Electric Power Systems Research 182 (2020) 106191

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

A proposed intelligent short-term load forecasting hybrid models of ANN, WNN and KF based on clustering techniques for smart grid

Hamed H.H. Aly

Mathematics and Statistics Department, Acadia University, Electrical and Computer Engineering Department, Dalhousie University and Electrical Power and Machines Engineering, Zagazig University, Egypt

ARTICLE INFO

Keywords:
Short-term load forecasting
Smart grid
ANN
KF
WNN
Clustering techniques

ABSTRACT

Smart grid is one of the most important topics to be covered with the increasing penetration of renewable energy in the power system grid to improve grid energy efficiency by managing the relationship between the demand and the generation. Load forecasting is playing a crucial role in this process as well as the output power generation from different renewable energy resources. The accuracy of the forecasting models is very important to deal with the new energy generation and consumption. Conventional approaches used in the literature work done for load forecasting can not handle the requirements of new generation of renewable energy and their uncertainties. This paper is proposing a novel technique for short-term load forecasting based on hybrid of different models and using clustering techniques to improve the overall system performance and quality. These models involve different combinations of Kalman filtering (KF), Wavelet and Artificial Neural Network (WNN and ANN) schemes. Six different models are proposed based on the clustering techniques. Simulations proved higher performance of the proposed models. The data used is commercial data, so it is scaled in this paper. The proposed work is validated by using different dataset for two different locations in Egypt and Canada.

1. Introduction

The accuracy of the forecasting models is very important to deal with the new energy generation and consumption. Many works done in the literature based on load forecasting. Time series forecasting technique, Neural networks and a Kalman filtering estimator are the commonly used techniques for load forecasting especially for smart grid applications [1–3]. For the meter-level load forecasting a deep learning-based technique with appliance behavior learning based on Long short-term memory (LSTM) is used to improve the system performance. LSTM is used to keep a memory cell to remember a significant paste state and maintain a gate to train and set the memory cell for some features [1]. A hybrid model of artificial neural networks and fuzzy expert systems is proposed for short-term load forecasting system [4]. Genetic-based multi-layered perceptions of ANN is used in Taiwan for power system short-term load forecasting [5]. Another model used time-series modeling for load forecasting based on peak load estimation capability [6].

The accuracy of load forecasting is very important for load management of energy demand. The distribution operators could reduce the demand during the peak time by attracting the customers to use the energy during a specific time by decreasing the tariff charge during these times. This could be done depending on an accurate load forecasting. Smart grid is playing a very important rule to manage this operation [7,8]. Time series analysis ARIMA models are used for short term load forecasting [9]. Different neural network models are used in literature for short term load forecasting [10]. Semi-parametric additive model is proposed for short term load forecasting [11]. Nonparametric function methods are used for forecasting of next day demand and price of electricity [12]. Short term household forecasting and load curve are formulated depending on clustering step of historical data that helps in describing the common shapes characteristics. Also, nonparametric curve discrimination scheme is used to cluster the segments [13]. Autoregressive integrated moving average (ARIMA) and artificial neural networks is used for natural gas consumption forecasting [14]. Genetic algorithm optimization and back propagation neural network is used for short-term load forecasting of natural gas [15]. System clustering techniques are used for day ahead and hour ahead load forecasting [16]. The demand of energy profile based on the day, weather condition and the location and all these factors are variables. A model based on a methodology that classifying the characteristics of crucial factors is proposed using Mean Absolute Error (MAPE) as a method of comparison [17]. Smart meters are used these days for many loads. Some companies prefer smart meters as they are accurate and sensitive to loads. Smart meters readings are used for short-term load forecasting for residential customers [18]. Neural network based on the construction of prediction intervals method is used for short-term load and wind

E-mail addresses: hamed.aly@dal.ca, hamed.aly@acadiau.ca, hhhendawy@zg.edu.eg.

https://doi.org/10.1016/j.epsr.2019.106191
Received 3 July 2019; Received in revised form 17 December 2019; Accepted 29 December 2019
Available online 06 February 2020

power forecasting. This method is called lower upper bound estimation. Lower upper bound estimation method converters the primary multi-objective problem into a constrained single one and particle swarm optimization approach is used for solving that problem. It has lesser number of parameters [19]. Radial basis neural network training algorithm is used for short-term electrical load forecasting [20]. Particle swarm optimisation based on least square support vector machine is proposed for load forecasting to improve the accuracy of the forecasting model [21]. Wavelet neural networks based on data prefiltering is used for very short-term load forecasting [22]. A combination of Wavelet decomposition and neural network along with augmented dickey fuller test approach is used for short-term load forecasting [23]. A hybrid of mutual information algorithm and cascaded neuroevolutionary technique is used for day-ahead price forecasting of electricity [24]. Two approaches are used for short term load forecasting; feature selection mutual information approach and enhanced differential evolution approach by making some modifications in the training process of the artificial neural network [25]. Fuzzy logic and wavelet transform integrated generalized neural network is used for short term load forecasting [26]. Different neural networks models are used for short term load forecasting to improve the system performance and its accuracy [27]. A new neuron model is used for rotating electrical machines and load forecasting problems short term forecasting [28].

Most of the previous techniques are using the data without making some statistical analysis like clustering to improve the proposed techniques performance. Due to uncertainties and fluctuation of the data; the models are not accurate especially if we are dealing with smart grid and renewable energy integration. This motivates introducing some statistical data analysis like clustering techniques to formulate different sets of data (in this work we used 6 clustering segments) to improve the overall system accuracy and executable time needed for the training. Hybrid models of WNN, ANN and KF are used in this work to add many desirable features to the proposed models. These models are designed such that the forecasting will be used for a week ahead or a day ahead. MAPE is used in this paper as a comparison method for different models. MAPE is defined as the absolute average difference between the actual and predicted values as shown in Eq. (1) [17,29–31].

$$MAPE = \frac{1}{N}\sum_{1}^{N}\left|\frac{x_i - y_i}{y_i}\right| \tag{1}$$

where, x_i is the predicted value, y_i is the actual value, and N is the number of observations.

$$r = \frac{n(\sum x_i y_i) - (\sum y_i)(\sum x_i)}{\sqrt{(n\sum x_i^2 - (\sum x_i)^2)(n\sum y_i^2 - (\sum y_i)^2)}} \tag{2}$$

where: r is describing the strength of the relation between the measured y_i and predicted x_i data and is defined in Eq. (2) as [28–31].

The most commonly used factor is normalized root mean square error (nRMSE) which is defined in Eq. (3) and is used also in this paper [30,31].

$$nRMSE = \frac{\sqrt{\frac{1}{N}\sum_{1}^{N}(x_i - y_i)^2}}{\bar{y}} \tag{3}$$

where $\bar{y}$ is mean of the actual data.

In this paper we used a clustering technique to discriminate between different shapes of data and organize them in different regions based on their common characteristics. Then we used six different hybrid models of WNN and ANN, ANN and WNN, ANN and KF, KF and ANN, WNN and KF and finally KF and WNN to choose the best hybrid model with the best performance. For each model; the clustered data is fed to the first technique and the error (which is the difference between the actual and the predicted data) is fed to the second model. Then the output is the summation of the output from the first technique and the output from the second technique. The paper is organized as follows. In Sections 2 and 3 the concept of clustering, ANN, WNN and Kalman filtering methodologies are introduced. In Section 4 ANN, WNN and KF are introduced. In Section 5 we defined six different hybrid proposed models as well as the results of the proposed hybrid models along with the simulations for two different dataset locations to validate the proposed work. Section 6 is showing the model parameters selections and the K-fold cross validation results to validate the proposed work.

2. Clustering techniques for the short-term load forecasting models

The load consumption features are varying from location to another and from time to time, so it is a good idea to make some analyses to examine the behaviours of the load and build common zones of the loads. This motivates researchers to use clustering techniques. This could be done using various statistical analysis with the help of different software. In this paper we used a clustering technique to create different zones with common features with the help of R-studio and MATLAB software [32,33]. After creating the clustering zones, we feed the data to ANN, WNN and/or KF for forecasting. This technique is reducing the error profile for the forecasted data. The number and size of clusters segments are playing a key role for improving the characteristics of the forecasting models. As the size is increased, the performance is improved but at the same time the number of clusters segments will be decreased, and the overall improvement will be affected. So, we are playing with two corelated factors; the size and the number of clusters (increasing the size will improve the performance but at the same time will minimize the clusters numbers which will have a negative impact on the performance and vise versa). So, it is a double edge weapon; that is why we have to be careful while we are dealing with it. In this paper we tried to use different combinations of clusters numbers and sizes and checked the error for the forecasted data compared to the actual one. In this step we tried to combine the hours or days that contain the same characteristics by looking into the historical dataset that we have and using clustering techniques with the help of R-studio software. After performing some statistical analyses we concluded that the best choice of the data is to cluster it into six different common regions (three weekdays and three weekends regions).These three regions are morning (7 am to 4 pm), the mid of the day (4 pm to 12 pm) and overnight (12 pm to 7 pm). After many trials we found that the improvement is not recognized if the clusters are above eight and below two, but the best choice is happened when the clusters numbers are six. In this paper we used different forecasting techniques to predict the short-term load consumption after clustering the data. Many ways of clustering techniques approaches are used in the past. In this paper we used K-means technique for clustering load segments. In the K-mean clusters approach we used *z* as data with *t* instances, and $S_1, S_2, \ldots, S_k$ are the *K* disjoint clusters of *z*. The error function is defined by Eq. (4).

$$\text{Error} = \sum_{i=1}^{k}\sum_{x\in Si} D(y,\ \mu(S_i)) \tag{4}$$

where $\mu(S_i)$ is the centroid of cluster S_i. $D(y,\mu(S_i))$ is the distance between *y* data point and $\mu(S_i)$, the Euclidean distance for any data points $y_i = (y_{i1}, y_{i2}, \ldots, y_{in})$ and $y_j = (y_{j1}, y_{j2}, \ldots, y_{jn})$ is defined by Eq. (5) [29].

$$D_{euc\,(yi,\,yj)} = (\sum_{k=1}^{n}(y_{ik} - y_{ik})^{2})^{1/2} \tag{5}$$

3. Proposed networks construction of the short-term load forecasting models

We used hybrid models of WNN, ANN and KF and we aggregated the clustered data into aggregated forecasting hybrid model. Fig. 1 shows the proposed clustering forecasting aggregated models. The data

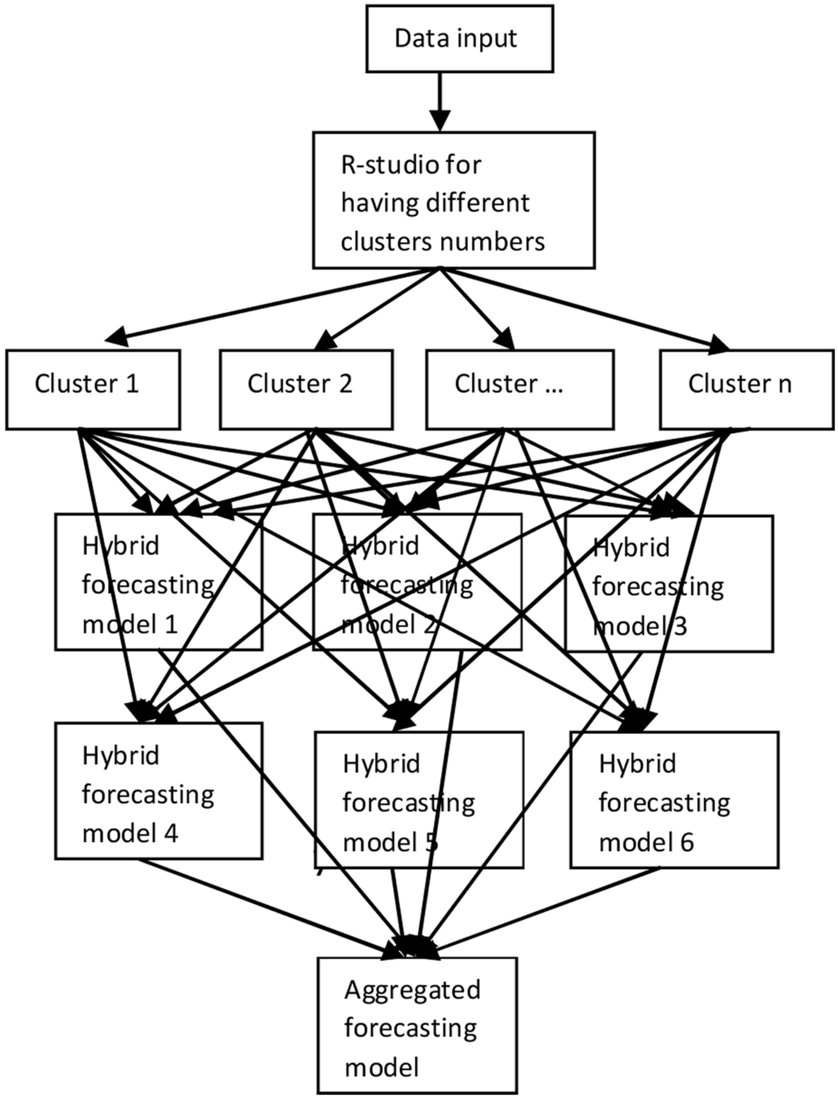

Fig. 1. Input output clustered forecasting aggregated model.

is fed first to clustering algorithm and the clustered data is fed to different forecasting models to choose the best forecasting model and then data is aggregated to have one model.

Algorithm

1: Load data file ⟵ day or week ahead load data.

2: Use R-studio for creating clustering regions.

3: Feed the clustered data to ANN then compare the forecasted data by the actual one to calculate the error data.

4: Feed the error data to WNN and find the error.

5: The forecasted load data is equal to the summation of output from ANN and WNN.

6: Compare between the forecasted load data and the actual load data and calculate MAPE, r, and nRMSE.

7: Repeat from 3 to 6 by using different approaches (WNN + ANN, ANN + Kalman filter, Kalman filter + ANN, WNN + Kalman filter, and Kalman filter + WNN).

8: Compare between different models and choose the best one that has minimum MAPE, and nRMSE.

4. ANN, WNN and Kalman filtering approaches

4.1. Artificial neural networks

Back propagation ANN is a supervised training approach and is used for improving the error derivative of the weights and biases. It takes partial derivatives for weight and force and back from hidden layers to output layers. The training is done in two different modes, the online mode and the batch mode. The number of weight updates are not the same for different modes. For online approach, the weight updates are computed for each input data sample and modified after each input. The batch training approach computes and stores the weight for each input sample and add all contributions at the end of each epoch which is depending on cumulative updates. The Radial Basis Function (RBF) network is practical application model. Linear transfer functions are used for layers and nonlinear transfer functions (normally Gaussian) are used for hidden layers. RBF requires more neurons compared to BP but it needs less time for the training. The weights of the hidden layer are depending on clustering techniques [31,34,35].

4.1.1. Back propagation

BP is a three-layer feed-forward neural network, which includes an input layer, a hidden layer/s and an output layer with linear neurons. The backpropagation algorithm consists of two phases; the forward phase where the activations are propagated from the input to the output layer, and the backward phase, where the error between the observed actual and the requested nominal value in the output layer is propagated backwards in order to modify the weights and bias values. BP algorithm needs many repetitions to converge; however, there is always

LOL MALES

the possibility to get stuck in a local minimum, often due to false weight dimension choice [35–39].

4.1.2. Radial basis function

RBF utilizes Cover's theorem which states that "A non-linearly separable problem is highly separable in high dimensional space than it is in low dimensional space." Nodes in hidden layer perform the radial basis transformation which is basically increasing the dimensions of input feature vector and by doing that defining optimal receptors and spread of the radial basis function. Input units distribute the values to the hidden layer units uniformly, without multiplying them with weights. Hidden units are known as RBF units because their transfer function is a monotonous radial basis function. The transformation performed are local and as a result or that their training is much faster. Radial basis networks may require more neurons than standard feed forward networks, but often they can be designed to take a fraction of time it takes to train standard feed forward networks [35–39]. In this paper we tried using both BB and RBF. RBF proved its strength in dealing with the clustered data. WNN and KF are used also to create hybrid of different approaches of these features to choose the best one for the two different set of data. Using WNN and then ANN improved the performance of the overall model especially if we use clustering techniques before training the system. In this work we used different number of neurons, different numbers of hidden layers as well as different number of clusters and we trained the model to choose the best performance based on genetic algorithm optimization. We found that the best number of clusters is six, the best numbers of neurons is seventy-five and the best number of hidden layers is two for better performance.

4.2. Wavelet neural networks

WNN has the same ability as of ANN with different activation functions in the hidden layer. Due to the localized wavelet activation functions WNN has more compact techniques and learning speed. The output is represented by sum of weighted wavelets in a similar way like ANN. w_{jk} is the weight between the hidden unit j and input unit k. w_{ij} is the weight between the output and hidden and hidden unit j. The sum of weighted inputs to the j^{th} hidden neuron, $x_k(n)$ is defined as the k^{th} input and is represented by $f_j(n)$ [39–41].

$$f_j(n) = \sum_{k=0}^{k=m} w_{jk}(n)^* x_k(n) \tag{6}$$

The output of each hidden neuron is defined by

$$\psi_{a,b}(f_j(n)) = \psi[(f_j(n) - b_j(n))/a_j(n)] \tag{7}$$

where ψ is the wavelet function, $a_j(n)$ is defined as the scaling, and $b_j(n)$ is the translation coefficients of the wavelet function in hidden neuron. The input $f(n)$ and output $y(n)$ of the output neuron is described by the following equations

$$f(n) = \sum_{k=0}^{k=m} w_{ij}(n)^* \psi_{a,b}(f_j(n)) \tag{8}$$

$$y(n) = \sigma[f(n)] \tag{9}$$

4.3. Kalman filtering

Kalman filtering is known as linear quadratic estimation. It is an algorithm that uses set of data containing some statistical noise and other inaccuracies and estimates some variables to represent the trend of these dataset by using some probability distribution over the variables.

Kalman filters are one of the most commonly used methods for state estimation using linear dynamical systems. The model is defined as

$$x_k = F\,x_{k-1} + B\,u_{k-1} + w_{k-1} \tag{10}$$

where F is the state transition matrix applied to the previous state vector x_{k-1}, B is the control-input matrix applied to the control vector u_{k-1}, and w_{k-1} is the process noise vector that is assumed to be zero-mean Gaussian. Kalman filter is used to estimate x_k. Kalman filter has two stages; the prediction and the updating. The Prediction is based on the state estimate $\hat{x}k- = F\hat{x}_{k-1} + B\mathbf{u}_{k-1}$ and the updating is using the residual $\tilde{y}k = \boldsymbol{Z_k} - H\hat{x}k$ where is z is the measurement vector and H is the measurement matrix [42,43].

5. Short-term load forecasting models

5.1. Hybrid model of ANN and WNN

In this model we used ANN feed forward back propagation and radial basis function to train the un-clustered (row) and clustered data. The input to that model is the time in hours and the output is the load power in MW. The forecasted data for the un-clustered model is compared to the actual data and we calculated the residuals (error) which is the forecasted minus the actual data. We did the same with the clustered model, but the data is aggregated before calculating the residuals. The residuals are then fed to WNN for both cases (clustered and un-clustered model) as seen in Fig. 2. The data used in this work is actual commercial data from Nova Scotia, Canada so it is scaled for confidentiality. The model is validated by using different set of data from different city with different weather conditions for different country (Cairo, Egypt) after scaling it.

The MAPE and nRMS are calculated for that model for both locations and it is shown in Table 1 for Nova Scotia and in Table 2 for Cairo. As we see from Table 1 the MAPE for un-clustered data is 2.247 and for clustered data is 2.115 and nRMS is 0.0280 for un-clustered data and is 0.0261 for clustered data. As we see the clustering technique is improving the performance of the forecasting model for both locations. Fig. 3 shows the relationship between the actual and forecasted data with and without clustering for short term load forecasting and the hours of weekdays for a certain period of time in Nova Scotia. Fig. 4 shows the same relationship but for another location as a way of validation of the proposed model. From Figs. 3 and 4 we conclude that the clustered hybrid forecasting model is improving the forecasted data compared to the un-clustered.

5.2. Hybrid model of WNN and ANN

In this model we used WNN to train the un-clustered (row) and clustered data. The input to that model is the time in hours and the output is the load power in MW. The forecasted data for the un-clustered model is compared to the actual data and we calculate the error which is the forecasted minus the actual. We did the same with the clustered data, but the data is aggregated before calculating the error. The error is then fed to ANN for both cases (clustered and un-clustered data). The MAPE and nRMS are calculated for that model for both cities and it is shown in Table 1 for Nova Scotia and in Table 2 for Cairo. As we see from Table 1 the MAPE for un-clustered data is 2.1771 and for clustered data is 1.98. Also, we calculated nRMS for both cases and it is 0.0232 and for un-clustered data and 0.01920 for clustered data. As we see the clustering technique is improving the performance of the forecasting proposed model for both locations. Fig. 5 shows the relationship between the actual and forecasted data with and without clustering and the hours of weekdays for a certain period of time in Nova Scotia. Fig. 6 shows the same relationship but for another location as a way of validation for the proposed model. From Figs. 5 and 6 we conclude that the clustered hybrid forecasting model is improving the forecasted data compared to the un-clustered. We could say that this model is ranked as the best model as it has the smallest MAPE and nRMS as we see in Tables 1 and 2.

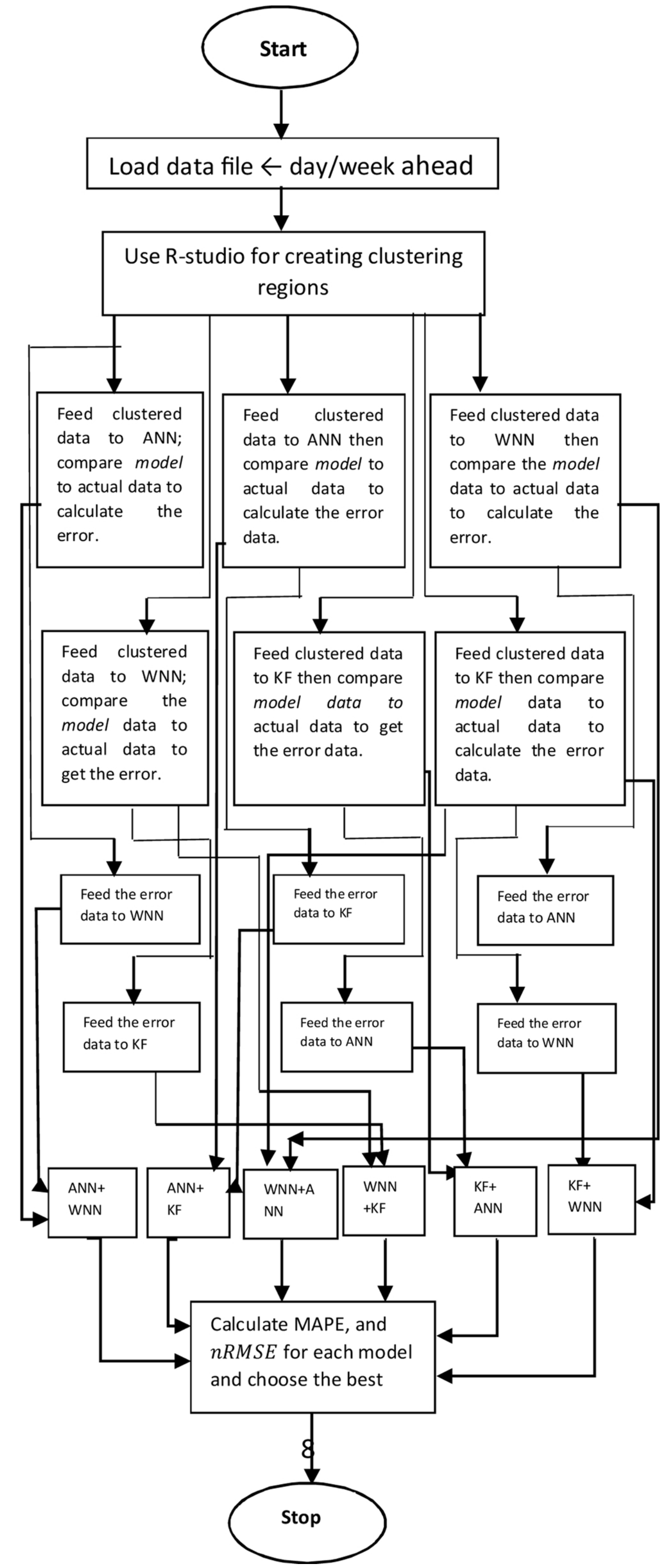

Fig. 2. Short-term load forecast proposed models.

5.3. Hybrid model of ANN and KF

In this model we used ANN to train the un-clustered (row) and clustered data. The forecasted data for the un-clustered model is compared to the actual data and we calculated the residuals and then feed it to KF. We did the same with the clustered data, but the data is aggregated before calculating the residuals. The error is fed then to KF for both cases (clustered and un-clustered data). The MAPE and nRMS are calculated for that model for both locations and it is shown in Table 1 for Nova Scotia and in Table 2 for Cairo. As we see from Table 1 the MAPE for un-clustered data is 2.238 and for clustered data is 2.208. Also, we calculated nRMS for both cases and it is 0.0258 for un-clustered data and 0.0210 for clustered data. As we see the clustering technique is improving the performance of the forecasting model for both cities. Fig. 7 shows the relationship between the actual data, forecasted data without clustering and the forecasted data with clustering technique for short term load forecasting and the hours of weekdays for a certain period of time in Nova Scotia. Fig. 8 shows the same relationship but for another location as a way of validation. From Figs. 5 and 6 we concluded that the clustered hybrid forecasting model is improving the forecasted data compared to the un-clustered.

Table 1
MAPE and nRMS for different models with and without clustering technique for Nova Scotia in case of weekdays forecasting for week ahead forecasting.

Nova scotia week ahead	Models without clustering		Models with clustering	
	MAPE	nRMSE	MAPE	nRMSE
ANN + WNN	2.247	0.0280	2.115	0.026
ANN + KF	2.238	0.0258	2.208	0.0210
WNN + ANN	**2.1771**	**0.0232**	**1.981**	**0.01920**
WNN + KF	2.4463	0.0278	2.4863	0.0296
KF + ANN	2.235	0.0291	2.215	0.027
KF + WNN	2.4208	0.0270	2.3208	0.0237

The bold values indicate the best model.

Table 2
MAPE and nRMS for different models with and without clustering technique for Cairo in case of weekdays forecasting for week ahead forecasting.

Cairo week ahead	Models without clustering		Models with clustering	
	MAPE	nRMSE	MAPE	nRMSE
ANN + WNN	2.9545	0.0292	2.7545	0.0083
ANN + KF	2.998	0.034	2.7838	0.0081
WNN + ANN	**2.3838**	**0.034**	**2.1538**	**0.0073**
WNN + KF	3.1293	0.0365	2.9293	0.0093
KF + ANN	2.6494	0.0335	2.7494	0.0082
KF + WNN	2.4524	0.0347	2.1824	0.0077

The bold values indicate the best model.

5.4. Hybrid model of KF and ANN

In this model we used KF to train the un-clustered (row) and clustered data. The forecasted data for the un-clustered model is compared to the actual data and we calculated the error which is the forecasted minus the actual. We did the same with the clustered data, but the data is aggregated before calculating the residuals. The residuals are then fed to ANN for both cases (clustered and un-clustered data). The MAPE and nRMS are calculated for that model for both cities and it is shown in Table 1 for Nova Scotia and in Table 2 for Cairo. As we see from Table 1 the MAPE for un-clustered data is 2.235 and for clustered data is 2.215. Also, we calculated nRMS for both cases and it is 0.0291 and for un-clustered data and 0.0270 for clustered data. As we see the clustering technique is improving the performance of the forecasting technique for both locations. Fig. 9 shows the relationship between the actual data, forecasted data without clustering and the forecasted data with clustering technique for short term load forecasting and the hours of weekday for a certain period of time in Nova Scotia. Fig. 10 shows the same relationship but for another location as a way of validation.

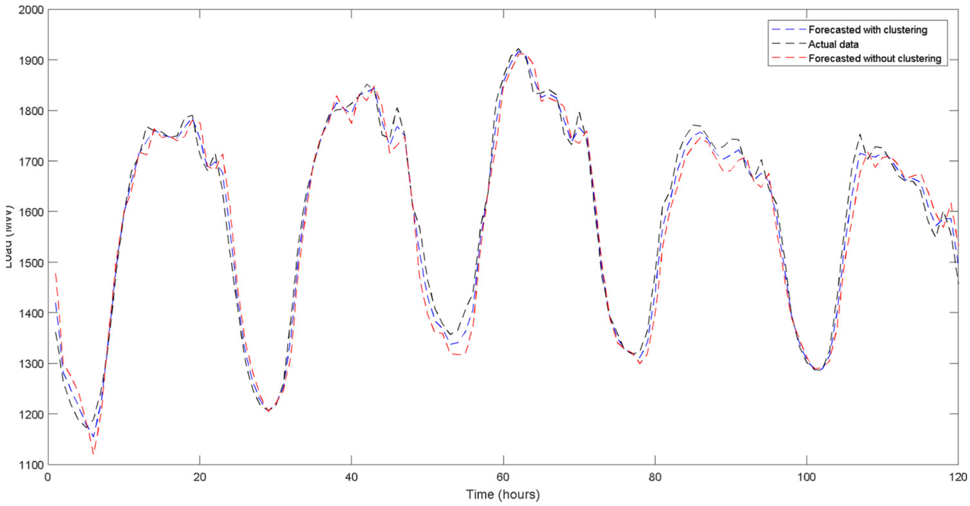

Fig. 3. The relationship between the actual data, forecasted data with and without clustering for short term load forecasting and the hours for weekdays for a certain period of time for Nova Scotia based on ANNWNN Hybrid model.

5.5. Hybrid model of WNN and KF

In this model we used WNN to train the un-clustered (row) and clustered data. The input to that model is the time in hours and the output is the load power in MW. The forecasted data for the un-clustered data is compared to the actual data and we calculated the error. We did the same with the clustered data, but the data is aggregated before calculating the error. The error is then fed to KF for both cases (clustered and un-clustered data). The MAPE and nRMS are calculated for that model for both provinces and it is shown in Table 1 for Nova Scotia and in Table 2 for Cairo. As we see from Table 1 the MAPE for un-clustered data is 2.235 and for clustered data is 2.215. Also we calculated nRMS for both cases and it is 0.0291 and for un-clustered data and 0.0270 for clustered data. As we see the clustering technique is improving the performance of the forecasting technique for both locations. Fig. 11 shows the relationship between the actual data, forecasted data without clustering and the forecasted data with clustering technique for short term load forecasting and the hours of weekday for a certain period of time in Nova Scotia. Fig. 12 shows the same relationship but for another location as a way of data validation (Table 3).

5.6. Hybrid model of KF and WNN

In this model we used KF to train the un-clustered (row) and clustered data. The forecasted data for the un-clustered data is compared to the actual data and we calculated the error. We did the same with the clustered data, but the data is aggregated before calculating the error. The error is fed then to WNN for both cases (clustered and un-clustered data). The MAPE and nRMS are calculated for that model for both provinces and it is shown in Table 1 for Nova Scotia and in Table 2 for Cairo. As we see from Table 1 the MAPE for un-clustered data is 2.4208 and for clustered data is 2.3208. Also, we calculated nRMS for both cases and it is 0.027 and for un-clustered data and is 0.0237 for clustered data. Fig. 13 shows the relationship between the actual data, forecasted data with and without clustering and the forecasted data with clustering technique for short term load forecasting and the hours of weekdays for a certain period of time in Nova Scotia. Fig. 14 shows the same relationship but for another location as a way of data validation (Table 4).

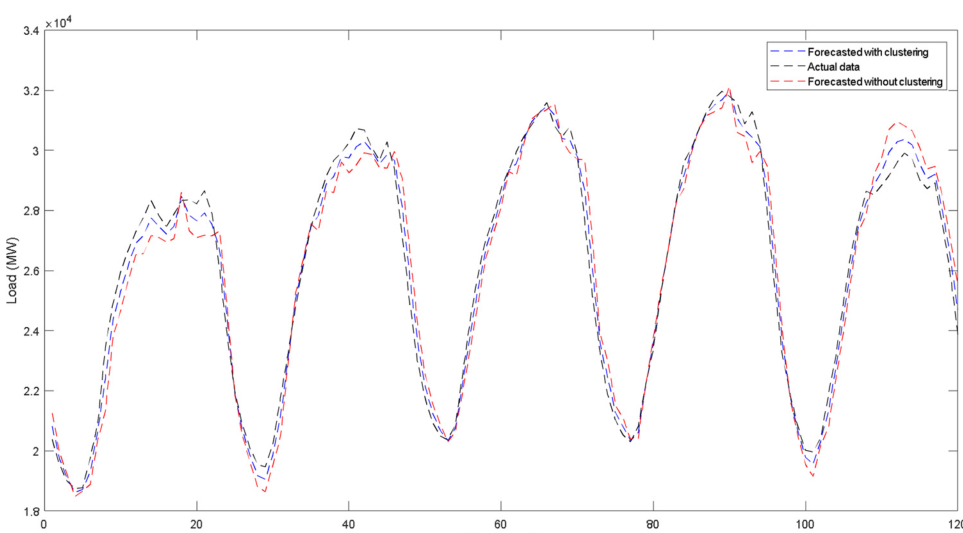

Fig. 4. The relationship between the actual data, forecasted data without and clustering for short term load forecasting and the hours for weekdays for a certain period for Cairo based on ANNWNN Hybrid model.

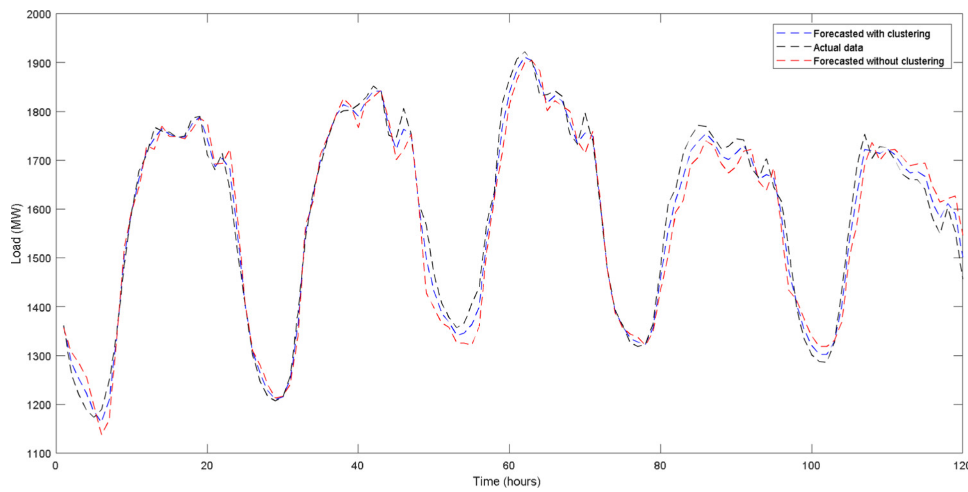

Fig. 5. The relationship between the actual data, forecasted data with and without clustering for short term load forecasting and the hours for weekdays for a certain period of time for Nova Scotia based on WNNANN Hybrid model.

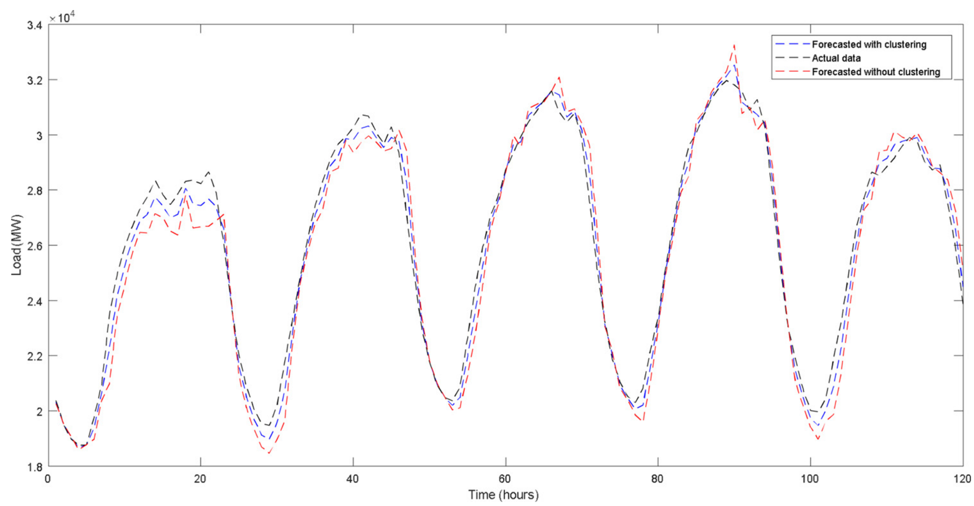

Fig. 6. The relationship between the actual data, forecasted data with and without clustering for short term load forecasting and the hours for weekdays for a certain period for Cairo based on WNNANN Hybrid model.

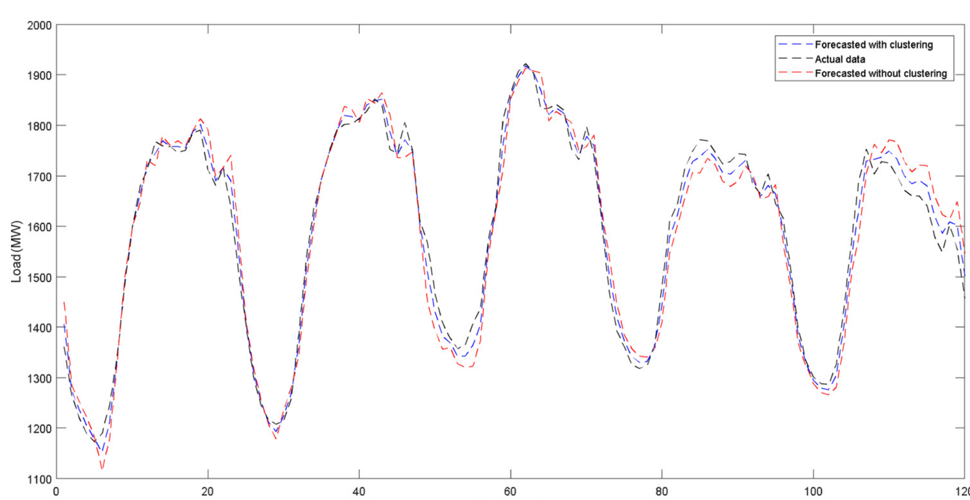

Fig. 7. The relationship between the actual data, forecasted data with and without clustering for short term load forecasting and the hours for weekdays for a certain period of time in Nova Scotia based on ANNKF Hybrid model.

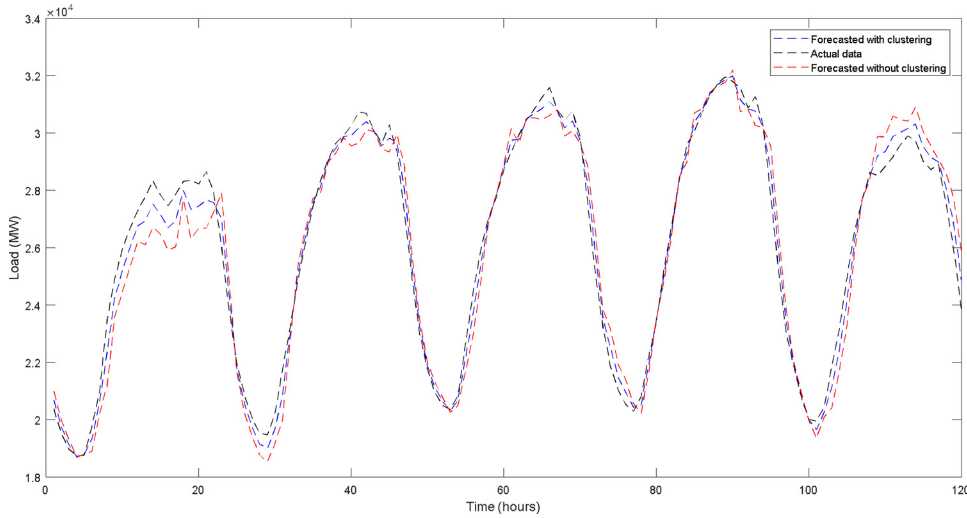

Fig. 8. The relationship between the actual data, forecasted data with and without clustering for short term load forecasting and the hours for weekdays for a certain period in Cairo based KFANN Hybrid model.

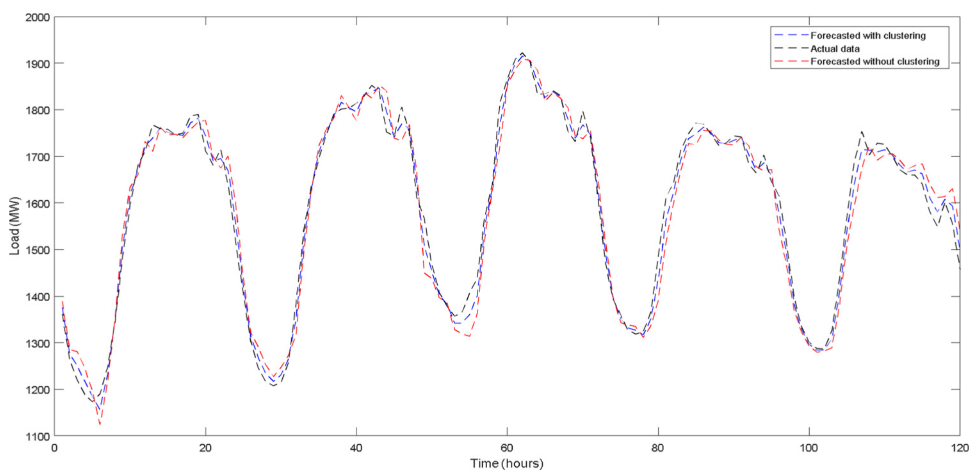

Fig. 9. The relationship between the actual data, forecasted data with and without clustering for short term load forecasting and the hours for weekdays for a certain period of time for Nova Scotia based on KFANN Hybrid model.

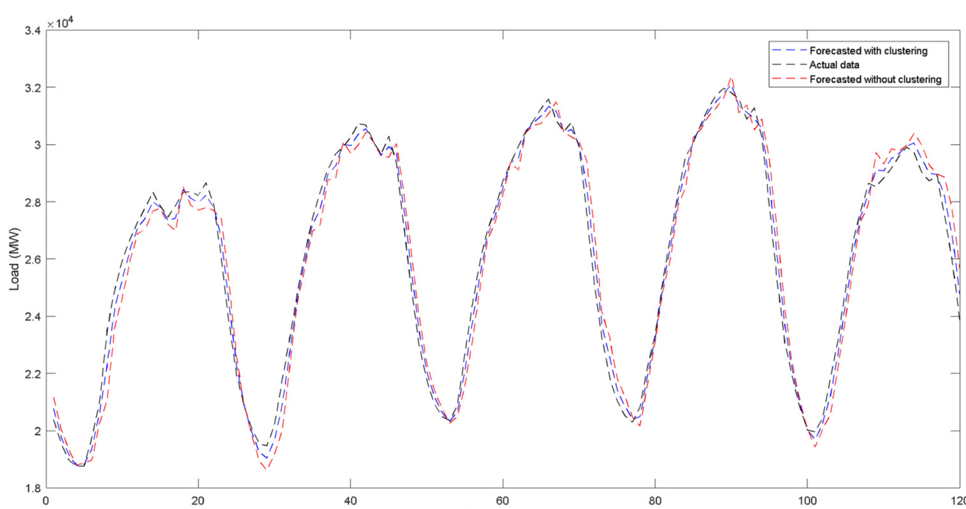

Fig. 10. The relationship between the actual data, forecasted data with and without clustering for short term load forecasting and the hours for weekdays for a certain period for Cairo based on KFANN Hybrid model.

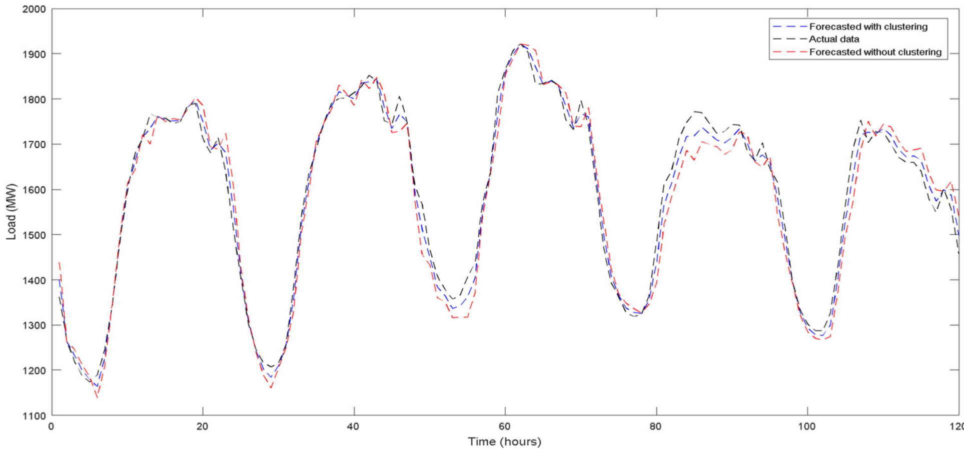

Fig. 11. The relationship between the actual data, forecasted data with and without clustering for short term load forecasting and the hours for weekdays for a certain period of time for Nova Scotia for WNNKF Hybrid model.

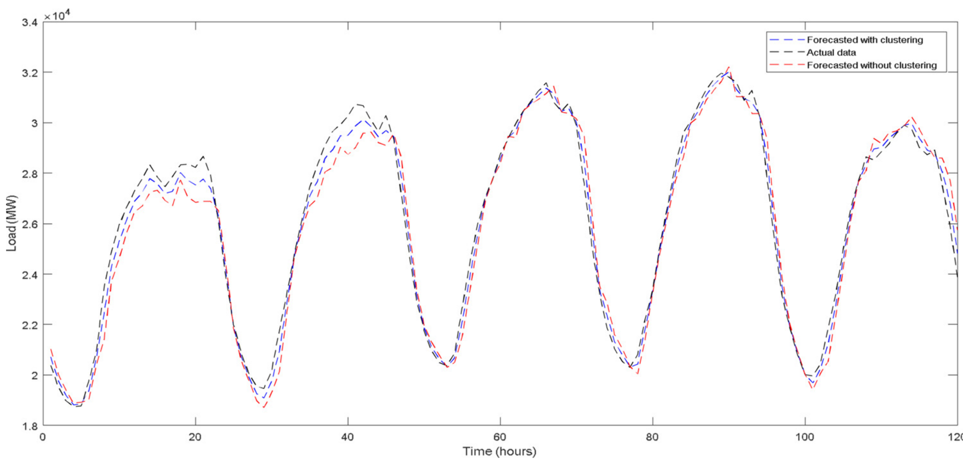

Fig. 12. The relationship between the actual data, forecasted data with and without clustering technique for short term load forecasting and the hours for weekdays for a certain period for Cairo based on KFWNN Hybrid model.

Table 3
MAPE and nRMS for different models with and without clustering technique for Nova Scotia in case of weekdays forecasting for day ahead forecasting.

Nova scotia day ahead	Models without clustering		Models with clustering	
	MAPE	nRMSE	MAPE	nRMSE
ANN + WNN	2.25791	0.0340	2.15791	0.0202
ANN + KF	2.4945	0.0361	2.3945	0.0252
WNN + ANN	**2.2322**	**0.0276**	**2.132**	**0.0172**
WNN + KF	2.5212	0.0298	2.5212	0.0302
KF + ANN 3	2.532	0.0311	2.486	0.0222
KF + WNN	2.9080	0.0471	2.5069	0.0292

The bold values indicate the best model.

5.7. Hybrid forecasting model based on day ahead

Since the hybrid model of WNN and ANN is the best model, so we used it to forecast day ahead for both locations. Figs. 15 and 16 show the forecasted data using WNNANN Hybrid model for the two locations. The MAPE is 2.001 for un-clustered data and is 1.7 for clustered data for Nova Scotia and is 2.05 and 1.8 respectively for Cairo. Also, nRMS is found to be 0.019 for un-clustered and is 0.0091 clustered data for Nova Scotia and is 0.023 for un-clustered and is 0.0051 clustered data for Cairo.

5.8. Hybrid forecasting model for weekends

The proposed hybrid model of WNN and ANN is used to forecast the weekends for both locations. Figs. 17 and 18 show the forecasted data using WNNANN Hybrid model for the two locations. The MAPE is 1.961 for un-clustered data and 1.75 for clustered data for Nova Scotia and is 2.10 and 1.8 for Cairo. Also, nRMS is found to be 0.029 for un-clustered and is 0.00501 clustered data for Nova Scotia and is 0.023 for un-clustered and is 0.00651 clustered data for Cairo.

From the above analysis and figures we conclude that using hybrid models is the best way for forecasting as the error is reduced. At the same time making some statistical analysis like clustering for the data will improve the performance of the forecasting model. Using clustering techniques before using the forecasting software technique is an efficient way as the overall error is reduced more. The error from the hybrid models is very small specially after using clustering techniques and the best hybrid model is the model of WNN and ANN.

6. Model parameters selection and validation

In this section we tried to optimize the model parameters of the individual algorithms by using different numbers of clusters, neurons, hidden layers and we choose the best one for our model. We tried to asses the output sensitivity for individual algorithm's error and we select the best model parameters then we applied the hybrid models based on the assessed parameters. We used KF alone at the beginning

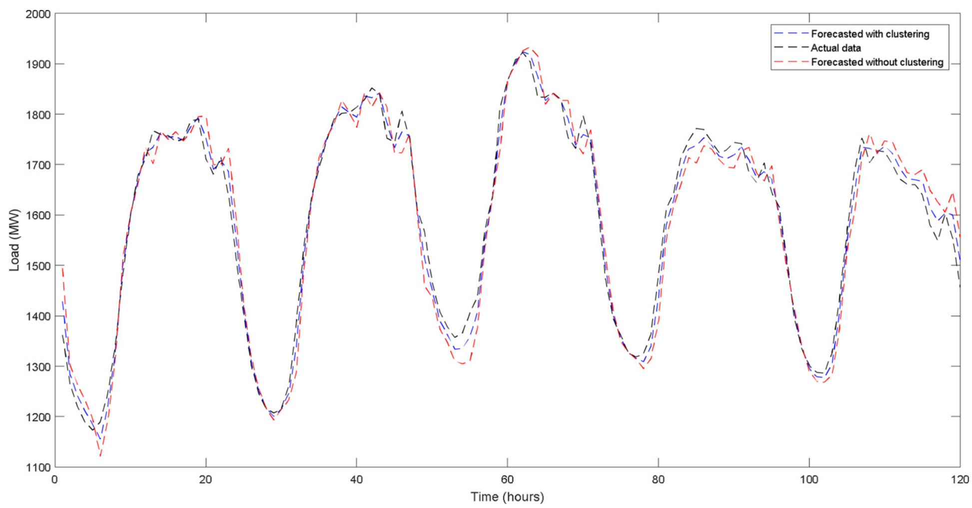

Fig. 13. The relationship between the actual data, forecasted data with and without clustering technique for short term load forecasting and the hours for weekdays for a certain period of time for Nova Scotia based on KFWNN Hybrid model.

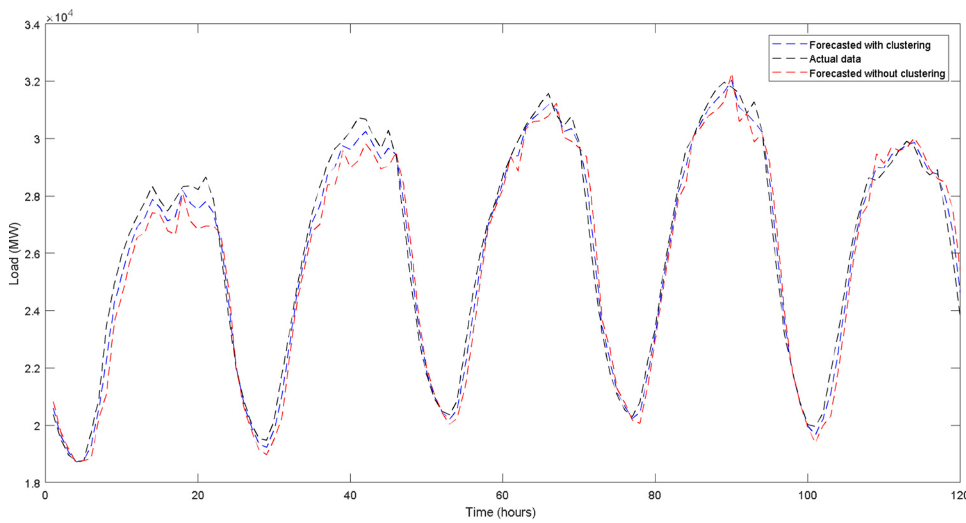

Fig. 14. The relationship between the actual data, forecasted data with and without clustering technique for short term load forecasting and the hours for weekdays for a certain period for Cairo based on WNNKF Hybrid model.

Table 4
MAPE and nRMS for different models with and without clustering technique for Cairo city in case of weekdays forecasting for day ahead.

Cairo day ahead	Models without clustering		Models with clustering	
	MAPE	nRMSE	MAPE	nRMSE
ANN + WNN	2.25791	0.0406	2.05791	0.0302
ANN + KF	2.3945	0.0421	1.9945	0.0292
WNN + ANN	**2.2322**	**0.0276**	**1.8322**	**0.0191**
WNN + KF	2.5212	0.0598	2.5212	0.0302
KF + ANN	2.386	0.0361	2.106	0.0282
KF + WNN	2.6069	0.0521	2.2069	0.02999

The bold values indicate the best model.

and we then changed the number of clusters and calculated MAPE and nRMSE as shown in Table 5. We found that the best performance of KF is when we have six clusters segments and the MAPE is 4.263 and nRMSE is 0.060. We tried to do the same using ANN and WNN. In Table 6 we trained ANN with different number of hidden layers and neurons to determine the best number of neurons and hidden layers for the best performance. As seen from Table 6; the best performance is when the number of neurons is 75 and the number of hidden layers is 2. We did the same with WNN and it gave the same result as shown in Table 7. After that we used the best parameters with the hybrid models as seen in Table 8. Table 8 shows the MAPE and nRMS for different models with different numbers of clusters, neurons, and hidden layers for Nova Scotia and Cairo in case of weekdays forecasting for week ahead for a sample of used data. The best performance is when the number of neurons is 75 and there are two hidden layers using six clusters regions.

Another way of validation is to use K-fold cross validation. In K-fold cross validation the data is first partitioned into K approximately equally sized segments. Data is commonly stratified prior to being split which means rearranging the data to ensure each fold is a good representative of the whole [44]. Table 9 shows a comparison between 8-fold 6-fold and 4-fold cross validation.

7. Conclusions

Due to the fluctuation and uncertainties of the requirements of new generation of renewable energy resources; smart grid is needed. Load forecasting is an important key for smart grid especially when we deal with renewable energy resources. Using either KF, ANN or WNN alone for the short-term load forecasting is not recommended as the error is

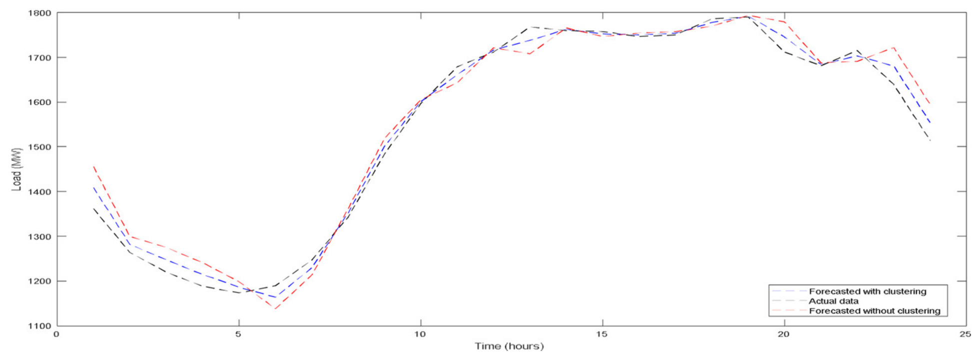

Fig. 15. The relationship between the actual, forecasted data with and without clustering and the hours in NS using WNNANN for day ahead.

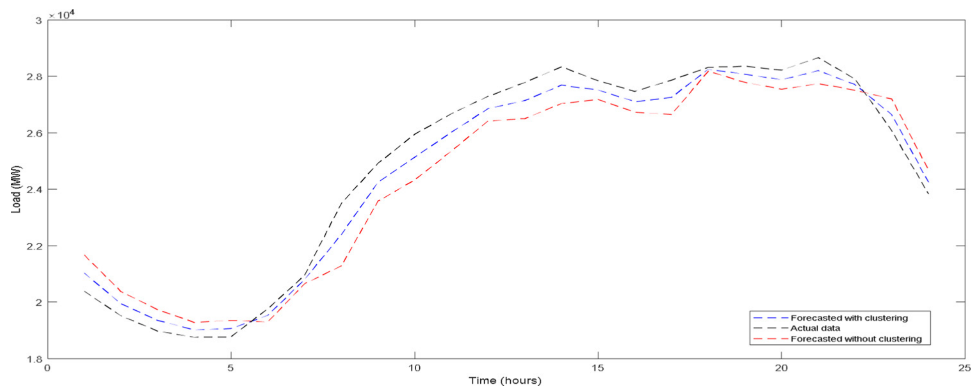

Fig. 16. The relationship between actual, forecasted data with and without clustering and hours in Cairo using WNNANN for day ahead.

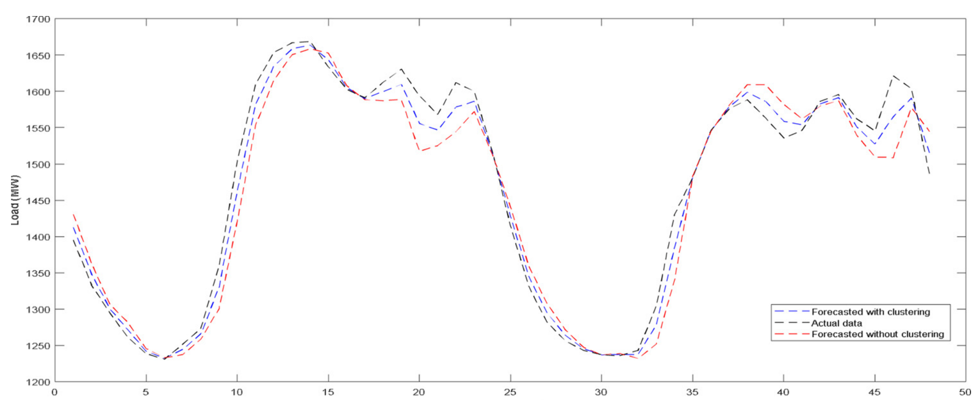

Fig. 17. The relationship between the actual, forecasted data with and without clustering and hours for weekend in NS based on WNNANN Hybrid model.

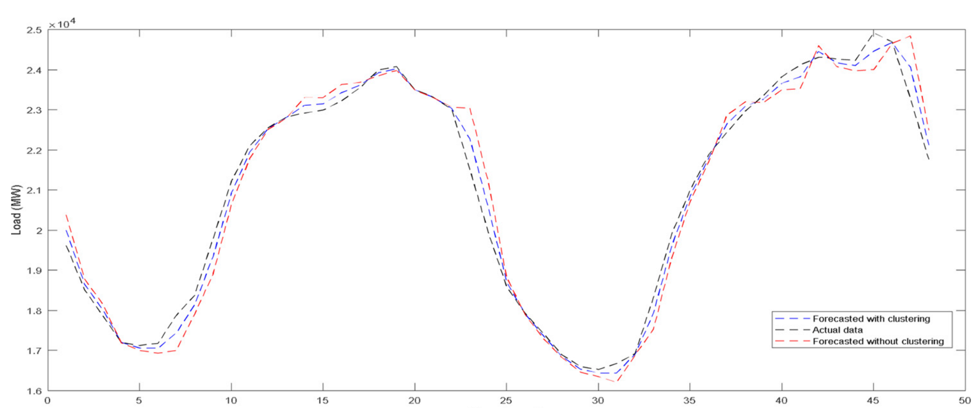

Fig. 18. The relationship between the actual, forecasted data with and without clustering and hours for weekend in Cairo based on WNNANN Hybrid model.

Table 5
KF with different clustering numbers.

Clusters No	MAPE	nRMSE
2	4.542	0.068
4	4.351	0.064
6	**4.021**	**0.042**
8	4.263	0.060
10	4.201	0.057

The bold values indicate the best model.

Table 6
ANN with different number of hidden layers and neurons.

Neurons no	Hidden layers no	MAPE	nRMSE
25	1	6.0921	0.0861
25	2	6.1152	0.08967
25	3	6.1656	0.09408
25	4	6.2475	0.09576
50	1	5.8401	0.08379
50	2	5.9976	0.08463
50	3	5.4642	0.08127
50	4	4.6326	0.05208
75	1	4.1601	0.04032
75	**2**	**4.2252**	**0.04263**
75	3	4.2714	0.04914
75	4	4.9476	0.07938
100	1	5.0841	0.08001
100	2	5.1366	0.08022
100	3	5.1891	0.08064
100	4	4.8221	0.05670
125	1	4.8321	0.05796
125	2	4.8552	0.05817
125	3	4.9245	0.06027
125	4	4.9392	0.07917

The bold values indicate the best model.

Table 7
WNN with different number of hidden layers and neurons.

Neurons no	Hidden layers no	MAPE	nRMSE
25	1	5.58289	0.07849
25	2	5.60368	0.081703
25	3	5.64904	0.085672
25	4	5.72275	0.087184
50	1	5.35609	0.076411
50	2	5.49784	0.077167
50	3	5.01778	0.074143
50	4	4.26934	0.047872
75	1	3.84409	0.037288
75	**2**	**3.90268**	**0.039367**
75	3	3.94426	0.045226
75	4	4.55284	0.072442
100	1	4.67569	0.073009
100	2	4.72294	0.073198
100	3	4.77019	0.073576
100	4	4.43989	0.05203
125	1	4.44889	0.053164
125	2	4.46968	0.053353
125	3	4.53205	0.055243
125	4	4.54528	0.072253

The bold values indicate the best model.

high. In this paper different novel models are proposed for short-term load forecasting based on clustering techniques. We proposed more than one hybrid model to choose from depending on the data input. Clustering techniques proved the effectiveness of the proposed models. It is recommended to use hybrid models of either WNN and ANN, WNN and KF, ANN and WNN, ANN and KF, KF and ANN or KF and WNN. All proposed models are accurate compared to the conventional models used in the literature work but for higher accuracy it is recommended to use hybrid model of WNN and ANN with two hidden layers, seventy-five neurons and six cluster regions. The order of the hybrid techniques used is making a difference in the error value.

Table 8
MAPE and nRMS for different models with different numbers of clusters, neurons, and hidden layers for Nova Scotia and Cairo in case of weekdays forecasting for week ahead.

Number of			Nova scotia data week ahead		Cairo data week ahead	
Clusters	Neurons	Hidden layers	MAPE	nRMSE	MAPE	nRMSE
2	50	1	2.975	0.0456	3.254	0.0363
2	75	2	2.936	0.0448	3.117	0.0343
2	100	3	2.901	0.0410	3.103	0.0327
2	125	4	2.912	0.0427	3.110	0.0334
4	50	1	2.856	0.0403	2.990	0.0239
4	75	2	2.781	0.0399	2.732	0.0208
4	100	3	2.602	0.0387	2.688	0.0190
4	125	4	2.471	0.0384	2.576	0.0186
6	50	1	2.345	0.0287	2.531	0.0130
6	**75**	**2**	**1.981**	**0.01920**	**2.1538**	**0.0073**
6	100	3	2.012	0.0203	2.231	0.0090
6	125	4	2.034	0.0234	2.312	0.0101
8	50	1	2.446	0.0382	2.567	0.0185
8	75	2	2.206	0.0248	2.407	0.0123
8	100	3	2.301	0.0270	2.630	0.0187
8	125	4	2.352	0.0377	2.687	0.0189
10	50	1	2.356	0.0378	2.690	0.0191
10	75	2	2.301	0.0276	2.632	0.0188
10	100	3	2.312	0.0277	2.547	0.0147
10	125	4	2.421	0.0381	2.560	0.0178

The bold values indicate the best model.

Table 9
Comparison between 8-fold 6-fold and 4-fold cross validation for day a head for Cairo.

WNN + ANN techniques for 8-fold 6-fold and 4-fold	Models with clustering	
	MAPE	nRMSE
4-fold	1.283	0.0320
6-fold	**1.0332**	**0.0081**
8-fold	1.190	0.0324

The bold values indicate the best model.

We validated this work by applying different proposed models on different dataset for two countries (Canada and Egypt). Also, we used K-fold cross validation method to validate the proposed work. From Table 9 we conclude that the best size is 6-fold as it has the smallest (MAPE and nRMSE) error compared to 8-fold and 4-fold. The simulation results proved the effectiveness of the proposed models. Using hybrid models improve the overall system accuracy as the advantages of all models are added together in one model as well as the advantages of using clustering techniques improve the variances for the data. Simulations results proved that using Hybrid of ANN and WNN with six segments clustering is considered as the best model for short-term load forecasting.

Author contribution

I am the only author for the paper and I did everything by myself without any help from authors.

Conflict of interest

I don't have conflict of interest or funding resources to declare.

Acknowledgment

To the soul of my academic father who supported me a lot; my PhD and post-doctorate supervisor Dr. Mohamed El-Hawary who had passed away last year. May his soul rest in peace.

References

[1] Weicong Kong, Zhao Yang Dong, David J. Hill, Fengji Luo, Yan Xu, Short-term residential load forecasting based on resident behaviour learning, IEEE Trans. Power Syst. 33 (January (1)) (2018).
[2] M. Ghofrani, M. Hassanzadeh, M. Etezadi-Amoli, M.S. Fadali, Smart meter based short-term load forecasting for residential customers, Proc. 2011 North Amer. Power Symp. (2011).
[3] M. Chaouch, Clustering-based improvement of nonparametric functional time series forecasting: application to intra-day household-level load curves, IEEE Trans. Smart Grid 5 (January (1)) (2014).
[4] K.H. Kim, J.K. Park, K.J. Hwang, S.H. Kim, Implementation of hybrid short-term load forecasting system using artificial neural networks and fuzzy expert systems, IEEE Trans. Power Syst. 10 (August) (1995).
[5] S.J. Huang, C.L. Huang, Genetic-based multi-layered perceptions for Taiwan power system short-term load forecasting, Int. J. Electr. Power Syst. Res. 38 (July (3)) (1996).
[6] N. Amjady, Short-term hourly load forecasting using time-series modeling with peak load estimation capability, IEEE Trans. Power Syst. 16 (November) (2001).
[7] Hamed H.H. Aly, M.E. El-Hawary, A proposed ANN and FLSM hybrid model for tidal current magnitude and direction forecasting, IEEE J. Ocean. Eng. 39 (January (1)) (2014).
[8] Hamed H.H. Aly, M.E. El-Hawary, A proposed algorithms for tidal in-stream speed model, Am. J. Energy Eng. 1 (1) (2013) 1–10.
[9] S. Fan, K. Methaprayoon, W.J. Lee, Multiregion load forecasting for system with large geographical area, IEEE Trans. Ind. Appl. 45 (August (4)) (2009) 1452–1459.
[10] H.S. Hippert, C.E. Pedreira, R.C. Souza, Neural networks for short-term load forecasting: a review and evaluation, IEEE Trans. Power Syst. 16 (February (1)) (2001) 44–55.
[11] S. Fan, R.J. Hyndman, Short-term load forecasting based on a semi-parametric additive model, IEEE Trans. Power Syst. 27 (2012) 134–141.
[12] J.M. Vilar, R. Cao, G. Aneiros, Forecasting next-day electricity demand and price using nonparametric functional methods, Int. J. Electr. Power Energy Syst. 39 (1) (2012) 48–55.
[13] Mohamed Chaouch, Clustering-based improvement of nonparametric functional time series forecasting: application to intra-day household-level load curves, IEEE Trans. Smart Grid 5 (January (1)) (2014).
[14] C.V. Cardoso, G.L. Cruz, Forecasting natural gas consumption using ARIMA models and artificial neural networks, IEEE Lat. Am. Trans. 14 (5) (2016).
[15] F. Yu, X. Xu, A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network, Appl. Energy 134 (2014) 102–113.
[16] I.P. Panapakidis, Electrical power and energy systems clustering based day-ahead and hour-ahead bus load forecasting models, Int. J. Electr. Power Energy Syst. 80 (2016) 171–178.
[17] J. Jiménez, K. Donado, C.G. Quintero, A methodology for short-term load forecasting, IEEE Lat. Am. Trans. 15 (March (3)) (2017).
[18] M. Ghofrani, M. Hassanzadeh, M. Etezadi-Amoli, M.S. Fadali, Smart meter based short-term load forecasting for residential customers, North American Power Symposium (NAPS), Boston, MA, USA, 2011.
[19] Hao Quan, Dipti Srinivasan, Abbas Khosravi, Short-term load and wind power forecasting using neural network-based prediction intervals, IEEE Trans. Neural Netw. Learn. Syst. 25 (February (2)) (2014).
[20] C. Cecati, J. Kolbusz, P. Rozcki, P. Siano, B.M. Wilamowski, A novel RBF training algorithm for short-term electric load forecasting and comparative studies, IEEE Trans. Ind. Electron. 62 (10) (2015).
[21] W.-M. Lin, C.-S. Tu, R.-F. Yang, M.-T. Tsai, Particle swarm optimization aided least-square support vector machine for load forecast with spikes, IET Proc. Gener. Transm. Distrib. 10 (5) (2016).
[22] C. Guan, P.B. Luh, L.D. Michel, Y. Wang, P.B. Friedland, Very short-term load forecasting: wavelet neural networks with data pre-filtering, IEEE Trans. Power Syst. 28 (1) (2013).
[23] Bowen Li, Jing Zhang, Yu He, Yang Wang, Short-term load-forecasting method based on wavelet decomposition with second-order gray neural network model combined with ADF test, IEEE Open Access J. 5 (2017) 16324–16331.
[24] N. Amjaday, F. Keynia, Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuroevolutionary algorithm, IEEE Trans. Power Syst. 24 (1) (2009).
[25] Ashfaq Ahmad, Nadeem Javaid, Mohsen Guizani, Nabil Alrajeh, Zahoor Ali Khan, An accurate and fast converging short-term load forecasting model for industrial applications in a smart grid, IEEE Trans. Ind. Inform. 13 (October (5)) (2017).
[26] D.K. Chaturvedi, A.P. Sinha, O.P. Malik, Short term load forecast using fuzzy logic and wavelet transform integrated generalized neural network, Int. J. Electr. Power Energy Syst. (May) (2015) 230–237.
[27] Kandil Nahi, Wamkeue Rene, Saad Maarouf, Georges Semaan, An efficient approach for short-term load forecasting using neural networks, Int. J. Electr. Power Energy Syst. 28 (8) (2006) 525–530.
[28] D.K. Chaturvedi, P.S. Satsangi, P.K. Kalra, New neuron model for simulating rotating electrical machines and load forecasting problems, Int. J. Electr. Power Energy Syst. 52 (1999) 123–131.
[29] Y. Jie, Z. Mingzhan, W. Qinglin, Correlative analysis of measured data between anemometer tower and WTG, 8th Int. Conf. on Computing and Networking Technology (ICCNT), Gueongju, 2001.
[30] T. Senjyu, H. Takara, K. Uezato, T. Funabashi, One-hour-ahead load forecasting using neural network, IEEE Trans. Power Syst. 17 (February) (2002) 113–118.
[31] Hamed H. Aly, A novel approach for tidal currents harmonic constitutions forecasting hybrid models based on clustering techniques for smart grid, Int. J. Renew. Energy 147 (March (1)) (2020) 1554–1564.
[32] https://www.mathworks.com/help/stats/cluster-analysis.html. Last access December 2019.
[33] https://www.statmethods.net/advstats/cluster.html. Last access December 2019.
[34] Franklin L. Quilumba, Wei-Jen Lee, Heng Huang, David Y. Wang, Robert L. Szabados, Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities, IEEE Trans. Smart Grid 6 (March (2)) (2015).
[35] Hamed Aly, Forecasting, Modeling and Control of Tidal Currents Electrical Energy Systems. Ph.D. Dissertation, Dept. Elec. Eng., Dalhousie Univ., Halifax, N.S, 2012.
[36] B.T. Xie, H. Yu, B. Wilamowski, Comparison between traditional neural networks and radial basis function networks, 2011 IEEE International Symposium on Industrial Electronics, Gdansk, 2011, pp. 1194–1199.
[37] J.B. MacQueen, Some methods for classification and analysis of multivariate observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1976.
[38] L. Rokach, A survey of clustering algorithms, in: O. Maimon, L. Rokach (Eds.), Data Mining and Knowledge Discovery Handbook, Springer, Boston, MA, 2009.
[39] Simon Haykin, Neural Networks: A Comprehensive Foundation, 2nd ed., Prentice-Hall, New Jersey, 1999.
[40] J. Moody, C.J. Darken, Fast learning in networks of locally-tuned processing units, Neural Comput. 1 (2) (1989) 281–294.
[41] Z. Zainuddin, O. Pauline, Improved wavelet neural networks for early cancer diagnosis using clustering algorithms, Int. J. Inform. Math. Sci. 6 (1) (2010) 30–36.
[42] D. Simon, Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches, John Wiley & Sons, Oxford, 2006.
[43] Fred C. Schweppe, Uncertain Dynamic Systems, Prentice-Hall, Englewood Cliffs, N.J, 1973.
[44] Juan D. Rodriguez, Aritz Perez, Jose A. Lozano, Sensitivity analysis of k-fold cross validation in prediction error estimation, IEEE Trans. Pattern Anal. Mach. Intell. 32 (March (3)) (2010).

Hamed H. H. Aly received the B.Eng. and M.ASc. degrees in electrical engineering with distinction, in 1999 and 2005, respectively, from Zagazig University, Egypt and the Ph.D. degree from Dalhousie University, Canada, in 2013. He was working as Postdoctoral Research Associate for one year and as a lecturer for four years at Dalhousie University. He worked as a lecturer at Zagazig University for eight years. His research interests include: Power System Dynamics and Stability; Power Electronics; Electric Machines, Power Quality Issues; Applications of Artificial Intelligence in Power Systems; Smart Grid; and the Integration of Renewable Energy into the Electrical Grid. He taught in different universities in Canada as well as in the Canadian Forces Naval Engineering School. He pioneered many computational and artificial intelligence solutions to problems in economic/environmental operation of power systems. He has written one textbook, three chapters books and more than 50 refereed journal and conference articles. He has consulted and taught for more than 16 years. He is a senior member of the Institute of Electrical and Electronics Engineers (IEEE). Dr. Aly is a faculty member at Acadia University since 2013. Dr. Aly is currently an Assistant Professor at Faculty of Engineering, Zagazig University and an adjunct Assistant Professor at Dalhousie University.