Energy Efficiency
https://doi.org/10.1007/s12053-019-09800-3

ORIGINAL ARTICLE

A hybrid approach to model and forecast the electricity consumption by NeuroWavelet and ARIMAX-GARCH models

Mehdi Zolfaghari · Bahram Sahabi

Received: 10 March 2018 / Accepted: 30 April 2019

Abstract Today, electrical energy plays a major role in production and consumption and is of special importance in economic decision-making process. Being aware of electrical energy demand for each period is necessary to correct planning. Therefore, the forecasting of electricity consumption is important among several economic sections. Besides the traditional models, in this paper, we offer a hybrid forecast approach that combines the adaptive wavelet neural network with the ARIMA-GARCH family models and uses the effective exogenous variables on electricity consumption. Based on this approach, two hybrid models are proposed. To assess the ability of the proposed models, we forecasted the daily electricity consumption by the hybrid and benchmark models for 60 days ahead in two separate seasons (summer and winter). The empirical results showed that the proposed models have more prediction accuracy compared with the other benchmark forecast models including neural network, adaptive wavelet neural network, and ARIMAX-GARCH family models.

Keywords Forecasting · Electricity consumption · ARIMAX-GARCH · Adaptive wavelet neural · Network wavelet · Hybrid models

M. Zolfaghari (✉) · B. Sahabi
Faculty of Management and Economics, Tarbiat Modares University, Tehran, Iran
e-mail: M.Zolfaghari@modares.ac.ir

B. Sahabi
e-mail: Sahabi_B@modares.ac.ir

Introduction

Electricity has distinct characteristics as compared with other commodities; it cannot be stored economically, and transmission congestion may prevent a free exchange of power among the control areas (Wu and Shahidehpour 2010). There are two insights about electricity consumption (EC): macroeconomics and microeconomics. From the macroeconomics perspective, electricity is one of the main inputs in industrial production. On the other hand, access to electricity is one of the vital needs for households and citizens. Therefore, providing sufficient electricity for households and manufactures will improve economic growth and development (Kavaklioglu et al. 2009). However, electricity generation is one of the major challenges in environmental studies. Analyzing the studies on electricity demand and factors affecting EC and the environmental studies shows that air temperature is one of the major exogenous variables on EC. Increases in temperature lead to increase in EC due to the use of cooling equipment and vice versa. Figure 1 is a strong evidence of this fact for Iran. The environmental studies emphasize that the earth is warming up due to increases in greenhouse gas emissions (Fig. 1, dotted line).

So, we expect that EC will be increased due to the earth warming. Since the most important source of electricity generation is fossil fuels, increase in EC leads to increase in fossil fuel consumption. On the other hand, the findings of the environmental researchers indicate that the major factor in increasing greenhouse

Published online: 28 May 2019

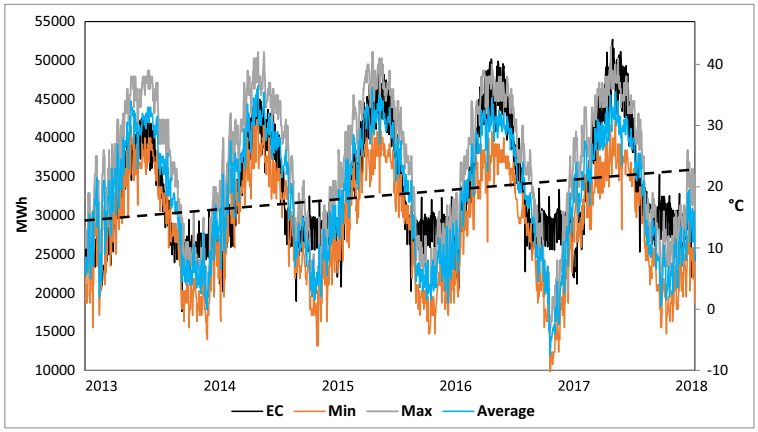

Fig. 1 The trend of electricity consumption (EC), max, min, and average temperature in Iran during January 01, 2013–March 02, 2018

Note: "MWh" denotes the Megawatt Hours and "°C" denotes the Centigrade.

gas emissions is the increase in fossil fuel consumption. So, there is a cycle among the earth warming up, electricity consumption, and fossil fuel consumption (Fig. 2).

The environmental challenge of EC is one of the key issues in energy policymaking. Hence, the estimation of potential environmental consequences in the future needs accurate forecasting of EC. It means that, with accurate prediction of EC, we will able to estimate its potential environmental consequences in the coming years. Indeed, an accurate prediction of future EC allows macro-policymakers to take appropriate measures in order to reduce the environmental consequences. In addition, prediction of EC is the basis for energy investment planning and establishing an energy policy either by international agencies or by the respective government. This energy demand will have to be satisfied by an optimum mix of the available energy sources, taking into account the restrictions imposed by future economic and social changes toward a sustainable world (Morales-Acevedo 2014).

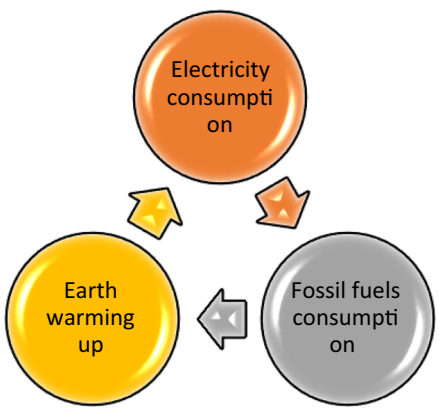

Fig. 2 Cycle among the earth warming up, electricity consumption, and fossil fuel consumption

From the microeconomics perspective, in recent years, electricity markets (EMs) in many countries have been deregulated to introduce competition in supply and demand activities. In a deregulated electricity market, generators compete to sell electricity and, at the same time, suppliers to consumers compete to purchase electricity (Liu and Shi 2013). In today's world, electricity forecasting is an important process in most utilities with the applications spread across several departments such as planning department, operation department, and trading department. (Hong 2010). An accurate prediction of EC can help the market participants (electricity generation and transmission companies and independent market operators) to make appropriate operational and management decisions such as resource scheduling and cost-effective risk management in EMs (Pindoriya et al. 2010). In this regard, energy forecasting covers a wide range of forecasting problems in the utility industry like generation forecasting, load forecasting, price forecasting, and demand response forecasting (Hong et al. 2014). For example, an overprediction leads to employing too many generators and unnecessary increase in the spin in reserve and operating costs. An underprediction may put the system reliability in danger due to insufficient resources to meet the security requirement or higher cost due to purchasing expensive spot market electricity to meet the demand (Rana and Koprinska 2016). As a result, energy forecasting, one of the most fundamental and classical problems, has found a new life in today's utility industry (Hong et al. 2014).

Improving the technical level of EC forecasting can not only accurately predict the demand of EM and be convenient for power companies to develop reasonable grid construction planning to improve the economic and social benefits of their system, but also be effective to predict the safety of power system operation and provide a reliable basis for the grid operation and maintenance (Zhai 2015).

There are two main groups of approaches for electricity consumption forecasting (ECF): (1) econometrics methods, and (2) artificial intelligence (AI) techniques (Rana and Koprinska 2016). The econometrics methods such as autoregressive integrated moving average (ARIMA), multiple linear regression, and stochastic time series are hard computing techniques due to the exact modeling of the system and utilizing linear analysis (Pindoriya et al. 2010). ARIMA is a popular procedure among the statistical models for time series analysis and forecasting applications (Sánchez et al. 2007). The ARIMA models with autoregressive conditional heteroskedastic (ARCH) (Engle 1982) or generalized autoregressive conditional heteroskedastic (GARCH) (Bollerslev 1986) processes are the widely used approaches to model the mean and volatility of electricity demand (Liu and Shi 2013). However, they have limited ability to capture non-linear and non-stationary characteristics of EC series and are not adaptive to rapid EC variations. Therefore, in recent years, the emphasis has shifted to the application of various AI techniques for ECF. The AI technique-based models like artificial neural networks (ANNs) (Mandal et al. 2013; Chen et al. 2010; Chaâbane 2014), fuzzy NNs (FNNs) (Kim et al. 1995), and knowledge-based expert system algorithms (Rahman and Hazim 1993) have been proposed in the literature for ECF. Among these models, ANNs are the most popular representatives of the second group, which can model non-linear relationships, and learn them from the asset of training examples (Pindoriya et al. 2010). However, one of the problems of ANNs is tackling the non-stationaries[1] (Wu and Shahidehpour 2010). The wavelet decomposition as a multiscale analysis tool can solve this problem through unfolding the inner EC characteristics. In recent years, with the rapid development of AI theory, its models (such as ANNs and wavelet analysis) have been widely used in load forecasting (Zhai 2015). Although these methods give accurate forecasts for time series, their weakness appears with noise or non-linear components, which characterize intra-day electricity demand. However, the combination of different models has proven to be an effective way to improve the forecasting accuracy (Chaâbane 2014).

In recent years, the hybrid forecast models with promising results have also been proposed in the literature. The benefit of hybrid models is to combine strengths of different techniques providing a robust modeling framework (Reston Filho et al. 2014). For instance, Reston Filho et al. (2014) proposed a wavelet-based approach for predicting the load from 1 to 24 h ahead from previous hourly data. They used the standard multilevel wavelet transform to decompose the load into several components that were inputs to two NN-based approaches. Chen et al. (2014) considered predicting the load 1-day ahead from previous hourly data using wavelets and NNs. They selected a day that was similar to the forecast day, and decomposed the load for it into two components using wavelet. Then, they built two separate NN models and combined their predictions. Non-wavelet features such as wind-chill temperature, humidity, cloud cover, and precipitation were also used as inputs for NNs. Amjady and Keynia (2008) proposed a combination of wavelet transform, NN, and an evolutionary algorithm for day-ahead price prediction in the PJM[2] market. The obtained results were compared with those of other traditional techniques and presented smaller errors. Nury et al. (2017) indicated that the wavelet-ARIMA model is more effective than the wavelet-ANN model. Valenzuela et al. (2008) presented a hybridization of intelligent techniques such as ANNs, fuzzy systems, and evolutionary algorithms. Interestingly, the final hybrid ARIMA-ANN model could outperform the prediction accuracy of those models when used separately. Examined on Spanish and PJM electricity markets, a novel price forecasting method based on wavelet transform combined with ARIMA and GARCH models was later suggested by Tan et al. (2010). Shafie-khah et al. (2011) conceived a

[1] A stationary process has the property that the mean, variance, and autocorrelation structure do not change over time. Most statistical forecasting methods are based on the assumption that the time series can be rendered approximately stationary through the use of mathematical transformations. Non-stationary data is, conceptually, the data that is very difficult to model because the estimate of the mean (and sometimes the variance) will be changing.

[2] PJM is a regional transmission organization (RTO) in the Eastern United States that operates one of the world's largest competitive wholesale electricity markets.

hybrid method that is based on wavelet transform, ARIMA models, and radial basis function neural networks (RBFNNs) for day-ahead electricity price forecasting.

Examination of the hybrid forecasting models used in previous studies showed that there are some important gaps, including the lack of considering the conditional heteroscedasticity in modeling the residual components (Conejo et al. 2005; Valenzuela et al. 2008; Bashir and El-Hawary 2009; Shafie-Khah et al. 2011; May et al. 2011; Zhang et al. 2012; Abedinia and Amjady 2015; Rana and Koprinska 2016), neglecting the leverage[3] and feedback[4] effects in time series modeling (Conejo et al. 2005; Valenzuela et al. 2008; Bashir and El-Hawary 2009; Wu and Shahidehpour 2010; Tan et al. 2010; Shafie-Khah et al. 2011; Zhang et al. 2012; Abedinia and Amjady 2015; Rana and Koprinska 2016; Kristjanpoller and Minutolo 2018), lack of applying the non-normal distributions such as Student *t* and generalized error distribution (GED) to achieve the best fitted model (Conejo et al. 2005; Valenzuela et al. 2008; Bashir and El-Hawary 2009; Tan et al. 2010; May et al. 2011; Shafie-Khah et al. 2011; Zhang et al. 2012; Abedinia and Amjady 2015 ; Rana and Koprinska 2016; Kristjanpoller and Minutolo 2018), and ignoring the effective exogenous variables in modeling and forecasting time series (Reis and Da Silva 2005: Conejo et al. 2005 ; Tan et al. 2010; Shafie-Khah et al. 2011; Zhang et al. 2012; Abedinia and Amjady 2015; Rana and Koprinska 2016; Kristjanpoller and Minutolo 2018). Also, the simple wavelet transform component has been used (Reis and Da Silva 2005; Conejo et al. 2005; Valenzuela et al. 2008; Tan et al. 2010; Chen et al. 2010; Shafie-Khah et al. 2011) to decompose the time series. So, in this paper, we tried to fulfill these gaps. In fact, the proposed hybrid models address all of the abovementioned limitations. The proposed methods consider the conditional heteroscedasticity, leverage, and feedback effects under normal, Student *t*, and GED distribution functions in the ARIMA-GARCH and ARIMAX-GARCH family models. In addition, the effective exogenous variables are used in the ARIMAX-GARCH family models, NN, and wavelet transform components. Some important features of the proposed EC forecasting models are:

- Combining the ARIMA-GARCH and ARIMAX-GARCH family models and adaptive wavelet neural network (AWNN) (these two hybrid models have not been found in other studies).
- Using a number of exogenous variables (temperature, month, day of week, and special days[5]) as well as the lag days (based on Auto Correlation Function (ACF)) in the structure of AWNN and ARIMAX-GARCH family models.
- Emphasizing on the leverage and feedback effects in the ARIMA-GARCH and ARIMAX-GARCH family models under normal, Student *t*, and GED distribution functions.
- Considering the abnormality of the conditional mean with the seasonal trends of time series.

For evaluation of the prediction accuracy of the hybrid models, we compared them with the common econometrics models (ARIMAX-GARCH family models) and AI models (ANN, AWNN). The motivation is that EC, at a given time, can be seen as a combination of components with different frequencies (e.g., low-frequency components represent more regular patterns, and high-frequency components represent irregular fluctuations). With identification of these components, combining them with other models, considering the number of exogenous variables, and predicting them separately, we will be able to build more accurate prediction models.

The structure of this paper is as follows: "Model specifications" introduces the econometrics and AI models. "Proposed models" describes the structure of the proposed hybrid models. The empirical analysis of the proposed hybrid and benchmark models is reported in "Empirical analysis." Results are presented in "Result," and a discussion is provided in "Discussion." Finally, the main conclusions are summarized in "Conclusion."

[3] The leverage effect means the phenomenon of a correlation of past returns with future volatility. This correlation is negative, which means the variance of EC with a decrease in EC.

[4] The feedback effect is based on the following dependence: if the volatility has its EC, the anticipated increase in volatility will increase in EC.

[5] Including holidays, special events, promotional activities, and system failures. We take it into account in the form of a binary dummy variable. The binary nature of the dummy contains the information about whether an event is occurring (1) or not (0).

Model specifications

This section briefly describes the models that are used widely by practitioners for modeling and forecasting the EC.

ARIMA-GARCH and ARIMAX-GARCH models

In the autoregressive (AR(p)) model, the current value of time series is expressed as a linear aggregate of p previous values and a random shock as follows:

$$y_t = a + \varnothing_1 y_{t-1} + \ldots + \varnothing_p y_{t-p} + \varepsilon_t \quad (1)$$

where, y_t is time series, $\varnothing_p$ is the parameter of the autoregressive part, and ε_t is white noise error terms ($\varepsilon_t \sim N(0, \sigma^2)$). Also, a is a constant term. The moving average (MA(q)) model, which expresses the current value of a time series as a current and q previous values of random shocks, is written as:

$$y_t = b + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \quad (2)$$

In Eq. (2), θ_q is the parameter of the moving average part. To increase flexibility when fitting actual time series, Eqs. (1) and (2) are combined to obtain a more general ARMA (p, q) model. If the time series are not stationary, ARMA turns to ARIMA. When an ARIMA model includes other time series as exogenous variables, the model is referred to as an ARIMAX model. In the ARIMAX (p, d, q), the specification of the electricity consumption time series (ECTS) is modeled as follows:

$$\Phi(B) y_t = c + \Theta(B)\varepsilon_t + \beta x_t \quad (3)$$

Where, y_t is ECTS, x_t is the vector of exogenous variables, and B is the backshift operator such that:

$$\begin{aligned} \Phi(B) &= 1 - \varnothing_1 B - \ldots - \varnothing_p B^p \\ \Theta(B) &= 1 - \theta_1 B - \ldots - \theta_q B^q \end{aligned} \quad (4)$$

Estimation of the abovementioned model requires covariance stationary ECTS and homoscedastic error terms. In the event that the conditional distribution differs from the unconditional distribution, the assumption of constant variance may no longer apply. While the unconditional variance must remain constant, as required of a covariance stationary series, the conditional variance may be conditionally heteroskedastic. In this situation, estimation of a time series needs to take into account the heteroskedastic. To solve this problem, GARCH model is useful. Some studies suggest ARIMA-GARCH–type models for explaining the volatility structure of the residuals obtained under the best suited mean models for the time series (Chand et al. 2012). Examples of these studies are Sumer et al. (2009), Erdogdu (2010), Hickey et al. (2012), Nury et al. (2017), and Zolfaghari and Sahabi (2017). In this study, we used the standard specifications of ARIMA-GARCH, ARIMA-GARCH-M (in mean), ARIMA-EGARCH, and ARIMA-EGARCH-M (in mean) models as follows:

$$y_t = \mu + \varepsilon_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \ \textit{for GARCH model} \quad (5)$$

$$y_t = \mu + \lambda\sigma_t + \varepsilon_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \quad \textit{for GARCH-M model} \quad (6)$$

$$y_t = \mu + \varepsilon_t, \quad \ln\left(\sigma_t^2\right) = \omega + \sum_{i=1}^{p} \alpha_i \left|\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right| + \sum_{k=1}^{r} \gamma_k \frac{\varepsilon_{t-k}}{\sigma_{t-k}} + \sum_{j=1}^{q} \beta_j ln\left(\sigma_{t-j}^2\right) \ \textit{for EGARCH model} \quad (7)$$

$$y_t = \mu + \lambda\sigma_t + \varepsilon_t, \quad \ln\left(\sigma_t^2\right) = \omega + \sum_{i=1}^{p} \alpha_i \left|\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right| + \sum_{k=1}^{r} \gamma_k \frac{\varepsilon_{t-k}}{\sigma_{t-k}} + \sum_{j=1}^{q} \beta_j ln\left(\sigma_{t-j}^2\right) \quad (8)$$

for EGARCH-M model

The first sentences of the above equations represent mean equations (ARIMA(p, d, q)). In these sentences, σ_t is conditional standard deviation (i.e., volatility), λ is coefficient of σ that represents the feedback effect, and ω is the intercept component. In Eqs. (7) and (8), the log of conditional variance implies that the asymmetric effect is exponential and that the forecasts of conditional variance are non-negative (Thomas and Mitchell 2005).

Neural networks (NNs)

NNs have been successfully applied to a variety of complex problems due to their ability to learn non-

linear relationships between the input and output patterns (Reston Filho et al. 2014). The structure of an artificial neuron is shown in Fig. 3. An ANN consists of multiple artificial neurons arranged in layers and joined together in different ways. Multi-layer perceptron (MLP) networks (also known as multilayer feed forward neural network (FFNN)) are the most popular and most widely used models of ANNs in many practical applications (Szoplik 2015).

The signals from all neurons located in the hidden layer are distributed to all the neurons of the output layer. In the neurons of the output layer, similar neurons from the hidden layer, firstly, aggregation of the weighted inputs to a neuron occurs upon activation of the output signal.

Adaptive wavelet neural network (AWNN)

Wavelets are mathematical tools for the analysis of time series and images. Wavelet neural network (WNN) is the extension of FFNNs. Comparing with ANN, WNN has the following advantages: (i) WNN has fast convergence because of the low correlation of wavelet neuron; (ii) expansion and contraction factors in the wavelet function make the approximation capability of the network more powerful; and (iii) with the good partial characteristic and multiresolution learning, the WNN matches commendably the signal, which can express function characteristic with different resolutions, and has higher forecast accuracy. The need for real-time processing has led to the development of adaptive wavelet neural networks (AWNNs), whose efficiency is determined by the rate of convergence of their learning algorithms.

A review of AWNN is described here and its advantages include faster network convergence speed and stable control performance. AWNN can handle controlling of systems without any prior information by learning it through time (Bouzari et al. 2011).

The general structure of an AWNN model is depicted in Fig. 4. It consists of an input layer, a wavelet layer, a product layer, and a linear output layer. The input data set $x = [x_1, x_2, \ldots, x_n]$, where n is the number of dimensions (input variables), which are directly transmitted into the wavelet layer. As the wavelet layer neurons make use of wavelets as activation functions, these neurons are usually referred to as "wavelons." Compared with other wavelet functions, the Mexican-hat wavelet function (the second derivative of the Gaussian function) has several distinguished characteristics, and has been used in this work due to the following advantages:

- It has an analytical expression; therefore, it can be used conveniently for multidimensional time series analysis.
- It is analytically differentiable; hence, any gradient-based techniques can be adopted to train the AWNN model.
- It is an admissible, real wavelet with two vanishing moments and computational efficiency (Pindoriya et al. 2010).

So, in this study, the Mexican-hat has been chosen as a mother wavelet. It is the second derivative of Gaussian function, which is defined as below:

$$\psi(x) = \left(1-x^2\right)e^{-0.5x^2} \tag{9}$$

Due to its symmetricity in shape, having explicit expression, and providing an exact time frequency analysis, it has been considered to be the most appropriate mother wavelet function (Pindoriya et al. 2008). If n represents the dimension of the input vector $x = [x_1, x_2,$

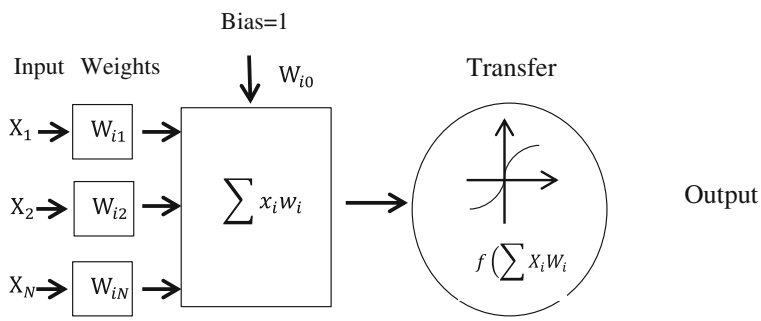

Fig. 3 The neuron model in a hidden or output layer of the artificial neuron network

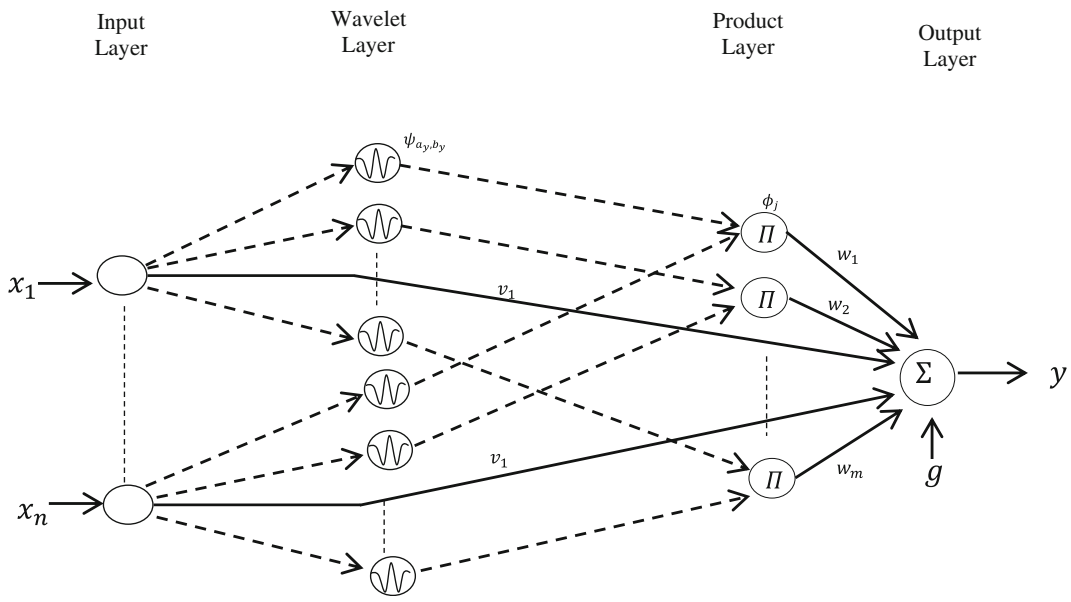

Fig. 4 General architecture of an AWNN model

$\ldots, x_n]$, then the wavelet family can be generated over the entire input space by translating and dilating the mother wavelet:

$$\phi_j = \prod_{i=1}^{n} \psi_{a_{ij},b_{ij}}(x_i) \qquad j \in m \tag{10}$$

where, m is the number of wavelons in the wavelet layer, a_{ij} and b_{ij} are the dilation and translation parameters, respectively, and $\psi_{a_{ij},b_{ij}}(x_i)$ is the dilated and translated mother wavelet of the jth wavelon connected with the ith input data, which can be expressed as follows:

$$\psi_{a_{ij},b_{ij}}(x_i) = \left(1 - \left(\frac{x_i - b_{ij}}{a_{ij}}\right)^2\right) e^{-0.5\left(\frac{x_i - b_{ij}}{a_{ij}}\right)^2} \tag{11}$$

With additional direct connection from the input to output nodes to map a linear input–output relation, the final output of an AWNN can be computed as:

$$y = \sum_{j=1}^{m} \omega_j \phi_j + \sum_{i=1}^{n} v_i x_i + g \tag{12}$$

where, ω_j is the layer weight between the jth node of product layer and output node, v_i is the input weight between the ith input variable and output node, and g is the bias at the output node. To predict the long-term electricity demand, usually a rolling mechanism is a useful tool (Ding et al. 2018). According to the AWNN structure, like other models, we apply the rolling mechanism for prediction of EC.

Proposed models

In our proposed approach, the time series $\{y_t\}$ is considered as a function of a linear component and a non-linear component. Thus, $y_t = f(L_t, N_t)$, where L_t denotes the linear component and N_t denotes the non-linear component. According to Zhang (2003), hybrid methodology is one of the most efficient models for improving forecast accuracy. We can establish an additive relationship between the linear and non-linear components. Consequently, we can write $y_t = L_t + N_t$, starting from the assumption that the linear and non-linear patterns of the considered time series can be modeled separately by different models.

According to some researchers (e.g., Zhang 2003; Chaâbane 2014), the ARMA has a high capability to estimate and forecast the linear (smooth) time series. Also, NN has a high capability to model and forecast non-linear time series.

In this paper, we propose two hybrid models. The structure of these models is shown in Fig. 5. In brief, the proposed hybrid methodology consists of (a) using discrete wavelet transforms (DWTs) to decompose the original time series into an approximate part and several detail parts; (b) using the ARIMA-GARCH and ARIMAX-GARCH family models to analyze and estimate the linear component of the time series (approximate part) based on three distribution functions, including normal, Student t, and

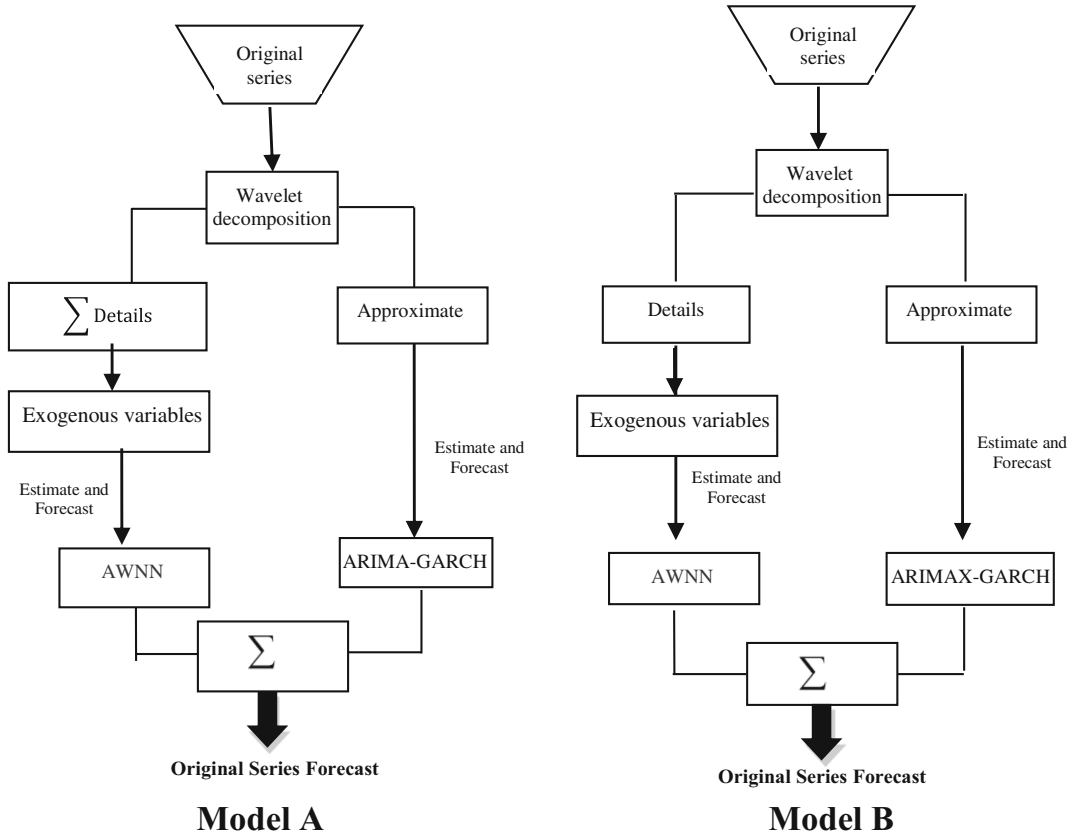

Fig. 5 The structure of proposed models

GED distributions; and (c) modeling the non-linear components of the time series (detail parts) through an AWNN model. Consequently, the obtained predictions from the models are recombined. Hence, the hypothesized hybrid models would exploit the strength of the ARIMA-GARCH and ARIMAX-GARCH family and AWNN models by taking into account the exogenous variables.

Both of the proposed models involve essentially five steps:

In model A At first, we decompose the historical EC data into an approximate (s) and several detail parts (d) through the multilevel wavelet decomposition as follows:

$$y_t = f(t) = \sum_{a=1}^{\infty} s_{b,ab,b}(t) + \sum_{a=1}^{\infty} d_{b,a}\psi_{b,a}(t) + \sum_{a=1}^{\infty} d_{b-1,a}\psi_{b-1,a}(t) + \ldots + \sum_{a=1}^{\infty} d_{1,a}\psi_{1,a}(t) \quad (13)$$

Coefficients of wavelet are obtained as follows:

$$s_{b,a} \approx \int W_{b,a}(t) f(t) dt \quad (14)$$

$$d_{b,a} \approx \int \psi_{b,a}(t) f(t) dt \quad (15)$$

where, $s_{b,\,a}$ is approximate trend in the bth level, and $d_{b,\,a}$ is the detail part.

$$W_{b,a}(t) = s^{-\frac{b}{2}}\left(\frac{t - s^b a}{s^b}\right) \quad (16)$$

$$\psi_{b,a}(t) = s^{-\frac{b}{2}}\psi\left(\frac{t - s^b a}{s^b}\right) \quad (17)$$

In the second step, we integrate the detail parts ($D = \sum d_{b,\,a}$) and feed them into the AWNN for the respective training and prediction. We apply some exogenous variables as neurons in the input layer. Exogenous variables include the calendar (month, day of

week, special days) and temperature (min, max, average) factors.

In this step, we use the standard backpropagation gradient descent algorithm (SBPGDA) as a training algorithm in AWNN. Training is based on minimization of cost function, also called as mean square error (MSE):

$$E = \frac{1}{2N}\sum_{p=1}^{N}[e(p)]^2 \qquad and \quad e(p) = y^d(p) - y(p) \tag{18}$$

where, $y(p)$ is the model output, $y^d(p)$ is the desired output for a given pth input pattern, and N represents the total training patterns.

During the training process, the over fitting of training data can be stopped at the right point using cross-validation. The available data set is divided into training, validation, and testing subsets. The training subset is used to compute the gradients and update all free parameters of the network. The error on the validation subset is monitored during the training session. To make the learning process converge more rapidly than the conventional method, where both learning rate and momentum parameters are kept constant during the learning process, an adaptive learning rate has been proposed by Pindoriya et al. (2008). The proposed adaption rule is as follows:

$$\eta(n+1) = \begin{cases} 1.05\eta(n) & , \Delta E_n > 0 \\ 1.05\eta(n) & , \text{otherwise} \end{cases} \tag{19}$$

where, $\eta(n)$ is the learning rate at iteration n, and $\Delta E_n = E(n-1) - E(n)$ with $E(n)$ being the mean squared error at the end of the nth iteration. Here, the basic idea is to increase η when ΔE_n is positive and decrease η when ΔE_n is negative. Note that for positive ΔE_n, the error is decreasing, which implies that the network parameters are updated to the correct direction. If the parameters move onto the wrong direction, causing error to increase, the direction will be ignored in the next iteration by decreasing the learning rate.

In the third step, we use ARIMA-GARCH family models to fit the linear component (the approximate time series) similar to Eqs. (5) to (8).

In the fourth step, we forecast the approximate time series (s_t) by the best fitted ARIMA-GARCH model and the integrated detail part of time series (D) by AWNN model for 60 days ahead in two separate seasons (summer and winter).

After prediction, in the last step, we recombine the predicted outputs to obtain the final predicted output:

$$y_t^f = s_t^f + D_t^f \tag{20}$$

where, s_t^f and D_t^f are forecasted values of the approximate and the integrated detail part of time series, respectively.

In model B At first, like model A, we decompose the historical EC data into an approximate part and several detail parts through the multilevel wavelet decomposition.

In the second step, each detail part (d) is fed into the AWNN for training and prediction separately. In this section, in addition to the exogenous variables (mentioned in model A), we apply the lag days (based on ACF) in the input layer for each detail part. Like model A, we use the SBPGDA as the training algorithm in the AWNN and the proposed adaption rules that were presented in Eq. (19).

In the third step, ARIMAX-GARCH family models are employed to fit the linear component (the approximate time series). Here, the exogenous variable is the average temperature. We do not use the calendar factors, because they are used into the ARIMA structure (primary phase of the ARIMAX-GARCH process). In fact, the weekly, monthly, and annual effects as coefficients of independent variables are considered in the ARIMA model.

In the fourth step, we forecast the approximate time series by the best fitted ARIMAX-GARCH model and each detail part of time series by AWNN for 60 days ahead in two separate seasons.

After prediction, in the last step, we combine the whole predicted detail parts of time series together. Finally, we recombine them with the predicted approximate time series to obtain the final predicted output.

Empirical analysis

Data description

The study focused on Iran's electricity market for the period January 1, 2013, to March 2, 2018. We applied daily time series of "electricity consumption at peak

time" in mid-day for EC. The data set was obtained from Iran Grid Management Company (IGMC). The weather data were obtained from Iran Meteorological Organization (IMO). Figure 6 shows the daily EC during the period January 1, 2013, to March 2, 2018.

According to Table 1, EC ranges from 17,616 to 52,692 MWh with the mean and standard deviation being 33,751 and 7319, respectively. The kurtosis statistics, which compares the peakedness and tailedness of the probability distribution with that of a normally distributed series, shows that the EC is low-peaked and fat-tailed. The Jarque and Bera (1980) statistic measuring the normality of the distribution using both skewness and the kurtosis statistics shows that we can reject the null hypothesis of normality for EC time series at all conventional significant levels.

For econometric modeling (like ARIMA model), we need to evaluate the stationary state of EC. By performing unit root tests, namely ADF (augmented Dickey Fuller), PP (Philips–Perron), and KPSS (Kwiatkowski–Phillips–Schmidt–Shin), we tested for stationarity. As illustrated in Table 2, for ADF on the 5% test level, and for PP and KPSS tests on the 1% test level, all the statistical values fall into the reject region, which means *H0* (time series is non-stationary) should be rejected. Thus, the considered time series is stationary and suitable for subsequent tests in our study. Therefore, ARIMA is turned to ARMA.

Main features of the daily EC and input variable selection

Figure 6 shows the daily ECTS during the period January 1, 2013, to March 2, 2018. As shown in Fig. 6, the ECTS exhibits multiple periodicities, corresponding to daily, weekly, monthly, and other periodicities. The daily ECTS is also influenced by the calendar effect (i.e., weekends and holidays). Some of these characteristics can be easily identified by inspecting the daily series. One can also easily observe that the value of EC in weekends is usually lower than in the working days. This is because the economic and social activities on the working days are at a higher level than on the weekend as well as the day before the weekend (due to reduction in working hours). In addition, the EC is a little higher on the first working day of the week than in other working days. It is well known that the daily EC has some seasonal patterns. According to Fig. 6, in summer, particularly from the second half of June until the first half of August, there are peak load demands because of heat wave. Air conditioning loads contribute to a large portion of the summer peak load demand. On the other side, during the months of December and January, the temperature reduces to the lowest level. Such information might be useful for developing EC models. Moreover, the temperature is the most dominant weather factor that drives the short-term EC.

Table 1 Descriptive statistics of EC (MWh)

Number of observations	1886
Mean	33,751
Std. Dev.	7319
Maximum	52,692
Minimum	17,616
Skewness	0.577
Kurtosis	2.23
Jarque–Bera	151.4
Probability	0.000

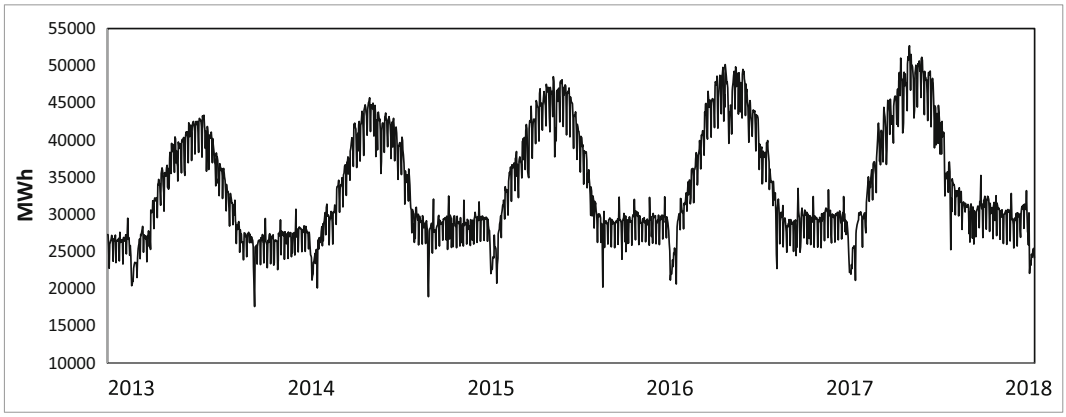

Fig. 6 Daily EC during January 01, 2013–March 02, 2018

Table 2 The unit root test for the ECTS

Test	1% critical value	5% critical value	10% critical value	Value of statistic
ADF test	− 3.433	− 2.862	− 2.567	− 3.065
PP test	− 3.433	− 2.862	− 2.567	− 6.96
KPSS	0.739	0.463	0.347	0.284

The relation between daily air temperature data and EC data is depicted in Fig. 7 (as well as Fig. 1). As shown, there is a non-linear relationship between EC and temperature. Figure 7 illustrates that at low temperature range, increase in temperature initially results in reduction of EC due to reduction in electric space heating, whereas, at higher temperatures (more than 23 °C), the effect of space cooling via air conditioners and the use of other appliances causes higher electricity demands.

It can be observed that the EC series presents multiple periodicities, and hence, the past EC could affect and imply the future EC. Therefore, besides the calendar date (7 days a week, 12 months a year, and special days) and temperature (min, average, and max), the historical EC series data should also be considered to model the trend and seasonality of the EC signal. To be more precise, if EC in day d (i.e., L_d) is to be forecasted, the EC information of previous days (up to p days) (i.e., $L_{d-1}, L_{d-2} \ldots L_{d-p}$) should be taken as a part of the inputs for modeling and forecasting the EC. ACF can be used to identify the degree of association between the data in EC series separated by different time lags. As shown in Fig. 1, in addition to cyclical behavior, the EC's overall trend is positive.

Implementation of hybrid models

Model A

In this model, we first decomposed the ECTS into an approximate part and five detail parts based on the mother wavelet. Figure 8 shows the decomposed ECTS at five levels. Selecting the number of detail parts is according to the nature of the time series' fluctuations. Usually, the number of detail parts has been selected between three and six previous studies (e.g., Pindoriya et al. 2008; Tan et al. 2010; Pindoriya et al. 2010; Nury et al. 2017).

In Fig. 8, Y is the original time series, and s and d are the time series of approximate part and detail parts, respectively. In the second step, we integrated the detail parts together to obtain D time series (see Fig. 9), and fed it into WANN for training and prediction. Figure 9 shows the sum of detail parts.

Also, six variables were taken into account in the input layer (equal to 23 neurons) including temperature (min, average, and max) and calendar date (7 days a week, 12 months a year, and special days).

The general structure and training of AWNN model have been already described in "Proposed models." Now, based on specific EC forecasting problem and the number of input variables, AWNN model is trained for different numbers of wavelons {2, 3, …, 20} with the initial learning rate of 0.01, the momentum parameter of 0.6, and the increment and decrement learning rate multipliers of 1.01 and 0.9, respectively. The results showed that AWNN gives minimum error for the wavelons equal to 6. Thus, the structure of AWNN is selected to be 23–6–1. In previous empirical studies, the above parameters (the number of wavelons, the initial learning rate, the momentum parameter, and the decrement learning rate multipliers) have been set near the above numbers (e.g., Pindoriya et al. 2010).

In the third step, we used ARIMA-GARCH family models to fit the linear component (the approximate time series). Based on ADF, PP, and KPSS tests, the approximate time series is stationary. Modeling the approximate time series consists of two phases according to the Box and Jenkins (1976) process.[6] The first phase involves the specification of the ARMA (p, q) model for mean returns, and the diagnostic tests of their residuals. The second phase is the specification of our strategy for modeling the conditional mean to search over alternative ARMA (p, q) models by varying the p and q parameters from 0 to 5 using the autocorrelation

[6] Include:
- The checking of stationary or non-stationary and transforming the data, if necessary;
- The identification of a suitable ARMA model;
- The estimation of the parameters of the chosen model;
- The diagnostic checking of the model's adequacy.

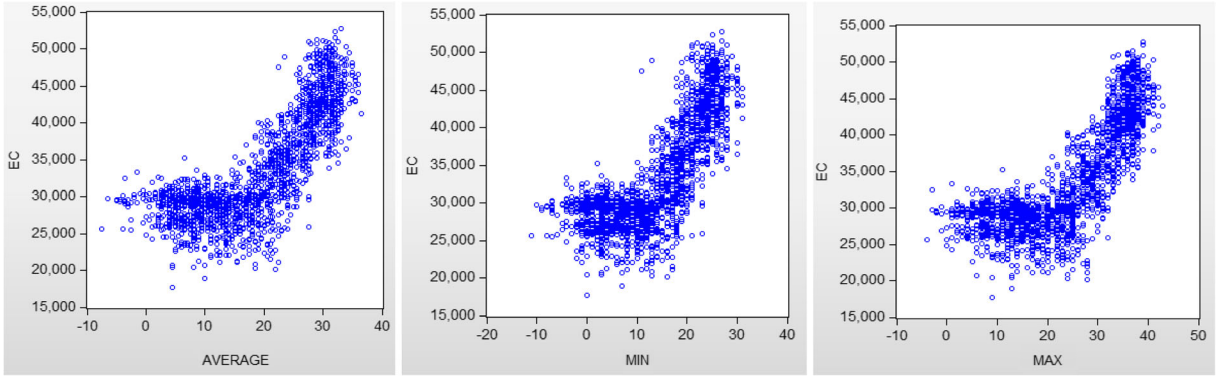

Fig. 7 Relation between daily temperature and daily EC

(AC) and partial autocorrelation (PAC) functions corresponding to the stages described by Box and Jenkins (1976) and identifying the best fitted model using the Schwartz Information Criterion (SIC) proposed by Schwarz (1978). To model ARMA and ARMAX, SIC is extensively adopted to guide the choice of alternative models. SIC uses a likelihood function to select the best fitted model. This criterion represents a trade-off between "fit" measured by the log likelihood value and "parsimony" as measured by the number of free parameters. The models with smaller SIC values are preferred (Liu and Shi 2013).

After estimating the ARMA model and identifying the number of p (equal to 1, 364) and q (equal to 1), the "ARCH effect" was tested on it. The results showed that this model has no ARCH effects on the base of F and χ^2 statistics. Therefore, the final model of ARMA(1)[7] with annual effect is as follows:

$$\underset{(122.6)}{y_t} = \underset{(16.5)}{34377} + \underset{(6.14)}{0.63y_{t-1}} + \underset{(6.58)}{0.19y_{t-364}} + 0.96u_{t-1} \quad (21)$$

In Eq. (21), the first sentence represents the ARMA model and the second one is the values of t-statistics that are reported in the parentheses. So, according to the t-statistic values, the estimated parameters of ARMA are significant at the level of 1%. The high significance of the estimated parameters in the ARMA model adequately discloses that the approximate time series has the apparent daily and yearly periodicities.

After modeling the approximate time series based on the ARMA model and the detail parts of time series based on AWNN, in the fourth step, we forecasted both of them by ARMA and AWNN, respectively, for 60 days ahead in two separate seasons (summer and winter). Finally, we recombined the predicted outputs to obtain the final predicted output.

Model B

In this model, first, we decomposed ECTS into an approximate part and five detail parts. The decomposing of ECTS was according to model A. In the second step, we modeled each detail part based on AWNN. For this purpose, we applied the lag days based on ACF in AWNN for each detail part. The process of training of AWNN model is like model A, but the numbers of input variables based on ACF are different. ACF analysis on EC series is carried out by considering the entire series and also the subsets of series at different intervals. Figure 10 shows that in the denoised signal of EC, there is a seasonal pattern, and the correlation factor decreases gradually as the lag times increases.

In addition to the lag days (based on ACF) in AWNN, we applied six factors[8] in the input layer for each detail part.

In the third step, we used ARMAX-GARCH family models to fit the linear component (the approximate time series) based on the Box and Jenkins (1976) process. The only exogenous variable was the daily average temperature (X (ave)). We did not use the calendar factors; rather, they were applied into the ARMA structure (the primary phase of ARMAX-GARCH process). In fact, the annual effect (y_{t-364} = AR(364) in Table 3) as the coefficient of independent variable was considered in the ARMAX model. The coefficients of weekly

[7] Considering the lack of integration order (I), the ARIMA models were turned to ARMA.

[8] These include 7 days a week, 12 months a year, temperature (min, average, and max), and special days.

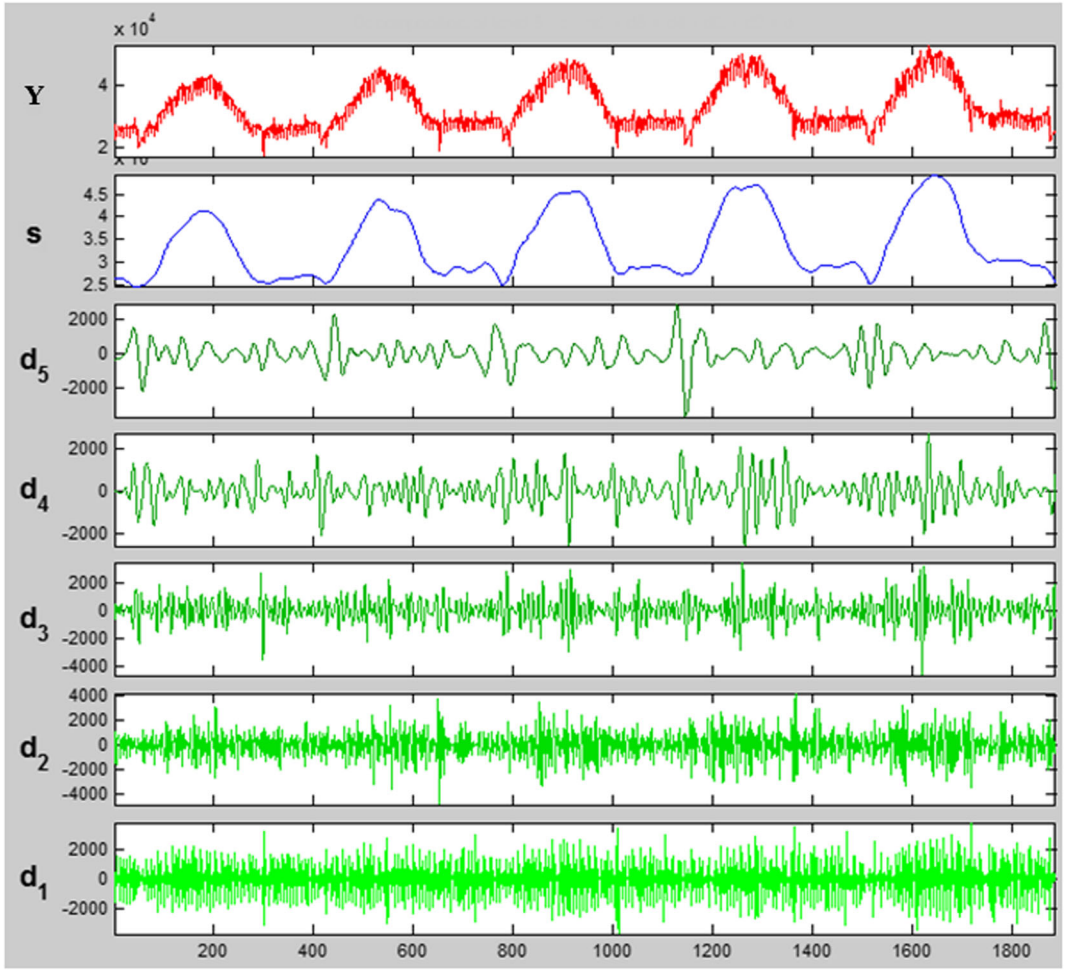

Fig. 8 Decomposed ECTS at 5 levels: $Y = s + d_5 + d_4 + d_3 + d_2 + d_1$

and monthly effects were insignificant. So, we removed them from the model.

After estimating the ARMAX model for approximate time series, the "ARCH effect" was tested on it. The results showed that there is an ARCH effect in the ARMAX model based on F and χ^2 statistics. Thus, the conditional variance of approximate time series was modeled based on four GARCH family models

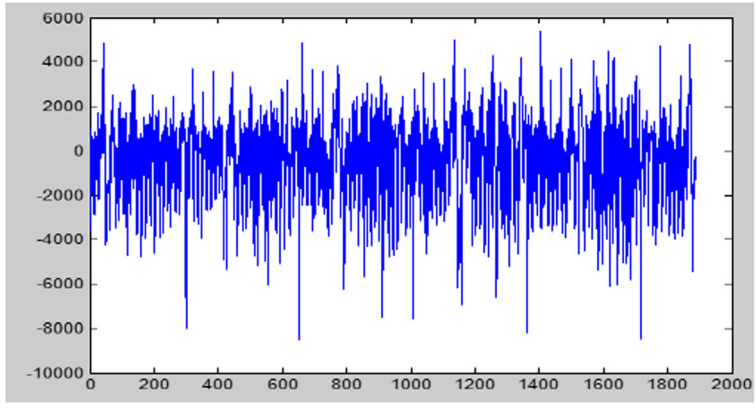

Fig. 9 Sum of detail parts: $D = d_5 + d_4 + d_3 + d_2 + d_1$

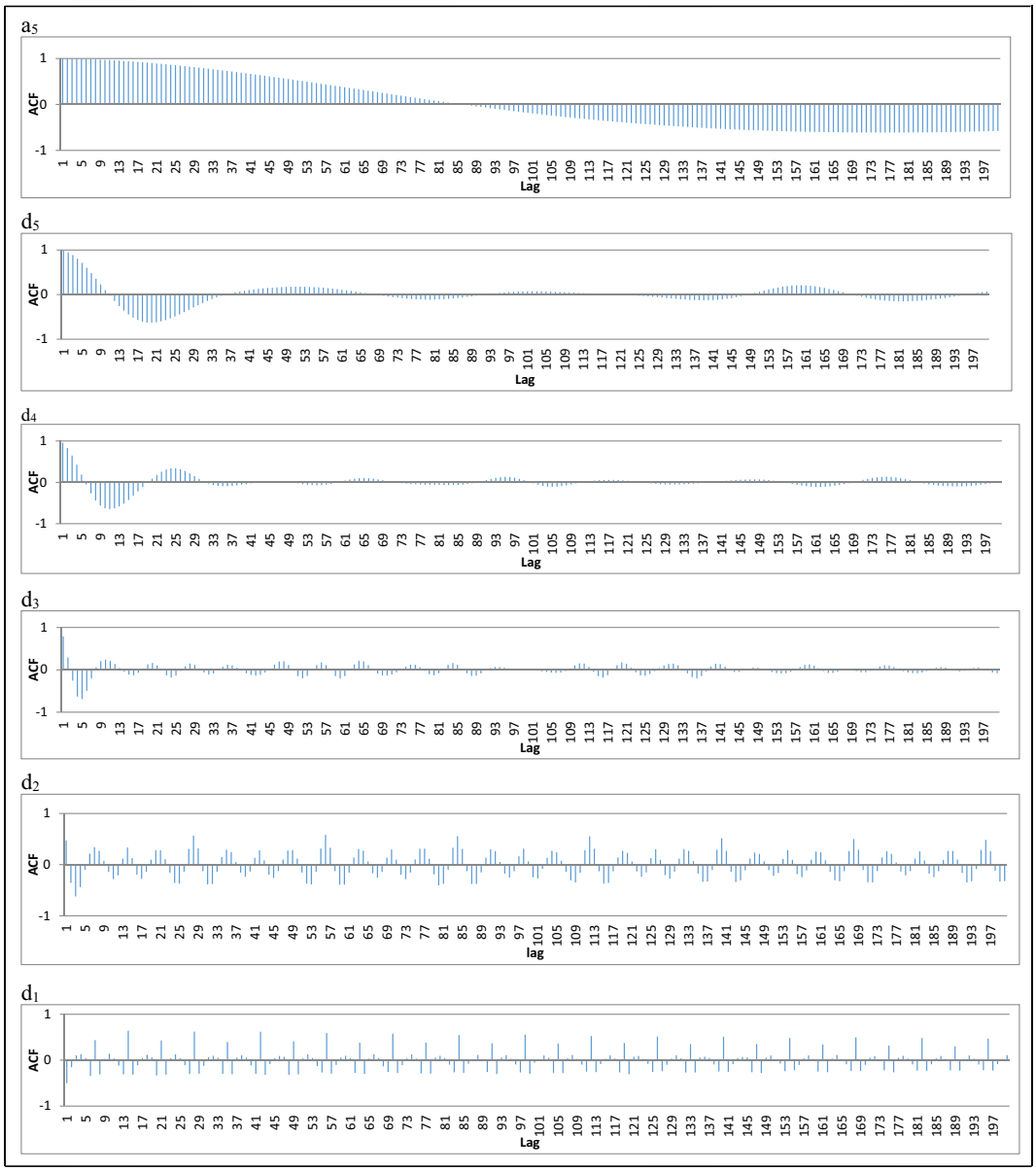

Fig. 10 ACF of detailed and approximate time series

under the normal, Student *t*, and GED distributions. Table 3 shows the conditional mean and variance models for the approximate time series.

As shown in Table 3, most of the estimated parameters are significant at 1% level. The high significance of the estimated parameters in the GARCH component shows that the volatilities of approximate time series are time-varying.

In this section, we used SIC and *R*-squared (R^2) criteria to choose the best fitted model among the ARMAX-GARCH family models. These statistics were given in the last two rows in Table 3. According to SIC and R^2 values, we chose ARMAX-EGARCH-in-mean model with GED distribution as an appropriate model. In the selected model, the approximate time series is significantly characterized by its past day (AR(1)) and its year (AR(364)) as well as the past day error (MA(1)). The exogenous variable has a positive and significant effect on it. In addition, although the original GARCH models are based on the Gaussian distribution, significant evidence suggests that the time series is rarely Gaussian. In fact, the approximate time series has a

Table 3 The conditional mean and variance models for the approximate time series

Variable	GARCH			GARCH-in-mean			EGARCH			EGARCH-in-mean		
	Normal	*t*	GED	Normal	*t*	GED	Normal	*t*	GED	Normal	*t*	GED
C	23,937* (5.49)	23,881* (11.66)	23,742* (295)	23,751* (2.95)	24,032* (5.36)	23,391* (56.76)	23,816* (19.6)	24,149* (89.03)	31,766* (59.39)	24,263* (44.21)	24,105* (2.18)	31,121* (35.80)
AR(1)	0.95* (46.90)	0.98* (418.53)	0.79* (823)	0.98* (115.01)	0.99* (438)	0.86* (3120)	0.90* (207)	0.97* (2475)	0.89* (1488)	0.93* (1093)	0.94* (94.27)	0.90* (1985)
AR(364)	− 0.06* (− 3.11)	0.02* (7.25)	0.22* (213.04)	0.03* (29.7)	0.06* (5.51)	0.01 (1.03)	0.11* (22.75)	0.02* (44.74)	0.01*** (1.79)	0.01 (0.45)	0.02* (6.77)	0.16* (589)
MA(1)	0.29* (5.67)	1.00* (923)	0.85* (536)	0.24* (6.90)	1.00* (628)	0.87* (219)	0.96* (1135)	0.98* (325)	0.89* (611)	0.98* (170)	0.94* (5.19)	0.88* (879)
X (ave)	367.77* (23.14)	0.13* (3.35)	16.02* (58.49)	130* (24.39)	0.01 (0.10)	2.81* (33.57)	0.01 (0.62)	0.01 (1.31)	0.05* (5.14)	0.03* (3.29)	− 0.07 (− 0.06)	0.03* (2.98)
λ				− 0.02 (− 1.89)	0.01 (− 0.27)	0.04 (199)				0.01 (5.39)	0.01 (1.76)	0.02 (2.50)
ω	10,809,957* (16.31)	16,630,702 (0.01)	16,630,702* (2.34)	10,788,648* (14.53)	16,597,909 (0.07)	16,597,923* (2.08)	11.96* (8.93)	− 0.26* (− 2.23)	0.22* (3.56)	− 0.20*** (− 1.78)	11.9* (2.07)	0.14** (1.96)
α	0.18* (11.68)	2826 (0.01)	− 2.49 (− 0.42)	0.02** (1.80)	5512 (0.07)	26.11 (1.19)	0.53* (6.01)	1.27* (8.56)	0.72* (16.83)	1.24* (9.48)	− 0.07 (− 0.28)	0.68* (17.62)
β	− 0.98* (− 11,653)	0.10 (0.20)	− 0.99* (− 40.5)	− 0.98* (− 446,680)	− 0.10 (− 0.22)	− 0.12 (− 0.29)	0.19* (3.58)	− 0.02 (− 0.24)	− 0.01 (− 0.61)	− 0.01 (− 0.22)	0.37 (1.60)	− 0.01 (− 0.69)
γ							− 0.39 (− 2.6)	0.88 (43.38)	0.87 (81.67)	0.88 (33.30)	− 0.19 (− 0.33)	0.88 (79.32)
R^2	0.954	0.955	0.955	0.953	0.954	0.955	0.954	0.954	0.955	0.954	0.955	0.955
SIC	17.338	11.351	14.215	16.093	11.207	13.676	11.806	10.393	10.063	10.420	12.322	10.061

*, **, and *** represent the levels of significance at 1%, 5%, and 10%, respectively. Also, the value of *t* statistics are reported in the parentheses

fat-tailed behavior. The conditional variance supports the inverse leverage effect and reveals the fact that the impact of a positive shock is larger than that of a negative one. In other words, the EC asymmetrically responds to both the positive and negative shocks. In addition to the leverage effect, the significant coefficient λ presents that there is a feedback effect in the selected model. It means that an anticipated increase in volatility (conditional deviation) leads to a rise in the approximate time series.

After modeling the approximate time series based on the ARMAX-EGARCH-in-mean model with GED distribution and the detail parts of time series based on AWNN, in the fourth step, we forecasted both of them by ARMAX-EGARCH-in-mean and AWNN, respectively, for 60 days ahead in two separate seasons (summer and winter). After prediction, in the last step, we recombined the predicted outputs to obtain the final predicted output.

Table 4 shows the number of inputs in the proposed models.

Implementation of benchmark models

ARMA-GARCH and ARMAX-GARCH models

Step 1: Estimation of the ARIMA and ARIMAX models EC modeling includes two phases that were mentioned in model A. Like the approximate time series modeling, we modeled EC based on ARMA and ARMAX models and chose the best fitted model using SIC and R-squared (R^2). Table 5 shows the estimation of ARMA and ARMAX models for EC.

According to the *t*-statistic values (in the parentheses), the estimated parameters for the mentioned two models are significant at the level of 1%, disclosing that EC has the apparent daily (AR(1)), weekly (AR(7)), and yearly (AR(364)) periodicities. Also, the coefficient of temperature (X (ave)) in the ARMAX model is positive and significant at the level of 1%. It means that when the temperature rises, the EC increases. In Table 5, the value of adjusted R^2 in the ARMAX model is higher than that in the ARMA model. Also, SIC value in the ARMAX model is lower than that in the ARMA model. As a result, we selected the ARMAX model as the best fitted model for EC. After estimating the ARMAX model, the "ARCH effect" was tested on it. The results showed that there is an ARCH effect in the ARMAX model based on F and χ^2 statistics. Thus, the conditional variance of EC was modeled based on four GARCH family models under the normal, Student *t*, and GED distributions. Table 6 shows the conditional mean and variance models for EC.

In Table 6, the values of adjusted R^2 in ARMAX-EGARCH-in-mean model under Student *t* distribution are higher than those in other models. Also, the SIC value in this model is lower than that in other models. Therefore, according to SIC and R^2 values, the ARMAX

Table 4 The number of inputs in the proposed models

Model	Components	Inputs
Model A	Integrated detail parts (D)	23 inputs include temperature (min, average, and max) and calendar date (7 days a week, 12 months a year, and special days)
	The approximate time series (s)	s_{t-1}, s_{t-364}, u_{t-1}
Model B	Detail part (d_5)	58 inputs include temperature (min, average, and max) and calendar date (7 days a week, 12 months a year, and special days) + $d_{5,\,t-1}$, $d_{5,\,t-2} \ldots d_{5,\,t-35}$
	Detail part (d_4)	53 inputs include temperature (min, average, and max) and calendar date (7 days a week, 12 months a year, and special days) + $d_{4,\,t-1}$, $d_{4,\,t-2} \ldots d_{4,\,t-30}$
	Detail part (d_3)	36 inputs include temperature (min, average, and max) and calendar date (7 days a week, 12 months a year, and special days) + $d_{3,\,t-1}$, $d_{3,\,t-2} \ldots d_{3,\,t-13}$
	Detail part (d_2)	33 inputs include temperature (min, average, and max) and calendar date (7 days a week, 12 months a year, and special days) + $d_{2,\,t-1}$, $d_{2,\,t-2} \ldots d_{2,\,t-10}$
	Detail part (d_1)	37 inputs include temperature (min, average, and max) and calendar date (7 days a week, 12 months a year, and special days) + $d_{1,\,t-1}$, $d_{1,\,t-2} \ldots d_{1,\,t-14}$
	The approximate time series (s)	s_{t-1}, s_{t-364}, u_{t-1}, temperature (average), σ_t, $\left\lvert\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right\rvert$, $\frac{\varepsilon_{t-1}}{\sigma_{t-1}}$, $\ln\left(\sigma_{t-1}^2\right)$

Table 5 ARMA and ARMAX models to EC

	C	AR(1)	AR(7)	AR(364)	MA(1)	*X* (ave)	R^2	SIC
ARMA	6290.2	0.202*	0.323*	0.496*	0.342*	–	95.51	17.582
	(0.604)	(11.27)	(16.46)	(23.67)	(11.86)			
ARMAX	6344.8	0.199*	0.328*	0.493*	0.338*	51.73*	95.56	17.579
	(0.615)	(11.18)	(16.66)	(23.51)	(11.71)	(3.33)		

*, **, and *** represent the levels of significance at 1%, 5%, and 10%, respectively

model with non-linear and asymmetric GARCH processes enjoys more potential to model EC.

NN structure

In the MLP network, data were split into three non-overlapping data sets: training (70%), validation (15%), and testing (15%). A number of neurons in the input layer of the network were equal to the number of exogenous variables including 7 days a week, 12 months a year, special days, and temperature (min, max, average), while a neuron in the output layer was equal to the amount of EC. One hidden layer with log sigmoid function for its neurons and single output neuron with linear function has been considered in the MLP structure of NNs. The number of neurons in the hidden layer of the network changed in the range of 5–40. We created 35 different NNs by varying the number of neurons in the hidden layer and selected the best NN in order to forecast the testing data. As a training algorithm, we

Table 6 The conditional mean and variance models for EC

Variable	GARCH			GARCH-in-mean			EGARCH			EGARCH-in-mean		
	Normal	*t*	GED	Normal	*t*	GED	Normal	*t*	GED	Normal	*t*	GED
C	12,102*	11,335*	19,695.3*	− 15,148	5796.3	24,418.6*	10,779.6*	9880.9**	12,552.9*	1579.9	1774.5	5475.8
	(2.4)	(2.4)	(28.6)	(− 1.0)	(1.1)	(58.3)	(2.1)	(1.8)	(3.1)	(0.2)	(0.3)	(1.1)
AR(1)	0.14*	0.11*	0.18*	0.20*	0.11*	0.15*	0.14*	0.10*	0.10*	0.13*	0.10*	0.11*
	(10.14)	(8.95)	(32.95)	(13.44)	(9.18)	(23.41)	(9.92)	(8.48)	(8.62)	(9.93)	(8.92)	(9.44)
AR(7)	0.35*	0.39*	0.37*	0.33*	0.37*	0.41*	0.35*	0.40*	0.40*	0.34*	0.37*	0.38*
	(26.42)	(27.94)	(69.01)	(22.71)	(27.19)	(58.20)	(26.89)	(28.64)	(29.49)	(25.67)	(27.39)	(27.69)
AR(364)	0.54*	0.53*	0.49*	0.49*	0.55*	0.50*	0.54*	0.53*	0.53*	0.55*	0.55*	0.55*
	(36.85)	(35.56)	(107.90)	(38.01)	(37.14)	(66.71)	(39.07)	(35.68)	(36.48)	(38.33)	(37.77)	(37.94)
MA(1)	0.34*	0.35*	0.29*	0.33*	0.35*	0.26*	0.34*	0.37*	0.35*	0.35*	0.36*	0.35*
	(9.91)	(11.81)	(45.06)	(13.75)	(11.91)	(25.57)	(10.12)	(12.65)	(13.22)	(10.47)	(12.70)	(12.96)
X (ave)	43.56*	38.66*	40.28*	54.32*	37.66*	40.79*	42.96*	35.16*	33.26*	44.09*	33.30*	30.60*
	(2.83)	(3.47)	(14.98)	(3.22)	(3.41)	(9.95)	(3.08)	(3.19)	(3.25)	(3.14)	(3.09)	(3.01)
λ	–	–	–	0.13*	0.10*	0.04*	–	–	–	0.10*	0.12*	0.11*
				(13.15)	(4.40)	(4.58)				(4.24)	(5.34)	(5.21)
ω	173.6*	129.5*	206.1*	475.6*	128.5*	205.9*	12.88*	7.44*	8.90*	12.93*	6.63*	8.64*
	(14)	(6.57)	(4.93)	(37.24)	(6.56)	(5.16)	(11.18)	(5.78)	(5.77)	(11.24)	(5.57)	(5.74)
α	0.32*	0.56*	0.05***	0.01*	0.55*	0.02	0.56*	0.71*	0.65*	0.57*	0.70*	0.65*
	(7.84)	(5.08)	(1.64)	(7.09)	(5.02)	(1.02)	(12.00)	(8.13)	(8.01)	(11.99)	(8.08)	(8.04)
β	0.002	0.13***	− 0.99*	− 1.01*	0.13***	− 1.00*	− 0.03	0.03	0.01	− 0.01	0.08*	0.04
	(0.04)	(1.69)	(− 239.87)	(− 520)	(1.73)	(− 784.21)	(− 0.84)	(0.47)	(0.11)	(− 0.36)	(1.67)	(0.81)
γ	–	–	–	–	–	–	0.09	0.46*	0.36*	0.09	0.52*	0.38*
							(1.16)	(5.18)	(3.36)	(1.11)	(6.27)	(3.60)
R^2	0.955	0.954	0.955	0.955	0.954	0.953	0.955	0.954	0.954	0.955	0.955	954
SIC	17.520	17.40	17.668	17.574	17.395	17.676	17.517	17.401	17.414	17.514	17.393	17.407

*, **, and *** represent the levels of significance at 1%, 5%, and 10%, respectively. Also, the values of *t*-statistic are reported in the parentheses

employed the Levenberg–Marquardt algorithm. We chose it over the SBPGDA due to its faster convergence. The Levenberg–Marquardt algorithm combines the advantages of the steepest gradient descent (stability when used with small learning rate) and the Gauss–Newton algorithm (fast convergence) (Yu and Wilamowski 2011).

AWNN model

Similar to the first step of the hybrid model B, we first decomposed the EC into an approximate part and five detail parts based on the mother wavelet. Similar to Fig. 9, ACF analysis on EC series was carried out by considering the entire series as well as the subsets of series at different intervals. AWNN architecture is similar to the one in the hybrid model B, with the exception that the approximate time series was modeled and forecasted by ARMAX-EGARCH model in the hybrid model B. Also in the AWNN model, the lag days (based on ACF) were taken into account in the input layer for detail parts, but in proposed model B, in addition to the lag days, we applied six factors (exogenous variables) in the input layer for detail parts.

Result

Error measures

In previous studies, in addition to the mean and variance of error time series, usually two evaluation criteria have been applied to examine the prediction accuracy of different models. They are root mean squared error (RMSE) and mean absolute percentage error (MAPE). In general form, RMSE and MAPE are computed as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{pi}-y_{ai}\right)^2}, \tag{22}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{y_{pi}-y_{ai}}{y_{ai}}\right| \times 100\right), \tag{23}$$

where, y_{ai} and y_{pi} represent the actual and predicted values, respectively. The evaluation criteria of prediction accuracy (RMSE and MAPE) are used to examine the prediction accuracy of different models. For both of the adopted measuring indices, the smaller the values, the more accurate the forecasts.

After calculating the error measures of the forecasting result, if the difference of RMSE and MAPE between the models is small, it is difficult to decide whether the result is due to chance or decisive. In fact, the answer cannot be simply concluded from the RMSE and MAPE values. As a result, whether the difference of forecasting performances is significant in the statistic view cannot be efficiently judged by the traditional evaluation criteria. To solve this problem, some statistic evaluation methods, such as Diebold–Mariano (DM), the asymmetric Diebold–Mariano (ADM), and Morgan–Granger–Newbold (MGN) tests, which offer a quantitative method to evaluate the forecast accuracy of EC forecasting models, were applied in this study.

The Diebold–Mariano test is based on the loss differentials, d_t:

$$d_t = L(e_{1.t}) - L(e_{2.t}) \tag{24}$$

where, $L(e_{1.\,t})$ and $L(e_{2.\,t})$ are the squared (or absolute) loss functions of forecasting models 1 and 2, respectively. Equivalently, the null hypothesis of equal predictive accuracy is shown as $H_0 : E[d_t] = 0$. The DM test statistic is:

$$DM = \frac{\overline{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{T}}} \rightarrow N(0,1) \tag{25}$$

where, $\overline{d}$ is the sample mean loss differential and $2\pi\hat{f}_d(0)$ is a consistent estimator of the asymptotic variance of $\sqrt{T}\,\overline{d}$.

Limited to symmetric structure, neither the squared-error loss nor the absolute-error loss can be an adequate description of the forecasting environment. In this case, the asymmetric loss function may help evaluate the forecasting accuracy. As a result, to make it practical, the DM tests based on the asymmetric loss functions, namely asymmetric DM tests, were applied in this study. Chen et al. (2014) proposed an asymmetric loss function as follows:

$$L_a(e_{i,t}) = \begin{cases} a\left|e_{i,t}\right|^p & if \quad e_{i,t} \geq 0 \\ \left|e_{i,t}\right|^p & if \quad e_{i,t} < 0 \end{cases} \tag{26}$$

where, p is a positive integer valued power parameter and a is the asymmetric index parameter. In this study,

we considered $p = 2$ and $a = 2$. Indeed, the asymmetric loss penalizes large positive forecasting errors, $e_{i,t}$.

The third test that we consider is the MGN test. Let $x_t = (e_{1,t} + e_{2,t})$ and $z_t = (e_{1,t} - e_{2,t})$. The null hypothesis of equal forecast accuracy is equivalent to zero correlation between x and z ($\hat{\rho}_{xy} = 0$). The MGN test statistic is as follows:

$$MGN = \frac{\hat{\rho}_{xy}}{\sqrt{\frac{1-\hat{\rho}_{xy}^2}{T-1}}} \tag{27}$$

where, $\hat{\rho}_{xy}$ is the contemporaneous correlation between x and y. If the forecasts are equally accurate, then the correlation between x and y will be zero.

Forecasting performance

Figures 11 and 12 show the actual and forecasted amounts of EC obtained by the proposed and benchmark models for 60 days ahead in two separate seasons (summer and winter). As mentioned above, the amount of EC in weekends is lower than in working days. This is because the business and social activities on working days are at a higher level than on weekends. According to Figs. 11 and 12, the proposed models accurately match with the actual data in working days as well as weekends (except mild drop), whereas the ARMAX-EGARCH-in-mean gives slightly more errors particularly in working days and weekends compared with other models.

The overall statistical forecasting performance measures (RMSE, MAPE, mean, and variance of the error time series) for EC forecasts using all the considered forecasting models for the winter and summer tests are presented in Table 7. Low values of RMSE, MAPE, mean, and variance of the error time series compared with the proposed models as well as the AWNN, NN, and ARMAX-EGARCH-in-mean models indicate the high forecast accuracy of model B for both of the summer and winter tests.

In Table 7, the proposed model B has the smallest RMSE, MAPE, mean, and variance of the error time series for both summer and winter tests. After model B, the proposed model A is also a good contender as it produces lower values of four statistical error measures than other benchmark models in the two test periods. Furthermore, based on RMSE and MAPE, the prediction performance of AWNN model is better than those of the NN and ARMAX-EGARCH-in-mean. Therefore, the proposed hybrid model forecast achieves better results than the single forecast. In addition, we learn that the consideration of exogenous variables series in modeling the approximate part and detail parts and non-linear characteristics would help improve the forecast accuracy. Also, consideration of each detail part in the structure of AWNN provides more accurate prediction than the integrated detail parts. One of the most important reasons is that, in addition to the calendar (7 days a week, 12 months a year and special days) and temperature (min, average, and max) variables, we applied the lag days (based on ACF) for each detail part. We used the ability of AWNN model to improve

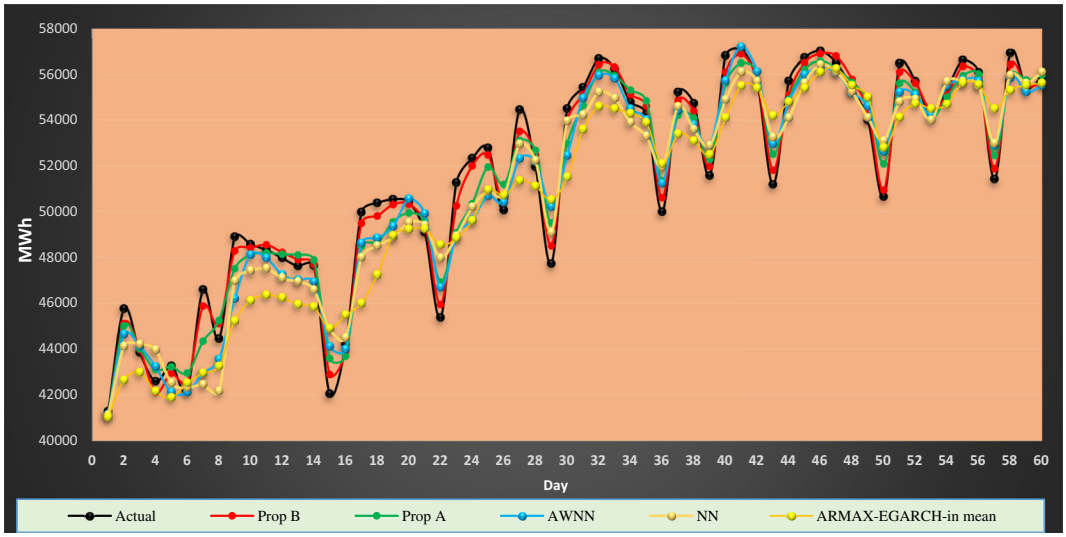

Fig. 11 Summer test (June 01, 2018–July 30, 2018)

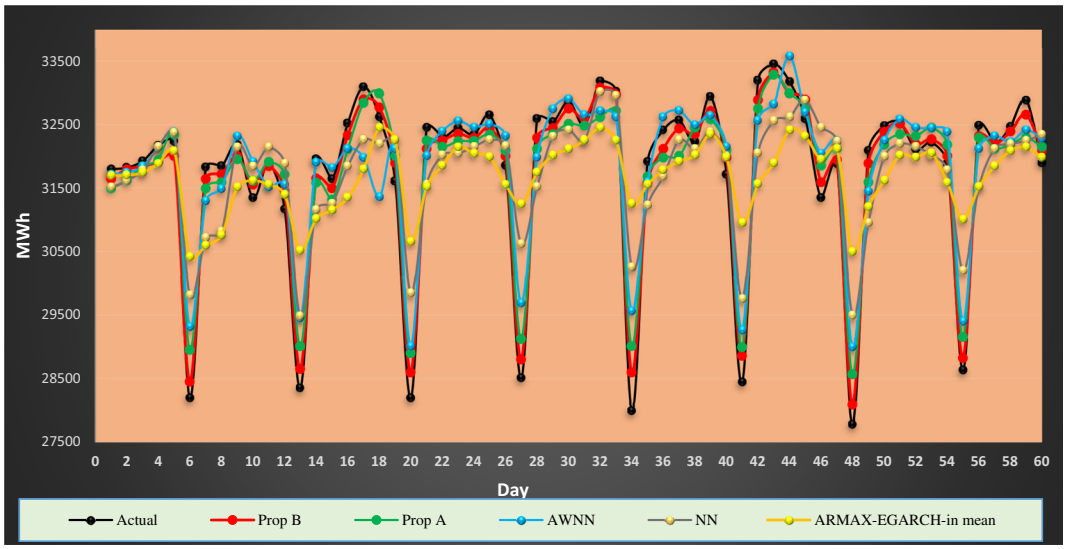

Fig. 12 Winter test (December 02, 2018–January 30, 2019)

prediction accuracy of the detail parts. The proper modeling of the approximate time series based on ARMAX-EGARCH-in-mean model with GED distribution would help improve the forecast accuracy of approximate part. So, in the mentioned reasons, the proposed model B has more prediction accuracy compared with the other benchmark forecast models.

Forecasting evaluation

We performed three different forecast-encompassing tests to determine whether the differences in forecast performance are statistically significant. We considered three different tests, since the underlying assumptions of each test may not hold in every case.

In this section, the forecasting performance of the five models is compared by DM, SDM, and MGN tests. Using the classical version of the DM test, the forecasting comparison of the proposed model B with other forecasting models is summarized in Table 8. The zero hypothesis, $H_0 : E[L(e_{1,t})] = E[L(e_{2,t})]$, means that the observed difference between the performance of two forecasting models (models 1 and 2) is not significant, while the alternative hypothesis, $H_1 : E[L(e_{1,t})] \neq E[L(e_{2,t})]$, means that the observed difference between the performances of two forecasting models is significant.

According to the DM and ADM tests based on the absolute-error loss and the squared-error loss, since the statistic values are more than |1.96|, the zero hypothesis is rejected at the 5% level of significance. In other words, the observed differences are significant and the forecasting accuracy of model B is better than the other benchmark models in both winter and summer tests.

Table 7 Overall statistical error measures for EC forecasts

Models	Winter test				Summer test			
	Mean error (MWh)	Variance error (MWh)	RMSE	MAPE	Mean error (MWh)	Variance error (MWh)	RMSE	MAPE
Model B	− 8.2	49,195	220.1	1.207	− 42.9	197,580	442.9	0.759
Model A	− 6.8	158,643	395	2.506	− 153.4	931,032	969.1	1.575
AWNN	94.7	313,824	563.5	3.029	− 390	1,429,459	1248.1	1.919
NN	− 500.9	15,105,660	3886.5	3.650	− 487.1	1,818,183	1423.1	2.314
ARMAX-EGARCH-in-mean	− 70.7	1,304,392	1134.7	5.082	− 821.1	3,111,343	1932.6	3.204

Table 8 Forecasting evaluation test based on model B and benchmark models

Models	Summer test				Winter test			
	Model A	AWNN	NN	ARMAX-EGARCH-in-mean	Model A	AWNN	NN	ARMAX-EGARCH-in-mean
DM-SE	− 4.561	− 4.063	− 5.143	− 6.704	− 7.93	− 9.297	− 11.06	− 10.532
p value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DM-AE	− 5.988	− 7.361	− 8.647	− 9.101	− 13.320	− 8.916	− 1.997	− 12.920
p value	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00
ADM-SE	− 3.97	− 3.494	− 3.758	− 6.378	− 3.360	− 3.228	− 3.837	− 4.888
p value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ADM-AE	− 5.172	− 5.431	− 6.212	− 8.344	− 8.948	− 7.43	− 1.985	− 11.185
p value	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00
MGN	− 14.46	− 14.04	− 14.07	− 22.10	− 11.226	− 12.131	− 67.349	− 32.211
p value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

DM-SE (*ADM-SE*) denotes the DM (ADM) test's statistics based on squared-error loss, and *DM-AE* (*ADM-AE*) denotes the DM (ADM-AE) test's statistics based on absolute-error loss

Similarly, according to the MGN test, the null hypothesis is rejected at the 1% level of significance. So, the forecasting accuracy of model B is better than the other benchmark models in both winter and summer tests.

Discussion

In the present works, we seek to find out why proposed model B (and model A) provides a more accurate prediction than the other models. First, the proposed models considered the conditional heteroscedasticity, leverage, and feedback effect in modeling the approximate time series under normal, Student *t*, and GED distributions. Although the original GARCH models are based on the Gaussian distribution, significant evidence suggests that the time series is rarely Gaussian. They are typically leptokurtic and exhibit fat-tail behavior. Therefore, each of these models will be estimated under three types of distribution including normal, Student *t*, and GED.

As shown in Table 3, the ARMAX-EGARCH-in-mean model with GED distribution was selected as the best fitted model for the approximate time series as follows:

$$\begin{aligned}\text{Conditional mean}: y_t &= 31121 + 0.90y_{t-1} \\ &+ 0.16y_{t-364} + 0.88u_{t-1} \\ &+ 0.03ave_t + 0.02\sigma_t \\ &+ \epsilon_t, \end{aligned} \quad (28)$$

$$\begin{aligned}\text{Conditional variance}: \ln\left(\sigma_t^2\right) \\ = 0.14 + 0.68\left|\frac{\epsilon_{t-p}}{\sigma_{t-p}}\right| - 0.01\frac{\epsilon_{t-p}}{\sigma_{t-p}} + 0.88\left(\sigma_{t-1}^2\right)\end{aligned} \quad (29)$$

In conditional mean equation, the approximate time series (y_t) is significantly affected by the previous amount of last day and year. Also, the air temperature has a positive and significant effect on it. The significant coefficient of σ_t proves the existence of feedback effect. So, we learn that the consideration of EC standard deviation would help improve the modeling and forecasting strength. Additionally, σ_t is a measure of uncertainty toward EC. Smaller standard deviation of error time series shows less uncertainly in EC. It means that with increase of the volatility of the electricity consumption, the average electricity consumption will increase. In this model, the coefficients of some calendar dates (such as weekly and monthly effects) were insignificant that it is logical because the weekly fluctuations are filtered by mother wavelet. So, we removed them from the conditional mean model. However, the annual effect and temperature had a significant effect on the approximate time series.

Since the EGARCH-mean was selected as the best fitted model, it supports inverse leverage effects and reveals the fact that in the context of electricity consumption, the impact of a negative shock to EC is lower than that of a positive shock. The positive or negative shock can be due to some events such as the announcement of a new

electricity price, technological progress, and unplanned outage. In fact, the log of conditional variance implies that the asymmetric effect is exponential and that the forecasts of conditional variance are non-negative (Tan et al. 2010). In fact, the asymmetry in the conditional variance is captured by the coefficient of $\frac{\epsilon_{t-p}}{\sigma_{t-p}}$ that is positive and implies an inverse leverage effect. If its coefficient is negative, it suggests a leverage effect.

In analyzing each detail part, we applied wavelet transform and AWNN model. The need for real-time processing led to use of AWNN, whose efficiency is determined by the rate of convergence of the learning algorithms. So, we benefited from this advantage of AWNN.

For detailed time series, we used the ACF analysis and identified the number of lags for each detail part. This helps to better identify the trend of each detail part. In addition to the lag days (based on ACF), we fed other six factors (including 7 days a week, 12 months a year, temperature (min, average, and max), and special days) into AWNN. So, one of the advantages of this model is that there are a lot of inputs in the input layer of the structure of each detail part.

The motivation is that EC at a given time can be seen as a combination of components with different frequencies (e.g., the low-frequency components representing the high regular patterns, and the high-frequency components representing the irregular fluctuations). In identifying these components, combining them with other models, considering the number of exogenous variables, and predicting them separately, we might be able to build more accurate prediction models.

Conclusion

In a deregulated electricity market, forecasting EC is critical for the market participants. However, due to some complex factors, it is difficult to predict EC accurately. According to the literature and empirical studies, the prediction methods based on hybrid models outperform the traditional models such as neural networks and econometric models (Szoplik 2015). In this paper, we proposed a hybrid forecast approach that combines the AWNN model with ARIMA-GARCH family models and uses the effective exogenous variables on electricity consumption as a prediction algorithm. In addition, the proposed models take into account the impact of calendar and weather factors on EC.

In the proposed models, at first, EC is decomposed into an approximate part and five detail parts based on the mother wavelet. Then, the wavelet components are modeled based on the AWNN and ARMA/ARMAX-GARCH family models. Afterwards, they predict the components separately. Finally, the results are combined. We compared the performances of the proposed models with those of the ARMAX-EGARCH-in-mean, NN, and AWNN models for Iran electricity market for 60 days ahead in two separate seasons. The EC prediction results showed that the proposed models outperform the existing approaches. Certainly, our results are based on only one electricity market; hence, it is necessary to conduct more empirical analyses on other electricity markets in the future. Additionally, EC probably depends on other weather variables such as humidity and sunlight, and we may investigate the possibility of including them as exogenous variables in our future works.

References

Abedinia, O., & Amjady, N. (2015). Day-ahead price forecasting of electricity markets by a new hybrid forecast method. *Modeling and Simulation in Electrical and Electronics Engineering, 1*(1), 1–7.

Amjady, N., & Keynia, F. (2008). Day ahead price forecasting of electricity markets by a mixed data model and hybrid forecast method. *International Journal of Electrical Power & Energy Systems, 30*(9), 533–546.

Bashir, Z. A., & El-Hawary, M. E. (2009). Applying wavelets to short-term load forecasting using PSO-based neural networks. *IEEE Transactions on Power Systems, 24*(1), 20–27.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307–327.

Bouzari, H., Šramek, M., Mistelbauer, G., & Bouzari, E. (2011). Robust adaptive wavelet neural network control of Buck converters. In *Recent advances in robust control-novel approaches and design methods*. IntechOpen.

Box, G. E., & Jenkins, G. M. (1976). Time series analysis: forecasting and control, revised ed. Holden-Day.

Chaâbane, N. (2014). A hybrid ARFIMA and neural network model for electricity price prediction. *International Journal of Electrical Power & Energy Systems, 55*, 187–194.

Chand, S., Kamal, S., & Ali, I. (2012). Modelling and volatility analysis of share prices using ARCH and GARCH models. *World Applied Sciences Journal, 19*(1), 77–82.

Chen, Y., Luh, P. B., Guan, C., Zhao, Y., Michel, L. D., Coolbeth, M. A., … & Rourke, S. J. (2010). Short-term load forecasting: similar day-based wavelet neural networks. *IEEE Transactions on Power Systems, 25*(1), 322–330.

Chen, H., Wan, Q., & Wang, Y. (2014). Refined Diebold-Mariano test methods for the evaluation of wind power forecasting models. *Energies, 7*(7), 4185–4198.

Conejo, A. J., Plazas, M. A., Espinola, R., & Molina, A. B. (2005). Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Transactions on Power Systems, 20*(2), 1035–1042.

Ding, S., Hipel, K. W., & Dang, Y. G. (2018). Forecasting China's electricity consumption using a new grey prediction model. *Energy, 149*, 314–328.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society, 50*, 987–1007.

Erdogdu, E. (2010). Natural gas demand in Turkey. *Applied Energy, 87*(1), 211–219.

Hickey, E., Loomis, D. G., & Mohammadi, H. (2012). Forecasting hourly electricity prices using ARMAX–GARCH models: An application to MISO hubs. *Energy Economics, 34*(1), 307–315.

Hong, T. (2010). Short term electric load forecasting.

Hong, T., Pinson, P., & Fan, S. (2014). Global energy forecasting competition 2012.

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters, 6*(3), 255–259.

Kavaklioglu, K., Ceylan, H., Ozturk, H. K., & Canyurt, O. E. (2009). Modeling and prediction of Turkey's electricity consumption using artificial neural networks. *Energy Conversion and Management, 50*(11), 2719–2727.

Kim, K. H., Park, J. K., Hwang, K. J., & Kim, S. H. (1995). Implementation of hybrid short-term load forecasting system using artificial neural networks and fuzzy expert systems. *IEEE Transactions on Power Systems, 10*(3), 1534–1539.

Kristjanpoller, W., & Minutolo, M. C. (2018). A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis. *Expert Systems with Applications, 109*, 1–11.

Liu, H., & Shi, J. (2013). Applying ARMA–GARCH approaches to forecasting short-term electricity prices. *Energy Economics, 37*, 152–166.

Mandal, P., Haque, A. U., Meng, J., Srivastava, A. K., & Martinez, R. (2013). A novel hybrid approach using wavelet, firefly algorithm, and fuzzy ARTMAP for day-ahead electricity price forecasting. *IEEE Transactions on Power Systems, 28*(2), 1041–1051.

May, R., Dandy, G., & Maier, H. (2011). Review of input variable selection methods for artificial neural networks. In *Artificial neural networks-methodological advances and biomedical applications*. InTech.

Morales-Acevedo, A. (2014). Forecasting future energy demand: electrical energy in Mexico as an example case. *Energy Procedia, 57*, 782–790.

Nury, A. H., Hasan, K., & Alam, M. J. B. (2017). Comparative study of wavelet-ARIMA and wavelet-ANN models for temperature time series data in northeastern Bangladesh. *Journal of King Saud University-Science, 29*(1), 47–61.

Pindoriya, N. M., Singh, S. N., & Singh, S. K. (2008). An adaptive wavelet neural network-based energy price forecasting in electricity markets. *IEEE Transactions on Power Systems, 23*(3), 1423–1432.

Pindoriya, N. M., Singh, S. N., & Singh, S. K. (2010). Forecasting of short-term electric load using application of wavelets with feed-forward neural networks. *International Journal of Emerging Electric Power Systems, 11*(1).

Rahman, S., & Hazim, O. (1993). A generalized knowledge-based short-term load-forecasting technique. *IEEE Transactions on Power Systems, 8*(2), 508–514.

Rana, M., & Koprinska, I. (2016). Forecasting electricity load with advanced wavelet neural networks. *Neurocomputing, 182*, 118–132.

Reis, A. R., & Da Silva, A. A. (2005). Feature extraction via multiresolution analysis for short-term load forecasting. *IEEE Transactions on Power Systems, 20*(1), 189–198.

Reston Filho, J. C., Affonso, C. D. M., & de Oliveira, R. C. (2014). Energy price prediction multi-step ahead using hybrid model in the Brazilian market. *Electric Power Systems Research, 117*, 115–122.

Sánchez, J. M. B., Lugilde, D. N., de Linares Fernández, C., de la Guardia, C. D., & Sánchez, F. A. (2007). Forecasting airborne pollen concentration time series with neural and neuro-fuzzy models. *Expert Systems with Applications, 32*(4), 1218–1225.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136.

Shafie-Khah, M., Moghaddam, M. P., & Sheikh-El-Eslami, M. K. (2011). Price forecasting of day-ahead electricity markets using a hybrid forecast method. *Energy Conversion and Management, 52*(5), 2165–2169.

Sumer, K. K., Goktas, O., & Hepsag, A. (2009). The application of seasonal latent variable in forecasting electricity demand as an alternative method. *Energy Policy, 37*(4), 1317–1322.

Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks. *Energy, 85*, 208–220.

Tan, Z., Zhang, J., Wang, J., & Xu, J. (2010). Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. *Applied Energy, 87*(11), 3606–3610.

Thomas, S., & Mitchell, H. (2005). GARCH modeling of high-frequency volatility in Australia's National Electricity Market. *System*, 1–39.

Valenzuela, O., Rojas, I., Rojas, F., Pomares, H., Herrera, L. J., Guillén, A., … & Pasadas, M. (2008). Hybridization of intelligent techniques and ARIMA models for time series prediction. *Fuzzy Sets and Systems, 159*(7), 821–845.

Wu, L., & Shahidehpour, M. (2010). A hybrid model for day-ahead price forecasting. *IEEE Transactions on Power Systems, 25*(3), 1519–1530.

Yu, H., & Wilamowski, B. M. (2011). Chapter 12. In *Levenberg–Marquardt training industrial electronics handbook* (pp. 12–11). CRC.

Zhai, M.-Y. (2015). A new method for short-term load forecasting based on fractal interpretation and wavelet analysis.

International Journal of Electrical Power & Energy Systems, 69, 241–245.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50*, 159–175.

Zhang, J., Tan, Z., & Yang, S. (2012). Day-ahead electricity price forecasting by a new hybrid method. *Computers & Industrial Engineering, 63*(3), 695–701.

Zolfaghari, M., & Sahabi, B. (2017). Impact of foreign exchange rate on oil companies risk in stock market: a Markov-switching approach. *Journal of Computational and Applied Mathematics, 317*, 274–289.