simulating the time series characterized by a large oscillation is difficult. The second-order gray model [GM (2, 1)] has two eigenvalues, which reflect the monotonous and nonmonotonous changes. As a result, the low-frequency components and high-frequency components with significant oscillation characteristics can be well simulated. Thus, the GM (2, 1) model is adopted in the current study.

A. GM (2, 1)

The components of the historical load data decomposed by a wavelet are shown as follows:

$$X^{(0)} = \left[x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\right], \tag{7}$$

where n is the number of sequences.

Superimposing these sequences produces new sequences, which are defined as 1-AGO [using superscript "(1)"], i.e.,

$$X^{(1)} = \left[x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)\right], \tag{8}$$

$x^{(1)}(\cdot)$ is defined as:

$$x^{(1)}(t) = \sum_{i=1}^{t} x^{(0)}(i), \quad t = 1, 2 \ldots n. \tag{9}$$

We set up the second-order differential model based on the first-order accumulated generation operation as follows:

$$\frac{\mathrm{d}^2 x^{(1)}}{dt^2} + a_1 \frac{d^2 x^{(1)}}{dt} a_2 x^{(1)} = b. \tag{10}$$

This equation is solved as follows:

$$x^{(1)}(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} + \frac{b}{a_2}. \tag{11}$$

Equation (11) is the analytical expression of the forecasted value λ_1 and λ_2 are the characteristic roots of the characteristic equation $\lambda^2 + a_1\lambda + a_2 = 0$. Finally, we obtain the final forecast results using the following:

$$x^{(0)}(t) = x^{(1)}(t) - x^{(1)}(t-1). \tag{12}$$

B. GRAY NEURAL NETWORK MODEL (GNNM)

The parameters of a gray model are usually computed by least square estimation, which needs to update the calculation. A large amount of computation is required to accelerate the training speed of the model. To improve the computation efficiency, least square estimation is used to estimate the initial value of the gray model. Then the gray neural network is trained to obtain the optimal model parameters.

To accelerate the training speed of the gray neural network and determine the initial value of a_1, a_2, and b in (10) (the initial weights of the gray neural network), the initial values are computed by least square estimation. The formula is as follows:

$$A = [a_1 \quad a_2 \quad b]^T = \left(B_N^T B_N\right)^{-1} B_N^T Y_N, \tag{13}$$

where

$$B_N = \begin{bmatrix} -x^{(0)}(2) & -z^{(1)}(2) & 1 \\ -x^{(0)}(3) & -z^{(1)}(3) & 1 \\ \vdots & \vdots & \vdots \\ -x^{(0)}(n) & -z^{(n)}(n) & 1 \end{bmatrix}$$

$$Y_N = \begin{bmatrix} x^{(0)}(2) & x^{(0)}(1) \\ x^{(0)}(3) & x^{(0)}(2) \\ \vdots & \vdots \\ x^{(0)}(n) & x^{(0)}(n-1) \end{bmatrix}$$

$$z^{(1)}(t) = 0.5x^{(1)}(t) + 0.5x^{(1)}(t-1)$$

$$t = 2, 3 \ldots n$$

Parameters C_1 and C_2 are obtained by the following derivation. We use the first-order difference to replace the integral term as follows:

$$\frac{dx^{(1)}}{dt} = x^{(1)}(t) - x^{(1)}(t-1) = x^{(0)}(t). \tag{14}$$

We take the derivative of (11) with respect to t, i.e.,

$$\frac{dx^{(1)}}{dt} = C_1\lambda_1 e^{\lambda_1 t} + C_2\lambda_2 e^{\lambda_2 t}. \tag{15}$$

Substituting (15) into (14) yields

$$x^{(0)}(t) = C_1\lambda_1 e^{\lambda_1 t} + C_2\lambda_2 e^{\lambda_2 t}. \tag{16}$$

Parameters C_1 and C_2 are obtained by solving (11) and (16).

To construct the GNNM, (11) is transformed as follows:

$$\begin{aligned} x^{(1)}(t) &= C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} + \frac{b}{a_2} \\ &= \left(\frac{C_1 e^{\lambda_1 t}}{1+e^{\lambda_1 t}} + \frac{b}{a_2}\frac{1}{1+e^{\lambda_1 t}}\right)\left(1+e^{\lambda_1 t}\right) \\ &\quad + \frac{C_2 e^{\lambda_2 t}}{1+e^{\lambda_2 t}}\left(1+e^{\lambda_2 t}\right) \\ &= \left(\frac{C_1}{1+e^{-\lambda_1 t}} + \frac{b}{a_2}\left(1 - \frac{1}{1+e^{-\lambda_1 t}}\right)\right)\left(1+e^{\lambda_1 t}\right) \\ &\quad + \frac{C_2}{1+e^{-\lambda_2 t}}\left(1+e^{\lambda_2 t}\right) \\ &= \left(\frac{C_1}{1+e^{-\lambda_1 t}} - \frac{b}{a_2}\frac{1}{1+e^{-\lambda_1 t}}\right)\left(1+e^{\lambda_1 t}\right) \\ &\quad + \frac{C_2}{1+e^{-\lambda_2 t}}\left(1+e^{\lambda_2 t}\right) - \left(-\frac{b}{a_2}\left(1+e^{\lambda_1 t}\right)\right) \end{aligned} \tag{17}$$

According to (17), a neural network map is constructed, as shown in Fig. 2.

Then, the optimal parameters of the model, i.e., (17), can be obtained. The neural network learning process follows this procedure.

Step 1: The initial weights [see (13) to (16)] and the thresholds of the network are entered.

$$\boldsymbol{U} = [U_1 \quad U_2] = [\lambda_1 \quad \lambda_2]$$

$$\boldsymbol{V} = \begin{bmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \end{bmatrix} = \begin{bmatrix} C_1 & -\frac{b}{a_2} & 0 \\ 0 & 0 & C_2 \end{bmatrix}$$

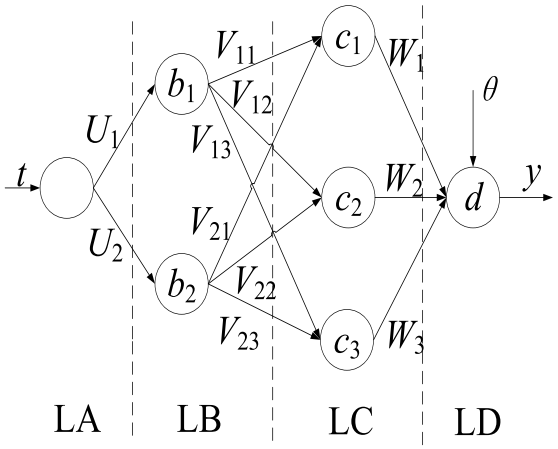

FIGURE 2. Schematic of the GNNM (2, 1) network. LA: Layer A. LB: Layer B. LC : Layer C. LD : Layer D.

$$\begin{aligned} \boldsymbol{W} &= [W_1 \quad W_2 \quad W_3]^{\mathrm{T}} \\ &= \left[1+\mathrm{e}^{U_1 t} \quad 1+\mathrm{e}^{U_1 t} \quad 1+\mathrm{e}^{U_2 t}\right]^{\mathrm{T}} \\ &= \left[1+\mathrm{e}^{\lambda_1 t} \quad 1+\mathrm{e}^{\lambda_1 t} \quad 1+\mathrm{e}^{\lambda_2 t}\right]^{\mathrm{T}} \end{aligned}$$

The threshold value of the LD layer is expressed as

$$\theta = -\frac{b}{a_2}(1+e^{\lambda_1 t}).$$

Step 2: The output of each layer is calculated.

The output of the LB neurons is defined as

$$b_1(t) = \frac{1}{1+e^{\lambda_1 t}}$$

$$b_2(t) = \frac{1}{1+e^{\lambda_2 t}},$$

The output of the LC neurons is defined as

$$c_1(t) = V_{11}b_1(t)$$

$$c_2(t) = V_{12}b_1(t)$$

$$c_3(t) = V_{23}b_2(t).$$

The output of the LD neurons is defined as

$$d(t) = y_1(t) = W_1c_1(t) + W_2c_2(t) + W_3c_3(t).$$

Step 3: The inverse error is calculated.

The LD layer error is defined as

$$\delta_d = y(t) - y_1(t).$$

Here, $y(t)$ is the actual data.

The LC layer error is defined as

$$\delta_{c_1} = \delta_d W_1$$

$$\delta_{c_2} = \delta_d W_2$$

$$\delta_{c_3} = \delta_d W_3.$$

The LB layer error is defined as

$$\delta_{b_1} = \frac{1}{1+e^{-\lambda_1 t}}\left(1-\frac{1}{1+e^{-\lambda_1 t}}\right)\left(\delta_{c_1}V_1+\delta_{c_2}V_2\right)$$
$$\delta_{b_2} = \frac{1}{1+e^{-\lambda_2 t}}\left(1-\frac{1}{1+e^{-\lambda_2 t}}\right)\delta_{c_3}V_3.$$

Step 4: The weights and thresholds are updated.

ΔU and ΔV are the correction weights of $\boldsymbol{U}$ and $\boldsymbol{V}$ respectively, η is the learning rate, and μ is the inertia coefficient.

$$\Delta U_1(\mathrm{s}) = \mu\Delta U_1(\mathrm{s}-1)+\eta\delta_{b_1}t$$
$$\Delta U_2(\mathrm{s}) = \mu\Delta U_2(\mathrm{s}-1)+\eta\delta_{b_2}t$$
$$\Delta V_1(\mathrm{s}) = \mu\Delta V_1(\mathrm{s}-1)+\eta\delta_{c_1}t$$
$$\Delta V_2(\mathrm{s}) = \mu\Delta V_2(\mathrm{s}-1)+\eta\delta_{c_2}t$$
$$\Delta V_3(\mathrm{s}) = \mu\Delta V_3(\mathrm{s}-1)+\eta\delta_{c_2}t.$$

The remaining correction for matrix $\boldsymbol{V}$ is zero, and s represents the number of trainings.

$$U(s+1) = U(s)+\Delta U(s)$$
$$V(s+1) = V(s)+\Delta V(s).$$

Matrix $\boldsymbol{W}$ is updated as follows:

$$W_1 = W_2 = 1+e^{U_1 t}$$
$$W_3 = 1+e^{U_2 t}.$$

Step 5: Steps 2–4 are repeated until the convergence condition is reached.

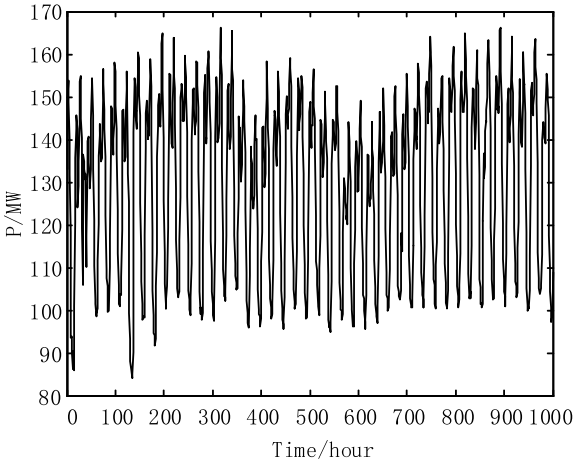

FIGURE 3. Load time series. The load profile of the time-series in 1000 time intervals (approximately 42 days).

IV. CASE STUDIES

A. HISTORICAL DATA

The 1000 data points of a continuous load of a substation in Southwest China in 2016 whose interval is 1 h are considered as the research object. The first 900 measured data in the sample are selected for the training, and the final 100 measured data are used for the test. Fig. 3 shows the samples.

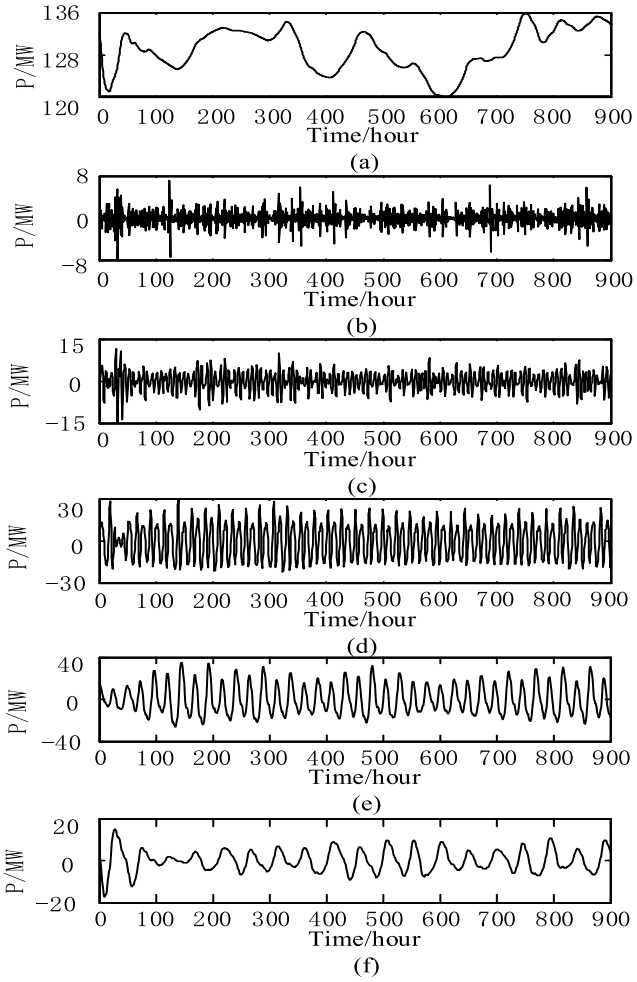

FIGURE 4. Decomposed electric load using discrete wavelet transform. (a) A5. (b) D1. (c) D2. (d) D3. (e) D4. (f) D5.

Some of the existing studies that use the wavelet transform in electrical load forecasting usually used the fourth-order Daubechies wavelet. From the ADF test, we choose the WD level as five. Given a signal s of length N, the discrete wavelet transform consists of log2 N stages at most, and the samples decompose 9 layers at most. Fig. 4 shows the electrical load decomposed by the frequency. Each component has a stable sequence.

B. FORECAST RESULTS

We clearly show the forecasted results based on WD with a second-order gray neural network model [WD–GNNM (2, 1)] of each component after the WD, as shown in Fig. 5.

The WD–Elman (five layers of WD combined with the Elman neural network) and GNNM (2, 1) are used to compare the performance of the proposed WD–GNNM (2, 1) [five layers of WD combined with the GNNM (2, 1)]. The load-forecasting results and the results of the different models are shown in Fig. 6.

C. RESULT ANALYSIS

First, to verify the ADF test to determine the optimal layer of the WD, the stability results based on the ADF test from

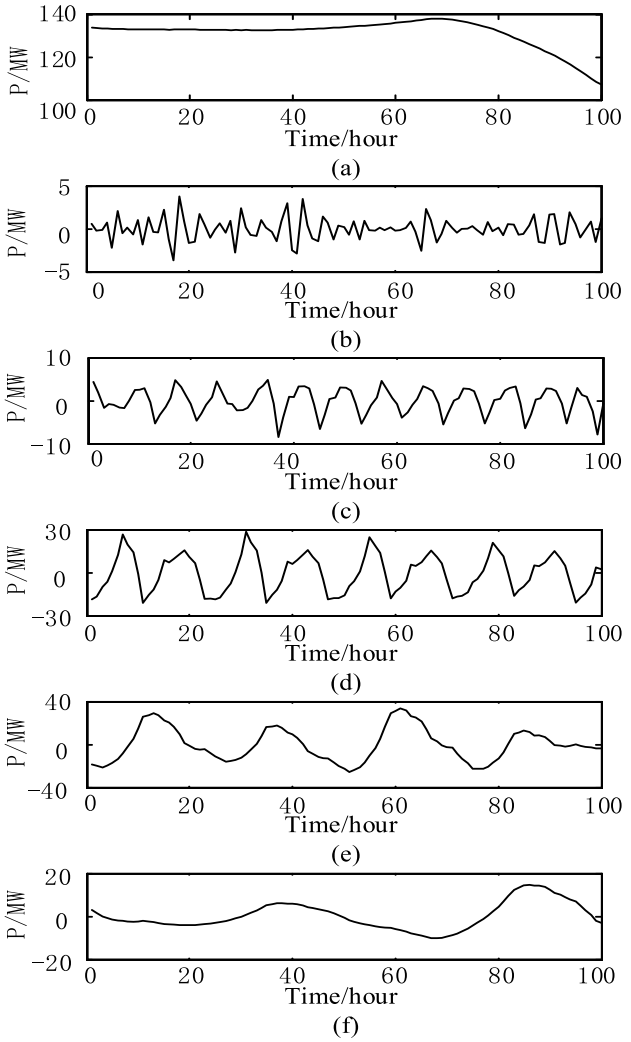

FIGURE 5. Forecasting results based on WD-GNNM (2, 1) of each component. (a) A5. (b) D1. (c) D2. (d) D3. (e) D4. (f) D5.

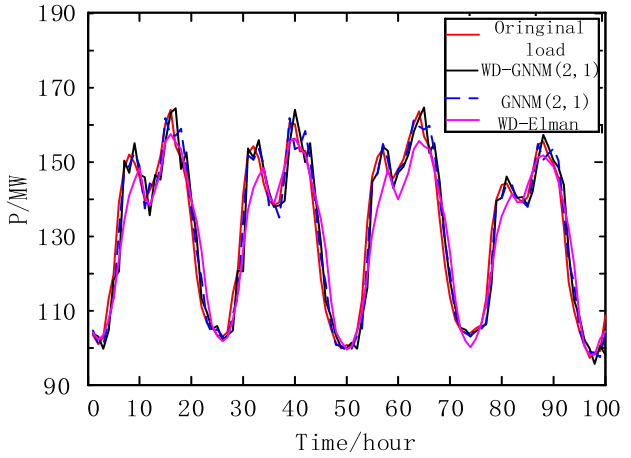

FIGURE 6. Load forecasting results of different models.

the single- to the nine-layer WDs are listed in Table 1. We evaluated our results using the mean absolute percentage error (MAPE). The computed MAPEs for different periods

TABLE 1. ADF test results.

component	*j*-layers of decomposition								
	j=1	*j*=2	*j*=3	*j*=4	*j*=5	*j*=6	*j*=7	*j*=8	*j*=9
A_j	0	0	0	0	1	1	1	1	1
D_1	1	1	1	1	1	1	1	1	1
D_2	-	1	1	1	1	1	1	1	1
D_3	-	-	1	1	1	1	1	1	1
D_4	-	-	-	1	1	1	1	1	1
D_5	-	-	-	-	1	1	1	1	1
D_6	-	-	-	-	-	1	1	1	1
D_7	-	-	-	-	-	-	1	1	1
D_8	-	-	-	-	-	-	-	1	1
D_9	-	-	-	-	-	-	-	-	1

TABLE 2. MAPE index of different decomposition layers.

decomposition layers	MAPE %
1	2.52
2	2.41
3	2.42
4	2.41
5	2.40
6	2.40
7	2.41
8	2.41
9	2.41

TABLE 3. MAPE index of different methods.

model	MAPE %
WD- Elman	3.94
GNNM(2,1)	3.23
WD-GNNM(2,1)	2.40

are listed in Table 2.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{|y' - y(i)|}{y(i)}\right) \times 100, \tag{18}$$

where $y'(i)$ is the forecast value, $y(i)$ is the real data, and N is the number of sequences.

In Table 1, "0" denotes that the sequence is nonstationary, and "1" denotes that the sequence is stationary. The ADF test results listed in Table 1. In the single-layer WD, A_1 is a nonstationary component, and D_1 is a stationary component. By continuing the decomposition of A_1, we obtain the results of the two-layer WD; however, A_2 remains a nonstationary component. With the continuous decomposition of A_j, each component becomes stable in the five-layer WD. As can be seen from Table 2, when the layer of the WD increases, the error is reduced. In the five-layer WD, the fundamental component is a stationary component. The cumulative errors is minimum at this time. When the decomposition level is more than five, the forecast accuracy of each component is guaranteed because all components are stable. However, with the progress of the WD, the cumulative errors increase.

Table 3 and Fig. 6 show the detailed error analysis of the proposed WD–GNNM (2, 1) and two other methods. The theoretical analysis and experimental evaluation reveal that

the proposed WD–GNNM (2, 1) is an optimal solution that will be beneficial for the establishment of a model with high accuracy for a load. We have shown in this paper that WD can improve the forecast accuracy.

V. CONCLUSION

To accurately predict a load with nonstationary characteristics, a short-term load-forecasting method based on WD with a second-order gray neural network was proposed in this paper. The ADF test was adopted as a test method for the stationarity of each decomposed component after WD. This paper draws the following main conclusions. First, the wavelet transform can reduce the nonstationarity of a load sequence and improve the forecast accuracy. Secondly, the method of determining the optimal WD level based on the ADF test can find the balance between the optimal WD level and the stability of the wavelet components; the forecast errors are reduced to the lowest. Third, the proposed WD–GNNM (2, 1) can effectively improve the forecasting accuracy. We believe that the proposed approach can be extended to other forecasting applications.

REFERENCES

[1] N. K. Singh, A. K. Singh, and M. Tripathy, "A comparative study of BPNN, RBFNN and ELMAN neural network for short-term electric load forecasting: A case study of Delhi region," in *Proc. Ind. Inf. Syst. (ICIIS)*, Dec. 2014, pp. 1–6.
[2] C. Cecati, J. Kolbusz, P. Różycki, P. Siano, and B. M. Wilamowski, "A novel RBF training algorithm for short-term electric load forecasting and comparative studies," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6519–6529, Oct. 2015.
[3] H. Kebriaei, N. Babak Araabi, and A. Rahimi-Kian, "Short-term load forecasting with a new nonsymmetric penalty function," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 1817–1825, Nov. 2011.
[4] Y.-H. Guo, X.-P. Shi, and X.-D. Zhang, "A study of short term forecasting of the railway freight volume in China using ARIMA and Holt-Winters models," in *Proc. Supply Chain Manag. Inf. Syst. (SCMIS)*, 2010, pp. 1–6.
[5] H. M. Al-Hamadi and S. A. Soliman, "Fuzzy short-term electric load forecasting using Kalman filter," *IET Proc.-Generat., Transmiss. Distrib.*, vol. 153, no. 2, pp. 217–227, Mar. 2006.
[6] C. Guan, P. B. Luh, L. D. Michel, and Z. Chi, "Hybrid Kalman filters for very short-term load forecasting and prediction interval estimation," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 3806–3817, Nov. 2013.
[7] W.-M. Lin, C.-S. Tu, R.-F. Yang, and M.-T. Tsai, "Particle swarm optimisation aided least-square support vector machine for load forecast with spikes," *IET Proc.-Generat., Transmiss., Distrib.*, vol. 10, no. 5, pp. 1145–1153, Apr. 2016.
[8] W. Dapeng and W. Bingwen, "Medium- and long-term load forecasting based on variable weights buffer grey model," *Power Syst. Technol.*, vol. 37, no. 1, pp. 167–171, Jan. 2013.
[9] G. Shaoyun, J. Ousha, and L. Hong, "A gray neural network model improved by genetic algorithm for short-term load forecasting in price-sensitive environment," *Power Syst. Technol.*, vol. 36, no. 1, pp. 224–229, Jan. 2012.
[10] M. O. Oliveira, D. P. Marzec, G. Bordin, A. S. Bretas, and D. Bernardon, "Climate change effect on very short-term electric load forecasting," in *Proc. IEEE Trondheim PowerTech*, Jun. 2011, pp. 1–7.
[11] C. E. Borges, Y. K. Penya, and I. Fernandez, "Evaluating combined load forecasting in large power systems and smart grids," *IEEE Trans. Ind. Informat.*, vol. 9, no. 3, pp. 1570–1577, Aug. 2013.
[12] Z. Shen, X. Wu, M. J. Guerrero, and Y. Song, "Model-independent approach for short-term electric load forecasting with guaranteed error convergence," *IET Control Theory, Appl.*, vol. 10, no. 12, pp. 1365–1373, Aug. 2016.
[13] S. K. Dash and D. Patel, "Short-term electric load forecasting using extreme learning machine—A case study of Indian power market," in *Proc. Power, Commun. Inf. Technol. Conf. (PCITC)*, 2015, pp. 961–966.
[14] F. M. Bianchi, E. D. Santis, A. Rizzi, and A. Sadeghian, "Short-term electric load forecasting using echo state networks and PCA decomposition," *IEEE Access*, vol. 3, pp. 1931–1943, 2015.
[15] P. Zhang, X. Wu, X. Wang, and S. Bi, "Short-term load forecasting based on big data technologies," *CSEE J. Power Energy Syst.*, vol. 1, no. 3, pp. 59–67, Sep. 2015.
[16] W. Dewen and S. Zhiwei, "Big data analysis and parallel load forecasting of electric power user side," *Proc. CSEE*, vol. 35. no. 3, pp. 527–537, Feb. 2015.
[17] C. Guan, P. B. Luh, L. D. Michel, Y. Wang, and P. B. Friedland, "Very short-term load forecasting: Wavelet neural networks with data pre-filtering," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 30–41, Feb. 2013.
[18] L. Zifa, P. Chengcheng, W. Zeli, and L. Ke, "Spatial load forecasting for distribution network based on cloud theory and cellular automata," *Proc. CSEE*, vol. 35, no. 3, pp. 98–105, Apr. 2013.
[19] Z. Deliang, L. Jiwei, L. Jizhen, L. Yu, and X. Xie, "Application of wavelet multi-scale analysis for wear characteristics," *Proc. CSEE*, vol. 32, no. 23, pp. 126–131, Aug. 2012.
[20] B.-G. Koo, H.-S. Lee, and J. Park, "A study on short-term electric load forecasting using wavelet transform," in *Proc. Innov. Smart Grid Technol. Conf. Europe (ISGT-Eur.)*, 2014, pp. 1–6.
[21] H. Shao, X. Deng, and F. Cui, "Short-term wind speed forecasting using the wavelet decomposition and AdaBoost technique in wind farm of East China," *IET Proc. Generat., Transmiss., Distrib.*, vol. 10, no. 11, pp. 2585–2592, Apr. 2016.
[22] Y.-X. Yang and D. Liu, "Short-term load forecasting based on wavelet transform and least square support vector machines," *Power Syst. Technol.*, vol. 29, no. 13, pp. 60–64, Jul. 2005.
[23] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

BOWEN LI received the B.S. degree from Wuhan University, Wuhan, China, in 2015. He is currently pursuing the M.S. degree with Guizhou University, Guiyang, China. His research interests include power system operation and control, power system load forecasting, wind power forecasting, and power system stability.

JING ZHANG received the B.S. and M.S. degrees from Guizhou University, Guiyang, China, and the Ph.D. degree from Huazhong University of Science and Technology, Wuhan, China. He is currently a Professor with Guizhou University. His research interests include power system operation and control and power system stability.

YU HE received the B.S. and M.S. degrees from Guizhou University, Guiyang, China. She is currently a Professor with Guizhou University. Her research interests include power system operation and control and power system voltage stability.

YANG WANG received the B.S. degree from Shanghai University of Electric Power, Shanghai, China, in 2015. He is currently pursuing the M.S. degree with Guizhou University, Guiyang, China. His research interests include power system operation and control and power system stability.