ORIGINAL ARTICLE

An adaptive backpropagation algorithm for long-term electricity load forecasting

Nooriya A. Mohammed[1] · Ammar Al-Bazi[2] (ID)

Abstract

Artificial Neural Networks (ANNs) have been widely used to determine future demand for power in the short, medium, and long terms. However, research has identified that ANNs could cause inaccurate predictions of load when used for long-term forecasting. This inaccuracy is attributed to insufficient training data and increased accumulated errors, especially in long-term estimations. This study develops an improved ANN model with an Adaptive Backpropagation Algorithm (ABPA) for best practice in the forecasting long-term load demand of electricity. The ABPA includes proposing new forecasting formulations that adjust/adapt forecast values, so it takes into consideration the deviation between trained and future input datasets' different behaviours. The architecture of the Multi-Layer Perceptron (MLP) model, along with its traditional Backpropagation Algorithm (BPA), is used as a baseline for the proposed development. The forecasting formula is further improved by introducing adjustment factors to smooth out behavioural differences between the trained and new/future datasets. A computational study based on actual monthly electricity consumption inputs from 2011 to 2020, provided by the Iraqi Ministry of Electricity, is conducted to verify the proposed adaptive algorithm's performance. Different types of energy consumption and the electricity cut period (unsatisfied demand) factor are also considered in this study as vital factors. The developed ANN model, including its proposed ABPA, is then compared with traditional and popular prediction techniques such as regression and other advanced machine learning approaches, including Recurrent Neural Networks (RNNs), to justify its superiority amongst them. The results reveal that the most accurate long-term forecasts with the minimum Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) values of (1.195.650) and (0.045), respectively, are successfully achieved by applying the proposed ABPA. It can be concluded that the proposed ABPA, including the adjustment factor, enables traditional ANN techniques to be efficiently used for long-term forecasting of electricity load demand.

Keywords MLP neural networks · Load demand · Long-term forecasting · Adaptive backpropagation · Linear regression · Radial basis function networks · Recurrent neural networks

1 Introduction

Different techniques have been developed for electricity load demand forecasting over short, medium, and long-term time scales during the past few years. These techniques range from statistical models such as regression and time-series approaches to Artificial Neural Networks (ANNs), machine learning, and expert systems [1]. ANNs, as a novel modelling technique, has been used in numerous studies of short and long-term forecasting. This technique has attributes, such as flexible computing frameworks and universal approximators, enabling it to solve forecasting problems in different fields with a high degree of accuracy [2].

Traditional ANN techniques represent complex nonlinear relationships between dependent variables and other influencing variables. For example, forecasting load demand of electricity is influenced by the Gross Domestic Product (GDP) time series. The latter is highly correlated

✉ Ammar Al-Bazi
aa8535@coventry.ac.uk

1 Planning and Studies Office, Ministry of Electricity, Baghdad, Iraq

2 School of Mechanical, Aerospace and Automotive Engineering, Coventry University, Coventry, UK

with different types of electricity consumption and the electricity unit's price [3]. The literature reveals that this ANN technique provides better short-term forecasting performance rather than long-term. This finding has been proven by authors in Ref. [4], who used a traditional ANN technique to forecast short-term load demand, and confirmed that this technique has several limitations, one of which is the accuracy of long-term forecasts. In Ref. [5], the authors concluded that ANN is more efficient in short-term electricity demand prediction than long-term forecasting. In Ref. [6], the authors confirmed that ANN techniques offer a powerful forecasting tool in planning energy usage and can be used to generate accurate forecasts only for short periods. This poor accuracy/performance has been attributed to insufficient training data and an increased accumulation of errors in longer-term estimations [7]. Also, ANNs do not generate better results than other techniques for long-term predictions [8].

Therefore, this paper aims to propose an Adaptive Backpropagation Algorithm (ABPA) for more accurate and robust long-term predictions. This accuracy is achieved by further improving the forecasting part of the traditional Backpropagation Algorithm (BPA) algorithm to be able to accommodate the impact of accumulated errors caused over long periods of prediction.

The main contributions of this study can be summarised as follows:

(1) To propose an Adaptive Backpropagation Algorithm, represented by new forecasting formulations that capture the deviation caused by different behaviours of trained and future input datasets (long term). The new formulation will accommodate the deviation caused while generating high accuracy forecast outputs.

(2) To measure the impact of the normalisation of data on the quality of long-term forecasting outcomes. This impact includes identifying factors in a specific range that, if considered by the proposed ABPA, will improve its performance for more reliable and sustainable long-term forecasts.

(3) To enable traditional ANN techniques with the proposed ABPA to be used for long-term electricity load demand forecasting.

This improved ANN model's benefits are that it will assist energy operators, including companies of generation and transmission, to predict the amount of energy needed for best consumer demand satisfaction. This assistance contributes to achieving the most efficient allocation of energy resources, best practice of electricity scheduling, and optimal replacement plans for current outdated energy sources (if found). Also, this work contributes to the formulation of several management policies on electricity

planning, scheduling, and distribution. These policies include encouraging suppliers to invest more in increasing their energy sources, use state-of-the-art computerised energy scheduling systems, and adopt other energy sources such as renewable energy in selected regions.

The rest of the paper is organised as follows: Sect. 2 reviews previous work on applications of traditional ANNs, their hybrids, and other advanced machine learning approaches in long-term furcating of load demand. In Sect. 3, the proposed ABPA is discussed alongside the traditional Backpropagation Algorithm's (BPA) limitations. In Sect. 4, a numerical case study is conducted, followed by a comparison study to justify the proposed algorithm's superiority. The last section addresses this study's main findings and recommendations for future work.

2 Literature review

A thorough review of machine learning approaches and their application in long-term load forecasting, including ANN and deep learning, is conducted in this section. This review provides a clear understanding of methodologies used in long-term forecasting.

2.1 Application of ANNs in long term-load forecasting

This section presents the applications of traditional ANNs and their hybrids in the long-term forecasting of electricity load demand.

The practice of using different methods for long-term load demand forecasting was demonstrated, and found that most of the basic ANNs were successfully applied in short-term forecasting. However, the most suitable ANN network model for long-term forecasting requires careful consideration of network architecture and its training method [9]. A Hierarchical Neural Network (HNN) model was proposed in Ref. [10] to forecast both short and long-term electricity demand. The developed HNN model generated more realistic predictions than the non-hierarchical simple ANN. In Ref. [11], a Hierarchical Hybrid Neural (HHN) model was developed to tackle the problem of long-term peak load forecasting. In Ref. [12], an Adaptive Multilayer Perceptron (AMP) algorithm consisting of eight steps was proposed for dynamic time series forecasting with minimal complexity. Results showed that the proposed scheme for the AMP algorithm could produce accurate and robust forecasting of electricity load consumption. In Ref. [13], a method that combines neural network models of load forecasting with Virtual Instrument Technology based on Radial Basis Function (RBF) neural networks was

introduced to build a virtual forecaster for reliable short, medium, and long-term load forecasting. Results indicated that the forecasting model based on the RBF has high accuracy and stability. In Ref. [14], a new method based on Multi-Layer Perceptron Artificial Neural Network (MLP-ANN) was proposed for long-term load forecasting. The comparison test results illustrate the advantages of the proposed method. In Ref. [15], the ANNs model was integrated with an Adaptive Neuro-Fuzzy Inference System (ANFIS) to analyse the effects of different weather variables and actual and previous energy for forecasting the electricity load. The developed ANFIS model produced more accurate and reliable long-term energy load forecasts than the basic ANN model. In Ref. [16], different Machine Learning (ML) approaches, including ANNs, Multiple Linear Regression (MLR), ANFIS and Support Vector Machine (SVM), were applied to forecast long-term electricity load. The results indicated that SVM provided more reliable and accurate results than other ML techniques. In Ref. [17], three models based on Multivariate Adaptive Regression Splines (MARS), ANNs, and Linear Regression (LR) methods were utilised for long-term load forecasting. The study concluded that the MARS model gives more accurate and stable results than the ANN and LR models. In Ref. [18], two approaches of ANNs and SVMs were compared to predict long-term electrical load and found that the performance of forecasting using SVMs is consistently better than ANN. In Ref. [19], Support Vector Regression (SVR) and ANN approaches were used to identify electricity's best prediction model. The outputs indicated that the seasonal SVR model outperformed the ANN model in terms of the generated forecasts' accuracy. In Ref. [20], different load forecasting techniques were reviewed and compared for power forecasting. These methods include ANN, Support Vector Regression (SVR), Decision Tree (DT), LR, and Fuzzy Sets (FS). In Ref. [21], a hybrid method based on the combination of Particle Swarm Optimisation (PSO), Auto-Regressive Integrated Moving Average (ARIMA), ANN, and the proposed SVR technique was used to forecast the long-term load and energy demands. The results showed that the proposed PSO-SVR and hybrid methods give more accurate results than ARIMA and ANN methods. In Ref. [22], accurate and precise medium and long-term district-level energy prediction models were proposed employing machine learning-based models. The ANN with nonlinear autoregressive exogenous multivariable inputs was one of these models. This model was identified as the best forecasting model with a minimum MAPE value. In Ref. [23], a Univariate Multi-Model (UMM) based on neural networks was proposed to forecast electrical load. The purpose was to increase the performance of mid to long-term forecasting. However, this technique was greedy, requiring large amounts of data for training purposes. Ref. [24] suggested three models based on Multivariate Adaptive Regression Splines (MARS), ANN, and LR methods for long-term forecasting of load demand. MARS was reported to be more computationally efficient than ANN. The MARS model gives more accurate and stable results than ANN and LR models by comparing these models. In Ref. [25], a Multiple Neural Network (MNN) model was proposed for long-term time series prediction. In this MNN model, each neural network component makes forecasts for a different length of time. The results revealed that the developed MNN model outperformed the single ANN model. In Ref. [26], a hybrid approach based on Wavelet Support Vector Machines (WSVM) and Chaos Theory were employed for mid-term load forecasting. In Ref. [27], the Prophet and Holt-Winters forecasting models were used for long-term peak loads forecasting. The Prophet model has proven to be more robust to noise than the Holt-Winters model. In Ref. [28], the Fuzzy Logic (FL) concept was applied to ANN for long-term electric load forecasting. A feedforward input-delay back propagation network was used to design the ANN model. It was concluded that, in long-term load forecasting, both ANN and FL are powerful tools with a minimal error rate. In Ref. [29], FS was applied to ANN for modelling long-term uncertainties, and the enhanced forecasting results were compared with those of traditional methods, including regression. Results indicated that although ANN has its flexibility in handling nonlinear systems, it cannot model all corresponding factors of long-term load. In Ref. [30], the accuracy of the results of statistic methods (Exponential Smoothing, ARIMA, Regression), FL, and ANN algorithms were used and compared to estimate electricity consumption. Double exponential smoothing was the best method for producing approximate data close to the target. In Ref. [31], a Variable Structure Artificial Neural Network (VSANN) was used to improve the forecast accuracy of electricity load for the mid to long-term. However, applying a simple ANN model did not lead to a minimum load forecast error. In Ref. [32], an approach based on dynamic Feedforward Backpropagation Artificial Neural Network (FBP-ANN) was presented for long-term forecasting of total electricity demand. The proposed approach proved its accuracy along with effectiveness in long-term forecasting. In Ref. [33], a modified technique based on an ANN combined with LR was applied to conduct long-term electrical load demand forecasting. Application results showed that the proposed hybrid technique is feasible and effective.

2.2 Application of deep learning techniques in long-term load forecasting

Although deep learning approaches are used to solve short-term load forecasting problems, their application for long-term load forecasting is still limited. In this section, up-to-date deep learning approaches for improved long-term load forecasting are investigated to understand the different architecture configurations used in such forecasting approaches. These include but are not limited to the authors of Ref. [34], who developed optimal design settings for Deep Neural Networks (DNNs) to generate accurate predictions of long-term electricity demand. The effectiveness of prediction using the developed DNNs' architecture, compared with traditional ANNs, was higher. In Ref. [35], a Deep-Feedforward Neural Network (Deep-FNN) with a sigmoid transfer function, resilient backpropagation training algorithm, and Deep-FNN with Rectified Linear Unit (ReLU) activation function was proposed for short-term and long-term load forecasting. The outcome demonstrated that Deep-FNN with sigmoid function and resilient back-propagation training algorithm performed better than other models. In Ref. [36], Deep Recurrent Neural Network (DRNN) models were proposed for a medium to long-term electric load prediction. The proposed DRNN's error is relatively lower when compared with the conventional multi-layered perceptron neural network. In Ref. [37], a bespoke DRNN configuration was explored for medium to long-term electricity and heat demand predictions. The DRNN model outperformed both Gradient Boosting Regression (GBR) and SVM approaches in terms of accuracy. Ref. [38] proposed a method for forecasting high energy demand using a Long Short-Term Memory (LSTM) neural network, and Exponential Moving Average (EMA) was proposed. The LSTM model proved its superiority compared with other ANN, SVM, Random Forests, Bayesian Regression, and Mean-Only Model approaches. Ref. [39] proposed a method for long-term load forecasting utilising RNN consisting of Long-Short-Term-Memory (LSTM-RNN) cells was proposed. The proposed model's performance could be further improved by incorporating more weather parameters. In Ref. [40], Feedforward Artificial Neural Network (FANN), SVM), RNN, Generalised Regression Neural Network (GRNN), K-Nearest Neighbours, and Gaussian Process Regression (GPR) were used for long-term load forecasting. The FANN method showed better results compared with others. In Ref. [41], LSTM was developed for long-term energy consumption prediction with distinct periodicity. This model introduced special units called memory blocks to overcome the vanishing/exploding gradient problems. In Ref. [42], different Machine Learning algorithms, including SVR,

Random Forest Regression, and Flexible Neural Tree, were compared for the best forecasting of a short and long-term load demand period. It was observed that the forecasting accuracy decreased with the increase in timescale. In Ref. [43], a new variant of RNNs based on the transform learning model, named Recurrent Transform Learning (RTL), was proposed for long-term load forecasting. Results showed that the proposed technique outperforms all other techniques.

The previous literature indicates that the forecasting efficiency of traditional ANNs for long-term predictions of load demand could only be increased if integrated with other forecasting approaches/algorithms such as ANFIS, SVR, and LR. The accuracy of deep learning approaches in long-term load forecasting requires developing sophisticated architecture configurations and careful selection of the training method, including best tuning of all the involved parameters.

However, improving the BPA algorithm of the traditional ANN and its forecasting capability has not been investigated yet. The difference in behaviours between trained and future datasets and how to encapsulate/quantify these behaviour differences to adapt the current BPA algorithm for better forecasts has not yet been studied. In addition, the modification suggested in the mathematical formulation of the forecasting formula and how to correct the behaviour gained after training the network to be familiar with the unexpected behaviour of other future values of the datasets that might have different behaviour has also not been investigated before. To further clarify the proposed mathematical derivation, an adjustment factor is proposed to capture and quantify any deviation between the trained dataset's behaviours and new/future datasets. This factor will then be added to the forecasting formula to smooth out any positive or negative deviations obtained in the forecasting values due to both trained and future dataset behaviour differences. Different adjustment factors and new forecasting equations for popular dataset ranges, including (0,1) and (-1,0), are proposed to obtain reliable and accurate long-term load forecasts.

It is also worth mentioning that the popular reasons behind the inability of such traditional ANNs to provide accurate long-term forecasts are the lack of training data and the increment of accumulated errors in long-term estimation. However, other reasons for the limitations of traditional ANNs as long-term forecasting require further discussion. Although authors in Ref. [44] addressed some possible reasons for the underperformance of many complex novel ANN architectures, none of these reasons relates to the problem's settings considered in this study.

Therefore, in this study, the reasons/limitations behind the poor performance of traditional ANNs in long-term forecasting will be further investigated and discussed. The

study will also propose an improved ABPA algorithm with a new formulation for reliable long-term forecasting.

3 Multiple-layer perceptron artificial neural networks

This section explains the BPA used in a multiple-layer traditional artificial neuron network. The BPA's limitation in terms of the validity of its optimal weight values for providing accurate long-term forecasts is also discussed. A proposed improvement on the BPA currently used is also introduced. Different forecasting formulations for different dataset ranges are suggested to identify any deviation caused by changed behaviours of trained and future datasets for better forecasting outcomes.

The proposed algorithm's main contribution is that it provides adaptive forecasting models/formulations to encapsulate the behaviour difference between trained and future datasets in terms of deviation. This deviation will be added as an adaptive factor to the current ANN forecasting model outcomes for more accurate long-term forecasts. This adaptive factor's value could be positive or negative, depending on the deviation between current and future datasets' behaviours.

3.1 Backpropagation algorithm (BPA)

This section discusses a Multi-Layer Perceptron (MLP) model. This model consists of j layers, and each layer involves several neurons. The first layer is the input layer, while the last layer is the output layer, and all layers in between are hidden layers, as shown in Fig. 1.

In the forward pass computations, the input variables/dataset represent individual independent variables $X_i, i = 1, ..., m$ (m is the number of independent variables), each variable involves (n) observations. These variables are fed into the network's input layer after normalising them in a specific range. In order to set the importance of each input, a weight $W_{i,k,j}$ is assigned to input index i, layer j-1, to the corresponding neuron index k, and next layer j.

These weights at the first term are selected randomly. For example, for the first hidden layer, the first output X_1 is calculated as below:

$$Y_{1,1} = f(net_{1,1}) \tag{1}$$

The value of $net_{1,1}$ is:

$$net_{1,1} = \sum_{i=1}^{m} X_i . W_{i,1,1} \tag{2}$$

where i is the input/neuron index at layer j-1, $net_{1,1}$ is the output of X_1 for all neurons of the first input layer.

The outputs $net_{k,j-1}$ of neuron k, layer j-1 becomes the inputs to the last layer j.

Each neuron output is normalised using the Activation function (Sigmoid). The final network output (calculated) is normalised using the below Sigmoid function:

$$\hat{Y}_d = \frac{1}{1 + e^{-(\sum_k net_{k,j} - \theta_j)}} \tag{3}$$

where k is the neuron index of the following corresponding layer j, $net_{k,j}$ is the input to the last hidden layer j, neuron k, $W_{i,kj}$ is the weight on the connection from neuron i to k at layer j, θ_j is the Threshold (bias) on layer j, $\hat{Y}_d$ is the desired/calculated output.

The validation stage is then applied to check the difference between the actual/calculated and desired/target outputs of the network.

The difference between the desired output Y_a(target) and the actual network's output $\hat{Y}_d$ (calculated) is calculated. An error function defined as Mean Square Error (MSE) is usually used as a typical error function.

$$MSE = \sum_d \frac{(Y_a - \hat{Y}_d)^2}{n}, d = 1, ..., n; a \tag{4}$$

where d is the output value index, n is the number of input values/observations. Y_a desired/target output (observations used in training).

The backpropagation computations involve three stages: *Training*, *Validation/Test,* and *Forecasting*. The dataset is divided into three parts according to these stages; the longest dataset's part is used for training, the second part to validate the model accuracy, and the third for model testing. The steps of backpropagation are summarised through the *Training* stage as follows: for a specific learning rate (η), the weights are updated using a learning algorithm (gradient descent). The weights are adjusted anywhere in the network after the error becomes known. The derivative of the error function for that weight must be found, and the delta rule Δw is generalised.

As shown in Eq. 3, the nonlinearity sigmoid function is the most used normalisation function in the multi-layered perceptron models [45]. Inadequate or inappropriate data normalisation to input variables may considerably worsen forecasting results [13]. The derivative being $\frac{dY}{dX} = Y = \hat{Y}(1 - \hat{Y})$ and the delta rule used to change the weight W_i in neuron k, hidden layer j with a sigmoid function is as follows:

$$\Delta W_{i,j,k} = \eta X_i \delta_{k,j} \tag{5}$$

where:

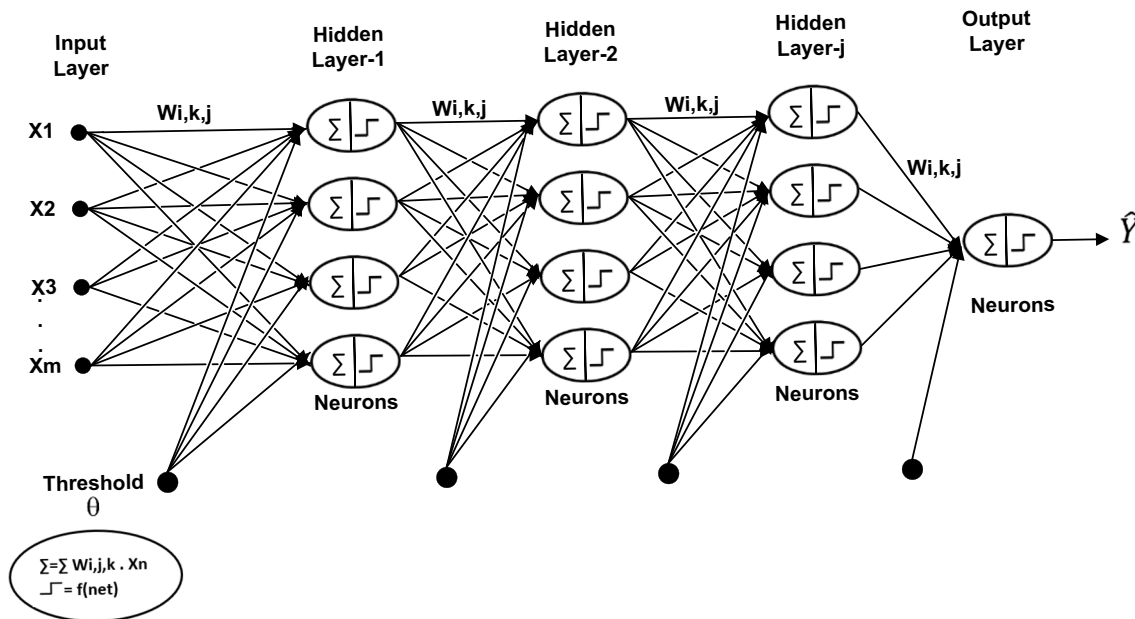$$\delta_{k,j} = Y_a(1 - Y_a)(\hat{Y}_d - Y_a) \tag{6}$$

Fig. 1 The architecture of the MLP model

The weights are then updated after each pattern is implemented by:

$$W_{i,k,j} \text{ update} = W_{i,k,j} - \Delta W_{i,j,k} \tag{7}$$

The newly updated weights will then be fed into the network for another training session. Each training session involves attempting different numbers of neurons and hidden layers to make model improvements and increase output accuracy. The training session stops when the MSE is minimal. The MLP is considered the best model and ready to be tested by exposing it and comparing it to the data's test part.

After the training stage is completed, the model is ready to forecast future values using the below equation:

$$\hat{Y}_{\text{Forecast}} = (\hat{Y}_d + 1)(\text{Max}(Y_a) - \text{Min}(Y_a)) + \text{Min}(Y_a) \tag{8}$$

where Y_a: desired output (within the training data set), (we enter this in the training session), Min (Y_a) is the minimum desired/target output, Max (Y_a) is the maximum desired/target output,

The forecasting model/formulation (8) is used to restore the data to their actual values. The BPA uses the forecasting formulation after the training and testing stages have been completed. This paper will improve this formulation by including the deviation/accumulated error caused by the behaviour difference between trained and future datasets for more accurate long-term forecasting outputs.

3.2 Memory of learning and long terms forecasting

Neural networks generally perform three functions. They supervise learning tasks, build knowledge, and memorise the correct answer to provide it in advance from datasets. As explained in Sect. 3, the networks then learn by tuning themselves to find the correct answer, increasing their prediction's accuracy. Also, they understand the behaviour of datasets and mimic their interchanging relationships, taking into consideration factors that could affect their behaviour and stability for further predictions.

However, the training stage is conducted on the available numerical information in input datasets. At the training stage, the behaviour of datasets is captured, including inputs of independent variables $X_i, i = 1, ..., m$ (m is the number of independent variables), each with (n) observations, assuming that some aspects of the past pattern will continue in the future. Hence, a memory of training is established. This memory enables backpropagation to predict datasets' future behaviour based on the current learned behaviour obtained from the training stage and supported by the validation/test stage. This training includes estimating and hence capturing the behaviour of each dataset of input X_i as it continues in the future. Although the bias term allows for some deviation/random variation and the effects of relevant variables that are not included in the model, the behaviour of future inputs X_i beyond the trained datasets could irritate and become unexpected due to unforeseen factors. This irritation could affect the validity of the optimal weights, make the

forecasting of the behaviour of future observations challenging, and subsequently elevate the bias factor further. Many researchers attribute the popular reason for such poor performance to a lack of training data and increased accumulated errors over long-term estimation.

We discuss in this paper other reasons related to the poor performance of the traditional BPA in long-term forecasting. One of these reasons is that optimal weights obtained after the training stage has been completed might become less sustainable in the long-term than short-term forecasting. This suitability issue escalates notably when future datasets' behaviour deviates from the trained datasets' current behaviour theme. Hence, these weights become less valid/sustainable, affecting the network learning memory and response to prediction decisions. If so, the memory of learning obtained at the training stage starts distracting in providing the correct forecast answer, resulting in incorrect responses. This inaccuracy of responses starts accumulating over time, causing "non-realistic" or "false" prediction decisions and ending with a significant deviation/bias value. The ANN model memory tends to vanish for longer-term forecasting resulting in a weak and unfit prediction model.

3.3 Adaptive backpropagation algorithm (ABPA)

This section discusses suggestions for improving forecast values' accuracy caused by learning memory disruption/deterioration. This deterioration of memory is attributed to the difference between the two behaviour practices of both training and future datasets. As explained in Sect. 3.2., this leads to wrong decisions to provide overestimated/underestimated forecasting outputs.

The difference between the behaviour of actual datasets (used to train the network) and any deviated behaviour caused by the new/future datasets used for forecasting is captured in an adjustment factor. This factor is calculated by finding the difference between the two behaviours. This difference is identified by adopting the last trained observation of the network datasets as a base of knowledge. This base represents the source of the datasets' behaviour that the ANN has captured during the training stage. Any deviation from the network will be captured by an adjustment factor $Max\left(Y_a^{adj}\right)$ or $Min\left(Y_a^{adj}\right)$ and will be added to or subtracted from the forecasting formulation/model, Eq. (8) based on the deviation type (positive or negative). A positive deviation $\Delta Max(Y_a)$ means that the generated values are overestimated forecasts, while a negative deviation $\Delta Min(Y_a)$ means that the provided forecasts are underestimated. Both deviations are caused by the different behaviour themes of future datasets that the trained network does not understand. The forecasting after adjustments/

adaptation should produce accurate forecast values considering any deviation/difference between the captured behaviour (after training) and other behaviours inherited by future dataset values.

This improvement includes adjustment of the network desired/target output Y_a. The adjustment could be made on either the Maximum Value of Y_a, Minimum Value of Y_a or both Maximum and Minimum Values of Y_a.

For the range of the dataset (-1,0), the adapted forecasting equation is developed as follows:

- Calculate the adjusted values of $Max(Y_a)$ and $Min(Y_a)$ using equations:

$$Max\left(Y_a^{adj}\right) = Max(Y_a) + \Delta Max(Y_a) \text{if } (\Delta Max(Y_a) \text{is} + \text{ve}), \text{or} - \Delta Max(Y_a)(\text{if } \Delta Max(Y_a) \text{is} - \text{ve})$$

$$\begin{aligned} Min\left(Y_a^{adj}\right) &= Min(Y_a) \\ &= Min(Y_a) + \Delta Min(Y_a) \text{if } (\Delta Min(Y_a) \text{is} + \text{ve}), \text{or} - \Delta Min(Y_a)(\text{if } \Delta Min(Y_a) \text{is} - \text{ve}) \end{aligned}$$

$$\Delta = Max_{\forall i}\left(\frac{X_{i,n+l} - X_{i,n}}{X_{i,n}}\right), l = 1, \ldots, L$$

$$\Delta = Min_{\forall i}\left(\frac{X_{i,n+l} - X_{i,n}}{X_{i,n}}\right), l = 1, \ldots, L$$

where: Y_a^{adj}: is the adjusted value of Y_a, Δ: is the variation between $X_{i,n}$ and $X_{i,n+l}$, n: is the index of last input/observation, l: is lag of forecasting, L: is the number of forecasting lags, $X_{i,n}$: the last input n of each dataset i, $X_{i,n}$, $X_{i,n+l}$: the forecasted input n + 1 of each input value X_i.

This adapted equation will adjust either $Max(Y_a)$ and/or $Min(Y_a)$ values to keep the bridge of the behavioural theme between the new datasets used for forecasting $X_{i,n+1}, X_{i,n+2}, \ldots, X_{i,n+L}$ for all i = 1,…,m and the last observation of the trained datasets of the independent variables $X_{i,n}$. This sort of behavioural bridging will subsequently improve the forecasting value if the learning memory starts distracting because of the confusion caused by the behaviour of future values of the input variables beyond the overall captured behaviour.

- In case adjustment of $Max(Y_a)$ is applied, the adapted forecasting equation is:

$$\hat{Y}_{Forecast} = \left(\hat{Y}_d + 1\right)\left(Max\left(Y_a^{adj}\right) - Min\left(Y_a\right)\right) + Min\left(Y_a\right) \tag{8.1}$$

- In case $Min(Y_a)$ is suggested for adjustment, the adapted forecasting equation is:

$$\hat{Y}_{\text{Forecast}} = (\hat{Y}_d + 1)(\text{Max}\,(Y_a) - \text{Min}\,(Y_a^{\text{adj}})) \\ + \text{Min}\,(Y_a^{\text{adj}}) \tag{8.2}$$

- In case both $\text{Min}(Y_a)$ and $\text{Min}(Y_a)$ are adjusted, the adapted forecasting equation is:

$$\hat{Y}_{\text{Forecast}} = (\hat{Y}_d + 1)(\text{Max}(Y_a^{\text{adj}}) - \text{Min}\,(Y_a^{\text{adj}})) \\ + \text{Min}\,(Y_a^{\text{adj}}) \tag{8.3}$$

Any of the above models/formulations could be applied to accommodate forecast value deviations caused by input datasets' behaviour (future). The quality of the forecasting output depends on the deviation in behaviour. Hence, a comparative study is required to decide which of the above forecasting models/formulations best adjust forecasting outputs to comply with the datasets' behaviour and provide the best forecasting outcomes.

This adjustment could also be applied to datasets with a range (0,1). The adapted forecasting equations are:

- In case adjustment of $\text{Max}(Y_a)$ is applied, the adapted forecasting equation is:

$$\hat{Y}_{\text{Forecast}} = ((\hat{Y}_d)(\text{Max}(Y_a^{\text{adj}}) - \text{Min}\,(Y_a))) + \text{Min}\,(Y_a) \tag{8.4}$$

- In case $\text{Min}(Y_a)$ is suggested for adjustment, the adapted forecasting equation is:

$$\hat{Y}_{\text{Forecast}} = ((\hat{Y}_d)(\text{Max}\,(Y_a) - \text{Min}\,(Y_a^{\text{adj}}))) \\ + \text{Min}\,(Y_a^{\text{adj}}) \tag{8.5}$$

- In case both $\text{Min}(Y_a)$ and $\text{Min}(Y_a)$ are adjusted, the adapted forecasting equation is:

$$\hat{Y}_{\text{Forecast}} = ((\hat{Y}_d)(\text{Max}(Y_a^{\text{adj}}) - \text{Min}\,(Y_a^{\text{adj}}))) \\ + \text{Min}\,(Y_a^{\text{adj}}) \tag{8.6}$$

4 Case study, results analysis, and discussion

4.1 Case study and ANN inputs

A case study is conducted to test the proposed forecasting formulations in Sect. 3.3. This case study considers the electricity load demand (monthly peak load) in the Republic of Iraq for 2011–2020 and electricity consumption in multiple sectors, including domestic (households), industrial, commercial, government, and agriculture. The electricity load demand of both Nov and Dec 2020 was supposed to be released in early 2021. However, this was

delayed due to the third wave of COVID-19 in the Republic of Iraq.

We considered six input variables to forecast electricity's load demand, each representing annual energy consumption by a sector/industry. The developed model also considers electricity cut periods, represented by an unsatisfied demand factor. This unsatisfied demand, leading to unplanned cuts, frequently occurs due to the many factors that have caused the current national electricity system in Iraq to become outdated. These factors include frequent wars and political interventions, lack of maintenance, unstable security conditions, and a market monopolised by private suppliers (mainly using diesel generators), all of which have led to an inability to satisfy the increasing electricity demand. This shortage is referred to as a cut period of electricity, most commonly unscheduled. Hence, this factor is vital in forecasting electricity consumption in Iraq and is considered in this study. The input and output parameters and respective symbols are summarised in Table 1.

The electricity load demand (monthly peak load) in the Republic of Iraq for 2011–2020 is considered as an output parameter, while electricity consumption in multiple sectors, including domestic (households), industrial, commercial, governmental, and agricultural, are considered as parameter inputs of the ANN. See Fig. 2 for the monthly electricity consumption rate in Iraq, including the load demand.

As seen in Fig. 2, the domestic sector, represented by households, consumes a significant portion of electricity. The second largest consumer of electricity is the industry, covering both private and public sectors. This consumption includes refineries, gas stations, and communication companies. Governmental consumption comes third. It can also be noticed that the industrial sector's energy consumption decreases after 2017 due to many political reasons, such as increased imports from neighbouring countries. Moreover, the governmental sector indicates high electricity consumption, increasing in 2011. This high consumption is attributed to the increasing number of government premises, employees and offices.

Table 1 Description of the input datasets and output parameter

Parameter	Description
Input	D: Domestic (MWh)
Output	C: Commercial (MWh)
	I: Industrial (MWh)
	G: Governmental (MWh)
	A: Agricultural (MWh)
	L: Load demand of electricity (MW)

Fig. 2 Load demand and consumption of electricity in Iraq
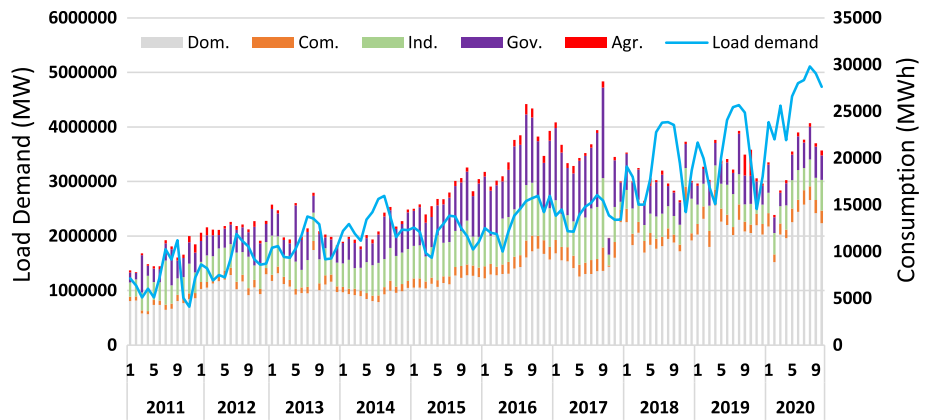
The load demand of electricity (MW) depends on the above sectors' energy consumption (MWh). This load demand (MW) increases due to increasing domestic sector requirements. This increment in the domestic sector requirements is attributed to changes in population growth and in economic and demographic patterns in Iraq.

The ANN is trained using input datasets (monthly) from 2011 to 2019. These datasets consist of load demand as the dependent variable and five other independent variables, including domestic, commercial, industrial, governmental, and agricultural consumption of electricity. A sample of the dataset used to train the ANN model is presented in Table 2.

The complete dataset inputs and the load demand as output are presented graphically in Fig. 2. Another set of inputs of the year 2020 from the months (Jan-Oct) are used to validate and evaluate the proposed ABPA performance. A sample of this testing data is presented in Table 3.

4.2 Experimental results

4.2.1 ANN architecture and training parameters

After the ANN has been trained, the optimal weight values are obtained and the optimal number of neurons and layers. The testing and forecasting stages then follow by adopting input datasets for the year 2020 from Jan to Oct for comparison and validation purposes. The architecture of five inputs, one hidden layer with five neurons and one output $(5 \times 5 \times 1)$, is the most suitable for the proposed ABPA. The proposed ANN architecture's performance is assessed using the MSE and MAPE criteria. See Table 4 for the optimal parameters used for ANN architecture and training.

Table 2 Sample training data set used for learning the ANN

Year	Month	D	C	I	G	A	L
2011	Jan	807,810	80,299	258,606	182,512	40,304	7142
	Dec	859,547	92,394	378,707	366,890	148,163	7222
2012	Jan	1,029,478	127,508	367,460	382,630	152,714	8625
	Dec	1,298,710	113,361	476,621	331,336	55,479	8708
2013	Jan	1,175,507	114,461	717,176	488,478	80,361	10,417
	Dec	972,583	105,221	431,702	287,158	31,213	10,500
2014	Jan	969,578	98,106	430,423	370,384	25,874	12,208
	Dec	1,046,479	127,509	589,467	671,879	51,525	12,292
2017	Jan	1,679,993	246,388	729,516	1,323,604	104,943	13,833
	Dec	2,255,290	222,741	149,700	353,834	19,539	13,417
2018	Jan	2,255,290	244,726	346,517	660,315	21,875	19,098
	Dec	1,917,974	130,057	633,546	298,107	29,103	18,696
2019	Jan	2,024,190	194,155	357,444	342,794	44,468	21,669
	Dec	2,024,190	194,155	357,444	342,794	44,468	18,014

Table 3 Sample testing data set used to evaluate the optimised ANN

Year	Month	D	C	I	G	A	L
2020	Jan	2,171,109	247,153	376,323	513,709	42,303	23,849
	Feb	1,521,127	142,113	389,312	292,361	40,375	21,988
	Mar	2,028,360	200,519	318,475	243,270	42,976	25,615
	Oct	2,232,481	235,083	561,184	449,513	88,824	27,615

Table 4 The architecture of ANN and training parameters

Architecture of ANN	
Number of layers	3
Number of neurons in the layer	5
Number of inputs	5
Number of hidden layers	1
Number of outputs	1
Initial weights and biases	Randomisation
Activation function	Sigmoid
Number of training data	108
Number of testing data	10
Training parameters	
Learning rule	Adaptive backpropagation (ABPA)
Learning rate	0.02
Momentum constant	1

4.2.2 Impact of input normalisation on the model's output

Having trained the ANN, obtained optimal parameters, and tested and validated the outputs as outlined above, normalisation of the data is applied.

Applying normalisation to the input datasets is then investigated to examine its effect on the weights'

calibration during the training stage. This assists in identifying the best range of input datasets for the best estimation of weights across the ANN and, subsequently, for high-quality forecasts. See Fig. 3 for a comparison of datasets before and after normalisation is applied.

In Fig. 3, the actual load demand datasets of the year 2020 starting from Jan to Oct are used to test the accuracy of long-term forecasts using Eq. 8, where adjustment is not applied (see Sect. 3.3). The normalisation is applied to electricity load and consumption datasets by converting them between (0,1) ranges. The normalisation of this range provides, as shown in Fig. 3, more accurate long-term forecasts, justified by its minimum MSE, 47.497.616 and MAPE, 0.24, than the dataset's range ($\pm\infty$) with MSE equal to 50.759.956 and MAPE equal to 0.26. This (0,1) range normalisation also reflects a well-training and best learning practice obtained after normalisation of range (0,1) is applied to the datasets. However, the datasets' range (−1,0) provides fluctuated forecasts and does not converge the actual load demand data, thereby achieving the highest MSE equal to 77.857.125 and MAPE equal to 0.287.

The traditional forecast model/formulation used in the BPA (Eq. 8) is improved by adding/subtracting the deviation represented by the adaptive factor to/from the forecast values obtained using Eq. 8 (the adjustment is applied).

Fig. 3 The evaluation of impacts of different ranges of input datasets
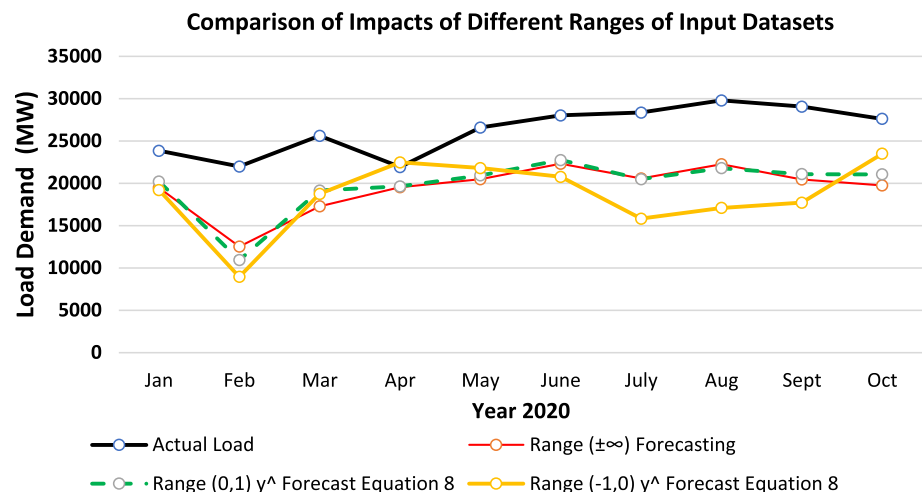
This adaptive factor equals the deviation between long-term forecasts and the behaviour theme the network becomes familiar with after the training stage has been completed.

For the range of the dataset between $(-1,0)$, formulations (8.1) to (8.3) are applied to improve the long-term forecast values obtained by Eq. (8). The dataset range $(0,1)$ shows more accurate forecasts using Eq. (8) (see Fig. 3 above). However, the purpose of the range $(-1,0)$ experiment is to justify the proposed formulations (8.1) to (8.3) by understanding the general theme of its behaviour, even though they are not the best forecast outcomes. See Fig. 4 for the impact of the proposed forecasting formulations (8.1) to (8.3), the dataset range $(-1,0)$ on the quality of forecasting outcomes.

Figure 4 shows that the best forecasts are obtained by applying Eq. (8.2), achieving minimum values of MSE 13.884.386 and MAPE 0.16 compared with the No-Adjustment Eq. (8), that scored MSE and MAPE equal to 77.857.125 and 0.287. Equation (8) generates forecasts that are below the forecasting theme, and hence $\text{Min}(Y_a)$ is suggested for adjustment of the $\hat{Y}_{\text{Forecast}}$ values. Equation (8.1) generates the worst forecasts with the highest MSE and MAPE values equal 105.350.141 and 0.48, respectively. However, Eq. (8.3) provides the second-best forecasts compared with the actual load data to achieve MSE and MAPE equal to 37.118.399 and 0.30, respectively.

The impact of the proposed forecasting adjustments for input datasets range $(0,1)$ is identified, and each adjustment is represented by the suggested forecasting formulations (8.4) to (8.6). See Fig. 5 for the impact of the proposed adjustments on forecasts' quality.

As presented in Fig. 5, the best forecasting of load demand is achieved by using the proposed formulation (8.4) for the best datasets of the $(0,1)$ range (the adjustment is applied). This best forecasting is attributed to having the minimum MAPE 0.045 and MSE 1.195.650. This adjustment is applied by adding the adjustment factor to the obtained forecasting values to improve long-term load forecasting further. The worst forecasts are generated using formulation (8), where no adjustment is applied to produce an MSE value equal to 47.497.616 and MAPE equal to 0.24. The second-best forecasts are generated using Eq. (8.6), where adjustment is applied to produce MSE equal to 1.850.110 and MAPE equal to 0.063, followed by forecasts generated by Eq. (8.5) with MSE and MAPE equal to 3.147.928 and 0.089, respectively.

In general, it can be concluded that the best forecasting practice of the load dataset inputs of the year 2020 over the long-term is achieved for input datasets range $(0,1)$ and by applying Eq. (8.4). These settings could change according to the type and number of the data and other learning requirements.

5 Comparison study

A comparison study is conducted to justify the proposed forecasting method's superiority in generating high-quality long-term forecasts. Five forecasting approaches were used to predict electricity's load demand from Jan to Oct 2020. These approaches include:

1. The traditional Backpropagation Algorithm (BPA)
2. The Adjusted Backpropagation Algorithm (ABPA)
3. Radial Basis Function Networks (RBFN)

Fig. 4 The evaluation of the impact of the proposed forecast adjustments – datasets range $(-1,0)$

Comparison of Impacts of the Proposed Forecasting Adjustments - Datasets Range (-1,0)
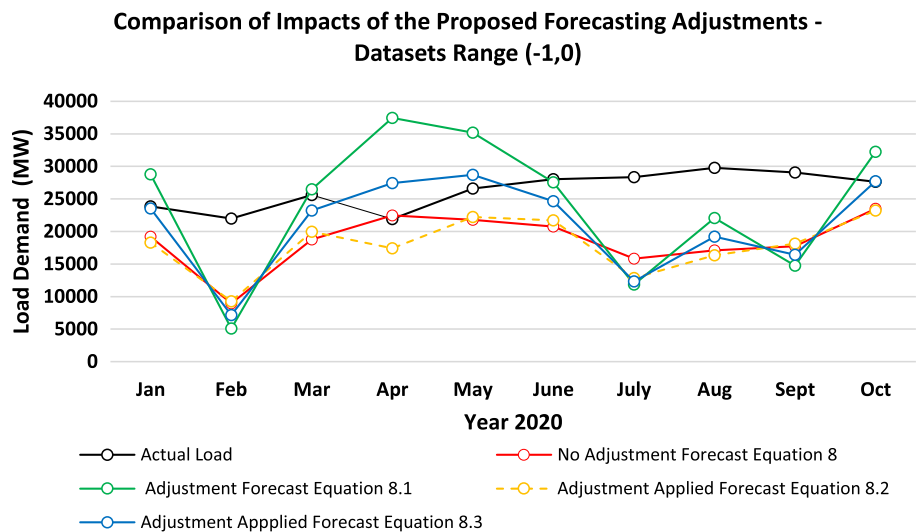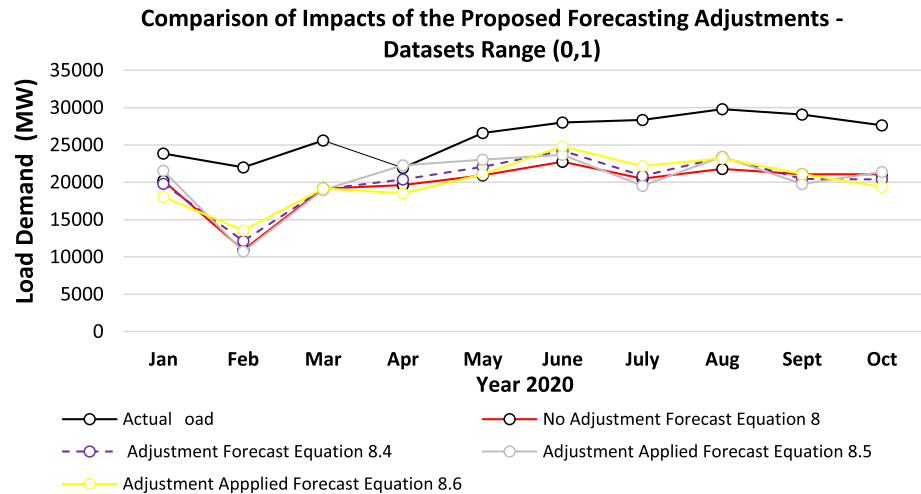
Fig. 5 The evaluation of the impact of the proposed forecast adjustments – datasets range (0,1)

4. Regression Analysis (REG)
5. Redcurrant Neural Networks – Long Short-Term Memory (RNNs-LSTM)

REG is used to forecast the load demand. The Least-Squares (LS) method is applied to estimate the regression model's parameters. After selecting the most influential independent variables, the model becomes:

$$L = 0.0099486D + 0.004962I + 0.002712G - 0.026445Cut$$

where L load demand, D: Domestic consumption, I: Industrial consumption, G: Governmental consumption, Cut: Unsatisfied demand, Cut (shortage of electricity).

After solving the above regression model, T test values for all parameters are zero, P = 0, which means that these parameters are statistically significant. The coefficient of determination R^2 is also calculated: $R^2 = 98\%$. This significant value refers to the strong relationships between the regression model variables. The F-test value is calculated as 574.3, which indicates that the regression model is a close fit to the actual one. However, the REG model is not accurate enough for long-term prediction due to the non-linearity relationship between the load demand dependent variable and other electricity consumption independent variables.

The RNNs-LSTM architecture is selected and used in this comparison study. This selection is attributed to its high efficiency in long-term forecasting [46]. In addition, such network architectures have a recurrent hidden state whose activation at each time is dependent on that of the previous time, which makes it specialised for processing sequential data (time series).

After the RNN has been trained, the best architecture, along with other relevant RNNs parameters, is found to be: number of inputs (5), number of layers (4), number of neurons (5), number of outputs (1), number of hidden layers (2), input delays (5), layer delays (1), learning rate (0.005), learn rate (Gradient), layer type (Dense, LSTM), layer activation (ReLU, Linear), data division (Random), training (Bayesian Regulation), sum square parameter (21.5), Calculations (MEX), epochs (134). See Table 5 for a comparison of load demand forecasting approaches.

The overall trend of the load demand curve for the trained datasets and long-term forecasting values obtained by each forecasting method is demonstrated in Fig. 6.

Figure 6 shows that the best long-term forecasts of load demand are obtained using the proposed ABPA. Based on forecasting value adjustments, the suggested forecasting formulations contributed significantly to amending the deviations obtained from the difference in behaviour between the trained and new forecast datasets. The long-term forecast values are further improved. This improvement was assessed considering the minimum MSE and MAPE equal to 1.195.650 and 0.045, respectively, compared with the previous version of forecasts using No Adjustment, i.e. Eq. 8, which achieved significantly higher MSE and MAPE values equal to 50.759.957 and 0.26, respectively.

The second-best forecast method based on the achieved MSE and MAPE values of 12.845.733 and 0.116 is the RNNs-LSTM. This method's performance, as discussed earlier, is attributed to its high capability for learning order dependence in sequence prediction problems. The RBFN method is considered the third-best forecasting method by achieving MSE and MAPE values equal to 19.197.960 and 0.144. The REG approach provides forecasts below the

Table 5 Load forecasting of four methods with actual data

Year 2020	Actual load	REG	BPA	ABPA	RBFN	RNNs-LSTM
Jan	23,849	18,324	19,484	19,784	25,344	20,924
Feb	21,988	14,100	12,510	12,138	23,212	16,126
Mar	25,615	17,117	17,276	19,080	17,812	19,858
Apr	21,923	18,707	19,529	20,394	15,960	23,365
May	26,592	19,653	20,494	22,047	21,271	25,437
June	28,015	22,830	22,319	24,235	25,539	28,868
July	28,354	21,903	20,558	20,847	27,043	23,375
Aug	29,792	23,829	22,243	23,326	27,329	27,746
Sept	29,060	21,355	20,445	20,436	26,513	24,679
Oct	27,615	19,997	19,753	20,376	21,052	27,327
MAPE	–	0.25	0.26	0.045	0.144	0.116
MSE	–	44.495.892	50.759.957	1.195.650	19.197.960	12.845.733

Fig. 6 The comparison of accurate forecasting of multiple approaches
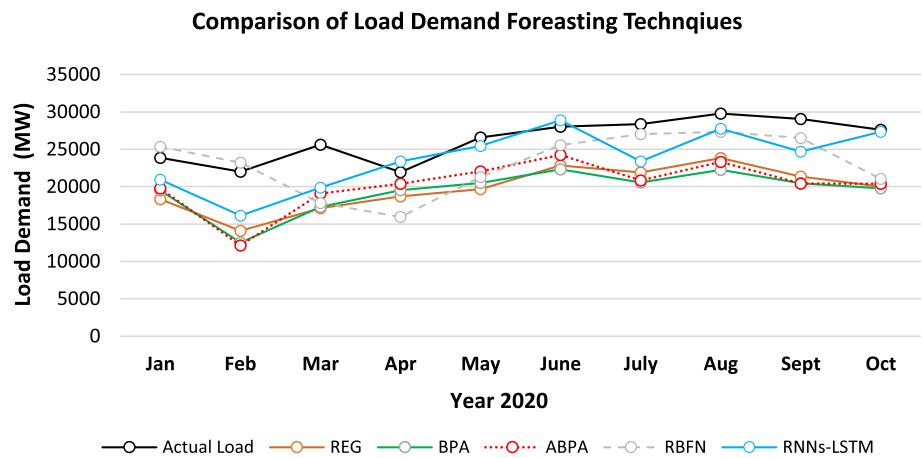
actual load and shows performance almost close to BPA with MSE and MAPE values equal to 44.495.892 and 0.25.

6 Conclusion and future work

This study proposed improving the classical forecasting formulation used in BPA for higher-quality long-term load demand forecasting. This improvement was achieved by capturing the behaviour differences within/outside the training input datasets and quantifying these differences as adjustment factors. This adjustment was possible by bridging the last observation's behaviour (within the trained datasets) and future datasets outside the training stage. This deviation (adjustment factor) was then quantified/encapsulated in a variation term, which was added to/subtracted from the classical forecasting model for best adjustments and higher accuracy outputs.

The new forecasting formulations, including the proposed adjustment factors, improved long-term forecasts' quality. This improvement was evident from the quality of load forecast values obtained using the ABPA compared with other forecasting methods. Compared with other methods, the APBA obtained the lowest values of MSE equal to 1.195.650 and MAPE equal to 0.045. The reduction in both MSE and MAPE values showed that the proposed ABPA, represented by the new forecasting formulations, successfully accommodated the deviation caused while generating high accuracy forecast outputs. The achieved high forecasting accuracy also verified the

traditional ANN techniques and the proposed ABPA's ability to successfully capture the unexpected behaviour of the year 2020, which was exceptionally different from other previous years due to the COVID-19 outbreak. The traditional BPA could not understand the significant change in the load dataset behaviour, especially for 2020. This distracted understanding (as explained in Sect. 3.2) led the BPA to generate inaccurate forecasts below the actual load inputs, resulting in significant MSE and MAPE values equal to 50.759.957 and 0.26, respectively. The RNNs-LSTM was the second-best approach in providing long-term forecast with MSE and MAPE equal to 12.845.733 and 0.116, followed by the RBFN approach with values of MSE and MAPE equal to 19.197.960 and 0.116, respectively.

The impact of data normalisation on the quality of long-term forecasting outcomes using the traditional forecasting equation was measured by applying three different ranges of input datasets. Using Eq. (8), the range (0,1) led to the best forecasts with minimum values of MSE equal to 47.497.616 and MAPE equal to 0.24. After the adjustment, Eq. (8.4) provided the best forecasts for the dataset range (0,1) with the lowest MSE equal to 1.195.650 and MAPE equal to 0.045.

This approach's limitation in terms of implementation is that it needs significant differences in behaviour between the training and forecasting/validation datasets (high non-stationarity). This dataset's behaviour differences enable the proposed ABPA to perform better and improve the expectations based on adjusting the dataset's behaviour.

Future work is suggested to develop more advanced forecasting models with more than one forecasting output for classical ANNs. The relationship between these models for different ANN outputs would be a promising opportunity to further investigate robust and reliable prediction purposes.

Declarations

Conflict of interest The authors confirm no conflict of interest with this submission, an original work.

References

1. Hahn H, Meyer-Nieberg S, Pickl SW (2009) Electric load forecasting methods: tools for decision making. Eur J Oper Res 199(3):902–907

2. Khashei M, Bijari M (2011) A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Appl Soft Comput 11(2):2664–2675

3. Mohammed NA (2018) Modelling of unsuppressed electrical demand forecasting in Iraq for long term. Energy 162:354–363

4. Majumdar S, Shukla PR (2007) Electric load forecasting using artificial neural networks and deficit management. Energy Sources 19(8):771–782

5. Ringwood JV, Bofelli D, Murray FT (2011) Forecasting electricity demand on short, medium and long time scales using neural networks. J Intell Rob Syst 31:129–147

6. Mihai C, Helerea E (2017) (2017) Load forecasting using artificial neural networks for the industrial consumer. International Conference on Modern Power Systems (MPS) 6–9:1–6

7. Baek S-M (2019) Mid-term load pattern forecasting with recurrent artificial neural network. IEEE Access 7:172830–172838

8. Mamun MA & Nagasaka K (2004) Artificial neural networks applied to long-term electricity demand forecasting. Fourth International Conference on Hybrid Intelligent Systems (HIS'04). IEEE; 204–209.

9. Ghods L, Kalantar M (2011) Different methods of long-term electric load demand forecasting a comprehensive review. Iran J Electr Electr Eng 7(4):249–259

10. Zheng S, Yue Y & Hobbs J Generating (2016) Long-term trajectories using deep hierarchical networks. Advances in Neural Information Processing Systems. 1543–1551.

11. Carpinteiro OAS, Lima I, Moreira EM, Pinheiro CAM, Seraphim E, Pinto JVL (2009) A hierarchical hybrid neural model with time integrators in long-term load forecasting. Neural Comput Appl 18:2960–2965

12. Lalis J.T., & Maravillas E. (2014) Dynamic forecasting of electric load consumption using the adaptive multi-layer perceptron(AMLP). International Conference on Humanoid, Nanotechnology, Information Technology, Communication, and Control, Environment, and Management (HNICEM). 11 1–7

13. Zhang GP, Kline DM (2007) Quarterly time-series forecasting with neural networks. IEEE Trans Neural Net 18:1800–1814

14. Faraji J, Hashemi-Dezaki H, Ketabi A (2020) Multi-year load growth-based optimal planning of grid-connected microgrid considering long-term load demand forecasting: A case study of Tehran, Iran. Sustainable Energy Technologies and Assessments. Volume 42, December 2020, 100827

15. Ammar, N., Sulaiman, M., Fateh, A. & Mohamad Nor, A. F. (2018) Long-Term Load Forecasting of Power Systems Using Artificial Neural Network. Journal of Engineering and Applied Sciences. 13.

16. Solyali D (2020) A comparative analysis of machine learning approaches for short-/long-term electricity load forecasting in cyprus. Sustainability 12:3612

17. Nalcaci G, Özmen A, Weber G-W (2019) Long-term load forecasting: models based on MARS, ANN and LR methods. CEJOR 27(4):1033–1049

18. Achant R (2012) Long term electric load forecasting using neural networks and support vector machines. Int J Comput Sci Technol 3(1):266–269

19. Oğcu G, Demirel OF, Zaim S (2012) Forecasting electricity consumption with neural networks and support vector regression. Procedia Soc Behav Sci 58:1576–1585

20. Jahan IS, Snasel V, Misak S (2020) Intelligent systems for power load forecasting: a study review. Energies 13:6105

21. Kazemzadeh M-R, Amjadian A., Amraee T. (2020). A hybrid data mining-driven algorithm for long-term electric peak load and energy demand forecasting. Energy. Volume 204, 1 August 2020, 117948

22. Ahmad T, Chen H (2018) Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment. Energy 160(1):1008–1020

23. Nezzar R.M., Farah N., Khadir M.T., Chouireb L. (2016) Mid-Long Terms Load Forecasting using Multi-Model Artificial

Neural Networks. International Journal on Electrical Engineering and Informatics. 8(2).

24. Nalcaci G, Özmen A, Weber GW (2019) Long-term load forecasting: models based on MARS, ANN and LR methods. CEJOR 27:1033–1049

25. Nguyen HH, Chan CW (2004) Multiple neural networks for a long-term time series forecast. Neural Comput Appl 13:90–98

26. Alirezaei HR, Salami A, Mohammadinodoushan M (2017) A study of hybrid data selection method for a wavelet SVR mid-term load forecasting model. Neural Comput Appl 31:2131–2141

27. Almazrouee AI, Almeshal AM, Almutairi AS, Alenezi MR, Alhajeri SN (2020) Long-term forecasting of electrical loads in kuwait using prophet and holt-winters model. Appl Sci 10(16):562

28. Melodi A.O., Adeniyi S.T., Oluwaniyi R.H. (2017) Long Term Load Forecasting for Nigeria's Electric Power Grid Using Ann And Fuzzy Logic Models. IEEE 3rd International Conference on Electro-Technology for National Development (NIGERCON). 962–968.

29. Daneshi H., Shahidehpour M., Choobbari A.L. (2008) Long-term load forecasting in the electricity market. IEEE International Conference on Electro/Information Technology, 2008.

30. Adhiswara R., Abdullah A.G., Mulyadi Y. (2019) Long-term electrical consumption forecasting using Artificial Neural Network (ANN). Journal of Physics: Conference Series, 1402 (2019) 033081.

31. Yu H. and Zhang Q. (2010) Application of variable structure artificial neural network for mid-long term load forecasting. 2nd IEEE International Conference on Information Management and Engineering.

32. Masoumi A., Jabari F., Zadeh S.G. Mohammadi-Ivatloo B. (2020) Long-Term Load Forecasting Approach Using Dynamic Feed-Forward Back-Propagation Artificial Neural Network. Optimization of Power System Problems; 233–257

33. Badran S. & Abouelatta O. (2012) Neural Network Integrated with Regression Methods to Forecast Electrical Load. International Conference on Electrical, Electronics and Biomedical Engineering (ICEEBE'2012) Penang (Malaysia) May 19–20, 2012

34. Hamedmoghadam, H., Joorabloo, N. & Jalili, M. Australia's (2018) Long-term electricity demand forecasting using deep neural networks. arXiv preprint

35. Nyandwi A., Kumar D. (2020) Neural Network Approach to Short and Long Term Load Forecasting Using Weather Conditioning. International Conference on Electrical and Electronics Engineering (ICE3). 02:.258–263.

36. Rahman A, Srikumar V, Smith AD (2018) Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Appl Energy 212:372–385

37. Taheri S., Jooshaki M., Moeini-Aghtaie M. (2021). Long-term planning of integrated local energy systems using deep learning algorithms. International Journal of Electrical Power & Energy Systems. 129, July 2021, 106855

38. Jiménez JM, Stokes L, Moss C, Yang Q, Livina VN (2020) Modelling energy demand response using long short-term memory neural networks. Energ Effi 13:1263–1280

39. Agrawal R.K., Muchahary F., Tripathi M.M. (2018). Long-term load forecasting with hourly predictions based on long-short-term-memory networks. 018 IEEE Texas Power and Energy Conference (TPEC).

40. Sangrody H., Zhou N., Tutun S., Khorramdel B., Motalleb M., Sarailoo M. (2018). Long Term Forecasting using Machine Learning Methods. IEEE Power and Energy Conference at Illinois (PECI). 2018.

41. Wang JQ, Du Y, Wang J (2020) LSTM based long-term energy consumption prediction with periodicity. Energy 197:117197

42. Vantuch T., Vidal A.G., Ramallo-González A.P., Skarmeta A.F., Misák S. (2018). Machine learning based electric load forecasting for short and long-term period. 2018 IEEE 4th World Forum on Internet of Things (WF-IoT).

43. Majumdar A, Megha G (2019) Recurrent transform learning. Neural Netw 118:271–279

44. Hewamalage H., Bergmeir C., and Bandara K. (2019) Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions. Preprint submitted to Elsevier, September 4, 2019.

45. Picton P. (2000). Neural Networks, Second Edition, Palgrave.

46. Choi E, Cho S, Kim DK (2020) Power demand forecasting using Long Short-Term Memory (LSTM) deep-learning model for monitoring energy. Sustainability 12:1109