Published by :
http://www.ijert.org
International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 11, November-2016

T-Test, Wavelet Transform and Fuzzy Logic in Load Curve Forecasting

P. T.T.Binh
Faculty of Electrical and Electronics Engineering,
HoChiMinh city University of Technology,
Ho chi Minh City, Viet Nam

Dinh X.Thu
HoChiMinh city University of Technology,
Ho chi Minh City, Viet Nam

Trong Nghia Le
Faculty of Electrical and Electronics Engineering,
HoChiMinh city University of Technology and Education,
Ho chi Minh City, Viet Nam

***Abstract*— The set of load curves can be regarded as the no stationary time series with some specialties due to the repetition of consumption behavior. These yield some correlation between the loads at specified hour and at other hours in the past. The T-test allows finding these correlations. The set of load curve may be treated as one times series if there are the correlations between load at t-hour and previous hours in the same day, load at t -hour but of the previous days. This set will be separated into two sets of load curves for working days and weekends in the case with eminent load difference between two kinds of day. And finally, 24 time series corresponded to 24 hours are examined if there is not correlation between load at t-hour and previous hours in the same day. This paper presented the load curve forecasting based on MODWT and fuzzy logic. The forecasting is carried out for one utility in the South of Vietnam**

Keywords*— *T-test, time series, MODWT; fuzzy-rules; cluster centers

I. INTRODUCTION

Recently the application of modern modeling techniques like Wavelet decomposition and fuzzy logic in forecasting time series is carried out. It is because theses techniques can overcome the troubles of no stationary series [1]. These techniques are also very successful for load forecasting [2][3][4].

The daily load curve set can be treated as the time series and the forecasting has some specialties due to the repetition of consumption behavior. In [5], the load forecasting was based on the maximal overlap discrete wavelet transform (MODWT) and the correlation between loads at one moment with another moment is carried out by the experiences. But basically, this correlation must be based on some mathematic analysis. Beside, regarding the load curve set as one time series may lead to a raw error when there is the eminent difference between load curves of working days and weekends.

In this paper, we will use the T-test to find out the correlation between the loads at different moments and propose three ways to load curve forecasting when there is the significant difference between load curves of working days and weekends. For implementation, the forecasting of one utility in the South of Vietnam is carried out.

This paper will use the MODWT (maximal overlap discreet wavelet transform) and fuzzy logic with subtractive clustering [6] for load curve forecasting

II. T-TEST AND WAVELET DECOMPOSITION

A. *MODWT*

For a time series *X* with *N* samples, the MODWT yields an additive decomposition or MRA given by**:**

$$X = \sum_{j=1}^{J_0} D_j + S_{J_0} \qquad (1)$$

where D_j is the detail and S_{J_0} is the trend:

$$D_{j,t} = \sum_{l=0}^{N-1} \tilde{h}^{\circ}_{j,l} \tilde{W}_{j,t+l \bmod N}$$

$$S_{j,t} = \sum_{l=0}^{N-1} \tilde{g}^{\circ}_{j,l} \tilde{V}_{j,t+l \bmod N} \qquad (2)$$

B. *T-test*

T-test is necessary to find out the relationship between one variable with another variables. And if there is some relationship between one variable with itself in the past, the auto regression model will be used. For daily load curve forecasting, T-test will find out whether the load at t-hour is depended on the loads at some previous hours, for example: *h* previous hours in the same day, t-hour but of *d*-previous days or *w* weeks before.

C. *The load curve set is regarded as one time series*

The set of daily load curves may be treated as one time series. From the T-test result, for each series $D_{j,t}$, at the current time t, there will be a relationship as the following:

$$D_{j,t} = f_j(D_{j,t-24*7*w},, D_{j,t-24*d}, D_{j,t-h}) \qquad (3)$$

The same way will be applied for $S_{j,t}$.

Equation (3) means that each $D_{j,t}$ is related to itself at moment *t-1,..t-h* in the same day, to itself in *1,2…d-* days before and *1,2,…w* weeks before.

D. The load curve set is divided in two time series

When the load curves have the eminent difference between the work days and weekends, separating the original time series is necessary to get higher accuracy. If there are strong correlations between load at t-hour with t-1 hour and t-24 hour, the following considerations are included:

The working day load curves time series: The forecasting for the t hour of the first working day in a week must use the load of the last weekend. So the load curve of this last day of weekend days must be included to involve the influence of t-1 or t-24 hours in this series. The left hand value in (3) or forecasted value will belong to the working days, not to weekends.

The weekend day load curves time series: Load curve of Friday must be included because the hourly loads of Saturday are related to Friday. The left hand value in (3) or forecasted belonged to the weekends, not to Friday.

To include the influence of w week, it is necessary to say that the day number in one week in (3) is not seven. Suppose the working day load curves time series consists of 5 days a week, then (3) will be:

$$D_{j,t} = f_j(D_{j,t-24*5*w},, D_{j,t-24*d}, D_{j,t-h}) \qquad (4)$$

E. The load curve set is regarded as 24 time series

If there is no correlation between load at t-hour and t-1 hour, the load curve set may be regarded as 24 time series as the following:

Series 1 consists of the loads at first hour. Series 2 consists of the loads at second hour and so on. Equation (3) shows only the influence of *d* previous days and *w* weeks before.

III. DETERMINING FUZZY RULES

Equation (3) can be approximated by some rules. The number of rules is the number of cluster centers. The paper will develop the subtractive clustering in [6]. The elements in (3) will be considered as members of the following vector:

$$\{\underbrace{D_{j,t-24*7*w},\dots,D_{j-24*d},\dots,D_{j,t-h},\dots D_{j,t-1}}_{\text{Input } y}\,\Big|\,\underbrace{D_{j,t}}_{\text{output } z}\} \qquad (5)$$

Examining the set of vectors x, each vector consists of two parts: input *y* and output *z*.

Consider a collection of *n* data points $\{x_1, x_2, \dots x_n\}$ in an M dimensional space. Using the subtractive clustering proposed by Chiu [6], the set of centers $\{x_i^*\}$ will be determined.

Each centre x_i^* of input y^* and output z^* will be regarded as one fuzzy rule. For each input vector *y*, its degree to satisfying the *i*-fuzy rule is:

$$\mu_i = e^{-\alpha\|y-y_i^*\|^2} \qquad (6)$$

The output will be:

$$z = \frac{\sum_{i=1}^{c}\mu_i z_i^*}{\sum_{i=1}^{c}\mu_i} \qquad (7)$$

Where c-number of centers

Yager và Filev suggested that Z_{ij} in (7) will be the linear function of the inputs as following:

$$z_{ij}^* = G_i y + h_i \qquad (8)$$

Here G_i –matrix of constants with *(N-1)x1*-dimension; *h*-column vector of constants with *N-1* elements where. *N-1*-dimension of input

Now denoting

$$\rho_i = \frac{\mu_i}{\sum_{j=1}^{c}\mu_j} \qquad (9)$$

Then (8) is rewritten as:

$$z = \sum_{i=1}^{c}\rho_i z_i^* = \sum_{i=1}^{c}\rho_i(G_i y + h_i) \qquad (10)$$

Or:

$$z^T = \begin{bmatrix}\rho_1 y^T & \rho_1 \dots \rho_c y^T & \rho_c\end{bmatrix}\begin{bmatrix}G_1^T \\ h_1^T \\ \vdots \\ G_c^T \\ h_c^T\end{bmatrix} \qquad (11)$$

With a set of n inputs $\{y_1,\dots,y_n\}$, the set of outputs will be:

$$\begin{bmatrix}z_1^T \\ \vdots \\ z_n^T\end{bmatrix} = \begin{bmatrix}\rho_{1,1}y_1^T & \rho_{1,1} & \cdots & \rho_{c,1}y_1^T & \rho_{c,1} \\ & & \vdots & & \\ \rho_{1,n}y_n^T & \rho_{1,n} & \cdots & \rho_{c,n}y_n^T & \rho_{c,n}\end{bmatrix}\begin{bmatrix}G_1^T \\ h_1^T \\ \vdots \\ G_c^T \\ h_c^T\end{bmatrix} \qquad (12)$$

The estimation of *G* and *h* in (12) can be realized by mean least square method. After evaluating *G* and *h*, for given *y* at moment *t+1*, we can calculate the output z_{t+1} as the one step ahead forecasting using (10).

The load forecasting for next moment will be carried out as:

$$X_{t+1} = \sum_{j=1}^{J} D_{j_{t+1}} + S_{J_{t+1}} = D_{1_{t+1}} + D_{2_{t+1}} + \dots + D_{J_{t+1}} + S_{J_{t+1}} \qquad (13)$$

For the case of 2.3 and 2.4, the forecasted load at moment t+1 will be used for forecasting load at t+2 and so on. That means this it the 24 steps ahead forecasting.

IV. APPLICATION

A. *The load curve set is an time series*

The data for training are hourly loads of one utility in the South of Vietnam from 3/1/2011 to 7/18/2011. The data for testing is 15 days, from 7/11 to 8/2. The T-*test* shows that the load at *t* hour is related to the loads at the following hours by decreased order of importance: t-1, t-24, t-168, t-338, t-672 hour.

The forecasted load at t+1 hour will be used for forecasting load at the next hour to get the whole daily load curve. The errors are displayed in the Table 1 with the mean error is 6.10%

B. *The load curve set is divided in two time series*

The T-test for the load curves from 03/01/2011 to 12/19/2011 has the best correlation with *h=1; d=1; w=4*

These load curves have the eminent difference between the working days (excluded Monday) and weekend. Here the load curve of Monday looks like those of Weekends

The first time series is the hourly loads for Mondays, Tuesdays, Wednesdays, Thursdays and Fridays. Monday is included to get the influence of the day before Tuesday. The results are presented in Table 2. The mean error is 3.07%. The concrete details of 24 steps ahead forecasting for one day are shown in Tabl.3 and Figure 1.

For the time series of Fridays, Saturdays, Sundays and Mondays, the load forecasting is carried out for Saturday, Sunday and Monday from 11/12 /2011 to 12/19/2011. The errors are shown in Table 4. The mean error for 18 weekends is 3.43%. The detail forecasting for one day are presented in Table 5.

TABLE I. The forecasting errors for 15 days (from 7/11 to 8/2/2011).

Date	Day	Error
7/19/11	Tuesday	0.019175
7/20/11	Wednesday	0.038723
7/21/11	Thursday	0.029726
7/22/11	Friday	0.025508
7/23/11	Saturday	0.020794
7/24/11	Sunday	0.047853
7/25/11	Monday	0.05848
7/26/11	Tuesday	0.064154
7/27/11	Wednesday	0.076005
7/28/11	Thursday	0.068913
7/29/11	Friday	0.059469
7/30/11	Saturday	0.104449
7/31/11	Sunday	0.172467
8/1/11	Monday	0.081312
8/2/11	Tuesday	0.04847

TABLE III. The errors for 20 working days.

Date	Tuesday	Wednesday	Thursday	Friday
Date	11/01	11/02	11/03	11/04
Error (%)	3.93	4.67	4.09	4.53
Date	11/08	11/09	11/10	11/11
Error (%)	2.77	3.35	2.16	5.01
Date	11/15	11/16	11/17	11/18
Error (%)	2.48	2.83	4.89	3.77
Date	11/22	11/23	11/24	11/25
Error (%)	1.93	3.16	2.20	1.66
Date	11/29	11/30	12/01	12/02
Error (%)	3.04	3.74	2.11	1.40
Date	12/06	12/07	12/08	12/09
Error (%)	1.89	1.89	2.57	3.57

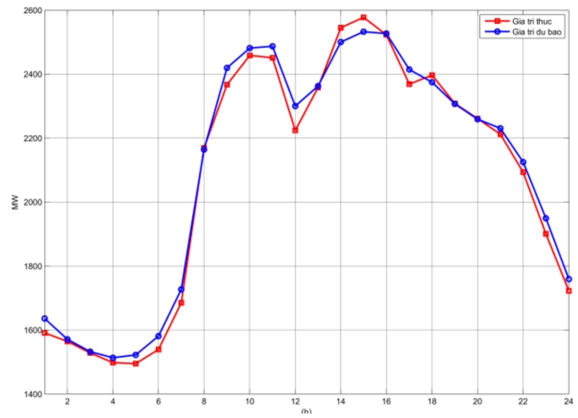

Fig.1. Real load curve and forecasted load curve at 12/2/2011

TABLE III. The forecasting load (MW) for 12/2/2011

Hour	1h	2h	3h	4h	5h	6h
Forecasting	1636.142	1571.178	1532.683	1513.131	1522.08	1580.978
Real load	1590.97	1564.42	1528.86	1498.32	1494.82	1539.51
Hour	7h	8h	9h	10h	11h	12h
Forecasting	1726.921	2164.257	2419.847	2481.365	2487.47	2300.3
Real load	1685.25	2169.66	2366.97	2458.47	2451.23	2224.27
Hour	13h	14h	15h	16h	17h	18h
Fore-casting	2362.799	2500.566	2532.633	2527.214	2414.654	2374.787
Real load	2358.43	2545.47	2577.82	2523.32	2369.02	2396.61
Hour	19h	20h	21h	22h	23h	24h
Forecasting	2306.952	2259.074	2230.387	2125.045	1949.436	1759.001
Real load	2308.53	2260.67	2212.38	2093.63	1900.8	1722.37

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 11, November-2016

C. The load curve set is expressed as 24 time series

The forecasting errors are higher than the previous case. The mean error for working days is 3.7%. It was not good because in this model, the influence of the previous hour was neglected meanwhile according to T-test, there is the correlation between load at t hour and t-1 hour.

TABLE IV. The errors for weekends and mondays from 11/12/2011 to 12/19/2011

	Saturday	Sunday	Monday
Day	11/12	11/13	11/14
Error (%)	3.4	5.35	2.93
Day	11/19	11/20	11/21
Error (%)	3.16	1.87	3.05
Day	11/26	11/27	11/28
Error (%)	1.51	2.21	2.55
Day	12/3	12/4	12/5
Error (%)	3.04	2.27	2.06
Day	12/10	12/11	12/12
Error (%)	3.38	5.48	7.53
Day	12/17	12/18	12/19
Error (%)	3.82	3.4	4.81

TABLE V. Forecasting loads (MW) for 11/26/2011

Hours	1h	2h	3h	4h	5h	6h
Forecasted	1619.956	1561.104	1518.818	1492.393	1493.631	1524.365
Real	1659.51	1569.69	1541.4	1518.2	1516.9	1546.8
Hours	7h	8h	9h	10h	11h	12h
Forecasted	1660.854	2028.948	2287.335	2380.874	2418.591	2212.458
Real	1688.3	2142.33	2335.52	2398.77	2407.61	2169.51
Hours	13h	14h	15h	16h	17h	18h
Forecasted	2252.476	2367.698	2364.746	2363.908	2177.314	2135.019
Real	2241	2384.9	2330.8	2305.82	2137.96	2181.98
Hours	19h	20h	21h	22h	23h	24h
Forecasted	2140.362	2099.144	2084.593	2007.182	1842.413	1652.698
Real	2110.05	2095.67	2113.4	2022.87	1844.55	1684

V. CONCLUSION

The T-test is necessary for finding the correlation between load at one moment and at the previous moments. It also leads to make the load curve set be treated as one time series, as two time series or 24 time series. If there are the eminent differences between the load curves of working days and weekend, separating in two time series will improve the forecasting accuracy. The MODTW allows finding the series of details and trend for time series of load curves. The details and trend at forecasted moment will be related with themselves in the past. These correlations are expressed by fuzzy rules based on the subtractive methods. Examining for one utility shows that the proposed approach based Wavelet Transform and Fuzzy Logic with T-test has the good result.

ACKNOWLEDGMENT

This work was supported by the Ho Chi Minh City University of Technology and Ho Chi Minh City University of Technology and Education.

REFERENCES

[1] Saif Ahmad, Ademola Popoola, and Khurshid Ahmad, Wavelet-based Multiresolution Forecasting, UniS Technical Report(June 2005).

[2] Wuhan, Hubei, Electrical Load Forecasting Based on Fuzzy Wavelet Neural Networks**,** *Conference on Future Biomedical Information Engineering* , pp.122-125, Dec. 2008

[3] Bhavesh Kumar Chauhan[1], Madasu Hanmandlu[2] , Load forecasting using wavelet fuzzy neural network , Journal International Journal of Knowledge-Based and Intelligent Engineering Systems, IOS Press, Volume 14, Number 2 / 2010.

[4] Yuancheng Li; Bo Li; Tingjian Fang; Short-term load forecast based on fuzzy wavelet**,** Intelligent Control and Automation, 2004. WCICA 2004. Fifth World Congress on , 15-19 June 2004.

[5] P.T.T.Binh, N.T.Hung, P.Q.Dung, Lee-Hong Hee, Load Forecasting Based on Wavelet Transform and Fuzzy Logic, POWERCON 2012, Aukland,2012.

[6] Chiu, S.. "Fuzzy Model Identification Based on Cluster Estimation." *Journal of Intelligent & Fuzzy Systems* 2 (3), 1994, 267-278.