INTERNATIONAL TRANSACTIONS ON ELECTRICAL ENERGY SYSTEMS
Int. Trans. Electr. Energ. Syst. 2014; **24**:736–760
Published online 25 February 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/etep.1734

A hybrid electricity price forecasting model for the Nordic electricity spot market

Sergey Voronin[1*,†], Jarmo Partanen[1] and Tuomo Kauranne[2]

[1]*LUT Energy, Laboratory of Electricity Market and Power Systems, Lappeenranta University of Technology, P.O.Box 20, 53851, Lappeenranta, Finland*
[2]*LUT Mafy, Laboratory of Technomathematics, Lappeenranta University of Technology, P.O.Box 20, 53851, Lappeenranta, Finland*

SUMMARY

A hybrid electricity price forecasting model for the Finnish electricity spot market is proposed. The daily electricity price time series is analyzed in two layers – normal behavior and spiky behavior. Two different data preprocessing techniques are applied to handle trend and seasonality in the time series. An ARMA-based model is used to catch the linear relationship between the normal range price series and the explanatory variable, a GARCH model is used to unveil the heteroscedastic character of residuals and a neural network is applied to present the nonlinear impact of the explanatory variable on electricity prices and improve predictions based on time series techniques. The probability of a price spike occurrence and the value of a price spike are produced by a Gaussian Mixture model and K-nearest neighboring model, respectively. Forecasts of normal range prices and price spikes are generated to form an overall price forecast up to one week ahead. The results show that hybridization of the normal range price and price spikes forecasts may provide comprehensive and valuable information for electricity market participants. Copyright © 2013 John Wiley & Sons, Ltd.

KEY WORDS: electricity price forecast; price spikes; ARMA; GARCH; neural network; Gaussian mixture; model; K-nearest neighboring

1. INTRODUCTION

Electricity price forecasting has become an important area of research in the aftermath of the worldwide deregulation of the power industry that launched competitive electricity markets embracing all market participants, including generation and retail companies, transmission network providers and market managers.

Companies trading in electricity markets require accurate electricity price forecasts to bid and hedge against price volatility and, thus, maximize their profitability. Certain unique characteristics of electricity markets make electricity price forecasting more complex than price forecasting of other commodities. Electric power cannot be stored economically, and transmission congestion influences the exchange of power. Unlike electricity demand time series, electricity price series can exhibit variable means, major volatility and significant outliers. Due to the extreme volatility reflected in so-called price spikes, electricity spot price modeling and forecasting face a number of challenges. Applications used to forecast the prices of other commodities are only of limited validity in electricity price forecasting and can produce large errors.

There is a clear need for a method that is able to predict price spikes, in addition to predicting the normal price range with a high degree of accuracy.

*Correspondence to: Sergey Voronin, Laboratory of Electricity Market and Power Systems, Lappeenranta University of Technology, Lappeenranta, Finland.
†E-mail: sergey.voronin@lut.fi

The paper is organized as follows. Section 2 reviews recent studies in the field. Section 3 specifies the market time framework used to forecast day-ahead prices. Section 4 briefly examines factors driving electricity price behavior and provides a definition of a price spike. Section 5 presents the mathematical framework of the study. The proposed hybrid model is described in Section 6. Section 7 gives forecasting results. The paper ends with conclusions and suggestions for future work.

2. LITERATURE REVIEW

Based on the needs of the market, a variety of approaches have been proposed in the last decades.

The first group of models applied to electricity price forecasting within the context of competitive electricity markets is based on simulation of power system equipment (transmission congestions, losses, etc.) and related cost information (marginal generation costs, heat rates or fuel efficiencies) [1,2]. A major drawback with this approach is the requirement of a large amount of real-time data about existing equipment. Nevertheless, the simulation methods presented could very well be effective if used by market operators and regulators who have the authority to collect reliable and accurate equipment and operational information.

The second group is game-theory-based models which focus on the impact of bidder strategic behavior on electricity prices. It has been claimed that spot market prices are closely related to the bidding and pricing strategies of market participants [3–5].

The third approach is based on stochastic modeling of finance. A modified version of geometric Brownian motion was proposed in [6] as a jump-diffusion model for the stochastic modeling of spot power prices. The robustness of various diffusion models when applied to electricity prices has been evaluated in [7]. The main conclusion of this work is that geometrical mean reverting jump-diffusion models provide the best performance and that all models without jumps appear inappropriate for modeling electricity prices. It should be noted that the main disadvantage of stochastic modeling approaches arises from difficulties in incorporating physical characteristics of power systems, such as losses and transmission congestions, into mathematical finance models, which may produce significant mismatch between the model output and the real power market.

The fourth approach is based on time series models and includes two major branches: regression-based models and artificial intelligence models such as artificial neural networks (ANN) and fuzzy logic. Regression models are considered to be functions of past price observations and exogenous explanatory variables such as electricity demand and meteorological conditions. Much work has been done on electricity demand and electricity price forecasting with an ARMA approach and transfer function. Five short-term forecasting techniques were analyzed and evaluated in [8], and autoregressive integrated moving averages (ARIMA) have been applied to load forecasting [9,10]. Transfer function, dynamic regression and ARIMA models to forecast next day electricity prices, using Spanish and Californian markets as test cases, were proposed in [11,12]. To overcome the restrictions of linear models and account for nonlinear patterns observed in real problems, several classes of nonlinear models have been proposed. These include threshold autoregressive models [13,14] and the autoregressive conditional heteroscedastic (ARCH) model of Engle [15] and its extended version GARCH [16–19]. More recently, ANNs have been suggested as an alternative to time series forecasting [20,21]. The main strength of ANNs is their flexible nonlinear modeling capability [22].

Linear regression based models and ANN models have both achieved successes in their own linear or nonlinear domains. However, none of them is a universal model that is suitable for all circumstances. For example, the approximation of ARMA models to complex nonlinear problems may not be adequate, and the use of ANNs to model linear problems has yielded mixed results.

Since it is difficult to thoroughly know the characteristics of the data in a real problem, hybrid methodology that has both linear and nonlinear modeling capabilities would appear to be a possibly productive strategy for practical use. By combining different models, different aspects of the underlying patterns may be captured. In [23], researchers compared various adaptive and non-adaptive linear and potentially nonlinear models and concluded that hybrid models consisting of multivariate adaptive linear and nonlinear models outperform other models for many variables. A model combining a neural network model with a seasonal time series ARIMA model has been developed in [24]. The model

DOI: 10.1002/etep

outperformed single ARIMA and ANN models in terms of performance accuracy measures. A hybrid model for day-ahead price forecasting composed of linear (ARMAX) and nonlinear (ANN) relationships of prices and explanatory variables such as electricity demand was developed in [25]. A day-ahead price forecasting model was implemented on the basis of a hybrid intelligent system in [26].

While most existing approaches to forecasting electricity prices are reasonably effective for normal range electricity prices, they disregard price spike events. Price spikes are caused by a number of complex factors and exist during periods of market stress. These stressed market situations are associated with extreme meteorological events, unusually high demand or, more often, unexpected shortfalls in supply, caused for example by generator failures [27]. In early research, price spikes were truncated before application of the forecasting model to reduce the influence of such observations on the estimation of the method parameters; otherwise, a very large forecast error would be generated at price spike events [28,29]. Electricity price spikes, however, are significant for electricity retailers, who cannot pass the spikes onto final customers, and improved analysis of spikes is thus important for risk management. Spikes were incorporated into a Markov-switching model in [27] by proposing different regimes: regular and spiky. Spikes were introduced into diffusion models in [18] by the addition of a Poisson jump component with time varying parameters. Data mining techniques have been applied to the spike forecasting problem and have achieved promising results [30,31]. These approaches have shown high forecasting accuracy and robustness, even with incomplete information and noisy market data. Data mining techniques based on a Bayesian classifier significantly outperformed other alternative techniques (decision tree, ANN, winnow, support vector machine (SVM) and K-nearest neighboring (K-nn)) for prediction accuracy and decision benefits [32].

Since there is little published research in the area of price forecasting in the Nordic electricity market, and especially in the Finnish electricity market, accurate and robust price forecast methods are of considerable interest. The novelty of this work lies in combining of existing methods to build a new comprehensive model for overall electricity price forecasting. Such an investigation of the Finnish electricity market has not been undertaken in the literature to date.

The paper presents a hybrid model for forecasting of daily electricity spot prices up to one week ahead. The hybrid model consists of two modules which are, respectively, used to predict electricity prices within a normal range and during price spikes. Depending on the data preprocessing approach, the first module utilizes an ARMAX or SARIMAX model to catch the linear relationship between the normal range price series and explanatory variable. A GARCH model is applied to simulate the heteroscedasticity of the residuals that exists even after price spikes have been extracted from the initial price time series, and an ANN is built to combine the predictions from the ARMAX/SARIMAX-GARCH models with historical price and demand data to improve predictions based on time series techniques and produce a final normal range price forecast. The second module exclusively focuses on forecasting extreme price events such as price spike occurrences.

Based on the hybrid model, this paper proposes a classification-based spike prediction framework. The framework can predict both the occurrence and the values of price spikes. The model is calibrated and evaluated with Finnish electricity spot market prices of Nord Pool (Elspot) for the period 1 January 2006 to 31 December 2009.

3. FORECASTING FRAMEWORK

The time framework to forecast day-ahead market prices in the Nordic electricity market is illustrated in Figure 1. The market price forecasts for day T are required on day T-1 (bidding hour: 12 a.m. CET on day T-1). Data for day T-1 are available on day T-2 (clearing hour: around 1 p.m. CET on day T-2). In hourly resolution, the actual forecasting of market prices for day T can take place between the clearing hour for day T-1 of day T-2 and the bidding hour for day T of day T-1.

Therefore, in hourly and daily resolution, to forecast prices for day T, price data of day T-1 are considered known. In multistep ahead prediction, the predicted price value of the current step is used to determine its value in the next step.

DOI: 10.1002/etep

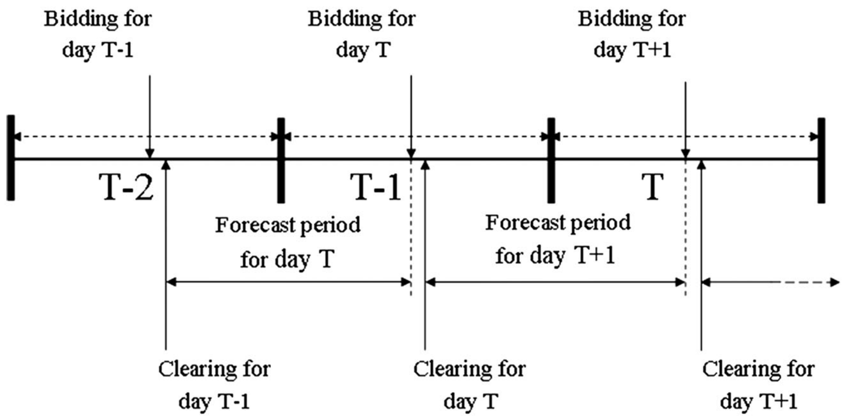

Figure 1. Time framework to forecast market prices for the day-ahead market.

4. FACTORS AFFECTING THE MARKET PRICE

Prices in the Nordic electricity market are highly volatile but are not purely stochastic and, therefore, can be explained, at least partly, by background variables. In order to get a high-performance forecasting model, adequate explanatory variables should be determined. Electricity prices are influenced by many factors, e.g. historical prices, electricity demand, weather conditions, imports, generation outages and operational reserves [21]. Obviously, some of the factors are more important than others.

The use of daily average data obtained from NordPool [33] is motivated by the search for the most appropriate and simple possible model relating electricity spot price characteristics and price spike occurrences to a limited number of exogenous factors. The use of average daily data simplifies the modeling of seasonal components of the series, and the influence of important driving factors is more easily captured in terms of the available data. The use of average daily data significantly reduces the computational costs when a number of different forecasting techniques is tested to select the most accurate approach and the set of external variables. In further work, the selected model could be extended to intraday price propagation keeping its robustness, but this would require more detailed modeling of the market structure.

4.1. Normal range prices

In all but completely regulated markets, the electricity price is strongly correlated with electricity demand. Electricity demand, therefore, is most often the input variable used to build price forecasting models [8,25,34]. Experiments demonstrate that including demand in price forecasting models can obtain significant improvement compared with the same model without demand [35].

A part of total electricity generation and consumption structures can be combined into so-called non-base electricity demand [36]. Hydro and nuclear power production are rather constant during the whole week and show low correlation with electricity prices. The non-base electricity demand is obtained by subtraction of nuclear power and hydro power generation from the total electricity demand. The new explanatory variable is the part of total electricity demand which is not covered by the base generation consisting of nuclear and hydro power generation.

Formally, in daily resolution, the explanatory variable is defined as:

$$Non - Base\ Demand(d) = System\ Demand(d) - Hydro\ power(d) - Nuclear\ power(d) \quad (1)$$

where $d = 1,\ldots,7$ (day of the week).

The value of non-base electricity demand is high if the electricity demand is high or there is a lack of base generation (values of hydro or/and nuclear power generation are low). The value of non-base electricity demand is low if electricity demand is low or there is a high level of hydro and/or nuclear power generation. This explanatory variable covers all possible cases and presents an adequate daily variation. Figure 2 shows the relationship between the variable and Finnish electricity spot prices from 1 January 2006 to 31 December 2009.

DOI: 10.1002/etep

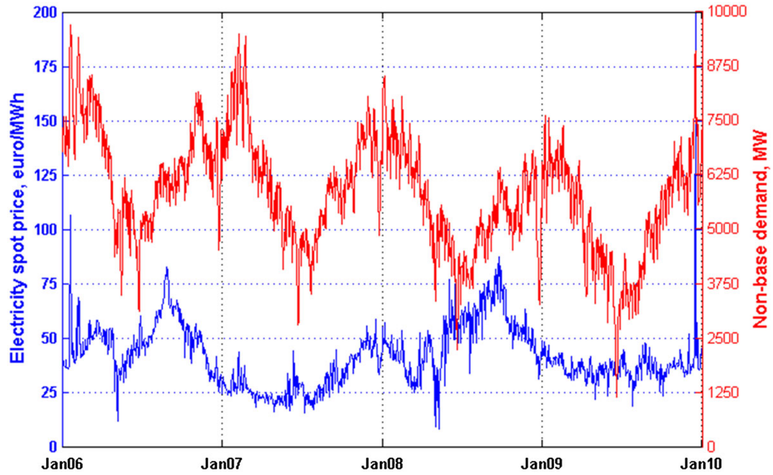

Figure 2. Finnish electricity spot price and non-base electricity demand.

In our paper, electricity demand and base generation forecasts were combined into the non-base electricity demand and used as an explanatory factor in the price forecasting model. The forecasting models for the total electricity demand and generation (i.e. internal supply) were based on the multiplicative double seasonal ARIMA model [37]. Values of hydro and nuclear power generation were forecasted by a simple random walk model.

Another very important source of variation in the electricity price is meteorology. Extreme weather conditions may cause an increase in electricity demand, which in turn may lead to a higher electricity price because more expensive production sources must be activated (see Figure 1). Difficult weather conditions are typical for the Nordic region, especially during winter time. Electricity demand is higher when the atmospheric temperature rises or falls from a base 'comfortable' level; temperature-dependent demand variations are more extreme if humidity is higher, since moisture increases the heat retention capability of air [38]. Atmospheric pressure variations generally cause air temperature variations and, as a consequence, load variations. The effect of temperature and other weather-related variables can be incorporated in the electricity demand, and they, therefore, were not used in the normal range price forecasting model to avoid variable collinearity.

Unit outage information, although clearly of importance, was not considered in the study because it is generally proprietary and not available to all market participants in real time.

4.2. Price spikes

In an ideal competitive electricity market, price spikes only occur when the demand exceeds supply. Most electricity markets, however, are not ideally competitive. Therefore, price spikes may happen even when the supply completely covers the demand. Electricity demand and supply, net interchange, i.e. the electricity transported from/to other countries, and a time index were chosen as inputs for the price spike forecasting model in [31]. The importance of electricity demand for electricity price forecasting was discussed in the previous section. The relationship between supply and price is similar to that of demand and price. This is because the power system requires supply and demand to be balanced constantly. A supply–demand relation, extreme weather conditions and transmission congestion events have been included in the list of main reasons causing price spikes in [30]. Although the case studies [30,31] were based on the Australian electricity market, the general approach can be considered to be applicable to other markets as well.

This study uses the composite relationship between spot price, demand and supply that was proposed in [30] and presented as a supply–demand balance index (SDI). The SDI on day T is defined by Equation (2):

$$SDI(T) = (Supply(T) - Demand(T))/Demand(T) \cdot 100\% \tag{2}$$

where *Demand*(*T*) is the market demand on day T, and *Supply*(*T*) is the electricity supply capacity on day *T*.

DOI: 10.1002/etep

In this paper, atmospheric temperature was chosen as a main indicator of weather extremity in electricity price spike study. The main electricity consumption areas in Finland are the south and central regions [39], and temperature data for Helsinki city were used because its geographical location indicates a temperature that is relevant to overall electricity consumption in the country. Temperature data forecasted for Helsinki city is available on the Weather Underground web site [40]. Capacity and power flow along the interconnectors between the Nordic countries and along the cables that connect the bidding areas in Norway are managed by Nord Pool Spot. Nord Pool Spot uses the transmission capacity to conduct power into high price areas and out of low price areas, thereby reducing the price in high price areas and raising the price in low price areas. Price differences between bidding areas (e.g. Finland and Sweden) occur when the surplus volume at the system price in one or more bidding areas is greater than the total export capacity from this area(s). The sales and purchase curves then have to be balanced taking the transmission capacity into account. This will lead to a relatively low price in the surplus area and a relatively high price in the deficit area – utilizing the maximum capacity between the areas [41]. The total Elspot power flow and the total Elspot capacity to Finland were calculated as a daily sum of the Elspot net exchange and Elspot capacities from Sweden, Norway and Estonia to Finland, respectively. Power flow data, and generation and demand data for the Finnish electricity system are available on the Fingrid web site [42]. Two regimes of the Finnish electricity system can be considered. One of the regimes is the regular regime; the other, the non-regular regime, is the capacity-limited regime and exists when the difference between the total Elspot power flow and the total Elspot capacity to Finland is close to zero. Congestion and thus extreme price changes are more likely to occur when the difference between the total Elspot power flow and the total Elspot capacity is small. Elspot power flow and capacity data for day T-1 are published by Fingrid and available on day T-2 (see Figure 1). Therefore, to forecast flow and capacity for day T, flow and capacity data of day T-1 are considered known. Elspot power flow data has strong seasonal patterns, which are captured by the time series forecasting approach. Elspot capacity is rather constant during the whole week.

4.2.1. Electricity price spike definition. For the purposes of study of price spikes, it is necessary to formulate a spike definition. A spike is defined as a price that surpasses a specified threshold. The main questions, therefore, are how high the threshold should be and whether the threshold should have a fixed or time-dependent value. The very volatile character of electricity price time series requires use of a varying threshold value instead of one global value to cut off global outliers. Specification of the boundary level is challenging. Some authors suggest the use of fixed log-price change thresholds [43] or a varying log-price range threshold [44]. This paper utilizes the methodology employed in [45], where the threshold value is based on the original price values. An electricity price spike is defined as a process realization which exceeds the price mean value in a specified window w by more than double standard deviation σ of the window. The defined spikes are extracted from the initial price series, and the mean value of the specified window w is used in their place. Based on the low frequency of the data and empirical results obtained in [18,45], $w = 90$ (3 months) was set as the value used to find electricity price spikes in the electricity price time series. The thus defined price spikes were extracted from the initial price series, as shown in Figures 3–5.

In addition to the physical factors given above, the day status of the sample needs to be implemented into the forecasting model. The whole data set was divided into weekdays, weekends and holidays. The following holidays in Finland were taken into account: Midsummer Day, Epiphany (Jan 6th), May Day, Ascension Day, Christmas, New Year and Independence Day (Dec 6th).

Table I shows basic distribution parameters for prices and spikes. It can be seen from the number of spikes that spikes constitute less that 1.5% of all the daily prices. However, their magnitude and unexpectedness cause them to have disproportionate significance in electricity markets. The statistics show that there is zero probability of an electricity price spike during weekends and holidays.

Based on the general analysis, the probability of price spikes has a close relationship with the electricity demand–supply balance, temperature, transmission congestions and day index. The distribution of the Finnish electricity spot prices versus the chosen driving factors is shown in Figure 6.

Note that the factors cannot exactly determine the occurrence of spikes. The Gaussian mixture model (GMM) proposed in this paper predicts spikes by evaluating their occurrence probability. The inputs of the model are not necessarily the determinants of spikes.

DOI: 10.1002/etep

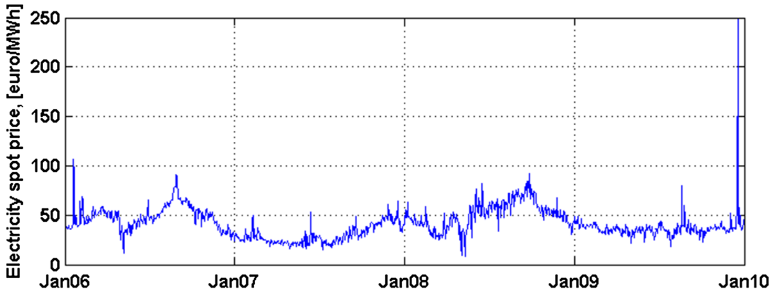

Figure 3. Original Finnish electricity spot price time series for 1 January 2006 – 31 December 2009.

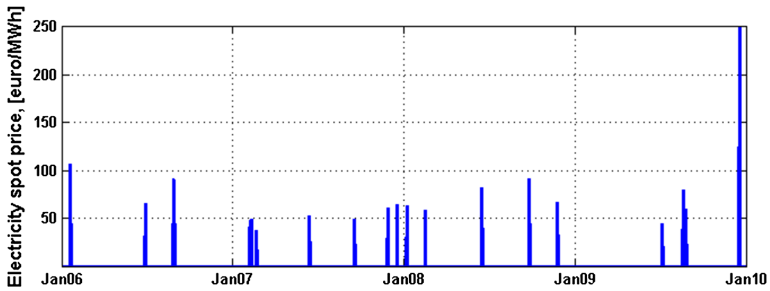

Figure 4. Extracted electricity price spikes for 1 January 2006 – 31 December 2009.

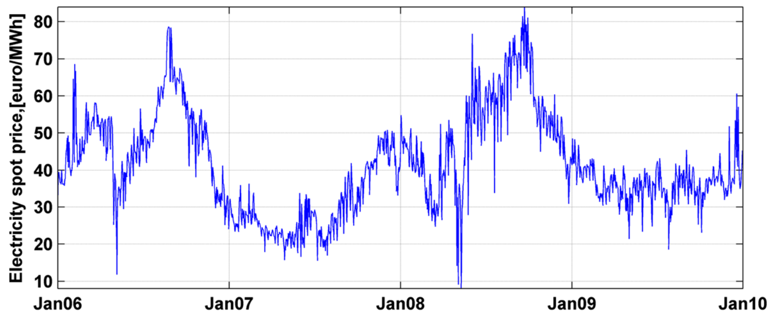

Figure 5. Normal range prices for 1 January 2006 – 31 December 2009.

5. METHODOLOGY

Before the prediction strategy is described, key features of ARMA, GARCH, neural networks, GMMs, the K-nn method and the seasonal decomposition technique are first introduced.

5.1. Normal range price forecasting

In normal range price forecasting the time series has to be detrended and deseasonalized before developing the forecasting models. Two competing approaches to handle trend and seasonality are

Table I. Basic statistics for Finland area spot price and price spikes.

	Number of observations	Mean	Std	Skewness	Kurtosis	Weekday (Numb. of spikes)	Weekend/ Holiday (Numb. of spikes)
Price	1436	41.26	13.28	0.55	2.91		
Spikes	25	71.86	41.88	3.26	14.65	25	0

DOI: 10.1002/etep

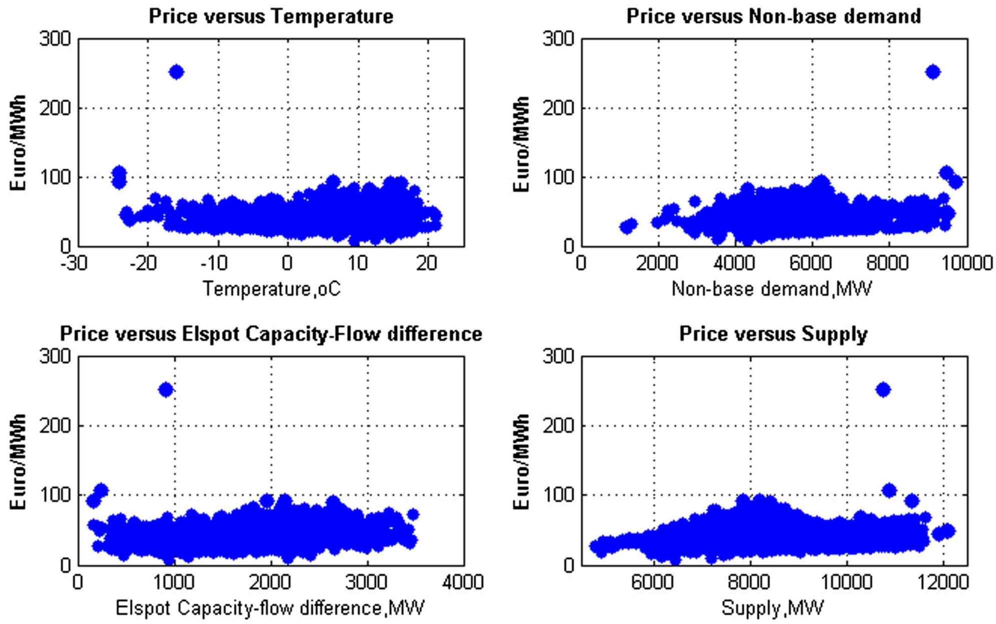

Figure 6. Scatter plots of Finnish electricity spot price versus potential price spike driving factors.

considered in this paper. In the first approach, trend and seasonality terms are directly captured by the forecasting model. A seasonal ARIMA model is a simple example of such an approach. Another approach to handle trend and seasonality in time series is application of additive or multiplicative decomposition. Both approaches were used in this work, and the forecasting performance of the model checked in each case.

5.1.1. Seasonal ARIMA model with explanatory variables. Data preprocessing is a standard requirement in the Box–Jenkins methodology for ARMA modeling. The seasonal ARIMA model with exogenous explanatory variables is based on a simple ARMA introduced in [46]. The ARMA model for a time series $\{X_t \mid t = 1, 2, \ldots, T\}$ is given in Equation (3):

$$\varphi_p(B)X_t = \theta_q(B)a_t \tag{3}$$

In the case of linear trend and/or seasonal behavior, non-stationary time series processes can be transformed by integration of the series to make them stationary. The model, therefore, is transformed to a seasonal autoregressive integrated moving average (SARIMA or SARIMAX). Without differencing operations, the model could be written more formally as

$$\varphi_p(B)\Phi_P\left(B^S\right)X_t = \theta_q(B)\Theta_Q\left(B^S\right)a_t \tag{4}$$

Since the model estimates differentiated data X_t and has external input U_t, it becomes a SARIMAX model by entering trend and seasonal integrating terms and an external variables term into the equation:

$$\varphi_p(B)\Phi_P\left(B^S\right)(1-B)^d(1-B)^D X_t = \eta_l(B)(1-B)^d(1-B)^D U_t + \theta_q(B)\Theta_Q\left(B^S\right)a_t \tag{5}$$

where B is a lag operator such as $BX_t = X_{t-1}$; $\varphi_p(B) = (1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p)$, $\theta_q(B) = (1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q)$, $\eta_l(B) = (1 + \eta_1 B + \eta_2 B^2 + \ldots + \eta_l B^l)$, $\Phi_P(B^S) = (1 - \Phi_1 B^S - \Phi_2 B^{2S} - - \ldots - \Phi_P B^{PS})$ and $\Theta_Q(B) = (1 + \Theta_1 B^S + \Theta_2 B^{2S} + \ldots + \Theta_Q B^{QS})$ are polynomials in B; p,q are regular orders of the AR and MA polynomials; P,Q are seasonal orders of the AR and MA polynomials; d is the number of regular differences, and D is the number of seasonal differences; l is an order of the external input polynomial and a_t is the estimated residual at time t.

The Box–Jenkins approach utilized uses an iterative model building strategy consisting of four stages. In the first stage, the structure of the model is identified. The basic principle of an ARMA process is that it should have theoretical autocorrelation properties. Utilization of the autocorrelation function (ACF) and the partial ACF (PACF) of the sample data as basic tools to identify the order

DOI: 10.1002/etep

of the ARMA best model, which is then estimated by maximum likelihood (ML) in the second step, has been proposed by Box and Jenkins [46]. The parameters are estimated such that an overall measure of errors is minimized. Goodness-of-fit tests on the estimated residuals is the third step. The a_t should be independently and identically distributed as normal random variables with mean = 0 and variance σ^2, and the roots $\varphi(X) = 0$ and $(X) = 0$ should all lie outside the unit circle. Plots of the residuals can be used to examine the adequacy of the tentative model to the historical data. If the model is not adequate, a new tentative model should be identified. Forecast future outcomes based on the known data are obtained in the fourth step.

5.1.2. Time series decomposition. Additive and multiplicative decomposition approaches allow separation of the underlying patterns in the data series from irregular components. In this paper, classical multiplicative time series decomposition is used:

$$X_t = T_t \cdot S_t \cdot I_t \tag{6}$$

where X_t is original data, T_t stands for the trend, and S_t and I_t for the seasonal and irregular components, respectively. The linear trend is estimated by the least-squares method, that is

$$T_t = a \cdot t + b \tag{7}$$

The identified trend is removed from the original data, and the remaining data are used for the seasonal decomposition. The number of seasonal indices equals the data seasonality order. In our case, these are 7 and 365 days. If there are different orders, decomposition is performed in several steps starting from the highest order. Then, for a given seasonality order k, its respective seasonal indices will be calculated as the following mean values

$$S_i = \mathrm{E}(S_i, S_{i+k}, S_{i+2k}, \ldots, S_{i+mk}) \tag{8}$$

where $i = 1,..,k$ and m is the number of all seasonal cycles within the total data horizon.

Eventually, the irregular component can be calculated by dividing the adjusted series by the trend and cycle components. The underlying patterns are projected into the future and used as the forecast.

5.1.3. GARCH model. ARMA models are used in many applied problems. The basic assumptions of the error terms of the models include zero mean and constant variance for σ, the traditional ARMA estimation. In practice, the homoscedasticity assumption of constant variance sometimes does not hold. Such time series are called heteroscedastic. Thus, when error terms are autocorrelated, the ordinary least-square estimator of ARMA model coefficients is no longer asymptotically unbiased and consistent. It is agreed that electricity price time series present nonconstant deviations over time as demonstrated in Figure 3. The GARCH model is an extension of the ARCH model introduced in [15] where conditional variance σ^2 is considered as time dependent:

$${\sigma_t}^2 = K + \sum_{i=1}^{m} \alpha_i {a_{t-i}}^2, \text{ at time } t, t = 1, 2, 3, \ldots, T \tag{9}$$

where $a_t = \varepsilon_t \sigma^2$ is an error term produced by ARMA at time t, $\varepsilon_t = N(0,1)$ and K is a variance constant.

As ε_t is white noise, which is assumed to be normally distributed, a_t will also be normally distributed with zero mean and variance σ^2. In practical applications, the current variance sometimes appears to be dependent not only on past squared disturbances, but the past variance of the errors as well. Such an extended model was introduced in [16] and comes as a GARCH(m,n) model

$${\sigma_t}^2 = K + \sum_{i=1}^{m} \alpha_i {a_{t-i}}^2 + \sum_{i=1}^{n} \beta_i {\sigma_{t-i}}^2, \text{ at time } t, t = 1, 2, 3, \ldots, T \tag{10}$$

The application of a GARCH model is an iterative procedure similar to the ARMA procedure and includes iteratively: order determination, parameter estimation and model diagnostic checking.

5.1.4. Neural network approach. A neural network, also called a multilayer perceptron, is a semi-parametric model and has been developed based on study of brain functions and the nervous system. Perceptrons are arranged in layers with no connections inside a layer, and each layer is fully connected

DOI: 10.1002/etep

to preceding and following layers without loops. The first and last layers are called input and output layers, respectively. Other layers are hidden layers. Each layer, therefore, consists of multiple computational elements, called neurons, which are connected to neurons in adjacent layers and capture complex nonlinear phenomena. A neural network should be trained by historical data of a time series to capture the characteristics of this time series. The output of each processing unit (or neuron) is propagated forward through each layer of the network using the equation

$$v_j = \sum_{i=1}^{N} x_i w_{ji} + w_{j0} = \sum_{i=0}^{N} x_i w_{ji} = w_j^T x \tag{11}$$

where v_j is a net activation for neuron j, w_{ji} is the weight on connection from the i^{th} to the j^{th} unit, x_i is input data from unit i to j, w_{j0} denotes a bias on the j^{th} unit and N is the total number of input units.

Neuron output is then $y_j = f(v_j)$ where $f(\cdot)$ is the transformation function that could be determined by a squashing (or sigmoid) function whose argument is v_j and whose value is between 0 and 1. A sigmoid function that is used in a hidden layer is in the form:

$$y_j = f\left(v_j\right) = 1/\left(1 + e^{-v_j}\right) \tag{12}$$

The procedure for developing neural networks is as follows: data preprocessing; definition of the architecture and parameters; weights initialization; training until the stopping criterion is reached (number of iterations, sum of squares of error is lower than a pre-determined value); finding the network with the minimum forecasting error on a validation data set that is usually 5–10% of the training data set; forecasting future outcomes.

An essential step before development of a neural network model is preprocessing of the input data. According to the computational intelligence literature, the use of deseasonalized data is very advantageous. Previous study shows that deseasonalization and detrending give significant improvement in the forecasting performance of neural network models [20]. Some authors have investigated whether prior statistical deseasonalization of data is necessary to produce more accurate neural network forecasts [47]. They concluded that neural networks were unable to adequately learn seasonality and their forecasting ability suffered accordingly. One possible reason for this finding was that if neural networks use deseasonalized time series, then the neural networks could focus on learning the trend and cyclical components. Neural networks that do not use deseasonalized time series have to learn trend and cyclical components, and seasonality. The latter is a more difficult task and requires a larger neural network. This work used several neural network models with different characteristics of input data, i.e. raw or deseasonalized and detrended, to check whether performance results support the previous study in the field.

The neural network toolbox of MATLAB was selected for ANN model building due to its flexibility and simplicity. The number of hidden nodes and size of the network were determined in order to minimize the forecasting error on a validation data set. The Levenberg–Marquardt algorithm [48] was used in this study, which is an advanced optimization algorithm and one of the more efficient for training neural networks ('trainlm' function in Matlab). Connection weights and node biases were adjusted iteratively by a process of minimizing the forecast errors ('mse' performance function in Matlab) on a validation data set. The maximum number of training epochs was 1000.

5.2. *Price spike forecasting. GMM*

As mentioned in Section 2, data mining techniques based on a Bayesian classifier have achieved promising results when applied to the spike forecasting problem. A GMM based on a Bayesian classifier was proposed in our paper to approximate the probability density function (pdf) of electricity prices and classify samples (non-spike or spike).

When the pdf that describes the data points in a class is not known, it has to be estimated prior to application of the Bayesian classifier. An arbitrary pdf can be modeled as a linear combination (weighted sum) of several pdfs. Therefore, if a high number of component distributions is used, any distribution can be approximated [49]. The pdf for the samples is then given by

DOI: 10.1002/etep

$$prob(x|\Delta) = \sum_{i=1}^{M} prob(x|\mu_i, \Sigma_i) P_{wi} \tag{13}$$

where x is a V-dimensional continuous-valued data vector (i.e. measurement of features), P_{wi}, $i = 1,\ldots, M$ are the mixture weights and $prob(x|\mu_i, \sum_i)$, $i = 1,\ldots,M$ are the component Gaussian densities. Each component density is a V-variate Gaussian function of the form,

$$prob(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{V/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu_i)'\Sigma_i{}^{-1}(x-\mu_i)\right\} \tag{14}$$

with mean vector μ_i and covariance matrix $\sum_i$. The mixture weights satisfy the constraint $\sum^{M}{}_{i=1}\mathrm{P_w}$ (ω_i). The complete GMM is parameterized by the mean vectors, covariance matrices and mixture weights from all the component densities. These parameters are represented as

$$\Delta = \{P_w(\omega_i), \mu_i, \Sigma_i\} \;\; i = 1, \ldots, M \tag{15}$$

Several techniques are available for estimating the parameters of the GMM. By far the most popular and well-established method is ML estimation. The aim of ML estimation is to find the model parameters which maximize the likelihood of the GMM given the training data. For a sequence of n training vectors $X = \{x_1,\ldots,x_n\}$, the GMM likelihood, assuming independence between the vectors, can be written as,

$$prob(X|\Delta) = \prod_{k=1}^{n} prob(x_k|\Delta) \tag{16}$$

This expression is a nonlinear function of the parameters of Δ, and direct maximization is not possible. However, ML parameter estimates can be obtained iteratively using a special case of the expectation-maximization (EM) algorithm. The basic idea of the EM algorithm is, beginning with an initial model Δ, to estimate a new model $\bar{\Delta}$ such that $prob(X|\bar{\Delta}) \geq prob(X|\Delta)$.The new model then becomes the initial model for the next iteration, and the process is repeated until some convergence threshold is reached. The EM algorithm for GMM was described in [50].

Predicting the occurrence of a spike is a typical binary classification problem. The factors relevant to spikes can be considered as the dimensions of the input vector $X = \{x_1,\ldots,x_n\}$ at each time point$t$ where x_j, $j = 1,2,\ldots,n$ is the value of a relevant factor. The object is to determine the label y for every input vector, where

$$y = \begin{cases} 1, non-spike \\ -1, spike \end{cases} \tag{17}$$

and y denotes whether a spike will occur.

GMM based on a Bayesian classification algorithm is used to mine the database to find out the internal relationships between electricity price spikes and external factors. Basically, for a given input vector $X = \{x_1,\ldots,x_n\}$ and its class label $y \in \{c_1, c_2,\ldots, c_m\}$, the probability classifier calculates the probability that X belongs to class c_j for $j = 1,2,\ldots,m$. X is labeled as class c_j, which has the largest probability. In [33], the probability of spikes was calculated for every input vector utilizing the naïve Bayesian classifier and then compared with a threshold. If the probability was larger than the threshold, a spike was predicted to occur, regardless of whether this probability was larger than the probability of non-spikes. This modification was performed because the price spike prediction problem is a serious imbalanced classification problem (i.e. some classes have many more samples than other classes). In fact, the probability of spikes is less than the probability of non-spikes on most occasions. Many spikes occur when their occurrence probabilities are smaller than 50%. Without setting a threshold smaller than 50%, many spikes will be misclassified. The threshold can be determined by historical data. A Bayesian classifier considering prior information, i.e. prior class probability, tends to be less prone to problems regarding sample class imbalance.

After having determined the probability of price spike occurrence, it is of considerable interest for market participants to be able to further predict the actual value of the price spike. The K-nn approach has been used for this task. From the training data samples, K-neighboring samples closest to the unknown sample are selected. Then, the sum of weighted values of the K-closest samples is computed

DOI: 10.1002/etep

as the unknown sample's value. The distance metric from the unknown sample $Z = \{z_1, \ldots, z_n\}$ to the neighboring sample $X = \{x_1, \ldots, x_n\}$ is determined by the Euclidian distance between two real valued vectors as given in Equation (18)

$$dist(X, Z) = \sqrt{\sum_{i=1}^{n} (x_i - z_i)^2} \tag{18}$$

where n is a vector dimensionality

6. HYBRID ELECTRICITY PRICE FORECASTING MODEL

6.1. Structure flowchart

The proposed methodology of a hybrid system was applied to analyze the given electricity price time series in two layers. The initial price time series was statistically divided into the normal range price set and the spikes set by the method introduced previously. Both data sets produced were analyzed independently, as shown in Figure 7. Depending on the type of data, i.e. raw or decomposed, the SARIMAX or ARMAX model was first applied to forecast the normal range prices, and the GARCH model was used to present the heteroscedastic characteristics of the residuals, resulting in the SARIMAX-GARCH or ARMAX-GARCH model.

A neural network has been utilized to present the nonlinear, non-stationary impact of the non-base electricity demand series on electricity prices and improve predictions based on time series techniques. Therefore, the set of inputs for the neural network includes both forecast and lagged values of electricity prices. If the data was initially decomposed, the output of the ARMAX-GARCH model was not transferred back to the original range values but fed to a neural network. This action gave an opportunity to compare the performance of the two models using input data with different parameters, i.e. original scale and decomposed data.

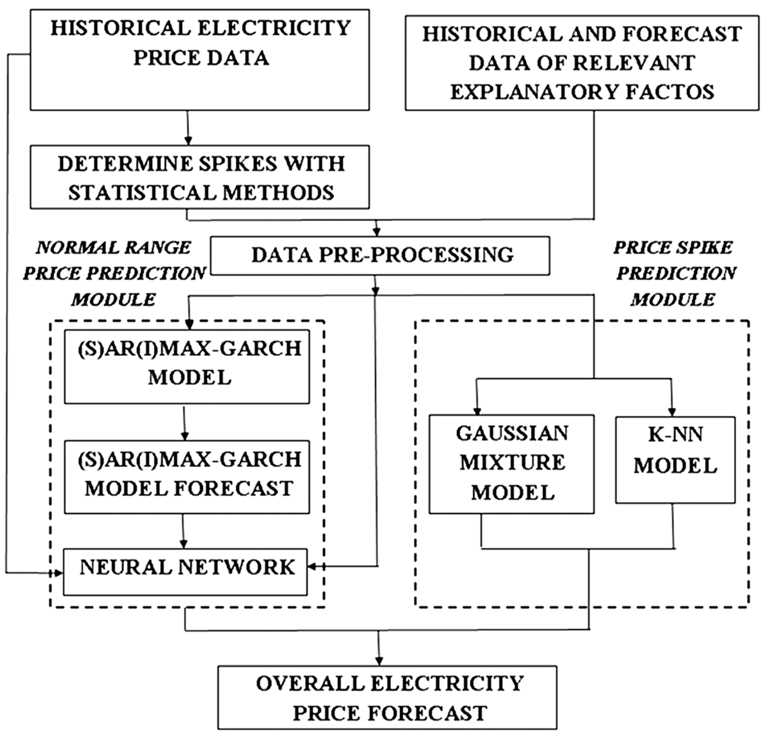

Figure 7. Scheme of the price forecasting model.

DOI: 10.1002/etep

The probability of a price spike occurrence and its value were produced by GMM and K-nn, respectively. The final output was the overall electricity price forecast consisting of a normal range price and a price spike forecast.

6.2. Forecast evaluation methods

Evaluation of the hybrid model was implemented separately for normal range prices and price spike modules of the hybrid model.

6.2.1. Normal range prices. Several evaluation criteria were used to examine the accuracy of the forecast results. Mean square error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) are considered in our paper to evaluate the performance of forecast results.

Evaluation criteria are listed below:

$$MSE = \sum_{t=1}^{T} (P_t - X_t)^2 / T \tag{19}$$

$$MAE = \sum_{t=1}^{T} |P_t - X_t| / T \tag{20}$$

$$MAPE = \left(\sum_{t=1}^{T} [|P_t - X_t| / X_t] / T \right) \cdot 100\% \tag{21}$$

Here, P_t is the predicted value at time t, X_t is the actual value at time t and T is the number of predictions.

The main disadvantage of MAPE criteria is the adverse effect accruing from small actual values. If the actual value is small, Equation (21) will contribute large terms to MAPE even if the difference between the actual and forecast values is small. In order to avoid such a negative effect, an adapted MAPE (AMAPE) was proposed in [26]:

$$AMAPE = \left(\sum_{t=1}^{T} \left[|P_t - X_t| / \left(\sum_{t=1}^{T} X_t / T \right) \right] / T \right) \cdot 100\% \tag{22}$$

6.2.2. Price spikes. Prediction of the occurrence of spikes has been stated as a classification problem [30,31]. It is important to define reliable measures to assess the performance of the classification model. Some classification performance measures have been proposed in [30]. A standard performance measure of a classification is the estimate of the probability of correct classification:

$$Classifier\ accuracy = N_{corr\_class\_patterns} / N \tag{23}$$

where $N_{corr\_class\_patterns}$ is the number of correctly classified patterns, and N is the total number of patterns. This measure is robust for many classification problems but not for the problem under discussion here. Since the data of our problem are extremely unbalanced, the value of the measure will stay high even if all the spikes are misclassified. Thus, other classification performance measures, namely, spike prediction accuracy and spike prediction confidence, are proposed to solve this problem.

Spike prediction accuracy is a ratio of the number of correctly classified spikes $N_{corr\_class\_spikes}$ to the number of actual spikes N_{spikes}:

$$Spike\ prediction\ accuracy = N_{corr\_class\_spikes} / N_{spikes} \tag{24}$$

This measure was introduced because the ability to correctly predict spike occurrence is the subject of greatest concern.

Spike prediction confidence is another measure which aims to account for the uncertainties and risks carried within the forecast. Spike prediction confidence is described as

$$Spike\ prediction\ confidence = N_{corr_{class}} spikes / N_{class_{spikes}} \tag{25}$$

DOI: 10.1002/etep

where $N_{corr\_class\_spikes}$ is the number of correctly classified spikes and N_{class_spikes} is the number of samples classified as spikes. As the classifier may misclassify some non-spikes as spikes, this definition is used to assess the percentile in which the classifier makes this kind of mistake.

7. CASE STUDY

This section presents structures and forecasting performance of the models implemented to predict normal range prices and price spikes.

7.1. Normal range prices prediction module

7.1.1. ARMA-based model building. To get a more accurate normal range price forecast, price spikes should be carefully eliminated from the historical price data. The proposed normal range price forecasting model was implemented using data from the Nord Pool spot market. Finnish electricity spot prices adjusted by the spike elimination method are given in Figure 5.

In the methodology, the data were divided into two sets: the training data set and the testing data set. It was opted to re-estimate the values of the model's parameters at each step of forecasting. Therefore, the training data set was readjustable in a rolling window of three years' length. At each step of forecasting, the model was fitted on the specific training data set. Using the model obtained, prices were predicted on the testing data set of length up to seven days.

The data set of the three years from 1 January 2006 to 31 December 2008 was used as the initial training data set. The data of one year, from 1 January 2009 to 31 December 2009, was utilized as the testing data set.

Two ARMA-based models with external variable (SARIMAX and ARMAX) were evaluated.

The SARIMAX and ARMAX models obtained from the initial training data set are presented in Equations (26) and (27), respectively.

$$(1-\phi_1 B)(1-B)\left(1-B^7\right)p_t = \eta_1 U_{n.b.dem}(1-B)\left(1-B^7\right) + \left(1+\theta_1 B^1\right)\left(1+\Theta_7 B^7\right)a_t + a_t \quad (26)$$

$$\left(1-\phi_1 B-\phi_7 B^7-\phi_8 B^8\right)p_t = \eta_1 U_{n.b.dem} + \left(1+\theta_7 B^7\right)a_t + a_t \quad (27)$$

Note that the SARIMAX (27) model was built using raw data, while the ARMAX (28) model utilized the irregular components of the original series of prices and demand. According to Equation (28), the model obtained for the resulting irregular series (ARMAX) still indicates the existence of some patterns in the irregular price series.

It is straightforward to check that the determined models deal successfully with the stationary and invertibility conditions required. The results of estimating (Equations (27)–(28)) can be seen in the second and third columns of Table II. All the coefficients are statistically significant at the 5% level. The models exhibit predictive power with R^2 ranging from 0.821 to 0.871. The residuals are free of serial correlation based on the chi-square Ljung–Box Q-statistics. However, the chi-square test statistic for autoregressive conditional heteroskedasticity is statistically significant at the 5% level.

7.1.2. (S)AR(I)MAX-GARCH model building. To recognize the presence of the ARCH in the residuals, (S)AR(I)MAX-GARCH models are estimated. The Akaike information criterion and the Bayesian information criterion were compared for an extensive range of different GARCH models. The methodology resulted in the following (S)AR(I)MAX-GARCH(1,1) models:

$$(1-\phi_1 B)(1-B)\left(1-B^7\right)p_t = \eta_1 U_{n\cdot b\cdot dem}(1-B)\left(1-B^7\right) + \left(1+\theta_1 B^1\right)\left(1+\Theta_7 B^7\right)a_t + a_t, \text{where} a_t \sim N\left(0,\sigma^2\right) \quad (28)$$

$$\left(1-\phi_1 B-\phi_7 B^7-\phi_8 B^8\right)p_t = \eta_1 U_{n.b.dem} + \left(1+\theta_7 B^7\right)a_t + a_t, \text{where} a_t \sim N\left(0,\sigma^2\right) \quad (29)$$

DOI: 10.1002/etep

Table II. ARMA-based models obtained from the initial training data set.

	SARIMAX	ARMAX	SARIMAX-GARCH(1,1)	ARMAX-GARCH(1,1)
φ_1	0.326	0.861	0.305	0.849
	(0.077)	(0.029)	(0.070)	(0.027)
φ_7		0.730		0.722
		(0.048)		(0.041)
φ_8		0.592		0.581
		(0.055)		(0.053)
η_1	0.003	0.506	0.003	0.495
	(0.000)	(0.065)	(0.000)	(0.054)
θ_1	−0.681	−0.222	−0.673	−0.235
	(0.062)	(0.045)	(0.057)	(0.041)
θ_7		−0.420		−0.440
		(0.054)		(0.061)
Θ_7	−0.859		−0.841	
	(0.029)		(0.021)	
Variance equation:				
K		26.785	21.670	26.785
		[0.141]	[0.359]	[0.141]
α		30.517	11.185	30.517
		[0.000]	[0.000]	[0.000]
β		0.871	0.821	0.871
Model diagnostics:				
LB_Q	21.670	26.785	26.413	29.001
	[0.359]	[0.141]	[0.203]	[0.105]
ARCH	11.185	30.517	0.559	0.888
	[0.000]	[0.000]	[0.455]	[0.346]
R^2	0.821	0.871	0.835	0.880

Notes: Standard errors are reported in parentheses and probability values in parentheses. R^2 is the coefficient of determination. LB_Q is the Ljung–Box Q-statistic to test for serial correlation in the residuals. ARCH tests for autoregressive conditional heteroscedasticity in the residuals.

$$\sigma_t{}^2 = K + \sum_{i=1}^{1} \alpha_i a_{t-i}{}^2 + \sum_{j=1}^{1} \beta_i \sigma_{t-j}{}^2, \text{ at time } t, t = 1, 2, 3, \ldots, T \tag{30}$$

The fourth and fifth columns of Table II report the results of the estimation (Equations (28)–(30)). All the coefficients are statistically significant at the 5% level. Residuals are free of ARCH.

Two further ARMA-based models without external variables (SARIMA-GARCH and ARMA-GARCH) were used to check whether inclusion of non-base electricity demand in the price forecasting model could result in significant improvement of forecasting performance.

7.1.3. Combination of SARIMAX/ARMAX and GARCH predictions through the neural network. A combination approach is proposed in Section 6.1 to be the normal range prices prediction module of the hybrid model. The module uses of both forecast and lagged values of the electricity prices as inputs to the neural network to increase the accuracy of the price forecast. The combined model was implemented on raw and decomposed data. There were six neurons in the input layer of the ANN of the combined model to predict price value on day T: *SARIMAX-GARCH/ARMAX-GARCH* model price prediction on day T; historical price on day T-1; historical price on day T-7; historical price on day T-14; non-base electricity demand on day T; non-base electricity demand on day T-1. The selection of input features was based on correlation analysis. Here, the historical price data utilized as the input for the neural network indicates trend and weekly periodicity of a price time series.

At the training stage, experiments were carried out with a number of different neurons in the hidden layer to minimize the forecasting error on a validation data set. The model with three hidden neurons resulted in the smallest forecasting error.

DOI: 10.1002/etep

7.1.4. Neural network model. In addition to the ANN utilizing predictions from other models, two new ANNs built on raw and decomposed data were applied to forecast electricity prices. The input features were chosen in order to be able to compare the forecasting performance of the ANNs to the previously mentioned models (ARMA-based models, the combined model). Five inputs were used for price forecasts on day T, namely: historical price on day T-1; historical price on day T-7; historical price on day T-14; non-base electricity demand on day T; non-base electricity demand on day T-1.

Both ANNs contained four hidden nodes and one output node that is a price forecast.

7.1.5. Simplistic benchmark method. The random walk was implemented as the naïve method since it is the most widely used and simplest naïve benchmark method used in forecasting studies. For the daily NordPool data, the forecast function is given in Equation (31)

$$p_t = p_{t-1}, \quad \text{at time } t, t = 1, 2, 3, \ldots, T \tag{31}$$

where p_t is the daily electricity spot price at time t.

7.1.6. Evaluation of the performance of the models. This part of the study compares the proposed combined model utilizing raw and decomposed data with the seven previously discussed simple models implemented to predict normal range prices. Table III summarizes statistical measures that characterize the prediction accuracy of the different techniques studied. This information allows ranking of the forecasting techniques. The first and second columns of the table identify the model and horizon of prediction. The third, fourth and fifth columns show the values of the MSE, MAE and AMAPE, respectively.

In addition to Table III, the forecasting accuracy of the nine models for forecasting horizons up to one week ahead is also shown in Figure 8. The error measure used for forecasting accuracy estimation in Figure 8 is AMAPE, which is easily interpreted as a proportionality between the forecast error and the actual electricity price and has performance similar to MSE and MAE.

From Figure 8, it can be seen that the naïve benchmark (random walk model) was substantially outperformed by all the other methods at all lead times. The SARIMA-GARCH model implemented on raw data has the poorest performance of the other models at all lead times, except for a lead time of one day when it shows very similar performance to the ANNs. The figure, furthermore, shows that models using decomposed data as an input performed better than analogical models trained on raw data. For ARMA-based methods, the decomposing preprocessing method is more effective than direct entering of trend and seasonal terms into a model. ANNs built with detrended and deseasonalized data can produce significantly more accurate forecasts than with raw data. This result suggests that neural networks built on raw data are unable to adequately learn seasonality and trend, a finding which supports previous studies on neural networks [20,47]. It is unsurprising that the performance of SARIMA-GARCH and ARMA-GARCH models was much improved after inclusion of exogenous factor information, resulting in SARIMAX and ARMAX, respectively.

The performance of the ANNs relative to the ARMA-based models was worse, which differs from the work in [20]. A possible explanation could be that specific characteristics of the initial price time series have become more linear after transformation of hourly data into daily data and spikes elimination. Moreover, forecasted values of prices on day T-1 were used as an input to the ANNs for the multistep predictions. In a nonlinear model, errors might be spread significantly. It must be noted that if a particular neural network fails to produce good results, this does not indicate that neural networks in general are poor predictors because a different specification of neural network might has performed better [37]. Problems may arise from the difficult tune-up of neural network algorithms, which need validation of the model by number of hidden layers, and the number of neurons in the input and hidden layers [51]. Of the remaining models, the combined model utilizing decomposed data performed considerably better than all the other models. Therefore, the combined model with decomposed data is used as the normal range price prediction module of the hybrid model.

Figure 9 focuses more closely on AMAPE results for lead times of one and seven days for the two best models, i.e. ARMAX-GARCH and the combined model with decomposed data, against the day of the week.

DOI: 10.1002/etep

Table III. Performance measurements of the models.

	Horizon, day	MSE	MAE	AMAPE, %
Naïve benchmark	1	15.27	2.70	7.43
	2	18.67	3.12	8.60
	3	20.84	3.34	9.19
	4	26.14	3.68	10.15
	5	26.30	3.69	10.18
	6	26.20	3.65	10.16
	7	26.44	3.73	10.29
SARIMA-GARCH	1	8.14	2.09	5.76
	2	10.55	2.53	6.96
	3	11.69	2.67	7.36
	4	15.13	2.92	8.05
	5	17.90	3.18	8.76
	6	19.43	3.32	9.15
	7	22.44	3.51	9.66
SARIMAX-GARCH	1	7.35	2.00	5.51
	2	8.57	2.23	6.15
	3	8.96	2.28	6.29
	4	10.97	2.42	6.67
	5	12.51	2.65	7.30
	6	13.95	2.82	7.77
	7	16.56	2.95	8.14
ARMA-GARCH[a]	1	6.30	1.94	5.34
	2	7.41	2.10	5.80
	3	8.09	2.17	5.97
	4	8.53	2.27	6.25
	5	8.83	2.29	6.38
	6	9.67	2.39	6.63
	7	9.76	2.40	6.75
ARMAX-GARCH[a]	1	6.02	1.90	5.23
	2	6.87	2.00	5.50
	3	7.00	2.02	5.57
	4	7.18	2.05	5.87
	5	8.36	2.23	6.14
	6	8.75	2.27	6.37
	7	8.79	2.31	6.47
ANN with raw data	1	10.14	2.14	5.89
	2	10.55	2.34	6.46
	3	10.80	2.45	6.76
	4	11.62	2.55	7.03
	5	14.33	2.78	7.68
	6	14.98	2.88	7.95
	7	15.73	3.00	8.27
ANN with decomposed data	1	7.09	2.08	5.73
	2	7.91	2.20	6.05
	3	8.83	2.27	6.25
	4	9.19	2.38	6.55
	5	9.92	2.48	6.82
	6	11.13	2.56	7.08
	7	11.16	2.64	7.27
Combined with raw data	1	7.41	1.96	5.40
	2	8.34	2.17	5.97
	3	8.56	2.19	6.05
	4	11.20	2.35	6.49
	5	11.48	2.51	6.92
	6	13.23	2.67	7.35
	7	13.70	2.78	7.66
Combined with decomposed data	1	6.01	1.89	5.20
	2	6.58	1.95	5.36
	3	6.63	1.97	5.39

(*Continues*)

DOI: 10.1002/etep

Table III. (Continued)

	Horizon, day	MSE	MAE	AMAPE, %
	4	7.18	2.05	5.63
	5	7.70	2.14	5.82
	6	8.16	2.16	5.95
	7	8.24	2.22	6.04

Notes: [a]Models tested by decomposed data

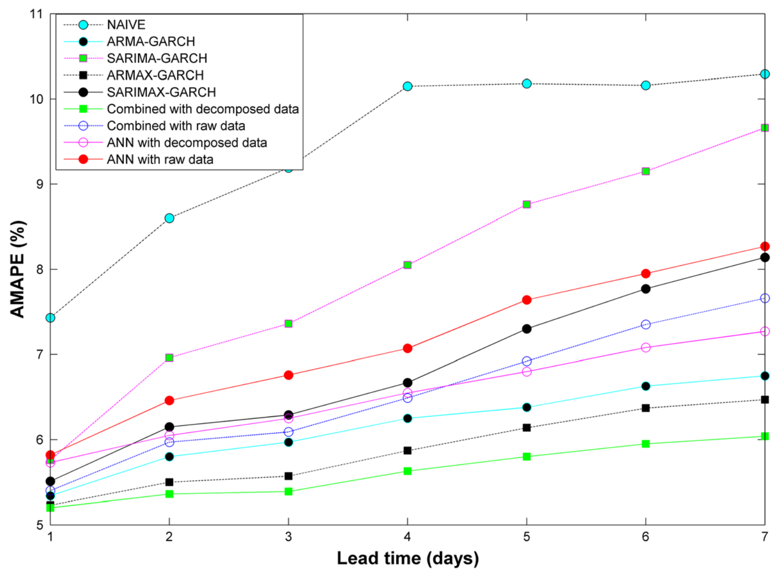

Figure 8. AMAPE results plotted against lead time.

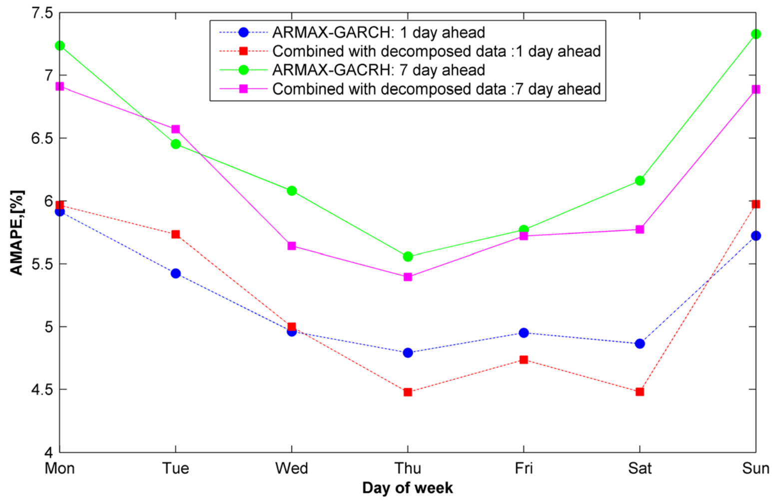

Figure 9. AMAPE results for lead time of one and seven days plotted against day of week.

Both techniques have relatively poor AMAPE results for Sundays and Mondays. This problem could arise from the model algorithm used in the paper. The price depends on the time of day and whether it is a weekday or weekend. Weekdays and weekends have substantially different characteristics in terms of electricity price distribution. Some authors divide the whole data set into

DOI: 10.1002/etep

weekday and weekend data sets to build different models for each data set [25,52], whereas others just report results for models fitted and evaluated on weekday observations only [53]. This paper did not make a division into weekdays and weekends. Therefore, it would seem that all the models have been somewhat challenged by being evaluated and tested on data with slightly different intraweek seasonal characteristics. In this content, the robustness of the models becomes important.

7.2. Price spikes prediction module

Based on statistical analysis, five attributes were chosen to be incorporated into the price spikes classification algorithm: Non-base electricity demand; SDI; temperature; Elspot capacity-flow difference; and day index.

The proposed price spike forecast model was used to predict the probability of price spikes in the Finnish electricity spot market of Nord Pool. Data for 1 January 2006 to 31 December 2008 were used to establish the initial training data sets. The initial training data set was adjusted in the same fashion as described in Section 7.1.1. Data for 1 January 2009 – 31 December 2009 were used as the testing data set. The spike threshold (see Section 5.2) resulting in the best overall performance of the model on a training data set was determined as 38%. Values of the three closest samples were considered to determine the unknown value of a price spike in the K-nn approach.

Actual electricity price spikes and the performance of the proposed approach for a lead time of one day and seven days are shown in Figures 10, 11, respectively.

Classifier accuracy, spike prediction accuracy and spike prediction confidence values decline as the forecasting horizon increases. Spike prediction accuracy and spike prediction confidence values vary between 60–80% and 43–50%, respectively, depending on the forecasting horizon (see Table IV). Of the five spikes in the testing data set, four of them were predicted by the model for a lead time of one day, and three of five spikes were predicted for a lead time of seven days.

From Table V, it can be seen that most of the error rates of the price spike forecast value are less than 35% for lead times of one and seven days, with only one case close to 70%. The forecast error for this particular day can easily be understood given that the actual price is much higher than the historical

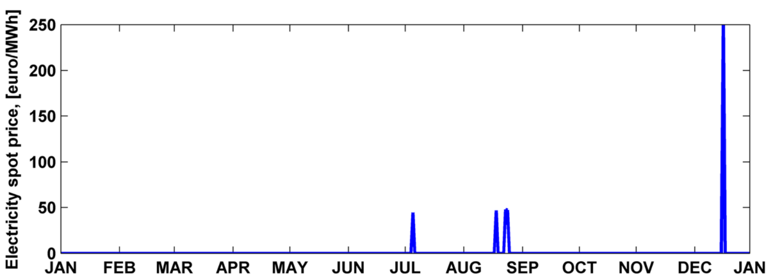

Figure 10. Actual electricity price spikes for 1 January 2009–31 December 2009 of the Finnish electricity spot market.

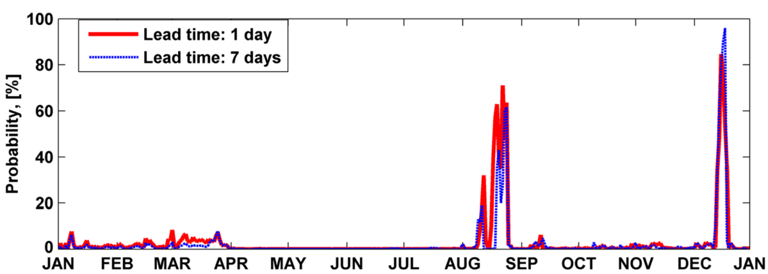

Figure 11. Forecasted probability of spikes for a lead time of one day and seven days for 1 January 2009–31 December 2009 of the Finnish electricity spot market.

DOI: 10.1002/etep

Table IV. Accuracy of the probability model on the testing data for lead times of one and seven days.

Horizon	Classifier accuracy	Spike prediction accuracy	Spike prediction confidence
1	360/365 = 98.63%	4/5 = 80.00%	4/(4 + 3) = 57.14%
7	359/365 = 98.36%	3/5 = 60.00%	3/(3 + 3) = 50.00%

Table V. Comparison between the actual and price spikes values forecasted by K-nn method on the testing data for lead times of one and seven days.

Horizon	Spike number	Forecasted price, euro/MWh	Original price, euro/MWh	Forecast error,%
1	1	59.40	79.12	24.92
	2	51.04	65.30	21.47
	3	65.20	56.35	15.71
	4	94.07	112.00	16.01
	5	88.57	251.04	64.71
				Mean: 28.56
7	1	53.46	79.12	32.43
	2	48.57	65.30	25.62
	3	61.34	56.35	8.85
	4	89.07	112.00	20.47
	5	73.14	251.04	70.86
				Mean: 31.65

average price spike. In fact, other forecasted price spike values are close enough to the real values to provided useful information for practitioners.

The output values of the spike forecasting model partially indicate the robustness of the model because of the small number of spiky samples in the testing data set. Model robustness and the stability of the model to changes in the training data may be tested by cross-validation. Here, the aim is to estimate how accurately the predictive model would perform in practice. In cross-validation, the original sample is randomly portioned into training and testing (or validation) subsamples of a specific size. This parting process is independently repeated H times to yield H portioned data sets, which are treated as independent sets. The cross-validation estimates matrix of the model performance measures for the testing data set, denoted $\widehat{\Omega}^{*(.)}$, is merely the mean matrix of the H estimates on the individual testing data sets executed by the data parting

$$\widehat{\Omega}^{*(.)} = \frac{1}{H}\sum_{h=1}^{H} \widehat{\Omega}^{*(h)} \tag{32}$$

where $\widehat{\Omega}^{*(h)}$ is a matrix of the model performance measures for the testing set on the portioned sample h.

At the cross-validation stage, the robustness of the model was checked on testing data sets of specific size. The whole data set from 1 January 2006 to 31 December 2009 (see Figure 4) was randomly divided into training data sets and testing data sets B times. The model was fitted using the training data set, and, using the model obtained, the values were predicted seven days ahead on the testing set of specific size. Note that in this study, we opted not to re-specify or re-estimate the model values at each step of forecasting. This action gave an opportunity to compare the performance of the model using the specific size of training and testing data sets. Table VI and Table VII show results of the GMM and K-nn models performance measures for testing data sets of specific size after $H = 500$ simulations of the cross-validation:

The seemingly not high rate in the values of the spike prediction accuracy and confidence for the probability model and relatively high error value for K-nn model are mainly because of the very limited number of price spike events in the historical data (1.5% of the whole sample). The values were obtained with insufficient data containing spikes. Many stochastic events causing spikes could not be considered in the model. Table V suggests that further training with historical data would improve the

DOI: 10.1002/etep

Table VI. Accuracy of the probability model on randomly selected testing data sets of specific size after H = 500 simulations.

	Size of a testing data set			
	365 days	180 days	30 days	7 days
Classifier accuracy	98.53%	97.27%	96.61%	91.64%
Spike prediction accuracy	41.01%	43.45%	48.76%	57.82%
Spike prediction confidence	29.24%	31.34%	36.93%	45.46%

Table VII. Error value of the K-nn model on randomly selected testing data sets of specific size after H = 500 simulations.

	Size of a testing data set			
	365 days	180 days	30 days	7 days
Error value	33.75%	28.17%	22.35%	18.23%

accuracy of the models. Spike prediction accuracy using the proposed probability model is above 50% for the testing data set of seven days, which means that more than 50% of spikes can be predicted. In the light of the fact that price spikes are highly stochastic, the achieved forecast accuracy level is sufficiently good to provide market participants with an ability to anticipate price spikes.

7.3. Hybridization of normal range price and spikes predictions

Integration of the spike probability and value forecasting results with the normal range price forecasting result obtained from the combined model with decomposed data (the best performing normal range price model) gives the complete electricity price forecast. Forecasted normal range prices, probability of price spikes and complete electricity price forecast for a lead time of one day are shown in Figures 12–14, respectively. Obviously, without the spike occurrence and spike value predictors, the performance of the normal range price forecasting model deteriorates when spikes occur.

Hybridization of the normal price range prediction and the price spike probability prediction provides valuable information about the electricity market and gives market participants the ability to manage their risks.

8. CONCLUSIONS

Since there is little published research in the area of price forecasting in the Nordic electricity market and especially in the Finnish electricity market, accurate and robust price forecast methods are of considerable interest. This paper presented a hybrid methodology for the prediction of normal range electricity market prices which also has the ability to predict the occurrence of electricity market price spikes. The performance of eight methods for forecasting normal range electricity prices was compared. A highlight of this study was the success of the combined model utilizing decomposed data. The proposed combined time series and ANN models, composed of linear and nonlinear relationships of prices and exogenous variables, improved the performance of normal range price forecast results.

In many other studies, price spikes were truncated before application of the forecasting model. When utilized in addition to the normal range price forecast, the proposed price spike forecast method can provide practically useful and reasonably accurate forecasts, enhancing the applicability of price forecasts in the actions of electricity market participants. In the light of the fact that price spikes are highly stochastic, the achieved spike forecast accuracy level is acceptable. Further training with historical data would improve the accuracy.

DOI: 10.1002/etep

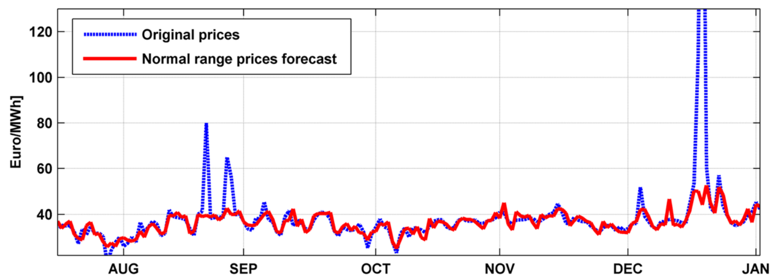

Figure 12. Forecasted normal range prices for a lead time of one day for July 2009–December 2009 of the Finnish electricity spot market.

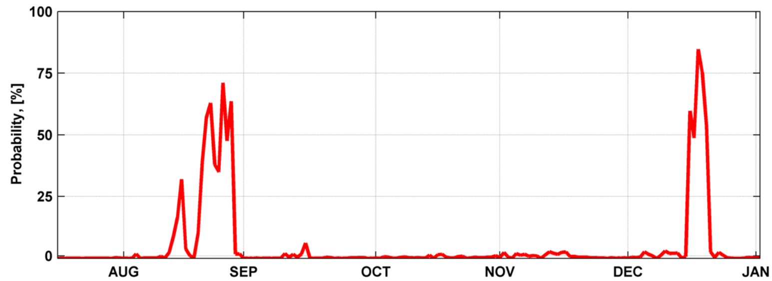

Figure 13. Forecasted probability of spikes for a lead time of one day for July 2009– December 2009 of the Finnish electricity spot market.

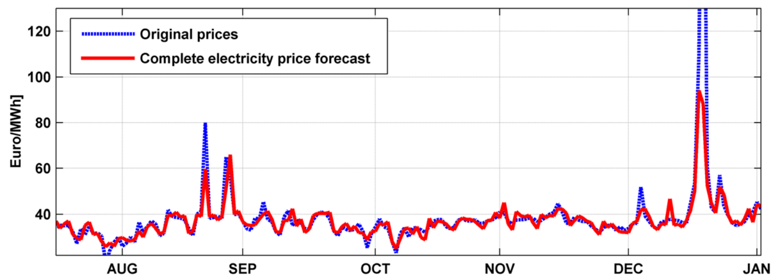

Figure 14. Complete electricity price forecast (normal range prices + GMM + K-nn forecasts) for a lead time of one day for July 2009– December 2009 of the Finnish electricity spot market.

9. LIST OF ABBREVIATIONS AND SYMBOLS

9.1. Abbreviations

ACF	autocorrelation function
AMAPE	adapted mean absolute percentage error
ANN	artificial neural network
ARCH	auto regressive conditional heteroskedasticity
ARMA	auto regressive moving average
ARMAX	auto regressive moving average with external variable
Élspot	day-ahead market for power
EM	expectation maximization algorithm
Fingrid	Finnish system operator
GARCH	generalized auto regressive conditional heteroskedasticity
GMM	Gaussian mixture model
K-nn	K-nearest neighboring

DOI: 10.1002/etep

ML	maximum likelihood
MAE	mean absolute error
MAPE	mean absolute percentage error
MSE	mean square error
PACF	partial autocorrelation function
pdf	probability density function
SARIMA	seasonal auto regressive integrated moving average
SARIMAX	seasonal auto regressive integrated moving average with external variable
SDI	supply–demand index

9.2. Symbols

B	lag operator
Σ	covariance matrix
LB_Q	the Ljung–Box Q-statistic
μ	mean vector
N	total number of patterns
N_{class_spikes}	number of actual spikes
$N_{corr\_class\_patterns}$	number of correctly classified patterns
$N_{corr\_class\_spikes}$	number of correctly classified spikes
N_{spikes}	number of spikes
p	regular order of the autoregressive polynomial
P	seasonal order of the autoregressive polynomial
P_w	mixture weight
R^2	coefficient of determination
σ	standard deviation
q	regular order of the moving average polynomial
Q	seasonal order of the moving average polynomial
$U_{n.b.dem}$	non-base electricity demand
w	window length

REFERENCES

1. Bastian J, Zhu J, Banunarayanan V, Mukerji R. Forecasting energy prices in a competitive market. *IEEE Computer Application, Power* 1999; **12**:40–45.
2. Fu Y, Li Z. Different models and properties on LMP calculations, In Proc. IEEE Power Eng. Soc. General Meeting, Montreal 2006; QC, Canada.
3. Guan X, Pepyne D, Ho YC. Gaming and price spikes in electric power markets. *IEEE Transactions on Power System* 2001; **16**:402–408.
4. Chandarasupsang T, Galloway S, Burt G, McDonald J, Siewierski T. Bidding behaviour and electricity market simulation. *European Transactions on Electrical Power* 2007; **17**:333–346.
5. Sadeh J, Mashhadi HR, Latifi MA. A risk-based approach for bidding strategy in an electricity pay-as- bid auction. *European Transactions on Electrical Power* 2009; **19**(1):39–55.
6. Barlow M. A diffusion model for electricity prices. *Mathematical Finance* 2002; **12**(4):287–98.
7. Barz G, Johnson B. Modeling the prices of commodities that are costly to store: the case of electricity. Proceedings of the Chicago Risk Management Conference 1998; Chicago, IL.
8. Moghram I, Rahman S. Analysis and evaluation of five short-term load forecasting techniques. *IEEE Transactions on Power System* 1989; **4**(4):1484–1491.
9. Hagan MT, Behr SM. The time series approach to short term load forecasting. *IEEE Transactions on Power System* 1987; **PWRS-2**(3):785–791.
10. Erdogdu E. Electricity demand analysis using cointegration and ARIMA modelling: a case study of Turkey. *Energy Policy* 2007; **35**(2):1129–46.
11. Nogales FJ, Contreras J, Conejo AJ, Espínola R. Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power System* 2002; **17**:342–348.
12. Contreras J, Espínola R, Nogales FJ, Conejo AJ. ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power System* 2003; **18**(3):1014–1020.
13. Robinson TA. Electricity pool prices: a case study in nonlinear time-series modelling. *Applied Economics* 2000; **32**(5):527–532.

DOI: 10.1002/etep

14. Rambharat BR, Brockwell AE, Seppi DJ. A threshold autoregressive model for wholesale electricity prices. *Journal of the Royal Statistical Society Series C* 2005; **54**(2):287–300.
15. Engle RF. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 1987; **50**:987–1007.
16. Bollerslev T. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 1986; **31**:307–327.
17. Garcia RC, Contreras J, van Akkeren M, Garcia JBC. A GARCH forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power System* 2005; **20**(2):867–874.
18. Jabłońska M, Nampala H, Kauranne T. The multiple-mean-reversion jump-diffusion model for Nordic electricity spot prices. *The journal of energy market* 2011; **4**(2):3–25.
19. Karandikar RG, Deshpande NR, Khaparde SA, Kulkarni SV. Modelling Volatility Clustering in Electricity Price Return Series for Forecasting Value at Risk. *European Transactions on Electrical Power* 2009; **19**:15–38.
20. Zhang GP, Qi M. Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research* 2005; **160**:501–514.
21. Catalão JPS, Mariano SJPS, Mendes VMF, Ferreira LAFM. An artificial neural network approach for short-term electricity prices forecasting. In Proc. 14th Int. Conf. on Intelligent System Applications to Power Systems 2007; 411–416.
22. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 2003; **50**:159–175.
23. Swanson NR, White HA. Model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *The Review of Economics and Statistics*, MIT Press, Cambridge, MA, 1997; **79**(4):540–550, November.
24. Tseng FM, Yu HC, Tzeng GH. Combining neural network model with seasonal time series ARIMA model. *Technological Forecasting and Social Change* 2002; **69**:71–87.
25. Wu L, Shahidehpour M. A hybrid model for day-ahead price forecasting. *EEE Transactions on Power System* 2010; **25**(3):1519–1530.
26. Amjady N, Hemmati M. Day-ahead price forecasting of electricity markets by a hybrid intelligent system. *European Transactions on Electric Power* 2009; **19**(1):89–102; doi:10.1002/etep.242.
27. Becker R, Hurn S, Pavlov V. Modelling Spikes in Electricity Prices. *The Economic Record* 2007; **82**(263):371–382.
28. Yamin HY, Shahidehpour SM, Li Z. Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets. *Electrical Power Energy System* 2004; **26**:571–81.
29. Weron R. Modeling and forecasting electricity loads and prices: A statistical approach. *Volume 396 of Wiley finance series*, Wiley, Chichester, West Sussex, England, 2006; 125–127.
30. Lu X, Dong ZY, Li X. Electricity Market Price Spike Forecast with Data Mining Techniques, *Electric Power Systems Research* 2005; **73**(1):19–29, ELSEVIER, Oxford, UK.
31. Zhao JH, Dong ZY, Li X, Wong KP. A framework for electricity price spike analysis with advanced data mining methods. *IEEE Transactions on Power System* 2007a; **22**(1):376–385.
32. Zhao JH, Dong ZY, Li X. Electricity market price spike forecasting and decision making, *IET Generation, Transmission and Distribution* 2007b; **1**(4):647–654.
33. NordPool Spot. Market data, http://www.nordpoolspot.com 2012a
34. Vucetic S, Tomsovic K, Obradovic Z. Discovering price – load relationships in California's electricity market. *IEEE Transactions on Power System* 2001; **16**(2):280–6.
35. Nogales FJ, Conejo AJ. Electricity price forecasting through transfer function models. *Journal of the Operational Research Society* 2006; **57**(4):350–356.
36. Calmarza M, de la Fuente JI. New forecasting method for the residual demand curves using lime series (ARIMA) models. In Proc. Int. Conf. Probabilistic Methods Applied to Power Systems 2002; Naples, Italy.
37. Taylor JW, de Menezes LM, McSharry PE. A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting* 2006; **22**:1–16
38. Willis H. *Spatial Electric Load Forecasting*. Second edition, Revised and Expanded. Marcel Dekker, Inc, New York,. 2002; 124–128
39. Statistics Finland. http://www.stat.fi 2012
40. Weather Underground. http://www.wunderground.com 2012
41. NordPool Spot. TSO congestion rent, http://www.nordpoolspot.com 2012b
42. Fingrid. Electricity market: cross-border transmission capacities, http://www.fingrid.fi 2012b
43. Bierbrauer ST, Weron R. *Lecture Notes in Computer Science* 2004; Springer Berlin, Heidelberg 859–867
44. Cartea A, Figueroa M. Pricing in electricity markets: A mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance* 2005; **12**(4):313–335.
45. Jablonska M. Analysis of outliers in electricity spot prices with example of New England and New Zealand markets 2008; Master's thesis, Lappeenranta University of Technology.
46. Box GEP, Jenkins GM. *Time Series Analysis: Forecasting and Control* 1976; Holden-Day, San Francisco
47. Nelson M, Hill T, Remus W, O'Connor M. Time series forecasting using neural networks: should the data be deseasonalized first? *Journal of Forecasting* 1999; **18**(5):359–367.
48. Catalão JPS, Pousinho HMI, Mendes VMF. An artificial neural network approach for short-term wind power forecasting in Portugal, 15th International Conference on Intelligent System Applications to Power Systems, ISAP '09, 2009.1–5.

49. Theodoridis S, Koutroumbas K. *An Introduction to Pattern Recognition: A MATLAB Approach*, Elsevier Inc, Burlington, USA. 2010; 79–84.
50. Reynolds D. Gaussian Mixture Models 2005; MIT Lincoln Laboratory.
51. Conejo AJ, Contreras J, Espínola R, Plazas MA. Forecasting electricity prices for a day-ahead pool-based electricity energy market. *International Journal of Forecasting* 2005; **21**(3):435–462.
52. Wang A, Ramsay B. A neural network based estimator for electricity spot pricing with particular reference to weekend and public holidays. *Neurocomputing* 1998; **23**:47–57.
53. Lora AT, Santos Jesus R, Santos Jose R, Exposito AG, Ramos JLM. A comparison of two techniques for next-day electricity price forecasting, *IDEAL, Lecture notes in computer science LNCS* 2002; Berlin, Heidelberg: SpringerVerlag. 384–90.

DOI: 10.1002/etep