Short-Term Load Forecasting: Similar Day-Based Wavelet Neural Networks

Ying Chen, Peter B. Luh, *Fellow, IEEE*, Che Guan, Yige Zhao, Laurent D. Michel, Matthew A. Coolbeth, Peter B. Friedland, and Stephen J. Rourke, *Senior Member, IEEE*

***Abstract*—In deregulated electricity markets, short-term load forecasting is important for reliable power system operation, and also significantly affects markets and their participants. Effective forecasting, however, is difficult in view of the complicated effects on load by a variety of factors. This paper presents a similar day-based wavelet neural network method to forecast tomorrow's load. The idea is to select similar day load as the input load based on correlation analysis, and use wavelet decomposition and separate neural networks to capture the features of load at low and high frequencies. Despite of its "noisy" nature, high frequency load is well predicted by including precipitation and high frequency component of similar day load as inputs. Numerical testing shows that this method provides accurate predictions.**

***Index Terms*—Neural network, short-term load forecasting, similar day, wavelet.**

I. Introduction

SHORT-TERM load forecasting (e.g., forecasting tomorrow's load at hourly intervals) has been essential for reliable power system operation. It becomes more important after power system deregulation since forecasted load is used by market operators to determine day-ahead market prices, and by market participants to prepare bids. Short-term load forecasting, however, is difficult in view of the complicated effects on load by a variety of factors.

Many short-term load forecasting methods have been developed, and representative methods include regressions, similar day methods, and neural networks. Regression methods [1], [2] assume that there are prespecified functional forms describing quantitative relationships between load and affecting factors (e.g., weekday index and weather). Functional coefficients are estimated through regression analysis of historical data.

Similar day methods are based on searching historical days that have weekday index and weather similar to the forecasted day [3], [4]. In these methods, the forecasted load is the load of one similar day or a combination of several similar days' load with appropriate adjustments. These methods are simple and intuitively appealing since load of the similar days and of the forecasted day are usually similar. However, these methods may not be sufficient enough to capture complex load features if used alone.

Neural network methods also assume a functional relationship between load and affecting factors, and estimate the functional coefficients by using historical data. Different from regression methods, the functional forms can be highly nonlinear. To forecast tomorrow's load, the standard neural network method [5]–[8] uses a single network with tomorrow's weekday index, predicted weather, and most recently available load (e.g., yesterday's load) as inputs. By using a single network, high frequency (fast changing) load features and low frequency (slow changing) load features are mixed together and treated without distinction. Since high frequency load features are usually less obvious than low frequency load features, these methods cannot well capture rapid changes in load. Additionally, since yesterday's load does not have a good enough correlation with tomorrow's load, the input load used in these methods has restricted prediction power, resulting in limited prediction accuracy.

In addition to the above representative methods, a method that combines wavelet decomposition and neural networks has been reported in the literature [9]. This method uses wavelet to decompose load into a low frequency component and three high frequency components. It opens the door for analyzing complex load features at different frequencies. However, since the high frequency load components are mistreated as noise, the high frequency load features are not appropriately captured.

Based on the above understanding, this paper presents a generic framework that combines similar day selection, wavelet decomposition, and neural networks to forecast tomorrow's load. The key idea is to select the similar day's load as the input load, apply wavelet to decompose it into a low frequency component and a high frequency component, and then use separate networks to predict the two components of tomorrow's load. The similar day's load is selected based on ISO New England's practice [3]. Correlation analysis shows that this load has higher prediction power than yesterday's load used in most NN methods. Also, high frequency load is not treated as noise following the literature [9]. Rather, the features of high frequency load are captured by including precipitation and the high frequency component of the similar day load as inputs to the high frequency network. The similar day load is

Manuscript received December 04, 2008; revised March 24, 2009. First published November 24, 2009; current version published January 20, 2010. This work was supported by a grant from ISO New England. Paper no. TPWRS-00929-2008.

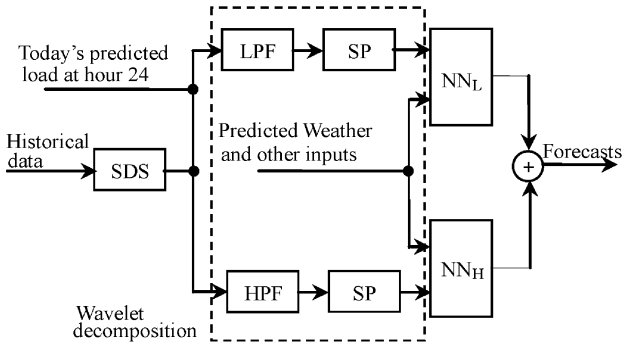
Y. Chen, P. B. Luh, C. Guan, and Y. Zhao are with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269-2157 USA.

L. D. Michel and M. A. Coolbeth are with the Department of Computer Science and Engineering, University of Connecticut, Storrs, CT 06269-2155.

P. B. Friedland and S. J. Rourke are with ISO New England, Holyoke, MA 01040 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2009.2030426



SDS: Similar day-based selection; LPF: Low pass filter; HPF: High pass filter; SP: Signal processing; NN_L: Low freq. network; NN_H: High freq. network.

Fig. 1. Overall structure of SIWNN.

SDS: Similar day-based selection; LPF: low-pass filter; HPF: high-pass filter; SP: Signal processing; $\mathrm{NN_L}$: Low freq. network; $\mathrm{NN_H}$: High freq. network.

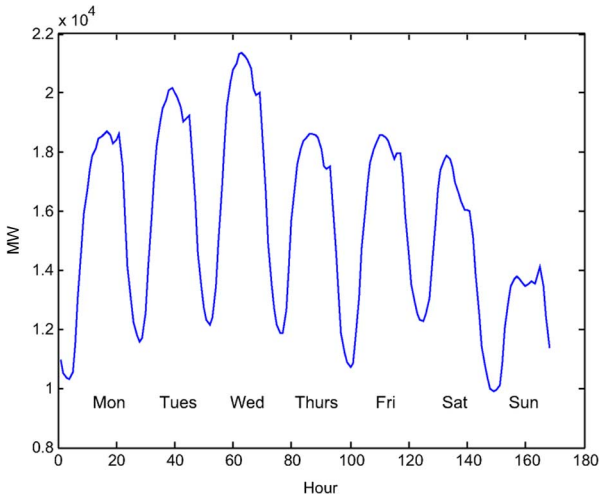


Fig. 2. Weekly patterns of New England load (August 25, 2003–August 31, 2003).

further supplemented by today's predicted load at hour 24 to capture the initial status of tomorrow's load. Numerical results for a simple example and for forecasting New England load in Section III show that our method produces accurate forecasts, and demonstrate the significant values of using similar day's load, wavelet decomposition, and separate neural networks. The method has been extended to include multilevel wavelet decompositions and holiday corrections for short-term load forecasting [10], and multilevel wavelet decompositions with data prefiltering for very short-term load forecasting [11] with very good results.

II. Similar Day-Based Wavelet Neural Networks

A similar day-based wavelet neural network method (SIWNN) is developed to predict tomorrow's load. It consists of similar-day based input selection, wavelet decomposition, and neural networks as depicted in Fig. 1. The inputs are selected

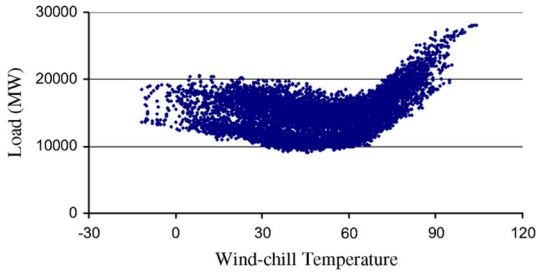


Fig. 3. Load versus T_{wc} (New England data for 2006).

from the information available by hour 9[1] today, and include historical actual load and weather until hour 8 today, predicted load for today's remaining hours (produced by SIWNN yesterday), and tomorrow's predicted weather.

In the following, Section II-A describes the major load affecting factors including weekday index and weather. Section II-B presents the selection of similar days and the use of their load as the input load, supplemented by today's predicted load at hour 24. Section II-C introduces wavelet decomposition used to decompose the input load into low and high frequency components. Section II-D then presents the two neural networks used to separately predict the low and high frequency load components. Forecasts obtained by using the two networks are added up to be the predicted load.

A. Weekday Index and Weather

Weekday index is an important load affecting factor in view that different days of a week generally have different load shapes as shown in Fig. 2. It can be seen that load shapes for Tuesday, Wednesday, and Thursday are alike, and load shapes for other days of a week are quite different.

Beyond weekday index, weather is the major drive for load. Weather information used in SIWNN includes wind-chill temperature, humidex, wind speed, cloud cover, and precipitation.

Since temperature that is felt could be much lower than air temperature in winter in view of the effects of wind, wind-chill temperature that measures the felt temperature is used for winter. According to [12], it is calculated based on air temperature and wind speed as follows:

$$T_{wc} = 35.74 + 0.6215T_a - 35.75v^{0.16} + 0.4245T_a v^{0.16} \quad (1)$$

where T_{wc}, T_a, and v, respectively, denote wind-chill temperature, air temperature, and wind speed. To quickly examine the effects of T_{wc} on load, the scatter plot of load versus T_{wc} is presented in Fig. 3 based on New England data, with horizontal axis representing T_{wc} and vertical axis representing load. It can be seen that the general shape is nonlinear, and has a V-shaped pattern. The large range of load values for a given T_{wc} is caused by the effects of other factors (weekday index, hour, and other weather information).

In view that a linear function is easier to learn than a nonlinear function due to its simplicity, the nonlinear data pattern in Fig. 3

[1]SIWNN is developed for ISO-NE's load forecasting performed around 9:00 am each day.

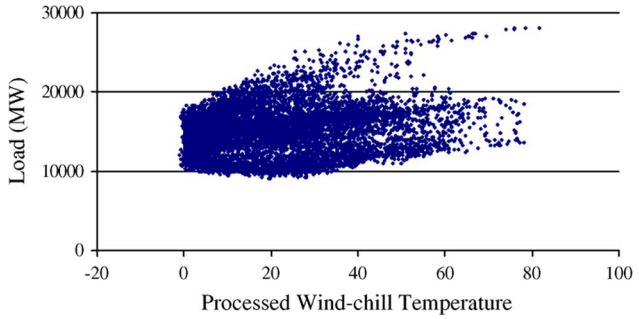


Fig. 4. Load versus T_{wc}^{p} (New England data for 2006).

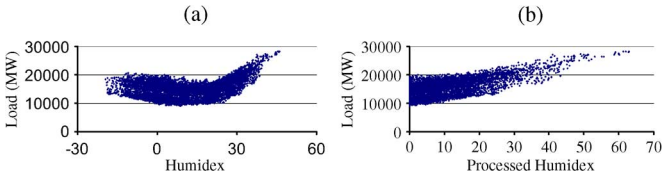


Fig. 5. Load versus humidex. (a) Original. (b) Processed.

is converted to a near linear pattern by processing T_{wc} with the following V-shaping function:

$$T_{wc}^{p} = |T_{wc} - c_{wc}| \tag{2}$$

where T_{wc}^{p} denotes the processed wind-chill temperature, and c_{wc} is a parameter determined based on Fig. 3. With this, load versus the processed wind-chill temperature is shown in Fig. 4.

The combined effects of heat and humidity cause high level of discomfort in summer. Therefore, humidex that measures the combined effect of heat and humidity is used for summer. According to [13], it is calculated based on air temperature T_a and dew point as follows:

$$H = T_a + 0.5555 \times \left(6.11 \times e^{5417.753 \times \left(\frac{1}{273.16} - \frac{1}{D}\right)} - 10\right) \tag{3}$$

where H denotes humidex, and D dew point. Similar to wind-chill temperature, H is also processed with a V-shaping function. Scatter plots of load versus H and load versus the processed H are, respectively, presented in Fig. 5(a) and (b).

For ease of neural network implementation, wind-chill temperature and humidex are used throughout the year. Cloud cover that measures sunshine intensity, and precipitation that measures rain/snow volume are also used.

B. Similar Day-Based Load Input Selection

Historical load is usually used as input for neural network-based prediction. A key question is how to properly select the days. To forecast the load of tomorrow (day D), the common practice is to use the most recently available load, i.e., the load of yesterday (D-2), and the load of one week ago (D-7, with the same weekday index) [5]–[8]. To examine the goodness of this practice, the scatter plot of actual load of the forecasted day (D) versus load of D-2 for New England 2006 data (including both weekdays and weekends) is presented in Fig. 6. It can be seen that the general pattern in Fig. 6 is linear. Correlation coefficient,

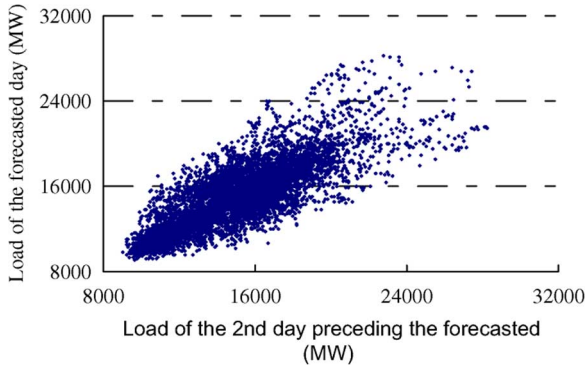


Fig. 6. Actual load of day D versus load of day $D-2$ (New England 2006).

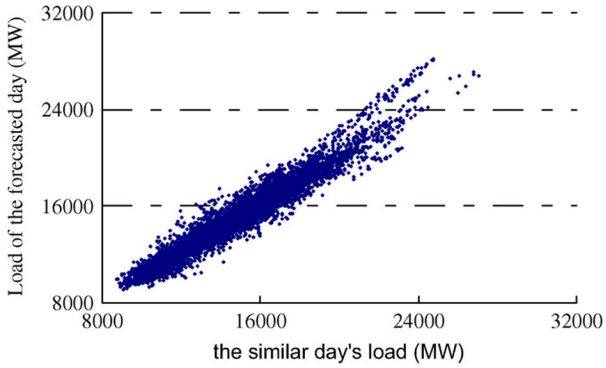


Fig. 7. Actual load of the forecasted day versus similar day's load (New England 2006).

a measure of prediction power contained in the D-2 load,[2] however, is only 0.67. This indicates that this selection may not be that good. To improve the prediction power, our idea is to use the similar day's load as the input load. For comparison purpose, the scatter plot of actual load of the forecasted day versus the similar day's load based on New England 2006 data is presented in Fig. 7 (also including both weekdays and weekends).

It is clear that Fig. 7 exhibits a much better linear pattern than Fig. 6. Correlation coefficient associated with Fig. 7 is 0.95, much higher than that for Fig. 6, indicating that the similar day load has a much higher prediction power than the most recent load.

The criteria to select similar days are based on ISO New England' operation procedures [3], and the selected day is required to have the same weekday index and similar weather to that of tomorrow. In this selection process, Tuesday, Wednesday, and Thursday are not differentiated following the rationale presented in Section II-A. To avoid seasonal variations, the selected day is also required to have its day-of-a-year index within a neighborhood of that of tomorrow. Let Θ denote the set of such historical days. Then the day with similar weather as tomorrow is selected by the following minimization process:

$$\min_{i} \sum_{t=1}^{24} \left| w^{f}(t) - w^{i}(t) \right|, \text{ where } i \in \Theta \tag{4}$$

[2]Correlation coefficient measures the association between the load of the day to be forecasted and the input load. The larger the association, the better the input load is [18].

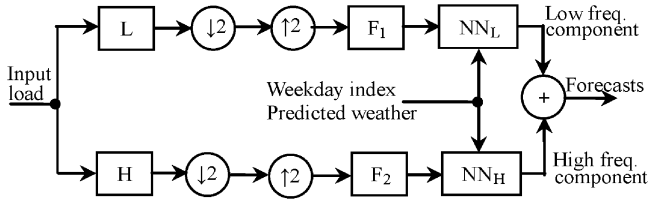


Fig. 8. Wavelet decomposition.

where the subscripts f and i, respectively, denote tomorrow (the forecasted day) and a historical day in Θ; w represents the weather factor under consideration: wind-chill temperature T_{wc} if tomorrow is a winter day, and humidex H if tomorrow is a summer day. For simplicity, T_{wc} is considered if tomorrow is a spring or fall day. If there are multiple solutions for (4), the day closest to tomorrow is then selected.

Despite of the high prediction power of the similar day's load, testing results (e.g., Example 3) show that prediction errors may be large if no other load is used. This is because the similar day's load does not have good information about the initial status of tomorrow's load, an important feature for the forecasting. To capture this feature, today's predicted load at hour 24 (produced by SIWNN yesterday) is supplemented to the similar day load. Testing results show that with this supplemental load, prediction errors are significantly decreased.

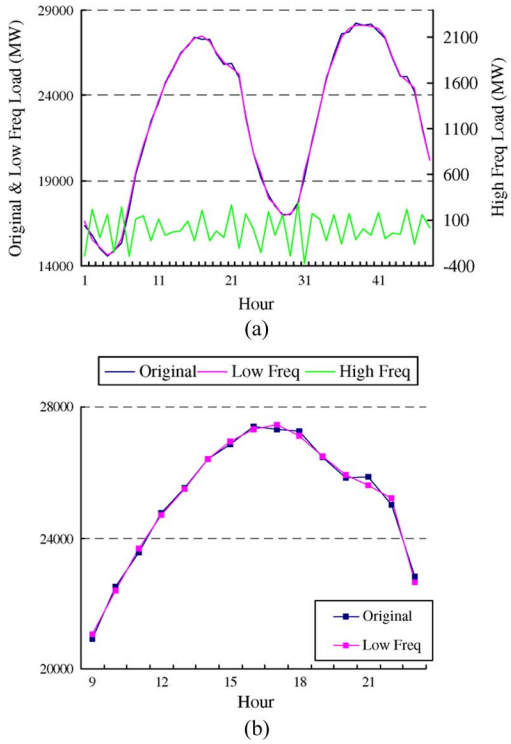


Fig. 9. (a) Wavelet decomposition results for New England load (August 1, 2006 and August 2, 2006). (b) Original and low frequency load (hour 10 to 24, August 2, 2006).

C. Decomposing Input Load

Daubechies wavelets are good for load forecasting since they are orthogonal wavelets, and will not cause information loss in the frequency domain. In SIWNN, Daubechies 4 wavelet (Db4) is used to decompose the input load into a low frequency component and a high frequency component. The decomposition is implemented by using a two-channel filter bank as depicted in Fig. 8. Application of other wavelets and multilevel decomposition are beyond the scope of this paper.[3]

For the low frequency channel in Fig. 8, L is the scaling function for Db4, and is basically a low-pass filter. It is used to filter out high frequency component from the selected similar day's load, and the output of L contains low frequency information of the load only. Following the standard filter bank design [14], three signal processing steps are after L: down-sampling to reduce data volume by dropping odd indexed data points, up-sampling to pad zeros to the down-sampled data so as to recover data length, and a low-pass filter F_1 to remove distortion caused by up-sampling (removing replicas of the signal spectrum) [15]. Symmetrically for the high frequency channel, H is the wavelet function for Db4, and is basically a high-pass filter. It is used to filter out low frequency component from the similar day's load. The wavelet function H is then followed by three signal processing steps similar to those for the low frequency channel, including down-sampling, up-sampling, and a high-pass filter F_2. The filters F_1 and F_2 are designed based on [14, Eq. (4.6)] for perfect data reconstruction, i.e., the similar day's load can be perfectly reconstructed from the decomposed low and high frequency components. This property is important since there will be no load feature loss by using such a decomposition. This decomposition is also applied to today's predicted load at hour 24. The decomposed load is then used in combination with other inputs for the two neural networks in the next subsection.

Computation in the above filtering steps requires the load temporally before and after the input load. In view of this, the entire historical load is decomposed in one shot, and the low and high frequency components for similar day's load are then identified from the decomposed results. To decompose today's predicted load at hour 24, today's predicted load at hours 21 to 23 is padded to its beginning, and today's predicted load at hours 1 to 3 is approximated as tomorrow's load at these hours and padded to its end.[4]

To illustrate the low and high frequency load features, New England load for August 1 and 2, 2006 (peak load days for summer 2006) is decomposed. The original load, and the decomposed low and high frequency load components, are depicted in Fig. 9(a), and the original and low frequency for some hours are zoomed in Fig. 9(b). The low frequency component has a clear pattern consistent with the original load but is smoother, while the high frequency component is noisy. Furthermore, the magnitude of the high frequency component, ranging from −400 MW to 400 MW, is much smaller than that of the low frequency component (ranging from 14 500 MW to 29 000 MW).

[3]Results on utilizing multilevel decomposition are reported in our recent papers [10] and [11] to be presented at the PES 2009 General Meeting.

[4]Db4 is a four-length wavelet, and operates on four adjacent load values each time. Therefore, three hours of load needs to be padded at each end of hour 24.

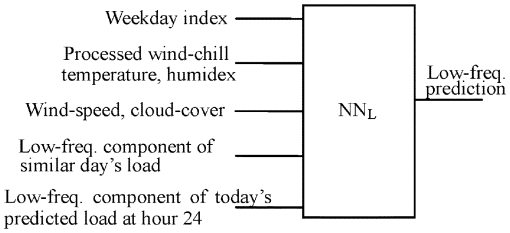


Fig. 10. Low frequency network.

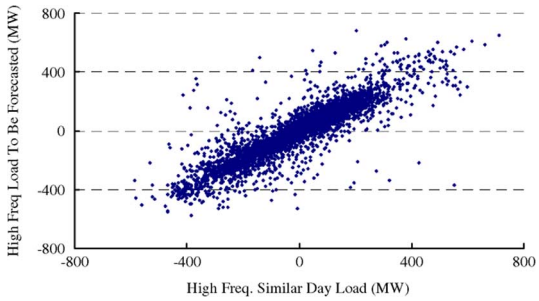


Fig. 11. Scatter plot of actual high frequency load of the forecasted day versus high frequency similar day's load (New England 2006).

D. Neural Networks

Two three-layer perceptron networks [16] are separately used for the low frequency component and the high frequency component.

For the low frequency network, inputs are selected based on our testing experience, and include weekday index, wind-chill temperature, humidex, wind speed, cloud cover, and low frequency components of the input load (the similar day's load and today's predicted load at hour 24) as shown in Fig. 10.[5] To improve network performance while providing the capabilities to forecast beyond (or below) the historical maximum (or minimum) load level, the input weather and load are normalized to values in (0.05, 0.95) by using the corresponding historical maximal (or minimal) values. Weekday index is coded by using a 7-length binary number, i.e., 1000000 for Monday, 0100000 for Tuesday, Wednesday and Thursday (following the rational presented in Section II-A), 0000100 for Friday, 0000010 for Saturday, and 0000001 for Sunday following [17].

For the high frequency network, the inputs are also selected based on testing experience, and include weekday index, wind-chill temperature, humidex, wind speed, cloud cover, high frequency components of the input load, and precipitation. The high frequency component of the similar day's load is a major input in view of its good correlation with tomorrow's high frequency load as shown in Fig. 11. Precipitation is another important input for this network since testing results (e.g., Example 3) show that prediction errors are large for rainy days if it is not used. The overall structure for the high frequency network is presented in Fig. 12.

The two networks are first trained by using historical actual data, with the similar day-based selection criteria presented in Section II-B applied to each day in the training period. For each network, the training process terminates when the training error is less than or equal to a specified threshold. To avoid over-fitting, for each network, its number of hidden neurons, input selection, and threshold value for terminating training are determined based on extensive testing. Furthermore, the two networks are trained with several years' historical data.

[5]Wind-chill temperature and humidex are used throughout the year for ease of implementation.

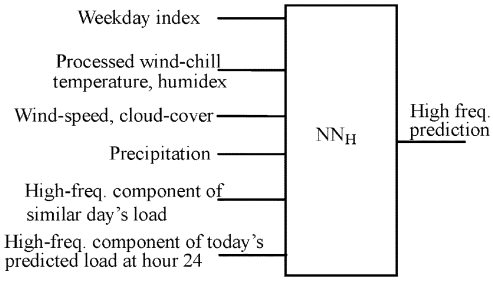


Fig. 12. High frequency network.

In the prediction phase, the two predictions generated by the two networks are added up to be tomorrow's forecasted load. The accuracy of the forecast can be evaluated when the actual values for tomorrow's load become available by using the standard measure mean absolute percentage errors (MAPE):

$$\text{MAPE} = \frac{1}{24}\sum_{t=1}^{24}\frac{\left|\widetilde{L}(t) - L(t)\right|}{L(t)} \times 100\% \tag{5}$$

where $L(t)$ and $\widetilde{L}(t)$, respectively, denote the actual and predicted values for tomorrow's load. MAPE is good when $L(t)$ [the denominator in (5)] is not close to zero.

III. Numerical Testing Results

The above method has been implemented in C++ on a Pentium-IV 2.67-GHz personal computer. Three examples are presented below. Example 1 uses a classroom-type problem to examine the value of using wavelet decomposition. Example 2 predicts New England 2006 load, demonstrates the values of wind-chill temperature, humidex, and weather preprocessing presented in Section II-A, and examines sensitivity of prediction to weather forecasting errors. Example 3 predicts New England 2007 load, and examines the effects of using wavelet decomposition, similar day's load, and supplemental load (today's predicted load at hour 24) on prediction accuracy. Predicted weather used in Example 2 and 3 were from ISO-NE (provided by weather stations through subscriptions). In each testing case, our method SIWNN is used as the base method, and the method used for comparison is implemented by removing the tested feature from SIWNN. The standard NN method is implemented by using a single network with weekday index, weather, and load of day D-2 used as inputs.

Example 1: Consider the following signal:

$$y(t) = 20sin(t) + sin(10t)$$

which is composed of a low frequency component $20sin(t)$ and a high frequency component $sin(10t)$. Five hundred noisy data sets $(t, \tilde{y}(t))$ were randomly generated for training with

$$\tilde{y}(t) = y(t) + \varepsilon(t)$$

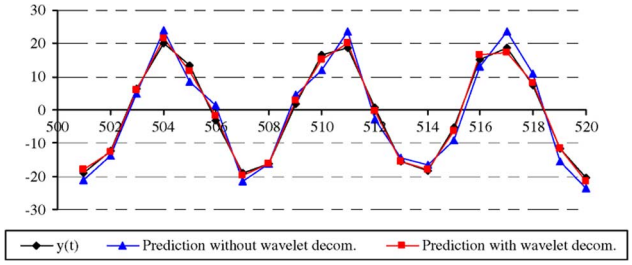


Fig. 13. Results for Example 1.

where $t \in [1, 2 \ldots 500]$, and $\varepsilon(t)$ is an i.i.d noise uniformly distributed in between 0 and 0.5. The objective is to predict $y(t)$ for $t \in [501, 502 \ldots 520]$.

Our method SIWNN and the standard NN method that uses a single neural network without wavelet decomposition are tested. Predictions obtained by using both methods and true $y(t)$ are plotted in Fig. 13. It can be seen that the prediction obtained by using SIWNN is closer to the true $y(t)$ as compared to the prediction obtained by using the standard NN method. To quantify the difference in predictions, mean absolute error (MAE)[6]

$$\text{MAE} = \frac{1}{20} \sum_{t=501}^{520} |\tilde{y}(t) - y(t)|$$

is calculated, and MAE is 0.86 for SIWNN, and 2.89 for the standard NN method. This shows the value of using wavelet decomposition.

Example 2: This example predicts New England 2006 load, and includes four cases. Case 1 shows the value of using wind-chill temperature as compared to air temperature (the standard practice) in predicting winter load. Case 2 shows the value of humidex as compared to the standard practice of using dew point in predicting summer load. Case 3 examines the value of data preprocessing (V-shaping wind-chill temperature and humidex), with SIWNN tested with and without the preprocessing. Case 4 then evaluates the sensitivity of prediction to weather forecasting errors.

Case 1: This case compares using wind-chill temperature versus using air temperature. The latter is implemented by replacing wind-chill temperature by air temperature in the similar day selection criterion (4) and for the neural networks' inputs. Training period is from 2003 to 2005, and prediction period is winter 2006 (January, February, and March). MAPE presented in Table I shows the value of wind-chill temperature.

Case 2: This case compares using humidex versus using dew point. The latter is implemented by replacing humidex by dew point in the similar day selection criterion (4) and for the neural networks' inputs. Training period is from January 2003 to May 2006, and prediction period is summer 2006 (June, July, and August). MAPE presented in Table II shows the value of humidex.

Case 3: SIWNN is used to predict New England 2006 load with and without data preprocessing [v-shaping wind-chill temperature and humidex based on (2)]. Training period is from 2003 to 2005. MAPEs presented in Table III show that the preprocessing improves prediction accuracy.

[6]MAPE is not good for this example since $y(t)$ is zero or negative sometimes.

TABLE I
MAPE (%) FOR NEW ENGLAND 2006 WINTER LOAD (CASE 1 IN EXAMPLE 2)

	MAPE(M1)	MAPE(M2)	MAPE(M1) - MAPE(M2)
Jan	1.68	1.65	0.03
Feb	1.52	1.48	0.04
March	1.55	1.52	0.03

M1: With air temperature; M2: With wind-chill temperature

TABLE II
MAPE (%) FOR NEW ENGLAND 2006 SUMMER LOAD (CASE 2 IN EXAMPLE 2)

	MAPE(M1)	MAPE(M2)	MAPE(M1) - MAPE(M2)
June	1.86	1.83	0.03
July	3.27	3.24	0.03
Aug	2.69	2.67	0.02

M1: With dew point; M2: With humidex

TABLE III
MAPE (%) FOR NEW ENGLAND 2006 LOAD (CASE 3 IN EXAMPLE 2)

	MAPE(M1)	MAPE(M2)	MAPE(M1) - MAPE(M2)
Jan	1.75	1.65	0.11
Feb	1.56	1.48	0.08
March	1.65	1.52	0.13
April	1.34	1.32	0.02
May	1.78	1.66	0.12
June	1.95	1.83	0.12
July	3.35	3.24	0.11
Aug	2.74	2.67	0.07
Sept	1.56	1.52	0.04
Oct	1.49	1.43	0.06
Nov	1.45	1.44	0.01
Dec	1.93	1.81	0.12

M1: Without the v-shaping; M2: With the v-shaping

The effects of the preprocessing on neural network training speed are also examined by comparing the training iteration numbers with and without the preprocessing under the same termination criteria (MAPE is below 1.25% for the low frequency network, and MAE is below 30 MW for the high frequency network considering the rationale presented in footnote 6). Without the preprocessing, training iteration numbers for the low and high frequency networks are, respectively, 1540 and 3127. With the preprocessing, these numbers drop to 700 and 1230, showing that the preprocessing speeds up networks' learning. Iterations numbers for high frequency networks are generally higher than those for low frequency networks due to the "noisy" nature of high frequency load.

Case 4: SIWNN and the standard NN method are both tested by using actual weather and predicted weather. For each method, historical actual data are used in training, and two predictions are then generated based on actual weather and predicted weather. Training and prediction periods are same as those in Case 3. MAPEs presented in Tables IV and V show that SIWNN is less sensitive to weather forecasting errors as compared with the standard NN method.

Example 3: This example predicts New England 2007 load, and includes four cases. Case 1 demonstrates the value of wavelet decomposition by comparing SIWNN with a single

TABLE IV
MAPE (%) FOR NEW ENGLAND 2006 LOAD USING SIWNN (CASE 4 IN EXAMPLE 2)

	MAPE(M1)	MAPE(M2)	MAPE(M2) - MAPE(M1)
Jan	1.60	1.65	0.05
Feb	1.43	1.48	0.05
March	1.47	1.52	0.05
April	1.26	1.32	0.06
May	1.61	1.66	0.05
June	1.79	1.83	0.04
July	2.70	3.24	0.54
Aug	2.62	2.67	0.05
Sept	1.48	1.52	0.04
Oct	1.38	1.43	0.05
Nov	1.39	1.44	0.05
Dec	1.75	1.81	0.06

M1: SIWNN with actual weather; M2: SIWNN with predicted weather

TABLE V
MAPE (%) FOR NEW ENGLAND 2006 LOAD USING THE STANDARD NN METHOD (CASE 4 IN EXAMPLE 2)

	MAPE(M3)	MAPE(M4)	MAPE(M4)-MAPE(M3)
Jan	2.01	2.12	0.11
Feb	1.5	1.66	0.16
March	1.55	1.66	0.11
April	1.51	1.59	0.08
May	1.69	1.76	0.07
June	2.3	2.42	0.12
July	3.72	3.86	0.14
Aug	3.33	3.48	0.15
Sept	1.6	1.66	0.06
Oct	1.52	1.58	0.06
Nov	1.73	1.83	0.10
Dec	1.91	2.03	0.12

M3: The standard NN with actual weather; M4: The standard NN with predicted weather

neural network without wavelet decomposition. For the fair comparison, the similar day load and supplemental load are used in both methods. Case 2 shows the effectiveness of our idea to predict high frequency load. Case 3 shows the benefits of using similar day's load as compared to the standard practice of using the most recently available load. Case 4 then demonstrates the importance of using the supplemental load. In all the cases, training period is from 2003 to 2006, and prediction period is 2007.

Case 1: SIWNN and a single network without wavelet decomposition are compared. MAPEs for both methods are presented in Table VI. It can be seen that wavelet decomposition improves prediction accuracy. This is consistent with the conclusion made in Example 1.

Case 2: Prediction of high frequency load is examined with accuracy measured by MAE (following the rationale presented in footnote 6). The results presented in Table VII are good considering the noisy nature of high frequency load. The actual and predicted high frequency load are plotted in Fig. 14 for two winter peak load days (February 7 and 8, 2007), and in Fig. 15 for two summer peak load days (July 10 and 11, 2007). It can be seen that predictions follow the trend of the actual well.

TABLE VI
VALUE OF WAVELET DECOMPOSITION MAPE (%) FOR 2007 NEW ENGLAND LOAD

	Single NN	SIWNN	MAPE (Single NN) – MAPE (SIWNN)
Jan	1.79	1.59	0.2
Feb	1.44	1.3	0.14
March	1.41	1.37	0.04
April	1.45	1.27	0.18
May	1.35	1.33	0.02
June	2.53	2.22	0.31
July	2.34	2.07	0.27
Aug	2.27	1.98	0.29
Sept	2.16	2.12	0.04
Oct	1.31	1.24	0.07
Nov	1.65	1.45	0.2
Dec	1.8	1.68	0.12

TABLE VII
ACCURACY (MAES) OF HIGH FREQUENCY LOAD (NEW ENGLAND 2007 DATA)

	MAE (MW)		MAE (MW)
Jan	44.81	July	48.14
Feb	48.19	Aug	52.99
March	55.99	Sept	60.56
April	55.72	Oct	47.35
May	48.25	Nov	49.47
June	55.38	Dec	50.52

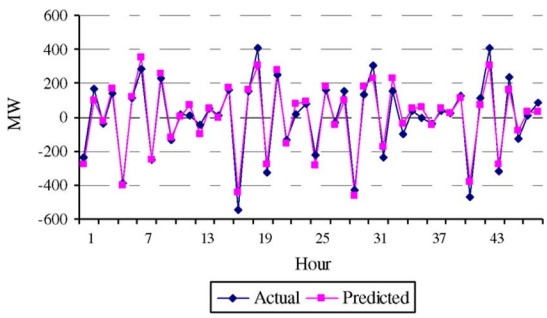


Fig. 14. Actual and predicted high frequency load (winter peak load days: February 7 and 8, 2007).

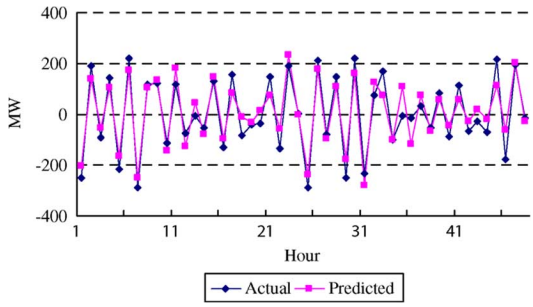
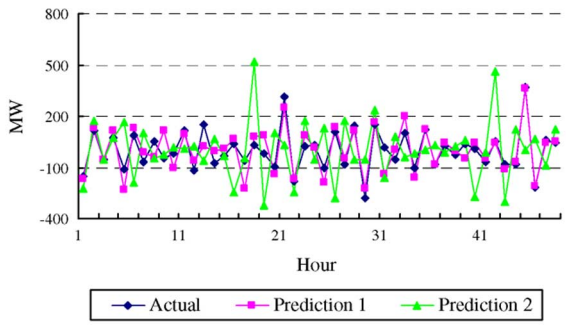


Fig. 15. Actual and predicted high frequency load (summer peak load days: July 10 and 11, 2007).

To examine the value of precipitation, the high frequency network is tested with and without precipitation. The actual and predicted high frequency load for two rainy days (April 15 and 16, 2007) are plotted in Fig. 16. The prediction obtained by using precipitation is clearly better than the other, demonstrating the value of precipitation.



Prediction 1: Prediction with using precipitation; Prediction 2: Prediction without using precipitation

Fig. 16. Actual and predicted high frequency load (rainy days: April 15 and 16, 2007).

Prediction 1: Prediction with using precipitation; Prediction 2: Prediction without using precipitation

TABLE VIII
VALUE OF USING THE SIMILAR DAY'S LOAD
MAPE (%) FOR NEW ENGLAND 2007 LOAD

	M1	M2	MAPE(M1) - MAPE(M2)
Jan	1.93	1.59	0.34
Feb	1.36	1.3	0.06
March	1.43	1.37	0.06
April	1.42	1.27	0.15
May	1.36	1.33	0.03
June	2.75	2.22	0.53
July	2.6	2.07	0.53
Aug	2.77	1.98	0.79
Sept	2.18	2.12	0.06
Oct	1.27	1.24	0.03
Nov	1.78	1.45	0.33
Dec	1.82	1.68	0.14

M1: Using load of day D-2; M2: Using similar day's load

TABLE IX
VALUE OF THE SUPPLEMENTAL LOAD MAPE (%)
FOR NEW ENGLAND 2007 LOAD

	M1	M2	MAPE(M1)-MAPE(M2)
Jan	4.79	1.59	3.20
Feb	4.16	1.30	2.86
March	4.41	1.37	3.04
April	3.79	1.27	2.52
May	4.37	1.33	3.04
June	5.76	2.22	3.54
July	5.89	2.07	3.82
Aug	5.12	1.98	3.14
Sept	4.57	2.12	2.45
Oct	4.68	1.24	3.44
Nov	3.18	1.45	1.73
Dec	4.92	1.68	3.24

M1: Without using supplemental load; M2: Using supplemental load

Case 3: This case compares using the similar day's load versus the standard practice of using the most recent load (load of day D-2). MAPEs presented in Table VIII show that using the similar day's load improves prediction accuracy.

Case 4: SIWNN is used with and without the supplemental load to examine its value. MAPEs presented in Table IX show that this load is important and significantly improves prediction quality.

IV. CONCLUSION

This paper presents a generic framework that combines similar day selection, wavelet decomposition, and neural networks to forecast tomorrow's load. The key idea is to use similar day's load supplemented by today's predicted load at hour 24 as input load, and use a synergistic combination of wavelet decomposition and neural networks to capture key features of load at low and high frequencies. Testing results show that this method provides accurate predictions. This method has been extended for holiday load forecasting, and very short-term load forecasting with very good results.

REFERENCES

[1] W. Charytoniuk, M. S. Chen, and P. V. Olinda, "Nonparametric regression based short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 13, no. 3, pp. 725–730, Aug. 1998.

[2] K. B. Song, Y. S. Baek, D. H. Hong, and G. Jang, "Short-term load forecasting for the holidays using fuzzy linear regression method," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 96–101, Feb. 2005.

[3] ISO New England, System Operation Procedures, Apr. 4, 2007. [Online]. Available: http://www.iso-ne.com/rules_proceds/operating/sysop/out_sched/sop_outsch_0040_0010.pdf.

[4] E. A. Feinberg and D. Genethliou, *Applied Mathematics for Restructured Electric Power Systems: Optimization, Control, and Computational Intelligence: Load Forecasting*. New York: Springer, 2005, pp. 269–285.

[5] K. Y. Lee, Y. T. Cha, and J. H. Park, "Short-term load forecasting using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 7, no. 1, pp. 124–132, Feb. 1992.

[6] C. N. Lu, H. T. Wu, and S. Vemuri, "Neural network based short term load forecasting," *IEEE Trans. Power Syst.*, vol. 8, no. 1, pp. 336–342, Feb. 1993.

[7] A. G. Bakirtzis, V. Petridis, S. J. Kiartzis, M. C. Alexiadis, and A. H. Maissis, "A neural network short term load forecasting model for the Greekpower system," *IEEE Trans. Power Syst.*, vol. 11, no. 2, pp. 858–863, May 1996.

[8] K. L. Ho, Y. Y. Hsu, and C. C. Yang, "Short term load forecasting using a multilayer neural network with an adaptive learning algorithm," *IEEE Trans. Power Syst.*, vol. 7, no. 1, pp. 141–149, Feb. 1992.

[9] R. R. Agnaldo and P. A. Alexandre, "Feature extraction via multiresolution analysis for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 189–198, Feb. 2005.

[10] Y. Zhao, P. B. Luh, C. Bomgardner, and G. H. Beerel, "Short-term load forecasting: Multi-level wavelet neural networks with holiday corrections," in *Proc. IEEE Power and Energy Soc. 2009 General Meeting*, Calgary, AB, Canada, Jul. 2009.

[11] C. Guan, P. B. Luh, M. A. Coolbeth, Y. Zhao, L. D. Michel, Y. Chen, C. J. Manville, P. B. Friedland, and S. J. Rourke, "Very short-term load forecasting: Multilevel wavelet neural networks with data pre-filtering," in *Proc. IEEE Power and Energy Soc. 2009 General Meeting*, Calgary, AB, Canada, Jul. 2009.

[12] R. Osczevski and B. Maurice, "The new wind chill equivalent temperature chart," *Bull. Amer. Meteorol. Soc.*, vol. 10, pp. 453–1458, 2005.

[13] J. M. Masterton and F. A. Richardson, Humidex, a Method of Quantifying Human Discomfort Due to Excessive Heat and Humidity, CLI 1-79, Environment Canada, Atmospheric Environment Service, 1979.

[14] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, 2nd ed. Wellesley, MA: Wellesley-Cambridge Press, 1997.

[15] B. Porat, *A Course in Digital Signal Processing*. New York: Wiley, 1997.

[16] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.

[17] L. Zhang, P. B. Luh, and K. Kasiviswanathan, "Energy clearing price prediction and confidence interval estimation with cascaded neural networks," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 99–104, Feb. 2003.

[18] N. R. Draper and H. Smith, *Applied Regression Analysis*, 3rd ed. New York: Wiley, 1998.

Ying Chen received the B.Eng. degree in electrical engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1999 and the Ph.D. degree in electrical and computer engineering from the University of Connecticut, Storrs, in 2008.

She is currently with Edison Mission Marketing and Trading, Boston, MA, as a Risk Analyst.

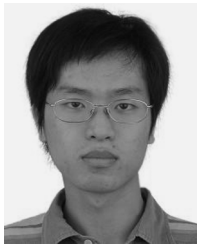
Peter B. Luh (F'95) received the Ph.D. degree in applied mathematics from Harvard University, Cambridge, MA, in 1980.

Since then, he has been with the University of Connecticut, Storrs. Currently, he is the SNET Professor of Communications and Information Technologies and the Head of the Electrical and Computer Engineering Department. He is interested in planning, scheduling, and coordination of design, manufacturing, and supply chain activities; configuration and operation of elevators and HVAC systems for normal and emergency conditions; and schedule, auction, portfolio optimization, and load/price forecasting for power systems.

Dr. Luh is Vice President of Publication Activities for the IEEE Robotics and Automation Society, the founding Editor-in-Chief of the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING (2003–2008), and was the Editor-in-Chief of the IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION (1999–2003).

Che Guan received the B.Eng. degree in electronics and information engineering from Changchun University of Science and Technology, Changchun, China, in 2004, and the M.Eng. degree in mechanical and electrical engineering from the Chinese Academy of Sciences, Beijing, China, in 2007. He is currently pursuing the Ph.D. degree in the Electrical and Computer Engineering Department at the University of Connecticut, Storrs.

He is interested in load forecasting for power systems.



Yige Zhao was born in Liaoning, China, in September 1984. He received the B.Eng. degree in communication and electrical engineering from the Beijing Institute of Technology, Beijing, China, in 2007. Currently, he is pursuing the M.S. degree in electrical engineering at the University of Connecticut, Storrs.

Laurent D. Michel received the Ph.D. degree in computer science from Brown University, Providence, RI, in 1999.

He currently is an Associate Professor in the Computer Science and Engineering Department at the University of Connecticut, Storrs. He specializes is combinatorial optimization with a particular emphasis on constraint programming. He has coauthored two monographs and more than 60 papers.

Dr. Michel sits on the Editorial Board of *Constraints, Mathematical Programming Computation and Constraint Letters*.

Matthew A. Coolbeth received the B.Eng. degree in computer science from the University of Connecticut, Storrs, in 2007, where he is currently pursuing the M.Eng. degree in the Department of Computer Science and Engineering.

Peter B. Friedland received the B.S.E.E. from the University of Connecticut (UCONN), Storrs, in 1986. He also attended the University of Washington and UCONN as an electrical engineering graduate student.

He has been working in the electric utility industry for the past 22 years, including work with an EMS Vendor, Transmission Operator, and ISO/RTO. For the past eight years, he has been with ISO New England, Holyoke, MA, in varying management capacities including Standard Market Design (SMD) Project Manager, IT-EMS Manager, and Operations Manager.

Mr. Friedland is a member of the UCONN Electrical Engineering Industrial Advisory Board.

Stephen J. Rourke (SM'94) received the B.S. degree in electrical engineering from Worcester Polytechnic Institute, Worcester, MA, and the M.B.A. degree from Western New England College, Springfield, MA.

He is Vice President, System Planning for ISO New England, Holyoke, MA, having previously served as the company's Director, Reliability & Operations Services. A former manager of the Rhode Island-Eastern Massachusetts-Vermont Energy Control center (REMVEC) in Westborough, MA, and former Manager of Marketing Operations for Northeast Utilities/Select Energy, Inc., in Berlin, CT, he has 30 years of experience in operations and planning of the New England bulk power system. He is responsible for overseeing development of the annual Regional System Plan, analysis and approval of new transmission and generation projects, implementing the FERC approved generator interconnection process, developing ISO findings for Transmission Cost Allocation, and supporting the capacity markets in New England.