Article

The Investigation of Monthly/Seasonal Data Clustering Impact on Short-Term Electricity Price Forecasting Accuracy: Ontario Province Case Study

Nazila Pourhaji [1], Mohammad Asadpour [1,*], Ali Ahmadian [2,3] and Ali Elkamel [3]

1 Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz 5166616471, Iran; n.pourhaji99@ms.tabrizu.ac.ir
2 Department of Electrical Engineering, University of Bonab, Bonab 5551761167, Iran; ahmadian@uwaterloo.ca
3 Department of Chemical Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada; aelkamel@uwaterloo.ca
* Correspondence: m_asadpour@tabrizu.ac.ir

Abstract: The transformation of the electricity market structure from a monopoly model to a competitive market has caused electricity to be exchanged like a commercial commodity in the electricity market. The electricity price participants should forecast the price in different horizons to make an optimal offer as a buyer or a seller. Therefore, accurate electricity price prediction is very important for market participants. This paper investigates the monthly/seasonal data clustering impact on price forecasting. To this end, after clustering the data, the effective parameters in the electricity price forecasting problem are selected using a grey correlation analysis method and the parameters with a low degree of correlation are removed. At the end, the long short-term memory neural network has been implemented to predict the electricity price for the next day. The proposed method is implemented on Ontario—Canada data and the prediction results are compared in three modes, including non-clustering, seasonal, and monthly clustering. The studies show that the prediction error in the monthly clustering mode has decreased compared to the non-clustering and seasonal clustering modes in two different values of the correlation coefficient, 0.5 and 0.6.

Keywords: clustering; LSTM; deep learning; price forecasting

1. Introduction

In the late 1900s, free trade in the energy market created a complex competitive market in which electricity price (EP) changed moment by moment. The slightest improvement in electricity price forecasting (EPF) accuracy saves millions of dollars in the industry [1]. Moreover, optimal management of the power system, due to the intermittent behavior of the EP, requires short-term forecasting of the EP with high accuracy [2].

Various parameters, such as weather conditions (wind speed, temperature, rainfall, etc.) and consumption patterns (peak hours, days of the week, seasonal characteristics, etc.), have a significant impact on the EPF [3].

The purpose of EPF is to forecast spot and ahead prices in wholesale markets at a point or probable position [4]. EPF methods in papers are generally divided into two main categories: statistical methods and methods based on artificial intelligence [5]. Most models in statistical methods for EPF rely on linear regression, i.e., the EP for a particular day and hour achieved with a linear combination of independent variables called regressors, inputs, or features [4]. The most related data models with statistical methods are linear regression models with high input characteristics that use regularization methods [6]. If there are many regressors, the least absolute shrinkage and selection operator (LASSO) [6] or its generalization of the elastic net [7] is used as an implicit feature selection method to improve the prediction results [4]. LASSO is an advanced statistical method [4] and

is utilized in [8–10] to increase the accuracy of the EPF models. The biggest drawback of the statistical methods is their inability to exactly track the sharp spikes of the EP in the market [11]. The scientists have tried to overcome this drawback by a combination of the statistical methods with other techniques [5].

Today, the artificial intelligence (AI)-based approaches are widely utilized for data mining and forecasting [12]. The deep learning methods are a subset of AI-based approaches. These methods have a strong ability to handle the uncertainties of modern power systems, including big data [13]. Deep learning approaches have different processing layers and can extract the main features of data by multiple abstract levels [14]. Accordingly, deeper networks with more hidden layers assembled with robust training methods are the most widely used approaches in many fields of research. These approaches have been widely used in various predictive topics, such as EP, load demand, wind speed, and so on, in various studies [5]. The authors in [4] have conducted a comprehensive review of different methods for EPF, such as statistical and deep learning methods. They have compared the deep neural network (DNN) and the Lasso estimated autoregressive (LEAR) with the relative mean absolute error (rMAE), the mean absolute error (MAE), the mean absolute percentage error (MAPE), the symmetric mean absolute percentage error (sMAPE), and the root mean square error (RMSE) measures on an open-access benchmark dataset. Comparison results in [4] show that models based on deep learning are the most probable to perform better than models based on statistical methods.

In the field of EPF, hybrid forecasting methods have been very popular in the last five years. Hybrid models have a complex framework. These models include at least two of the five algorithms for data analysis, feature selection, data clustering, one or more prediction models, and several heuristic optimization algorithms for estimating models or their hyperparameters [4]. It is feasible for hybrid methods created from one or more conventional methods and machine learning to work even better than a single method for deep learning [15]. The disadvantage of hybrid methods is that, by combining one or more models, the number of parameters increases [16].

In recent years, many models have been proposed and examined to predict the load and EP. In [17], the efficiency of an ensemble-based method for prediction of short-term electricity spot prices in the Italian electricity market (IPEX) is examined. In this paper, the price time series is partitioned into deterministic and stochastic components. Semi-parametric methods are used to estimate the deterministic component and time series, and various machine learning algorithms are utilized to estimate the stochastic component. Based on the results, the performance of the group-based model is better than random forest (RF) and autoregressive moving average (ARMA). A comprehensive analysis of linear and nonlinear parametric modeling techniques for short-term electrical load prediction has been performed in [18]. This study demonstrated the effectiveness of artificial neural network–Levenberg–Marquardt (ANN–LM) with a hidden layer compared to autoregressive with exogenous inputs (ARX), output error (OE), support vector machine (SVM), autoregressive moving average with exogenous inputs (ARMAX), K-nearest neighbor (KNN), artificial neural network–particle swarm optimization (ANN–PSO), ANN–LM with two hidden layers, and bootstrap aggregation models. In [19], the prediction of several models' performance for one-day forecast of price and demand in four electricity markets is examined and, finally, a double functional model is selected as the best model. A new approach for EPF based on prediction aggregated purchase and sale curves has been introduced in [20]. In this paper, modeling and forecasting of sales curves are performed using functional data analysis methods. Parametric (FAR) and nonparametric (NPFAR) functional autoregression models have compared with some criteria. Finally, based on the application results presented in the paper, NPFAR models perform significantly better. In [21], linear and nonlinear models for one-day-ahead EPFs using component estimation techniques are considered. In this paper, linear and nonlinear models are used for deterministic and stochastic components, and final prediction is obtained by merging the predictions of both these components. When the deterministic component is estimated by the parametric

approach, the best result is obtained. Due to the growing shortage of electricity in Pakistan, in [22], a forecasting method based on the component estimation method is implemented to predict medium-term electricity consumption. Based on the efficiency measurement criteria in the paper, the proposed method has a good performance in predicting power consumption. Different modeling techniques are compared in [23] to predict short-term electricity demand. For this purpose, the time series of electricity demand is partitioned into deterministic and stochastic components, and both components are estimated by regression methods and different time series by parametric and nonparametric estimation techniques. The component-wise estimation method proposed in the paper is very effective for predicting electricity demand. A comprehensive analysis of EPFs has been performed in [24], comparing several outlier filtering techniques. The results show that the outliers' pre-filtering increases the accuracy of EPF. To deal with India's illegal electricity market, an algorithm has been proposed in [25] to predict the next day's electricity load, and the performance of the proposed method has been confirmed at the seasonal level. In [26], a linear model for predicting long-term EP in the Danish market is presented. This model estimates the price on an hourly basis for the next 12 days based on the parameters of wind power plant production, thermal power plant production, electricity consumption, and previous EPs. The performance of ANN and SVM learning methods against multiple linear regression (MLR) and adaptive neuro-fuzzy inference system (ANFIS) is confirmed in [27] to predict electricity load. In [28], a hybrid method called CNN-GRU is proposed to predict the price and load of electricity. In this method, effective features for forecasting are selected and transferred to convolutional neural network (CNN) by coronavirus herd immunity optimization (CHIO) approach and SVM for prediction.

Among the various prediction models in recent years, long short-term memory (LSTM) has been used mainly for EPFs [4]. LSTM is a new type of deep recurrent neural network (DRNN) that has more accuracy and stable performance [5]. LSTM in [29] has been used to predict short-term load demand and photovoltaic (PV) power output in a community microgrid, and the results have showed that the DRNN–LSTM model performs better than the multilayer perception network (MLP) and SVM. In [30], LSTM is used to predict power fluctuations in real time. Based on the simulation results of this paper, the proposed LSTM algorithm achieved the best performance compared to curve fitting, simple back-propagation neural network, radial basis function neural network (RBFNN), and gated recurrent unit (GRU). In [29,30], a small number of parameters and one-year data are used for prediction. Meanwhile, the use of an appropriate number of effective parameters and samples related to several years increases the accuracy of the forecasting. Toubeau et al. [31] have combined the LSTM method with probabilistic methods and, using this combined method, have predicted the parameters of the probability distribution function to generate the scenarios. The results of this paper show that the proposed method obtains accurate and calibrated prediction distributions from the historical dataset, and the generated scenarios are able to increase the economic profitability of electricity market participants. In this paper, only the RMSE criterion is used to evaluate the proper performance of the proposed method, while, to ensure the correct operation of the model, several performance criteria are usually used. A model for identifying time-varying parameters for composite load modeling (CLM) with ZIP load and induction motor in [32] has been developed using a multi-modal LSTM (M-LSTM) deep learning method and has been a successful model. LSTM has been utilized to predict air quality in [33] and the spatial and temporal stability and accuracy of multi-stage regional air quality forecasts have been improved. The wavelet transform technique with the LSTM network and the Adam optimizer in [34] is used for EPF. In this paper, after the wavelet transform, the nonlinear sequence of the EP is decomposed after the wavelet conversion, and the combination of Adam and LSTM can record the appropriate behaviors exactly for the EP. Based on the results of this paper, the proposed model can significantly improve the forecast accuracy. The deep LSTM (DLSTM) has been used to predict the price and load of electricity with time series data in [35], and the superior performance of this network compared to the nonlinear autoregressive

network with exogenous variables (NARX) and extreme learning machine (ELM) has been confirmed. In [36], using the Australian electricity market time series data, LSTM is used for EPF. LSTM performance was better than SVM, regression tree (RT), and NARX. The results in [34–36] show an improvement in LSTM performance in the forecasting. However, time series data are used in prediction and the effect of other parameters in EPF is ignored. Therefore, this issue has reduced the accuracy of the models. Liu et al. [37] predicted the wind speed with the combined experimental wavelet transform (EWT) model, the German neural network, and the LSTM, and the proposed method predicted the wind speed with high accuracy. Although all these papers mentioned show the successful performance of LSTM in forecasting operations, there is a common problem in all of them. In these papers, the historical data are utilized for the network training without clustering. The main purpose of clustering is to create clusters of similar data from a large dataset to accurately represent the behavior of a model. Thus, utilization of non-clustered data for network training increases the error in forecast results and decreases forecast accuracy.

In this paper, the historical data are clustered seasonally and monthly at first and they are utilized for the network training. Due to the different fluctuations of the input data at different hours, the lack of data clustering may show an error in the predicted results, which is shown more at the spike points. By clustering data, data that have similar values over a specific period of time fall into a category. Thus, clustering input data into seasonal and monthly clusters helps to obtain more accurate results in EPFs and makes the network more reliable and efficient. The EP is forecasted for day-ahead horizon using Ontario province, Canada data. The purpose of this study is to optimally manage the power system and help electricity market participants to make a profitable deal. In all the mentioned studies, the positive effect of clustering on increasing the accuracy of LSTM prediction has been ignored. Therefore, in this study, the input data were clustered in two groups: seasonal and monthly. Then, in the feature selection stage, the effective parameters in EPFs are selected by grey correlation analysis (GCA) method. LSTM is used to predict the EP for the next day. Finally, the accuracy of LSTM forecasting in three modes, including without clustering, seasonal clustering, and monthly clustering, based on four criteria, MAE, MAPE, RMSE, and R-squared, is shown.

In the following, the contents of the article are divided as follows. Section 2 provides an explanation of the problem. Section 3 describes the methodology used for implementation. The data used in the implementation, the criteria for calculating the output error, and the results of the model implementation are presented in Section 4. In Section 5, the model implementation results are discussed. The paper is concluded in Section 6.

2. Electricity Price Forecasting

The competitive electricity markets can be classified into single-settlement or two-settlement electricity markets [38]. In a single-settlement market or real-time market, EPs are determined on an hourly, half-hour, or five-minute basis based on available supply and demand. Moreover, in a two- or multi-settlement electricity market, the EP for supply and demand depends on the next day and the real-time operation of the market. To determine EP in a two-settlement market, one-day electricity demand and existing supply are utilized, and the real-time market covers the difference between the suggested and real demand and supply. In these competitive markets, electricity business is accomplished via spot markets, forward markets, or two-sided agreements. Therefore, the future EP prediction is significant to optimize the market contributor operation [39].

Since the beginning of competitive power markets 20 years ago, EPF has gently become an essential process for decisions of energy companies [40]. So, accurate EPFs play a very important role in the non-monopoly electricity market. Using EPF models, electricity market participants can have a clear understanding of future EPs [41]. Electricity market participants need to EPF to make the right decisions about their day-to-day market activities, such as trading, risk management, and future planning [21]. If a buyer in the electricity market has an accurate forecast of the next day EP, they can offer a reasonable

bidding strategy to maximize its yield on trading. Accurate forecasting of EP also helps to identify the customer needs and effectively regulate the power grid by power generators. However, it is difficult to develop highly accurate prediction models due to high-frequency data, fluctuations, nonlinearity, and seasonality [42], which are discussed in this paper.

3. Materials and Methods

The proposed methodology consists of three main parts: clustering, selecting effective parameters, and forecasting. Improving the final results of the proposed method has a lot to do with clustering [5]. Therefore, at first, the input data are clustered in seasonal and monthly groups. EPF is a "big data" issue and various parameters are involved in determining it. The GCA method is used to select the effective parameters in the prediction [1]. This method is applied to select effective features to the input data of each season and month. Finally, due to the proper performance of LSTM in forecasting tasks [5], this network has been used for EPF.

3.1. Clustering

EP and input data for EPF fluctuate in different months and seasons. Therefore, clustering the data in the form of different months and seasons increases the accuracy of the prediction. In this paper, clustering is carried out to increase prediction accuracy. Input data before the feature selection step are clustered in seasonal and monthly groups, and the information of similar seasons and months is classified in separate Excel files. For seasonal data clustering, the EPF input data for the spring seasons of 2016, 2017, and 2018 are stored in an Excel file, and the same is carried out for the other seasons. Moreover, for monthly EPF data clustering, for example, November input data for 2016, 2017, and 2018 are stored in an Excel file and, for other months, clustering operations continue in this way.

3.2. Selecting Effective Parameters Using GCA

Each of the input parameters has a different degree of correlation with EP. To increase the accuracy of EPF, it is necessary to use the data of high correlation parameters. Using the feature selection method and determining the appropriate correlation value help to select effective parameters for EPFs. GCA is a suitable feature selection method and tool for selecting effective parameters in EPF. In this method, the degree of correlation between different input data and the target is determined [43]. After determining the correlation value of EPF input data with GCA, parameters with low correlation coefficient are ignored.

The input data are defined as an input matrix R [1]. The dimensions of the input matrix are m × n:

$$\text{R} = \begin{bmatrix} \beta_1(1) & \cdots & \beta_z(1) & \cdots & \beta_n(1) \\ \vdots & & \vdots & & \vdots \\ \beta_1(ti) & & \beta_z(ti) & & \beta_n(ti) \\ \vdots & & \vdots & & \vdots \\ \beta_1(m) & \cdots & \beta_z(m) & \cdots & \beta_n(m) \end{bmatrix} \quad (1)$$

In the input matrix, the rows represent the time sample and the columns represent the index of input data [43].

First, the normalization operation is performed to change the data values of each column to use a common scale. In data normalization, differences in the range of data values are also maintained. The input data normalization operation can be performed to apply GCA with Formula (2):

$$\beta_z^*(ti) = \frac{\beta_z(ti) - \min\beta_z(ti)}{\max\beta_z(ti) - \min\beta_z(ti)} \quad (2)$$

In Equation (2), each numeric value in the column is subtracted from the minimum column value and divided by the difference between the largest and smallest column values.

In the next step, the Grey coefficient is obtained based on Formulas (3)–(6) [2]:

$$\pi_z(\beta_o^*(ti), \beta_z^*(ti)) = \frac{\Delta_{min} + \varepsilon\Delta_{max}}{\Delta_{oz} + \varepsilon\Delta_{max}} \qquad \varepsilon \in (0,1) \tag{3}$$

$$\Delta_{oz}(ti) = |\beta_o^*(ti) - \beta_z^*(ti)| \tag{4}$$

$$\Delta_{max} = \max|\beta_o^*(ti) - \beta_z^*(ti)| \quad z = 1, \ldots, n \tag{5}$$

$$\Delta_{min} = \min|\beta_o^*(ti) - \beta_z^*(ti)| \quad z = 1, \ldots, n \tag{6}$$

The value of ξ is equal to the value of 0.5 to maintain the useful features as much as possible and to avoid sharp selection in the GCA task [43,44]. To calculate the delta matrix, in the normal data values matrix, the values of the input parameter columns are subtracted from the values of the target data column, and the maximum and minimum values of the delta matrix are calculated based on Equations (5) and (6). Using Equation (3), the matrix of Grey coefficients is obtained.

Finally, using Formula (7), the final correlation values between different input data and the target are calculated using the grey coefficient matrix [1]:

$$\Gamma_z' = (\beta_o^*(ti), \beta_z^*(ti)) = \frac{\sum_{ti=1}^{m} \pi_z(\beta_o^*(ti), \beta_z^*(ti))}{m} \tag{7}$$

3.3. Forecasting Module

The LSTM network is used to predict EP in the proposed method. The LSTM neural network is a special type of R-NN and a strong memory unit. This network has three main gates: input, output, and forgetting. The information input gate remembers new and previous steps. The task of the forget gate is to remove unimportant information from the memory cell, and the output gate can extract useful information from the memory cell [5]. The structure of an LSTM cell is shown in Figure 1.
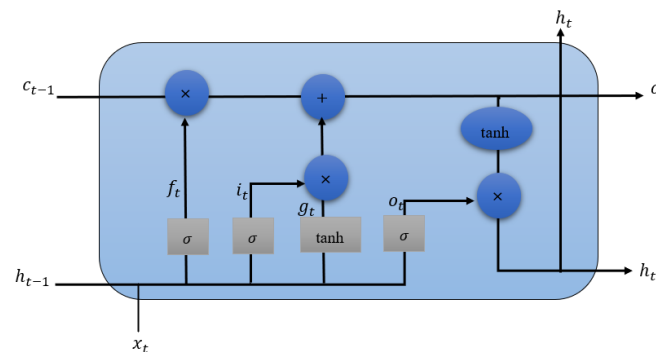

Figure 1. LSTM cell.

The main equations of the LSTM cell are expressed in Formulas (8)–(13). σ is a sigmoid activation function [45]:

$$i_t = \sigma(Wt_i.[h_{t-1}, x_t] + bi_i) \tag{8}$$

$$f_t = \sigma\left(Wt_f.[h_{t-1}, x_t] + bi_f\right) \tag{9}$$

$$c_t = f_t.c_{t-1} + i_t.\bar{c}_t \tag{10}$$

$$\bar{c}_t = tanh(Wt_c.[h_{t-1}, x_t] + bi_c) \tag{11}$$

$$o_t = \sigma(Wt_o.[h_{t-1}, x_t] + bi_o) \tag{12}$$

$$h_t = o_t.\tanh(c_t) \tag{13}$$

4. Numerical Studies

4.1. Data Description

In this study, the EP records of Ontario province, Canada are used as a dataset for simulation and the proposed methodology's evaluation [46]. The dataset parameters include data for each hour of the day (1–24) of the year for the years 2016, 2017, and 2018. The required energy for this province is supplied from different energy resources, such as nuclear, gas, wind, solar energies, etc., and the amount of production of each power plant affects the EP. Moreover, the climatic conditions of the region determine the amount of consumption and EP. For example, in high temperatures, widespread use of air conditioners, refrigerators, etc., increases consumption and EPs. In addition, the EP for the past few hours also plays a role in determining the EP for the next hour. Accordingly, the parameters in the dataset are divided into several categories, such as generation power data (gas, nuclear, wind, hydro, biofuel, and solar that are based on MW), total generation output power (based on MW), predicted EP data (hour 1 pre-dispatch (H1P), hour 2 pre-dispatch (H2P), and hour 3 pre-dispatch (H3P) that are the forecasted price for one, two, and three hours ahead, respectively, with CAD/MWh unit), weather condition data (dew point temperature (°C), real humidity (%), temperature (°C)), and Ontario demand (based on MW) [1]. The data from 1 January 2016 to 31 December 2018 with 1-h time intervals are utilized for this simulation.

4.2. Error Measurement Strategy

In this study, MAE, MAPE, and RMSE [47] have been used to measure the error rate as follows:

$$MAE = \frac{1}{n_0} \sum_{g=1}^{n_0} \left(\left| \hat{Y}_g - Y_g \right| \right) \tag{14}$$

$$MAPE = \frac{1}{n_0} \sum_{g=1}^{n} \left(\left| \frac{\hat{Y}_g - Y_g}{\hat{Y}_{gmean}} \right| \right) \tag{15}$$

$$RMSE = \sqrt{\frac{1}{n_o} \sum_{g=1}^{n_o} \left(\left| \hat{Y}_g - Y_g \right|^2 \right)} \tag{16}$$

Moreover, the R-squared criterion, which expresses the square of the correlation between the target and the predicted values, is utilized. The value of R-squared is between 0 and 1, and a higher value of R-squared indicates a better prediction result [48].

$$R^2 = 1 - \frac{\sum_{g=1}^{n_0} \left(Y_g - \hat{Y}_g \right)^2}{\sum_{g=1}^{n_0} \left(Y_g - \overline{Y}_g \right)^2} \tag{17}$$

4.3. Simulation Results

In this section, GCA results for the three studied modes, including non-clustering, seasonal clustering, and monthly clustering, are explained. Data from 2016, 2017, and 2018, before the specific date defined in the article, have been used for LSTM network training in non-clustering mode. In seasonal clustering, information of similar seasons and in monthly clustering and information of similar months in 2016, 2017, and 2018, before the specific date defined in the article, have been used for network training. To test the network, the specific day and month data defined in the article are used. In total, 90% of the data are used for training and 10% for testing. The Ontario hourly EP data plot for 2016, 2017, and 2018 is shown in Figure 2. Then, the accuracy of LSTM prediction in these three cases (non-clustering, seasonal, and monthly) is presented with two different correlation values. For LSTM, the value of the term L2 regulation is adjusted by a coefficient of 0.001 to prevent sharp and rapid changes in weights during the training procedure. The optimizer used in the training method is stochastic gradient descent with momentum (SGDM) optimizer.

SGDM is one of the most common optimizers, which accelerates gradient vectors in the right directions, resulting in faster convergence. To avoid over-fitting, the dropout layer is implemented with a probability of 0.5. The maximum repetition for LSTM training is 500.
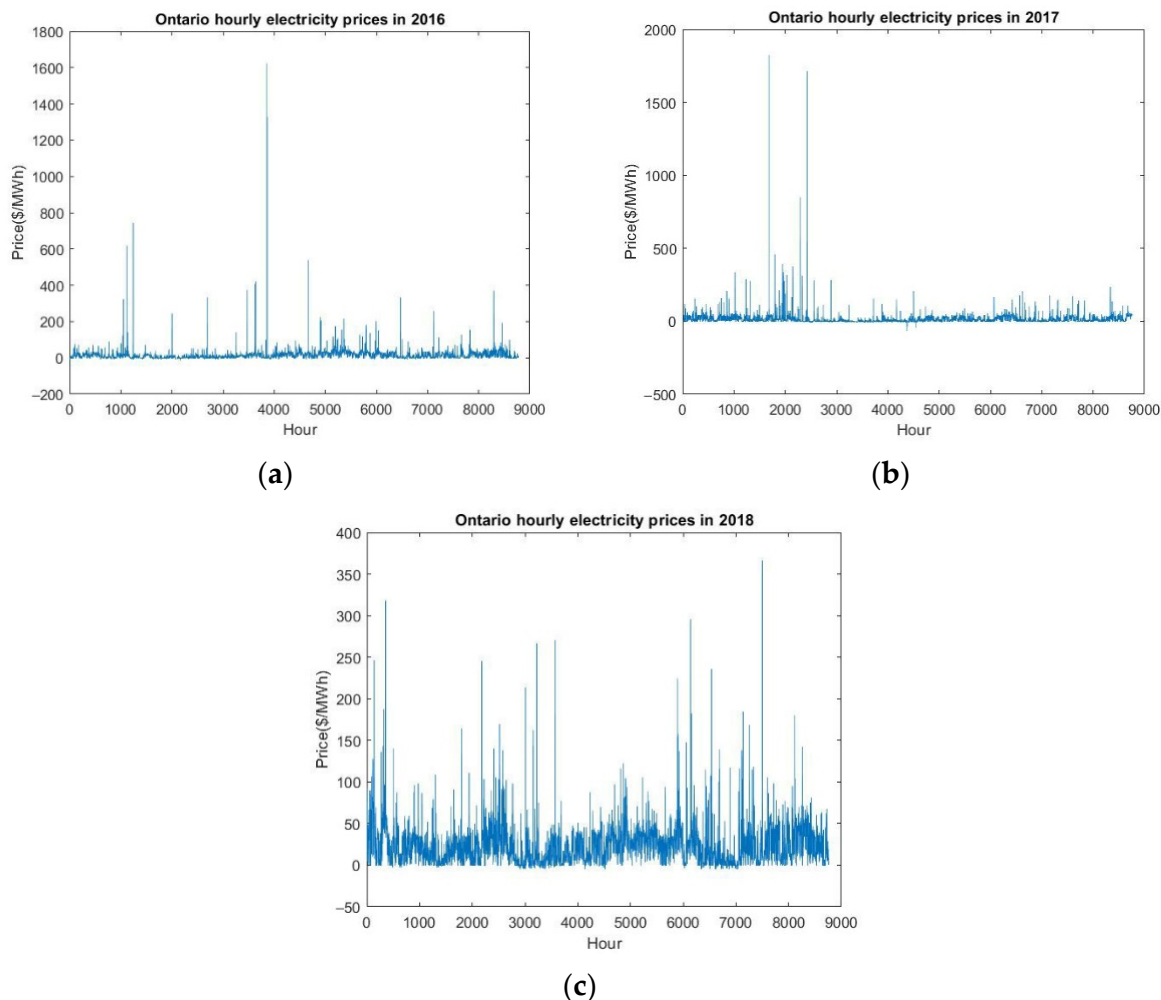

Figure 2. Ontario Hourly EP data plots for (**a**) 2016, (**b**) 2017, and (**c**) 2018.

The numerical study of this article is carried out by MATLAB R2020b software on a PC with an Intel Core i5, 2.71GHz CPU and 8 GB of RAM.

4.3.1. GCA Results

The results of GCA for non-clustering, seasonal (spring), and monthly (July) clustering are shown in Figures 3–5. In these plots, the horizontal axis represents the input parameters and the vertical axis the correlation value of each parameter with the output variable of EP. As can be seen, the results show a considerable difference in the studied cases. The correlation value of wind power with EP, for example, for data without clustering is 0.73, for spring in seasonal clustering is 0.7, and for July in monthly clustering is 0.8. The reason is that the wind speed is significantly different during a year and it may apply an error on the forecasted values. This fact is true for most of the other parameters too, such as solar, hydro, temperature, etc.
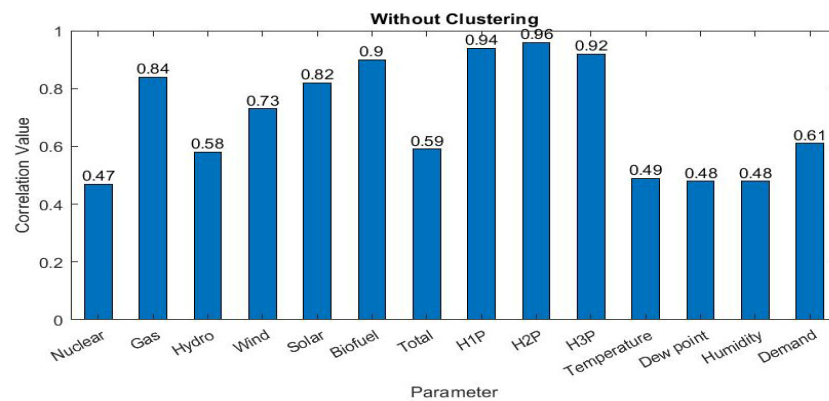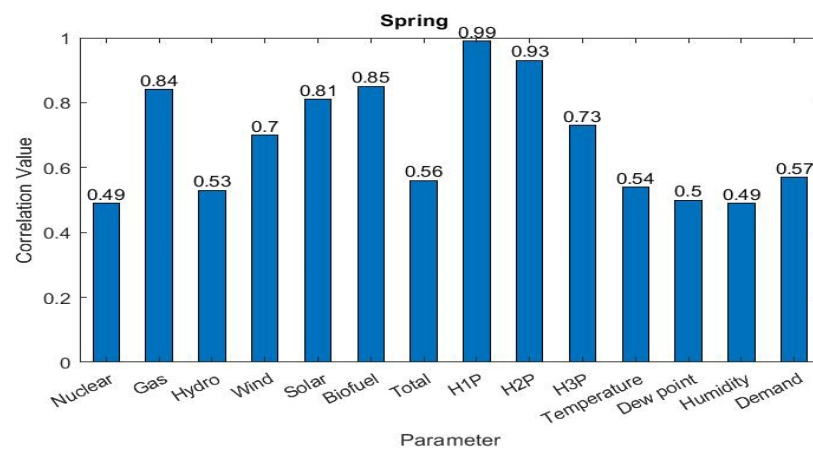
Figure 3. GCA result for non-clustering.

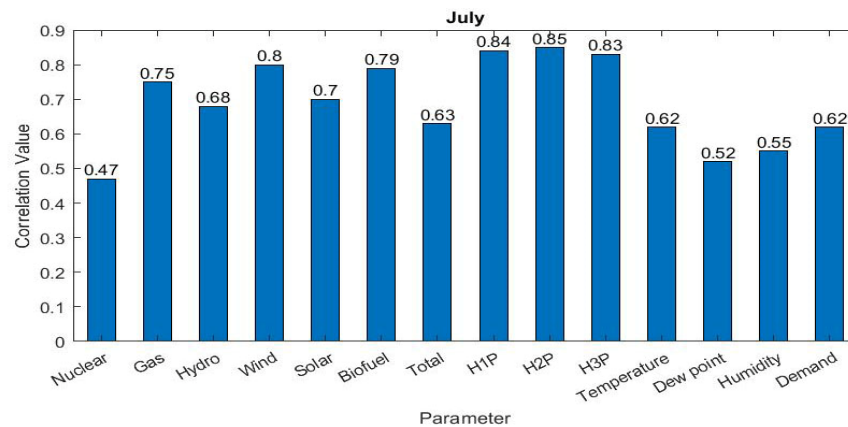**Figure 4.** GCA result for spring.

Figure 5. GCA result for July.

4.3.2. Results for Correlation Value 0.5

In this section, input parameters with a correlation coefficient of less than 0.5 are omitted to obtain the LSTM prediction output. To confirm the performance of the clustering technique, the results of LSTM prediction for the three modes, non-clustering, seasonal clustering, and monthly clustering, are shown in Table 1. For example, on 30 August 2018, MAE, MAPE, and RMSE decreased from 0.718319, 0.965788, and 0.755829 in non-clustering mode to 0.707927, 0.951815, and 0.747863 in seasonal clustering. This reduction in the number of errors for this date in the monthly clustering compared to the season has also occurred and MAE, MAPE, and RMSE values in the monthly clustering have reached 0.590662, 0.794151, and 0.617849. The reduction in monthly clustering error is due to the

reduction in data fluctuations compared to the seasonal clustering and without clustering mode. This error reduction trend at other dates can also be seen in Table 1. This indicates the positive effect of clustering in reducing prediction error. The R-squared values for 30 August 2018 in the three cases are also shown in Figure 6.

Table 1. Results for correlation value 0.5.

	Non-Clustering			Seasonal Clustering			Monthly Clustering		
Date	**MAE**	**MAPE**	**RMSE**	**MAE**	**MAPE**	**RMSE**	**MAE**	**MAPE**	**RMSE**
30 March 2018	0.278435	1.03064	0.401885	0.274898	1.01755	0.399363	0.271402	1.004607	0.37668
March 2018	0.100155	0.919093	0.134502	0.097938	0.898752	0.134302	0.095492	0.876302	0.126315
30 August 2018	0.718319	0.965788	0.755829	0.707927	0.951815	0.747863	0.590662	0.794151	0.617849
30 July 2018	0.334843	0.916085	0.387628	0.322518	0.882367	0.373947	0.212374	0.581027	0.266549
July 2018	0.220278	0.852426	0.254799	0.208578	0.80715	0.24533	0.128541	0.497425	0.163983
27 September 2018	0.177126	0.878413	0.26119	0.123373	0.611835	0.205115	0.11053	0.548144	0.196724
September 2018	0.076773	0.7422	0.119532	0.058819	0.56863	0.095234	0.05384	0.520497	0.09442

(**a**)

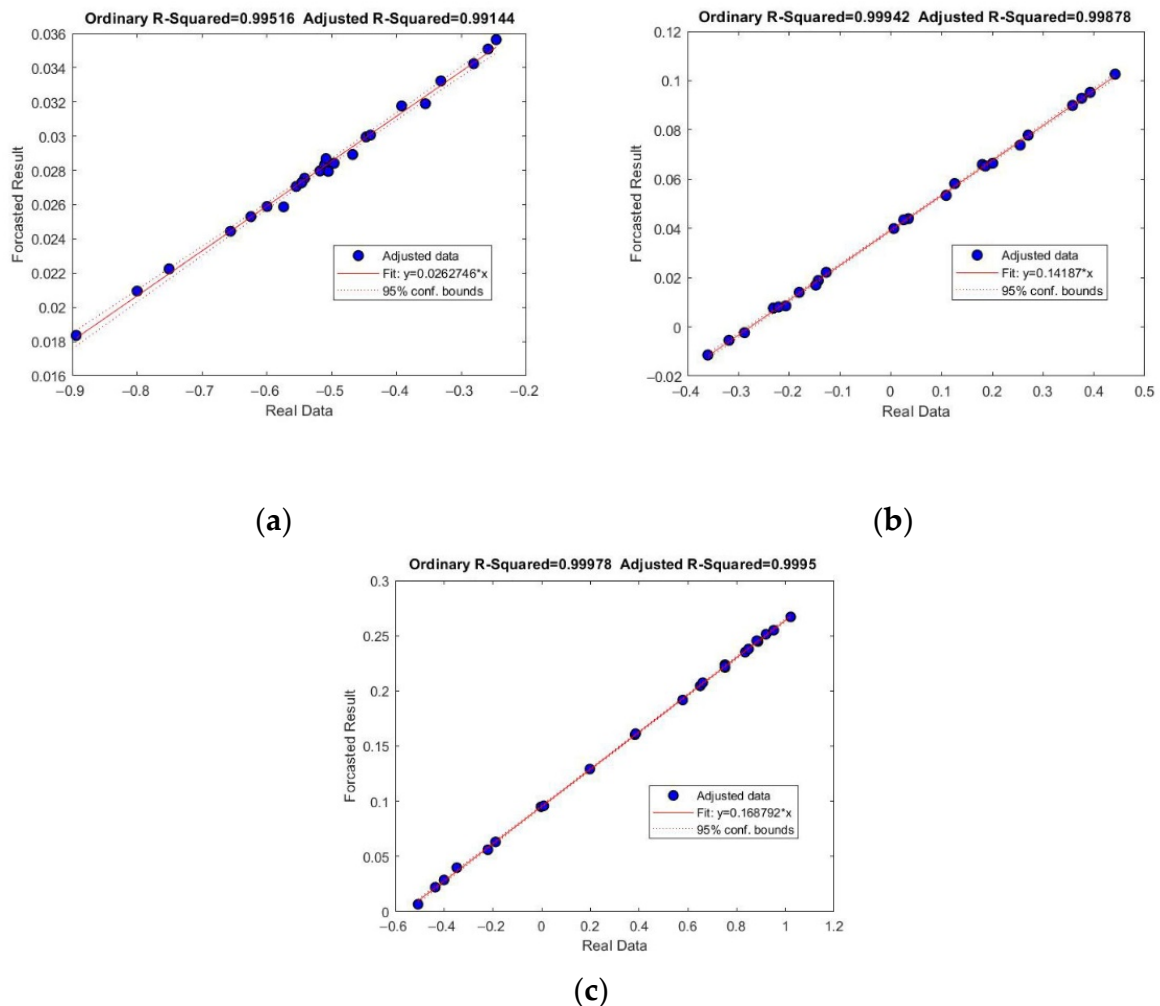(**b**)

(**c**)

Figure 6. R-squared value on 30 August 2018 for a correlation coefficient of 0.5: (**a**) non-clustering, (**b**) seasonal clustering, (**c**) monthly clustering.

Normalization operations are performed for input and output data and the data are in the range of zero and one. The number of hidden layers of LSTM is 15 and the learning rate is defined as 0.09. A total of 10% of training data are allocated for validation.

4.3.3. Results for Correlation Value 0.6

In this section, input parameters with a correlation coefficient of less than 0.6 are omitted to obtain LSTM prediction results. The LSTM prediction results for the three modes, non-clustering, seasonal clustering, and monthly clustering, are presented in Table 2. In this table, similar to Table 1, a reduction in the error rate is observed by limiting the clustering range from non-clustering to seasonal, and then monthly clustering. For example, on 30 August 2018, MAE, MAPE, and RMSE decreased from 0.788079, 1.059581, and 0.834331 in non-clustering mode to 0.783233, 1.053065, and 0.823699 in seasonal clustering, and then to 0.622135, 0.836467, and 0.65346 in monthly clustering. The results presented in this table, similar to Table 1, show the positive effect of clustering in reducing data fluctuations. The R-squared values for these three modes for 30 August 2018 are also shown in Figure 7.

Table 2. Results for correlation value 0.6.

Date	Non-Clustering			Seasonal Clustering			Monthly Clustering		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
30 March 2018	0.335223	1.240845	0.453206	0.288832	1.069127	0.408023	0.278811	1.032032	0.396027
March 2018	0.120099	1.102112	0.15613	0.117517	1.078423	0.151843	0.104502	0.958984	0.135949
30 August 2018	0.788079	1.059581	0.834331	0.783233	1.053065	0.823699	0.622135	0.836467	0.65346
30 July 2018	0.392924	1.074989	0.455631	0.390127	1.067337	0.445071	0.257015	0.703159	0.309999
July 2018	0.26909	1.04132	0.31132	0.240731	0.931574	0.276083	0.152377	0.589666	0.188007
28 February 2018	0.241559	1.3833	0.357837	0.222776	1.275739	0.341735	0.218931	1.253718	0.337652
February 2018	0.169699	0.911453	0.221781	0.166475	0.894133	0.21366	0.164636	0.884258	0.211801

(**a**)
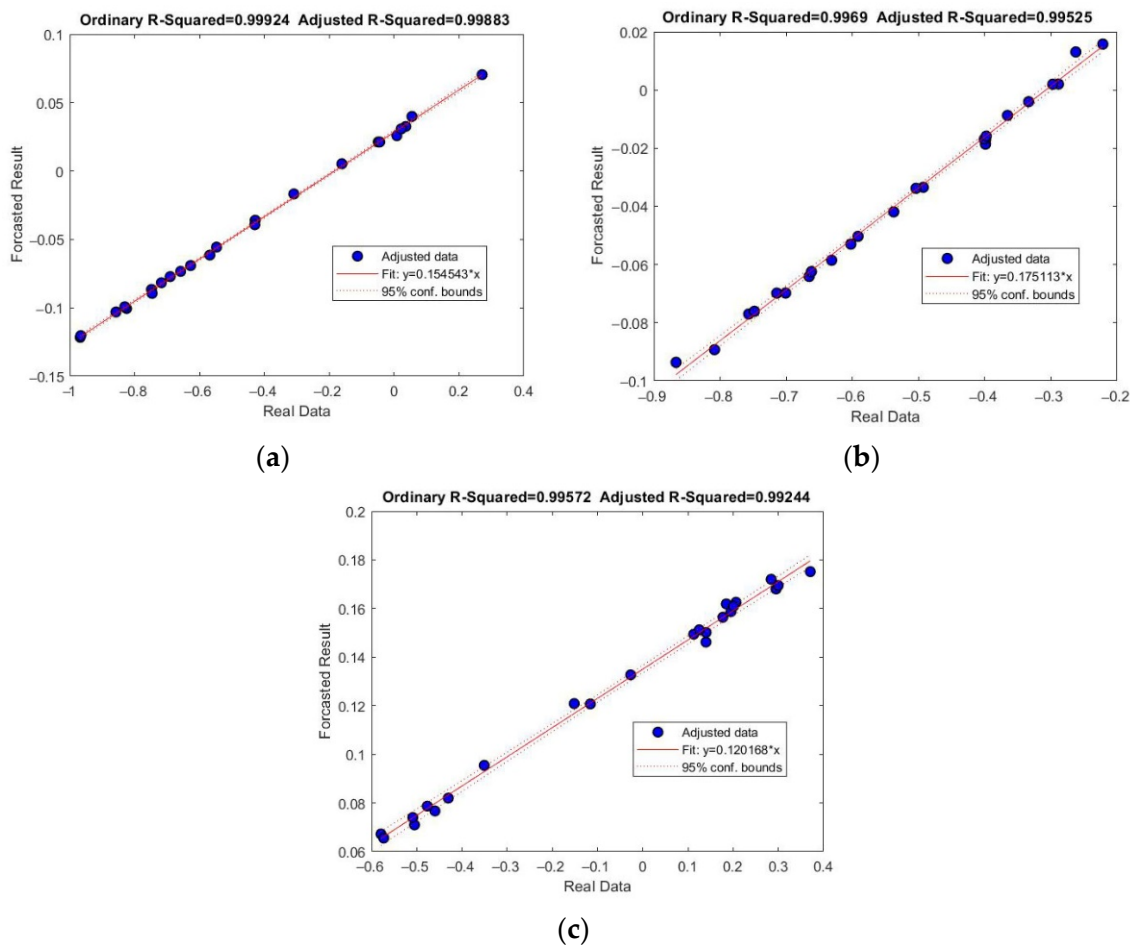

(**b**)

(**c**)

Figure 7. R-squared value on 30 August 2018 for a correlation coefficient of 0.6: (**a**) non-clustering, (**b**) seasonal clustering, (**c**) monthly clustering.

All input and output data are normalized and the data are in the range of zero and one. The number of hidden layers of LSTM is 13 and the learning rate is defined as 0.01. A total of 5% of the training data are intended for validation.

5. Discussion

Accurate EPFs are very difficult due to the different effects of specific parameters in each region on EPs. Using the clustering technique, low-volatility data can be used to predict EPs over a period of time. The smaller the clustering range, the lower the amount of data fluctuations, and this reduces the forecast error. Moreover, using the GCA feature selection method and selecting the appropriate value for the correlation coefficient causes the data that have the most correlation with the EP in a certain time period to be selected for network training. As a result, the accuracy of the forecast increases. In this article, LSTM network prediction accuracy was improved using seasonal and monthly clustering techniques. Based on the forecast results presented in Tables 1 and 2, the accuracy of LSTM prediction in monthly clustering is improved compared to seasonal clustering and non-clustering mode. This improvement in accuracy is also seen in the results of seasonal clustering compared to the non-clustering mode. In addition to the large effect of clustering in reducing the error rate, determining the appropriate value of the correlation coefficient is also effective in reducing the error. By comparing the error values in Table 1 for the correlation coefficient of 0.5 and Table 2 for the correlation coefficient of 0.6, it is observed that, by reducing the coefficient from 0.6 to 0.5, the amount of error in the common dates between the two tables has decreased.

6. Conclusions and Future Work

This paper examines short-term electricity price forecast (EPF). Seasonal and monthly clustering was used to improve long short-term memory (LSTM) prediction accuracy. The next step to increase the accuracy of the forecast was to select the effective parameters in determining the price of electricity by Grey correlation analysis (GCA). To prove the effect of clustering and GCA method on increasing accuracy, LSTM output was presented in three modes, including non-clustering, seasonal, and monthly clustering, with two different correlation values of 0.5 and 0.6. According to the result tables, it can be concluded that the accuracy of predicting monthly clustering is the best in both cases of correlation coefficient 0.5 and 0.6. Moreover, in correlation coefficient 0.5, there is an increase in prediction accuracy for all common dates between these two coefficients compared to correlation coefficient of 0.6.

As a future work, the combination of LSTM and convolutional neural network (CNN) can be used to predict electricity price (EP). Moreover, prediction accuracy can also be improved with conventional clustering algorithms, such as K-means.

Nomenclature

Parameters

n_0	Entire number of input data components	Y_g	Forecasted vector for sample g
m	Entire number of input data in GCA	$\hat{Y}_g$	Actual vector for sample g
n	Entire kind of input data parameters in GCA	$\hat{Y}_{gmean}$	Mean value of actual vector
ξ	Distinguishing factor in GCA.		

Variables

bi_i, bi_c, bi_f, bi_o	Offset vectors	$\beta_z(ti)$	Z-th data sample data at time ti for GCA
c_t	Current cell memory information in LSTM cell	$\beta_z^*(ti)$	Z-th data sample normalized data at time ti for GCA
$\bar{c}_t$	The temporary of the memory cell in the memory block	$\beta_o^*(ti)$	Normalized target data at time ti for GCA.
f_t	The output of the output gate in LSTM cell	$\pi_z(\beta_o^*(ti), \beta_z^*(ti))$	Grey coefficient between sequence $\beta_o^*(ti)$ and $\beta_z^*(ti)$
h_t	Hidden layer current state in LSTM cell	$\Gamma_z(\beta_o^*(ti), \beta_z^*(ti))$	Grey correlation grade between sequence $\beta_o^*(ti)$ and $\beta_z^*(ti)$
i_t	The output of the input gate	Indices	
o_t	The output of forget gate	g	Output layer sample index
Wt_c, Wt_i, Wt_f, Wt_o	Weight matrices connecting the input signal x and the hidden layer output signal y	ti	Time sample index for Input data in GCA
x_t	The current input in LSTM cell	z	Input data type index in GCA

References

1. Jahangir, H.; Tayarani, H.; Baghali, S.; Ahmadian, A.; Elkamel, A.; Golkar, M.A.; Castilla, M. A novel electricity price forecasting approach based on dimension reduction strategy and rough artificial neural networks. *IEEE Trans. Ind. Inform.* **2019**, *16*, 2369–2381. [CrossRef]

2. Xu, F.Y.; Cun, X.; Yan, M.; Yuan, H.; Wang, Y.; Lai, L.L. Power market load forecasting on neural network with beneficial correlated regularization. *IEEE Trans. Ind. Inform.* **2018**, *14*, 5050–5059. [CrossRef]

3. Li, X.; Li, X. Big data and its key technology in the future. *Comput. Sci. Eng.* **2018**, *20*, 75–88. [CrossRef]

4. Lago, J.; Marcjasz, G.; De Schutter, B.; Weron, R. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Appl. Energy* **2021**, *293*, 116983. [CrossRef]

5. Jahangir, H.; Tayarani, H.; Gougheri, S.S.; Golkar, M.A.; Ahmadian, A.; Elkamel, A. Deep Learning-Based Forecasting Approach in Smart Grids With Microclustering and Bidirectional LSTM Network. *IEEE Trans. Ind. Electron.* **2020**, *68*, 8298–8309. [CrossRef]

6. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [CrossRef]

7. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 301–320. [CrossRef]

8. Uniejewski, B.; Weron, R. Regularized quantile regression averaging for probabilistic electricity price forecasting. *Energy Econ.* **2021**, *95*, 105121. [CrossRef]

9. Jędrzejewski, A.; Marcjasz, G.; Weron, R. Importance of the long-term seasonal component in day-ahead electricity price forecasting revisited: Parameter-rich models estimated via the LASSO. *Energies* **2021**, *14*, 3249. [CrossRef]

10. Marcjasz, G.; Uniejewski, B.; Weron, R. Beating the naïve—Combining LASSO with naïve intraday electricity price forecasts. *Energies* **2020**, *13*, 1667. [CrossRef]

11. Gilanifar, M.; Wang, H.; Sriram, L.M.; Ozguven, E.E.; Arghandeh, R. Multitask Bayesian spatiotemporal Gaussian processes for short-term load forecasting. *IEEE Trans. Ind. Electron.* **2019**, *67*, 5132–5143. [CrossRef]

12. Palhares, R.M.; Yuan, Y.; Wang, Q. Artificial intelligence in industrial systems. *IEEE Trans. Ind. Electron.* **2019**, *66*, 9636–9640. [CrossRef]

13. Rusk, N. Deep learning. *Nat. Methods* **2016**, *13*, 35. [CrossRef]

14. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]

15. Hajirahimi, Z.; Khashei, M. Hybrid structures in time series modeling and forecasting: A review. *Eng. Appl. Artif. Intell.* **2019**, *86*, 83–106. [CrossRef]

16. Hamdoun, H.; Sagheer, A.; Youness, H. Energy time series forecasting-analytical and empirical assessment of conventional and machine learning models. *J. Intell. Fuzzy Syst.* **2021**, *40*, 12477–12502, Preprint. [CrossRef]

17. Bibi, N.; Shah, I.; Alsubie, A.; Ali, S.; Lone, S.A. Electricity Spot Prices Forecasting Based on Ensemble Learning. *IEEE Access* **2021**, *9*, 150984–150992. [CrossRef]

18. Javed, U.; Ijaz, K.; Jawad, M.; Ansari, E.A.; Shabbir, N.; Kütt, L.; Husev, O. Exploratory Data Analysis Based Short-Term Electrical Load Forecasting: A Comprehensive Analysis. *Energies* **2021**, *14*, 5510. [CrossRef]

19. Lisi, F.; Shah, I. Forecasting next-day electricity demand and prices based on functional models. *Energy Syst.* **2020**, *11*, 947–979. [CrossRef]

20. Shah, I.; Lisi, F. Forecasting of electricity price through a functional prediction of sale and purchase curves. *J. Forecast.* **2020**, *39*, 242–259. [CrossRef]

21. Shah, I.; Bibi, H.; Ali, S.; Wang, L.; Yue, Z. Forecasting one-day-ahead electricity prices for italian electricity market using parametric and nonparametric approaches. *IEEE Access* **2020**, *8*, 123104–123113. [CrossRef]

22. Shah, I.; Iftikhar, H.; Ali, S. Modeling and forecasting medium-term electricity consumption using component estimation technique. *Forecasting* **2020**, *2*, 163–179. [CrossRef]

23. Shah, I.; Iftikhar, H.; Ali, S.; Wang, D. Short-term electricity demand forecasting using components estimation technique. *Energies* **2019**, *12*, 2532. [CrossRef]

24. Shah, I.; Akbar, S.; Saba, T.; Ali, S.; Rehman, A. Short-term forecasting for the electricity spot prices with extreme values treatment. *IEEE Access* **2021**, *9*, 105451–105462. [CrossRef]

25. Uppal, M.; Garg, V.K.; Kumar, D. Weather biased optimal delta model for short-term load forecast. *IET Smart Grid* **2020**, *3*, 835–842. [CrossRef]

26. Schütz Roungkvist, J.; Enevoldsen, P.; Xydis, G. High-resolution electricity spot price forecast for the Danish power market. *Sustainability* **2020**, *12*, 4267. [CrossRef]

27. Solyali, D. A comparative analysis of machine learning approaches for short-/long-term electricity load forecasting in Cyprus. *Sustainability* **2020**, *12*, 3612. [CrossRef]

28. Aslam, S.; Ayub, N.; Farooq, U.; Alvi, M.J.; Albogamy, F.R.; Rukh, G.; Haider, S.I.; Azar, A.T.; Bukhsh, R. Towards Electric Price and Load Forecasting Using CNN-Based Ensembler in Smart Grid. *Sustainability* **2021**, *13*, 12653. [CrossRef]

29. Wen, L.; Zhou, K.; Yang, S.; Lu, X. Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting. *Energy* **2019**, *171*, 1053–1065. [CrossRef]

30. Wen, S.; Wang, Y.; Tang, Y.; Xu, Y.; Li, P.; Zhao, T. Real-time identification of power fluctuations based on lstm recurrent neural network: A case study on singapore power system. *IEEE Trans. Ind. Inform.* **2019**, *15*, 5266–5275. [CrossRef]

31. Toubeau, J.F.; Bottieau, J.; Vallée, F.; De Grève, Z. Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets. *IEEE Trans. Power Syst.* **2018**, *34*, 1203–1215. [CrossRef]

32. Cui, M.; Khodayar, M.; Chen, C.; Wang, X.; Zhang, Y.; Khodayar, M.E. Deep learning-based time-varying parameter identification for system-wide load modeling. *IEEE Trans. Smart Grid* **2019**, *10*, 6102–6114. [CrossRef]

33. Zhou, Y.; Chang, F.J.; Chang, L.C.; Kao, I.F.; Wang, Y.S. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *J. Clean. Prod.* **2019**, *209*, 134–145. [CrossRef]

34. Chang, Z.; Zhang, Y.; Chen, W. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. *Energy* **2019**, *187*, 115804. [CrossRef]

35. Mujeeb, S.; Javaid, N.; Ilahi, M.; Wadud, Z.; Ishmanov, F.; Afzal, M.K. Deep long short-term memory: A new price and load forecasting scheme for big data in smart cities. *Sustainability* **2019**, *11*, 987. [CrossRef]

36. Fatema, I.; Kong, X.; Fang, G. Electricity demand and price forecasting model for sustainable smart grid using comprehensive long short term memory. *Int. J. Sustain. Eng.* **2021**, *14*, 1714–1732. [CrossRef]

37. Liu, H.; Mi, X.-W.; Li, Y.-F. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. *Energy Convers. Manag.* **2018**, *156*, 498–514. [CrossRef]

38. Veit, D.J.; Weidlich, A.; Yao, J.; Oren, S.S. Simulating the dynamics in two-settlement electricity markets via an agent-based approach. *Int. J. Manag. Sci. Eng. Manag.* **2006**, *1*, 83–97. [CrossRef]

39. Sandhu, H.S.; Fang, L.; Guan, L. Forecasting day-ahead price spikes for the Ontario electricity market. *Electr. Power Syst. Res.* **2016**, *141*, 450–459. [CrossRef]

40. Nowotarski, J.; Weron, R. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1548–1568. [CrossRef]

41. Uniejewski, B.; Marcjasz, G.; Weron, R. Understanding intraday electricity markets: Variable selection and very short-term price forecasting using LASSO. *Int. J. Forecast.* **2019**, *35*, 1533–1547. [CrossRef]

42. Yang, Z.; Ce, L.; Lian, L. Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. *Appl. Energy* **2017**, *190*, 291–305. [CrossRef]

43. Wang, K.; Xu, C.; Zhang, Y.; Guo, S.; Zomaya, A.Y. Robust big data analytics for electricity price forecasting in the smart grid. *IEEE Trans. Big Data* **2017**, *5*, 34–45. [CrossRef]

44. Liu, S.; Lin, Y. Introduction to grey systems theory. In *Grey Systems*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 1–18.

45. Zhou, S.; Zhou, L.; Mao, M.; Tai, H.M.; Wan, Y. An optimized heterogeneous structure LSTM network for electricity price forecasting. *IEEE Access* **2019**, *7*, 108161–108173. [CrossRef]

46. The Independent Electricity System Operator (IESO). 2018. Available online: http://ieso.ca/ (accessed on 5 December 2018).

47. Wang, L.; Zhang, Z.; Chen, J. Short-term electricity price forecasting with stacked denoising autoencoders. *IEEE Trans. Power Syst.* **2016**, *32*, 2673–2681. [CrossRef]

48. Hashemi, S.M.; Sanaye-Pasand, M. A new predictive approach to wide-area out-of-step protection. *IEEE Trans. Ind. Inform.* **2018**, *15*, 1890–1898. [CrossRef]