HAL
open science

A Combination of Variational Mode Decomposition with Neural Networks on Household Electricity Consumption Forecast

Vanessa Haykal, Hubert Cardot, Nicolas Ragot

To cite this version:

Vanessa Haykal, Hubert Cardot, Nicolas Ragot. A Combination of Variational Mode Decomposition with Neural Networks on Household Electricity Consumption Forecast. Proceedings of ITISE 2017, Sep 2017, Granada, Spain. hal-01588198

HAL Id: hal-01588198

https://hal.science/hal-01588198v1

Submitted on 15 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Combination of Variational Mode Decomposition with Neural Networks on Household Electricity Consumption Forecast

Vanessa Haykal, Hubert Cardot, and Nicolas Ragot

Université François Rabelais Tours, Computer Science Lab (LI, EA 6300), France

Abstract.
Recently, there has been a significant emphasis on the forecasting of the electricity demand due to the increase in the power consumption. This paper presents the computational modeling of electricity consumption based on Neural Network (NN) training algorithms. The noise in signals, which are caused by various external factors, often corrupt demand series and influence consequently on the model performance. For accurate electricity demand forecasting, we propose a novel approach that combines a NN MLP (multilayer perceptron) with VMD (variational mode decomposition)-based signal filtering. Using the daily electricity demand series of EDF (Electricté De France) obtained from the UCI machine learning repository, this paper demonstrates that the proposed VMD-NN model greatly improves the forecasting error comparing to existing stationary stochastic process such as the autoregressive moving average (ARMA) model.

Keywords: neural network algorithms, time series, household electricity consumption forecast, variational mode decomposition, multiresolution analysis

1 Introduction

Domestic energy consumption [1] is the total amount of energy used in a house for household work. The amount of energy used per household varies widely depending on the standard of living of the country, the climate, and the age and type of residence. Energy demand forecasting is a very important task in the electric power distribution system to enable appropriate planning for future power generation. Quantitative and qualitative methods have been utilized previously for the electricity demand forecasting. These methods fail to provide effective results. With the development of the advanced tools, these methods are replaced by efficient forecasting techniques. According to common classifications [2], demand forecasting models are classified based on two different criteria: the forecasting horizon and the aim of the forecast, also we can divide them into linear and nonlinear models and a third group consists of models that use a combination of both.

adfa, p. 1, 2011.
© Springer-Verlag Berlin Heidelberg 2011

This paper presents an improved method for forecasting, we use the VMD-NN model. The VMD is a fully adaptive method for the analysis of nonlinear and non-stationary properties of time series. The original series will be decomposed by the VMD method into several high and low frequency signals. These sub-series will be used in the NN model in order to make the prediction. The forecasting results of this work have revealed that the VMD-NN model outperforms the NN itself and the ARMA models.

The rest of this paper is organized as follows: Section 2 introduces the notions of the classical forecasting models namely the autoregressive moving average and the artificial neural networks. Section 3 shows in details the theory of the recently developed variational mode decomposition. Section 4 contains our experiments and results, we start this part by defining the practical error measurement, we describe different steps to get the optimal training algorithm on our dataset, and we sum up our work with some simple quantitative performance evaluations compared to the baseline models. Section 5 concludes on the effectiveness of our novel approach, and includes some future directions and expected improvements.

2 Classical Forecasting Methods

2.1 ARMA Process

The various researches [3] have used these methods with time series data for the electric power consumption. In [4] Zhu, Guo, and Feng studied the issue of household energy consumption in China from the year 1980 to 2009 with construction on of VAR model. There were two forecasting methods that used ARIMA and BVAR. The results showed that both of them can predict the sustained growth of household energy consumption (HEC) trends. Ediger and Akar [5] applied SARIMA (Seasonal ARIMA) methods to estimate the future primary fuel energy demand in Turkey from the years 2005 to 2020. The research work of Contreras et al. [6] applied ARIMA methods to predict next day electricity price in Californian markets. Conejo et al. [7] applied wavelet transform and ARIMA models to predict day-ahead electricity price of mainland Spain in year 2002.

In this paper, we use the ARMA process [8, 9] as a reference model. It has become a popular linear statistical model for stationary time series analysis and forecasting. The ARMA *(p,q)* generating process is given by

$$\varphi(B)\, v_t = \theta(B)\, e_t$$

where v_t and e_t are respectively the actual value and random error at time period t, B is the backshift operator. The error term e_t are assumed to be independently and identically distributed *(iid)* with a mean $E(e_t) = 0$ and a variance $V(e_t) = \sigma^2$. The polynomials $\varphi(B)$ and $\theta(B)$ are given by

$$\varphi(B) = (1-\varphi_1 B-\cdots-\varphi_p B^p)$$
$$\theta(B) = (1-\theta_1 B-\cdots-\theta_q B^q)$$

where p is the number of autoregressive orders, q is the number of moving average orders, θ is the autoregressive coefficient, and φ is the moving average coefficient.

In particular, the autoregressive (AR) component is expressed by the coefficients φ that represent a linear relationship between the value predicted by the model at time t and the past values of the interest rate variation time series. Similarly, the moving average (MA) component is expressed by the coefficients θ that represent a linear relationship between the value predicted by the model at time t and the error term e.

2.2 Artificial Neural Networks

Based on some literature reviews, the non-linear models, derived from the artificial neural networks (ANNs), have gained more and more attention since the second half of the 80's. This evolution is due to the fact that certain researchers achieved great advances on ANNs.

Artificial neural networks in Fig.1 [10] are a class of statistical learning models inspired by the physiology of biological neural networks. Each neuron performs a specific kind of computation. First, a weighted sum of the input variables and the bias term b is built, with the result being then processed by an activation function $f(t)$. Once the single neuron operation is specified, one can easily calculate the network outputs given an input vector by evaluating the output of each layer by forward input propagation. The result is a function of the network configuration, i.e. its topology and the value of the connection weights. It will be the job of the training phase to learn the weights in order to induce the desired computation.

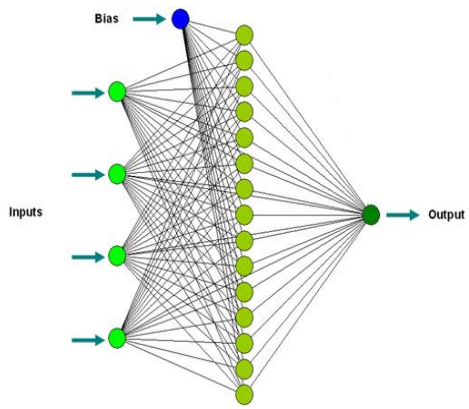

Fig. 1. Schematics of a fully connected multilayer perceptron with four inputs and a bias unit. The weighted input sum is added to the bias term and then enters as argument of the activation function f which generates the output (Neuromaster, 2015)

This has been overcome by the back-propagation [11, 12] algorithm; nowadays it is widely applied in training multilayer perceptron. Given a supervised training set $\{x_i, t_i: i = 1...N\}$ with x_i input variables and t_i target variables, we denote by y_i the correspondent output computed by the network when x_i is fed forward. In general, we have $t_i \neq y_i$. A global error on the training set can be then defined as a quadratic function of the form

$$E(\boldsymbol{w}) = \frac{1}{2N}\sum_i ||\boldsymbol{t}_i - \boldsymbol{y}_i||^2$$

and can be seen as a function of the network weights *w*. Other error definitions are possible, for example by choosing a different norm. The idea behind back propagation is to minimize this error by updating the weights using the gradient descend [13] method (with *k* as iteration index), i.e.

$$w_{ij}^{(k)} \rightarrow w_{ij}^{(k)} - \alpha \frac{\partial E(\boldsymbol{w})}{\partial w_{ij}^{(k)}}$$

The calculation of the partial derivatives is thus crucial for the algorithm. It is done by using directly the dependence of the error function on the training set instances. When all the instances have been used, one 'epoch' of training is completed. Usually many epochs of training are needed in order for the error function to converge to a local or global minimum, resulting in longer training periods.

3 Variational Mode Decomposition

More recently, a new multiresolution technique called variational mode Decomposition (VMD) was introduced by Dragomiretskiy and Zosso (2014) yields better results in signal processing domain specifically in the case of signals without prior knowledge. In [14], they propose an entirely non-recursive variational mode decomposition model. The model looks for an ensemble of modes and their respective center frequencies. We apply this technique on our dataset before using the predictive neural network model described in Section 2.2. In this part, we mentioned a few concepts and tools from signal processing that will constitute the building blocks of the VMD model.

3.1 Denoising Problem

Using a simple denoising problem, an underlying signal f_0 consist of an unknown signal f corrupted by an additive noise, namely the zero-mean Gaussian noise. The use of the Wiener filter is to estimate the unknown signal using an original signal as input. The filter is based on a statistical theory in order to minimize the mean squared error classically addressed using Tikhonov regularization [15].

$$\min_f \left\{ \|f - f_0\|_2^2 + \alpha \|\partial_t f\|_2^2 \right\},$$

The Euler-Lagrange equations are typically solved in Fourier domain:

$$\hat{f}(\omega) = \frac{\hat{f}_0}{1 + \alpha\omega^2},$$

where $\hat{f}$ is the fourier transform of the signal f. This solution corresponds to convolution with a Wiener filter, where α represents the variance of the white noise, and the signal has a low-pass $1/w^2$ power spectrum prior.

3.2 Constrained Model

For a multicomposition real valued signal f, VMD assumes that f is composed of a given number of subsignals u_k (modes). Each mode is regarded as an amplitude-modulated and frequency-modulated (AM-FM) signal and has mostly compact frequency ω_k around a center pulsation [16].

To assess the bandwidth of the modes, the following scheme is proposed by Dragomiretskiy and Zosso (2014):

(a) Compute the associated analytic signal by means of the Hilbert transform to obtain a unilateral frequency spectrum for each mode.

(b) Shift the frequency spectrum of each mode to the baseband by mixing with an exponential tuned to the estimated center frequency.

(c) Estimate the bandwidth through the H^1 Gaussian smoothness of the frequency translated function, that is, the squared L^2-norm of the gradient.

The resulted constrained variational problem is the following:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\sigma(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}$$
$$\text{s.t.} \quad \sum_k u_k(t) = f.$$

where f is the signal, u is its mode, w is the frequency, σ is the Dirac distribution, t is time script, k is number of modes, and $*$ denotes convolution.

Thus, we intend to minimize the sum of the bandwidths defined as the squared L^2-norm of the gradient of the demodulated signal components. To solve the constrained variational problem [16], the augmented Lagrangian is introduced and the non-constrained variational problem is gotten by

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) \\ = \alpha \sum_k \left\| \partial_t \left[\left(\sigma(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 \\ + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle, \end{aligned}$$

where α denotes the balancing parameter of the fidelity constraint and λ is the lagrangian multiplier. The saddle point here is to get the optimal solutions of u_k and w_k using the alternate direction method of multipliers (ADMM).

3.3 Minimization with respect to u_k

The subproblem can be written as the following equivalent minimization problem:

$$u_k^{n+1} = \arg\min_{u_k \in \mathbb{R}} \left\{ \alpha \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f - \sum u_i + \frac{\lambda}{2} \right\|_2^2 \right\}.$$

The solution of this quadratic optimization problem is found by using Parseval/Plancherel Fourier isometry and exploiting the Hermitian symmetry [16]. All the modes can be obtained from the below equation in the frequency domain through updating each mode and its center frequency ω_k constantly:

$$\widehat{u}_k^{n+1}(\omega) = \frac{\widehat{f}(\omega) - \sum_{i \neq k} \widehat{u}_i(\omega) + \widehat{\lambda}(\omega)/2}{1 + 2\alpha \left(\omega - \omega_k\right)^2}.$$

This equation is regarded as the Wiener filtering result of the current residue with signal prior $1/(w - w_k)^2$. Consequently, the mode in time domain is obtained as the real part of the inverse Fourier transform of this filtered analytic signal.

3.4 Minimization with respect to w_k

As before, the optimization takes place in the Fourier domain. The relevant subproblem thus reads:

$$\omega_k^{n+1} = \arg\min_{\omega_k} \left\{ \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}.$$

The new center frequency is put at the center of gravity of the corresponding mode's power spectrum, which can be updated by

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left|\widehat{u}_k(\omega)\right|^2 d\omega}{\int_0^\infty \left|\widehat{u}_k(\omega)\right|^2 d\omega}.$$

4 Experimentation

4.1 Performance Measure

The forecasting performance [9] is examined using the root mean of squared errors (RMSE). It measures the deviation between actual and predicted values. A small value of RMSE means that the predicted time series values are closed to the actual values. Thus, it can be used to evaluate the prediction error. The computation of this criterion is given as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (v_i - p_i)^2}{n}}$$

where v and p represent respectively the actual and predicted value and n is the total number of the sample data points.

4.2 Data Preprocessing and Training

This research used dataset [17] about the electric power consumption in one household that has a sampling rate one minute over a long period of time from the years 2006 to 2010. Following the existing study [3], we use the "Global Active Power" variable which is the household global minute-averaged active power (in kilowatt).

The raw data were not ready for constructing the forecast model because some values are missing and the recorded time frames are inappropriate. The lack of some information [3] may decrease the predictive efficiency of the forecasting model. To fill the missing data, we use the previous value, where we assume that the current data will be similar to the previous ones as shown in Fig.2.

Date	Time	Global_active_power
21/12/2006	11:19:00	0.244
21/12/2006	11:20:00	0.244
21/12/2006	11:21:00	0.242
21/12/2006	11:22:00	0.244
21/12/2006	11:23:00	0.244
21/12/2006	11:24:00	0.244
21/12/2006	11:25:00	0.246
21/12/2006	11:26:00	0.246
21/12/2006	11:27:00	0.244
21/12/2006	11:28:00	0.244

Fig. 2. Fill the missing data by the previous value "0.244" (extracted from [3])

We aggregate the minute-by-minute data into daily observations, and then we got a new sample of 1442 data points. In Fig.3, we represent the shape of the daily HEC time series. It is clearly shown that the data points are highly fluctuated and non-stationary, since their means and variances change over time.

The data is divided into two parts. The first group is the training dataset which contain data from 26/12/2006 to 31/12/2009 (1112 observations) for the construction of the predictive models. The second group is the test sample which contain data from 01/01/2010 to 26/11/2010 (330 observations).

We want to predict the future value based on *n* previous days, where *n* is considered as the window size of the time series. There is no specific rule to define the window, in our case, we choose the first thirty lags for training the artificial neural networks.

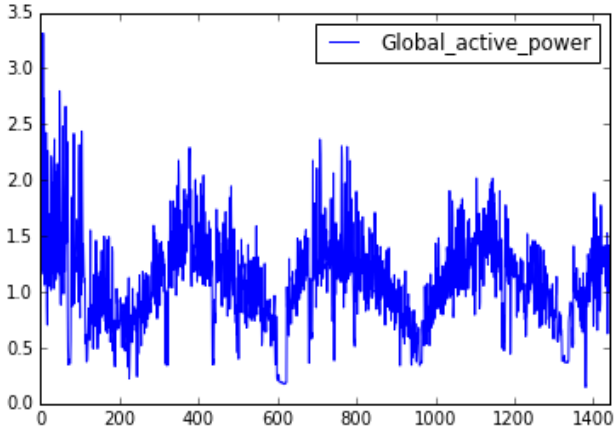

Fig. 3. The daily time series of the Global Active Power from 2006 to 2010

The main aim of this work is to determine the optimal NN for electricity demand forecasting. The model is generated by using the grid search cross validation technique. In brief, the grid search is simply an exhaustive searching algorithm with a manual specified subset for hyperparameter optimization. It must be guided by some performance metric, typically measured by the k-fold cross-validation. The cross validation algorithm does this by splitting the training dataset into k subsets and takes turns training models on all subsets except one which is held out, and evaluating model performance on the held out validation dataset. The process is repeated until all subsets are given an opportunity to be the held out validation set. The performance measure is then averaged across all models that are created.

We used this technique to identify the number of layers (length) and the number of neurons in each layer (width). First, we vary the length from 1 to 5 layers with a width of 10, 60, 110 neurons per layer. We found that the optimal model structure consists of a hidden layer with 110 neurons. Then, we tried again the algorithm for one layer with a range of neurons between 90 and 120; the record gives 100 neurons as the best fit. Thus, the optimal neural network is composed of one layer with 100 neurons. Throughout the training, we use an epoch equal to 1000. Considering that, one epoch is a multiple number of iterations for the gradient descent updates until we show all the data to the NN, and then start again.

As proposed in this paper, we work on the VMD-based signal filtering method to reduce the noise. The VMD algorithm requires predetermining the number of variational modes to be extracted. However, it is not easy to set a rule to determine an appropriate number of modes.

On the one hand, we tried our experiments on the training sample with mode = 6 in order to exemplify the theory described in Section 3. We illustrate in Fig.4 the results of the VMD algorithm on the non-stationary HEC dataset in order to assess the clarification of the proposed approach (Dragomiretskiy & Zosso, 2014).

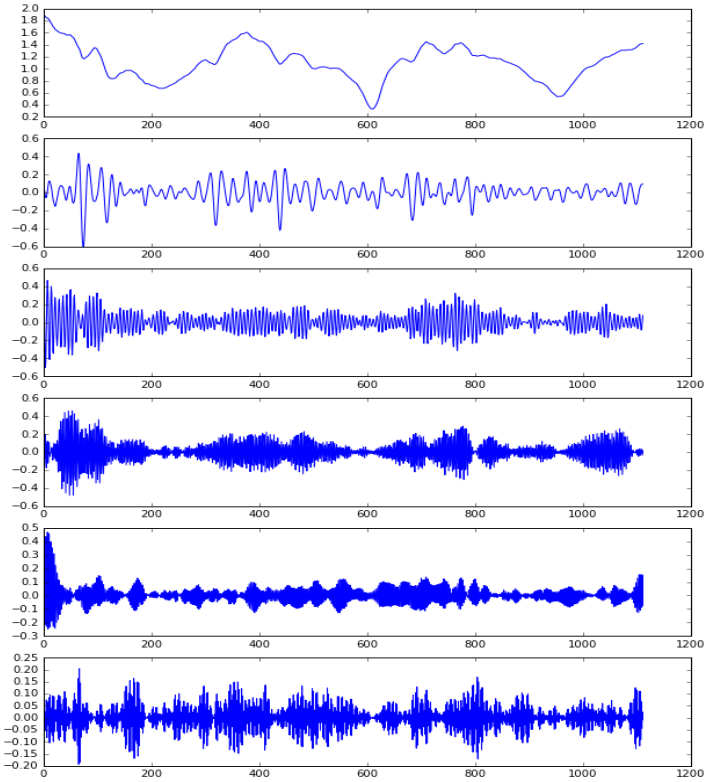

Fig. 4. Applying the VMD = 6 on the train sample

One can clearly see an oscillating low-frequency pattern. The first mode captures the low-frequency oscillation of the baseline. Then, the distinct spikes of the train sample create important higher-order harmonics in the next modes. The 6th mode is the highest frequency mode and contains the most noise with a highly non-sinusoidal spikes.

On the other hand, in order to contribute to the forecasting system, we repeatedly apply the VMD algorithm on every window of the time series to elaborate the modes. We tried our experiments on different levels of decomposition: 8, 10, 15, and 20. In each case, the modes are integrated in the optimized MLP model for 7 day-ahead forecast. Based on the RMSE values explained in the Section 4.1, we found that VMD = 15 gives the most efficient result.

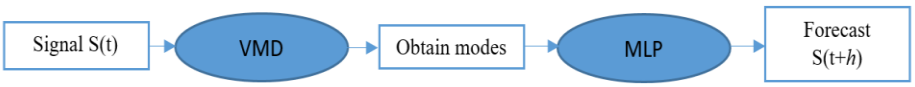

Fig. 5. The VMD-NN model components

As a partial conclusion, the following diagram in Fig.5 explains the different steps of the hybrid combination VMD-NN model in order to build this novel forecasting approach. We mean by signal S(t), the series values that correspond to the window range from 1 to 30. Applying the VMD on each signal, we obtain the 15 different variational modes. Then, using the previous optimized MLP (100,) as indicated in Fig.5, we considered the variational modes as inputs of the model in order to obtain the prediction on horizon *h*.

4.3 Results

We define the Carbon Copy as the model that takes exactly the same value of the previous day, which means that the predicted value is equal to the actual value. For comparison purpose with ARMA models (Chujai et al., 2013), the horizon $h = 7$ was employed.

Based on the test sample and the optimal models, we make a summary of the RMSE values shown in the table 1. Using the NN model, we found a significant improvement in the error comparing to ARMA process for the proposed dataset. The RMSE decreases from 0.34 to 0.272. However, the novel approach VMD-NN model clearly shows an efficient reduction in the error among all the previous studies. Its corresponding RMSE is equal to 0.077. The VMD-NN greatly outperforms the NN itself by decreasing the RMSE from 0.272 to 0.077 for predicting the household electricity consumption dataset.

Models	**RMSE *h* = 7**
Carbon Copy	0.374
ARMA (Chujai et al. [3])	0.340
NN	0.272
VMD-NN	0.077

Table 1. RMSE comparative analysis

Thus, based on the VMD-NN model, the RMSE analysis shows that we divide the error by 4.4 (~ 0.34/0.077) comparing to ARMA, and by 3.5 (~ 0.272/0.077) comparing to NN.

The VMD technique cannot be applied to ARMA model. The variational modes cannot be implemented in its algorithm, since it only involves regressing the variable on its own lagged (i.e. past) values.

The RMSE of the VMD-NN could be also minimized by making a new optimization of the hyperparameters of the neural network. But, as long as we get a significant decrease in the RMSE, we restrict our study to the MLP (100,) to show the effectiveness of the VMD comparing to the same previous MLP.

In Fig. 6, we only plot the first 150 observations of the test sample to clearly show the difference between the curve of the predicted and original values. The curves are very close to each other, thus the VMD-NN model fits the data very well.

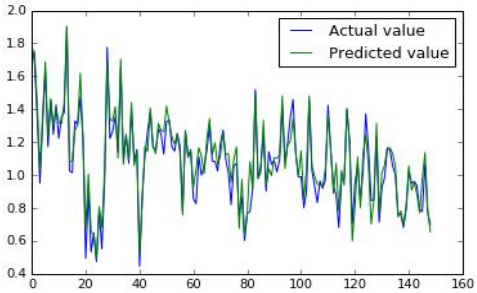

Fig. 6. The actual versus predicted values

Besides the RMSE measures, we also examined the distribution of the forecast errors in order to check the normality of the distribution. The histogram in Fig.7 shows that the errors are normally distributed between [-0.3, 0.3] where the highest point on the curve represents the most probable event in the error close to zero, while all other possible occurrences are equally distributed around the center, creating a downward-sloping line on each side of the peak.

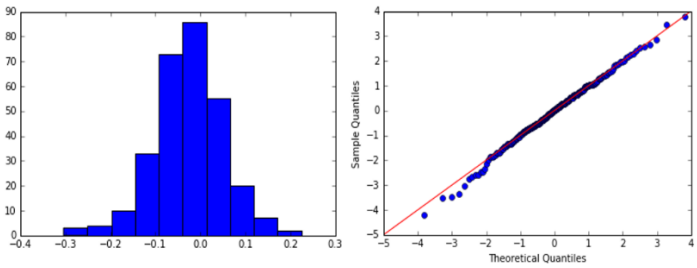

Fig. 7. The histogram (left) and the Q-Q plot (right) of the forecast errors

We use the Q-Q plot in Fig.7 as a test to verify the normality. Roughly speaking, the Q-Q plot take the sample data, sort it in ascending order, and then plot them versus quantiles calculated from a theoretical distribution known as the standard normal distribution with mean 0 and standard deviation 1. If both sets of quantiles come from the same distribution, we should see the points converge to the straight line. As long as the blue points in Fig.7 are close to the red line, the normality can be assumed, and we have stability in the model error.

5 Conclusion

Due to the lack of research on this UCI dataset, our objective was to build a forecast system to make some improvement comparing to existing studies specifically on a daily level. Experiments with RMSE statistical criteria, clearly demonstrate that VMD-based Neural Network model significantly achieved the lowest forecasting error among models. This indicates that this novel approach can be used as a very promising methodology specifically for non-stationary and noisy time series. The VMD is considered as a

new adaptive multiresolution technique, and this is the main advantage of adopting this approach.

Finally, a comparative study of accuracy of the VMD combined with other machine learning models such as support vector machines could be considered for future works to also examine its effectiveness.

6 References

1. T.M. Usha and S. Appavu alias Balamurugan, "Computational modeling of electricity consumption using econometric variables based on neural network training algorithms," 2016
2. L. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Lloret and Joaquim Massana, "A Survey on Electric Power Demand Forecasting: Future Trends in Smart Grids, Microgrids and Smart Buildings," IEEE Communications Surveys & Tutorials, pp. 14601495, Third Quarter 2014.
3. P.Chujai and N. Kerdprasop "Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models," 2013
4. Q. Zhu, Y. Guo and G. Feng, "Household energy consumption in China forecasting with BVAR model up to 2015," 2012
5. Volkan Ş. Ediger, Sertaç Akar, "ARIMA forecasting of primary energy demand by fuel in Turkey," Energy Policy, vol.35, 2007, pp.1701-1708.
6. Javier Contreras, Rosario Espinola, Francisco J. Nogales, and Antonio J. Conejo,"ARIMA models to predict next-day electricity prices. Power Systems," IEEE Transactions on 2003, vol.18, no. 3, pp. 1014-1020.
7. Antonio J. Conejo, Miguel A. Plazas, Rosa Espinola, and Ana B.Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," IEEE Trans. Power Syst., vol. 20, no. 2, 2005, pp. 1035-1042.
8. G. Box, G. Jenkins, "Time Series Analysis: Forecasting and Control," Holden-Day, San Francisco, 1970.
9. S. Lahmiri, "Interest rate next-day variation prediction based on hybrid feedforward neural network, particle swarm optimization, and multiresolution techniques," 2016
10. Q. Yi Feng, R. Vasile, M. Segond, A. Gozolchiani, Y. Wang, M. Abel, S. Havlin, A. Bunde, and H. A. Dijkstra, "ClimateLearn: A machine-learning approach for climate prediction using network measures," 2016
11. Bishop, C. M. "Pattern recognition and machine learning," Springer, New York, 2006.
12. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation. In parallel distributed processing: Exploration in the microstructure of cognition," volume I, Bradford Books, Cambridge, MA, 1986
13. Yann A LeCun, L. Bottou, Genevieve B Orr, K.-R. Müller, "Efficient backprop," 2012
14. S. Liu, G. Tang, X. Wang, and Y. He, " Time-Frequency Analysis Based on Improved Variational Mode Decomposition and Teager Energy Operator for Rotor System Fault Diagnosis," 2016
15. A. N. Tichonov, "Solution of incorrectly formulated problems and the regularization method,"Soviet Mathematics, vol. 4, pp. 1035–1038, 1963.
16. K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," IEEE Transactions on Signal Processing, vol.62, no.3, pp. 531–544, 2014.
17. https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption