Long-Term Energy and Peak Power Demand Forecasting Based on Sequential-XGBoost

Tingze Zhang, *Student Member, IEEE*, Xinan Zhang, *Member, IEEE*, Osaka Rubasinghe, *Student Member, IEEE*, Yulin Liu, *Student Member, IEEE*, Yau Hing Chow, Herbert H. C. Iu, *Senior Member, IEEE*, and Tyrone Fernando, *Senior Member, IEEE*

***Abstract*—Long-term energy and peak power forecast are essential tasks for the effective planning of power systems. Utilities often conduct long-term energy consumption and peak power demand forecasting separately through different forecasting frameworks, resulting in high complexities. To overcome this issue, this paper proposes a complete long-term forecasting model using an eXtreme Gradient Boosting algorithm with different sequential configurations. Firstly, it contributes to establish a 1-3 years ahead monthly energy consumption forecasting model, considering some external drivers such as macro-economic and climatic conditions. Based on the nature of energy consumption profile, a multi-input multi-output sequential strategy is applied. Then, the forecasted energy consumption forms an influencing input of a multivariate 1-3 years ahead monthly peak power demand forecast model. A hybrid direct-recursive sequential configuration is adopted to handle the highly fluctuating nature of peak power demand. By forecasting peak power demand using the information of forecasted energy consumption, better forecasting accuracy was obtained. The validity of the proposed long-term forecasting model was tested using the data from New South Wales (NSW) power network. The results were compared with several state-of-the-art long-term forecast models to show its superiority.**

***Index Terms*—Energy consumption, eXtreme Gradient Boosting, feature selection, long-term forecast, peak power demand, sequential configuration.**

I. Introduction

ENERGY and demand forecasting are essential for power system operation and planning. Based on the time horizon, the forecasting can be categorised as short-term, medium-term and long-term. Short-term forecast considers a forecasting horizon of several hours to a week. If the time horizon is few weeks or months ahead, then it is considered to be medium-term forecasting. Long-term energy and demand forecasts usually correspond to a forecast horizon of one year to several years ahead [1]. Long-term forecasting is the essential foundation towards power system planning for future generation deployment, system upgrade and network expansion [2]. Energy consumption and peak power demand are two key forecasting targets of long-term forecasting. On one hand, peak power demand is important for capacity planning. It represents the worst case scenario and needs to be tested against system capacity limits. Utilities can plan long-term infrastructure upgrades and retrofits based on the forecasted peak power demand [3]. On the other hand, energy consumption forecast is required for economic evaluation and financial planning. Some articles and reports have investigated the effect of long-term electricity consumption on economic and financial developments [4], [5]. The energy consumption forecast will support utilities for financial planning and to inform market policy changes. Hence, power utilities regularly publish monthly or yearly energy consumption and peak power demand forecasts for the next one to several years to support system planning [2].

Compared to short-term forecasting, long-term forecasting must take into account the long-term trends of different factors, which makes the task complicated. Various factors influence the forecasted energy and peak power demand profiles. Some examples of these influencing factors include the historical data trend, the macro-economic indices, and the climatic conditions. Hence, to enhance the long-term forecasting accuracy, the most significant influencing factors need to be identified. For instance, in [6], a growth rate index decomposition scheme is used to identify the key factors that influence the long-term energy consumption evolution and then the result is fed into the estimated intensity models to accomplish long-term energy forecasting. The results show that consumption past trends and Gross Domestic Product (GDP) growth have a direct relationship, which can improve the accuracy of long-term energy forecasts. Ang et al. adopted stepwise regression and multiplicative decomposition models to study how temperature increase affects the long energy consumption changes in different consuming sectors [7]. Results indicate positive correlation between temperature and energy consumption changes. A deep neural network based long-term forecasting model is proposed to predict 12-months ahead national level electricity consumption [8]. In the model, GDP, Population (POP) and average temperature are identified as exogenous variables and historical consumption trend is considered as the endogenous variable. Melodi et al. proposed a

Manuscript received 23 September 2022; revised 9 February 2023 and 1 June 2023; accepted 17 June 2023. Date of publication 26 June 2023; date of current version 21 February 2024. This work was supported in part by Western Power Ltd. and in part by the University of Western Australia. Paper no. TPWRS-01428-2022. *(Corresponding author: Xinan Zhang.)*

Tingze Zhang, Xinan Zhang, Osaka Rubasinghe, Yulin Liu, Herbert H. C. Iu, and Tyrone Fernando are with Electrical, Electronic and Computer Engineering, The University of Western Australia, Crawley, WA 6009, Australia (e-mail: tingze.zhang@research.uwa.edu.au; xinan.zhang@uwa.edu.au; osaka.rubasinghe@research.uwa.edu.au; yulin.liu@research.uwa.edu.au; herbert.iu@uwa.edu.au; tyrone.fernando@uwa.edu.au).

Yau Hing Chow is with the Grid Transformation of Western Power, Perth, WA 6000, Australia (e-mail: Yau.Chow@westernpower.com.au).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPWRS.2023.3289400.

Digital Object Identifier 10.1109/TPWRS.2023.3289400

probabilistic long-term peak power forecast model using Artificial Neural Network (ANN) and Monte Carlo simulation considering the historical peak power data, POP and GDP growth as the predominant driving factors [9]. Nezzar et al. improved the long-term peak forecasting results using Nonlinear AutoRegressive (NAR) combined with Feed Forward Network (FNN) [10]. As a univariate model, only the past peak power values are considered as input features. In [11], a Multiple Linear Regression (MLR) model that considers hourly information was established to achieve long-term peak forecasting. The main effects considered in this work include historical trend, GDP and Cooling/Heating Degree Days (CDD/HDD). Among all the aforementioned literature on long-term energy consumption or peak power demand forecasting, the historical trend is identified to be the most dominant input feature of the forecasting algorithm. Moreover, commonly used macro-economic indices include GDP and POP, while the climatic conditions are captured by CDD/HDD and temperature. However, very limited work investigated the relationship between the energy consumption and peak power demand. In theory, energy consumption implies the trend of demand fluctuations and thus, might be useful for the forecasting of peak power demand. Motivated by this idea, the authors propose a complete long-term forecasting model which combines energy consumption forecasting and peak power demand forecasting. In this model, regional monthly energy consumption measured in MWh is firstly forecasted for the next one to several years. Then, the correlation between energy and peak power is studied and the result is used as one of the influencing factors to forecast the same region's monthly peak power demand measured in MW over the same forecasting horizon. The obtained monthly forecasts over future one-year or multiple years ahead are meaningful for utilities operating in a deregulated electricity market. For energy consumption, the forecasts provide utilities with valuable information about the energy market needs, enabling them to plan fuel purchases and negotiate contracts with other businesses advantageously, lowering the financial risks. For peak results, they are essential for future peak capacity planning and major plants maintenance scheduling. Also, the proposed model establishes the relationship between peak power demand and energy consumption in long-term forecasting context, enabling operators to better understand their dependency.

In addition to the influencing factors, an excellent forecasting algorithm is also needed to produce accurate results. The state-of-the-art long-term forecasting techniques can be broadly categorised into two streams: statistical and Artificial Intelligent (AI) methods. For statistical approaches, AutoRegressive Integrated Moving Average (ARIMA) has been used for long-term energy forecasting due to its simplicity [12], [13]. However, this method is a typical univariate approach, which can only analyse the historical trend. This feature makes ARIMA unsuitable for long-term forecasting because the long-term energy consumption change is also highly affected by some other external factors. Moreover, statistical multivariate models, such as MLR and Vector AutoRegression (VAR), are also adopted in long-term forecasting work [14], [15], [16]. Nevertheless, the linear regression models have limitations in rejecting unimportant explanatory variables and addressing multicollinearity issues [14]. This disadvantage limits the applicability of such models in forecasting of time series with random fluctuations, such as monthly peak power demand. In contrast, AI methods are capable to deal with nonlinearities introduced by external factors. For example, neural network based models have been adopted to forecast the long-term peak power demand and the long-term electricity price [17], [18]. Although the forecasting accuracy is higher than the traditional linear regression models, the neural network based models lack interpretability, which makes the internal process of a trained model hard to be explained. For example, when the dataset changes or more external factors (e.g., renewable penetration rate and electric vehicle adoption) need to be considered, the model is not capable to interpret the impact of external factors on the forecasting results. In addition, some neural network models, such as Long-short Term Memory (LSTM), require large scale training datasets to fully demonstrate their advantage in forecasting [19]. However, since long-term forecasting usually adopts monthly or yearly data samples, the size of dataset will be limited to few hundreds. This might result in degraded LSTM forecasting performance. To overcome the aforementioned issues, tree-based ensemble learning algorithm, such as eXtreme Gradient Boosting (XGBoost) is a good option. Such algorithm can also effectively handle nonlinearities among external variables and most importantly it provides interpretability for the external factors [20]. Furthermore, the good forecasting performance of the XGBoost with different sizes of datasets has been reported in [21]. Nonetheless, to the best of the authors' knowledge, XGBoost has been mainly applied to short-term power demand forecasting [22], [23], [24]. The literature on long-term energy and power demand forecasting using XGBoost algorithm is very sparse.

As we mentioned, the forecast targets in this work are monthly energy consumption and monthly peak power demand over the next one to several years. In theory, this long-term forecasting work can be regarded as a multi-step sequence prediction problem. Many-to-one and many-to-many sequential configurations were adopted to forecast yearly peak power demand in [17]. More accurate forecasting results were obtained by many-to-many structure when the actual peak power demand fluctuates significantly. In [25], some conventional sequential configurations, such as single-year recursive and singe-year with interval were used to forecast the yearly peak power demand. The analysis demonstrated that different configurations significantly affect the forecast accuracy. Nonetheless, few research articles have studied the impact of sequential configurations on monthly energy consumption and peak power demand forecasting. Also, peak power demand forecasting error shown in [17], [25] is as high as 5%. Hence, better sequential configurations for both energy and peak power forecasting are studied in this work.

Therefore, the main contributions of this research can be summarised as follows:

1) A unified long-term forecasting framework is proposed to sequentially forecast energy consumption and peak power demand over next one to several years. By considering the internal relationship between both targets in peak power demand forecasting, the forecasting accuracy is greatly improved.

2) A two-stage feature selection process is done for the proposed model based on qualitative domain knowledge analysis and quantitative feature importance calculation. This increases the interpretability of a machine learning model in the context of long-term energy and peak power forecasting research.
3) This work combines sequential configurations with XGBoost algorithm, which is termed as "Sequential-XGBoost", to achieve long-term monthly energy consumption and peak power demand forecasts. A thorough one-year ahead comparison of the proposed model and several state-of-the-art forecasting algorithms is done using the same public dataset. Also, the forecasting performance of a longer forecasting horizon (i.e., three years ahead) is tested to prove the scalability of the proposed model. To the best knowledge of authors, this work is the first of its kind to apply XGBoost algorithm on long-term energy and peak power forecasting work.

Rest of the article is structured as follows: Section II analyses the relationship between energy consumption and peak power demand. Section III introduces the XGBoost algorithm, sequential configurations and the process of sequential strategy selection for different forecast targets. Section IV illustrates the detailed framework of the proposed long-term forecasting model. Section V discusses the forecast results and compares them with several other state-of-the-art forecasting algorithms. Section VI draws the conclusions.

II. Relationship Between Energy Consumption and Peak Power Demand

The correlation between energy consumption and peak power demand is usually not fully explored when they are forecasted separately. Hence, to figure out the relationship between energy and peak power demand, firstly, energy consumption is computed by taking the difference between metered exports (i.e., energy to customers) and metered imports (i.e., energy to network). Then, the energy consumption can be converted to average power demand by further dividing it by time duration. For example, monthly average power is computed by (1), where MAP represents monthly average power, MEE is the monthly energy export to customers and MEI is the monthly energy import to the network. Eq. (1) indicates that the average power demand can be used as a reference for energy consumption. Mainly because the integration of average power demand in certain period is equal to energy consumption in that period.

$$\text{MAP} = \frac{\text{MEE} - \text{MEI}}{\text{Hours in Month}} \tag{1}$$

In this study, the raw energy and power demand data of the New South Wales (NSW) electricity network with 30-minute resolution is used [26]. Fig. 1 demonstrates time series of daily peak and monthly average power demand at NSW network during the 20-year period. It is obvious that average power demand curve, which is in response to long-term profile of weather conditions, is relatively smooth. On the other hand, the peak power demand curve demonstrates high volatility caused by renewable intermittency, extreme weather conditions and unpredictable load changes. Noticeably, there is a strong correlation between the average and peak power demands, which is evidenced by the Pearson correlation value of 0.73. The strong correlation verifies the internal relationship between the average power demand (i.e., a reference for energy consumption) and peak power demand, which implies that the average power can be used as an input in the forecasting of peak power demand.

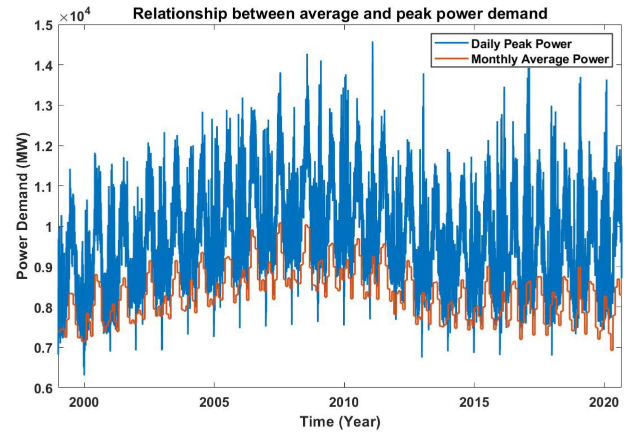


Fig. 1. Time-series of average power and peak power demand from 1999 to 2020.

III. Methodology

A. XGBoost Algorithm

XGBoost is a Gradient Boosting Decision Tree (GBDT) based ensemble learning algorithm [27], which can be used to complete regression and classification tasks. The essence of XGBoost is to stack a set of weak learners (i.e., Tree 1 to t) into a strong learner to achieve accurate forecasting. In other words, the final forecasting result $\hat{y}_i$ is equal to the sum of all weak learners' results, represented by (2).

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in \mathscr{F} \tag{2}$$

where x_i is the input features for the ith sample; K is the number of decision trees; f_k denotes the kth weak learner; $\mathscr{F}$ represents the function space containing all decision trees. Then, the regression problem is regarded as an optimization problem, and the objective function includes a loss function and a regularization term:

$$Obj = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{K} \Omega(f_k),\ f_k \in \mathscr{F} \tag{3}$$

where loss function is the indicator to measure the forecasting error between $\hat{y}_i$ and the label y_i; the regularization term Ω is used to control model complexity; n is the number of labels. XGBoost adopts the additive model in each iteration of adding a decision tree, each new tree learns the loss (i.e., negative gradient) of the sum of all previous residuals. For example, the model f_t at iteration t is built by focusing on the error between

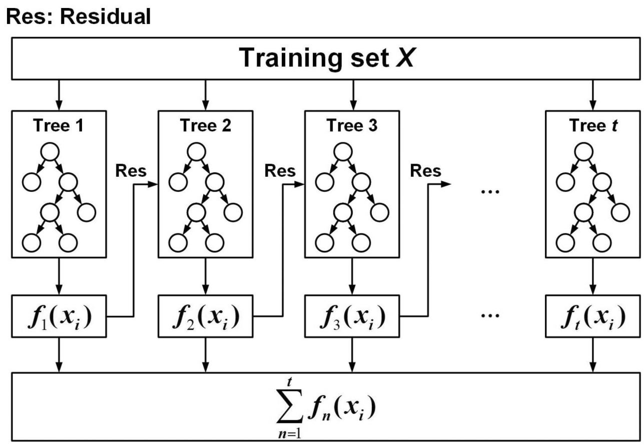


Fig. 2. Structure diagram of XGBoost algorithm.

the actual label y_i and the forecasted label $\hat{y}_i^{(t-1)}$ till iteration $(t-1)$. In the tth iteration, the predicted value is shown as $\hat{y}_i^t = \hat{y}_i^{t-1} + f_t(x_i)$, $f_t(x_i)$ represents the weak learner added at tth iteration. Thus, (3) can be rewritten as:

$$\begin{aligned} Obj^{(t)} &= \sum_{i=1}^{n} l\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{K} \Omega(f_k) \\ &= \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{t-1} + f_t\left(x_i\right)\right) + \Omega(f_t) \end{aligned} \tag{4}$$

According to the objective function shown in (4), the greedy algorithm is used to construct the decision tree. Then, by continuously adding decision trees, a complete XGBoost model is built. The structure of XGBoost algorithm is shown in Fig. 2.

To conclude, XGBoost is a decision tree-based ensemble learning model which adopts gradient boosting framework. This technique exhibits the following advantages: 1) XGBoost has a regularization term which prevents the over-fitting issue in the weak learner functions (i.e., f_t). This feature improves the generalization ability of the model. 2) XGBoost applies second-order Taylor expansion to expand the loss function [7], which reduces the complexity of the model and enhances the computational speed. 3) XGBoost has a function to sort the features based on the feature importance to the model output [28]. This is crucial for regression problems that consider multiple features as inputs, as it will support feature selection and improve the forecasting accuracy while enhancing the interpretability of the regression model.

B. Sequential Configurations Selection

Since long-term forecasting is a multi-step time series forecasting problem, to the best of authors' knowledge, several sequential strategies have been proposed in literature [29] to tackle such a multi-step ahead forecasting task. Nevertheless, few work investigates the most suitable sequential strategies for long-term energy consumption and peak power demand (i.e., two different forecast targets) forecasting. In this work, based on the correlation analysis of the time series data and forecasting accuracy, multi-input multi-output (MIMO) strategy and direct-recursive hybrid (DirRec) strategy are selected to construct the long-term (i.e., one to three years ahead) monthly energy and peak power demand forecasting framework, respectively.

1) Direct-Recursive Hybrid Strategy: This is a hybrid strategy which combines the traditional direct multi-step forecasting approach and the recursive multi-step approach [29]. The direct method involves developing a separate model for each forecasting time step. In other words, the direct strategy forecasts each horizon independently from the others. For example, in the case of forecasting next H values of a time series $[y_t, \ldots, y_{t-d+1}]$, H models are learned from the time series:

$$\hat{y}_{t+h} = f_h(y_t, \ldots, y_{t-d+1}),\ h \in \{1, \ldots, H\} \tag{5}$$

where $\hat{(\cdot)}$ refers to the predicted values, d denotes the number of past values used to forecast future horizons, and f denotes the trained model. It is obvious that the direct strategy does not utilise any predicted values to compute the forecasts, thus is not affected by the accumulation of forecasting errors. However, the independent H models lead to a conditional independence of the H forecasts. This may affect the forecast accuracy as it loses the dependencies between the forecasting values $\hat{y}_{t+h}$, such as forecasting $\hat{y}_{t+2}$ is usually dependent on $\hat{y}_{t+1}$, which is common for time series forecasting. Also, this strategy requires a large computational time as H models need to be learned (one for each horizon) during the forecast process.

On the other hand, the recursive approach builds a one-step model f which is applied to forecast each future time step. The forecasting from the previous time step will be considered as an input to forecast the next time step using the same one-step ahead model. Hence, when forecasting H steps ahead, the forecasts are expressed as:

$$\hat{y}_{t+h} = \begin{cases} f(y_t, \ldots, y_{t-d+1}),\ h = 1 \\ f(\hat{y}_{t+h-1}, \ldots, \hat{y}_{t+1}, y_t, \ldots, y_{t-d+h}),\ h \in \{2, \ldots, d\} \\ f(\hat{y}_{t+h-1}, \ldots, \hat{y}_{t+h-d}),\ h \in \{d+1, \ldots, H\} \end{cases} \tag{6}$$

This configuration considers dependencies between adjacent forecasting results. Nonetheless, since the established one-step model will be used recursively, the forecasting error at each future time step may accumulate when applying this strategy for a long forecasting window. This is especially true when the forecasting horizon h is greater than d, as when $h \in \{d+1, \ldots, H\}$, all the inputs are predicted values instead of any actual observations. Also, the computational time is small since only one model needs to be trained during the forecast process.

To combine the advantages of both direct and recursive methods, a hybrid strategy is proposed. The hybrid approach (DirRec) constructs separate models for future time steps and considers the forecasting result in the previous step for future forecasting. The H steps ahead forecasts are shown as:

$$\hat{y}_{t+h} = \begin{cases} f_h(y_t, \ldots, y_{t-d+1}),\ h = 1 \\ f_h(\hat{y}_{t+h-1}, \ldots, \hat{y}_{t+1}, y_t, \ldots, y_{t-d+1}),\ h \in \{2, \ldots, H\} \end{cases} \tag{7}$$

It is worth noting that unlike the direct and recursive strategies, the size of d (i.e., the number of past values used to forecast future steps) of the hybrid approach enlarges as forecasting horizon increases. This feature prevents the strategy from only considering forecasting values as the input set to forecast future horizon, thereby ensuring the forecasting accuracy for each time step. Also, a large computational time is needed since H models are trained (like the direct method).

2) Multi-Input Multi-Output Strategy: Multi-input multi-output (MIMO) configuration is another strategy that can be used to achieve multi-step time series forecasting [29]. It can forecast an entire signal sequence with many steps at one time. Let the trained multiple-output model be F, then the H steps ahead forecasting is given by:

$$[\hat{y}_{t+H}, \ldots, \hat{y}_{t+1}] = F(y_t, \ldots, y_{t-d+1}) \tag{8}$$

The MIMO strategy establishes a model to learn dependencies between historical and forecasting time series. The correlations between historical time steps are preserved and mapped to forecasting time steps. Compared with the hybrid strategy, this configuration focuses more on the overall forecasting accuracy of the forecast window. This characteristic makes it effective in forecasting time series that have a strong correlation with the shape of the historical curve. In other words, the MIMO strategy can learn the general trend of a time series well. Also, since only one model is required, the strategy requires a small computational time.

3) Sequential Strategy Selection for Energy Consumption and Peak Power Demand Forecasting: As introduced in the previous two subsections, different sequential strategies have different input/output structures to achieve multi-step forecasting. The different input/output pairs have their own strength in forecasting different targets. On the one hand, the DirRec approach emphasises the forecast accuracy of each future time step. This feature can help to forecast time series with large fluctuations compared to its historical curve. On the other hand, the MIMO strategy works well for forecasting time series with stable historical trends because the approach focuses on accuracy over the entire forecast window. Therefore, it is necessary to learn energy consumption and peak power demand time series, so as to select appropriate sequential strategy for them.

Fig. 3 shows the annual energy consumption and peak power demand curves for the years from 1999 to 2019. It is worth noting that both energy and peak power curves have clear seasonal trends. Nonetheless, the shape of the annual energy curve fluctuates small from year to year, while the annual peak curve demonstrates random fluctuations. This is proven by calculating the Pearson correlation between every two years and the result is shown in Fig. 4. According to the heatmaps, for any two years, the correlation of the energy curves is higher than 0.6, whereas the peak curve correlation can be as low as 0.2. Furthermore, the average correlation is calculated among all the instances (i.e., every two years indicate one instance). Values of 0.874 and 0.631 are obtained for energy consumption and peak power demand, respectively. Therefore, combining the above correlation analysis with the characteristics of the introduced sequential strategies, the MIMO strategy and DirRec strategy are suggested for one-year to several years ahead monthly energy consumption and peak power demand forecasting, respectively.

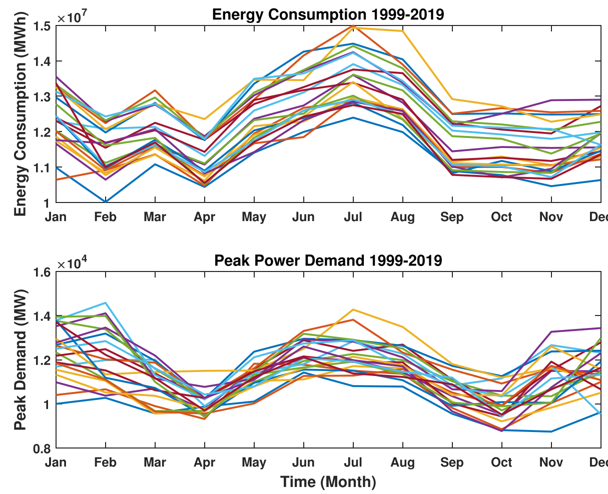


Fig. 3. Annual energy consumption and peak power demand 1999-2019.

IV. Long-Term Energy and Peak Power Forecasting Framework

The long-term energy and peak power forecasting framework based on the Sequential-XGBoost model is illustrated in Fig. 5. The framework is mainly established by three stages which include data processing, feature and sequential strategy selection and XGBoost-based forecasting. Each stage is described in detail in this section.

A. Data Description and Data Processing

The data used for this work comes from three data sources. 1) The historical demand data for the state of NSW was collected from Australian Energy Market Operator (AEMO) [26]; 2) The historical and future economic/population data for NSW was obtained from Australian Bureau of Statistics [30]; 3) The historical temperature data for the same region was collected from Bureau of Meteorology (BOM) [31]. For the future temperature data, which is normally difficult to forecast, the historical average temperature observed in the same area in the past ten years is used for conservative estimation [17], [32]. The usage of historical average temperature complies with the conservative nature of power system planning. In total, twenty years (1999-2019) data were used in the case study. It is worth noting that the data from different sources are sampled at different rates, i.e., some are sampled with an interval of 30-min, some are sampled monthly or three monthly. In this work, all data were re-sampled to monthly resolution. For example, monthly peak power was calculated from picking the largest value from the 30-minute load values that were obtained for a one-month period. The raw load data has a small amount of missing data due to interrupted signal transmission or instrument error. To effectively diminish the impact of missing data on forecasting, missing data was filled with the mean value calculated from the previous and next

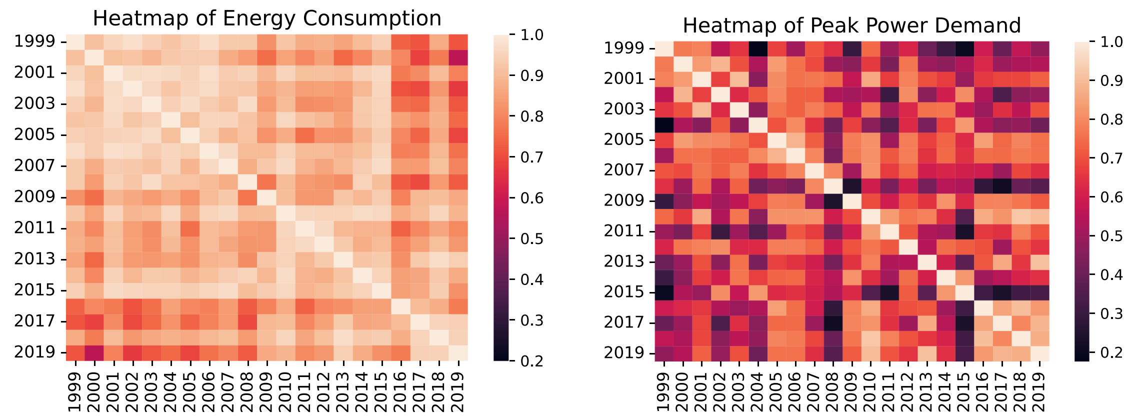


Fig. 4. Correlation analysis for the year from 1999 to 2019 a) Correlation heatmap of annual energy consumption and b) Correlation heatmap of annual peak power demand.

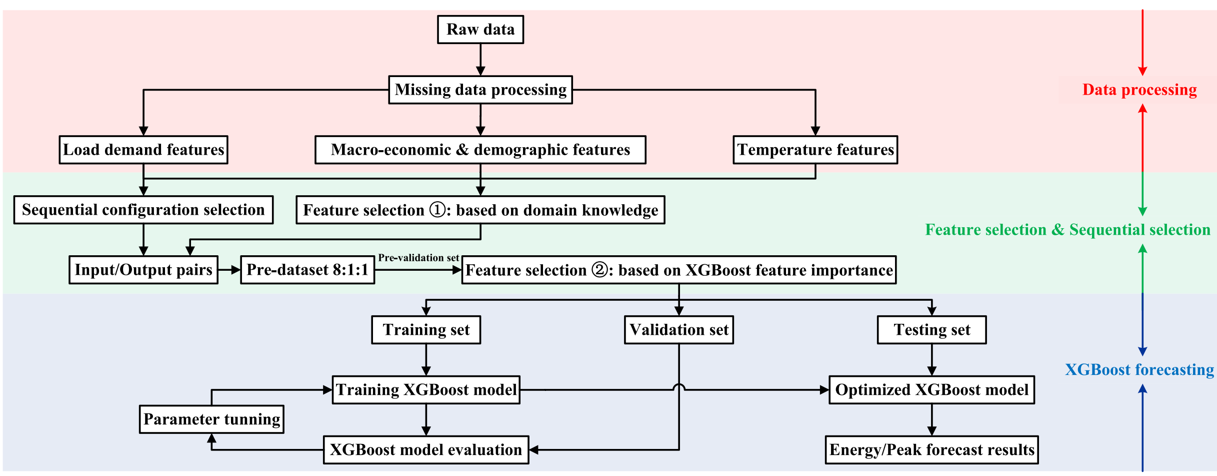


Fig. 5. Long-term energy and peak power forecasting framework based on the Sequential-XGBoost model.

moment of the same day. Furthermore, raw temperature data are monthly values extracted directly from [31], so no further processing is required for temperature data. Also, for the three monthly economic raw data, the database provides up to 82 economic factors including demographic variables, labour market variables, national accounts variables, etc. However, missing values are visible, since the statistical records of some variables cannot be traced back to 1999. For an example, Dwelling Stock (DWS) has only been recorded since 2006. For such missing data, piecewise linear interpolation was adopted to complete the three monthly data from 1999 to 2005. After that, Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) technique was used in this work to convert the economic data to a monthly resolution.

B. *Feature Selection I: Based on Domain Knowledge*

After data processing, all features are unified to a monthly resolution. The next step of building a forecasting model is feature selection. In this article, the feature selection process consists of two stages: I. Feature selection based on domain knowledge; II. Feature selection through XGBoost feature importance calculation. Domain knowledge is important, especially when the pool of input features are large (e.g., nearly 90 features here), it is the empirical knowledge pertaining to a particular specialised field [33]. We employed domain knowledge in this work to qualitatively analyse and select input features related to energy consumption and peak power demand, which helps to reduce the data dimensionality and keeps only the important features. Noticeably, in this work, the input features are broadly categorised as exogenous and endogenous features according to their nature. The exogenous features refer to external factors that influence the change of forecast target; whereas the endogenous feature refers to the historical data trend of the forecast target.

1) Exogenous Features: Exogenous features are defined as the overall external drivers in a forecast region, which are used as the inputs of a load forecasting model. In the proposed method, the macro-economic, demographic and temperature factors are taken as the exogenous inputs.

1) *Macro-economic Inputs:* A variety of economic indicators can affect the electricity usage of NSW. Firstly, Gross State Product (GSP) is added to the forecast model. In practice, long-term and short-term dynamics between energy consumption and economic output (i.e., GSP) for NSW over a

25-year period is investigated in [34]. The results demonstrate that energy consumption is positively related to the GSP. Moreover, several other economic indicators such as the Total Employment (EMP) and DWS are identified. High employment results in more electricity consumption. DWS describes the number of residential units in a region. Large DWS implies more energy consumption as well as greater peak power demand.

2) *Population Inputs:* POP is another important driver of the long-term energy consumption. As the population grows, the electricity usage will experience a significant increase. Even when the economy slows down, the residential energy consumption level remains quite stable. This is because, the residential electricity usage is mainly affected by daily household activities, such as cooking, laundry, lighting, etc. These activities are less affected by economic conditions. Moreover, labor force as part of the population affects economic activities and it is related to EMP in **Macro-economic Inputs**.

3) *Weather Inputs:* Temperature is another key factor to be considered. It should be noted that the temperature indicators considered in this work are different for energy consumption and peak power demand. On the one hand, as we discussed in Section II, energy consumption can be described by average power demand. Average power demand is correlated with the averaged weather condition (i.e., averaged temperature). Hence, monthly mean min/max temperatures are used [31]. They present the average minimum/maximum temperatures of the entire month. On the other hand, peak power demand shows high volatility in response to extreme weather scenarios. Thus, monthly lowest/highest temperatures are selected [31]. This is because peak power demand often coincides with extreme temperatures caused by cooling and heating electricity usage [35]. Furthermore, to cope with the high volatility characteristics of peak power demand, min/max temperature changes are also taken into account. For example, if the highest temperatures in January and February are 32 °C and 30.5 °C, respectively, then the change of temperature is −1.5 °C.

2) Endogenous Features - Historical Energy and Peak Power Demand: In the proposed model, the forecasting of future energy consumption uses monthly energy consumption data of the previous year as the endogenous input. When forecasting future peak power demand, the monthly peak power as well as monthly average power demand of the previous year or more historical months are considered.

Overall, the input features discussed above are summarised in Table I based on either the forecasting target is energy consumption or peak power demand. These features are important because they are the primary features that affect the customer energy usage in a forecast region. They are readily available from a utility application prospective. Also, for some technological breakthroughs, such as Distributed Energy Resources (DER) and Electric Vehicle (EV) adoption, they can significantly affect the future load. However, these features are specific to different regions and can be included if applicable.

TABLE I
INPUT FEATURES FOR ENERGY CONSUMPTION AND PEAK POWER DEMAND

Input Feature (Unit)	Group	Forecast Target
GSP ($ million)	Exogenous	Energy Consumption
EMP ('000 persons)	Exogenous	Energy Consumption
DWS ('000)	Exogenous	Energy Consumption /Peak Power Demand
POP ('000 persons)	Exogenous	Energy Consumption
Mean Max/Min Temperature (°C)	Exogenous	Energy Consumption
Max/Min Temperature (°C)	Exogenous	Peak Power Demand
Change of Max/Min Temperature (°C)	Exogenous	Peak Power Demand
Historical Monthly Energy Consumption (MWh)	Endogenous	Energy Consumption
Historical Monthly Peak Power Demand (MW)	Endogenous	Peak Power Demand
Historical Monthly Average Power Demand (MW)	Endogenous	Peak Power Demand

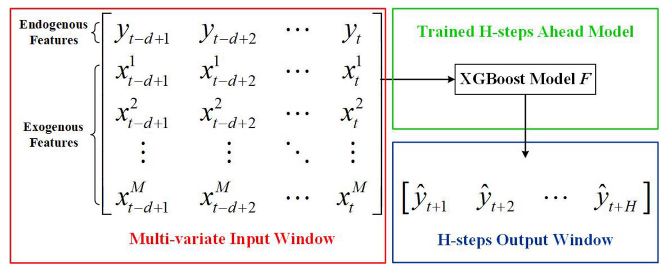


Fig. 6. Input/output mapping structure of long-term energy forecasting model with MIMO strategy.

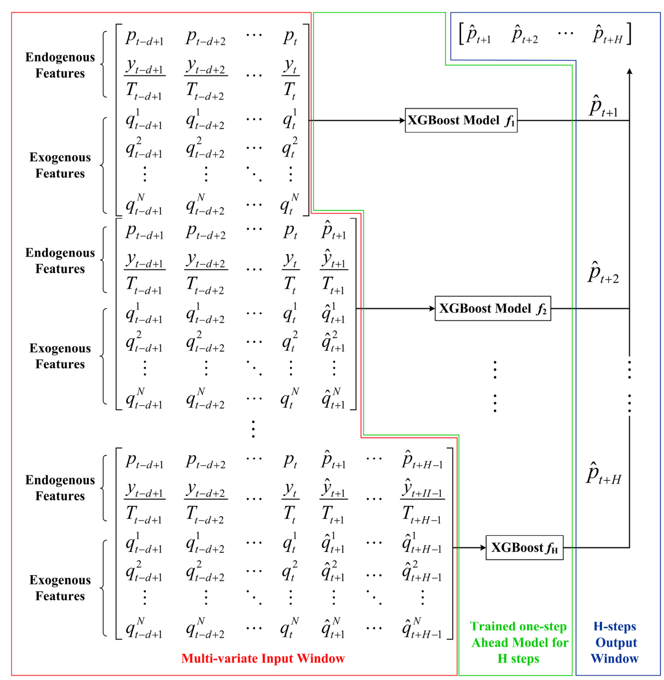


Fig. 7. Input/output mapping structure of long-term peak power forecasting model with DirRec strategy.

C. Input/Output Formats Based on Sequential Configuration

As discussed in Section III-B, MIMO strategy and DirRec strategy are suggested for one to three years ahead monthly energy consumption and peak power demand forecasting, respectively. To fit into the different sequential configurations, the raw monthly features need to be grouped into specific input/output formats. Figs. 6 and 7 show the input/output mapping structure

TABLE II
DESCRIPTION OF SYMBOL IN FIGS. 6 AND 7

Symbol	Description
$y_{(\cdot)}$	Energy consumption at historical time step
$x_{(\cdot)}^{M}$	M-th exogenous feature of energy model at historical time step
$\hat{y}_{(\cdot)}$	Forecast energy consumption at future time step
$p_{(\cdot)}$	Peak power demand at historical time step
$T_{(\cdot)}$	The number of hours in a month
$q_{(\cdot)}^{N}$	N-th exogenous feature of peak power model at historical time step
$\hat{p}_{(\cdot)}$	Forecast peak power demand at future time step
$\hat{q}_{(\cdot)}^{N}$	Forecast N-th exogenous feature of peak power model at future time step
d	The number of past values used to forecast future time step
t	t-th month
H	H-th forecast value
F	XGBoost energy forecast model
f_H	H-th XGBoost peak power forecast model

of long-term energy and peak power forecasting models using MIMO and DirRec strategies respectively. Let us take one-year ahead forecasting as an example. H is set to 12 representing future twelve months. Furthermore, as both annual energy and peak power curves show clear seasonal trends (Fig. 3), the features of previous year may have a significant impact on forecasting of the future year. Hence, d is set to 12, the previous 12 months are used to forecast the future. The descriptions of other symbols in Figs. 6 and 7 are given in Table II.

According to Fig. 6, the input/output format of the one-year ahead energy forecasting model with MIMO strategy can be regarded as a "12 to 12" time series forecasting structure. The model is trying to figure out the correlation between the previous year and the current year. The configuration makes this model very efficient since no model update (i.e., only F is trained) is required and it does not utilise any future data. Table III shows an example of how input/output pairs are generated based on the "12 to 12" energy forecasting structure using the available data. In the table, time duration (e.g., 1999.1-1999.12) is used instead of original data. It is worth noting that a fixed-length (i.e., 12 steps) rolling window is adopted to generate the input/output pairs. On the other hand, Fig. 7 shows the input/output format of the peak power forecasting model with DirRec strategy. Compared with the energy model, it has the following main differences: 1) Input/output pairs follow a "many to one" mapping structure. The multi-variate input window is no longer a fixed size, it enlarges at each forecasting step until all 12 steps are predicted. 2) The forecasting values are used as inputs to forecast the next step, and this process will recursively run for all 12 steps. 3) Twelve models are learned (one for each forecasting step), each model has different sizes of rolling windows (e.g., size 12, 13, etc.) to generate input/output pairs. 4) Average power demand as a representative of energy consumption is considered as another endogenous feature when predicting the peak power demand. The average power (i.e., $\frac{y_{(\cdot)}}{T_{(\cdot)}}$ and $\frac{\hat{y}_{(\cdot)}}{T_{(\cdot)}}$) is converted from energy consumption through (1), and the energy consumption is extracted from one-year ahead energy forecasting model. The forecasted average power demand (i.e., $\frac{\hat{y}_{(\cdot)}}{T_{(\cdot)}}$) at previous time step will also serve as an input to forecast the peak power demand at current time step. Table IV shows the example of generating input/output pairs based on the first peak power forecasting model f_1 which is "12 to 1" structure. The length of rolling window is 12 steps. Similarly, we can write the input/output of the first data record of the second peak power forecasting model as "1999.1-2000.1/2000.2". The rolling window is set to 13 steps. In the same way, the input/output pairs of the H-th peak power forecasting model (i.e., f_H) can be deduced.

D. Feature Selection II: Feature Importance Calculation Based on XGBoost Model

In Section IV-B, domain knowledge is adopted to select input features for energy consumption and peak power demand in a qualitative manner. To further prove the reliability of the selection in a quantitative way, permutation importance is employed to calculate the feature importance of XGBoost model. The rationale of permutation importance is the decrease in a model score when a single feature is randomly shuffled [36]. The calculation process breaks relationship between features and targets, thus changes in the model forecast accuracy indicate how much the model depends on the features. In this work, all features listed in Table I are included in energy consumption and peak power demand forecasting models. The generation of input/output pairs based on these identified features are defined as pre-datasets and they are split into pre-training, pre-validation and pre-testing sets with a proportion of 8:1:1. Then, the pre-validation set is used to complete the second stage of feature selection. The results shown in Fig. 8 and Table V demonstrate how each feature contributes to the trained models. The importance is normalised and the sum of all features' importance equals to one.

In the energy consumption training model illustrated in Fig. 8(a) and Table V, the historical total energy consumption is the most important feature for future energy forecasting. Furthermore, among all macro-economic and demographic features, POP and DWS exhibit significant impacts on electricity usage. EMP and GSP are also non-negligible. In terms of temperature, the energy usage of a month is more sensitive to the mean value of min/max temperatures. On the other hand, in the peak power demand training model illustrated in Fig. 8(b) and Table V, the historical peak power demand serves as the most important input feature. Moreover, historical average power demand is considered as the second most important feature, which verifies the previously highlighted relationship between average power demand and peak power demand. Also, minimum/maximum temperatures are much more important for peak power demand forecasting compared to their mean values. In fact, even the influence of change minimum/maximum temperatures is greater than that of the mean temperatures. However, all temperature change features do not contribute much to peak power forecasting model. In terms of economic and population indicators, the impact of DWS is much larger than other factors. Therefore, the input features of the energy and peak power demand forecast models are selected based on domain knowledge as well as feature importance analysis on the training XGBoost models. The final features used in the one-year ahead energy and peak power forecasting models are highlighted in Fig. 8.

TABLE III
EXAMPLE OF INPUT/OUTPUT FORMAT OF ONE-YEAR AHEAD ENERGY FORECASTING MODEL WITH MIMO STRATEGY

Data Record Number	Endogenous Features (Input)	Exogenous Features (Input)				Target (Output)
	Hist_Tot_Energy $y_{(\cdot)}$	$x^1_{(\cdot)}$	$x^2_{(\cdot)}$	...	$x^M_{(\cdot)}$	Tot_Energy $\hat{y}_{(\cdot)}$
1	1999.1-1999.12	1999.1-1999.12	1999.1-1999.12	1999.1-1999.12	1999.1-1999.12	2000.1-2000.12
2	1999.2-2000.1	1999.2-2000.1	1999.2-2000.1	1999.2-2000.1	1999.2-2000.1	2000.2-2001.1
...	...	...	...	...	...	...
193	2015.1-2015.12	2015.1-2015.12	2015.1-2015.12	2015.1-2015.12	2015.1-2015.12	2016.1-2016.12
...	...	...	...	...	...	...

TABLE IV
EXAMPLE OF INPUT/OUTPUT FORMAT FOR THE FIRST MODEL OF ONE-YEAR AHEAD PEAK POWER DEMAND FORECASTING WITH DIRREC STRATEGY

Data Record Number	Endogenous Features (Input)		Exogenous Features (Input)				Target (Output)
	Hist_PeakPDem $p_{(\cdot)}$	Hist_AvePDem $\frac{y_{(\cdot)}}{T_{(\cdot)}}$	$q^1_{(\cdot)}$	$q^2_{(\cdot)}$	...	$q^N_{(\cdot)}$	PeakPDem $\hat{p}_{(\cdot)}$
1	1999.1-1999.12	1999.1-1999.12	1999.1-1999.12	1999.1-1999.12	1999.1-1999.12	1999.1-1999.12	2000.1
2	1999.2-2000.1	1999.2-2000.1	1999.2-2000.1	1999.2-2000.1	1999.2-2000.1	1999.2-2000.1	2000.2
...	...	...	...	...	...	...	...
204	2015.12-2016.11	2015.12-2016.11	2015.12-2016.11	2015.12-2016.11	2015.12-2016.11	2015.12-2016.11	2016.12
...	...	...	...	...	...	...	...

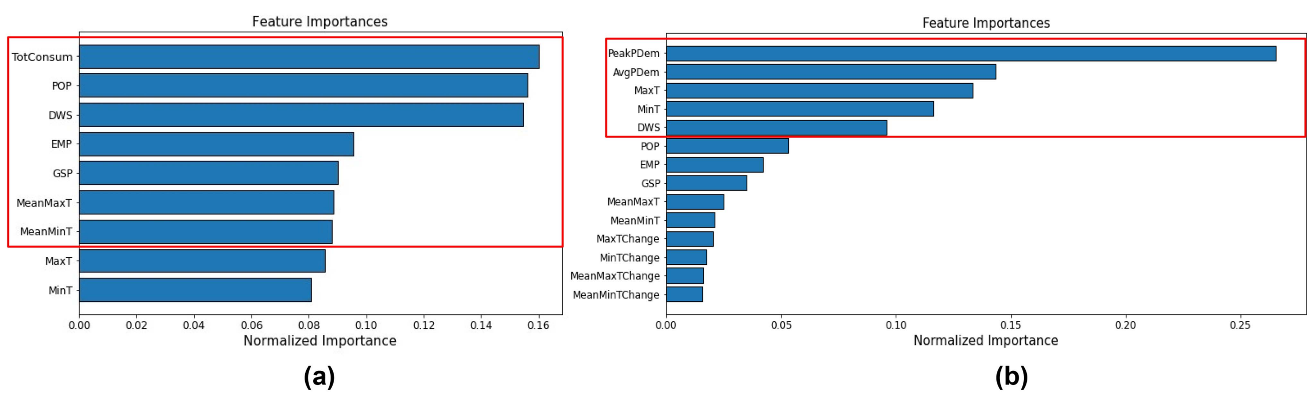


Fig. 8. (a) Feature importance for energy forecast model (b) Feature importance for peak power forecast model.

TABLE V
FEATURE IMPORTANCE STATISTICS

Energy Forecasting Model		**Peak Power Forecasting Model**	
Feature Ranking	Importance (p.u.)	Feature Ranking	Importance (p.u.)
TotConsum	0.160	PeakPDem	0.258
POP	0.156	AvgPDem	0.139
DWS	0.155	MaxT	0.133
EMP	0.096	MinT	0.120
GSP	0.090	DWS	0.103
MeanMaxT	0.089	POP	0.051
MeanMinT	0.088	EMP	0.041
MaxT	0.085	GSP	0.035
MinT	0.081	MeanMaxT	0.030
-	-	MeanMinT	0.021
-	-	MaxTChange	0.020
-	-	MinTChange	0.018
-	-	MeanMaxTChange	0.016
-	-	MeanMinTChang	0.015

E. XGBoost Based One-Year Ahead Energy and Peak Power Forecasting

After feature selection and sequential strategy selection, XGBoost was applied to forecast one-year ahead energy consumption and peak power demand of the NSW network. It is worth noting that XGBoost as a tree model does not require data normalisation process [27]. The algorithm was implemented with the XGBoost library [37] in Python. For energy consumption model with MIMO sequential configuration (i.e., 12 to 12 forecast structure), 193 input/output pairs were recorded based on the data format shown in Table III. The 193 records were split into a training set, a validation set and a testing set based on the proportion of 8:1:1. In terms of peak power demand model with DirRec sequential configuration (i.e., many to one forecast structure), different lengths of input/output pairs for each forecasting step were recorded based on the input/output format described in Section IV-C. Nonetheless, the data division ratio is still set to 8:1:1. Overall, the training set is used to

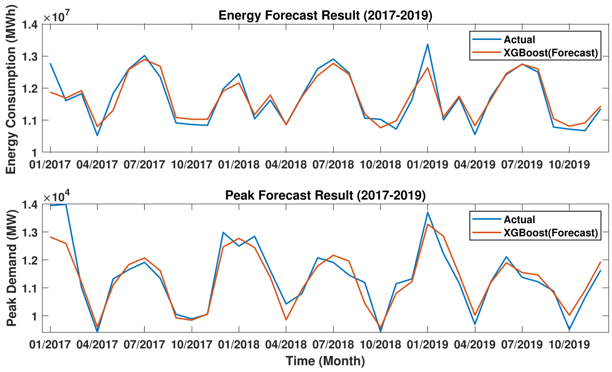


Fig. 9. One-year ahead energy and peak power forecast results.

train the XGBoost models and the validation set is adopted for hyperparameter tuning. After that, the testing set is input to the optimised XGBoost models to obtain the forecasting results. To evaluate the performance of one-year ahead forecast models, the energy consumption and peak power demand of year 2017-2019 were forecasted and compared to the actual values. In this article, statistical metrics Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) are used to evaluate the model performance and they are defined as follows:

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{9}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{10}$$

where N is the number of forecasting months, y_i is the actual energy consumption or peak power demand of month i, $\hat{y}_i$ is the forecasted energy consumption or peak power demand of month i. Moreover, the following XGBoost hyperparameter settings were identified to be the most optimal using grid search approach [38]: the number of sequential trees (i.e., number of estimators) is 80, the learning rate is 0.1 and the max tree depth is set to 6. Also, squared error is selected as the loss function. In addition, early stopping rounds with a patience of 15 instances is set to prevent the over-fitting issue. If the error of validation set does not decrease for 15 consecutive times, the iteration will stop and the configuration of parameters will be saved.

V. Forecast Results Discussion

A. XGBoost Point and Probabilistic Forecast Results

Fig. 9 and Table VI demonstrate the one-year ahead forecasting results for year 2017-2019. A few observations can be drawn from the outcomes: 1) The result of energy forecasting is quite accurate with a 1.93% three-year average MAPE. Although average MAE value (2.29×10^5 MWh) of energy forecasting looks large, it is a small number compared to the raw energy data reaching 10^7 MWh. 2) The overall forecasting result of peak power demand is also accurate with a 2.8% average MAPE

TABLE VI
One-Year Ahead Forecast Accuracy

Year	Energy Forecasting		Peak Power Forecasting	
	MAE (MWh)	MAPE (%)	MAE (MW)	MAPE (%)
2017	2.67×10^5	2.25	325.36	2.89
2018	1.78×10^5	1.5	335.01	2.94
2019	2.42×10^5	2.04	287.33	2.56
Mean	$\mathbf{2.29 \times 10^5}$	**1.93**	**315.9**	**2.80**

and 315.9 MW MAE. The proposed peak power model has the capability to forecast some irregular fluctuations accurately. This characteristic is attributed to the DirRec sequential strategy that we adopted. Furthermore, it was found that relatively large peak power demand forecasting errors occur in January and February 2017. The reason is that, the start of 2017 experienced a continuous heatwave, with the highest temperature reaching about 40 degrees for two consecutive months [31], [39]. This phenomenon resulted in a significant increase of cooling system usage on hot days of these two months, which increased the amount of power consumed to 14,000 MW (i.e., almost the record highest MWs in NSW history). This heatwave condition is a historically extreme event and therefore leads to less accurate forecasting. In the future, the extreme weather conditions may happen more often and thus, having an accurate prediction of the temperature or other climate event may support us to obtain more accurate peak power forecasting in case of extreme events. This can be further researched in future work.

After analysing the deterministic forecasting results shown in Fig. 9, a statistical analysis of forecasting errors was conducted to generate probabilistic forecasting results with a certain confidence interval. The overall process contains 1) Collect forecasting errors of the sample dataset and draw the frequency histogram. 2) Apply a suitable probability distribution function to fit the frequency histogram. 3) Calculate confidence interval based on the parameters (i.e. (μ, σ)) of the selected probability function and the assumed confidence level. As illustrated in Fig. 10(a) and (b), forecasting error distributions of both energy and peak power models well fit with the normal distribution function (also known as Gaussian distribution). The probability density function of the normal distribution is expressed as:

$$\text{Forecast Error}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{11}$$

where μ is the mean and σ is the standard deviation of the forecast errors. Then, 95% is chosen as the confidence level in this work and the upper and lower forecast boundaries are calculated by $c\frac{\sigma}{\sqrt{n}}$, where c is 1.96 refers to 95% confidence level and n indicates the number of samples. Therefore, the probability forecast results of the energy and peak power models are shown in Fig. 11. To evaluate the performance of probabilistic forecasting result, the prediction interval coverage probability (PICP) metric is used. PICP represents the probability that actual values fall within the upper and lower bounds of the confidence interval. It is a widely used statistical metric for density forecasting [40] and the formula is defined as follows:

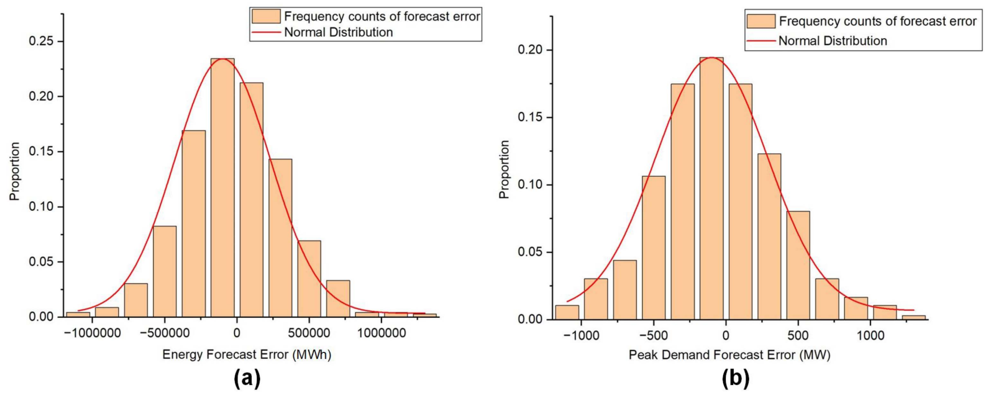


Fig. 10. (a) Energy forecast model prediction error distribution (b) Peak power forecast model prediction error distribution.

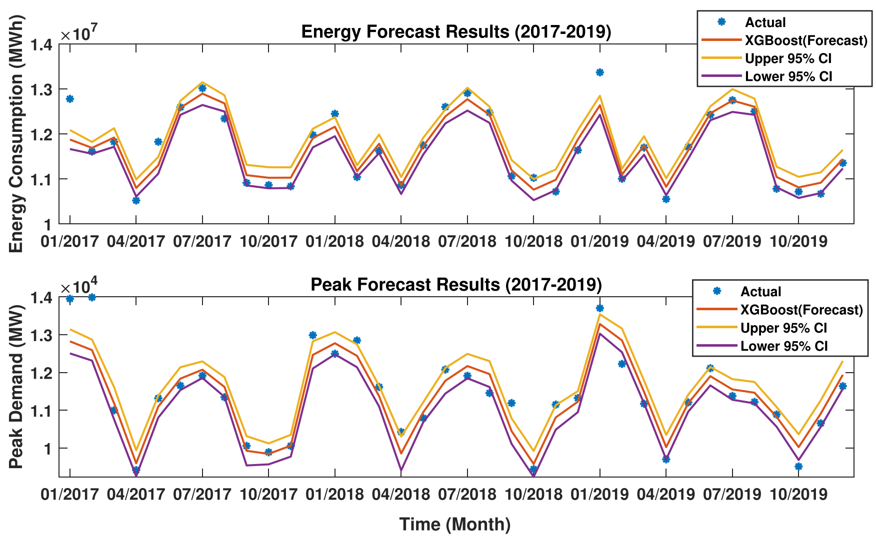


Fig. 11. One-year ahead energy and peak power demand probability forecast results.

$$\text{PICP} = \frac{1}{N}\sum_{i=1}^{N}\mu_i \tag{12}$$

where N is the number of forecasting months, μ_i is a Boolean variable and it is expressed as:

$$\mu_i = \begin{cases} 1, \; y_i \in [L_i, U_i] \\ 0, \; y_i \notin [L_i, U_i] \end{cases} \tag{13}$$

if the actual observation y_i lies within the interval $[L_i, U_i]$, μ_i is 1, otherwise is 0. The actual energy or peak power values are labelled as black dots in Fig. 11. After the calculation, the PICP obtained by energy and peak power probabilistic forecast results are 0.83 and 0.805 respectively for 95% confidence interval. The 5% is caused by the aforementioned extreme events (e.g., consecutive heatwave conditions) that were quite rare in the past several decades. In summary, the probabilistic forecasting results demonstrate the great forecasting performance of proposed models.

TABLE VII
EFFECT OF DIFFERENT SEQUENTIAL CONFIGURATIONS

Forecasting Model	3-year Average MAPE	3-year Average MAE
Energy-MIMO (Proposed)	1.93%	2.29×10^5 MWh
Energy-DirRec	1.98%	2.35×10^5 MWh
Peak-DirRec (Proposed)	2.80%	315.9 MW
Peak-MIMO	4.68%	553.01 MW

B. *Effect of Different Sequential Configurations*

One contribution of this work is that different sequential configurations were selected to forecast energy and peak power demand, respectively. To test the effectiveness of the selected configurations, DirRec approach was applied on energy consumption forecasting and MIMO strategy was employed on peak power demand forecasting. The results are compared to the proposed configurations by measuring the average MAPE and MAE of the year 2017 to 2019. It can be seen from Table VII,

TABLE VIII
EFFECT OF ENERGY ON PEAK POWER FORECASTING

Year	Peak Forecast without Energy Info		Peak Forecast with Energy Info	
	MAE (MW)	MAPE (%)	MAE (MW)	MAPE (%)
2017	470.65	3.95	325.36	2.89
2018	382.98	3.32	335.01	2.94
2019	410.22	3.51	287.33	2.56
Mean	421.28	3.59	**315.9**	**2.80**

3-year average MAPE and MAE of different sequential configurations for energy consumption forecasting are close. This is to be expected because, yearly (i.e., within 12 months) energy consumption is relatively stable within each year. Nonetheless, the performance of the proposed configuration is slightly better than using DirRec structure. In addition, although the accuracy is close, the proposed configuration does not require model updating, which makes it easier to implement and therefore, it is more suitable for energy consumption forecasting. In terms of peak power demand forecasting, the proposed configuration outperforms the MIMO configuration by measuring both MAPE and MAE. This can be explained by the fact that MIMO strategy can well capture the annual trends, but can not perform well with forecasting variables that are highly volatile.

C. Effect of Energy Consumption on Peak Power Demand Forecasting

Another contribution of this work is that the forecasted energy consumption is used for long-term peak power demand forecasting. Table VIII shows the one-year ahead peak power forecasting performance with or without considering average power demand (converted by energy consumption) as an input feature. It can be seen that "Peak Forecasting with Energy Info" has the lower forecasting error for each forecasting year. The average MAPE and MAE values of the model with energy are improved by 0.79% and 105.38 MW respectively. This improvement may have a significant impact when the one-year ahead forecast result is implemented for maintenance scheduling or coordination of load dispatch. Also, this result is matching with the conclusion drawn by Fig. 8(b), which is average power demand (i.e., a reference for energy consumption) making a non-negligible contribution to long-term peak power forecasting.

D. Comparison With Different Forecast Models

In this section, the proposed Sequential-XGBoost forecast models are compared with several forecasting models that use different algorithms. To ensure the completeness of comparison, the selected contrast models cover a seasonal persistence model, univariate/multivariate models based on statistical methods, and neural network based deep learning models. The seasonal persistence model is set as a baseline to evaluate the forecasting performance of other algorithms. It is worth noting that, for the models whose inputs are multivariate, the input features highlighted in Fig. 8 are applied to all contrast models to ensure fairness. Furthermore, since one-year ahead peak power demand forecasting requires the forecasted energy consumption as input, the XGBoost-based energy forecast results are used for all contrast models. This ensures that the input data of all multivariate models are consistent when forecasting peak power demand. In addition, same sequential strategies are applied to the neural network based models, which is MIMO and DirRec strategies are used for energy consumption and peak power demand forecasting, respectively. Also, to compare the forecasting performance, the experiments are conducted for the same forecasting period (i.e., 2017-2019). A desktop PC with a 10^{th} generation Intel Core-i9 processor with 16 GB RAM is used to run the algorithms along with software configurations of Windows 10 and Python 3.8 with Tensorflow backend.

1) *Seasonal persistence model:* Persistence models are simple methods of using past data to forecast future data points. They are developed to benchmark the performance evaluation of more complex models. In this work, motivated by the seasonal characteristics of monthly energy consumption and peak power demand, seasonal persistence method is selected. This method is useful for highly seasonal time series, it assumes the condition of the current situation is the same as the one in target future, and therefore the future value is considered equal to the current value [41]. For example, with monthly data, the energy consumption at Jan 2017 is equal to the last observed January value which is Jan 2016. Based on this principle, the forecasted energy consumption and peak power demand of 2017 to 2019 were obtained.
2) *ARIMA model:* ARIMA is a univariate regression model which only analyses the time-series variable itself. To forecast one-year ahead energy consumption and peak power demand, as an example for year 2017, the historical data from 1999 to 2016 were fed into the ARIMA model for training. Same process was adopted to forecast the energy and peak power of the year 2018 and 2019. ARIMA (1,1,1) was selected based on the augmented Dickey-Fuller test, Partial AutoCorrelation Function (PACF) and AutoCorrelation Function (ACF) [42]. Then, energy and peak power demand of the year 2017/2018/2019 were forecasted and the statistical metrics were calculated for evaluation.
3) *Vector AutoRegression (VAR) model:* A statistical based multivariate VAR model is considered for comparison. An important parameter for VAR model is the number of lagged responses, in here VAR (12) was chosen to create one-year-ahead forecast models. After that, the energy and peak power forecast results of 2017-2019 were obtained through Monte Carlo simulations.
4) *Neural network based deep learning models:* In this article, several deep learning models are selected for the comparison purpose. As we mentioned, the input features as well as the sequential strategies are exactly the same as the proposed Sequential-XGBoost forecasting structure. Nonetheless, the deep learning models need data normalisation process, where Min-Max Normalisation is adopted to standardise the input features. Also, for all deep learning models, a number of hyperparameters need to be

TABLE IX
ONE-YEAR AHEAD FORECAST MODELS PERFORMANCE COMPARISON

Algorithm	Energy Forecast (2017-2019)				Peak Power Demand Forecast (2017-2019)			
	MAE (MWh)	MAPE (%)	Execution Time (s)	Improvement (%)	MAE (MW)	MAPE (%)	Execution Time (s)	Improvement (%)
Sequential-XGBoost	$\mathbf{2.29 \times 10^5}$	**1.93**	**11.75**	**0.04**	**315.90**	**2.80**	**23.47**	**2.88**
Seasonal Persistence	2.33×10^5	1.97	–	–	645.32	5.68	–	–
ARIMA	2.44×10^5	2.06	2.51	-0.09	619.14	5.25	3.47	0.43
VAR	2.30×10^5	1.94	6.1	0.03	535.41	4.54	15.11	1.14
DNN	2.57×10^5	2.19	22.31	-0.22	493.97	4.12	138.96	1.56
LSTM	2.76×10^5	2.33	23.34	-0.36	550.71	4.66	124.5	1.02
CNN-LSTM	2.66×10^5	2.27	25.03	-0.3	483.18	4.03	195.6	1.65
Attention-LSTM	2.73×10^5	2.31	20.25	-0.34	533.05	4.52	203.92	1.16

determined, such as number of hidden layers, number of neurons in each layer, epochs and batch size. In this work, the grid search algorithm [38] is used to obtain the best parameter combination that resulted in the best forecasting outcome. The determined parameters of energy and peak power forecasting models are summarised in Appendix. The considered deep learning models are as follows:

a) Deep Neural Network (DNN) model: This is the basic deep learning framework, also known as fully connected neural network.
b) LSTM model: LSTM is a sequence-based deep learning framework which can well determine the relationships between previous and current time steps.
c) CNN-LSTM model: This is another deep learning model we considered for comparison purpose. The reason for this is the advantages of both CNN and LSTM networks. The CNN network can efficiently extract the relationships among multi-dimensional time series features, then the processed data will fed into LSTM network to accomplish forecasting.
d) Attention-LSTM model: Attention module is added on the basis of the original LSTM model to improve its ability to allocate different attentions on certain parts of the input sequence when forecasting an output sequence.

The comparison among different models is measured by several metrics and the results are shown in Table IX. The improvement (%) column is a comparison between the seasonal persistence model (i.e., the benchmark model) and all the other models. On the one hand, for energy results, the persistence benchmark model shows a good forecasting accuracy with MAPE and MAE values of 2.33×10^5 MWh and 1.97%. This is expected because of the shape of annual energy curve which fluctuates insignificantly from year to year. Compared with the benchmark, XGBoost and VAR models have improvements of 0.04% and 0.03% in the three-year average MAPE value. For this close improvement, seasonal forecast accuracy analysis was conducted to evaluate the performance of XGBoost and VAR models. The average MAPE values for the four seasons in the test period were calculated and the results are shown in Table X. Based on the results, we can observe that summer is the most difficult season to forecast, the error is over 2.3% for both algorithms. Furthermore, XGBoost forecasting is more stable across seasons. Although VAR shows advantages in spring forecast, it has become worse in the following summer forecast. This drawback has affected the overall reliability of VAR forecasting performance. Thus, we can draw the conclusion that both algorithms have advantages in forecasting different seasons, while XGBoost shows more reliable forecasting ability than VAR. In addition to this, compared with the benchmark, all the other models including ARIMA and neural network based deep learning models show no improvement. The reason for this is that ARIMA is a univariate model and it is not capable of considering the effect of external factors on long-term energy forecasting. For deep learning models with no improvement, the reason is firstly the fluctuation of energy itself is relatively stable throughout the years, the deep learning models with more complex network does not mean high efficiency when predicting such target. Secondly, deep learning models are more suitable to deal with large datasets and perform relatively poor on small datasets (i.e., only hundreds of training samples). In terms of time complexity, the statistical methods (i.e., ARIMA and VAR) result in the smallest execution time, and the neural network-based approaches require an execution time over 20 seconds. The execution time of proposed XGBoost algorithm ranks in the middle. On the other hand, for one-year ahead peak power forecasting results, the seasonal persistence model shows a low forecasting accuracy with 5.68% MAPE and 645.32 MW MAE. This is expected because the annual peak power curve demonstrates random monthly fluctuations, therefore being limited to using the last observed same months' value to represent the future values are not reliable. Compared with this benchmark, all the algorithms show improvements and it is reasonable for us to compare through all the tested algorithms. Based on the results, all multivariate models perform better than the univariate one (i.e., ARIMA). Furthermore, XGBoost has the most accurate forecast result, which is 2.80% MAPE and 315.9 MW MAE. Also, the deep learning models exhibit

TABLE X
SEASONAL FORECAST PERFORMANCE METRIC

Algorithm	Forecast error % across different seasons			
	Spring (Sep-Nov)	Summer (Dec-Feb)	Autumn (Mar-May)	Winter (Jun-Aug)
XGBoost	1.87	2.32	1.85	1.67
VAR	1.33	2.88	1.90	1.66

TABLE XI
PROPOSED ONE-YEAR AHEAD MODELS COMPARE WITH EXISTING WORK

Method	Forecast Target	Forecast Horizon	MAPE (%)	Ref
SARIMAX	Monthly energy	One-year ahead	4.1	[44]
Pattern Similarity based model(PSFM)	Monthly energy	One-year ahead	2.4	[45]
MLR	Monthly energy	One-year ahead	2.9	[11]
AutoRegressive	Monthly energy	One-year ahead	2.7	[46]
FOA-MHW	Monthly energy	One-year ahead	3.65	[47]
MIMO-XGBoost	Monthly energy	One-year ahead	**1.93**	**this work**
Adaptive BP	Monthly peak	One-year ahead	4.5	[48]
MLR Model L	Monthly peak	One-year ahead	3.3	[11]
ANN	Monthly peak	One-year ahead	3.53	[43]
Fuzzy Logic	Monthly peak	One-year ahead	6.9	[49]
DirRec-XGBoost	Monthly peak	One-year ahead	**2.8**	**this work**

relatively lower accuracy ranging from 4.03% to 4.66% MAPE. CNN-LSTM demonstrates potential in highly fluctuated target forecasting. Nonetheless, for the small dataset problem in this article, the performance is still not as good as the XGBoost model. Overall, the proposed XGBoost model shows the largest improvement among all the tested algorithms in both one-year ahead energy and peak power demand forecasting. These improvements, especially peak power forecast improvement which is 2.88% or approximately 330 MW, will support efficient future peak capacity planning and maintenance scheduling efficiently. Although energy forecasting improvement is only 0.04%, the result can still provide valuable information about the market needs of future monthly energy in a one-year period, enabling the utilities to properly schedule the fuel purchases.

E. Comparison With Existing Work

Table XI compares the results of this work against some of the existing work. All selected work have the same forecasting horizon of one-year ahead for both monthly energy consumption and peak power demand. MAPE value is the only evaluation metric compared here considering the different scales of the energy markets in these work. From the algorithm perspective, the published work cover most categories of forecasting algorithms, including statistical methods, neural network based methods and fuzzy logic methods. As per the results comparison, our proposed Sequential-XGBoost models outperform other approaches, the main reasons are as follows: 1) We have combined the XGBoost algorithm with sequential configurations, different configurations were conducted on different forecasting targets based on their characteristics. Then, taking the advantages of the XGBoost simple structure and strong model interpretability, excellent forecasting results were obtained. 2) The average power demand was considered as an input in the peak power forecasting model. It is worth noting that none of the works in Table XI considered energy information when forecasting the peak power demand. According to the analysis of the effect of energy on peak power forecasting shown in Section V-C, the MAPE value of peak power forecasting without considering energy information is 3.59 %, which is close to the MAPE values shown in work [11], [43]. Therefore, it is reasonable to say that considering average power demand (i.e., a reference for energy consumption) as an input feature improves the peak power forecasting accuracy.

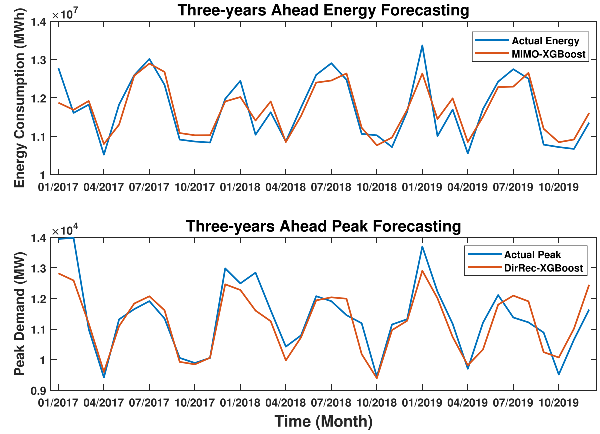


Fig. 12. Three-years ahead energy and peak power forecast results.

TABLE XII
THREE-YEARS AHEAD FORECAST PERFORMANCE

Forecast Target	Metrics	Values
Energy Consumption (2017-2019)	MAE (MWh)	2.65×10^5
	MAPE (%)	2.25
	Time (s)	14.05
Peak Power Demand (2017-2019)	MAE (MW)	426.28
	MAPE (%)	3.61
	Time (s)	74.16

F. Application Results for Three Years Ahead Forecasting

To verify the scalability of the proposed long-term energy and peak power demand forecasting models, we extend the forecast horizon to three years ahead. In other words, the forecasting window contains 36 steps representing all months in three years. Specifically, for the energy forecasting model, “12 to 12” MIMO structure is still adopted and the recursive method is used to achieve the expansion from one-year to three years. For example, if we want to forecast the energy from 2017 to 2019, the 2017 forecasting result is obtained by using the one-year ahead model, then the result is used to forecast the 2018 energy consumption, and the result for 2019 is obtained upon the result of 2018. For the peak power demand forecasting model, 36 separate models (i.e., one for each forecasting step) need to be studied during the forecasting. The input and output mapping relationship starts at “12 to 1”, and ends at “47 to 1” based on the mechanism shown in Fig. 7. The three years ahead forecasting results are shown in Fig. 12 and Table XII. It can be seen that the forecast accuracy of both energy and peak power has declined to some extent, however, the model is still capable of forecasting the energy consumption and peak power demand for the next three years

TABLE XIII
PROPOSED THREE-YEARS AHEAD MODELS COMPARE WITH EXISTING WORK

Method	Forecast Target	Forecast Horizon	MAPE (%)	Ref
Multi-CNN	Monthly energy	Three-years ahead	2.5	[50]
Multi-AR Splines	Monthly energy	Three-years ahead	3.5	[51]
ANN	Monthly energy	Three-years ahead	4.65	[52]
MIMO-XGBoost	Monthly energy	Three-years ahead	**2.25**	**this work**
Fuzzy-ANN	Monthly peak	Three-years ahead	5.73	[53]
CNN-LSTM	Monthly peak	Three-years ahead	4.29	[54]
GRU	Monthly peak	Three-years ahead	6.02	[54]
DirRec-XGBoost	Monthly peak	Three-years ahead	**3.61**	**this work**

with a very high accuracy of 97.75 % and 96.39 %, respectively. In fact, the decrease in forecasting accuracy happens in the last two years (i.e., 2018 and 2019), because the forecasting for 2017 is exactly the same for the one-year ahead and three years ahead model. The execution time is still acceptable, the execution time of three years ahead forecasting model is 74.16 s, which is about three times that of the one-year ahead model. This is expected as three times of forecasting models need to be built in the three years model. Also, to further prove the forecasting performance of the extended models, the results are compared with some existing works. Again, all selected works have the same forecasting horizon of three-years ahead for both monthly energy consumption and peak power demand. Based on the results shown in Table XIII, our proposed Sequential-XGBoost models outperform others in terms of the forecasting accuracy. This indicates the ability of the proposed models in forecasting longer period.

VI. CONCLUSION

This article proposes a new complete long-term forecasting model to sequentially forecast the energy consumption and peak power demand of power systems with high accuracy. The impact of external factors, including economic indicators and the climatic conditions, are considered. Furthermore, the correlation between energy consumption and peak power demand is studied and utilised to improve the accuracy of peak power demand forecasting. This complete long-term forecasting model implies a unified forecasting framework which is a simplification compared to the traditional methods that require separate forecasting of energy and peak power demand through different algorithms or frameworks. In addition, based on the nature of energy and peak power demand profiles, different sequential configurations are explored to achieve superior forecasting results. Also, the XGBoost algorithm is employed to obtain elevated long-term energy and peak power forecasting accuracy with a small training dataset. In our future work, the results obtained from the proposed method will be applied to power system applications, such as major plants maintenance scheduling.

APPENDIX

Model	Parameter	Value
DNN(Energy)	Hidden layers	1
	Number of input layer nodes	16
	Number of hidden layer nodes	32
	Number of output layer nodes	12
	Activation function	ReLU
	Activation function output	Linear
	Optimization function	Adam
	Learning rate	0.01
	Epochs of training	250
	Batch size	16
DNN(Peak)	Hidden layers	1
	Number of input layer nodes	16
	Number of hidden layer nodes	32
	Number of output layer nodes	1
	Activation function	ReLU
	Activation function output	Linear
	Optimization function	Adam
	Learning rate	0.01
	Epochs of training	150
	Batch size	48
LSTM(Energy)	Hidden layers	2
	Number of input layer nodes	50
	Number of hidden layer nodes	50
	Number of output layer nodes	12
	Activation function	ReLU
	Optimization function	Adam
	Learning rate	0.001
	Epochs of training	180
	Batch size	72
LSTM(Peak)	Hidden layers	1
	Number of input layer nodes	10
	Number of hidden layer nodes	64
	Number of output layer nodes	1
	Activation function	ReLU
	Optimization function	Rmsprop
	Learning rate	0.001
	Epochs of training	200
	Batch size	32
CNN-LSTM(Energy)	1D-CNN filter size	34
	Number of dense layer nodes	32
	Number of LSTM layer nodes	19
	Number of output layer nodes	12
	Activation function	LeakyRelu
	Negative slope coefficient	0.3
	Optimization function	Rmsprop
	Learning rate	0.001
	Epochs of training	300
	Batch size	40
CNN-LSTM(Peak)	1D-CNN filter size	41
	Number of dense layer nodes	22
	Number of LSTM layer nodes	25
	Number of output layer nodes	1
	Activation function	Relu
	Optimization function	Rmsprop
	Learning rate	0.001
	Epochs of training	80
	Batch size	72
AttentionLSTM(Energy)	Number of LSTM layer nodes	64
	Number of Attention layer nodes	40
	Number of Dense layer nodes	12
	Activation function	ReLU
	Optimization function	Adam
	Learning rate	0.001
	Epochs of training	120
	Batch size	24

Model	Parameter	Value
AttentionLSTM(Peak)	Number of LSTM layer nodes	64
	Number of Attention layer nodes	32
	Number of Dense layer nodes	1
	Activation function	ReLU
	Optimization function	Adam
	Learning rate	0.001
	Epochs of training	200
	Batch size	16

REFERENCES

[1] R. J. Hyndman and S. Fan, "Density forecasting for long-term peak electricity demand," *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 1142–1153, May 2010.

[2] Y. Li and B. Jones, "The use of extreme value theory for forecasting long-term substation maximum electricity demand," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 128–139, Jan. 2020.

[3] F. Elkarmi, *Power System Planning Technologies and Applications: Concepts, Solutions and Management*. Hershey, PA, USA: IGI Glob., 2012.

[4] A. Matar, "Does electricity consumption impacting financial development? Wavelet analysis," *Future Bus. J.*, vol. 6, no. 1, pp. 1–9, Dec. 2020.

[5] M. Shahbaz, T. H. Van Hoang, M. K. Mahalik, and D. Roubaud, "Energy consumption, financial development and economic growth in India: New evidence from a nonlinear and asymmetric analysis," *Energy Econ.*, vol. 63, pp. 199–212, 2017.

[6] J. Pérez-García and J. Moral-Carcedo, "Analysis and long term forecasting of electricity demand trough a decomposition model: A case study for Spain," *Energy*, vol. 97, pp. 127–143, 2016.

[7] B. Ang, H. Wang, and X. Ma, "Climatic influence on electricity consumption: The case of Singapore and Hong Kong," *Energy*, vol. 127, pp. 534–543, 2017.

[8] H. Hamedmoghadam, N. Joorabloo, and M. Jalili, "Australia's long-term electricity demand forecasting using deep neural networks," 2018, *arXiv:1912.1801.02148*.

[9] A. O. Melodi, J. A. Momoh, and O. M. Adeyanju, "Probabilistic long term load forecast for Nigerian bulk power transmission system expansion planning," in *Proc. IEEE PES PowerAfrica*, 2016, pp. 301–305.

[10] R. Nezzar, N. Farah, M. Khadir, and L. Chouireb, "Mid-long term load forecasting using multi-model artificial neural networks," *Int. J. Elect. Eng. Inform.*, vol. 8, pp. 389–401, Jun. 2016.

[11] T. Hong, J. Wilson, and J. Xie, "Long term probabilistic load forecasting and normalization with hourly information," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 456–462, Jan. 2014.

[12] E. M. de Oliveira and F. L. Cyrino Oliveira, "Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods," *Energy*, vol. 144, pp. 776–788, 2018.

[13] Y. Li, D. Han, and Z. Yan, "Long-term system load forecasting based on data-driven linear clustering method," *J. Modern Power Syst. Clean Energy*, vol. 6, no. 2, pp. 306–316, 2018.

[14] S. Zheng, G. Huang, X. Zhou, and X. Zhu, "Climate-change impacts on electricity demands at a metropolitan scale: A case study of Guangzhou, China," *Appl. Energy*, vol. 261, Mar. 2020, Art. no. 114295.

[15] D. Willingham, "Long term energy forecasting with econometrics in MATLAB," 2022. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/49279-long-term-energy-forecasting-with-econometrics-in-matlab

[16] J. Xie, T. Hong, and C. Kang, "From high-resolution data to high-resolution probabilistic load forecasts," in *Proc. IEEE/PES Transmiss. Distrib. Conf. Expo.*, 2016, pp. 1–5.

[17] M. Dong and L. Grumbach, "A hybrid distribution feeder long-term load forecasting method based on sequence prediction," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 470–482, Jan. 2020.

[18] B. Oh, D.-H. Lee, and D. Lee, "Oil-price based long-term hourly system marginal electricity price scenario generation," *IEEE Access*, vol. 10, pp. 25051–25061, 2022.

[19] L. Li, C. J. Meinrenken, V. Modi, and P. J. Culligan, "Short-term apartment-level load forecasting using a modified neural network with selected auto-regressive features," *Appl. Energy*, vol. 287, 2021, Art. no. 116509.

[20] C. Chen et al., "Improving protein-protein interactions prediction accuracy using XGBoost feature selection and stacked ensemble classifier," *Comput. Biol. Med.*, vol. 123, 2020, Art. no. 103899.

[21] A. Ogunleye and Q. -G. Wang, "XGBoost model for chronic kidney disease diagnosis," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 6, pp. 2131–2140, Nov./Dec. 2020.

[22] Y. Wang et al., "Short-term load forecasting of industrial customers based on SVMD and XGBoost," *Int. J. Elect. Power Energy Syst.*, vol. 129, 2021, Art. no. 106830.

[23] Y. Liu, H. Luo, B. Zhao, X. Zhao, and Z. Han, "Short-term power load forecasting based on clustering and XGBoost method," in *Proc. IEEE 9th Int. Conf. Softw. Eng. Service Sci.*, 2018, pp. 536–539.

[24] B.-W. Liu, D. Xu, and Y. He, "Short-term power load forecasting during the COVID-19 pandemic using XGBoost and copula theory," in *Proc. IEEE China Automat. Congr.*, 2021, pp. 5789–5794.

[25] M. Dong, J. Shi, and Q. Shi, "Multi-year long-term load forecast for area distribution feeders based on selective sequence learning," *Energy*, vol. 206, 2020, Art. no. 118209.

[26] "Aggregated price and demand data," 2022. [Online]. Available: https://aemo.com.au/en/energy-systems/electricity/national-electricity-market-nem/data-nem/aggregated-data

[27] T. Chen and C. Guestrin, "XGBoost," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794.

[28] G. Suo, L. Song, Y. Dou, and Z. Cui, "Multi-dimensional short-term load forecasting based on XGBoost and fireworks algorithm," in *Proc. IEEE 18th Int. Symp. Distrib. Comput. Appl. Bus. Eng. Sci.*, 2019, pp. 245–248.

[29] S. B. Taieb, G. Bontempi, A. F. Atiya, and A. Sorjamaa, "A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7067–7083, 2012.

[30] "Australian bureau of statistics data by region," 2022. [Online]. Available: https://dbr.abs.gov.au/

[31] Climate data online, 2022. [Online]. Available: http://www.bom.gov.au/climate/data/index.shtml?bookmark=200

[32] "AccuWeather," 2022. [Online]. Available: https://www.accuweather.com/en/au/sydney/22889/july-weather/22889

[33] M. S. Kandil, S. M. El-Debeiky, and N. E. Hasanien, "Long-term load forecasting for fast developing utility using a knowledge-based expert system," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 491–496, May 2002.

[34] S. Awaworyi Churchill and K. Ivanovski, "Electricity consumption and economic growth across Australian states and territories," *Appl. Econ.*, vol. 52, no. 8, pp. 866–878, 2020.

[35] M. Waite, E. Cohen, H. Torbey, M. Piccirilli, Y. Tian, and V. Modi, "Global trends in urban electricity demands for cooling and heating," *Energy*, vol. 127, pp. 786–802, 2017.

[36] A. Altmann, L. Toloşi, O. Sander, and T. Lengauer, "Permutation importance: A corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, 2010.

[37] "Distributed (deep) machine learning community XGBoost," 2021. [Online]. Available: https://github.com/dmlc/xgboost

[38] P. Liashchynskyi and P. Liashchynskyi, "Grid search, random search, genetic algorithm: A big comparison for NAS," 2019., *arXiv:1912.06059*

[39] "Queensland and NSW put the heat on demand," 2017. [Online]. Available: https://www.energycouncil.com.au/analysis/queensland-and-nsw-put-the-heat-on-demand/

[40] Y. He, Y. Qin, X. Lei, and N. Feng, "A study on short-term power load probability density forecasting considering wind power effects," *Int. J. Elect. Power Energy Syst.*, vol. 113, pp. 502–514, 2019.

[41] P. Wang, B. Liu, and T. Hong, "Electric load forecasting with recency effect: A Big Data approach," *Int. J. Forecasting*, vol. 32, no. 3, pp. 585–597, 2016.

[42] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. Melbourne, VIC, Australia: OTexts, 2018.

[43] Y. Aslan, S. Yavasca, and C. Yasar, "Long term electric peak load forecasting of Kutahya using different approaches," *Int. J. Tech. Phys. Problems Eng.*, vol. 3, no. 2, pp. 87–91, 2011.

[44] F. Alasali, K. Nusair, L. Alhmoud, and E. Zarour, "Impact of the COVID-19 pandemic on electricity demand and load forecasting," *Sustainability*, vol. 13, no. 3, 2021, Art. no. 1435.

[45] G. Dudek and P. Pełka, "Pattern similarity-based machine learning methods for mid-term load forecasting: A comparative study," *Appl. Soft Comput.*, vol. 104, 2021, Art. no. 107223.

[46] S. Mirasgedis et al., "Models for mid-term electricity demand forecasting incorporating weather influences," *Energy*, vol. 31, no. 2/3, pp. 208–227, 2006.

[47] W. Jiang, X. Wu, Y. Gong, W. Yu, and X. Zhong, "Holt–Winters smoothing enhanced by fruit fly optimization algorithm to forecast monthly electricity consumption," *Energy*, vol. 193, 2020, Art. no. 116779.

[48] N. A. Mohammed and A. Al-Bazi, "An adaptive backpropagation algorithm for long-term electricity load forecasting," *Neural Comput. Appl.*, vol. 34, no. 1, pp. 477–491, 2022.

[49] D. Ali, M. Yohanna, M. Puwu, and B. Garkida, "Long-term load forecast modelling using a fuzzy logic approach," *Pacific Sci. Rev. A: Natural Sci. Eng.*, vol. 18, no. 2, pp. 123–127, 2016.

[50] M. Elkamel, L. Schleider, E. L. Pasiliao, A. Diabat, and Q. P. Zheng, "Long-term electricity demand prediction via socioeconomic factors–A machine learning approach with florida as a case study," *Energies*, vol. 13, no. 15, 2020, Art. no. 3996.

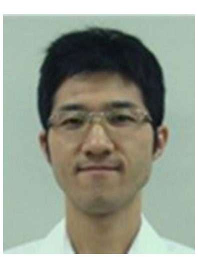
[51] E. İlseven and M. Göl, "Medium-term electricity demand forecasting based on MARS," in *Proc. IEEE PES Innov. Smart Grid Technol. Conf. Europe*, 2017, pp. 1–6.

[52] P. Bunnoon, "Mid-term load forecasting based on neural NetworkAlgorithm: A comparison of models," *Int. J. Comput. Elect. Eng.*, vol. 3, no. 4, p. 600–605, 2011.

[53] H. Daneshi, M. Shahidehpour, and A. L. Choobbari, "Long-term load forecasting in electricity market," in *Proc. IEEE Int. Conf. Electro/Inf. Technol.*, 2008, pp. 395–400.

[54] O. Rubasinghe, X. Zhang, T. K. Chau, Y. Chow, T. Fernando, and H. H.-C. Iu,, "A novel sequence to sequence data modelling based CNN-LSTM algorithm for three years ahead monthly peak load forecasting," *IEEE Trans. Power Syst.*, early access, Apr. 28, 2023, doi: 10.1109/TPWRS.2023.3271325.

Tingze Zhang (Student Member, IEEE) received the B.E. degree in electronic engineering from Dalian Polytechnic University, Dalian, China, in 2017, and the MPE degree in electrical and electronic engineering in 2020 from The University of Western Australia (UWA), Perth, WA, Australia, where he is currently working toward the Ph.D. degree in electrical and electronic engineering. His research interests include load and renewable energy forecasting, energy storage management and power system operation and planning under high penetration of renewables.

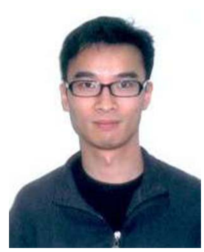


Xinan Zhang (Member, IEEE) received the B.E. degree in electrical engineering and automation from Fudan University, Shanghai, China, in 2008 and the Ph.D. degree from Nanyang Technological University (NTU), Singapore, in 2014. Then, he was a Postdoc Researcher with NTU and the University of New South Wales, Sydney, NSW, Australia, from 2014 to 2017. From 2017 to 2019, he was a Lecturer with NTU. In September 2019, he joined the University of Western Australia as a Senior Lecturer. His research interests include electrical machine drives, control and modulation of power electronic converters and management of hybrid energy storage systems.

Osaka Rubasinghe (Student Member, IEEE) received the B. Eng. (First Class Hons.) in electrical power engineering from Curtin University, Perth, WA, Australia, in 2019. He is currently working toward the Ph.D. degree with The University of Western Australia, Perth. He was an Electrical Engineer with the Mining, Oil and Gas Industry. His research interests include the renewable energy integration and forecasting, energy storage management, and power system analysis.

Yulin Liu (Student Member, IEEE) received the B.E. degree in automation from the Southwest University (SWU), Chongqing, China, in 2018 and the MPE degree in electrical and electronic engineering in 2020 from The University of Western Australia (UWA), Perth, WA, Australia, where he is currently working toward the Ph.D. degree in electrical engineering. His research interests include advanced model-based control and artificial intelligence control in grid-connected fuel cells.

Yau Hing Chow received the B.Eng. degree in electrical engineering from Curtin University, Perth, WA, Australia, in 2000. In 2005, he joined Western Power. He is currently a Principal Planning Engineer with Western Power in the Grid Transformation Area.



Herbert H. C. Iu (Senior Member, IEEE) received the B.Eng. (Hons.) degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong, in 1997 and the Ph.D. degree from the Hong Kong Polytechnic University, Hong Kong, in 2000. In 2002, he joined the School of Electrical, Electronic and Computer Engineering, The University of Western Australia, Perth, WA, Australia, as a Lecturer, where he is currently a Professor. He has authored or coauthored more than 100 papers in these areas. His research interests include power electronics, renewable energy, nonlinear dynamics, current sensing techniques, and memristive systems. He was the recipient of the two IET Premium Awards in 2012 and 2014. In 2014, he was also the recipient of the Vice-Chancellor's Mid-Career Research Award, and 2019 IEEE Transactions on Very Large Scale Integration Systems Best Paper Award. He is currently an Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, IEEE TRANSACTIONS ON POWER ELECTRONICS, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, and the Editor of IEEE TRANSACTIONS ON SMART GRIDS. He is the Co-editor of *Control of Chaos in Nonlinear Circuits and Systems* (Singapore: World Scientific, 2009) and a co-author of *Development of Memristor Based Circuits* (Singapore: World Scientific, 2013).

Tyrone Fernando (Senior Member, IEEE) received the bachelor of engineering degree (with Hons.) and the degree of doctor of philosophy from the University of Melbourne, Melbourne, VIC, Australia, in 1990 and 1996 respectively. In 1996 he joined the Department of Electrical Electronic and Computer Engineering, University of Western Australia, Perth, WA, Australia, where he is currently a Professor. His research interests include power systems, renewable energy and control theory. He was an Associate Editor for IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE and the Guest Editor of the Journal of *Optimal Control Applications and Methods*. He is an Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II and IEEE ACCESS.