A Note on the Practical Efficiency of Using Wavelet Transform for Short-term Load Forecasting in Smart Grids

M. Karkhaneh[1] and S. Ozgoli[2*]

[1] Electrical & Computer Engineering department, PhD candidate, Tarbiat Modares University, Tehran, Iran (e-mail: m.karkhaneh@modares.ac.ir).

[2] Electrical & Computer Engineering department, Faculty of Electrical & Computer Engineering, Tarbiat Modares University, Tehran, Iran (e-mail: ozgoli@modares.ac.ir).

*Corresponding Author

Received 25 Apr. 2022 Received in revised form 12 July 2022 Accepted 18 Aug. 2022

Type of Article: Research paper

***Abstract*— In the smart grid era, load forecasting is the building block of a secure, reliable, and economic power system. Therefore, many researchers have spent a lot of time trying different methods to improve load forecasting accuracy. In recent years, one of the rather frequently used methods is the decomposition of load series into high and low-frequency components using wavelet transform, which reportedly has shown impressive results in some articles. In this paper, through several simulations, it's demonstrated that despite some of the benefits of the wavelet transform, it can produce unrealistic results due to the border distortion problem. In fact, our work investigates the practical efficiency of wavelet transform in the load forecasting task from the viewpoint of a system operator who is forecasting the next day's load profile every day. To this end, Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) models are used with wavelet transform to conduct experiments on New York City electric load dataset.**

***Keywords*: Artificial Neural Network, Load Forecasting, Multiple Linear Regression, Wavelet Transform.**

I. Introduction

In the smart grid era, short-term load forecasting (STLF), which provides electric load forecasts up to two weeks ahead, is a vital part of a secure, reliable, and economic power system operation. Accurate production coordination of electrical generators in the smart grids greatly depends on STLF accuracy. Furthermore, STLF is necessary for the utilities and retailers to purchase the correct amount of required energy in electricity markets, which results in a lower energy cost for the utility or retailer.

With the development of smart grids, active participation of energy consumers, smart charging of electric vehicles, renewable energy sources and other smart grid elements are making load forecasting a more difficult task. For this reason, despite the extensive literature on load forecasting, this topic is still a subject of active research.

Load forecasting techniques can be classified into two major groups: 1- statistical techniques: e.g., Multiple Linear Regression (MLR), Auto Regressive Moving Average (ARMA), exponential smoothing models, and 2- artificial intelligence techniques: e.g., Artificial Neural Networks (ANN), Support Vector Machines (SVM) [2], gradient boosting machines [3] and fuzzy systems [4].

From the statistical techniques, MLR [5] and ARMA [6] models have received the most attention [7]. In MLR, the electric load is explained by combining two or more independent variables, such as temperature and calendar variables, which is a great feature where there is a tangible relationship between temperature and electricity consumption. ARMA models are based on the electric load only; these models do not include other factors like temperature in the model, so this technique is suitable for regions where the electric load is not affected by other factors like weather conditions.

From artificial intelligence techniques, ANN [8] has received the most attention in the load forecasting area [9], which is partially due to the fact that ANN doesn't require much prior knowledge of the relationship between load and affecting variables; because ANN is a black box technique that can infer underlying relationships between input and output variables.

Besides the mentioned techniques, several useful methods such as similar day methods [10], [11], variable selection methods [12], and load decomposition methods [13], [14] can be applied to the techniques to increase the load forecasting accuracy. In the load decomposition methods, the goal is to decompose the load series into several components, for example, using wavelet transform [15], which results in extracting extra features of the load. These extra features provide additional useful information for the forecasting system and may increase overall accuracy. For more than two decades wavelet transform method has been used with different load forecasting techniques to decrease the forecasting error. Among these techniques, artificial neural networks were much appreciated by the researchers; for example, in [16], [17], and [18], authors utilized Multi-Layer Perceptron, Echo State Networks, and Bayesian Neural Networks, respectively, to perform load forecasting with wavelet transform method. In [19], the authors proposed Kalman filter models based on wavelet transform to improve short-term load forecasting at the system level. A combination of wavelet transform and gray model for the purpose of load forecasting is presented in [20]. Recently, for predicting the electrical load of New England Independent System Operator (ISO-NE), a combination of MLR with the wavelet transform method is proposed in [21].

Our experimental results show that despite some of the benefits of the wavelet transform in electrical load forecasting, one cannot expect impressive results. Therefore, in this paper, the purpose is to show why the impressive results of some articles may not be reproduced in practice, and where the key problem is.

The rest of the paper is organized as follows: Section II is devoted to a short review of the papers which have used wavelet transform as a part of their forecasting procedure. Section III briefly explains wavelet transform, MLR, and ANN. The problem statement is reported in section IV. Simulation results and further discussions are reported in section V. Finally, Section VI concludes the paper.

II. LITERATURE REVIEW

For more than four decades, researchers have been looking to find the best techniques and methods in electric load forecasting; one of the solutions found to reduce the load prediction error is the use of wavelet transform in which the electric load sequence is decomposed into a low and several high-frequency components to extract more features out of it.

In [22], Zhang & Dong used wavelet transform and neural network (MLP structure) to predict Queensland electricity demand; at first, they decomposed the load into different scales, then each scale was predicted by a separate NN and, at the final stage a NN used these predicted values to obtain the final result. In the end, they concluded that a plain MPL offers a reasonable level of performance compared to wavelet-based methods.

Reis & Da Silva [16] discussed two different strategies for embedding wavelet transform into NN-based load forecasting. In the first strategy (proposed one), decomposition of the differenced load in addition to the actual load is given as inputs to a NN to produce the final load forecast, while the second strategy uses separate NNs to forecast different load components; hence a reconstruction phase is needed in the second strategy. The authors concluded that the first strategy shows a balanced performance in both one-step and one-day ahead load forecasting; it’s worth mentioning that despite the ample complexity of the two strategies, both hadn’t any improvement in one-step ahead load forecasting compared to simple non-wavelet NN based methods. In this paper, for the first time, the problem of border distortion has been tackled. Also, they used Daubechies wavelets of order 2 (Db2) with three decomposition levels for signal decomposition.

Amjady & Keynia [23] proposed an STLF method in which, after using wavelet transform (Db4 with three decomposition levels), decomposed components were predicted by a combination of NN and evolutionary algorithm (EA) to produce the final predicted load after reconstruction i.e., using inverse wavelet transform. They compared their results with the proposed methods in [16] and reported a better forecasting accuracy.

Deihimi et al. [17] utilized wavelet transform with 5 levels of decomposition, then individual echo state networks were implemented to predict decomposed components of the load; to produce the final load forecast a separate echo state network was used as reconstruction engine. For day-ahead load prediction, 24 of the mentioned forecaster have been used to produce a 24 hour prediction (one for each hour); a comparison of the proposed method with [16] and [23] showed better forecasting accuracy.

Chen et al. [24] presented a similar day based method to forecast tomorrow’s electric load. The idea was to find similar days in the load history based on the weekday index and tomorrow's weather. Then Db4 wavelet was used to decompose the similar day's load and the predicted load of tomorrow (at hour 24) into a high and a low-frequency component; then, each of the two high and low-frequency components was predicted using a separate neural network and added together to produce the final predicted load of tomorrow.

Pandey et al. [25] proposed a wavelet neural network (WNN) in which wavelet transform was used as a smoothing method. In the smoothing stage, they first decomposed the load and temperature into high and low-frequency components (using Db2 with three decomposition levels). Then, the smoothed data were created by removing high-frequency components. In the forecasting stage, the smoothed data were fed to an RBFNN to produce the final load forecast. Besides, this paper compares time series, RBFNN, and fuzzy inference neural network models with their wavelet-based counterparts, which shows the superiority of wavelet-based models.

Liu et al. [26] used multiwavelet transform to extract more information of the electric load series. Then this information goes to three different neural networks (BPNN, RBFNN, and WNN) to produce three other inputs for the final neural network (a three-layer feed-forward NN), whose output is the forecasted load. It's been alleged that the forecasting error of the proposed method is 0.3504 in terms of Mean Absolute Percentage Error (MAPE), which is far better than a BPNN with 1.5773 MAPE for daily load forecasting of some districts in the Sichuan Grid of China.

Bahrami et al. [20] combined the gray model (optimized by PSO) and wavelet transform for load forecasting. In the first step, they used wavelet transform (Haar with scale 10) to eliminate high-frequency components of the electric load. In the second step, the PSO algorithm was used to determine the parameters of the gray model. In the final step, the electric load of tomorrow was predicted using the gray model and filtered load.

Ghofrani et al. [18] approach in load forecasting was based on wavelet decomposition and Bayesian neural network (BNN). For this purpose, they first classified the input data into a bunch of sub-series based on a new input selection method; then, these subseries were ranked based on correlation analysis and L2-norm calculation. The sub-series with the least L2-norm with respect to the desired correlation coefficient were decomposed using wavelet transform with four levels of decomposition to provide proper inputs for the first BNN; the other sub-series with L2-norm close to the least L2-norm were selected as the inputs of the separated BNNs. In the end, a weighted sum of the BNNs outputs was used to provide the final forecast. They argued that their approach outperforms the proposed method of [24] and ANN by 75.7% and 79.5%, respectively.

Alipour et al. [27] proposed a structure based on wavelet transform (Dmey wavelet with 10-level decomposition) for feature extraction and deep neural network as a model for electric net-load forecasting. In their proposed deep neural network structure they used some sparse autoencoders and a cascade neural network. A comparison made by other forecasting techniques shows that as they said, their method has extraordinary accuracy. In terms of MAPE, simple NN and SVR have 0.365 and 0.393 percent error while their proposed method has only 0.039 percent error for the DE region of Germany.

By reviewing the above articles, one can find that some of them stated that they could dramatically reduce the load forecasting error using wavelet transform. In the following sections, the validity of their results is studied.

III. Theoretical Background

A. Wavelet Transform

Readers are familiar with the Fourier Transform (FT). The main problem of FT is that it doesn't give any information on the time position of a specific frequency; in order to solve this problem, wavelet transform has been developed to provide a time-frequency (more precisely, time-scale) representation of a signal. Wavelet analysis enables us to discover aspects of signals that other signal analysis tools miss (e.g., trends, breakdown points, discontinuities in higher derivatives, and self-similarity). There are two kinds of wavelet transforms: the continuous wavelet transform and the discrete one. The continuous wavelet transform of a signal x(t) is defined as follows [23]:

$$\mathrm{w}(a,b) = \int_{-\infty}^{+\infty} x(t)\,\psi_{ab}(t)dt \tag{1}$$

In (1), $w(a,b)$ is the wavelet transform, and $\psi_{ab}(t)$ is the mother wavelet which is defined as:

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{2}$$

Where a and b are the scale and the translation parameters.

The continuous wavelet transform is calculated by continuously scaling and translating $\psi_{ab}(t)$; As a consequence, in addition to the cumbersome computations, a lot of redundant information is generated. Therefore, discrete wavelet transform (DWT) is invented using certain scales and translations to reduce computation complexity while keeping desired performance. DWT is defined as:

$$\mathrm{W}(m,n) = \frac{1}{\sqrt{2^m}}\sum_{t=0}^{T-1} x(t)\,\psi\left(\frac{t-n.2^m}{2^m}\right) \tag{3}$$

Where T and t are the length and index of the signal $x(t)$, and scaling and translation parameters are functions of the integer variables m and n ($a = 2^m, b = n.2^m$). In practice, the discrete wavelet transform of a signal is computed by Mallat's pyramidal algorithm [28], in which a signal is decomposed into low-frequency (approximation) and high-frequency (details) components using consecutive low pass and high pass filters. Mallat's algorithm is composed of a decomposition stage, and a reconstruction stage. In Fig. 1, the structure of a multi-resolution analysis system via Mallat's pyramidal algorithm, which computes two-level DWT, is shown [29]. In the decomposition stage, the original signal (S) is convolved with high pass filter (H) and low pass filter (L), and then these filtered data are down-sampled by removing odd numbered points to produce the first level detail (cD1) and approximation (cA1) coefficients, respectively. By performing the same procedure on the cA1, the second level coefficients i.e.

cD2 and cA2 can be obtained. In the reconstruction stage, cD1 is up-sampled by padding zeros between cD1 elements to recover the original data length, and then these up-sampled data are convolved with the corresponding reconstruction filter (H') to produce high frequency components (D1) of the input signal. To obtain the second level decomposition (i.e. D2 and A2) from cD2 and cA2, up-sampling and convolving with the proper reconstruction filters should be performed twice. Note that original signal can be reconstructed by summation of the approximation component of the last level and all the detail components e.g. in Fig. 1, $\mathrm{S} = \mathrm{A2} + \mathrm{D2} + \mathrm{D1}$.

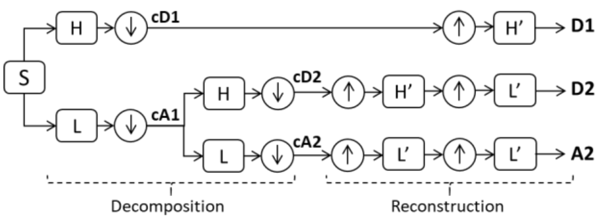

Figure 1. Multiple-level decomposition of signal S (A and D denote approximation and detailed components of S, respectively)

B. *Multiple Linear Regression*

Multiple linear regression is a statistical technique in which multiple explanatory variables are used to describe a response variable. This technique aims to model the linear relationship between the explanatory variables and a response variable. Despite simple computations, MLR has proven to be a capable technique in predicting future load [30]; even in recently published papers in the load forecasting area, satisfactory results have been reported [31], [32].

The MLR model is given by:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{i,p} + e_i \qquad (4)$$

Where Y_i denotes a dependent variable, $X_{i1}, \ldots, X_{i,p}$ denote explanatory variables, $\beta_0, \ldots, \beta_p$ are model parameters in which β_0 is the intercept term, and $\beta_1, \ldots, \beta_p$ are slope coefficients for each explanatory variable, e_i is model's error term (residuals), and i is the number of observations. Model parameters can be estimated using least-squares estimation techniques.

Note that in MLR, the following assumption should be held: 1- Dependent variable is a linear combination of the explanatory variables and the model parameters; in fact, this linearity is in terms of the parameters rather than explanatory variables, so any form of explanatory variables can be used. 2- The explanatory variables should not be highly correlated with each other. 3- Error terms should be normally distributed with zero mean and constant variance.

C. *Artificial Neural Network*

Artificial Neural Networks are computing systems that work similarly to the human brain. Just like neurons in the human brain that are responsible for processing the received information, in ANNs, units called neurons perform a function similar to that of the brain neurons. AN ANN consists of three groups of layers: 1-input layer, 2- hidden layers, and 3- output layer. For better illustration, Fig. 2 shows a simple feed-forward neural network with three layers in which received information from the input layer is first linearly combined by a weighted sum of all the inputs, then these values are fed to the neurons of the hidden layer where they are passed through a generally nonlinear activation function to produce the outputs of hidden layer; at last, these outputs of the hidden layer neurons are linearly combined together and are fed to the neurons of the output layer to produce the final result. For a hyperbolic tangent (tanh) Activation Function ($h(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$), the mathematical formulation of Fig. 2 can be written as follows:

$$u_j = \alpha_{0j} + \sum_{i=1}^{I} \alpha_{ij} x_i \qquad (5)$$

$$w_j = h(u_j)$$

$$v_k = \beta_{0k} + \sum_{j=1}^{J} \beta_{jk} w_j$$

$$y_k = h(v_k)$$

Where α_{0j} and β_{0k} are the bias terms of the hidden and output neurons, which are not shown in Fig. 2 for simplicity.

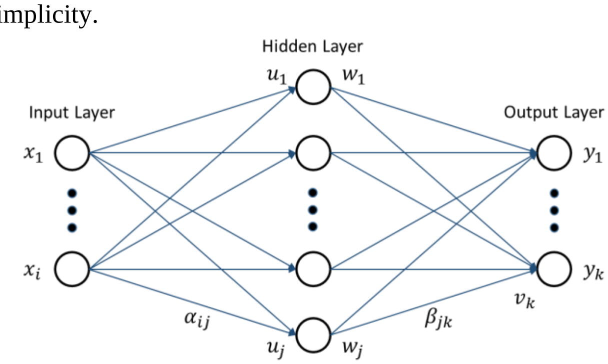

Figure 2. A simple feed-forward neural network with three layers

ANNs have been widely used in load forecasting since the '90s [33]; even a simple structure like feed-forward neural network is still being used in practice [34], [35]. The reason for the extensive use of the ANN technique, in addition to the simplicity of implementation, lies in the fact that it does not require much prior knowledge in the field under study.

IV. PROBLEM STATEMENT

A. Border Effect

As stated before, the wavelet transform is achieved by convolution of the input signal with low pass and high pass filters. As usual, when a convolution is performed on finite-length signals, border distortions appear. In other words, border distortion arises due to the fact that to calculate the wavelet coefficients at the beginning and the end of the signal, part of the filter goes beyond the extent of the signal; this makes transformed values close to the border of the signal be tainted by the unavailable data of the signal edge. This undesirable effect at the borders of the signal is called "border effect".

Normally, the electric load obtained at the day before the forecasting day has the highest correlation with the load of the forecasting day; therefore in the STLF, one of the main inputs is the obtained load of the day before the forecasting day. Unfortunately, since this obtained load is located at the border of load data, the border effect distorts transformed values of this important input. Hence, despite the many advantages of the wavelet transform, the problem of border distortion dramatically reduces the effectiveness of this method for STLF.

Some signal extension methods (padding) are proposed in the literature to reduce the border distortion problem [36]. The most widely used signal extension methods are as follows: 1) Zero-padding, in which additional zeros are added outside the signal boundaries; 2) symmetric padding, which is done by symmetric boundary value replication; 3) smooth padding, which corresponds to padding by use of a linear extension of the first and the last two values of the signal; 4) periodic padding, which is done by periodic extension of the signal. However, as shown in the next section, these methods are not suitable for short-term load forecasting and significantly increase the forecasting error.

Considering the mentioned problem, why have some brilliant results been reported in the field of STLF using wavelet transform? The results of the proposed simulations and the author's practical experiments in this field show that these obtained results are due to a missing point at the implementation phase of the wavelet transform, where the decomposed data are provided for the forecasting model (illustrated in Fig. 3); More precisely, authors suspect that some researchers have applied wavelet transform to their entire electric load data series (all the test data) at once to reduce the computational burden; then they have fed these decomposed data to the forecasting model. But since wavelet decomposition is achieved by wavelet filters, as stated before, some of the future data enter the model invisibly, which incorrectly masks the border distortion problem and leads to outstanding results in the simulations.

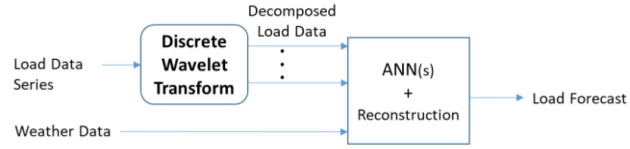

Figure 3. A typical structure of load forecasting using wavelet transform and ANN

The correct implementation is to apply wavelet transform to all the available data before the forecasting day. For example, if we want to test our load forecasting model by forecasting 365 days of a given year that its load data is available, and one of the inputs of our forecasting model is the decomposed load of the previous day; to predict the electric load of the 21st day of the year, the decomposed load of the 20th day should be provided by getting wavelet transform of the entire available data until the 20th day and extracting the decomposed load of the 20th day from this set (not getting wavelet transform from all the 360 available days in the test set and extracting the decomposed load of the 20th day from this set). Hence, predicting tomorrow's load profile in a test period means re-computing wavelet coefficients to the number of days that are going to be predicted in the test period. Of course, this has a high computational burden, and as shown in the next section, it dramatically affects the results due to a disturbing phenomenon (border effect). Unfortunately, there is no other choice because the realized load of the coming days is not available in real practice, and this is the case with the operators who deal with STLF in their company every day. In section V, some simulation results are provided to describe the problem more precisely.

B. Level of Decomposition

Another issue is that better results are observed in some articles that have used higher levels of wavelet decomposition. To investigate this issue, at first, the relation between the decomposition level and the number of required data as padding (padding length) should be determined. Based on the convolution theory, every convolution between a signal and a filter causes a distortion with the length of $fl-1$ in which fl is the filter length; since in Fig. 1, two convolutions must be performed (with H and H') to calculate the D1 component; therefore, the total distortion length would be $2(fl-1)$. Likewise, since four convolutions must be performed in the second level of decomposition, the total distortion length for A2 and D2 components would be $4(fl-1)$. Therefore, in the multi-level decomposition with wavelet, the padding length is recommended to be at least $2^{lvl}(fl-1)$ to reduce the border distortions [29], where lvl is the maximum decomposition level.

In regard to the above discussion, in the wrong implementation, as the number of wavelet decomposition levels increases, the wavelet filters will be placed on more load data in the future (instead of the padded data); as a result, more information about the future load imperceptibly enters the model which improperly further reduces the load forecasting error. But as shown in the next section, if one uses the correct method described in the previous section, that is, acts like an electric utility operator who does not have access to the future electricity consumption data, Since the filter is placed further on the padded data (not the real data in the future), the prediction error happens to get worse instead of improving.

V.SIMULATION RESULTS AND DISCUSSIONS

In this section, the impact of the wrong and correct implementation of wavelet transform and the impact of the wavelet decomposition level on the STLF accuracy have been analyzed. For this purpose, two of the most favorable techniques in the STLF area, i.e., MLR and ANN, have been used.

MLR model- We've used 24 of the following MLR model to forecast 1 to 24 hours ahead:

$$\mathrm{Y}_i = \beta_0 + \beta_1 Month + \beta_2 Day + \beta_3 Month * T + \beta_4 Month * T^3 + \beta_5 Month * T_d + \beta_6 L_d + \beta_7 L_D * Day + \beta_8 L_T * Day + \beta_9 Trend + \beta_{10} Holiday + \beta_{11} Holiday_d \quad (6)$$

In which $Month$ and Day are class variables that correspond to the month of the year and day of the week, T is the average temperature of the forecasting day, T_d is the average temperature of the day before the forecasting day, L_d is load profile of the day before forecasting day, L_D is yesterday's load at hour i, L_T is the load of the same day as the forecasting day in the previous week at hour i, $Holiday$ and $Holiday_d$ are class variables which show national holidays of the forecasting day and the day before forecasting day, respectively. Trend is a natural number that captures the increasing trend of the load by assigning a separate number to each day in historical data, i.e., '1' to the first day of 2015, '2' to the second day of 2015, and '1826' to the last day of 2019. Note that, the sign '$*$' denotes the interaction between explanatory variables.

ANN- In this case, 24 ANNs with feedforward structure have been used to forecast 1 to 24 hours ahead. ANN's inputs are: $Month, Day, T, T_d, L_d, L_D, L_T$, $Holiday, Holiday_d$ & $Trend$. Since ANN automatically captures the interaction between input variables, no interaction term is used in ANN-based models. It should be noted that the simulations are performed on the MATLAB R2020b, and Bayesian regularization backpropagation is used as the training algorithm.

Dataset- Our dataset is the hourly electric load [37] and the daily average temperature of New York City over the period of January 1, 2015, to December 31, 2019.

Decomposition method- In this paper, discrete wavelet transform with Daubechies wavelet of order four (Db4) is used to decompose the electric load series into high and some low-frequency components. Note that in all the wavelet-based techniques, decomposed load series are used as model inputs instead of the original load series.

Metrics- In all the simulations, the results are reported in terms of mean absolute percentage error (MAPE), which is defined as follows:

$$\mathrm{MAPE}(\%) = \frac{1}{N}\sum_{n=1}^{N}\frac{|L_A(n) - L_F(n)|}{L_A(n)} \times 100 \quad (7)$$

Where $L_A(n)$ and $L_F(n)$ denote the actual and forecasted load at the nth hour, and N is the total number of the forecasted hours.

A. Impact of signal extension method

To investigate the best signal extension method for STLF, commonly used signal extension methods (explained in section IV) in addition to the method proposed in [16] (i.e., Reis & Da Silva Method), which uses measured values at the beginning and forecasted values at the end of the load series, are compared in Table I using MLR technique. Furthermore, to better illustrate the impact of the border effect at the end of a signal, the approximation component of the transformed signal for different signal extension methods and different decomposition levels are shown in Fig. 4.

TABLE I
COMPARISON OF MAPE VALUES IN SHORT TERM LOAD FORECASTING USING DIFFERENT SIGNAL EXTENSION METHODS

Decomposition level	Symmetric padding	Smooth padding	Periodic padding	Reis & Da Silva Method
Level 1	2.24	1.80	7.53	1.746
Level 2	10.24	6.00	25.20	1.752
Level 3	11.50	26.85	14.70	2.07

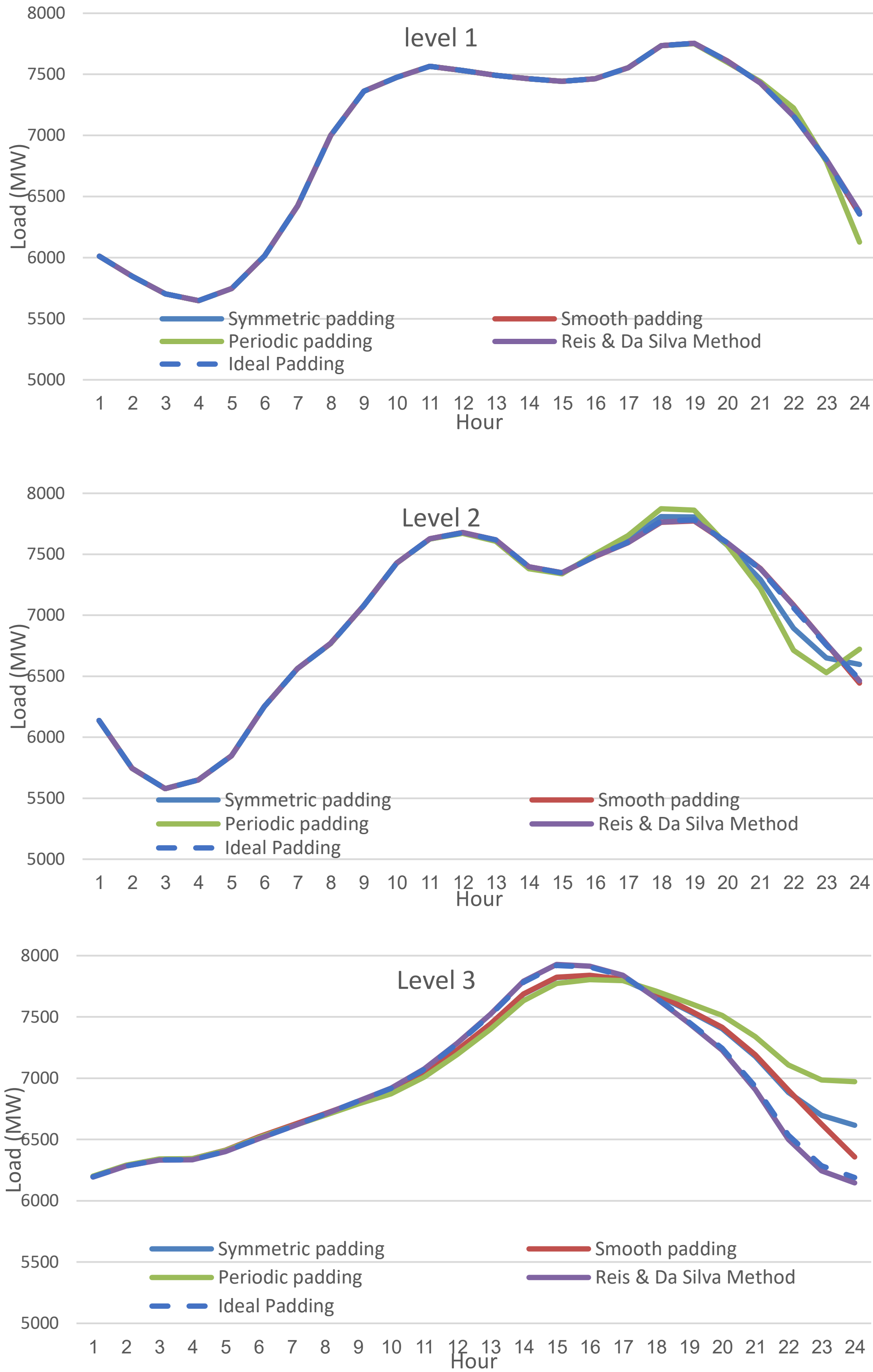

Figure 4. Approximation component of the transformed signal for different signal extension methods and different decomposition levels

From the results of Table I and Fig. 4, it can be concluded that the best available padding method for STLF, which produces more consistent and better results, is the method proposed by Reis & Da Silva; so from now on, this method will be used as signal extension method required for wavelet decomposition in the correct implementation scheme. Another important point from Fig. 4 is that as the level of decomposition increases, the boundary distortions expand, so one can expect degradation of forecasting accuracy due to giving more unreliable data to the model.

B. Wrong and correct implementation of wavelet transform

As mentioned in the previous section, implementing wavelet transform requires high delicacy. In this subsection, two cases have been analyzed. Case 1: in which the whole electric load dataset is decomposed in one shot, then these decomposed components are fed to the model part by part when it's needed (wrong implementation). Case 2: in this case, the electric load data until the day before the forecasting day is decomposed, which corresponds to applying 365 times wavelet transform for testing over a year (correct implementation). Comparative results of these two cases and the third case, in which no wavelet decomposition is used, are given in Table II. To better illustrate the forecasting accuracy of Case 1 to Case 3, the real and forecasted load of 13/2/2019 are shown in Fig. 5 and Fig. 6 using MLR and ANN techniques, respectively.

Note: Case 1 and Case 2 are performed using three decomposition levels; also, in Case 2, required values for padding are provided from the predicted load of Case 3.

TABLE II
COMPARISON OF MAPE VALUES OF WRONG AND CORRECT IMPLEMENTATION OF WAVELET TRANSFORM

	MAPE (%) value of MLR	MAPE (%) value of ANN
Case 1	1.04	1.11
Case 2	2.07	1.79
Case 3	1.74	1.75

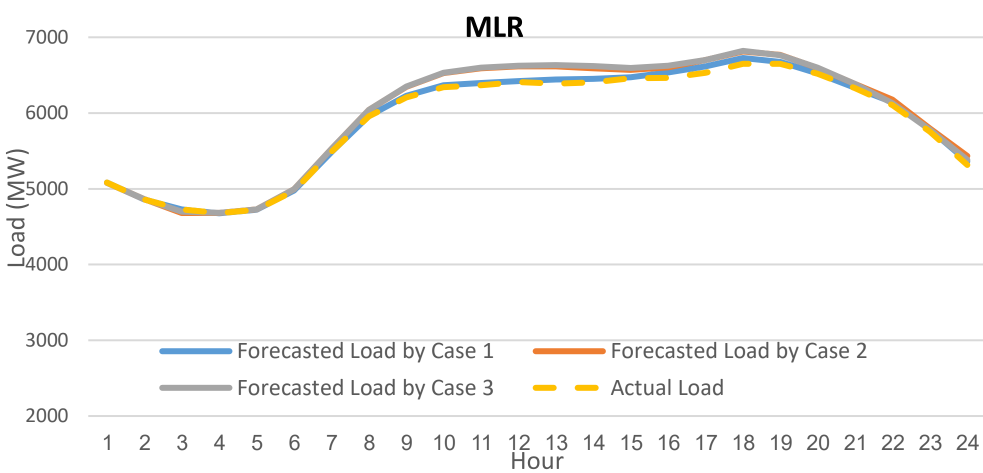

Figure 5. Actual and forecast load of New York City on 13/2/2019 using MLR technique

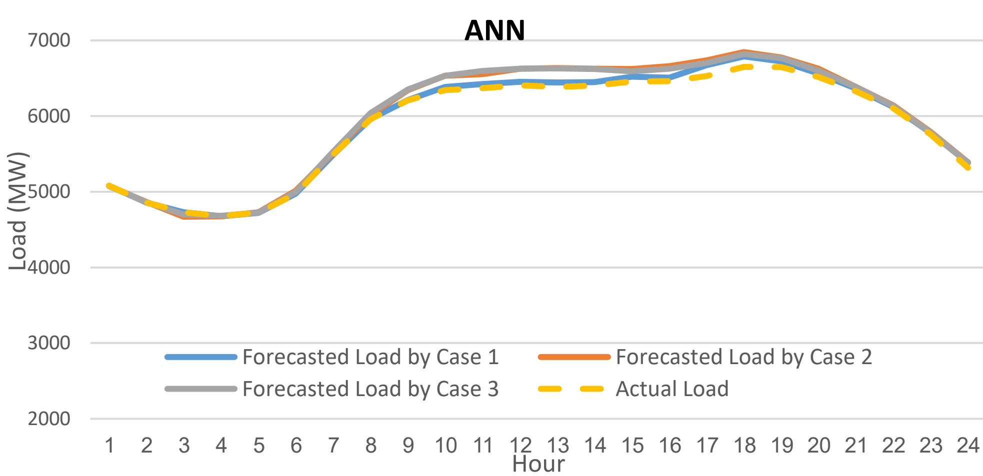

Figure 6. Actual and forecast load of New York City on 13/2/2015 using ANN technique

Remark 1: in Case 1, no future load is given explicitly to the model, but since the wavelet filter moves beyond the realized electric load series, imperceptibly, some form of the future load is exposed to the model; the authors think that's the reason of stellar results in some papers. Unfortunately, these stellar results can never happen in the real world. Moreover, from Table II, it can be seen that the correct implementation of wavelet transform has no benefit over Case 3, which is due to the border distortions caused by the padding method.

C. Impact of the decomposition level

The level of decomposition is a crucial choice in electric load forecasting using the wavelet transform. To better analyze the impact of the decomposition levels on STLF accuracy, MAPE values of correct and wrong implementation schemes are reported in Table III for one to three levels of decomposition. Since the MLR technique produced more consistent results compared to ANN, only MLR results are reported in Table III.

TABLE III
COMPARISON OF MAPE VALUES IN SHORT TERM LOAD FORECASTING USING DIFFERENT DECOMPOSITION LEVELS

Decomposition level	Correct implementation	Wrong implementation
Level 1	1.746	1.70
Level 2	1.752	1.48
Level 3	2.07	1.04

Remark 2: From the results of Table III, it can be concluded that by increasing the number of decomposition levels, the accuracy of load forecasting in the correct implementation scheme decreases; this low accuracy is due to the more severe border distortion. Meanwhile, by increasing the number of decomposition levels, accuracy of load forecasting in the wrong implementation scheme increases since more of the future load is exposed to the model in the simulations; as mentioned before, due to the border distortions, these excellent results of the wrong implementation can never happen in real practice. It is now obvious why higher levels of wavelet decomposition have led to far better results in some reports.

Remark 3: This article's simulations were performed on different datasets (some of them were confidential); in all cases, the same results could be inferred. But for the briefness, just the simulations on the New York City dataset are provided in this article. New York City electricity consumption data are publicly available in [37].

VI.CONCLUSIONS

In the smart grid environment, load forecasting plays a critical role in decision-making and increasing the stability of the power system. In this paper, the practical efficiency of using wavelet transform in the load forecasting field has been investigated. Through a case study on New York City electric load consumption data, it's shown that the reason for some reported extraordinary results is probably due to an inadvertent mistake that have occurred during the implementation phase of the wavelet transform; and such a result should never be expected in practice because the problem of border distortions severely affects the performance of wavelet transform in the STLF area. Furthermore, the performance of various signal extension methods in the literature was analyzed, and the best method for load forecasting was selected to perform the case study.

It should be noted that this paper doesn't deny the valuable benefits of wavelet transform for load forecasting. The authors of this paper believe that the wavelet transform is of significant importance, for example, in the pre-processing stage of load forecasting where outliers should be detected to prevent performance degradation. Furthermore, since STLF is highly affected by the border distortion problem, further researches are necessary to discover a more suitable padding method for STLF.

References

[1] D. Bunn and E. D. Farmer, "Comparative Models for Electrical Load Forecasting.," *Comp Model. Electr Load Forecast*, 1985.

[2] W. C. Hong and G. F. Fan, "Hybrid empirical mode decomposition with support vector regression model for short term load forecasting," *Energies*, vol. 12, no. 6, p. 1093, 2019.

[3] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, "A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting," *Energy*, vol. 214, p. 118874, 2021.

[4] G. Cerne, D. Dovzan, and I. Skrjanc, "Short-Term Load Forecasting by Separating Daily Profiles and Using a Single Fuzzy Model Across the Entire Domain," *IEEE Trans. Ind. Electron.*, vol. 65, no. 9, pp. 7406–7415, 2018.

[5] J. Kim, S. Cho, K. Ko, and R. R. Rao, "Short-term Electric Load Prediction Using Multiple Linear Regression Method," in *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, SmartGridComm 2018*, 2018, pp. 1–6.

[6] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, "A review on time series forecasting techniques for building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 74, pp. 902–924, 2017.

[7] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int. J. Forecast.*, vol. 32, no. 3, pp. 914–938, 2016.

[8] F. Y. Xu, X. Cun, M. Yan, H. Yuan, Y. Wang, and L. L. Lai, "Power Market Load Forecasting on

Neural Network With Beneficial Correlated Regularization," *IEEE Trans. Ind. Informatics*, vol. 14, no. 11, pp. 5050–5059, 2018.

[9] T. Hong, "Energy forecasting: Past, present, and future," *Foresight Int. J. Appl. Forecast.*, no. 32, pp. 43–48, 2014.

[10] W. Huang, D. Zhang, Z. Song, H. Wang, and H. Liu, "Short-term load forecasting based on similar day approach and intelligent algorithm using analytic hierarchy process," in *Proceedings of 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2019*, 2019, pp. 2507–2512.

[11] R. J. Park, K. Bin Song, and B. S. Kwon, "Short-term load forecasting algorithm using a similar day selection method based on reinforcement learning," *Energies*, vol. 13, no. 10, p. 2640, 2020.

[12] J. Xie and T. Hong, "Variable selection methods for probabilistic load forecasting: Empirical evidence from seven states of the United States," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6039–6046, 2018.

[13] J. Li *et al.*, "A Novel Hybrid Short-Term Load Forecasting Method of Smart Grid Using MLR and LSTM Neural Network," *IEEE Trans. Ind. Informatics*, vol. 17, no. 4, pp. 2443–2452, 2021.

[14] F. He, J. Zhou, Z. kai Feng, G. Liu, and Y. Yang, "A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm," *Appl. Energy*, vol. 237, pp. 103–116, 2019.

[15] M. Rafiei, T. Niknam, J. Aghaei, M. Shafie-Khah, and J. P. S. Catalao, "Probabilistic load forecasting using an improved wavelet neural network trained by generalized extreme learning machine," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6961–6971, 2018.

[16] A. J. da Rocha Reis and A. P. Alves da Silva, "Feature extraction via multiresolution analysis for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 189–198, 2005.

[17] A. Deihimi, O. Orang, and H. Showkati, "Short-term electric load and temperature forecasting using wavelet echo state networks with neural reconstruction," *Energy*, vol. 57, pp. 382–401, 2013.

[18] M. Ghofrani, M. Ghayekhloo, A. Arabali, and A. Ghayekhloo, "A hybrid short-term load forecasting with a new input selection framework," *Energy*, vol. 81, pp. 777–786, 2015.

[19] T. Zheng, A. A. Girgis, and E. B. Makram, "Hybrid wavelet-Kalman filter method for load forecasting," *Electr. Power Syst. Res.*, vol. 54, no. 1, pp. 11–17, 2000.

[20] S. Bahrami, R. A. Hooshmand, and M. Parastegari, "Short term electric load forecasting by wavelet transform and grey model improved by PSO (particle swarm optimization) algorithm," *Energy*, vol. 72, pp. 434–442, 2014.

[21] H. Aprillia, H. T. Yang, and C. M. Huang, "Optimal decomposition and reconstruction of discrete wavelet transformation for short-term load forecasting," *Energies*, vol. 12, no. 24, p. 4654, 2019.

[22] B. L. Zhang and Z. Y. Dong, "An adaptive neural-wavelet model for short term load forecasting," *Electr. Power Syst. Res.*, vol. 59, no. 2, pp. 121–129, 2001.

[23] N. Amjady and F. Keynia, "Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm," *Energy*, vol. 34, no. 1, pp. 46–57, 2009.

[24] Y. Chen *et al.*, "Short-term load forecasting: Similar day-based wavelet neural networks," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 322–330, 2009.

[25] A. S. Pandey, D. Singh, and S. K. Sinha, "Intelligent hybrid wavelet models for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1266–1273, 2010.

[26] Z. Liu, W. Li, and W. Sun, "A novel method of short-term load forecasting based on multiwavelet transform and multiple neural networks," *Neural Comput. Appl.*, vol. 22, no. 2, pp. 271–277, 2013.

[27] M. Alipour, J. Aghaei, M. Norouzi, T. Niknam, S. Hashemi, and M. Lehtonen, "A novel electrical net-load forecasting model based on deep neural networks and wavelet transform integration," *Energy*, vol. 205, p. 118106, 2020.

[28] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *Fundam. Pap. Wavelet Theory*, pp. 494–513, 2009.

[29] 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, 2013.

[30] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *International Journal of Forecasting*, vol. 30, no. 2. Elsevier, pp. 357–363, 2014.

[31] N. Zhang, Z. Li, X. Zou, and S. M. Quiring, "Comparison of three short-term load forecast models in Southern California," *Energy*, vol. 189, p. 116358, 2019.

[32] C. E. Kontokosta and C. Tull, "A data-driven predictive model of city-scale energy use in buildings," *Appl. Energy*, vol. 197, pp. 303–317, 2017.

[33] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, 2001.

[34] M. Talaat, M. A. Farahat, N. Mansour, and A. Y. Hatata, "Load forecasting based on grasshopper optimization and a multilayer feed-forward neural network using regressive approach," *Energy*, vol. 196, p. 117087, 2020.

[35] R. Houimli, M. Zmami, and O. Ben-Salha, "Short-term electric load forecasting in Tunisia using

artificial neural networks," *Energy Syst.*, vol. 11, no. 2, pp. 357–375, 2020.

[36] M. Misiti, "Wavelet Toolbox," *MathWorks Inc., Natick, MA*, vol. 15, p. 21, 1996.

[37] NYISO, "New York Independent System Operator," 2020. http://www.nyiso.com (accessed Sep. 03, 2020).