Monthly electricity demand forecasting using empirical mode decomposition-based state space model

Article *in* Energy & Environment · April 2019

DOI: 10.1177/0958305X19842061

5 authors, including:

Jing Ma
Sichuan University

Liangwei Yang
University of Illinois at Chicago

Liming Yao
Sichuan University

Journal Title
XX(X):1–11

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Monthly Electricity Demand Forecasting using Empirical Mode Decomposition-Based State Space Model

Zhineng Hu[1] Jing Ma [1] Liangwei Yang [2] Liming Yao [1] Meng Pang [4]

Abstract

Guaranteeing stable electricity demand forecasting is paramount for the conservation of material resources. However, because electricity consumption data are often made up of complex and unstable series, it is very hard for a simple single method to always obtain accurate predictions. To improve electricity demand forecasting robustness and accuracy, a hybrid empirical mode decomposition and state space model is proposed, for which the empirical mode decomposition is applied to decompose the total time series (noise filtering), and the state space model is employed to forecast every sub-series (feature extraction), with the state space model parameters being optimized using maximum likelihood via a Kalman filter. Compared with autoregressive integrated moving average model and artificial neural networks, the proposed model had more stable and accurate forecasting. This method could be broadly applied to not only forecast electricity demand, being a key step for developing electricity generation plans and formulating energy policy, but also forecast any similar time series data with noise and substantive latent features, making a new step toward solving such a problem.

Keywords

Electricity demand forecasting; Empirical Mode Decomposition; Hybrid model; State Space Model; Time series analysis

Introduction

As a long-term important hot issue, Electricity demand forecasting research can date back to Heinemann et al. (1966). Hobbs et al. (1999) stated that with a reduction of 1% in the forecasting mean absolute percentage error, 10,000 MW of electricity energy can be saved, which means as power grid planning, investment and transactions are based on accurate electricity demand forecasting, an accurate forecasting model could save up to $1.6 million a year.

Electricity forecast horizons can be short term from an hour or a week, medium-term from a week to a year, or long-term for over a year (Jiang et al. 2017; Khuntia et al. 2016). Medium-term electricity demand forecasting and especially monthly forecasting is a critical index for many business and government decision-making processes (Guo et al. 2018), as it is designed to maintain a balance between supply and demand (Yang A & X 2018).

Due to the complexity and instability of electricity consumption data, no one best individual method has been found to always perform well. However, as ensemble learning and hybrid methods seek to obtain better forecasting performance by strategically combining multiple algorithms (Qiu et al. 2017b), this paper develops a hybrid monthly electricity demand forecasting model based on empirical mode decomposition (EMD) and a state space model (SSM), which will provide a new step for solving time series data with noise and substantive latent features. The EMD as a noise filter is applied to decompose the total time series, and the SSM is employed for the feature extraction and forecasting of each every sub-series; with the SSM being optimized using maximum likelihood Kalman filter.

The remainder of this paper is organized as follows. Section gives a brief literature review, Section introduces the proposed method, Section reviews the data and forecasting results, and Section concludes the paper.

Literature review

Modeling and forecasting electricity demand is an indispensable part of demand side management (Shao et al. 2017). Even though researchers have been developing and enhancing many electricity demand forecast models, they have been modeled around two main forecasting methods, individual forecasting and hybrid forecasting.

Individual forecasting models such as time series analysis and machine learning models have been the most widely used to date. Time series analysis seeks to extract meaningful statistics and other characteristics from the time series data by analyzing the data itself (Qiu et al. 2017a), and therefore time series models have been widely used for electricity

[1]Business School, Sichuan University, Chengdu, 610064, P. R. China
[2]Big Data Research Center, University of Electronic Science and Technology of China, Chengdu, 611731, P. R. China
[3] Sichuan Regulatory Commission, State Administration of Energy of China, Chengdu 641000,P. R. China

Corresponding author:
Liming Yao, Business School, Sichuan University, Chengdu, 610064, P. R. China
Email: lmyao@scu.edu.cn

load forecasting using linear regression (Papalexopoulos & Hesterberg 1990), seasonal autoregressive (Mbamalu & El-Hawary 1993), ARIMA (Chen et al. 1995), threshold autoregressive (Huang 1997), Kalman filtering (Lynch et al. 2016), and seasonal autoregressive integrated moving average methods (SARIMA) (Tarsitano & Amerise 2017). As it does not account for the exogenous variables, time series analysis models only require the consumption data to be complete for the forecasting. Machine learning models, such as artificial neural networks (Kouhi & Keynia 2013), support vector machines (Hong 2010), Gaussian process (Van der Meer et al. 2018), grey forecasting model (Wang et al. 2018b), Markov model (Xie et al. 2015) and ensemble learning methods (Burger & Moura 2015) have also been employed to forecast electricity demand. Although machine learning models have proven to be very accurate, these models are unsuitable for analyzing electricity demand as they are unable to provide any insights into the causes of structural changes (Takeda et al. 2016).

However, because of the many factors that affect electricity demand as well as the non-stationary nature of electricity consumption data, using only one individual model may not produce good results (Che et al. 2012). To address this challenge, highly accurate forecasting models and hybrid technologies have been considered to improve model performance. Hybrid models, which are an integration of different models, can improve the forecasting accuracy. When models are combined, they can capture the different electricity consumption features and overcome the defects in the individual models(Zhang et al. 2018). Hybrid models also have better performance than individual models because of bias-variance decomposition and strength-correlation (Qiu et al. 2017a). Therefore, hybrid models have generally proven to be superior to single models. Hybrid electricity load forecasting models can be classified into three main categories.

(1) In the first category models, the electricity demand is forecast separately by the different models, after which the weight of each model is calculated using a suitable method. The final forecasting value can is determined by adding each model's forecasting value multiplied by its weight. Nowotarski et al. (2016) investigate the performance of combining so-called sister load forecasts which share similar model structure but are built based on different variable selection processes. Besides, Xiao et al. (2017) used a multi-objective flower pollination algorithm (MOPFA) to optimize the weights of single models that included several artificial neural networks.

(2) The second model category supposes that there is a linear part and a nonlinear part. The data are forecast using the linear model and the residuals checked for a nonlinear pattern, after which the residuals are forecast using a nonlinear model; and vise versa. The model's final forecast is determined from the sum of the linear and nonlinear model results. Barak & Sadegh (2016) combined ARIMA and an Adaptive Neuro Fuzzy Inference System (ANFIS) and found that there were good in linear and nonlinear structural performances. Moreover, Wang et al. (2018a,c) uses the linear ARIMA to correct nonlinear metabolic grey model forecasting residuals to improve forecasting accuracy.

(3) The third category model decomposes the electricity consumption data into several sub-series, after which each sub-series is forecast using a suitable model. The final forecast results are the sum of each of the sub-series' forecasting results. Popular decomposition technologies are empirical mode decomposition, wavelet transforms, and Fourier transformation. Li et al. (2016) used wavelet transformation to decompose original electricity consumption data into several sub-series, after which each sub-series was forecast by an extreme learning machine combined with partial least squares regression, and Zhang et al. (2018) proposed a hybrid model based on improved empirical mode decomposition (IEMD), autoregressive integrated moving average (ARIMA), and a wavelet neural network (WNN) that had been optimized using a fruit fly optimization algorithm (FOA).

While EMD, Fourier, and wavelets are all used to decompose signals, EMD is fundamentally different from the other two. EMD makes no assumptions a priori about the composition of the signal. Rather, it uses spline interpolation between maxima and minima to successively trace out "Intrinsic Mode Functions". since it makes no assumptions about signal, the results might be more meaningful. Also, since the IMFs can change over time, EMD makes no assumptions about the stationarity of the signal (or the signal components) and is therefore better suited to non-linear signals than either Fourier or Wavelets. This makes EMD particularly attractive when analyzing signals from complex systems; for instance in electricity demand analysis. To increase the efficiency of decomposition, EMD was used by many researchers (Rios & Mello 2013; Fan et al. 2016; Liu et al. 2014). (Qiu et al. 2017a) used EMD to decompose the electricity load into some detail parts and an approximate part. Then the deep belief network (DBN) model was used for forecasting. Xiong et al. (2014) used EMD to increase decomposition efficiency and introduced an EMD-based support vector regression modeling framework.

Unlike most studies, this paper focused on the basic characteristics of electricity consumption data; such as trend, season, cycle, and random effect. In addition, every IMF and residue component are forecast by SSM next because a simpler model can make less cost. Therefore, a simpler calculation and a higher accuracy forecasting ensemble are constructed to forecast electricity demand. In addition, one of the advantages of SSMs is that the individually created models (submodels or components) can be easily incorporated into a single model (Takeda et al. 2016). Consequently, SSMs were developed to assess national electricity demand in France (Dordonnat et al. 2008). Therefore, to capture the different characteristics and to overcome the shortcomings of a single model, this paper presents a hybrid combined EMD and SSM model to forecast monthly electricity demand, which has the following advantages.

(1) To more accurately and effectively extract the electricity demand characteristics, the EMD is used to reduce information loss, which has rarely been discussed in electricity demand decomposition research.

(2) Based on its specific features and fit characteristics in the original electricity consumption, each sub-series

decomposed from the original electricity consumption is forecast using a suitable SSM.

Methodology

Because of the specific monthly electricity consumption data characteristics, this section presents the key components for the proposed forecasting method (EMD-SSM). The EMD extraction process first deals with the noise filtering, after which the SSM model combined with the EMD interpolates and extrapolates the latent components such as the trends, seasonal effects and irregularities.

Noise filtering (EMD)

Known as the Hilbert-Huang transformation (HHT) (Huang et al. 1998), EMD, an empirical approach to obtaining instantaneous frequency data from non-stationary and nonlinear data sets, decomposes a signal into several intrinsic mode functions (IMF) and a residue.

As it is influenced by several factors, electricity consumption is a random, non-stationary process composed of thousands of individual components, with the data having trend, season, cycle, and random effect characteristics; therefore, the EMD algorithm can be very effective for noise filtering.

An IMF is a function that has only one extreme between zero crossings and a mean value of zero. The shifting process EMD uses to decompose the signal into IMFs is described in the following:

(1) For a time series $x(t)$, let m_1 be the mean of its upper and lower envelopes as determined by a cubic-spline interpolation of the local maxima and minima.

(2) The first component h_1 is computed by subtracting the mean from the original time series: $h_1 = x(t) - m_1$.

(3) In the second shift process, h_1 is treated as the data, and m_{11} is the mean of h_1's upper and lower envelopes: $h_{11} = h_1 - m_{11}$.

(4) This shifting procedure is repeated k times until one of the following stop criterion is satisfied: i) m_{1k} approaches zero, ii) the number of zero-crossings and extrema for h_{1k} differs at most by one, or iii) the predefined maximum iteration is reached. h_{1k} can be treated as an IMF in this case and computed by: $h_{1k} = h_{1(k-1)} - m_{1k}$.

(5) Then it is designated $c_1 = h_{1k}$, the first IMF component from the data, which contains the shortest signal period component, which is then separated from the rest of the data: $x(t) - c_1 = r_1$. This procedure is repeated for r_j: $r_1 - c_2 = r_2, \ldots, r_{n-1} - c_n = r_n$.

As a result, the original time series signal is decomposed into a set of functions: $x(t) = \sum_{i=1}^{n}(c_i) + r_n$, where the number of functions n in the set depends on the original signal.

Feature extraction (SSM)

SSM, also known as Kalman filtering, was first introduced by Kalman (Kalman 1960). State-space models consist of two equations to simultaneously account for two variation distinct sources: the state equation (latent process) deals with the process uncertainty caused by the unobserved factors, and the observation model incorporates the effect of the error caused by the mismeasurement of outcomes. The general form for a state-space model is as follows.

$$\begin{cases} \text{Observation equation:} & \boldsymbol{Y_t} = \boldsymbol{Z_t}\boldsymbol{\alpha_t} + \boldsymbol{\epsilon_t}, \\ \text{State transition equation:} & \boldsymbol{\alpha_{t+1}} = \boldsymbol{T_t}\boldsymbol{\alpha_t} + \boldsymbol{\eta_{t+1}}, \end{cases} \tag{1}$$

where $n * 1$ vector $\boldsymbol{\alpha_t}$ is the state vector, denoting the state at time t; and $k * 1$ vector $\boldsymbol{y_t}$ is the observations at time t; and $\boldsymbol{\epsilon_t}$ and $\boldsymbol{\eta_t}$ are the observation and state noises, which are respectively drawn from Gaussian distributions with zero means, variances of $\boldsymbol{R_t}$ and $\boldsymbol{Q_t}$, and $\boldsymbol{S_t}$ covariances. The $n * n$ matrix $\boldsymbol{T_t}$ called the transition matrix and the $k * n$ matrix $\boldsymbol{Z_t}$ called the output matrix in Eq. (1) are referred to as the system matrices with appropriate dimensions.

It is well known in the time series analysis that unobserved components models has been proven to be very effective (Koopman & Ooms 2006; Golpe et al. 2012; Mnif 2017). And a very successful and widely accepted unobserved components model comprises trend, cycles, seasonal and irregular components (Pedregal & Young 2006). Hereafter, the SSM model is employed to forecast each sub-series, the general form for which is:

$$\boldsymbol{y_t} = \mu_t + \gamma_t + c_t + \varepsilon_t,$$

where $\boldsymbol{y_t}$ is the observation vector at time t, μ_t is the trend component, γ_t is the seasonal component, c_t is the cycle component, and ε_t is an irregular component. The modeling details for these components are described in the following.

The trend component is a dynamic regression model extension that includes an intercept and linear time-trend and is given as::

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_{t-1}, \\ \beta_t &= \beta_{t-1} + \zeta_{t-1}, \end{aligned}$$

where the level is a generalization of an intercept term that can dynamically vary across time, and the trend is a generalization of the time-trend so that the slope can dynamically vary across time; $\eta_t \sim N(0, \sigma_\eta^2)$ and $\zeta_t \sim N(0, \sigma_\zeta^2)$.

The seasonal component is modeled as:

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t+1-j} + \omega_t,$$

where s is the number of seasons (12 in monthly data), and $\omega_t \sim N(0, \sigma_\omega^2)$. This component results in one parameter to be selected via maximum likelihood: σ_ω^2 and one parameter to be chosen, the number of seasons s.

The cyclical component is intended to capture the cyclical effects at time frames much longer than those captured by a seasonal component.

$$\begin{aligned} c_{t+1} &= \rho_c(\tilde{c}_t \cos\lambda_c t + \tilde{c}_t^* \sin\lambda_c) + \tilde{\omega}_t, \\ c_{t+1}^* &= \rho_c(-\tilde{c}_t \sin\lambda_c t + \tilde{c}_t^* \cos\lambda_c) + \tilde{\omega}_t^*, \end{aligned}$$

where c_t and c_t^* are independent, zero-mean, Gaussian disturbances with variance $\sigma_\omega^2, \omega_t, \tilde{\omega}_t$ iid $N(0, \sigma_{\tilde{\omega}}^2)$, damping factor ρ_c can be any value in the interval (0, 1), including one but excluding zero and parameter λ_c (the frequency of the cycle) is an additional parameter to be estimated using maximum likelihood estimation (MLE).

An autoregressive component (that is often used as a replacement for the white noise irregular term):

$$\varepsilon_t = \rho(L)\varepsilon_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2).$$

In summary, the state space model can be written as follows:

$$\begin{aligned}
\text{Observation equation: } y_t &= \mu_t + \gamma_t + c_t + \varepsilon_t, \\
\text{Transition equation: } \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_{t-1}, \\
\beta_t &= \beta_{t-1} + \zeta_{t-1}, \\
\gamma_t &= -\sum_{j=1}^{s-1} \gamma_{t+1-j} + \omega_t, \\
c_{t+1} &= \rho_c(\tilde{c}_t \cos \lambda_c t + \tilde{c}_t^* \sin \lambda_c) + \tilde{\omega}_t, \\
c_{t+1}^* &= \rho_c(-\tilde{c}_t \sin \lambda_c t + \tilde{c}_t^* \cos \lambda_c) + \tilde{\omega}_t^*, \\
\varepsilon_t &= \rho(L)\varepsilon_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2).
\end{aligned}$$

EMD-based SSM model

The proposed hybrid EMD and SSM electricity demand forecasting model decomposes the electricity consumption data into several IMFs and one residue using the EMD method, after which the SSM model is applied to each IMF and the residue. When the forecasting for all sub-series is the complete, the forecasting results are aggregated by simple addition to obtain the final forecasting. Fig. 1 shows the overall schematic for this ensemble method:

(1) The time series data is decomposed using EMD into several IMFs and one residue.

(2) For each IMF and the residue, one training data is constructed as the SSM input.

(3) The SSM is then trained to obtain the forecast results for each of the extracted IMFs and the residue.

(4) All forecast results are added to formulate the ensemble output for the time series.

Error Measurement

As suggested by Taylor (2010) and Shaikh & Ji (2016), forecast model performance can be evaluated using the absolute relative error (ARE), the root-mean-square error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE), each of which is defined as follows.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|, \quad \text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2},$$

$$\text{ARE} = \left|\frac{\hat{y}_i - y_i}{y_i}\right|, \quad \text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|,$$

where $\hat{y}_i$ is the forecast value of the corresponding y_i, and n is the number of data points in the testing time series.

Results and Discussion

Data handling

This study was based on electricity consumption data from January 2007 to October 2010 in Sichuan Province, China. The data sets were generated by the Electricity Saving Association, which is the major source of power data in the province and covers almost all sectors of the economy. Total electricity consumption was composed of residential and nonresidential (industrial, transportation, manufacturing, and information service) electricity demand consumption. To make the results more convincing, all electricity consumption variables; that is, the total electricity consumption and its components were forecast. For a fair comparison analysis, each consecutive three months in 2010 were used as test set and the 36 months before the test month was used as training set. The specific data divisions are shown in Table 1.

Table 1. Data division

Training set	Test set
Training set 1: 2007.01-2009.12	Test set 1: 2010.01-03
Training set 2: 2007.02-2010.01	Test set 1: 2010.02-04
Training set 3: 2007.03-2010.02	Test set 1: 2010.03-05
Training set 4: 2007.04-2009.03	Test set 1: 2010.01-06
Training set 5: 2007.05-2009.04	Test set 1: 2010.01-07
Training set 6: 2007.06-2009.05	Test set 1: 2010.01-08
Training set 7: 2007.07-2009.06	Test set 1: 2010.01-09
Training set 8: 2007.08-2009.07	Test set 1: 2010.01-10
Training set 9: 2007.09-2009.08	Test set 1: 2010.09
Training set 10: 2007.10-2009.09	Test set 1: 2010.10

Model adaptability

Fig. 2 shows that mixed with various of signals (trend, season and cycle), the original data seems to be volatile and unstable. EMD first is used to decomposed original electricity consumption for noisy filtering. Fig. 3 shows the decomposition results by EMD for training set 1. After decomposition, every subseries is stable with different feathers; for example, cycle in 3decomposition and 4decomposition, trend in 5decomposition. SSM then is used to fit every subseries for feature extraction. Finally, the forecast model could be obtained by adding all sub-models. The adaptability of the proposed method for total electricity consumption is shown in Fig. 4 and Table 2. The adopted modelling approaches were found to be a good fit for the training sets. Fig. 4 shows the results of EMD-SSM for the three-period-ahead total demand forecasting for training sets $1 \sim 4$. Satisfactory results were obtained using the proposed method as the shape of the forecast curves were very similar to the shape of the actual curves. Starting at zero, the fitting curves were poor for the first few points; therefore, all measurements (in Table 2) were calculated using the remaining 26 consumption data. Table 2 shows the results for the three error measurements for total electricity consumption in the training sets. As all MAPEs were lower than 0.03 and the MAEs and RMSEs were also low, the proposed method was in a good fit for the total power consumption in the training sets.

To verify the adaptability of EMD-SSM, one-period-ahead forecasting was conducted for different electricity variables using ten test sets (Table 3). As shown, nearly 100% of the ARE outputs for the total power demand forecasting were lower than 0.1 and 70% were lower than 0.05. For the other electricity variables, most ARE outputs were lower than 5% or around 5% and only several were higher than 0.1, all

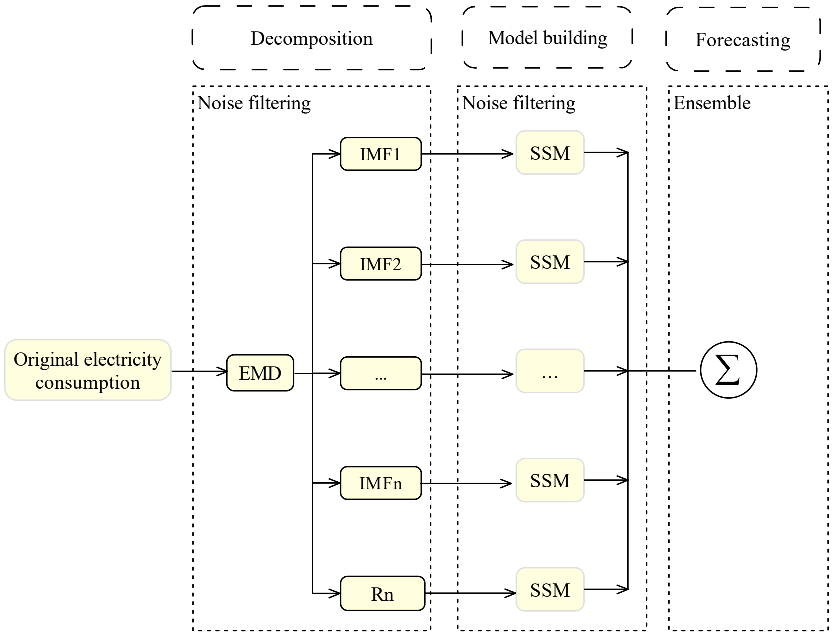

Figure 1. Schematic diagram of the proposed EMD-SSM method

Table 3. One-period-ahead AREs of EMD-SSM for each month in 2010 (for testing sets)

Electricity variable	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.
Total	0.015	0.037	0.014	0.096	0.040	0.105	0.008	0.048	0.046	0.052
Nonresidential	0.037	0.066	0.024	0.041	0.035	0.125	0.057	0.030	0.026	0.006
Residential	0.030	0.030	0.108	0.044	0.078	0.011	0.100	0.109	0.047	0.144
Industry	0.066	0.072	0.083	0.077	0.011	0.062	0.100	0.007	0.039	0.018
Transportation	0.013	0.018	0.049	0.076	0.019	0.017	0.004	0.034	0.045	0.051
Manufacturing	0.147	0.123	0.033	0.035	0.028	0.097	0.018	0.016	0.091	0.093
Information service	0.022	0.081	0.012	0.078	0.023	0.065	0.165	0.100	0.101	0.009

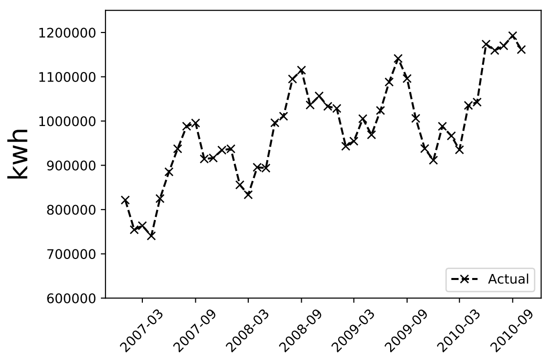

Figure 2. Total electricity consumption from 2007.01 to 2010.12

Table 2. Adaptability of the proposed method for total power consumption (for training sets)

Training set	MAPE	MAE	RMSE
Training set 1	0.025559574	24435.23271	32185.59497
Training set 2	0.026291536	25172.88265	30662.9301
Training set 3	0.023462413	22691.31679	27795.67893
Training set 4	0.028083031	26454.65447	42586.04947
Training set 5	0.021264841	20716.21956	25108.66479
Training set 6	0.023301599	23450.06357	29007.42461
Training set 7	0.019287419	19597.95449	25375.25747
Training set 8	0.019326223	19681.43743	23459.46975
Training set 9	0.021036301	22204.09311	31201.56442
Training set 10	0.019359572	20131.52036	23643.17369

of which indicated that the EMD-SSM was able to achieve precise and stable forecasting.

The other statistical results in Table 4 further illustrate the performance of the proposed model. The proposed model performed well for the different electricity variables as there were low average values and ARE variances. The maximum AREs were lower than 0.1 for the total, industry, and transportation power consumption, indicating satisfactory results. However, the results were unsatisfactory for residential, manufacturing and information service electricity demand forecasting as the maximum ARE were respectively 0.144, 0.147 and 0.165. These results can be explained by the amount of training data available for each electricity variable; that is, there was a large amount of training data for the total, industry, and transportation power consumption, but significantly less for the residential, manufacturing and information service power consumption. Despite these results, according to Williams (2013), actual power generation must be able to accommodate up to a 15% increase in extra demand at any time, which means that even in the worst situation, the proposed method is in the highly accurate and stable forecast range.

Comparison analysis with benchmark methods

Well-known and accurate, benchmark methods are very useful when seeking to understand relative accuracy and can assist in identifying the weaker areas in proposed

Table 4. Statistical ARE values for the different electricity variables (one-period-ahead)

Electricity variable	Average value	Variance	Minimum	Maximum
Total	0.046	0.0010	0.008	0.105
Nonresidential	0.044	0.0010	0.001	0.125
Residential	0.070	0.0018	0.011	0.144
Industry	0.054	0.0011	0.007	0.100
Transportation	0.032	0.0005	0.004	0.076
Manufacturing	0.068	0.0022	0.016	0.147
Information service	0.066	0.0025	0.009	0.165

models (Takeda et al. 2016). Over the years, various forecasting models have been developed in literature, of which the ARIMA and ANN are widely popular. ARIMA models are well-known for their notable forecasting accuracy and flexibility in representing several different types of time series (Khandelwal et al. 2015). Meanwhile, ANN is proved to be suitful for complex nonlinear time series modeling (Gnay 2016). To further substantiate EMD-SSM the forecasting results, comparison analyses with three benchmark methods; ARIMA, ANN-1 (with 1 hidden layer) and ANN-2 (with 2 hidden layer); were conducted.

The basic steps for the ARIMA are as follows: first, the nonstationary series is transformed to a stationary series using differencing, after which p (the order of the autoregressive part) and q (the order of the moving-average process) are determined using using ACF and PACF, and the optimal model determined by checking for white noise residuals. The ANN models use the six month electricity consumption prior to the three months to be forecast as the input data and has six input neurons, twelve hidden neurons for each hidden layer, and three output neurons. A rectified linear unit (ReLU) activation function is used as the hidden layer activation function, a linear function is used in the output layer, and adaptive moment estimation (Adam) is used as the learning algorithm.

To verify the effectiveness of the proposed model, three electricity demand periods were forecast. Every three months in 2010 were used as the test set and the 36 months before the test set was used as the training set. Table 5 presents the AREs for every month for the test set. As can be seen, the same training set is used in the same column of Table 5. Table 6 shows the other evaluation criteria for the three months for every test set using the corresponding training set.

As can be seen in Table 5, the forecast accuracy of the proposed model was higher than the comparison models. The ARE for the EMD-SSM was in the range 0.0014–0.1428 which was the narrowest of the four methods. Further, no matter what the forecast period, the ARE averages for the proposed method were the lowest (0.455, 0.0608 and 0.0608, respectively). In addition, as shown in Table 6, the EMD-SSM achieved the lowest MAPE, MAE and RMSE for most of the test sets, which indicated that the hybrid model had the best stability. Both Table 5 and Table 6 demonstrate that the proposed model had more accurate and stable results compared to the other models.

To further illustrate the effectiveness of the EMD-SSM, Table 7 summarizes the number of times the four methods ranked 1, 2, 3 or 4. The EMD-SSM was ranked 1 three times, ranked 2 four times, ranked 3 once and was never ranked 4, again demonstrating the superiority of the proposed model.

Table 7. Number of times the three methods ranked 1, 2, 3 or 4

Rank	ARIMA	ANN-1	ANN-2	EMD-SSM
1	2	2	1	3
2	1	2	1	4
3	4	2	1	1
4	1	2	4	0

The comparison of the descriptive statistical values for the different evaluation criterions found that the average values for the MAPE, MAE and RMSE obtained using the proposed method were lower than the comparative models (Table 8). The proposed model accuracy was able to attain an operational level and outperformed the benchmark methods. The average MAPE values for the ANN-1 (0.0938), ANN-2 (0.1099) and ARIMA (0.0602) were improved by 40.62%, 49.32% and 7.48% using the proposed model (0.0557). Further, the MAPE, MPE, and RMSE variances obtained using the proposed model were all lower than the other models, further indicating that the proposed hybrid model stability was superior.

To verify the above, boxplots were drawn to show the MAPE distribution and are presented in Fig. 5. Even though there were no significant differences between the benchmarks and the proposed model, the boxplots show that the proposed method performed better than the ARIMA and the ANNs because it achieved the lowest Q_{25} ($25th$ percential), median, and Q_{75} ($75th$ percential) and had the narrowest inner fence. Table 8 and Fig. 5 demonstrate that the proposed model was able to provide more accurate and more stable results than the benchmark models.

To present the advantages of the proposed model with a deep analysis, Table 6 shows that the ANN-2 achieved the highest MAPE, MAE and RMSE for most of the test sets. Moreover, ANN-2 was ranked four times in the four models as shown in Table 7. Also, Fig. 5 shows that the ANN-2 obtained the stable but very bad results because of the higher MAPE means, median and the minimum. These all illustrate that the ANN-2 is not suitable for this data set, which may be caused by overfitting or local minimum. Meanwhile, all results show that ANN-1 and ARIMA are also disappointing, compared with the EMD-SSM. In addition, a further analysis using original data indicated that when data set size used to train the simple ANN was altered, there was an obvious impact on its generalization ability (Moyo & Sibanda 2015; Foody et al. 1995). Therefore, even though there was a relatively small 36-month training set for each forecasting process, the proposed model still performed well, which indicated that the EMD-SSM could perform well on smaller data sets.

Table 5. AREs for the different methods for each month in 2010 (for test sets)

One-period-ahead	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Average
ARIMA	0.0816	0.0438	0.0346	0.0983	0.0207	0.1073	0.0273	0.0056	0.0524
ANN-1	0.0353	0.0235	0.0555	0.1595	0.0136	0.2705	0.0809	0.0886	0.0909
ANN-2	0.0416	0.0287	0.0480	0.0767	0.0508	0.0203	0.0961	0.1000	0.0578
EMD-SSM	0.0147	0.0368	0.0140	0.0963	0.0398	0.1054	0.0082	0.0484	0.0455
Two-period-ahead	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Average
ARIMA	0.0606	0.0895	0.0611	0.1004	0.0946	0.0924	0.0250	0.0205	0.0680
ANN-1	0.0075	0.0031	0.1612	0.0487	0.0025	0.2723	0.0849	0.1798	0.0950
ANN-2	0.2358	0.1094	0.0846	0.0513	0.0225	0.0825	0.1550	0.1353	0.1096
EMD-SSM	0.0014	0.0542	0.0638	0.0970	0.0870	0.1111	0.0019	0.0700	0.0608
Three-period-ahead	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Average
ARIMA	0.0274	0.0086	0.0626	0.1959	0.0777	0.0966	0.0116	0.0101	0.0613
ANN-1	0.0284	0.0551	0.0272	0.0567	0.0379	0.2292	0.1592	0.1704	0.0955
ANN-2	0.0129	0.1834	0.0906	0.1866	0.1944	0.1630	0.2654	0.2036	0.1625
EMD-SSM	0.0488	0.0302	0.0778	0.1428	0.0326	0.0738	0.0687	0.0121	0.0608

Table 6. Different evaluation criterion for the different models (for test sets)

test set 1				test set 2			
method	MAPE	MAE	RMSE	method	MAPE	MAE	RMSE
ARIMA	0.0565	54935	59398	ARIMA	0.0473	44982	54401
ANN-1	0.0237	22900	25661	ANN-1	0.0272	27561	35497
ANN-2	0.0968	93681	133884	ANN-2	0.1072	106668	125576
EMD-SSM	0.0216	20501	27642	EMD-SSM	0.0404	39196	40071
test set 3				test set 4			
method	MAPE	MAE	RMSE	method	MAPE	MAE	RMSE
ARIMA	0.0528	53638	55713	ARIMA	0.1316	145497	157277
ANN-1	0.0813	82382	102232	ANN-1	0.0883	94165	106898
ANN-2	0.0744	75658	78767	ANN-2	0.1049	117320	138024
EMD-SSM	0.0519	53426	60868	EMD-SSM	0.1120	122830	126851
test set 5				test set 6			
method	MAPE	MAE	RMSE	method	MAPE	MAE	RMSE
ARIMA	0.0611	74228	83492	ARIMA	0.0988	115382	115649
ANN-1	0.0180	20319	26696	ANN-1	0.2573	300427	301295
ANN-2	0.0893	101616	134572	ANN-2	0.0886	103432	123977
EMD-SSM	0.0532	60491	67303	EMD-SSM	0.0968	112949	114528
test set 7				test set 8			
method	MAPE	MAE	RMSE	method	MAPE	MAE	RMSE
ARIMA	0.0213	24913	26133	ARIMA	0.0121	14239	16091
ANN-1	0.1083	127677	135072	ANN-1	0.1463	172002	178783
ANN-2	0.1722	203116	220248	ANN-2	0.1463	171641	178588
EMD-SSM	0.0263	31241	47661	EMD-SSM	0.0435	51383	58809

Table 8. Descriptive statistics for the different evaluation criteria for the different methods

Method	Criterion	Average value	Variance	Minimum	Maximum
ARIMA	MAPE	0.0602	0.0015	0.0121	0.1316
	MAE	65976	44578	14238	145497
	RMSE	71019	46698	16091	157277
ANN-1	MAPE	0.0938	0.0064	0.0179	0.2573
	MAE	105929	95413	20319	300427
	RMSE	114017	93730	25661	301295
ANN-2	MAPE	0.1099	0.001	0.0744	0.1722
	MAE	121641	1858543370	75658	203116
	RMSE	141704	1741510642	78767	220248
EMD-SSM	MAPE	0.0557	0.0010	0.0216	0.1120
	MAE	61502	37170	20501	122830
	RMSE	67967	35024	27642	126851

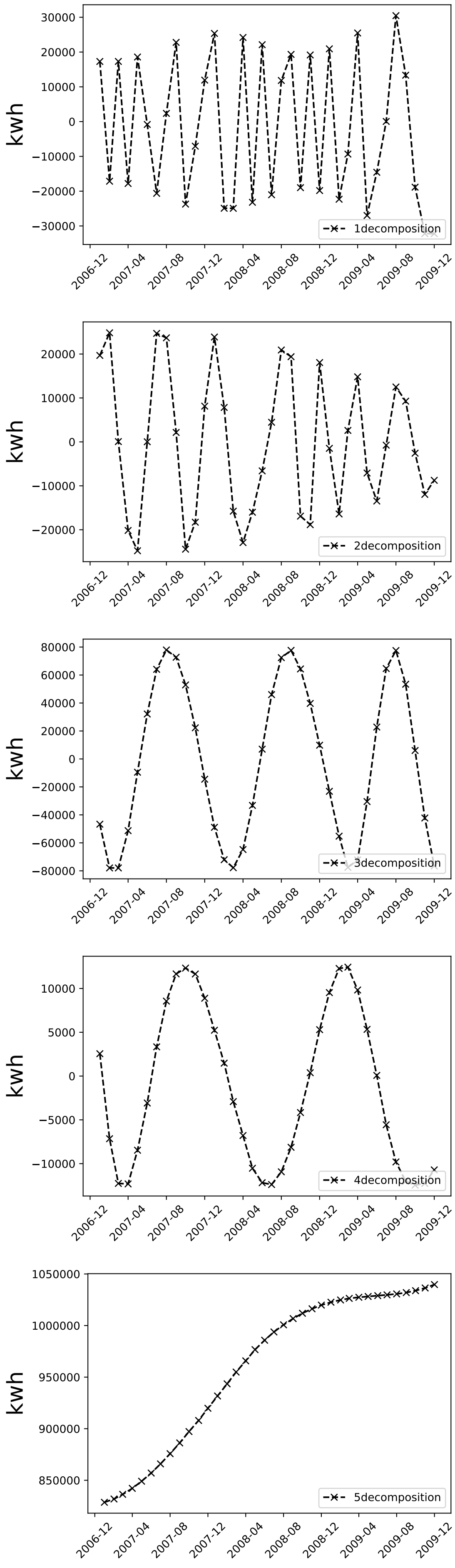

Figure 3. The decomposition results by EMD for training set 1

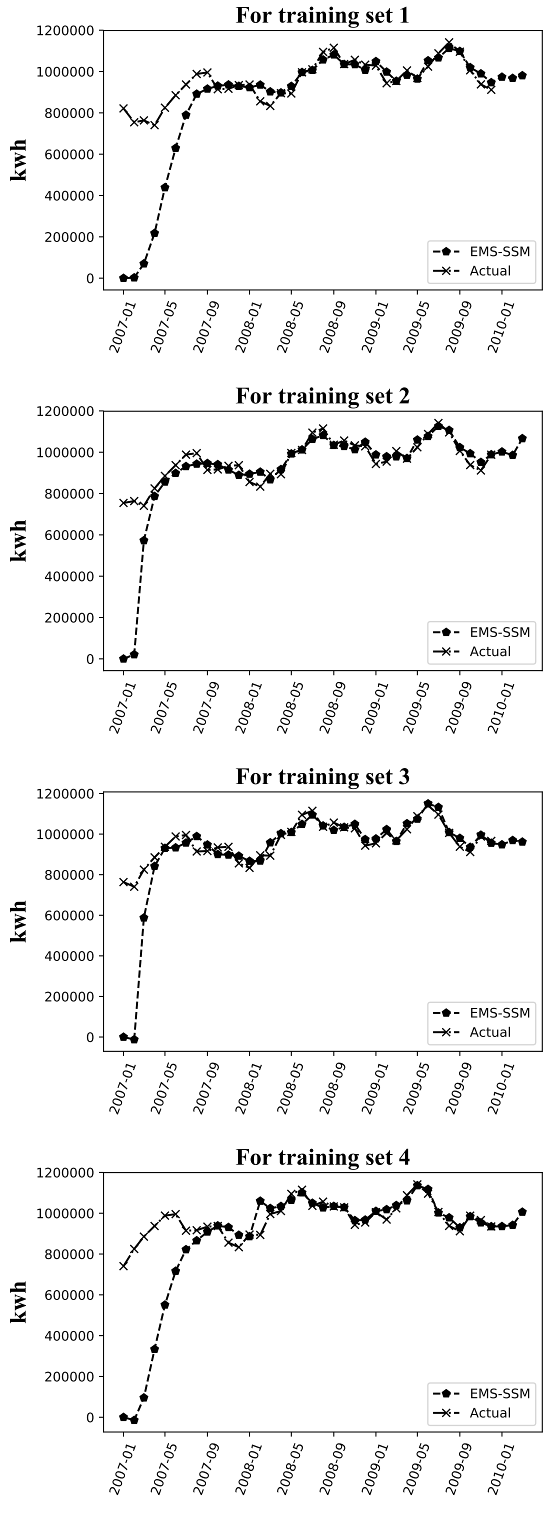

Figure 4. Adaptability of the proposed method

Conclusion

For monthly electricity demand forecasting, a long-term important hot issue, this paper proposed a hybrid model , formulating an empirical mode decomposition-based state space model. (1) Empirical mode decomposition was introduced for noise filtering in the original electricity consumption, from which several IMFs with one residue were obtained. After decomposition, every IMF shows obvious features, such as trends, cycles and seasoning. (2)

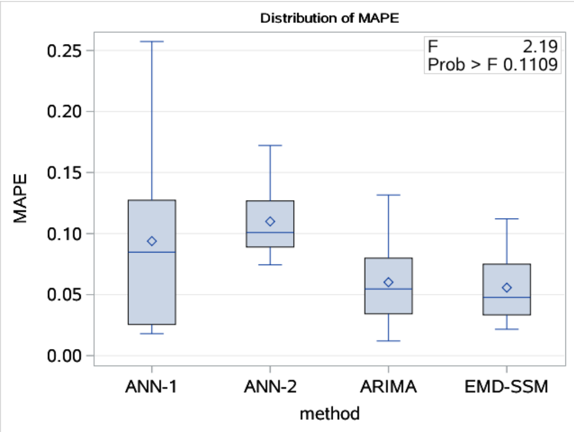

Figure 5. Boxplot for the MAPE distribution

The state space model was used to extract the latent features of each IMFs as well as the residue. (3) The electricity demand forecasting was then obtained through ensemble learning. Electricity consumption data from the Electricity Saving Association of Sichuan Province in China were used to demonstrate the performance of the proposed model, with all results indicating that the proposed model was able to improve electricity demand forecasting accuracy compared to the ARIMA and ANN models.

Based on the common forecast model performance evaluation together with statistical comparison analysis, four facts clearly emerged from the results: (1) EMD can effectively reduce original electricity consumption data noise; (2) the state space model can flexibly extract the time series features and fit well with the decomposed components; (3) by harnessing the advantages of both the EMD and SSM, the hybrid model is able to capture the different features associated with electricity demand; and (4) the EMD-SSM also performs well on small data sets. The EMD-SSM method was firstly proposed by combining the advantages of EMD and SSM, and the results indicated that the ensemble-learning method is valid and achieves good performance.

Obviously, although this paper was focused on the forecasting of monthly series data sets, the methods shown here could also be applied to quarterly or yearly data. Basically, this paper provided a new step for any time series data with noise and substantive latent features, within electricity demand forecasting or in other similar forecast fields. And the approach applied in this work can be implemented for other provinces or countries to make accurate predictions for the future. The results stated that the proposed methods, dealing with noise filtering and feature extracting, have the potential to generate growth in an important new forecast sub-area that may be small now but is expected to grow. For example, to further improve the forecast power, different models such as machine learning methods could be used on decomposed series.

Acknowledgements

We thank those who provided constructive comments and feedback to help improve this paper. This work was supported by the National Natural Science Foundation of China [Grant Nos. 71771157, 71301109].

References

Barak, S., & Sadegh, S. S. (2016). Forecasting energy consumption using ensemble arimacanfis hybrid algorithm. *International Journal of Electrical Power & Energy Systems*, *82*, 92–104. doi:https://doi.org/10.1016/j.ijepes.2016.03.012.

Burger, E. M., & Moura, S. J. (2015). Gated ensemble learning method for demand-side electricity load forecasting. *Energy & Buildings*, *109*, 23–34. doi:https://doi.org/10.1016/j.enbuild.2015.10.019.

Che, J., Wang, J., & Wang, G. (2012). An adaptive fuzzy combination model based on self-organizing map and support vector regression for electric load forecasting. *Energy*, *37*, 657–664. doi:https://doi.org/10.1016/j.energy.2011.10.034.

Chen, J., Wang, W., & Huang, C. (1995). Analysis of an adaptive time-series autoregressive moving-average (arma) model for short-term load forecasting. *Electric Power Systems Research*, *34*, 187–196. doi:https://doi.org/10.1016/0378-7796(95)00977-1.

Dordonnat, V., Koopman, S. J., Ooms, M., Dessertaine, A., & Collet, J. (2008). An hourly periodic state space model for modelling french national electricity load. *International Journal of Forecasting*, *24*, 566–587. doi:https://doi.org/10.1016/j.ijforecast.2008.08.010.

Fan, G. F., Peng, L. L., Hong, W. C., & Sun, F. (2016). Electric load forecasting by the svr model with differential empirical mode decomposition and auto regression. *Neurocomputing*, *173*, 958–970. doi:https://doi.org/10.1016/j.neucom.2015.08.051.

Foody, G. M., Mcculloch, M. B., & Yates, W. B. (1995). The effect of training set size and composition on artificial neural network classification. *International Journal of Remote Sensing*, *16*, 1707–1723. doi:https://doi.org/10.1080/01431169508954507.

Gnay, M. E. (2016). Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: Case of turkey. *Energy Policy*, *90*, 92–101. doi:https://doi.org/10.1016/j.enpol.2015.12.019.

Golpe, A. A., Carmona, M., & Congregado, E. (2012). Persistence in natural gas consumption in the us: An unobserved component model. *Energy Policy*, *46*, 594–600. doi:https://doi.org/10.1016/j.enpol.2012.04.021.

Guo, H., Chen, Q., Xia, Q., Kang, C., & Zhang, X. (2018). A monthly electricity consumption forecasting method based on vector error correction model and self-adaptive screening method. *International Journal of Electrical Power & Energy Systems*, *95*, 427–437. doi:https://doi.org/10.1016/j.ijepes.2017.09.011.

Heinemann, G. T., Nordmian, D. A., & Plant, E. C. (1966). The relationship between summer weather and summer loads-a regression analysis. *IEEE Transactions on Power Apparatus and Systems*, *11*, 1144–1154. doi:10.1109/TPAS.1966.291535.

Hobbs, B. F., Jitprapaikulsarn, S., Konda, S., & Chankong, V. (1999). Analysis of the value for unit commitment of improved load forecasts. *IEEE Transactions on Power Systems*, *14*, 1342–1348. doi:10.1109/59.801894.

Hong, W. C. (2010). Application of chaotic ant swarm optimization in electric load forecasting. *Energy Policy*, *38*, 5830–5839. doi:https://doi.org/10.1016/j.enpol.2010.05.033.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N. C., Chi, C. T., & Liu, H. H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings Mathematical Physical & Engineering Sciences*, *454*, 903–995. doi:10.1098/rspa.1998.0193.

Huang, S. (1997). Short-term load forecasting using threshold autoregressive models. *IEE Proceedings-Generation, Transmission and Distribution*, *144*, 477–481. doi:10.1049/ip-gtd:19971144.

Jiang, P., Liu, F., & Song, Y. (2017). A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short-term load forecasting. *Energy*, *119*, 694–709. doi:https://doi.org/10.1016/j.energy.2016.11.034.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, *82*, 35–45. doi:10.1115/1.3662552.

Khandelwal, I., Adhikari, R., & Verma, G. (2015). Time series forecasting using hybrid arima and ann models based on dwt decomposition. *Procedia Computer Science*, *48*, 173–179. doi:https://doi.org/10.1016/j.procs.2015.04.167.

Khuntia, S. R., Rueda, J. L., & van der Meijden, M. A. (2016). Forecasting the load of electrical power systems in mid- and long-term horizons: a review. *Iet Generation Transmission & Distribution*, *10*, 3971–3977. doi:10.1049/iet-gtd.2016.0340.

Koopman, S. J., & Ooms, M. (2006). Forecasting daily time series using periodic unobserved components time series models. *Computational Statistics & Data Analysis*, *51*, 885–903. doi:https://doi.org/10.1016/j.csda.2005.09.009.

Kouhi, S., & Keynia, F. (2013). A new cascade nn based method to short-term load forecast in deregulated electricity market. *Energy Conversion and Management*, *71*, 76–83. doi:https://doi.org/10.1016/j.enconman.2013.03.014.

Li, S., Goel, L., & Wang, P. (2016). An ensemble approach for short-term load forecasting by extreme learning machine. *Applied Energy*, *170*, 22–29. doi:https://doi.org/10.1016/j.apenergy.2016.02.114.

Liu, N., Tang, Q., Zhang, J., Fan, W., & Liu, J. (2014). A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids. *Applied Energy*, *129*, 336–345. doi:https://doi.org/10.1016/j.apenergy.2014.05.023.

Lynch, C., O'Mahony, M. J., & Guinee, R. A. (2016). Electrical load forecasting using an expanded kalman filter bank methodology. *IFAC-PapersOnLine*, *49*, 358–365. doi:https://doi.org/10.1016/j.ifacol.2016.12.069.

Mbamalu, G., & El-Hawary, M. (1993). Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation. *IEEE Transactions on Power Systems*, *8*, 343–348. doi:10.1109/59.221222.

Van der Meer, D., Shepero, M., Svensson, A., Widén, J., & Munkhammar, J. (2018). Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using gaussian processes. *Applied Energy*, *213*, 195–207. doi:https://doi.org/10.1016/j.apenergy.2017.12.104.

Mnif, A. T. (2017). Political uncertainty and behavior of tunisian stock market cycles: Structural unobserved components time series models. *Research in International Business and Finance*, *39*, 206–214. doi:https://doi.org/10.1016/j.ribaf.2016.07.029.

Moyo, V., & Sibanda, K. (2015). Training set size for generalization ability of artificial neural networks in forecasting tcp/ip traffic trends. *International Journal of Computer Science & Application*, *4*, 9–17. doi:10.5120/19885-1902.

Nowotarski, J., Liu, B., Weron, R., & Hong, T. (2016). Improving short term load forecast accuracy via combining sister forecasts. *Energy*, *98*, 40–49. doi:https://doi.org/10.1016/j.energy.2015.12.142.

Papalexopoulos, A. D., & Hesterberg, T. C. (1990). A regression-based approach to short-term system load forecasting. *IEEE Transactions on Power Systems*, *5*, 1535–1547. doi:10.1109/PICA.1989.39025.

Pedregal, D. J., & Young, P. C. (2006). Modulated cycles, an approach to modelling periodic components from rapidly sampled data. *International Journal of Forecasting*, *22*, 181–194. doi:https://doi.org/10.1016/j.ijforecast.2005.03.001.

Qiu, X., Ren, Y., Suganthan, P. N., & Amaratunga, G. A. J. (2017a). Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. *Applied Soft Computing*, *54*, 246–255. doi:https://doi.org/10.1016/j.asoc.2017.01.015.

Qiu, X., Suganthan, P. N., & Amaratunga, G. A. J. (2017b). Short-term electricity price forecasting with empirical mode decomposition based ensemble kernel machines. *Procedia Computer Science*, *108*, 1308–1317. doi:https://doi.org/10.1016/j.procs.2017.05.055.

Rios, R. A., & Mello, R. F. D. (2013). Improving time series modeling by decomposing and analyzing stochastic and deterministic influences. *Signal Processing*, *93*, 3001–3013. doi:https://doi.org/10.1016/j.sigpro.2013.04.017.

Shaikh, F., & Ji, Q. (2016). Forecasting natural gas demand in china: Logistic modelling analysis. *International Journal of Electrical Power & Energy Systems*, *77*, 25–32. doi:https://doi.org/10.1016/j.ijepes.2015.11.013.

Shao, Z., Chao, F., Yang, S. L., & Zhou, K. L. (2017). A review of the decomposition methodology for extracting and identifying the fluctuation characteristics in electricity demand forecasting. *Renewable & Sustainable Energy Reviews*, *75*, 123–136. doi:https://doi.org/10.1016/j.rser.2016.10.056.

Takeda, H., Tamura, Y., & Sato, S. (2016). Using the ensemble kalman filter for electricity load forecasting and analysis. *Energy*, *104*, 184–198. doi:https://doi.org/10.1016/j.energy.2016.03.070.

Tarsitano, A., & Amerise, I. L. (2017). Short-term load forecasting using a two-stage sarimax model. *Energy*, *133*, 108–114. doi:https://doi.org/10.1016/j.energy.2017.05.126.

Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. *European Journal of*

Operational Research, *204*, 139–152. doi:https://doi.org/10.1016/j.ejor.2009.10.003.

Wang, Q., Li, S., & Li, R. (2018a). China's dependency on foreign oil will exceed 80% by 2030: Developing a novel nmgm-arima to forecast china's foreign oil dependence from two dimensions. *Energy*, *163*, 151–167. doi:https://doi.org/10.1016/j.energy.2018.08.127.

Wang, Q., Li, S., & Li, R. (2018b). Forecasting energy demand in china and india: Using single-linear, hybrid-linear, and non-linear time series forecast techniques. *Energy*, *161*, 821–831. doi:https://doi.org/10.1016/j.energy.2018.07.168.

Wang, Q., Song, X., & Li, R. (2018c). A novel hybridization of nonlinear grey model and linear arima residual correction for forecasting us shale oil production. *Energy*, *165*, 1320–1331. doi:https://doi.org/10.1016/j.energy.2018.10.032.

Williams, H. P. (2013). *Model building in mathematical programming*. WILEY.

Xiao, L., Shao, W., Yu, M., Ma, J., & Jin, C. (2017). Research and application of a combined model based on multi-objective optimization for electrical load forecasting. *Energy*, *119*, 1057–1074. doi:https://doi.org/10.1016/j.energy.2016.11.035.

Xie, N. M., Yuan, C. Q., & Yang, Y. J. (2015). Forecasting chinas energy demand and self-sufficiency rate by grey forecasting model and markov model. *International Journal of Electrical Power & Energy Systems*, *66*, 1–8. doi:https://doi.org/10.1016/j.ijepes.2014.10.028.

Xiong, T., Bao, Y., & Hu, Z. (2014). Interval forecasting of electricity demand: A novel bivariate emd-based support vector regression modeling framework. *International Journal of Electrical Power & Energy Systems*, *63*, 353–362. doi:https://doi.org/10.1016/j.ijepes.2014.06.010.

Yang A, L. W., & X, Y. (2018). Short-term electricity load forecasting based on feature selection and least squares support vector machines. *Knowledge-Based Systems*, . doi:https://doi.org/10.1016/j.knosys.2018.08.027.

Zhang, J., Wei, Y., Li, D., Tan, Z., & Zhou, J. (2018). Short term electricity load forecasting using a hybrid model. *Energy*, *158*, 774–781. doi:https://doi.org/10.1016/j.energy.2018.06.012.