Short-term Load Forecasting Method Based on Wavelet and Reconstructed Phase Space

Zunxiong Liu[1, 2], Zhijun Kuang[2], Deyun Zhang[1]

1. *Dept. of Information and Communication Eng, Xi'an Jiaotong University. Xi'an, Shanxi, China.*

2. *Dept. of Information Eng, East China Jiaotong University. Nanchang, Jiangxi, China*

Abstract: This paper proposed wavelet combination method for short-term forecasting, which makes merit of wavelet decomposition and different local approximation algorithm in reconstructed phase space. The load series are non-stationary, observed from a chaotic dynamical system. With the view of wavelet multi-resolution analysis, the original load series are decomposed with A Trous algorithm into a relatively stationary residual and a set of wavelet coefficients, which are taken as independent dynamical subsystems. The subsystems are coped with different local approximation algorithms, with corresponding parameters, producing predicted coefficients. The coefficients are reconstructed with wavelet method to obtain prediction value. Case study demonstrates that the wavelet Combination method has a good performance for less-step predictions, compared with the result from direct method to the original series.

Keywords: *Short Term Load Prediction; Wavelet Decomposition; Local Approximation Algorithm; Chaos Theory*

1. Introduction

STLF(Short-tem load forecasting) plays an important role in the operation of electric power system, especially in a deregulated power market, where many systems are being pushed in a stressed situation close to their security margin. So the system security is more concerned in electricity industry. It's urgent to take the operational planning seriously based on accurate and effective short-tem load forecasting, enforcing the system security. The aim of short-tem load forecasting is to predict future electricity demands, based usually on historical data and other factors, such as weather conditions. In general the techniques for STLF have two approaches, they are regression analysis and time series analysis, which only take historical data. In this paper, we consider a 48-point-per-day electricity demand data series. In view of these, we will employ the prediction techniques related to time series.

Short-tem load series is nonlinear non-stationary time series typically produced by a dynamic system with chaotic behavior [1,2]. Traditionally, short-term load forecast techniques use load shape models[3] which relay on time series analysis techniques, such as Autoregressive moving average(ARMA) model, but basically these techniques are linear methods. In fact, because non-linearity exists in the load series, artificial neural network (ANN) and dynamical system approach have been widely used in the area of short-term load forecasting [4-8]. Moreover, wavelet multi-scale analysis is introduced to build wavelet network model for STLF and other applications in order to improve the forecasting accuracy [9,10]. In this paper, we study local prediction algorithm in the reconstructed phase space combined with wavelet transformed data. Our strategies make use of the merits of wavelet analysis and chaos, and approximate the decomposed time-series at different levels of resolution, where wavelet decomposition is implemented with A Trous algorithm[10]. With the help of this technique, a time series can be expressed as an additive combination of the wavelet coefficients at different resolution levels, resulting in a residual coefficient series and a set of wavelet coefficient series. We analyze and make predictions for each of the series with chaos theory, and get predicted values with wavelet reconstruction.

This paper is organized as follows. In Sect.2, wavelet transform A Trous algorithm is introduced. In Sect.3, we discuss the reconstructed embedding phase space and local approximation algorithm for analyzing nonlinear dynamical system. Procedure for the proposed method is presented in Sect.4. Application to real load series, involving parameters election method and numerical simulation, were presented in Sect.5 followed by discussion and conclusions in Sect.6.

2.Discret Wavelet Transform Algorithm

Wavelet transform provides a way of analyzing a signal in both time and frequency domains, categorized into CWT (continuous wavelet transform) and DWT (discrete wavelet transform). CWT is mainly used for theory research, but DWT is more popular in engineering area, because the observed time series are discrete in real world, including short-term load series. For a suitably chosen mother wavelet function ψ, a function $f(t)$ can be expressed as:

$$f(t)=\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}W_{jk}2^{j/2}\psi(2^{j}t-k) \qquad (1)$$

where the functions $\psi(2^{j}t-k)$ are all orthogonal to each other. The coefficient W_{jk} gives information on the behavior of the function $f(t)$ concentrating on the effects of scale around 2^{-j} near time $2^{-j}\times k$. This wavelet decomposition of a function is closely related to a similar decomposition (the discrete wavelet transform, DWT) of a signal observed in discrete time.

It's well known that DWT has many advantages in signal compression, but DWT often suffer from a lack of translation invariance in time series analysis. Hence a redundant or non-decimated wavelet transform is introduced to tackle the problem. A redundant transform based on a n-length input time series has a n-length series for each of the resolution levels of interest. Therefore, information at each resolution scale is directly related at each time point. A Trous algorithm is always used to realize the shift-invariant wavelet transform.

Let $Z(t)$ (or $C_0(t)$) denote the original discrete time series. A Trous decomposition algorithm as following:

$$C_i(t)=\sum_{l=-\infty}^{+\infty} h(l)C_{i-1}(t+2^i l) \quad (l=1,2,\cdots) \tag{2}$$

$$W_i(t)=C_{i-1}(t)-C_i(t) \tag{3}$$

where $h(l)$ is the discrete low-pass filter, $C_i(t)$, $W_i(t)$ $(i=1,2,\cdots p)$ are approximation-coefficient (residual) series and detail-coefficient series at the resolution level i respectively. $W_1(t)$, $W_2(t)$, $\cdots$, $W_p(t)$ and $C_p(t)$ are called discrete wavelet transform with the resolution level P. In equation (2), extending of boundaries may be dealt with the approach of taking $C(n+k)=C(n-k)$. After the wavelet decomposition, some interesting characteristics about the signal, such as period, hidden period, dependence and jump can be diagnosed easily from the wavelet components.

Given wavelet decomposition coefficient series, the wavelet reconstruction of the original time series is obtained by

$$Z(t)=C_p(t)+\sum_{i=1}^{p} W_i(t) \tag{4}$$

A Trous decomposition and reconstruction algorithm are simple and rapid. It's key to determine the discrete low-pass filter for it.

Given a short-term load series data, We can get a residual series and a set of wavelet-coefficient series by way of above-mentioned wavelet decomposition. The number of resolution level p is empirically determined by way of the inspection of the smoothness of derived scale-coefficient series .

3. Phase Space Reconstruction Forecasting Method

There are many variables in real dynamic systems, such as electric power system. These variables make coordinates of the phase space diagram that describe the system. The trajectories of the phase-space diagram describe the evolution of the system from some initial state (assumed to be known) and hence represent the history of the system. A point in the phase-space represents the state of the system at a given time.

When dealing with a dynamic system, it's impossible to get information about all the variables. Under such circumstances, one way to represent the dynamics of the system is by way of PSR (Phase Space Reconstruction), i.e embedding of a single-variable time series in a multi-dimensional phase-space, because a nonlinear system is characterized by self-interaction, so a scalar time series of a single variable can carry the information about the dynamics of the entire multi-variable system.

A reconstructed phase space is an *m*-dimensional metric space into which a time series is embedded. Given a time series $\{x_t, t=1,2,\cdots n\}$, according to Takens Theorem [11], if *m* (number of dimension of phase space) is large enough, the phase space is homeomorphic to the state space that generated the time series, i.e. the phase space contains the same information as the original state space. The time-delayed embedding of a time series maps a set of *m* time series observations taken from $\{x(t)\}$ onto $X(t)$, where $X(t)$ is a vector or point in the phase space, defined as:

$$X(t)=(x(t),x(t-\tau),\cdots x(t-(m-1)\tau)) \tag{5}$$

Where $t=(m-1)\tau+1,\cdots,n$, m is embedding dimension, τ is delay time. Methods for choosing appropriate τ and *m* together with a discussion can be found in [12,13]. By this process, a one-dimensional time series can be unfolded into *m*-dimensional phase space.

In the reconstructed phase space there is a smooth map $f:\boldsymbol{R}^m \to \boldsymbol{R}$ such that

$$\begin{aligned} x(t+\tau) &= f\left(x(t),x(t-\tau),\cdots x(t-(m-1)\tau)\right) \\ &= f\left(X(t)\right) \end{aligned} \tag{6}$$

where $X(t)$ is state vector at time t (current state), $x_{t+\tau}$ is time series value at time $t+\tau$ (future time). Thus time series prediction can be carried out as long as an appropriate expression for f is found.

There are several approaches for determining f, the local approximation method being most widely used[14]. In order to predict $x(t+\tau)$ into the future from the last *m*-dimensional vector point $X(t)$, we have to find all the nearest neighbors $\{X^r(t)\}_{r=1}^{nr}$ in the ε-neighbourhood of this point. The number of the nearest neighbors can be selected by a trial and error procedure. To be more specific, let $B_\varepsilon(X(t))$ be the set of points within ε of $X(t)$ (i.e. the ε-ball). Thus any point in $B_\varepsilon(X(t))$ is closer to the $X(t)$ than ε. All these points $B_\varepsilon(X(t))$ come from the previous trajectories of the system and hence we can follow their evolution into the future state points $\{X^r(t+\tau)\}_{r=1}^{nr}$. The final prediction for the point $X(t)$ can be obtained by suitable local approximation algorithm, weighted projection averaging and weighted base function algorithms are employed in this paper.

The weighted projection averaging algorithm can be written as

$$x(t+\tau) = \frac{1}{nr}\sum_{r=1}^{nr} w^r x^r(t+\tau) \qquad (7)$$

$$w^r = \frac{\exp(-(d^r - d_{\min}))}{\sum_{r=1}^{nr}\exp(-(d^r - d_{\min}))} \qquad (8)$$

where d^r is the distance between $X^r(t)$ *and* $X(t)$, $d_{\min}$ is minimum value among $\{d^r\}_{r=1}^{nr}$.

The weighted base functions algorithm works in another way. $x(t+\tau)$ can be predicted with $X(t)$ via the evolutionary relationship,

$$\begin{aligned} x(t+\tau) &= \mathbf{C}(\mathrm{t})\Phi(\mathbf{X}(\mathrm{t})) = (\mathrm{c}_1(\mathrm{t}), \mathrm{c}_2(\mathrm{t})\cdots \mathrm{c}_M(\mathrm{t}))\times \\ &(\phi_1(\mathrm{X}), \phi_2(\mathrm{X}), \cdots \phi_M(\mathrm{X}))^{\mathrm{T}} \end{aligned} \qquad (9)$$

where $\boldsymbol{C}(t) = (c_1(t), c_2(t), \cdots c_M(t))$ is coefficient vector that need to be determined, $\Phi(\mathbf{X}(\mathrm{t}))$ is a vector of M local base functions, consisting of polynominals or radial basis functions. In this paper linear basis functions are adopted as special cases of the polynominals, i.e.

$$\begin{aligned} \Phi(X(t)) &= (\phi_1(X), \cdots \phi_M(X))^T \\ &= (1, x(t), x(t-\tau)\cdots x(t-(m-1)\tau))^T \end{aligned} \qquad (10)$$

Here $M = m+1$.

The set of nearest neighbors $\{X^r(t)\}_{r=1}^{nr}$ can be used to estimate the coefficient vector $\boldsymbol{C}(t)$.

$$X^r(t) = (x^r(t), x^r(t-\tau)\cdots x^r(t-(m-1)\tau)) \qquad (11)$$

At time level $t+\tau$, $X^r(t)$ evolve to $X^r(t+\tau)$ which should be in the neighborhood of $X(t+\tau)$. Then the coefficient vector $\boldsymbol{C}(t)$ can be determined by minimizing

$$\sum_{r=1}^{nr}\left| w^r(x^r(t+\tau) - \sum_{m=1}^{M} c_m(t)\phi_m(\boldsymbol{X}^r(t)))\right|^2 \qquad (12)$$

Once the basis functions are known, the above minimization can be tackled by weighted least squares, resulting in $\boldsymbol{C}(t)$.

4. Procedure for the proposed method

Our proposed method made merit of wavelet transform and different local prediction techniques in reconstructed phase space. Every series produced from wavelet decomposition of the original series are taken as independent chaotic sequences, and predict next-time coefficients for these series with different local approximation algorithms. The weighted projections averaging is used to predict coefficients for the higher frequecncy coefficients series because they have lower amplitude and usually taken as noise, but for other series we will apply weighted base functions algorithm. The procedure for this paper is formulated as follows.

- choose a suitable resolution level with wavelet transform for given load series;
- perform wavelet transform on the load series $\{x_t, t = 1, 2, \cdots n\}$ with a trous algorithm , the detail and approximation coefficients at scale j is denoted by $s_{j,t}$;

$$D = \left\{\left(\left(s_{j,k-(m-1)\tau}, \ldots, s_{j,k-\tau}, s_{j,k}\right), s_{j,k+\tau}\right) \in \mathbf{R}^m \times \mathbf{R}\right\}$$

- choose optimal m, τ (scale with discrete sample units) for every series, such as $s_{j,t}$, forming sample space data
- analyze every series with dynamical system theory. use weighted projections averaging algorithm for the higher frequecncy coefficients series, and weighted base functions algorithm for other series. The corresponding predicted coefficients can be obtained;
- produce predicted target values through wavelet reconstruction with above obtained coefficients.

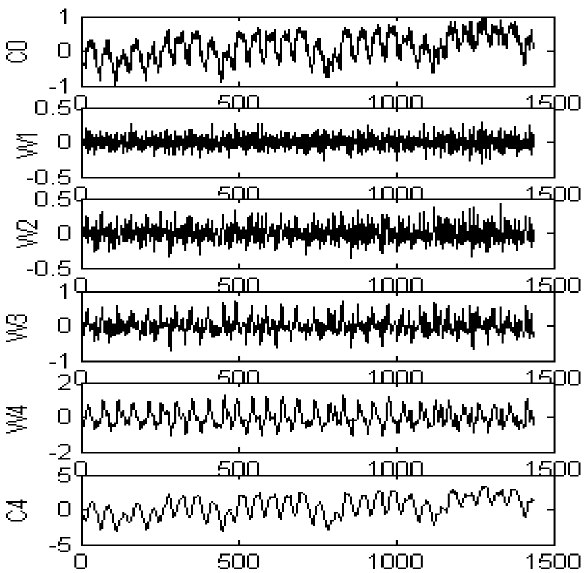

Fig.1. Illustration of the *a trous* wavelet decomposition of a series of electricity demand

load series, where the data have been normalized. In addition, the original series and residual are plotted in the same figure for comparison. The higher the resolution level, the smoother the corresponding wavelet coefficient series.

At the second stage, different local approximation algorithm are employed to predict coefficients with different series. We will apply the weighted projection averaging for the first two high frequency series, just because their amplitudes are relative lower and series burst more often. For other series, the weighted base functions algorithm is employed.

The choice the embedding dimension m and time delay τ for every series are crucial to our algorithm. Here the time delays are determined with the time delayed mutual information suggested by Fraser and Swinney[12]. A certain value of τ corresponding to the minimum value of mutual information is a good candidate for a reasonable time delay. In the instance, the time delays for all the series are set to 1 by caculation. For the model structure, we adopted the Cao's algorithm to estimate the embedding dimension for all the series[13].

Table 1 Load forecast performance on training data on APE measure

	1	2	3	4	5	6	7
μ_w	0.021265	0.026508	0.03547	0.040665	0.06059	0.1053	0.20633
$\sigma^2_w (\times 10^{-1})$	0.002909	0.0039591	0.016583	0.057693	0.22719	0.61227	0.84906
μ_{c0}	0.023016	0.031128	0.037354	0.042209	0.047332	0.060692	0.054106
$\sigma^2_{c0} (\times 10^{-1})$	0.0032876	0.0068307	0.012868	0.013642	0.017745	0.047762	0.024452

5. Application

In this section, the proposed algorithm is applied to predict a real electricity demand series $\{x_t, t = 1, 2, \cdots 1400\}$, one-month data recorded every half hour, that is a series of 1400 values. We hope to extrapolate into the future with 1 or more than 1 subsequent values. Before wavelet transform, the low-pass filter $h(l)$ should be determined, here B_3 spline is used,

$$h(l) = (\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16})$$

We select the resolution level for wavelet transform as 4 by observing the smoothness of the final residual series, and carry out a trous wavelet transform on values x_1 - x_{1400}. In Fig.1, we show the behavior of the four wavelet coefficients over a 1400 points for a

Thus we cope with every series one by one. The sample data space for each of the series is divided into two subsets referred to the training and test subset, the training subset first contains 800 time entries, where the local approximation methods take the nearest neighbor. The other entries constitutes the test subset.

In this experiment, one-to-seven-step predictions were all carried out over test subset. The one-step prediction does take real coefficients for every prediction time, but other two-to-seven-step prediction refer to taking real coefficients for the first prediction, and doing forecasting for other steps with predicted coefficients as input values, and the training subset will be expanded with the known information every circulation. For instance, a circulation for seven-step predictions contains 7 prediction time points.

At the third stage, the predicted results of all the scales for one-to-seven-step predictions are combined with the linear additive reconstruction property of the *a trous*, see Eq.(4).

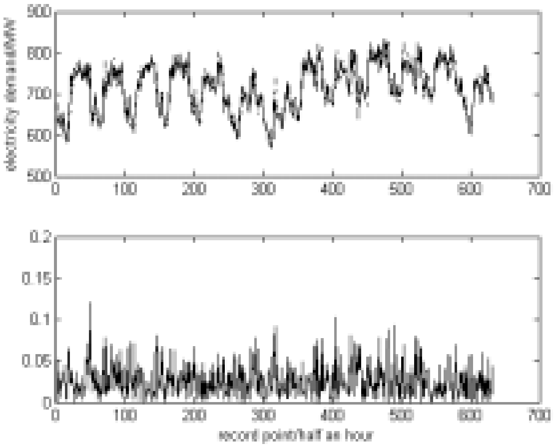

Fig.2 Forecasting the electricity demand with two step on test data set and APE for it with combination method

In comparison, the direct predictions for the original load series are made with one-to-seven steps, the corresponding embedding dimension and delay time are set as above-mentioned method. The local weighted base function approximation algorithm was employed with the predictions.

Moreover, this paper adopted the absolute percentage error (APE) to assess the forecasting performance. APE is defined as:

$$e_{APE}^{k} = 100\frac{L^k - \hat{L}^k}{L^k} \quad k = 1, 2, \cdots n$$

where L^k and $\hat{L}^k$ is the real and predicted load at time instance k, respectively. The average and standard deviation (S.D) for APE over the forecasting range of interests are calculated, meaningfully representing forecasting error.

In Tables 1 are presented the forecasting results for two methods, denoted by the different subscripts. c0 indicates the direct forecasting schemes for the original series with chaos theory, and w means the proposed wavelet combination prediction method, referred to as direct and combination method respectively. The results in Table 1 are the mean, variance over the absolute percentage error (APE) for different approaches, in view of which we observe that the combination method is better that the direct method for less-than-four-step predictions. As the prediction step increases, the performance reduces, even dramatically. In Fig.2, we show the curve of prediction values(solid line), the actual value(dotted line) and APE for the proposed combination method on test data set with two steps.

Delving into the above phenomena, we study that the reason for it may arise from the related aspects. We know, the chaotic time series is born with the typical feature of sustained irregularity and long-term prediction impossibility. Moreover, combination prediction method uses a different local approximation model for each of the produced series, resulting in certain errors. The kind of error will increase with the prediction step, making the performance unsatisfactory. In the meantime, noise existing in original series from dynamical system will impact the forecasting performance.

6. Conclusion and discussion

In this paper, wavelet transform combination method with chaos theory is applied in short term load forecasting, improving the accuracy for less-step predictions. Wavelet decomposition techniques are used to gain deeper insight into the demand data series, based on which different local approximation algorithms are employed to predict.

Although The weighted projection averaging algorithm is used with the high frequency sun-series, the prediction accuracy for it is not so good. It resulted from noise in the electricity demand series to some degree, because the noise distorts the trajectories in state space, and make the modeling of dynamics more difficult and biased. In the future work, we will focus on noise reduction and outliers filtering to improve prediction performance, a hopeful and applicable research direction with dynamical system.

Sustained irregularity

Reference

[1]Jae Gyun Choi, Jong Keun Park, Kwang-Ho Kim, *et al.* A daily peak load forecasting system using a chaotic time series[C]. roceedings of International Conference on Intelligent Systems Applications to Power Systems. France：1996.

[2]Mori, H, Urano S. Short-term load forecasting with chaos time series analysis[C]. Proceedings of International Conference on Intelligent Systems Applications to Power Systems. France：1996.

[3]D.K. Ranaweera, G.G. Karady, R.G. Farmer, Effect of probabilisticinputs on neural network-based electric load forecasting,IEEE Trans. Neural Works 7 (6) (1996) 1528–1532.

[4]LEEKY, CHAYT, PARKJ H. Short - term load forecasting using an artificial neural network [J] . IEEE Transactions on Power Systems , 1992 ,7(1) :124 - 132.

[5]Kholtanzad A,Atlhami-Rohani,ANNSTLF-A Neural Network-Based Electric Load Forecasting System[J]. IEEE Trans. On Neural Network, 1997,8(4):835-846

[6]Aussen A et Murtagh F. Combining neural network forecasts

on wavelet2transfomed time series. Connection Science,9(1997),113-121.
[7] Kiartzis S, Bakirtzis A, Petridis V. Short-term load forecasting using neural networks. Electric Power Syst Res 1995;33(1):1– 6.
[8]Sun Hongbo, Qin Yihong, et. The artificial Neural Network Approach for short-term Load forecasting, Journal of Chongqing University(Natural Science Edition), 7(1995),pp.42-47, in Chinese.
[9]B.Zhang, "Adaptive Neural-Wavelet Model for Short Term Load Forecasting, *International Journal of Electric Power Systems Research*, Vol.59, pp.121-129, 2001.
[10]Wensheng Wang, Ding, , Wavelet Network Model Wavelet Network Model and Its Application to the Prediction of Hydrolog, Nature and Science, 1(1), 2003.
[11]F. Takens, in: Springer lecture notes in mathematics, Vol. 898, eds. D.A.Rand and L.S.Young (Springer, Berlin, 1981) pp. 366-381.
[12] A. M. Fraser and H. L. Swinney, Independent coordinates for strange attractors from mutual information, Phys. Rev. A **33**, 1134 (1986).
[13] Cao, L. (1997), Practical method for determining the minimum embedding dimension of a scalar time series, Physcai D, 110, pp. 43-50.
[14]Farmer, D.J., Sidorowich, J.J., 1987. Predicting chaotic time series.Phys. Rev. Lett. 59, 845–848.