Peer-reviewed version available at Energies 2018, 11, 1255; doi:10.3390/en11051255

Article

Electricity Price Forecasting Using Recurrent Neural Networks

Umut Ugurlu [1,†], Ilkay Oksuz [2,†*] and Oktay Tas [1]

1 Istanbul Technical University, Management Engineering Department, Besiktas, Istanbul, Turkey
2 King's College London, Biomedical Engineering Department, London, UK
* Correspondence: ilkay.oksuz@kcl.ac.uk
† These authors contributed equally to this work.

Abstract: Accurate electricity price forecasting has become a substantial requirement since the liberalization of the electricity markets. Due to the challenging nature of the electricity prices, which includes high volatility, sharp price spikes and seasonality, various types of electricity price forecasting models still compete and can not outperform each other consistently. Neural Networks have been successfully used in machine learning problems and Recurrent Neural Networks (RNNs) have been proposed to address time-dependent learning problems. In particular, Long Short Term Memory and Gated Recurrent Units (GRU) are tailor-made for time series price estimation. In this paper, we propose to use Gated Recurrent Units as a new technique for electricity price forecasting. We have trained a variety of algorithms with rolling 3-year window and compared the results with the RNNs. In our experiments, 3-layered GRUs outperformed all other neural network structures and state of the art statistical techniques in a statistically significant manner in the Turkish day-ahead market.

Keywords: electricity price forecasting; deep learning; gated recurrent units; long short term memory; artificial intelligence, turkish day-ahead market

---

1. Introduction

Since the liberalization of the electricity markets, electricity price forecasting has become an essential task for all the players of the electricity markets due to several reasons. Energy supply companies, especially dam-type hydroelectric, natural gas, and fuel oil power plants could optimize their procurement strategies according to the electricity price forecasts. As the share of the regulated electricity markets, such as day-ahead and balancing markets, increase day by day; bilateral contracts also take the regulated-market prices as a benchmark [1]. Moreover, prices of the energy derivatives are also based on electricity price forecasts [2]. From the demand side, some of the companies can schedule their operations according to the low-price zones and operate in these hours or months. Zareipour et al. [3] stress the importance of the short-term electricity forecasting accuracy. A 1% improvement in the mean absolute percentage error (MAPE) would result in about 0.1% - 0.35% cost reductions from short term electricity price forecasting, which results to circa $1.5 million per year for a medium-size utility with a 5-GW peak load [4], [5].

Electricity prices differ from all the other assets and even commodities due to its' unique features such as requirement of having constant balance between the supply and demand sides, demand inelasticity, oligopolistic generation side, and non-storability [6]. These features cause to some important characteristics of the electricity prices: High volatility, sharp price spikes, mean reverting process, and seasonality in different frequencies [7]. Due to all these idiosyncratic features and characteristics, forecasting the electricity prices accurately becomes a very challenging task. Machine learning models are able to solve very complicated classification and regression problems with great success. Recently, deep learning models have become the state of the art in speech recognition [8], handwriting recognition [9] and image classification [10].

This paper presents a Gated Recurrent Unit (GRU) based method for electricity price estimation with the goal of using the valuable time series information fully in a neural network architecture.

Peer-reviewed version available at Energies 2018, 11, 1255; doi:10.3390/en11051255

Neural network based methods showed great promise in computer vision, speech recognition and natural language processing [8]. In particular, Recurrent Neural Networks are capable of faithfully preserving the key time-dependent patterns for natural language processing type problems. This motivated us to propose a thorough analysis of multiple features for the electricity prices estimation using Recurrent Neural Networks (RNNs). In particular, the main contributions of this paper are:

- Use of a GRU Recurrent Neural Network setup for estimating electricity prices.
- A wide analysis of multiple feature settings for neural networks, Convolutional Neural Networks (CNN), Long Short Term Networks (LSTM) and state of the art statistical methods.
- Extensive electricity price estimation performance analysis with both daily and monthly comparisons.
- Detailed analysis between the state of the art statistical models and the neural network based methods.

1.1. Literature

Electricity price forecasting literature started to develop in the beginning of 2000s by [11], [12], [13], [14], [15], [16], and [17]. Following the review of Weron [18], we partition the main methods of electricity price forecasting to five: Multi-agent, fundamental, reduced-form, statistical, and computational intelligence models.

Multi-agent models simulate the operation of the system and build the price process by matching the demand and the supply. Shafie-Khah et al. [19] and Ziel and Steinert [20]'s papers are very good and recent examples of these type of papers. Shafie-Khah et al. [19] model wind power producers, plug-in electricity vehicle owners and customers, who participated into demand response programs, as independent agents in a small Spanish market. Furthermore, Ziel and Steinert [20] propose a model for the German EPEX market, which takes all the supply and demand information of the system and discusses the effects of the changes in supply and demand.

Fundamental or structural methods discuss the effects of the physical and economic factors on the electricity prices. In this part of the literature, variables are modeled and predicted independently, often via other methods such as reduced-form, statistical or machine learning methods. For example, Coulon and Howison [21] develop a model for electricity spot prices by using the stochastic processes of the independent variables. Their method also takes the bid stack function of the price drivers and the electricity prices into account. In another study, Carmona and Coulon [22] focuses on the role of the energy prices and the effect of the fundamental factors on the electricity prices in a survey about the structural methods. Carmona et al. [22] also discuss the superiority of the fundamental models to the reduced-form models. Both Carmona and Coulon [2] and Füss et al. [23] construct fundamental models to achieve the final aim of electricity derivatives pricing.

Reduced-form models mainly consist of two methods: Markov regime-switching and jump diffusion. These models are relatively better than structural and statistical models in terms of handling the spikes. Geman and Roncoroni [24] use mean-reverting jump diffusion (MRJD) model. Their approach captures both trajectory and statistical components of the electricity prices. Cartea and Figueroa [25] and Janczura et al. [26] use more hybrid methods. First of all, they filter out the jumps by using a jump diffusion model; then they propose more statistical methods to model the remaining, stationary part of the series. Hayfavi and Talasli [7] apply a hybrid-jump diffusion model to the Turkish market and compares the results with [25] and [27]. Janczura and Weron [27] compares some of the examples in the literature with their own 3-regime-switching Markov model, which captures both positive and negative spikes, in addition to exhibiting the inverse leverage effect of the electricity spot prices. Furthermore, Eichler and Türk [28] propose a semi-parametric Markov regime-switching model. In their method, model parameters are employed by robust statistical techniques. Moreover, it is easier to estimate, needs less computational time and distributional assumptions. Keles et al. [29] and Bordignon et al. [30] are some examples, which use jump diffusion and Markov regime-switching, respectively, in hybrid works.

Peer-reviewed version available at Energies 2018, 11, 1255; doi:10.3390/en11051255

Statistical and computational intelligence, are the most common models in the electricity price forecasting literature. Statistical models are in great variety from basic naive method [14] to very developed methods [31]. As Ziel and Weron [31] discuss, there are univariate and multivariate frameworks in the electricity price forecasting. In the day-ahead electricity price forecasting, players bid the prices and the quantities for the 24 hours of the next day. In this sense, first way is to predict all the prices in a univariate framework from a single price series as a 24-step-ahead forecast. On the other hand, forecasting the prices from 24 different time series as 1-step-ahead forecasts is another option, which is called multivariate framework. Since Weron and Misiorek [32] apply the univariate framework to the Nordic data, Kristiansen [33] utilizes the multivariate framework on the same dataset in a follow-up study and argues that using univariate framework increases the prediction accuracy. However, it contradicts with the findings of [16], which mentions that using the multivariate framework presents better forecasting results than univariate method. In the same Nordpool market, Raviv et al. [34] has a different point of view. It compares the 1-step-ahead daily average price forecasts in a univariate framework with the aggregated 24-step-ahead forecasts of the hourly prices. From the empirical evidence, Raviv et al. [34] state that multivariate framework has lower out-of-sample errors than the univariate one. Nogales et al. [14], Contreras et al. [13], and Conejo et al. [35] are some substantial examples of the auto-regressive models. Nogales et al. [14] propose the naive method and as mentioned in [14], [13] and [35], not-well calibrated forecasting methods can not outperform the naive method. Although Conejo et al. [35] find out that ARIMA model is worse than the model with exogenous variables in the American PJM market, Contreras et al. [13] state that adding an exogenous variable does not necessarily increase the prediction accuracy.

Many types of computational intelligence models are applied in the electricity price forecasting literature. Some of the early stage papers are [36], [37] and [38]. Mandal et al. [36] forecast the electricity loads and prices in the Australian market by applying Artifical Neural Network (ANN) model for 1-6 hours ahead. MAPE increases from 9.75% to 20.03%, while 1-step ahead forecast increases to 6-step ahead forecast. In another study, Catalao et al. [37] utilize a three-layered feed-forward neural network, which is trained by Levenberg-Marquardt method, and forecast 168-step-ahead in the Spanish and Californian markets. Although they give the results for all the seasons of the Spanish market; in the Californian market, results are available only for the Spring term. Therefore, it is difficult to compare the results of both markets. The main difference in [38] is that it forecasts the daily average prices and requires only 1-step-ahead forecast. In the Nordpool market, a standard error back-propagation method is used, which is improved by self-adaptive learning rate and momentum coefficient algorithms. Results indicate that ANN model outperforms the standard ARIMA method. Recent studies of Keles et al. [1] and Panapakidis and Dagoumas [39] apply mainly ANN methods. Keles et al. [1] propose ANN models with different variables by utilizing the clustering methods. Their ANN based method outperforms the benchmark naive-type models and the Seasonal Auto-regressive Integrated Moving Average (SARIMA) model. An important contribution of this work is the thorough analysis of the forecast accuracy according to the months, extreme price levels, and small and extreme price changes. Panapakidis and Dagoumas [39] compare the forecast performances of different ANN models with various numbers of variables, layers and neurons. The main approach they apply is the clustering of the groups. According to their results, clustering gives 20% better results. Amjady et al. [40] apply fuzzy neural network, Zhao et al. [41] perform support vector machines and Pindoriya et al. [42] utilize adaptive wavelet-neural network.

1.2. Turkish market

Electricity markets differ from country to country due to several reasons. Main difference is the supply share of different production methods. When share of the renewables, i.e. wind and solar, as well as hydro power plants increase, prices tend to decrease. As Diaz and Planas [43] mention, Spanish market has many zeros, which is the minimum price allowed, as well as the price floor in the Canadian market, 0 [44]. Turkish market has the same price floor of 0 and the price cap of 2000 Turkish

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0286.v1

Peer-reviewed version available at Energies 2018, 11, 1255; doi:10.3390/en11051255

Liras/MWh (about 598 Euros/MWh, by the 2016 average exchange rate). Furthermore, as Fanone et al. [45] and Keles et al. [29] mention, many negative prices occur due to increased wind share in the German market and it needs a special attention. Ugurlu et al. [6] mention some information about the shares of the installed capacity in the Turkish market, 34.2% for hydro and 7.6% for wind. In addition to the improved technology in the other supply methods, increasing shares of hydro and wind trigger the decrease in the Turkish day-ahead electricity prices, which causes many zeros in the price series. These zeros require a special treatment and transformations prior the forecasting procedure [46], [43], [6]. Avci-Surucu et al. [47] and Ozozen et al. [48] give some information about the working mechanism of the Turkish day-ahead market. Day-ahead market is used to balance the electricity requirement one day before the physical delivery of the electricity. As in many other markets, market participants give their bids in terms of quantity and price until 11.00, and the price for each hour of the next day is determined by the market maker until 14.00 according to the intersection of the supply and demand curves. It is aimed to meet the required demand with the lowest possible price.

Turkish day-ahead electricity market has an improving literature. Hayfavi and Talasli [7] is one of the first works, which proposes a multifactor model and compares the model with [25] and [27]. The stochastic model composed of three jump processes outperforms [25] and [27] according to the comparison of the empirical moments and model moments in the daily Turkish data. Kolmek and Navruz [49] compare an artificial neural network (ANN) model with the ARIMA model. According to their results, performance of the models differ widely in respect to the selected evaluation period. But, overall, ANN model is a little better than the ARIMA model. In another work, Ozguner et al. [50] propose an ANN model to forecast the hourly electricity prices and loads in the Turkish market and compare the results with multiple linear regression. Findings of this paper is very similar to [49]; in both papers, ANN model outperforms ARIMA model with a small difference. Ozyildirim and Beyazit [51] compare another machine learning method, radial basis function, with the multiple linear regression. In their work, difference between the prediction performance of the models are negligible. [48] adapts a method from the literature to Turkish electricity prices and takes the residuals of the SARIMA forecast and puts it into ANN procedure. However; simple model of [6], which even doesn't include an exogenous variable, outperforms [48]. In our opinion, the reason of the better performance is the factorial ANOVA application of [6] on the electricity price series prior forecasting. Although the best model varies from period to period, SARIMA is chosen as the best statistical model for the Turkish day-ahead market in [6].

1.3. Deep Learning

Neural networks transform into deep neural networks (deep learning) with the addition of more layers into the neural network mechanisms. Besides, recurrent neural networks such as LSTM and GRU have started to give better results in the time series data, which triggered the application of these methods in the electricity price forecasting and related literature. RNNs have showed great success in speech recognition, handwriting recognition and polyphonic music modeling [8]. In the electricity load forecasting literature, Zheng et al. [52] apply similar days selection and empirical mode decomposition methods in addition to long short-term memory and their method outperforms many state of the art methods such as support vector regression, ARIMA or ANN. Xiaoyun et al. [53] make wind power forecast by combining principal component analysis (PCA) with LSTM. In a solar power forecast research, Gensler et al. [54] apply LSTM method with AutoEncoder and the results show that LSTM usage gives much better results than ANN. In another work, Bao et al. [55] apply very similar method to the stock price forecasting and use wavelet transformation, stacked AutoEncoders and LSTM. Hosein et al. [56] have similar findings as the superiority of the deep neural networks (various deep neural networks including LSTM ones are used) in the power load forecasting, but mention the computational complexity as a drawback. To the best of our knowledge, only deep neural networks (deep learning) application in the day-ahead electricity price forecasting literature is [57]. However, authors use only a simple multi-layer perceptron with more than single layer and do not propose a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0286.v1

Peer-reviewed version available at Energies 2018, 11, 1255; doi:10.3390/en11051255

RNN algorithm such as LSTM or GRU. Another point is that the paper's main research question is the effect of the market integration on the electricity price forecasting in Europe and deep neural network is only used as the forecast model and not compared with any other method.

In this paper, we propose to use RNNs for the time-dependent problem of electricity price estimation. To the best of our knowledge, our paper is the first study in the electricity price forecasting literature, which applies the deep RNNs, LSTM and GRU. Furthermore, these models are compared with simple deep neural networks (multi-layer ANN), single layer neural networks and the statistical time series methods. In addition to the lagged values of the price series, forecast Demand/Supply (D/S), temperature, realized D/S and balancing market prices are used as the exogenous variables. Various combinations of these features are selected to measure the effects of the variables. Moreover, Diebold-Mariano (DM) test [58] is applied to evaluate the statistical significance of the performance difference achieved with all different architectures and features.

The remainder of the paper is structured as follows. Chapter 2 gives information about the data. The neural networks based methods are described in Chapter 3 with a particular interest in RNNs. Experimental setup, methods of comparison and corresponding results are shared in Chapter 4. We conclude the paper with a detailed discussion on the results in Chapter 5.

2. Data

Turkish Day-ahead Market electricity prices are effected by various types of seasonality. Early morning hours (2.00-7.00) have relatively low prices, even some zeros. Moreover, there are double peaks in the day, one before and one after the lunch time, 11.00 and 14.00, respectively, as visualized in Figure 1. In weekly terms, Saturday morning prices are as high as the other weekdays, which shows the working pattern on the Saturday mornings. Furthermore, there are two minimums on Saturday night and Sunday night. From a seasonal point of view, both heating and cooling requirements cause high prices in winter and summer, respectively. However, due to the high share of hydro power plants in the electricity production, prices tend to decrease in spring time. An example data from each season is visualized in Figure 2.

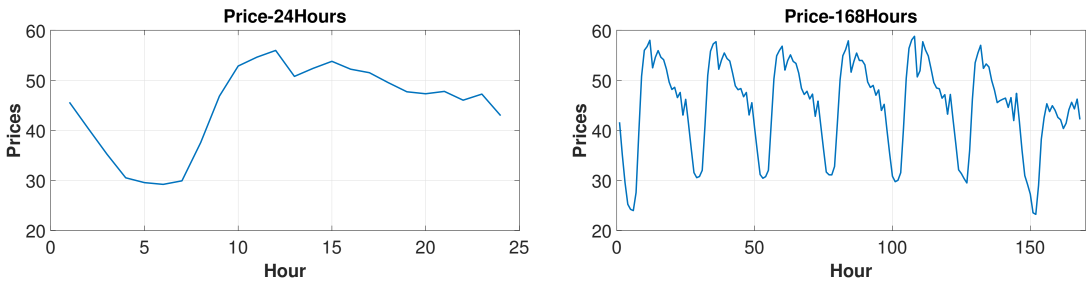


Figure 1. Left panel: Price Distribution of Hourly Prices (Euro/MWh) According to the Hours of the Day. Right panel: Price Distribution of Hourly Prices (Euro/MWh) According to the Hours of the Week (Based on 168 hours)

Hourly day-ahead electricity prices of the Turkish Day-Ahead Market are obtained from 01.01.2013 to 21.12.2016 [59]. Establishment of the Turkish Day-Ahead Market is on the 1^{st} of the December, 2011. First 13 months is excluded due to the learning-by-doing process, which limited us to start our data from 01.01.2013.

In the neural network applications first 3 years (01.01.2013-31.12.2015) are used for training and the each and every hour of the next day (01.01.2016) is predicted by using the 24-step-ahead forecast scheme. This process is repeated by using rolling window method by moving the window 24 hours in every forecast. Training period remained as 3 years and the forecast period as 24-hours of the following day. This process is repeated for 356 days of 2016. The reason of not including the last 10 days of 2016 in the forecast procedure is the very high prices, which occurred in this term due to the natural gas

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0286.v1

Peer-reviewed version available at Energies 2018, 11, 1255; doi:10.3390/en11051255

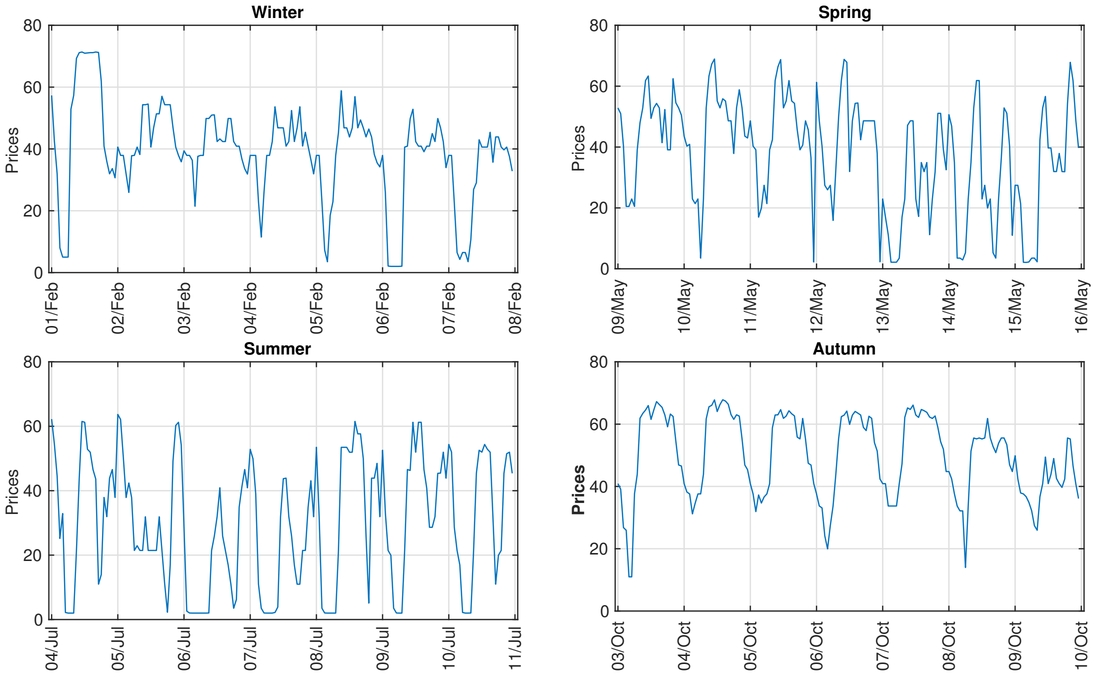


Figure 2. Price time series of sample weeks from each season

shortage and inactivity of the natural gas power plants. Prices increased up to 515 Euro/MWh on the 23rd of December, 14.00; which is 14 times higher than the average price level.

In the statistical time series methods, such as Markov, TAR and SARIMA; due to non-stationary nature of the price series and zeros; factorial ANOVA [6] transformation is applied and the series split into deterministic and stochastic parts. Then, stationary stochastic part is forecasted and added to the deterministic part values, which include the hour, weekday, month, holiday and year components. This all process is repeated in the rolling window scheme for 356 days like in the neural network methods.

Variable selection is one of the most important topics in the electricity price forecasting. In our paper, we have chosen the lagged price values as variables according to auto-correlation, partial auto-correlation functions. The chosen lags are also coherent with the lagged price series used in the literature. Furthermore, exogenous variables are also selected according to the electricity price literature [5] and [31]. Due to the high correlation between them and the independent variable, forecast D/S, temperature and the 24^{th} lags of realized D/S and balancing market price are selected as exogenous variables. One advantage is that the market maker (EPIAS) provides forecast D/S before the bids are given into the system for the next day. Another variable is temperature, which is taken from the Turkish State Meteorological Service as 81 city-based hourly temperatures. Then, annual energy consumption for all the cities is taken from EPDK [60] and energy consumption-weighted hourly temperatures (T) are calculated for every hour. Furthermore, we have also taken the 24^{th} lags of realized D/S and balancing market prices into account, because both have very high correlation with the price series and are used as variables in the literature. In addition to the above mentioned exogenous variables; 1, 23, 24, 48, 72, 168 and 336 hour lagged prices are also utilized as features to estimate the day ahead prices for the upcoming 24 hours. To report the results with aforementioned features we will use the symbols stated in Table 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0286.v1

Peer-reviewed version available at Energies 2018, 11, 1255; doi:10.3390/en11051255

Table 1. Utilized features for electricity price estimation

Symbol	Feature
F1	24-hour lagged price
F2	168-hour lagged price
F3	1-hour lagged price
F4	48-hour lagged price
F5	23-hour lagged price
F6	72-hour lagged price
F7	336-hour lagged price
F8	Forecast demand over supply
F9	Temperature
F10	Realized demand/supply with 24 hours lag
F11	Balancing market price with 24 hours lag

3. Methods

In this section, we describe the Neural Network architectures we used for electricity price estimation. A simple neural network with three input neurons is visualized in Figure 3. The guiding equation of a neuron can be described as:

$$Y = f\left(\sum_{i}^{Inputs} (x_i w_i + b_i) \right) \tag{1}$$

Equation 1 calculates the output of a neuron, where x is the input of the neuron, w is the weight on each connection to the neuron, b is the bias and f is the activation function and the Rectified Linear Unit (ReLU) is the activation function in our experiments.

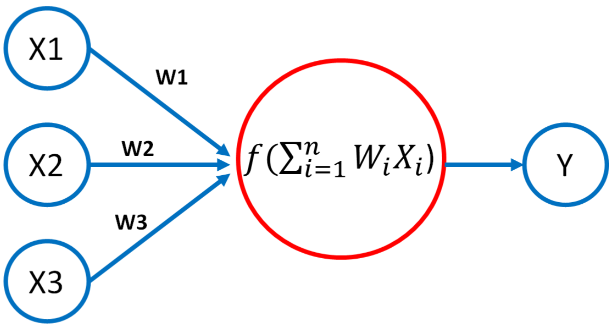


Figure 3. Simple Neural Network

In section 3.1 basic neural network structure, Artificial Neural Networks, is defined. In section 3.2 we give a brief definition of Convolutional Neural Networks and their application on the time series data for electricity price estimation. Then, we move to RNNs in section 3.3, which is the focal point of our work. In section 3.3.1 we define the LSTM networks and their benefits for time series prediction tasks. Finally, in section 3.3.2 we define the GRUs and their fundamental differences from LSTMs.

3.1. Artifical Neural Networks

ANN is a basic architecture of a neural network, which consists of multiple fully connected layers of neurons [61]. This type of networks are also known as Multi-layer Perceptrons (MLP) and they are early examples of the neural networks. We use a shallow network with a single layer with 10 neurons and a deeper 3-layer network each consisting of 10 neurons for our experiments. We add a final layer to estimate the target values.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0286.v1

Peer-reviewed version available at Energies 2018, 11, 1255; doi:10.3390/en11051255

3.2. Convolutional Neural Networks

Convolutional Neural Networks have been successfully applied to many problems in computer vision [10] and medical image analysis [62]. In our application, the convolutional layers are constructed using one-dimensional kernels that move through the sequence (unlike images where 2D convolutions are used). These kernels act as filters which are being learned during training. As in many CNN architectures, the deeper the layers get, the higher the number of filters become. We use two convolutional layers and a final fully connected layer for prediction. Each convolution is followed by pooling layers to reduce the sequence length.

3.3. Recurrent Neural Networks

RNNs are networks with loops in them, allowing information to persist. They [63] are used to model time-dependent data, like words in a sentence. We feed in words one by one, and the nodes in the network store their state at one time step and use it to inform the next time step. Unlike MLP, RNNs use temporal information of the input data, which make them more appropriate for time series data. An RNN realizes this ability by recurrent connections between the neurons. A general equation for RNN hidden state h_t given an input sequence $x = (x_1, x_2, \ldots, x_T)$:

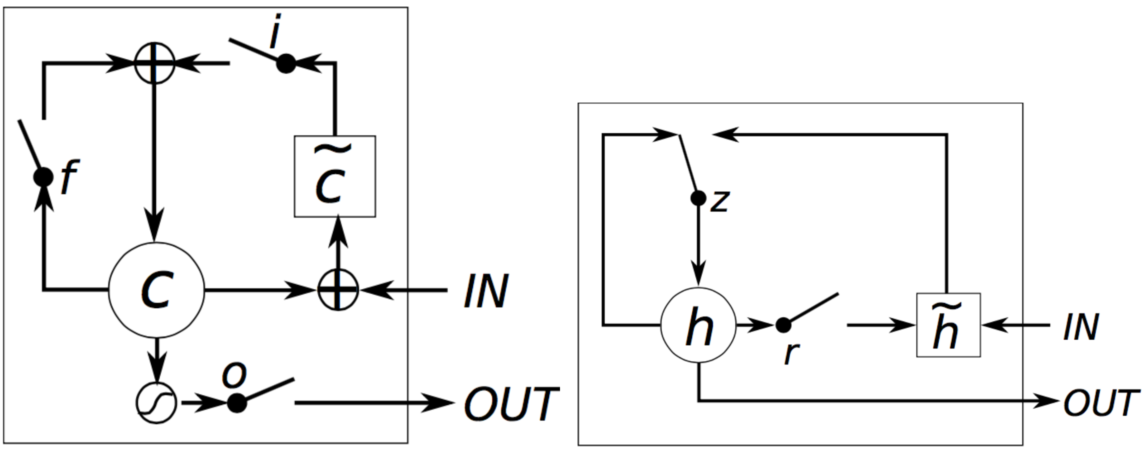
$$h_t = \begin{cases} 0, & \text{if } (t = 0) \\ \phi(h_{t-1}, x_t), & \text{otherwise} \end{cases} \tag{2}$$

where ϕ is a non-linear function. The update of recurrent hidden state is realized as:

$$h_t = g(Wx_t + Uh_{t-1}) \tag{3}$$

where g is a hyperbolic tangent function.

In general this generic setting of RNN without memory cells suffer from vanishing gradient problems. In this paper, we investigate the performance of two RNNs with memory cells for electricity price forecasting namely, LSTMs and GRUs.



(a) Long Short Term Memory (LSTM) **(b)** Gated Recurrent Unit (GRU)

Figure 4. Illustration of (a) LSTM and (b) GRU. (a) i, f and o are the input, forget and output gates, respectively. c and $\tilde{c}$ denote the memory cell and the new memory cell content. (b) r and z are the reset and update gates, and h and $\tilde{h}$ are the activation and the candidate activation. (figure adapted from [64])

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0286.v1

Peer-reviewed version available at Energies 2018, 11, 1255; doi:10.3390/en11051255

3.3.1. Long Short-Term Memory Networks

LSTM (Long Short-Term Memory Networks) [65] is a special subset of RNN that is able to deal with remembering information for much longer periods of time. The idea behind an LSTM is using each node as a memory cell that can store other information instead of being simply a node with a single activation function. Specifically, it maintains its own cell state. Normal RNNs take in their previous hidden state and the current input, and output a new hidden state. An LSTM does the same, except it also takes in its old cell state and outputs its new cell state c_t^j. This property helps LSTMs to address the vanishing gradients problem from the previous time-steps.

LSTM has three gates: input gate i_t, forget gate f_t and output gate o_t as visualized in Figure 4a. All gates are generated by a sigmoid function over the ensemble of input s_t and the preceding hidden state h_{t-1}. In order to generate the hidden state at current step t, it first generates a temporary result q_t by a tanh non-linearity over the ensemble of input x_t and the preceding hidden state h_{t-1}, then combines this temporary result g_t with history p_{t-1} by input gate i_t and forget gate, f_t respectively, to get an updated history p_t, finally uses output gate o_t over this updated history p_t to get the final hidden state h_t.

We visualize the LSTM structure in Figure 4a to define the guiding equations of LSTM. The hidden state h_t^j of LSTM unit is defined as:

$$h_t^j = o_t^j \tanh(c_t^j)$$

where o_t^j modulates the memory influence on the hidden state. The output gate is computed as:

$$o_t^j = \sigma(W_o x_t + U_o h_{t-1} + V_o ct)^j \quad ,$$

where σ is the logistic sigmoid function and V_o is a diagonal matrix. The memory cell c_t^j is updated partially following the equation;

$$c_t^j = f_t^j c_{t-1}^j + i_t^j \tilde{c}_t^j \quad ,$$

where the memory content is defined by a hyperbolic tangent function:

$$\tilde{c}_t^j = \tanh(W_c x_t + U_c h_{t-1})^j$$

The extent to which the existing memory is forgotten is modulated by a forget gate f_t^j, and the degree to which the new memory content is added to the memory cell is modulated by an input gate i_t^j. Gates are computed by:

$$f_t^j = \sigma(W_f x_t + U_f h_{t-1} + V_f c_{t-1})^j$$
$$i_t^j = \sigma(W_i x_t + U_i h_{t-1} + V_i c_{t-1})^j$$

Unlike the traditional recurrent unit which overwrites its content at each time-step, an LSTM unit is able to decide whether to keep the existing memory via the introduced gates. Intuitively, if the LSTM unit detects an important feature from an input sequence at early stage, it easily carries this information (the existence of the feature) over a long distance, hence, capturing potential long-distance dependencies.

3.3.2. Gated Recurrent Units

A GRU [66] has two gates, a reset gate r, and an update gate z as visualized in Figure 4b. The update gate defines how much of the previous memory to be kept and the reset gate determines how to combine the new input with the previous memory. GRUs become equivalent to RNNs, if the reset gates are all 1 and update gates all 0.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0286.v1

Peer-reviewed version available at Energies 2018, 11, 1255; doi:10.3390/en11051255

Following [64], we formulate the guiding equations. The activation h_t^j of the GRU at time t is a linear interpolation between the previous activation h_{t-1}^j and the candidate activation h_t^j:

$$h_t^j = (1 - z_t^j)h_{t-1}^j + z_t^j \tilde{h}_t^j$$

where an update gate z_t^j decides how much the unit updates its activation, or content. The update gate is computed by:

$$z_t^j = \sigma(W_z x_t + U_z h_{t-1})^j$$

This procedure of taking a linear sum between the existing state and the newly computed state is similar to the LSTM unit. The GRU, however, does not have any mechanism to control the degree to which its state is exposed, but exposes the whole state each time. The candidate activation $\tilde{h}_t^j$ is computed similarly to that of the traditional recurrent unit

$$h_t^j = \tanh(W x_t + U(r_t \odot h_{t-1}))^j$$

where r_t is a set of reset gates and $\odot$ is an element-wise multiplication. When off (r_t^j close to 0), the reset gate effectively makes the unit act as if it is reading the first symbol of an input sequence, allowing it to forget the previously computed state. The reset gate r_t^j is computed similarly to the update gate:

$$r_t^j = (W_r x_t + U_r h_{t-1})^j$$

GRUs have the same fundamental idea of gating mechanism to learn long-term dependencies compared to LSTM, but there are couple of significant differences. First of all, GRU has two gates and less parameters compared to LSTM. The input and forget gates are coupled by an update gate z and the reset gate r is applied directly to the previous hidden state in GRUs. In other words, the responsibility of the reset gate in an LSTM is divided into both reset gate r and the update gate z. GRUs don't possess any internal memory that is different from the exposed hidden state. They don't have the output gate that is present in LSTMs. Also, in LSTMs there is a second non-linearity applied when computing the output, which is not present in GRUs.

4. Results

This section offers a qualitative and quantitative analysis of the proposed method, as well as quantitative comparison of our proposed method w.r.t. state-of-the-art methods, to demonstrate its effectiveness for electricity price estimation.

Our quantitative analysis consists of comparing our method with others and also looking into monthly and weekly performance. In the following we describe the evaluation metrics section 4.1 and then describe the state of the art statistical methods in section 4.2. We report the quantitative results achieved by all network types with a different combination of layers in section 4.3 and evaluate the statistical significance in section 4.4. Finally, we mention some implementation details about the neural network training and hyper-parameters in section 4.5.

4.1. Evaluation metrics

In the performance evaluation of the forecast techniques, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) are the most used techniques. Although MAPE gives opportunity to compare the electricity price forecasts' performances from various markets; for the prices around zero, it does not give interpretable results. For zeros, MAPE can not be calculated; for negative prices, there are negative values, which are meaningless; and for small positive prices, MAPE values are very high. In the comparisons, there is not an important difference between the MAE and RMSE values, because both of them are based on the absolute errors

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0286.v1

Peer-reviewed version available at *Energies* **2018**, *11*, 1255; doi:10.3390/en11051255

[6]. Therefore, MAE method is used as the performance evaluation criterion in this paper. Equation 4 shows the MAE formula.

$$MAE = \frac{1}{T}\sum_{i=1}^{T}\left|P_i - \hat{P}_i\right| \tag{4}$$

4.2. State of the art statistical methods

Traditionally, Naive method, SARIMA, Markov regime-switching and Self exciting threshold auto-regressive regression (SETAR) have been used with great success for time series estimation in the electricity price forecasting literature. We compare the robustness of these techniques with the neural network architectures.

4.2.1. Naive method

One of the most important benchmark techniques in the electricity price forecasting literature, naive method [14], can be found below in Equation 5. According to [14] and [35], forecasting methods, which are not calibrated well, can not outperform the naive method.

$$P_{d,h} = \begin{cases} P_{d-7,h} + \epsilon_{d,h}, & \text{Monday, Saturday, Sunday} \\ P_{d-1,h} + \epsilon_{d,h}, & \text{Tuesday, Wednesday, Thursday, Friday} \end{cases} \tag{5}$$

$P_{d,h}$ states the price of the selected day and hour. $\epsilon_{d,h}$ stands for the noise term.

4.2.2. Markov regime-switching auto regressive (MS-AR) model

As another benchmark method, 2-state Markov regime-switching auto regressive model [67] with the $1^{st}, 24^{th}, 48^{th}$and168^{th} lags of the price series are used in the estimation. This method allows the observations to be distributed into different states by a latent variable. Equation 6 related to MS-AR model can be found below.

$$y_t = a_s + \sum_{i=1}^{p}\phi_{s,i}\, y_{t-i} + \epsilon_t, \tag{6}$$

where s_t is a 2-state discrete Markov-chain with S= 1,2 and $\epsilon_t \sim i.i.d.\ N(0,\sigma^2)$. The estimation of the MS-AR model is performed by maximum likelihood algorithm [68].

4.2.3. Self-exciting threshold auto-regressive (SETAR) model

Threshold auto-regressive (TAR) models are similar to Markov regime-switching models in terms of placing the observations into different groups. The main difference of the TAR models is that the threshold variable is observable compared to the latent one in the Markov models. TAR models allow to choose the threshold according to an exogenous variable. If the threshold variable is selected according to a lagged value of the dependent variable, then it is called SETAR model. In Equation 7, SETAR model is given.

$$x_t = \phi_0^{(j)} + \phi_1^{(j)}\, x_{t-1} + ... + \phi_p^{(j)}\, x_{t-p} + a_t^{(j)},\ \text{if } \gamma_{j-1} \le x_{t-d} \le \gamma_j \tag{7}$$

where k and d are positive integers; j= 1, ... , k; γ_i are real numbers such that $-\infty = \gamma_0 < \gamma_1 < ... < \gamma_{k-1} < \gamma_k = \infty$, the superscript (j) is used to signify the regime, and $\alpha_t^{(j)}$ are i.i.d. sequences with mean 0 and variance σ_j^2 and are mutually independent for different j. The parameter d is the delay parameter for different regimes [69].

As in Markov model, $1^{st}, 24^{th}, 48^{th}$and168^{th} lags of the price series are used in the estimation, in addition to the delay parameter, d=1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0286.v1

Peer-reviewed version available at Energies 2018, 11, 1255; doi:10.3390/en11051255

4.2.4. Seasonal auto-regressive integrated moving average (SARIMA) model

ARIMA is a special kind of regression, which takes the past prices (AR), previous values of the noise (MA) and the integration level (I) of the price series into account. In SARIMA, seasonal component (S) also gets involved into the estimation process. Generally only intra-weekly nature of the series is incorporated as a seasonal component, but in the electricity price series, it is required to deal with the intra-daily and intra-yearly seasonality as well. Therefore, triple SARIMA model of [70] is performed by maximum likelihood assuming Gauss-Newton optimization. Equation 8 refers to the triple SARIMA model.

$$\phi_p(L)\Phi_{p_1}(Ł^{s_1})\Omega_{P_2}(Ł^{s_2})\Gamma_{P_3}(Ł^{s_3})(y_t - a - bt) = \theta_q(L)\Theta_{Q_1}(Ł^{s_1})\Psi_{Q_2}(Ł^{s_2})\Lambda_{Q_3}(Ł^{s_3})\epsilon_t \quad (8)$$

y_t is load in period t, a is a constant term, b is the coefficient of linear deterministic trend term, ϵ_t is a white noise error term, Ł is the lag operator, $\phi_p, \Phi_{p_1}, \Omega_{P_2}, \Gamma_{P_3}, \theta_q, \Theta_{Q_1}, \Psi_{Q_2}$ Λ_{Q_3} are the polynomial functions of orders $p, p_1, P_2, P_3, q, Q_1, Q_2$ and Q_3, respectively [70].

Our triple SARIMA model can be stated as $(1,0,1)_1 x(1,0,1)_{24} x(1,0,1)_{168}$. To comply with the other statistical methods, ARMA(48,48) component is also added to this model.

4.3. Quantitative Analysis

In this section, we report the performance analysis of neural networks in comparison with the state of the art methods. We also use a different combination of features for shallow and deep networks to analyze the prediction accuracy. Finally, we report the monthly average results and illustrate the price estimation accuracy of GRU on a graph.

4.3.1. Comparison with the state of the art methods

In our first experimental setup, we use key features of lagged price values 1, 24, 48 and 168 on all described algorithms to compare the 1-layered neural network algorithm performance with the state of the art methods. Results in Table 2 indicate the neural network models' success compared to the statistical ones. Recurrent neural networks, LSTM and GRU are the best methods in this comparison. As a note, naive method outperforms 2 other methods, which is in line with the findings of [14], [13] and [35], mentioning the relatively good performance of naive method.

Table 2. Single-layer day ahead prediction results comparison of neural network based methods with state of the art techniques

Features	Markov	Naive	SETAR	SARIMA	CNN	ANN	LSTM	GRU
F1-4	8.04	7.95	7.89	7.29	9.82	6.37	5.91	5.71

4.3.2. Shallow Network Comparison

Our first comparison is on shallow network architectures to see the performance of each neural network method. We experiment different network architectures using many different combination of features from Table 1 following the findings of the literature. Table 3 demonstrates, addition of new variables into the single-layer neural networks. It should be stated that the addition of 1^{st} and 48^{th} lagged values of the price series to the 24^{th} and 168^{th} lags decrease the MAE values, but addition of the exogenous variables do have a very little or even negative effect.

Table 3. Single-layer day ahead prediction MAE results. Each network of 1-layer and a final fully connected layer for prediction. CNNs have been implemented 2 convolutional layers stacked together.

Features	CNN	ANN	LSTM	GRU
F1-2	9.82	8.51	7.79	7.70
F1-4	8.57	6.37	5.91	5.71
F1-7	9.47	6.65	6.01	5.64
F1-8	10.05	8.05	6.22	5.83
F1-9	10.51	9.27	6.16	5.83
F1-10	10.64	9.85	6.02	5.58
F1-11	10.58	9.48	5.93	5.55

4.3.3. Deep Network Comparison

To showcase the performance of deeper networks we concatenate 3-layers for simple ANNs, LSTMs and GRUs. It is evident in Table 4, the GRU still performs the best compared to other techniques. The multiple layer structure comes up with an additional computational cost and to find the optimal number of layers, we do a test on the algorithms.

In this deep neural networks comparison, CNN is excluded due to the low performance. Addition of the new layers increase the performance in every neural networks mechanism. However, positive effect of the additional variables are still very small, which is in line with our findings in the shallow network comparison section.

Table 4. Multi-layer day ahead prediction MAE results. Each network of stacked 3-layers and a final fully connected layer for prediction.

Features	ANN-3	LSTM-3	GRU-3
F1-2	7.63	7.66	5.86
F1-4	5.66	5.66	5.68
F1-7	5.59	5.58	5.57
F1-8	5.84	5.62	5.56
F1-9	6.08	5.70	5.57
F1-10	6.29	5.51	5.41
F1-11	6.20	5.47	**5.36**

4.3.4. Monthly Comparison

We also evaluate the monthly performance of each technique in Table 5. The results for each month are generally consistent with the overall average performance with some exceptional cases. Results demonstrate the relatively good performance of the LSTM and GRU models. Although there are some months that single-layer is better than the multi-layer neural networks, in most of the months deep neural networks give much better results. With the exception of Naive method in August and ANN-3-layer in October, recurrent neural networks, LSTM and GRU, have the best results in every month.

Peer-reviewed version available at Energies 2018, 11, 1255; doi:10.3390/en11051255

Table 5. Monthly MAE compassion of all the price estimation methods

Months	Markov	Naive	SETAR	SARIMA	CNN	ANN	ANN-3	LSTM	LSTM-3	GRU	GRU-3
January	9.15	10.27	8.84	8.92	10.73	8.73	5.75	7.16	5.82	6.58	5.61
February	8.59	10.85	8.58	7.89	9.35	7.28	5.89	6.93	5.85	6.47	5.67
March	8.66	8.93	8.58	8.17	10.77	6.37	6.15	6.56	6.10	6.23	6.09
April	10.35	10.18	10.21	9.33	9.23	7.56	7.11	7.16	7.17	7.04	7.22
May	10.53	11.32	10.40	9.45	10.60	8.03	7.68	8.04	7.64	7.66	7.70
June	8.80	7.63	8.54	7.62	9.89	6.92	6.68	6.14	6.64	6.16	6.87
July	10.08	9.05	9.89	8.37	10.45	7.90	6.88	6.85	6.86	6.66	6.97
August	6.11	4.42	5.92	5.00	6.32	5.14	4.99	4.60	5.03	4.66	5.16
September	6.84	6.59	6.70	6.06	6.89	5.56	5.32	5.16	5.27	5.12	5.37
October	5.75	4.71	5.64	5.45	5.40	4.25	3.84	3.95	3.87	3.86	3.92
November	5.55	4.93	5.39	5.18	5.64	3.97	3.84	3.96	3.82	3.84	3.83
December	6.36	9.68	6.17	6.44	8.20	4.95	3.79	4.64	3.84	4.38	3.76

4.3.5. Seasonal Prediction Results

We illustrate the prediction results of GRU for the sample weeks from each season we defined in Section 2. Figure 5 shows the successful performance of GRU with a good match to the original prices. The ability of capturing the spikes as well as the good performance in relatively calm periods. It is clear that the performance of the GRU model is great in the relatively calmer autumn week. Moreover, the performance in the summer week, which has a high volatility, gives evidence about the spike detection of the model.

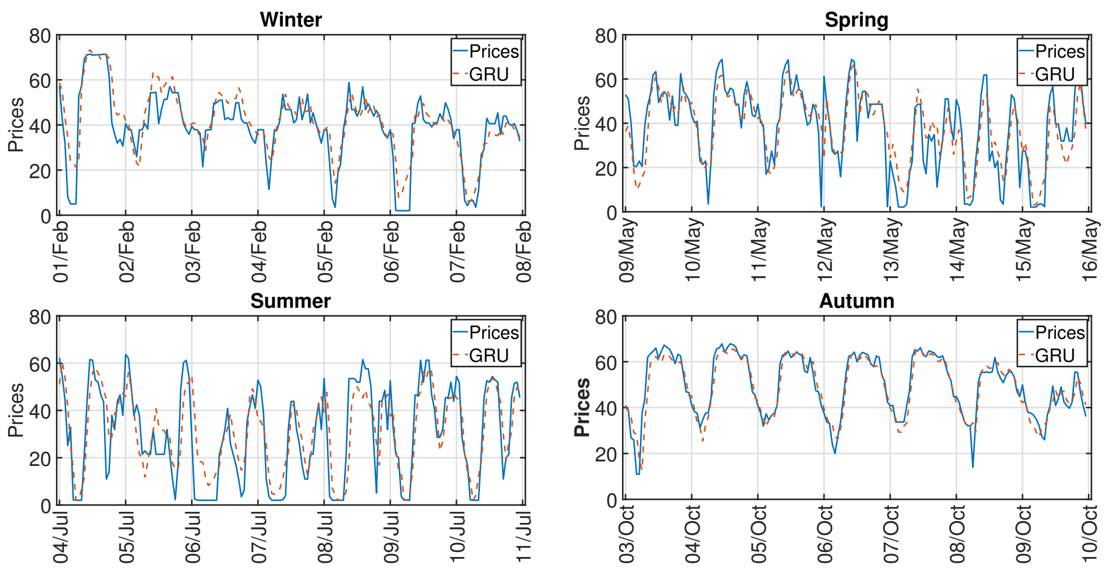


Figure 5. Prediction results of GRU for a sample day from each season

4.4. Diebold-Mariano Tests

The Tables 2, 3 and 4 can be used to provide a ranking of the various methods, but not statistically significant conclusions on the performance of the forecasts of one method compared to others. To showcase the statistical significance of the performance difference between all model variations and features combinations, we use a Diebold-Mariano test [58], which takes the correlation structure into account. In Figure 6, we show the p-values for the Diebold-Mariano tests between neural network based methods and the state of the art statistical methods. In Figure 7 we repeat the same tests for shallow and deep networks using different number of features. It tests the forecasts of each pair of transformations against each other and uses a color map to show p-values. The low p-values show

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0286.v1

Peer-reviewed version available at *Energies* **2018**, *11*, 1255; doi:10.3390/en11051255

statistically significant better performance of the methods in X-axis. For example, F1-11 GRU model outperforms all the other models significantly in the 3-layer networks comparison, Figure 7b.

Figure 6 demonstrates the performance of the neural networks models, except CNN, compared to the statistical methods. Especially, good performance of the recurrent neural network models, GRU and LSTM, is statistically proven by Diebold-Mariano test

In Figure 7a, single layer networks are compared with each other. F1-10 GRU and F1-11 GRU are significantly better than all the other models. Performance of F1-7 GRU and F1-4 LSTM, which do not include any exogenous variables, should also be mentioned. In Figure 7b, in 3-layer networks, addition of new features has a much more significant effect than the single layer network. As demonstrated; F1-11 GRU, F1-10 GRU, F1-11 LSTM, F1-10 LSTM are the best methods in 3-layer networks.

4.5. Implementation Details

The training of a neural network can be viewed as a combination of two components, a loss function or training objective, and an optimization algorithm that minimizes this function. In this study, we use the Adam optimizer to minimize the mean absolute error loss function. The training ends when the network does not significantly improve its performance on the validation set for a predefined number of epochs (300).

During training, a batch-size of three years is used. The momentum of the optimizer is set to 0.90 and the learning rate was 0.001. The parameters of the fully-connected, convolutional, and recurrent layers are initialized randomly from a zero-mean Gaussian distribution. In each trial, training is continued until the network converged. Convergence is defined as a state, in which no substantial progress was observed in the training loss.

We have done multiple tests to see the performance of different number of layers in ANN, LSTM and GRU architectures for selecting the optimal number of layers. Figure 8 shows that the optimal results can be achieved using 3-layers. Additional layers increase the total number of parameters and adds to the computational cost without achieving a significant gain in the performance.

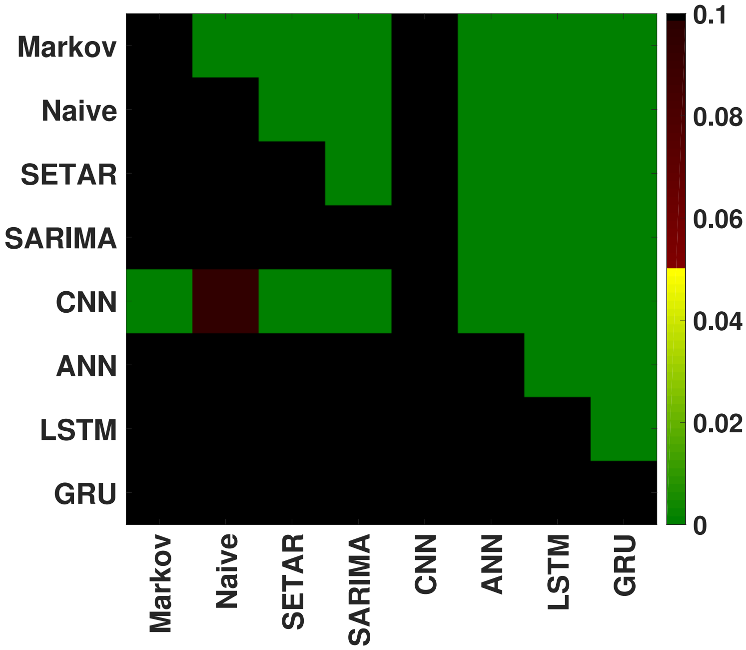
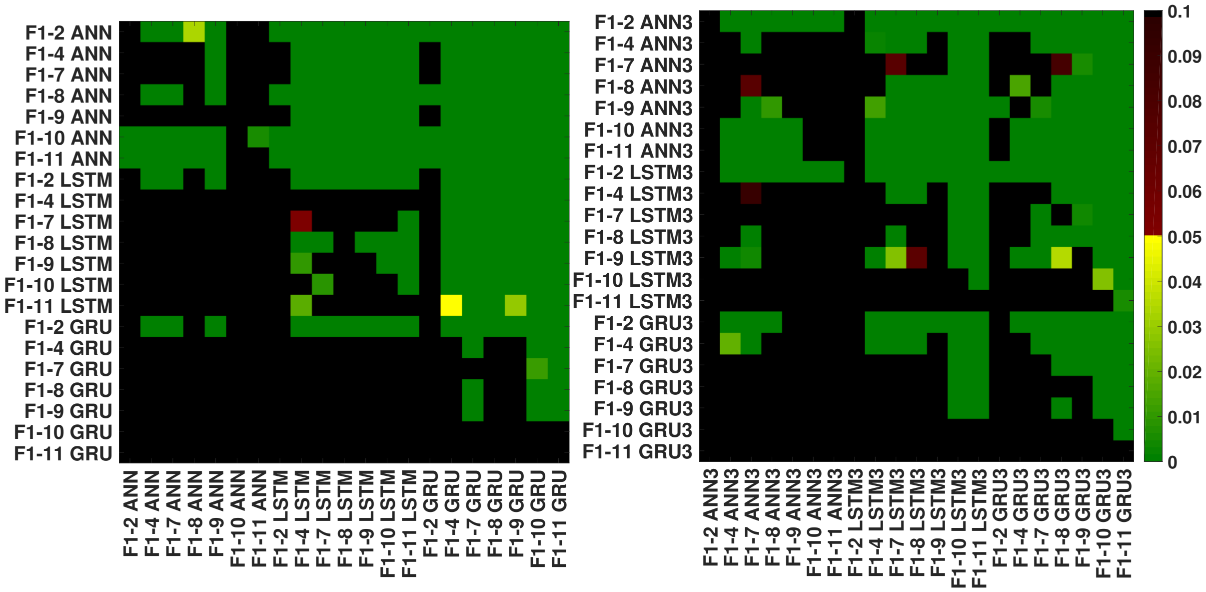


Figure 6. Results of the Diebold-Mariano tests defined by the loss differential series in between all investigated parameters for F1-4. The figure indicates the statistical significance (green) for which the forecasts of a model on the X-axis are significantly better than those of a model on the Y-axis.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0286.v1

Peer-reviewed version available at *Energies* **2018**, *11*, 1255; doi:10.3390/en11051255



(a) Single Layer Networks **(b)** 3-layer Networks

Figure 7. Results of the Diebold-Mariano tests defined by the loss differential series in between all investigated parameters and used features for different number of layers. The figure indicates the statistical significance (green) for which the forecasts of a model on the X-axis are significantly better than those of a model on the Y-axis.

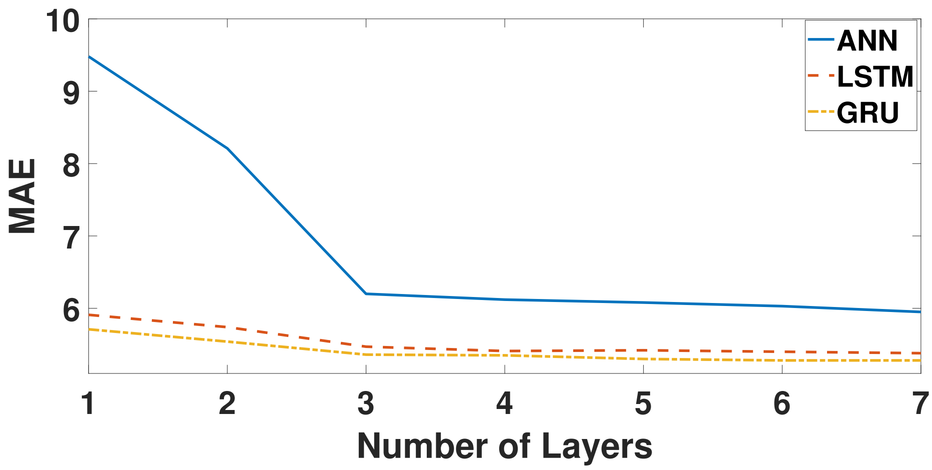


Figure 8. Performance change when applying different number of layers to ANN, LSTM and GRU algorithms

5. Discussion

With this paper, we investigate the application of various neural network architectures on electricity price forecasting. Our experiments in Table 2 highlight that neural network based methods produce better results compared to the state of the art statistical forecasting methods in the literature such as SARIMA and Markov models. We use simple artifical neural networks (ANNs), CNNs, LSTMs and GRUs to estimate the electricity prices in the Turkish market. We see that the RNN models namely LSTM and GRU are able to separate themselves in terms of performance compared to CNNs and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 April 2018 doi:10.20944/preprints201804.0286.v1

Peer-reviewed version available at Energies 2018, 11, 1255; doi:10.3390/en11051255

simple ANNs in Table 3. This is due to the fact that RNN models have memory about the previous time steps, which makes them the method of choice for time series type problems. They keep a memory of the previous instances effectively, which is crucial for estimating electricity prices of the day-ahead market.

The deep learning paradigm of stacking multiple layers increases the performance for ANNs, LSTM and GRUs as highlighted in Table 3 in comparison with Table 4. GRUs still give the best performance among all available techniques and we reached the best results of 5.36 Euros/MWh MAE using 3-layered GRUs. The results show good alignment with the prices as illustrated in Figure 5.

Neural networks are data-driven models and their performance heavily depends on the availability of the large training data. The limited data is a deteriorating factor for all training based methods, but in particular for neural network based methods. We show in Figure 8 that the performance does not improve after 3-layers for any of the networks due to the limited data. With the availability of further data, we believe the overall performance of LSTM and GRU methods will be better.

Another significant observation is the fact that GRUs perform better than the LSTM models. This can explained by the fewer number of parameters that are needed to be learned by GRUs. In the literature [71], [64] have compared the two models for polyphonic music modeling and speech signal modeling task and have seen the better performance of GRU for these tasks. Moreover, GRUs train faster due to the fact that they require fewer parameters.

We see that the key features are lagged price values for estimating the electricity prices, which is in line with the findings of [5]. In terms of single layer, addition of 1^{st} and 48^{th} lagged values to the 24^{th} and 168^{th} lagged values have an important effect. Especially for LSTM single layer using the 1^{st}, 24^{th}, 48^{th} and 168^{th} lagged values is as good as using all the variables. For GRU adding 23^{rd}, 72^{nd} and 336^{th} lagged values give better results. Addition of exogenous variables have a very small effect in LSTM. Although addition of forecast D/S and temperature do not have a significant effect in GRU, further addition of 24^{th} lags of realized D/S and Balancing market price have significant effects. In 3-layer networks, results are similar, but addition of features help much more to have better results. Instead of F1-4, F1-7 give better results; if we don't use any exogenous variables. In GRU-3-layer networks, addition of all the variables, except temperature, change the performance significantly. On the other hand, LSTM F1-7 is only worse than LSTM F1-10 and F1-11, which is similar to the single layer results. To conclude, endogenous variables are the most important ones and using the 1^{st}, 24^{th}, 48^{th} and 168^{th} lagged prices give relatively good results. In most cases, adding one or two exogenous variables don't help to have better results, but if we use the lagged values of the other exogenous variables, in addition to forecast D/S and temperature, then these models with all the variables significantly outperform the models with less variables.

One additional comparison we made was grouping the results in terms of months. It is possible to say that the general error levels are lower in autumn and winter months compared to spring and summer months. In relatively mild weather months of Turkey; October, November and December, GRU-3-layer networks' MAE values are lower than 4 Euros/MWh. On the other hand, relatively hot weather months of Turkey; May, June, July have MAE values around 7 Euros/MWh, which is almost double of the mild weather months. It must be mentioned that in most of the countries, summer months' prices are not high compared to the other months, but as mentioned in the Turkish market section 1.2, due to the requirement of air conditioning, summer months prices are very close to the winter months prices. We can conclude that the MAE values show a similar pattern with the price levels, which demonstrate the effect of the seasonality.

Generalization capability of machine learning models is promising for applying our model for different market data. The GRU network architecture is capable to accurately predict the electricity prices in the Turkish Market. With the availability of the multiple feature data for each market, the model can be applied to various markets. However, Aggarwal et al. [72] underline the superiority of different methods in different markets and combination of multiple methods might be promising in this type of problems. We would like to investigate possibility of using hybrid models to merge benefits of multiple methods. Zhang [73] proposed to combine ARIMA and ANN models to froecast the linear and non-linear components of price separately. Chaabanae [74] develops the Zhang [73]'s method and combines auto-regressive fractionally integrated moving average (ARFIMA) with neural networks model. These types of hybrid approaches can aid the performance of RNNs.

One avenue of improvement for our method is to investigate the decomposition techniques. Hong and Wu [75] apply principal component analysis (PCA) as a dimension reduction method, Ziel [76] and Ludwig et al. [77] use Lasso shrinkage method for variable selection. Zheng et al. [52] propose to use empirical mode decomposition for decomposing the signal to several intrinsic mode functions (IMFs) and residuals. They used these IMFs to train LSTM to forecast short-term load. In future, we would like to include dimension reduction algorithms and investigate their contribution to seasonality of the data, in particular in RNN setting.

In conclusion, this study instigates the utility of neural networks for electricity price estimation. Development of new conditions in electricity markets across the world brings new challenges. Accurate price estimation is a crucial task for adapting to the new market conditions, and the machine learning methods are capable to address these issues with high accuracy. Recurrent Neural Networks set the state of the art in addressing time-dependent problems. With this work, we show a detailed analysis on RNNs for prices forecasting and highlight the superior performance of GRUs in comparison to various neural network based methods and state of the art statistical techniques.

Acknowledgments: Ilkay Oksuz was supported by an EPSRC programme Grant (EP/P001009/1) and the Wellcome EPSRC Centre for Medical Engineering at School of Biomedical Engineering and Imaging Sciences, King's College London (WT 203148/Z/16/Z). This work was supported by Research Fund of the Istanbul Technical University; project number: SDK-2018-41160. Furthermore, Umut Ugurlu was supported by The Scientific and Technological Research Council of Turkey, 2214/A Programme. The GPU used in this research was generously donated by the NVIDIA Corporation. We also thank Tolga Kaya and Anirban Mukhopadhyay for fruitful discussions.

Author Contributions: Ilkay Oksuz and Umut Ugurlu basically conducted all numerical simulations for the current manuscript which included all figures and tables under the supervision of Oktay Tas. In particular, Umut Ugurlu wrote the Introduction and Data sections. Ilkay Oksuz wrote the Methods and the Results section. Ilkay Oksuz has implemented the algorithms and Umut Ugurlu has done the experiments with the implementations. Ilkay Oksuz has generated the figures for the manuscript and has done the statistical significance analysis. Oktay Tas has provided useful suggestions for data analysis and discussed research progress.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Keles, D.; Scelle, J.; Paraschiv, F.; Fichtner, W. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. *Applied Energy* **2016**, *162*, 218–230.
2. Carmona, R.; Coulon, M.; Schwarz, D. Electricity price modeling and asset valuation: a multi-fuel structural approach. *Mathematics and Financial Economics* **2013**, *7*, 167–202.
3. Zareipour, H.; Canizares, C.A.; Bhattacharya, K. Economic Impact of Electricity Market Price Forecasting Errors: A Demand-Side Analysis. *IEEE Transactions on Power Systems* **2010**, *25*, 254–262.
4. Hong, T. Crystal ball lessons in predictive analytics. *EnergyBiz Mag* **2015**, pp. 35–37.
5. Uniejewski, B.; Nowotarski, J.; Weron, R. Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies* **2016**, *9*, 621.
6. Ugurlu, U.; Tas, O.; Gunduz, U. Performance of Electricity Price Forecasting Models: Evidence from Turkey. *Emerging Markets Finance and Trade* **2018**.
7. Hayfavi, A.; Talasli, I. Stochastic multifactor modeling of spot electricity prices. *Journal of Computational and Applied Mathematics* **2014**, *259*, 434–442.
8. Greff, K.; Srivastava, R.K.; Koutnik, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems* **2017**, *28*, 2222–2232.
9. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **1998**, *86*, 2278–2324.

10. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, pp. 1097–1105.
11. Szkuta, B.; Sanabria, L.A.; Dillon, T.S. Electricity price short-term forecasting using artificial neural networks. *IEEE Transactions on Power Systems* **1999**, *14*, 851–857.
12. Bunn, D.W. Forecasting loads and prices in competitive power markets. *Proceedings of the IEEE* **2000**, *88*, 163–169.
13. Contreras, J.; Espinola, R.; Nogales, F.J.; Conejo, A.J. ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems* **2003**, *18*, 1014–1020.
14. Nogales, F.J.; Contreras, J.; Conejo, A.J.; Espinola, R. Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems* **2002**, *17*, 342–348.
15. Shahidehpour, M.; Yamin, H.; Li, Z. Market Operations in Electric Power Systems, New York, NY: IEEE, 2002.
16. Cuaresma, J.C.; Hlouskova, J.; Kossmeier, S.; Obersteiner, M. Forecasting electricity spot-prices using linear univariate time-series models. *Applied Energy* **2004**, *77*, 87–106.
17. Bunn, D.W. *Modelling prices in competitive electricity markets*; Wiley Finance, John Wiley & Sons, 2004.
18. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting* **2014**, *30*, 1030–1081.
19. Shafie-khah, M.; Catalão, J.P. A stochastic multi-layer agent-based model to study electricity market participants behavior. *IEEE Transactions on Power Systems* **2015**, *30*, 867–881.
20. Ziel, F.; Steinert, R. Electricity price forecasting using sale and purchase curves: The X-Model. *Energy Economics* **2016**, *59*, 435–454.
21. Coulon, M.; Howison, S. Stochastic behavior of the electricity bid stack: from fundamental drivers to power prices. *The Journal of Energy Markets* **2009**, *2*, 29–69.
22. Carmona, R.; Coulon, M. A survey of commodity markets and structural models for electricity prices. In *Quantitative Energy Finance*; Springer, 2014; pp. 41–83.
23. Füss, R.; Mahringer, S.; Prokopczuk, M. Electricity derivatives pricing with forward-looking information. *Journal of Economic Dynamics and Control* **2015**, *58*, 34–57.
24. Geman, H.; Roncoroni, A. Understanding the fine structure of electricity prices. *The Journal of Business* **2006**, *79*, 1225–1261.
25. Cartea, A.; Figueroa, M.G. Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance* **2005**, *12*, 313–335.
26. Janczura, J.; Trück, S.; Weron, R.; Wolff, R.C. Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling. *Energy Economics* **2013**, *38*, 96–110.
27. Janczura, J.; Weron, R. An empirical comparison of alternate regime-switching models for electricity spot prices. *Energy Economics* **2010**, *32*, 1059–1073.
28. Eichler, M.; Tuerk, D. Fitting semiparametric Markov regime-switching models to electricity spot prices. *Energy Economics* **2013**, *36*, 614–624.
29. Keles, D.; Genoese, M.; Möst, D.; Fichtner, W. Comparison of extended mean-reversion and time series models for electricity spot price simulation considering negative prices. *Energy Economics* **2012**, *34*, 1012–1032.
30. Bordignon, S.; Bunn, D.W.; Lisi, F.; Nan, F. Combining day-ahead forecasts for British electricity prices. *Energy Economics* **2013**, *35*, 88–103.
31. Ziel, F.; Weron, R. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics* **2018**, *70*, 396–420.
32. Weron, R.; Misiorek, A. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International journal of forecasting* **2008**, *24*, 744–763.
33. Kristiansen, T. Forecasting Nord Pool day-ahead prices with an autoregressive model. *Energy Policy* **2012**, *49*, 328–332.
34. Raviv, E.; Bouwman, K.E.; van Dijk, D. Forecasting day-ahead electricity prices: Utilizing hourly prices. *Energy Economics* **2015**, *50*, 227–239.
35. Conejo, A.J.; Plazas, M.A.; Espinola, R.; Molina, A.B. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE transactions on power systems* **2005**, *20*, 1035–1042.

36. Mandal, P.; Senjyu, T.; Funabashi, T. Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market. *Energy Conversion and Management* **2006**, *47*, 2128–2142.
37. Catalão, J.P.d.S.; Mariano, S.J.P.S.; Mendes, V.; Ferreira, L. Short-term electricity prices forecasting in a competitive market: A neural network approach. *Electric Power Systems Research* **2007**, *77*, 1297–1304.
38. Zhang, J.; Cheng, C. Day-ahead electricity price forecasting using artificial intelligence. Electric Power Conference, 2008. EPEC 2008. IEEE Canada. IEEE, 2008, pp. 1–5.
39. Panapakidis, I.P.; Dagoumas, A.S. Day-ahead electricity price forecasting via the application of artificial neural network based models. *Applied Energy* **2016**, *172*, 132–151.
40. Amjady, N. Day-ahead price forecasting of electricity markets by a new fuzzy neural network. *IEEE Transactions on power systems* **2006**, *21*, 887–896.
41. Zhao, J.H.; Dong, Z.Y.; Xu, Z.; Wong, K.P. A statistical approach for interval forecasting of the electricity price. *IEEE Transactions on Power Systems* **2008**, *23*, 267–276.
42. Pindoriya, N.; Singh, S.; Singh, S. An Adaptive Wavelet Neural Network-Based Energy Price Forecasting in Electricity Markets. *IEEE Transactions on Power Systems* **2008**, *23*, 1423–1432.
43. Díaz, G.; Planas, E. A note on the normalization of Spanish electricity spot prices. *IEEE Transactions on Power Systems* **2016**, *31*, 2499–2500.
44. Filipovic, D.; Larsson, M.; Ware, T. Polynomial processes for power prices. *arXiv preprint arXiv:1710.10293* **2017**.
45. Fanone, E.; Gamba, A.; Prokopczuk, M. The case of negative day-ahead electricity prices. *Energy Economics* **2013**, *35*, 22–34.
46. Uniejewski, B.; Weron, R.; Ziel, F. Variance stabilizing transformations for electricity spot price forecasting. *IEEE Transactions on Power Systems* **2017**.
47. Avci-Surucu, E.; Aydogan, A.K.; Akgul, D. Bidding structure, market efficiency and persistence in a multi-time tariff setting. *Energy Economics* **2016**, *54*, 77–87.
48. Ozozen, A.; Kayakutlu, G.; Ketterer, M.; Kayalica, O. A combined seasonal ARIMA and ANN model for improved results in electricity spot price forecasting: Case study in Turkey. Management of Engineering and Technology (PICMET), 2016 Portland International Conference on. IEEE, 2016, pp. 2681–2690.
49. Kolmek, M.A.; Navruz, I. Forecasting the day-ahead price in electricity balancing and settlement market of Turkey by using artificial neural networks. *Turkish Journal of Electrical Engineering & Computer Sciences* **2015**, *23*, 841–852.
50. Ozguner, E.; Tor, O.B.; Guven, A.N. Probabilistic day-ahead system marginal price forecasting with ANN for the Turkish electricity market. *Turkish Journal of Electrical Engineering & Computer Sciences* **2017**, *25*, 4923–4935.
51. Ozyildirim, C.; Beyazit, M.F. Forecasting and Modelling of Electricity Prices by Radial Basis Functions: Turkish Electricity Market Experiment. *İktisat İşletme ve Finans* **2014**, *29*.
52. Zheng, H.; Yuan, J.; Chen, L. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies* **2017**, *10*, 1168.
53. Xiaoyun, Q.; Xiaoning, K.; Chao, Z.; Shuai, J.; Xiuda, M.; of Smart Grid, S.K.L.; University, X.J. Short-Term Prediction of Wind Power Based on Deep Long Short-Term Memory. 2016 IEEE PES Asia-Pacific Power and Energy Conference - Xi'an - China. IEEE, 2016.
54. Gensler, A.; Henze, J.; Sick, B.; Raabe, N. Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks. Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on. IEEE, 2016, pp. 002858–002865.
55. Bao, W.; Yue, J.; Rao, Y. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLOS ONE* **2017**, *12*, e0180944.
56. Hosein, S.; Hosein, P. Load forecasting using deep neural networks. Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2017 IEEE. IEEE, 2017, pp. 1–5.
57. Lago, J.; Ridder, F.D.; Vrancx, P.; Schutter, B.D. Forecasting day-ahead electricity prices in Europe: The importance of considering market integration. *Applied Energy* **2018**, *211*, 890–903.
58. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics* **1995**, *13*, 253–263.
59. EPIAS (Epias Transparency Platform). https://seffaflik.epias.com.tr/transparency.

60. EPDK (Republic of Turkey Energy Market Regulatory). http://www.epdk.org.tr/TR/Dokumanlar/Elektrik/YayinlarRaporlar/ElektrikPiyasasiGelisimRaporu.
61. Wasserman, P.D.; Schwartz, T. Neural networks. II. What are they and why is everybody so interested in them now? *IEEE Expert* **1988**, *3*, 10–15.
62. Oksuz, I.; Ruijsink, B.; Anton, E.; Sinclair, M.; Rueckert, D.; Schnabel, J.; King, A. Automatic Left Ventricular Outflow Tract Classification For Accurate Cardiac MR Planning. 2018 IEEE International Symposium on Biomedical Imaging (ISBI).
63. Dorffner, G. Neural Networks for Time Series Processing. *Neural Network World* **1996**, *6*, 447–468.
64. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* **2014**.
65. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural computation* **1997**, *9*, 1735–1780.
66. Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259* **2014**.
67. Hamilton, J.D. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society* **1989**, pp. 357–384.
68. Özkan, H.; Yazgan, M.E. Is forecasting inflation easier under inflation targeting? *Empirical Economics* **2015**, *48*, 609–626.
69. Tsay, R.S. *Analysis of financial time series*; Vol. 543, John Wiley & Sons, 2005.
70. Taylor, J.W. Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research* **2010**, *204*, 139–152.
71. Yin, W.; Kann, K.; Yu, M.; Schütze, H. Comparative study of cnn and rnn for natural language processing. *arXiv preprint arXiv:1702.01923* **2017**.
72. Aggarwal, S.K.; Saini, L.M.; Kumar, A. Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power & Energy Systems* **2009**, *31*, 13–22.
73. Zhang, G.P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **2003**, *50*, 159–175.
74. Chaâbane, N. A hybrid ARFIMA and neural network model for electricity price prediction. *International Journal of Electrical Power & Energy Systems* **2014**, *55*, 187–194.
75. Hong, Y.Y.; Wu, C.P. Day-ahead electricity price forecasting using a hybrid principal component analysis network. *Energies* **2012**, *5*, 4711–4725.
76. Ziel, F. Forecasting electricity spot prices using lasso: On capturing the autoregressive intraday structure. *IEEE Transactions on Power Systems* **2016**, *31*, 4977–4987.
77. Ludwig, N.; Feuerriegel, S.; Neumann, D. Putting Big Data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests. *Journal of Decision Systems* **2015**, *24*, 19–36.