

# A comparative climate-resilient energy design: Wildfire Resilient Load Forecasting Model using multi-factor deep learning methods

Weijia Yang<sup>\*</sup>, Sarah N. Sparrow, David C.H. Wallom

University of Oxford, Department of Engineering Science, Oxford, OX1 3PJ, United Kingdom

## ARTICLE INFO

### Keywords:

Climate resilience  
Deep learning  
Distribution network  
Extreme weather  
Load forecast

## ABSTRACT

Power grid damage and blackouts are increasing with climate change. Load forecasting methods that integrate climate resilience are therefore essential to facilitate timely and accurate network reconfiguration during periods of extreme stress. Our paper proposes a generalised Wildfire Resilient Load Forecasting Model (WRLFM) to predict electricity load based on operational data of a Distribution Network (DN) in Australia during wildfire seasons in 2015–2020. We demonstrate that load forecasting during wildfire seasons is more challenging than during non-wildfire seasons, motivating an imperative need to improve forecast performance during wildfire seasons. To develop the robust WRLFM, comprehensive comparative analyses were conducted to determine proper Machine Learning (ML) forecast structures and methods for incorporating multiple factors. Bi-directional Gated Recurrent Unit (Bi-GRU) and Vision Transformer (ViT) were selected as they performed the best among all 13 recently trending ML methods. Multi-factors were incorporated to contribute to forecast performance, including input sequence structures, calendar information, flexible correlation-based temperature conditions, and categorical Fire Weather Index (FWI). High-resolution categorical FWI was used to build a forecasting model with climate resilience for the first time, significantly enhancing the average stability of forecast performances by 42%. A sensitivity analysis compared data set patterns and model performances during wildfire and non-wildfire seasons. The improvement rate of load forecasting performance during wildfire seasons was more than two times greater than in non-wildfire seasons. This indicates the significance and effectiveness of applying the WRLFM to improve forecast accuracy under extreme weather risks. Overall, the WRLFM reduces the Mean Absolute Percentage Error (MAPE) of the forecast by 14.37% and 20.86% for Bi-GRU and ViT-based models, respectively, achieving an average forecast MAPE of around 3%.

## 1. Introduction

On the path to Net Zero, power systems face significantly growing uncertainty due to generation diversification and demand electrification. Furthermore, the escalating incidence of extreme weather events poses substantial challenges to the reliable and resilient operation of energy systems. In the case of generation, not only does the output of all types of renewable generators become increasingly intermittent, but conventional power generation may also experience substantial change. Fossil fuel and nuclear-based power stations are affected due to shortages and higher temperatures of water supplies used in cooling. For example, the frequency of climate-related nuclear plant outages has increased nearly eightfold since the 1990s, as reported by the American Nuclear Society [1]. On the consumer side, demand surges during periods of extreme weather, particularly for cooling and heating applications. For instance, it is projected that the peak electricity demand in the U.S. will increase by 7.2% by the end of this century due to

increasing extreme temperature events [2]. Therefore, a pressing need exists to develop intelligent and climate-resilient energy forecasting tools to help local energy operators properly dispatch power in advance or swiftly restore power connections when facing extreme weather.

With the combined risks from climate change and the ageing power systems, extreme weather events increasingly cause unintentional blackouts, resulting in significant economic and social costs. Among these consequences, power grid damage due to wildfire has become a growing concern, e.g., reduced line capacity, conductor line sag, and physical infrastructure damage. For example, various regions across Australia confront annual wildfire seasons at varying times. According to the Australian Bureau of Meteorology [3], the bushfire season typically spans from December to May in Southern Australia, May to October in Northern Australia, and August to March in Central Australia. Southeastern Australia, in particular, stands out as one of

<sup>\*</sup> Corresponding author.

E-mail address: [weijia.yang@eng.ox.ac.uk](mailto:weijia.yang@eng.ox.ac.uk) (W. Yang).

URL: <https://eng.ox.ac.uk/people/weijia-yang/> (W. Yang).

<https://doi.org/10.1016/j.apenergy.2024.123365>

Received 19 December 2023; Received in revised form 23 March 2024; Accepted 30 April 2024

Available online 20 May 2024

0306-2619/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

the most fire-prone areas globally. During the 2019–2020 wildfire season, numerous wildfires erupted in Australia, with a notable impact in Victoria State, leading to 33 fatalities and 19 million hectares of land destruction [4]. The National Emergency Management Agency declared the commencement of the 2019–2020 bushfire season for Victoria on 21<sup>st</sup> November 2019 [5]. Furthermore, during the New Year Period in New South Wales, over 20,000 households experienced power outages [6]. Considering the geographical location and documented prevalent wildfire risk periods, the annual wildfire season in Victoria, Australia, is defined as November to March in our study.

As extreme wildfire risks are characterised by high temperature, low humidity and high wind conditions, our study picks multi-year wildfire seasons in Australia as a case study, aiming to develop the Wildfire Resilient Load Forecasting Model (hereafter the WRLFM) at the Distribution Network (DN) level. In terms of extreme weather, it is a regional and short-term phenomenon rather than a large-area and long-term climate pattern. Therefore, load forecasting at the transmission network or higher national level is out of scope.

In our paper, we develop the WRLFM through a two-stage approach. First, we assess, compare, and adapt trending load forecasting models to identify the most effective ones. Subsequently, we incorporate multiple factors into the model, updating some previously employed variables to enhance flexibility. Additionally, we introduce the Canadian Forest Service Fire Weather Index Rating System (FWI), which improves the overall load forecast accuracy and resilience.

There are two primary categories of load forecasting methods: traditional statistical methods and Artificial Intelligence (AI) based methods. Commonly used statistical techniques include Multiple Linear Regression (MLR), Autoregressive Integrated Moving Average (ARIMA), Semi-parametric Additive Model (SAM), and Exponential Smoothing Model (ESM) [7,8]. However, these methods face challenges capturing demand fluctuations induced by non-linear characteristics during extreme weather events. To address non-linearity, data-driven methods such as shallow Machine Learning (ML) and Deep Learning (DL) have emerged in recent research. Common shallow ML methods encompass Support Vector Machine (SVM), Random Forest (RF) and Extreme Gradient Boosting (XGB) [9]. DL methods have further evolved in two main directions: Recurrent Neural Network (RNN) specialised in handling sequential data and Convolutional Neural Network (CNN) proficient in capturing spatial data features, such as image recognition [10,11]. While RNNs and CNNs enable machines to glean information and patterns from extensive data sets, Deep Reinforcement Learning (deep RL) has also found application in several energy management scenarios, facilitating system operators in making optimised decisions [12]. In addition to the advanced load forecasting models, RL techniques and optimisation strategies can be integrated to enhance system efficiency and associated benefits. Such applications include scenarios like hybrid electric vehicle integrated energy systems [13] and wind-thermal-bundled power systems [14], among others.

In addition to comparing and adapting various forecast models, the inclusion of multiple factors is crucial for enhancing forecast accuracy. Few studies have investigated the internal relationship between various factors and load behaviours before incorporating them as model inputs. For example, prior studies [15,16] typically used 24-hour ahead load data as the default input due to its daily periodicity but had yet to explore the optimised design of input structures. Calendar effects were applied to load forecasting studies in different classification ways, such as working and non-working day characteristics in [17]. While most studies adopt a specific method for calendar effects, there is room to explore and compare various approaches. Furthermore, multiple climate-related factors, such as temperature, humidity, wind speed, and precipitation, have been considered before. Daily mean or instantaneous temperature was commonly used [18–21]. However, the relationship between load and temperature varies by region, necessitating a thorough investigation into the efficient utilisation of temperature and its temporal dynamics. Our paper comprehensively evaluates and

compares existing approaches to incorporate multi-factors, including input sequence lengths, calendar effects, and temperatures. We aim to optimise the utilisation of each factor. Additionally, we introduce the use of the FWI, a factor calculated based on temperature, relative humidity, wind, and rainfall, to enhance the performance of the WRLFM during extreme wildfire seasons [22].

This paper introduces a novel climate-resilient multi-factor load forecasting model, known as the WRLFM, based on RNN and CNN-derived methods. The WRLFM has been developed using operational data from a DN in Victoria, Australia, during the wildfire seasons in 2015–2020. Initially, we compare and adapt 13 contemporary ML and DL models to identify the most effective ones with suitable input structures. Subsequently, we explore the internal relationship between energy consumption behaviours and various factors. Each factor's contribution is optimised, including the input matrix structure, calendar effects, correlation-based preceding temperature conditions, and the innovative inclusion of a high-resolution factor: the Fire Weather Index. Finally, the generalised WRLFM is further evaluated using the wildfire and non-wildfire data sets. This evaluation underscores the challenges associated with load forecasting during extreme stress periods and the significance of our work.

The rest of this paper is organised as follows. Section 2 presents the core methodology used to develop the WRLFM, including pre-processing steps to generate the input matrix and an in-depth discussion of how each factor contributes to improved forecast performance. Additionally, we review the mechanisms of each ML/DL method and the metrics used for performance evaluation. Section 3 introduces the background and data acquisition methods for our case study. Our model's generalisation is outlined in Section 4 through five key steps based on the performance enhancements. Section 5 summarises the key findings and advises future work.

## 2. Methodology

In Fig. 1, the main steps to develop the WRLFM are described in a flow chart on the left-hand side, with the detail of how each factor contributes to the model resilience and accuracy improvement shown on the right. The pre-processing methods for each factor and details to generate the input matrix are explained in Section 2.1. Then, the input matrix is fed into various trending ML/DL methods to forecast sequential electricity demand. The mechanism of each forecast structure is reviewed in detail in Section 2.2. Finally, the metrics used for performance evaluation are discussed in Section 2.3.

### 2.1. Pre-processing steps to generate the multi-factor input matrix

As displayed in Fig. 1, the input matrix is created through four stages: input structure selection, classification of calendar effects, consideration of flexible preceding temperature conditions, and introduction of the Fire Weather Index. In Step 5, the final generalised model is tested with data sets corresponding to wildfire and non-wildfire seasons to ascertain the effectiveness and significance of applying the proposed methods during periods of extreme weather.

#### Step 1 Input sequence structure

Step 1 comprises two distinct functions: comparing 13 trending methods and selecting the appropriate input sequence length for each method. The structure and mechanism of each AI-based forecasting model will be discussed in Section 2.2. In a sequential forecasting scenario, historical data with a fixed length must be defined to predict the next one or several data points. In Step 1, we exponentially vary the input sequence length to identify the optimal input length  $n$ , ensuring both high accuracy and acceptable computational burden. The column number of the input matrix is then determined by the input sequence length, set as  $n$  columns.

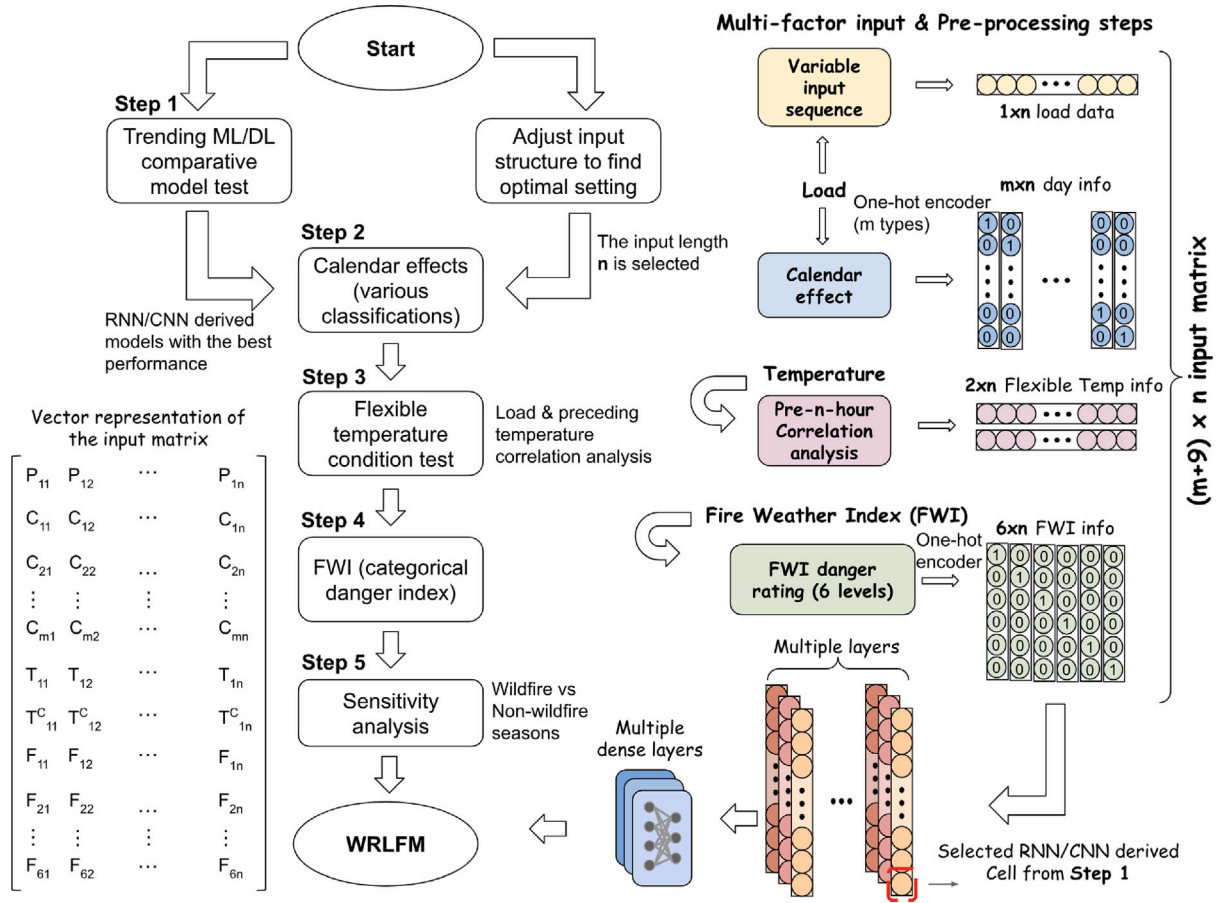


Fig. 1. Main structures of the Wildfire Resilient Load Forecasting Model (WRLFM). The vector representation of the input matrix is placed in the bottom left corner.

## Step 2 Calendar effect

In Step 2, calendar information is categorised into  $m$  types using a One-hot encoder, which converts categorical day-type features into numeric variables [23]. The principle of One-hot encoding is as follows: a categorical variable  $x$  with  $n$  discrete values is expressed as  $x_1, x_2, \dots, x_n$ . Each category is transformed into a vector  $v$ , where every element of the vector is zero except for the  $i^{th}$ , which has a value of 1. For each sampling point, the calendar vector is transposed into a column vector according to its category, arranged as an  $m \times n$  matrix, and placed in the middle of the input matrix.

## Step 3 Flexible preceding correlation based temperature

As depicted in Fig. 1, temperature information occupies two rows within the input matrix. The first row comprises a sequence of the preceding  $n$  temperature data for each energy sampling point, akin to extracting the  $1 \times n$  load data in Step 1. The second row is organised based on flexible preceding temperature conditions determined through dynamic correlation analyses. Recognising that temperature changes may induce fluctuations in energy demand, we perform correlation analyses involving preceding instantaneous, maximum, and average temperatures with energy consumption. These analyses sort out the potential flexible temperature conditions to improve the WRLFM performance. The temperature condition with the strongest correlation to energy consumption data is selected. For the second row, the time series is shifted such that the final value of the row is the point of the strongest correlation, with the preceding values being the time points prior to this.

Fig. 2 illustrates the results of a flexible correlation analysis between energy demand and various preceding temperatures, using operational data from the Horsham DN in Victoria, Australia, during the wildfire seasons in 2015–2020. The analysis considers the instantaneous

temperatures at different hours ahead, the maximum and the average temperatures over various preceding hours, with increments of 0.5 h, ranging from 0.5 to 24 h ahead. Correlation coefficients, the 1<sup>st</sup> order  $r^2$  value, and the 2<sup>nd</sup> order  $r^2$  value are plotted for load and temperature. While the correlation coefficient represents the strength of the relationship between two variables, the  $r^2$  value explains the extent to which the variance of one variable affects another. Some correlation coefficients exceed 0.7 for various instantaneous and average temperature lead times. For example, in Fig. 2(a), the blue-marked point denotes that the instantaneous temperature at 5 h ahead exhibits the highest correlation with energy demand among all preceding 24-hour conditions (with a correlation coefficient of 0.7224). The 2<sup>nd</sup> order  $r^2$  value curves follow a similar trend to the correlation coefficient, peaking at 0.647 under the pre-7-hour average temperature condition. According to preliminary work of this paper [24], a comparative analysis was conducted between the temperature conditions yielding the highest correlation coefficients and  $r^2$  values, revealing that the optimal correlation condition holds a greater potential to improve forecast performance. Thus, the temperature condition leading to the strongest correlation among the three temperature types is identified and selected as the method for processing temperature information, placing it in the second row to potentially enhance its contribution to forecast performance.

## Step 4 High-resolution Fire Weather Index

As our case study aims to design a load forecasting tool resilient to wildfire risks, we introduce a novel feature, the Fire Weather Index (FWI), to aid the WRLFM. The FWI is a meteorological index designed to estimate the risk of fire globally. It encompasses the effects of fuel moisture and wind impact on wildfire spread. A higher FWI indicates a greater likelihood of fire initiation and a more intense fire risk. Illustrated in Fig. 3, FWI calculation necessitates five climate variables

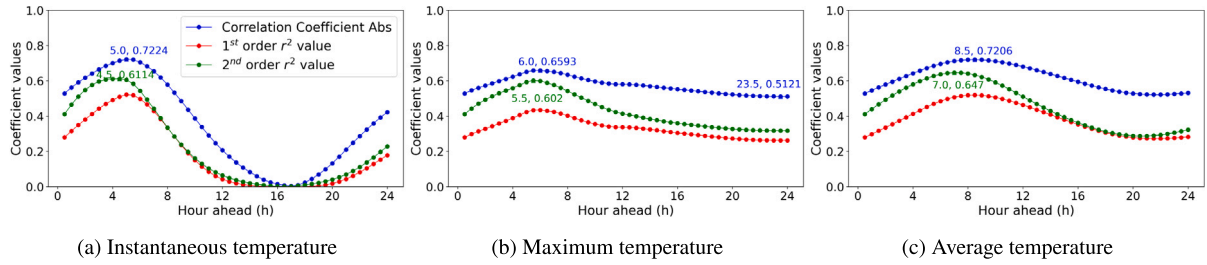


Fig. 2. Correlation analyses for load and various preceding temperatures of Horsham, Victoria, Australia during wildfire seasons in 2015–2020. (a) Instantaneous temperatures from 0.5 to 24 h ahead (b) Maximum temperatures within the preceding 0.5 to 24 h (c) Average temperatures within the preceding 0.5 to 24 h.

as initial inputs for subsequent computation of moisture codes and fire behaviour indices. These variables include temperature, 10 m wind speed (u direction), 10 m wind speed (v direction), total precipitation, and relative humidity.

The off-the-shelf FWI is available with a daily temporal resolution and a spatial resolution of  $0.25^\circ \times 0.25^\circ$  from the European Centre for Medium-Range Weather Forecasts (ECMWF) Copernicus database [25]. However, these resolutions are inconsistent with energy load requirements; higher temporal and spatial resolutions are needed for the load forecasting tool. The procedures for calculating the high-resolution FWI are detailed in Fig. 3. Firstly, NetCDF files for five climate variables are obtained from the ECMWF Reanalysis v5 (ERA5) hourly databases on single and pressure levels. Subsequently, each variable is refined with a 15-minute temporal resolution and a  $0.01^\circ \times 0.01^\circ$  spatial resolution using the interpolation tool of the Climate Data Operator (CDO) [26]. As the data are in NetCDF format, they are then transformed into CSV format, facilitating subsequent calculations. A unit conversion procedure is carried out since the original units for variables differ from the unit requirements in the primary FWI calculation process.

As shown at the bottom of Fig. 3, the primary FWI calculation process consists of three stages. Firstly, four fire observational variables are required, each with specific units. Subsequently, moisture codes are computed based on various combinations of these observational variables. Moisture codes serve as numeric ratings for the moisture content of the forest floor and other deceased organic matter, with values increasing as moisture content decreases. The Fine Fuel Moisture Code (FFMC), the Duff Moisture Code (DMC), and the Drought Code (DC) indicate the moisture levels of litter and other fine fuels, loosely compacted organic layers of moderate depth, and deep, compact organic layers, respectively. The remaining three fire behaviour indices are the Initial Spread Index (ISI), the Buildup Index (BUI), and the Fire Weather Index (FWI), signifying the rate of fire spread, the fuel available for combustion, and the frontal fire intensity, respectively. Detailed calculation methods and coding references for generating FWI can be found in [27,28], respectively.

To validate the effectiveness of our methods in generating high-resolution FWI, we compared the daily FWI downloaded directly with the 15-minute resolution FWI curves we calculated. The daily FWI, available for download, aims to capture the peak burning condition of a day, typically around 16:00 local standard time (LST) [29]. Fig. 4 illustrates the results of the Quantile–Quantile (Q–Q) comparative plots for the downloaded daily FWI and the calculated FWI at 4 pm, as well as the calculated daily peak FWI. Fig. 4(a) demonstrates that the calculated FWIs at 4 pm do not align well with the distribution of downloaded daily FWI. This discrepancy arises because peak burning conditions may occur at times different from 16:00 LST in Victoria, Australia. In Fig. 4(b), the distribution of the calculated peak daily FWI fits very well with the downloaded daily FWI (peak burning condition), proving the effectiveness and reliability of using our methods to generate high-resolution FWI curves. Fig. 5 depicts the distribution and variability of the calculated high-resolution FWI by plotting the mean FWI curve and percentile lines. The blue rug plot at the bottom demonstrates the density distribution of the peak FWI times, indicating

Table 1

FWI Danger Rating levels (upper bound excluded).

Danger Rating	FWI range
Very low	(0, 5.2)
Low	(5.2, 11.2)
Moderate	(11.2, 21.3)
High	(21.3, 38.0)
Very high	(38.0, 50.0)
Extreme	(50.0, $+\infty$ )

that FWIs are most likely to peak around 6 pm. Again, this underscores the importance of using high-resolution FWI to represent the fire risk trend within a day rather than directly using the fixed daily FWI value.

While the FWI is a numeric index, the FWI Danger Rating (DR) is classified into six levels, as illustrated in Table 1. Similar to calendar effects, FWI information is integrated into the input matrix using a One-hot encoder. The categorical FWI information is transformed into a  $6 \times n$  matrix, positioned at the bottom of the input matrix.

Finally, the input matrix, with a size of  $(m + 9) \times n$ , is composed, where  $m$  represents the category number of calendar effects, and  $n$  denotes the optimal length of historical input (look\_back). The vector representation of the input matrix is placed in the bottom left corner of Fig. 1. P represents energy demand, C denotes calendar-related information, T indicates temperature,  $T^C$  is the temperature condition with high correlations, and F stands for Fire Weather Index-related input. The comprehensive input matrix is then fed into the selected multi-layer RNN-derived or CNN-related forecast structures to generate the load forecast.

## 2.2. Trending ML/DL based forecast methods

This section reviews the mechanisms of trending AI-based methods designed for addressing sequential data forecasting tasks. The description will be expanded in detail, commencing with shallow ML methods, progressing to basic DL methods, and then further exploring RNN-derived and CNN-related methods.

### Shallow machine learning: SVM

The Support Vector Machine (SVM) was first proposed by Vapnik [30]. This technique involves mapping the data into an N-dimensional space (N features) and constructing an optimal separating hyperplane to classify all data points. The optimal hyperplane is selected to be the one with the maximum margin, meaning the distance between the hyperplane and each data group is maximised. Hinge loss is employed as the loss function to aid in maximising the margin. The kernels (linear, polynomial, radial basis function) and parameters (regularisation, gamma, degree) should be adjusted in different cases [31].

### Shallow machine learning: RF

Random Forest (RF) is an ensemble learning method that combines the predictions of multiple decision trees to enhance overall performance and reduce overfitting. Each decision tree in the forest is



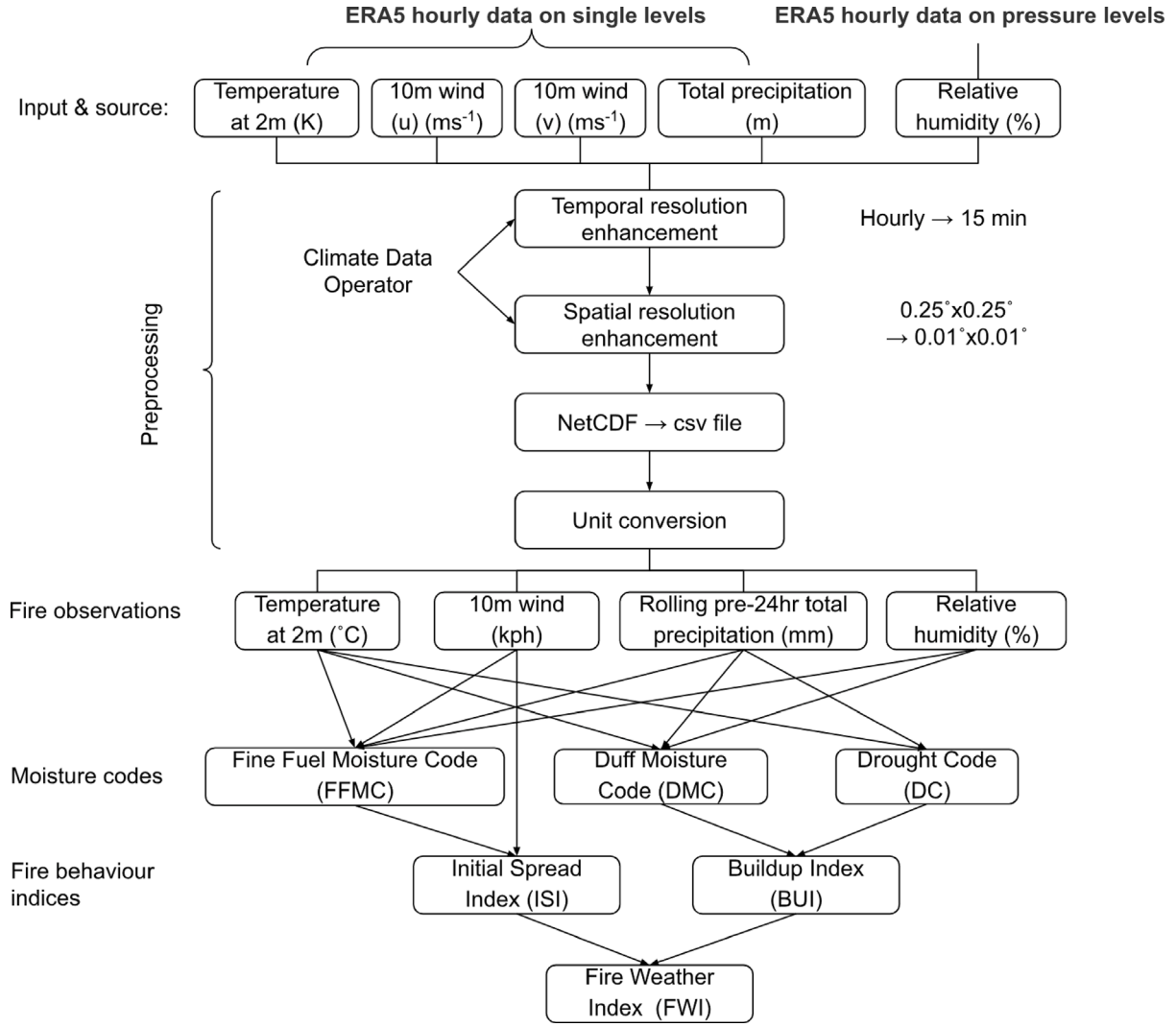


Fig. 3. Methodology to generate Fire Weather Index (FWI) data with high spatial and temporal resolutions.

trained on random subsets of the data and features at each split. This approach enhances the diversity of the systems and improves classification performance. For the RF model, three parameters need adjustment and selection: the input feature period, the ratio of randomly selected features, and the number of predictors [32].

#### Shallow machine learning: XGB

Operating on the gradient boosting framework, Extreme Gradient Boosting (XGB) is renowned for its efficiency and effectiveness in generating accurate predictions. XGB sequentially builds a series of decision trees, introducing a new tree in each iteration to minimise the loss function, with each tree correcting the errors of its predecessors. Parameters, including the learning rate, maximum depth of trees, and regularisation terms, can be tuned to enhance forecast performance [33].

#### DL basis: MLP

A Multilayer Perceptron (MLP) is a category of artificial neural networks characterised by a minimum of three layers of nodes: an input layer, one or more hidden layers, and an output layer. Neurons serve as the fundamental computational units within each layer. Each neuron receives input, processes it utilising weights adjusted during training, applies an activation function, and generates an output. The activation function introduces non-linearity into the network, with common activation functions encompassing the sigmoid, Hyperbolic Tangent (tanh), and Rectified Linear Unit (ReLU). The loss function gauges the disparity between the predicted output and the actual target. Throughout the

training process, the objective is to minimise this loss by iteratively adjusting the weights through optimisation algorithms such as gradient descent. The fine-tuning of activation functions, hidden layer numbers, and the quantity of neurons in each layer is imperative for effectively addressing diverse data sets [34,35].

#### RNN-derived methods

Recurrent Neural Network (RNN) constitutes a category of DL methods that have demonstrated notable effectiveness in tasks dealing with sequences, such as natural language processing, time series analysis, and speech recognition [36]. Let  $x_t$ ,  $h_t$ ,  $y_t$  denote the input, hidden state and the output at time step  $t$ . The hidden state at each time step is calculated as a function of the input and the preceding hidden state:

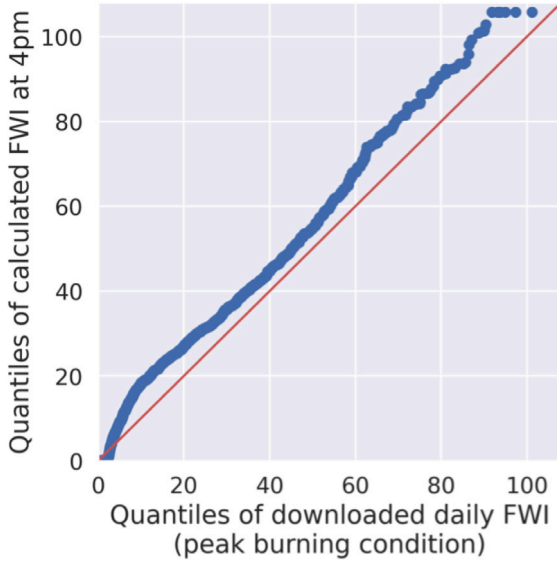
$$h_t = f(W_{hh} * h_{t-1} + W_{xh} * x_t + b_h) \quad (1)$$

where  $W_{hh}$  represents the weight matrix for recurrent connections,  $W_{xh}$  is the weight matrix for input connections,  $b_h$  denotes the bias term, and  $f$  stands for the activation function. The current state output  $y_t$  is subsequently computed based on the current hidden state:

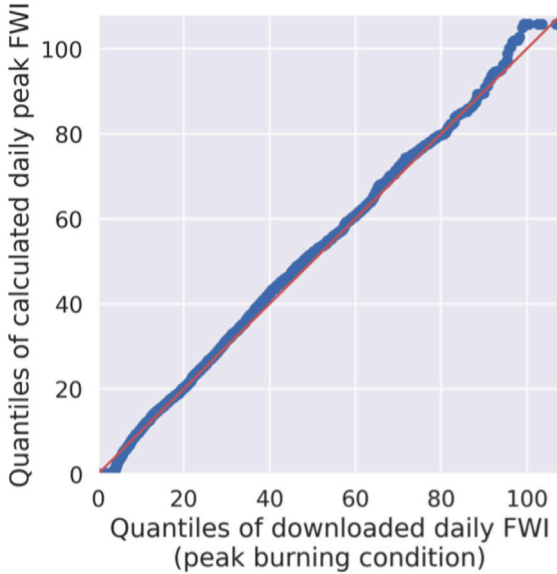
$$y_t = W_{yh} * h_t + b_y \quad (2)$$

where  $W_{yh}$  is the weight matrix for output connections and the  $b_y$  denotes the bias term.

Traditional RNNs face challenges in capturing long-range dependencies due to the vanishing or exploding gradient problem. More



(a) Q-Q plot for calculated FWI at 4 pm & downloaded daily FWI



(b) Q-Q plot for calculated daily peak & downloaded daily FWI

Fig. 4. Comparison between the downloaded daily FWI and the calculated FWI at 4 pm & the calculated daily peak FWI, based on operational data of Horsham, Victoria, Australia, from 2015 to 2020.

advanced architectures, such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Unit (GRU) networks, have been introduced. These architectures incorporate gating mechanisms that help control the flow of information through the network, allowing for more effective handling of long sequences and mitigating gradient-related problems [37]. Both methods comprise different gates, helping determine which information is valuable and should be retained to produce the next memory cell. LSTM has three gates to control the memory cell state  $s_t$ , including the input gate  $\Psi_{it}$ , the forget gate  $\Psi_{ft}$ , and the output gate  $\Psi_{ot}$ , following the formulae (3), (4), (5), (6), (7) and (8) [11,38].

$$\Psi_{ft} = \text{sigmoid}(w_{fx}x_t + w_{fh}h_{t-1} + b_f) \quad (3)$$

$$\Psi_{it} = \text{sigmoid}(w_{ix}x_t + w_{ih}h_{t-1} + b_i) \quad (4)$$

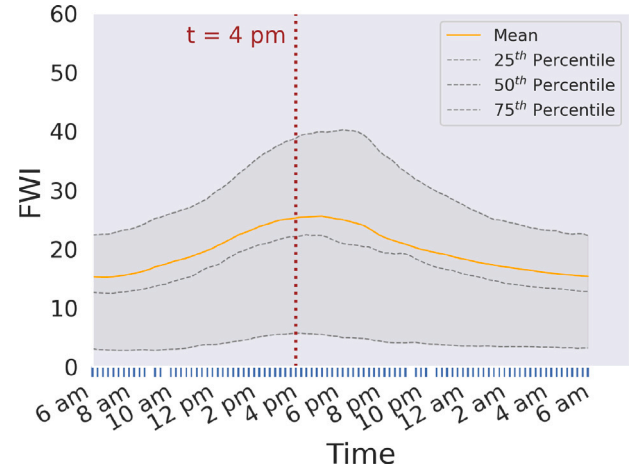


Fig. 5. The distribution of the calculated daily 15-min resolution FWI curves, based on operational data of Horsham, Victoria, Australia, from 2015 to 2020. The 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> percentile lines and the daily mean FWI curve are plotted to show variability of distribution.

$$\Psi_{gt} = \tanh(w_{gx}x_t + w_{gh}h_{t-1} + b_g) \quad (5)$$

$$\Psi_{ot} = \text{sigmoid}(w_{ox}x_t + w_{oh}h_{t-1} + b_o) \quad (6)$$

$$s_t = \Psi_{gt} * \Psi_{it} + s_{t-1} * \Psi_{ft} \quad (7)$$

$$h_t = \tanh(s_t) * \Psi_{ot} \quad (8)$$

where  $\Psi_{gt}$  is the input node,  $w$  and  $b$  represent the corresponding connection weight and bias matrices.

The GRU was proposed by [39] in 2014 to address deficiencies found in traditional RNNs, including the vanishing gradient problem and the overhead reduction in the LSTM architecture. To diminish computational complexity, the GRU employs two gates to propagate the information flow: the update gate  $\Psi_{zt}$  and the reset gate  $\Psi_{rt}$ . The computation of each GRU cell can be expressed using Eqs. (9), (10), (11) and (12).

$$\Psi_{zt} = \text{sigmoid}(W_z \cdot [h_{t-1}, x_t]) \quad (9)$$

$$\Psi_{rt} = \text{sigmoid}(W_r \cdot [h_{t-1}, x_t]) \quad (10)$$

$$\hat{h}_t = \tanh(W \cdot [\Psi_{rt} * h_{t-1}, x_t]) \quad (11)$$

$$h_t = (1 - \Psi_{zt}) * h_{t-1} + \Psi_{zt} * \hat{h}_t \quad (12)$$

where  $W$  represents the corresponding weight matrices.

Further details of functional equations and weight matrices for LSTM and GRU can be found in [11,40]. While both LSTM and GRU belong to the RNN category, they exhibit advantages over each other in specific cases. GRU, a more recent invention than LSTM, features a simplified gate structure that utilises less memory and is faster. Additionally, it has been demonstrated that GRU is more resilient to gradient exploding and vanishing issues [41]. However, for data sets with long and high-complexity sequences, LSTM has exhibited superior forecast performance [42]. Consequently, both structures are tested and compared based on the WRLFM.

While RNN, LSTM and GRU models are tested, their corresponding bi-directional structures are also introduced and compared. Bi-directional RNN (Bi-RNN), Bi-directional LSTM (Bi-LSTM), and Bi-directional GRU (Bi-GRU) process sequential data in both forward

and backward directions to comprehend the context from both past and future. Combining information from the two hidden layers (forward and backward) effectively captures long-term dependencies in sequences [43].

#### CNN-related methods

Given that the load forecast is calculated based on historical load and other features, the generated multi-factor matrix can be processed using deep learning methods designed for image feature extraction. Convolutional Neural Network (CNN), Temporal Convolutional Network (TCN), and Vision Transformer (ViT) are tested in our paper to understand the multi-factor input and predict future electricity demand.

CNN is designed to process structured grid data, such as images, specialising in extracting spatial features in data sets [44]. Convolutional layers consist of filters (also known as kernels) that slide across the input data, capturing local patterns. Pooling layers are then utilised to decrease spatial dimensions, retaining important information [45].

TCN is an extension of CNN designed for sequential data, making it suitable for time series tasks. It maintains the convolutional architecture but is adapted for 1D input sequences. In addition to the normal CNN structure, dilated convolutions allow TCN to capture various dependencies across different time scales. Skip connections help address the vanishing gradient problem and facilitate the flow of information [46].

ViT represents a novel approach to address image processing and multivariate time series forecasting tasks [47]. It leverages the transformer architecture, originally designed for sequence-to-sequence tasks in natural language processing, for image classification [48,49]. ViT divides the input into fixed-size, non-overlapping patches, treating each patch as a token. These patches are linearly embedded and fed into the transformer encoder, which uses self-attention mechanisms to capture global contextual information in parallel. As our case study is to forecast time series energy demand using multiple variables, we select the ViT structure in the comparative analyses, which upgrades the traditional Transformer method with a patching process to capture global information. Positional encoding is used to record spatial information [50]. The self-attention mechanism allows the model to weigh various parts of the input sequence differently, capturing dependencies and relationships between elements [51]. The attention mechanism is computed using three sets of vectors: Query ( $Q$ ), Key ( $K$ ), and Value ( $V$ ), as expressed in Eq. (13), (14) and (15).

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (13)$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^o \quad (14)$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V), i \in 1, 2, 3, \dots, h \quad (15)$$

$d_k$  is the dimensionality of the Query and Key vectors. Softmax is a mathematical function used for normalisation by converting raw scores into probabilities. Multi-head attention comprises multiple sets of  $Q$ ,  $K$ , and  $V$ . Each head operates in parallel, and their results are concatenated to obtain the final output of the attention mechanism. The motivation behind using multiple heads is to enable the model to attend to different parts of the input sequence simultaneously, capturing diverse patterns and relationships.  $W^o$  is the output weight matrix after multi-head attention splicing. Function *Concat* concatenates the attention output from each head.  $W_i^Q$ ,  $W_i^K$  and  $W_i^V$  are the weight parameter matrices for model training.

#### 2.3. Evaluation methods

In our paper, Mean Squared Error (MSE) calculates the loss at each epoch to adjust the DL model training direction [50]. Additionally, MSE serves as one metric to evaluate the forecast performance. Another commonly used metric for assessing forecast accuracy is the Mean Absolute Percentage Error (MAPE) [52]. While MSE is a scale-based metric, the

unit of MAPE is percentages, rendering it comparable among data sets with different units and scales.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (16)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (17)$$

where:  $y$  is the actual value,  $\hat{y}$  is the predicted value, and  $n$  is the number of data points.

MSE and MAPE calculate the average error over the entire testing period. To evaluate the daily forecast performance, the daily mean error plots with the temporal resolution of 30 min are discussed in [Results and Discussion](#). The daily error variance is computed to assess the fluctuation in forecast performance within a day. The error variance  $\sigma^2$  is calculated using the formula (18):

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N} \quad (18)$$

where  $\sigma^2$  is the error variance,  $X$  is the average half-hourly error,  $\mu$  is the mean of the average half-hourly error over a day, and  $N$  is the total sampling points for a day. The smaller the daily error variance  $\sigma^2$ , the better robustness and stability of the forecast model [53]. Studying the stability of forecast accuracy within a day is valuable for identifying daily weak points and enhancing performance by addressing hard-to-predict periods.

### 3. Case study

To validate and demonstrate the development of the WRLFM, we selected the Horsham Distribution Network as the case study region. The area covered by this network includes the city of Horsham as well as the administrative district of the same name. This is located in the Wimmera region in the western part of Victoria State, covering an area of 4267 km<sup>2</sup> with a population of 20,429 (2021) [54]. The Horsham district economy is predominately rural industry-based with little to no high-consuming industry. This was selected as the study region due to the availability of high-quality data, its relatively central location, and the annual wildfire seasons it encounters [55]. Five wildfire seasons (Nov–Mar) from 2015 to 2020 were chosen to train and assess model performance. The first three wildfire seasons (2015–2018) were selected as training data sets, while the last two were used as testing data sets. This section reviews the five main steps in developing the WRLFM, providing a detailed explanation of relevant data sources.

#### Overview of the case study

As depicted in [Fig. 1](#), the development of the WRLFM comprises five main steps: (1) Compare and identify the optimal input structures for 13 recently trending ML/DL methods. (2) Test the effect of applying different calendar classification methods. (3) Conduct a correlation analysis to elucidate the internal relationship between various preceding temperatures and load, selecting highly correlated temperature conditions to enhance forecast performance. (4) Apply categorical FWI features to the WRLFM. (5) Conduct sensitivity analysis on wildfire and non-wildfire periods to substantiate the effectiveness and resilience of the WRLFM against extreme wildfire risks.

#### Original data set acquisition and processing

The original data sources required to construct the WRLFM encompass historical energy demand data, calendar data, temperature, and FWI data (derived from temperature, wind speed, 24-hour total precipitation, and relative humidity). To consolidate all data into a unified input matrix, preprocessing steps are essential to ensure consistent resolutions. The preprocessing methods for generating the input matrix have been comprehensively discussed in [Section 2.1](#). [Table 2](#) details all data sets' sources and formats.

Powercor is one of the five primary energy distributors in Victoria, Australia, covering western and central Victoria. The energy consumption data set is obtained from the Powercor Zone Substation

**Table 2**  
Original data formats and sources. N/A represents non-applicable.

Name	Spatial resolution	Temporal resolution	Unit	Source
Energy demand	N/A	30 min	MW	Powercor [56]
Calendar data	N/A	30 min	day-month-year	Powercor [56]
Temperature at 2 m	$0.25^\circ \times 0.25^\circ$	hourly	K	ERA5 single level [57]
10 m wind speed (u)	$0.25^\circ \times 0.25^\circ$	hourly	$\text{ms}^{-1}$	ERA5 single level [57]
10 m wind speed (v)	$0.25^\circ \times 0.25^\circ$	hourly	$\text{ms}^{-1}$	ERA5 single level [57]
Total precipitation	$0.25^\circ \times 0.25^\circ$	hourly	m	ERA5 single level [57]
Relative humidity	$0.25^\circ \times 0.25^\circ$	hourly	%	ERA5 pressure level [58]
Fire Weather Index	$0.25^\circ \times 0.25^\circ$	daily	unit-less	Copernicus Database [59]

Reports [56], with a temporal resolution of 30 min. Following the pre-processing steps, the temporal resolutions of all data sets are standardised to 30 min. For improved location accuracy of the DN, all climate data sets are spatially interpolated with a spatial resolution of  $0.01^\circ \times 0.01^\circ$  using CDO.

#### 4. Results and discussion

This section assesses and discusses the results, following the steps described in Fig. 1. With each factor added, the improvement in load forecasting accuracy is evaluated to determine the appropriate method of embedding various factors into the input. After completing the four steps, a sensitivity analysis is conducted to compare the model's performance during wildfire and non-wildfire seasons, demonstrating the greater difficulty of load forecasting during extreme weather events and highlighting the significance of employing our proposed method to contribute to a more resilient and accurate load forecasting during wildfire seasons.

##### 4.1. Step 1 comparative analyses of trending ML techniques and corresponding optimal input structures

Following the description of Step 1 in Section 2.1, the MSE and MAPE based on 13 trending ML methods were plotted and compared in Fig. 6. In each subplot, the input length was exponentially raised from 1 to 64 to determine the appropriate input sequence size with both good accuracy and acceptable computational burden. Overall, DL methods outperformed shallow ML methods (SVM, RF, and XGB) in terms of both MSE and MAPE.

Within the DL methods covered in Section 2.2, Bi-GRU exhibited the lowest error among all RNN-derived techniques, with an MAPE of 3.16% and an MSE of  $0.44 \text{ MW}^2$ . Concerning CNN-related methods, the ViT demonstrated the best performance with an MAPE of 3.20% and an MSE of  $0.45 \text{ MW}^2$ . As previously mentioned, RNN-derived and CNN-related methods have different strengths. While the Bi-GRU model outperformed all other methods, the forecast performance of the ViT method was only a little inferior to the Bi-GRU case. Moreover, additional factors were anticipated to be incorporated into the input matrix in the following stages. Given the ViT methods' superior capability to handle multi-row features compared to Bi-GRU's, they may contribute to enhanced forecast performance. Consequently, we selected the top-performing models from both RNN-derived (Bi-GRU) and CNN-related (ViT) categories as the foundational forecasting structures for the WRLFM in all the following steps.

Errors decreased as the input length increased and reached a stable level when the input length was approximately 16 h. Although several models achieved their minimum error with an input sequence length of 64, the errors were close between the 32-lookback and 64-lookback operating cases. Therefore, the pre-16-hour period (lookback = 32) was determined as the optimal input sequence length for subsequent steps, i.e., setting the column number of the input matrix to 32.

##### 4.2. Step 2 calendar information impact on forecast accuracy

In Step 2, two distinct approaches for assessing calendar effects on load forecasting were juxtaposed with the scenario lacking calendar information. The initial operational scenario categorised day types into three groups: weekdays, weekends, and holidays. The alternative method under consideration further refined the classification, encompassing each day of the week from Monday to Sunday, along with a separate category for holidays (comprising a total of eight types).

To evaluate performance, Probability Density Function (PDF) and Cumulative Distribution Function (CDF) plots, along with MAPEs and MSEs for all operational scenarios, are presented in Fig. 7. Bi-GRU and ViT, identified as the top performers in Step 1, were both subjected to testing. It is evident that the incorporation of calendar information significantly enhanced forecast accuracy, as the error points in the case without calendar information accrued at higher values for both forecasting structures. Furthermore, the three-category calendar classification method surpassed the eight-category calendar classification method. Upon applying the three-category calendar classification method, the MSE decreased by 8.1% and 18.8%, and the MAPE decreased by 5.3% and 15.4% for the Bi-GRU and ViT models, respectively, compared to the original case without considering calendar information in the input.

##### 4.3. Step 3 adopt flexible temperature conditions based on correlation analysis

In Step 3, our proposed flexible temperature conditions based on correlations were compared with typical temperature applications in existing load forecasting methods. The assessments were conducted using the best-performing conditions identified in the first two steps: Bi-GRU and ViT forecast models with three-calendar classification methods. Load forecasting performances were evaluated under various temperature conditions, including scenarios with no temperature, simultaneous temperature conditions, previous 24-hour mean temperature conditions, and flexible correlation-based temperature conditions. For our case study, we identified the temperature condition exhibiting the strongest correlation coefficient with the load as a novel and flexible condition. In Fig. 2, the instantaneous temperature five hours ahead of the load demonstrates the most robust relationship, with a correlation coefficient of 0.7224. Consequently, the five-hour-ahead instantaneous temperature condition was selected for comparison with three alternative methods.

Fig. 8 illustrates the comparative daily mean absolute error curves for Bi-GRU and ViT-based models. The bold blue lines depict Horsham's real power curve, while the shaded area surrounded by blue dotted lines represents the 5<sup>th</sup> to 95<sup>th</sup> percentile interval of the daily load, indicating the change in dispersion level. The additional uptick in consumption observed between 10:30 pm and 1:30 am is attributed to dedicated loads from off-peak electrical hot water systems [60]. It is noticeable that the proposed best correlation temperature conditions yielded optimal performance in terms of MSE and MAPE for both Bi-GRU and ViT models. Compared to the no-temperature operating cases,



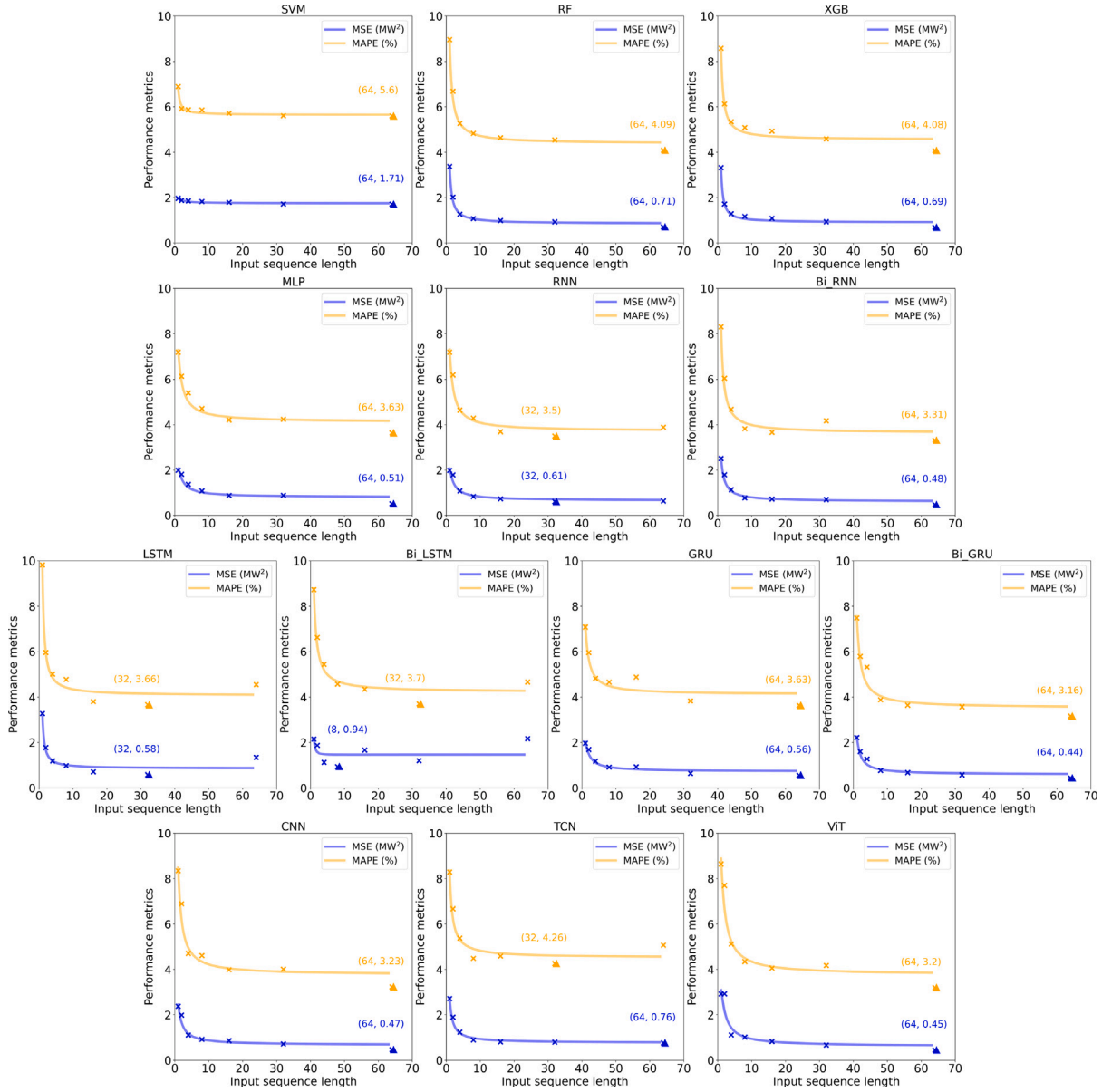


Fig. 6. Horsham electricity load forecasting performances using the state-of-the-art ML methods and various lengths of historical data as input. The input length was raised exponentially from 1–64 to find the lowest error point. (the lowest error points are marked with triangles).

our suggested best correlation temperature conditions contributed 9.4% & 5.6% of reductions in MSEs and 7.4% & 3.7% of decreases in MAPEs for the Bi-GRU and ViT-based models, respectively.

While our proposed methods, utilising the best correlation temperature conditions, demonstrated superior performance in terms of the two error evaluation metrics, the 24-hour mean temperature methods exhibited slightly better results with regard to daily error variation. Specifically, in Fig. 8(b), the error variance of our proposed method is twice as high as that of the 24-hour mean temperature condition. One potential explanation could be the influence of loads from the off-peak electrical hot water systems. Given that we selected the previous five-hour instantaneous temperature as the flexible condition, it posed a challenge to adequately capture the relationship between load and temperature during manually controlled large-scale heating periods. Future work is required to explore methods to mitigate error variance without compromising accuracy.

#### 4.4. Step 4 upgrade load forecasting with categorical fire weather index

As described in Section 2.1, high-resolution FWI curves were generated to contribute to the inaugural WRLFM. According to Table 1, FWI Danger Ratings are classified into six levels. The One-hot encoder was employed to convert the categorical wildfire risks into a  $6 \times n$  matrix, thereby incorporating wildfire effects into load forecasting. Step 4 confirmed that the inclusion of high-resolution FWI contributed to an enhancement in forecast performance.

Fig. 9 demonstrates the comparative outcomes for Bi-GRU and ViT models. Applying our proposed categorical FWI resulted in reductions across all three evaluation metrics. For the Bi-GRU-based model, MSE, MAPE, and error variance were decreased by 4.2%, 2.2%, and 24.6% following the incorporation of FWI. In the case of the ViT-based model, the introduction of FWI led to reductions of 3.9%, 2.9%, and 59.4% in MSE, MAPE, and error variance, respectively. While the decreases in MSE and MAPE were more modest, the use of categorical FWI

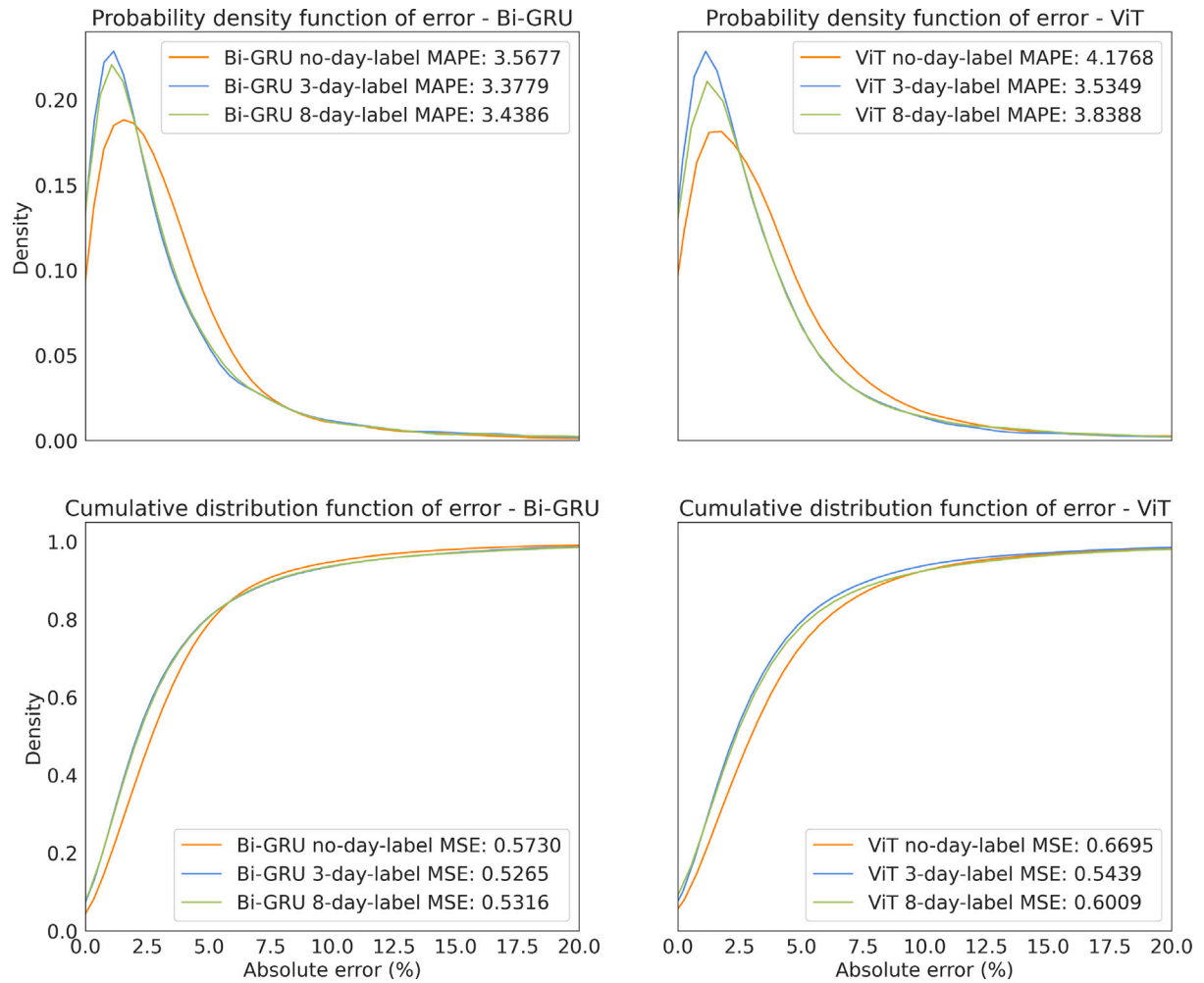


Fig. 7. Impact of applying various calendar classification methods on load forecasting performance. Probability density functions and Cumulative distribution functions for forecast errors, based on Bi-GRU and ViT structures. (The units of MSE and MAPE are  $\text{MW}^2$  and %).

notably diminished error variances, thereby enhancing the stability and reliability of forecasting results.

#### 4.5. Step 5 sensitivity analysis: data patterns and forecast performances during wildfire & non-wildfire seasons

Drawing conclusions from the four aforementioned steps, the optimal configuration for the WRLFM model can be summarised as follows.

(1) Configuring the input sequence length as 32 historical data points.

(2) Implementing the three-calendar classification method.

(3) Choosing the best correlation-based temperature condition.

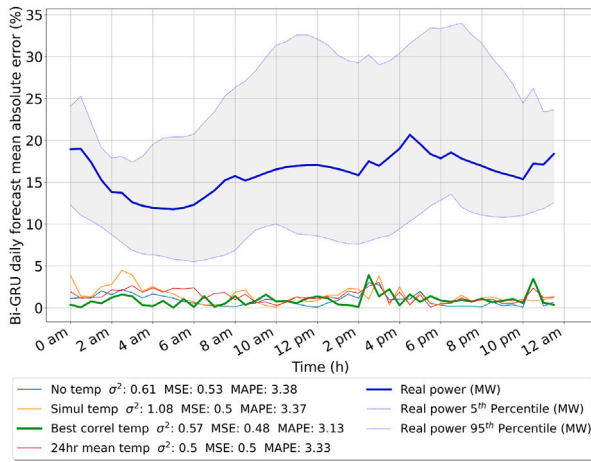
(4) Incorporating high-resolution categorical FWI as an input.

The subsequent Step 5 is to assess the performance of the WRLFM during both wildfire and non-wildfire seasons, aiming to enhance our understanding of the importance of focusing on load forecasting amidst extreme weather events.

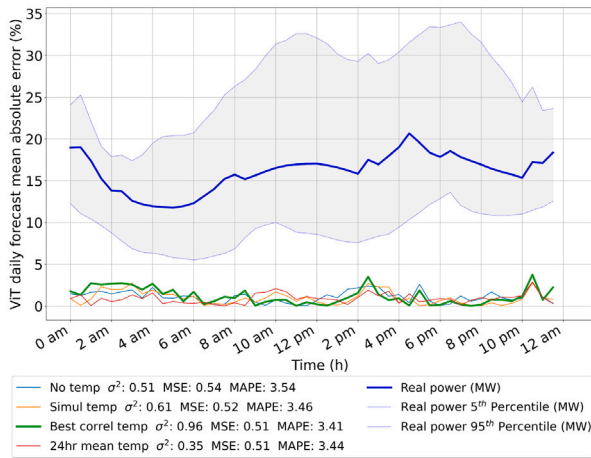
Before evaluating changes in forecast performances, it is essential to analyse the inherent characteristics of the data sets. As depicted in Fig. 10, the relationship between power and temperature is assessed and compared based on various classification standards. The distributions of points exhibit distinctive features, with data sets having higher FWI values or occurring during wildfire seasons demonstrating greater absolute values of correlation coefficients compared to those with lower FWI values or during non-wildfire seasons. This suggests a potential closer relationship between climate factors and demand behaviour during wildfire seasons.

In Step 5, the original power-only forecast model from Step 1 and the ultimately best-performing WRLFM from Step 4 were applied to data sets corresponding to wildfire and non-wildfire seasons. Given that the annual wildfire season spans five months, we selected a comparable duration of non-wildfire seasons for testing. The comparative results are presented in Table 3. Notably, as indicated in the last column of Table 3, regardless of the DL structures and factor settings employed, load forecasting accuracy during non-wildfire periods consistently surpasses that during wildfire seasons under equivalent conditions. This phenomenon can be explained by different inherent data dispersion levels observed for energy demands during wildfire and non-wildfire seasons. As marked in Fig. 10, both variances for power and temperature data sets are higher with a greater FWI or during wildfire seasons. The increased deviation in input data serves as one potential factor contributing to heightened prediction complexity. These results underscore the increased difficulty of load forecasting during extreme weather periods.

Additionally, we gauged the enhancements in forecast performance achieved by employing our proposed novel feature selection methods for both the Bi-GRU and ViT models. In the case of Bi-GRU, MAPEs were reduced by 14.37% and 2.8% when applying the WRLFM based on wildfire and non-wildfire season data sets, respectively. Similarly, the MAPE reductions from using the WRLFM were 20.86% and 14.01% during wildfire and non-wildfire seasons, respectively, for operating cases based on the ViT structure. The more pronounced improvement in forecast accuracy during wildfire season scenarios underscores the



(a) Bi-GRU based model



(b) ViT based model

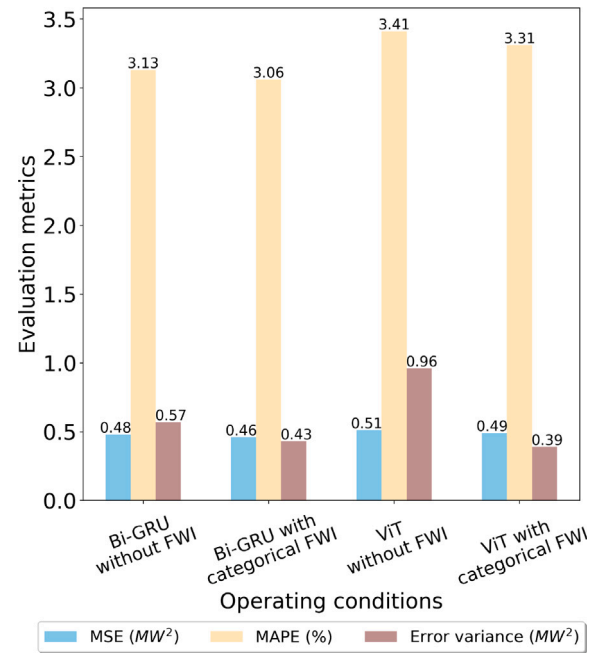
**Fig. 8.** Impact assessment of various temperature conditions on load forecasting performances. Daily forecast mean absolute error curves, along with MSEs, MAPEs, and error variances, are presented in each subfigure (Bi-GRU & ViT). The analysis utilises operational data from Horsham, Victoria, Australia, during the wildfire seasons in 2015–2020. (The units of error variance, MSE and MAPE are  $\text{MW}^2$ ,  $\text{MW}^2$  and %).

effectiveness and robustness of employing our proposed WRLFM for predicting electricity load during extreme wildfire seasons.

#### 4.6. Practical challenges

The forecast performance of the proposed WRLFM has been enhanced through the preceding steps, showcasing superior forecast accuracy and stability compared to other trending forecast models. This section explores the challenges that may arise when implementing the WRLFM in real-world settings, encompassing considerations such as computational cost, target users, and dynamic availability of data for the WRLFM.

It is imperative to consider the computational cost and efficiency when deploying models in real-world applications. Since all training processes are conducted offline beforehand, the model does not necessitate training during online deployment. In real-world scenarios, the trained model only undergoes forward propagation during testing. Processes with relatively high computational complexity, such as backward propagation for gradient calculation, are not involved in online applications. For our proposed WRLFM, we have examined the model summary for the two operational cases generated in Step 4, which represent the two most intricate models in our study. The total



**Fig. 9.** Impact of considering categorical Fire Weather Index on load forecasting performance.

**Table 3**

The load forecasting performance comparison between the proposed WRLFM and the original setting during wildfire and non-wildfire seasons — based on Bi-GRU and ViT structures.

Model	Factor setting	Data set period	MSE ( $\text{MW}^2$ )	MAPE (%)
Bi-GRU	Original	Wildfire	0.5730	3.5677
		Non-Wildfire	0.6232	2.9154
	WRLFM	Wildfire	0.4639	3.0551
		Non-Wildfire	0.5812	2.8337
ViT	Original	Wildfire	0.6695	4.1768
		Non-Wildfire	0.6773	3.7878
	WRLFM	Wildfire	0.4924	3.3057
		Non-Wildfire	0.6465	3.2571

parameter counts are 135,329 and 67,873 for the Bi-GRU and ViT-based models, respectively. Given the magnitude of these parameter counts, the trained model can even be executed on mobile devices for online deployment. Computational cost and efficiency are unlikely to impact online forecast performance. Hence, detailed calculations regarding computational cost and efficiency fall outside the scope of our paper.

Bearing in mind the target audience for the information provided by WRLFM being electricity Distribution System Operators (DSO), practical considerations would be mainly around access to data as input, computational resources to execute the method, and the distribution of the output for action. For input data, the DSO has immediate and timely access to relevant consumption data. The DSO will need to enter into a relationship with a Meteorological Service provider (Australian Bureau of Meteorology for our case study area). As previously discussed, the computational requirements of WRLFM are minimal to the extent that no special resources are required. The output of WRLFM can be presented to the system operator control room staff in a manner that is most easily understood such that required action can be taken in time.

## 5. Conclusion and future work

This paper has developed the Wildfire Resilient Load Forecasting Model (WRLFM) through a series of comprehensive comparative

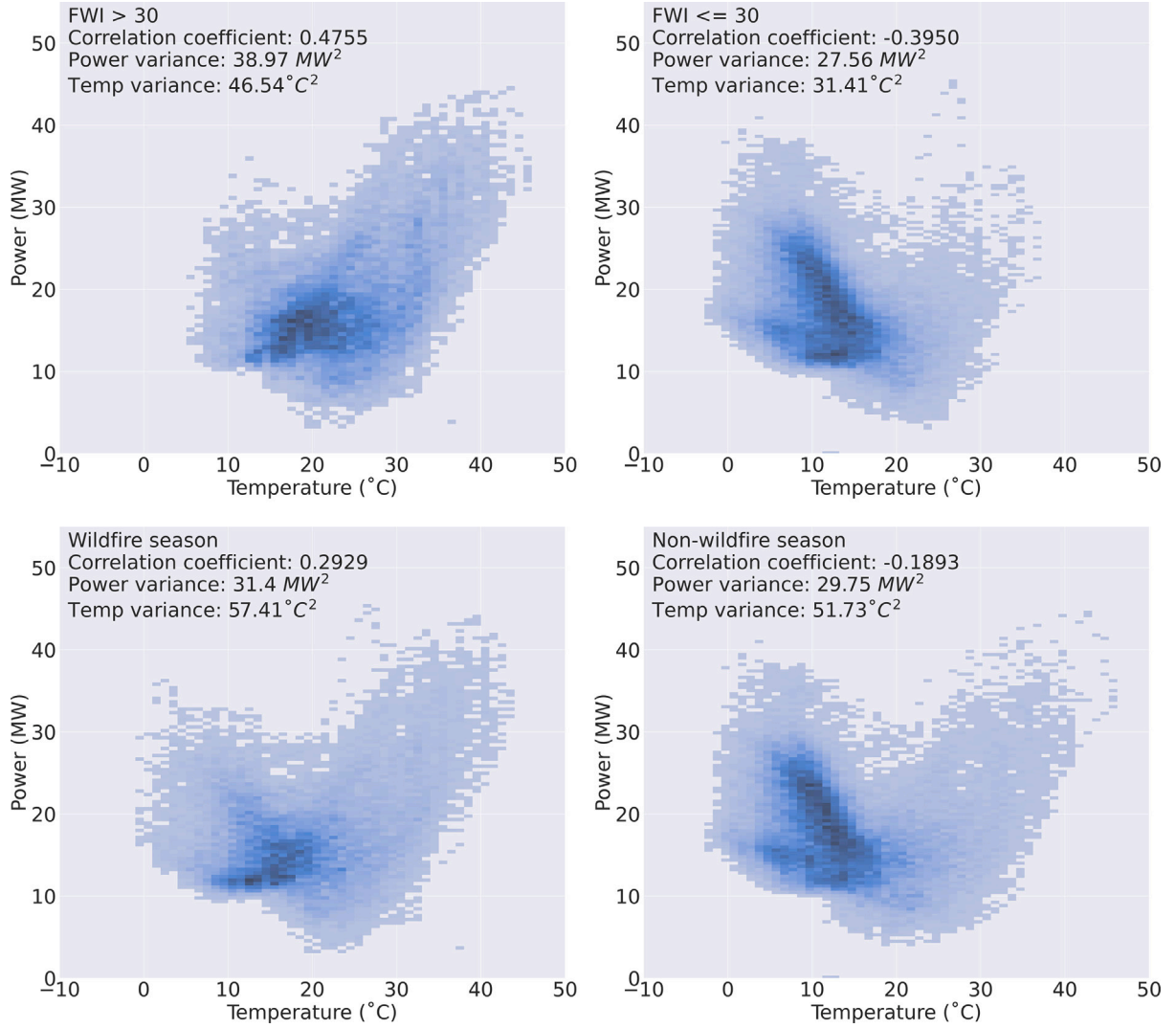


Fig. 10. Different patterns of power-temperature relationships during wildfire and non-wildfire seasons — Horsham, Victoria, Australia, 2015–2020, grouped by operating periods and FWI values.

analyses based on operational data from a distribution network in Australia from 2015 to 2020. Visual interpretation of the data sets has demonstrated that load forecasting under wildfire risks is more challenging than during normal times, and there exists a closer relationship between load behaviour and climate variables during wildfire seasons. Therefore, we tested and optimised multiple forecast structures and factors, enhancing load forecasting performance during wildfire seasons.

In the beginning, 13 recently trending ML-based forecast models were compared to identify the best performers with appropriate input sequence structures. Bi-GRU and ViT emerged as the top performers among RNN-derived and CNN-related methods, respectively, with 32 historical data points selected as the optimal input sequence lengths for subsequent steps.

After determining appropriate forecast structures, multiple factors were incorporated into the input matrix to enhance forecast performance. Concerning calendar information, classifying days into three types (weekdays, weekends, and holidays) outperformed other common methods, resulting in an average reduction of 13.5% in MSE and 10.4% in MAPE, compared to the scenario without calendar information.

Furthermore, various temperature applications in load forecasting were tested, demonstrating that our proposed best correlation-based temperature conditions yielded optimal performances. The average reductions in MSE and MAPE were 7.5% and 5.6%, respectively, compared to the previous step.

Lastly, we introduced the novel use of the categorical high-resolution FWI as a factor to enhance the robustness and accuracy of load forecasting in the face of high wildfire risks. As the original FWI data set is generated daily, detailed methods to generate half-hourly FWI curves were provided to align with the electricity data resolution in our paper. Categorical FWI information was embedded in the input using a One-hot encoder, leading to significant reductions in all three error metrics: 4% in MSE, 2.6% in MAPE, and 42% in daily error variance. To clarify, all error reductions calculated in this paragraph are the averages of the Bi-GRU and ViT results.

The final optimised setting of the WRLFM was identified as the Bi-GRU/ViT-based model with (1) an input sequence length of 32, (2) a three-calendar classification method, (3) a best-correlation-based temperature, and (4) categorical high-resolution FWIs. The optimised WRLFM reduced the MAPE of load forecasting during wildfire seasons by 14.37% and 20.86% for the Bi-GRU and ViT-based models,



respectively. In contrast, the improvement rate of load forecasting performance during non-wildfire seasons was more than two times less, indicating the significance of applying the WRLFM, particularly to operational data during wildfire seasons.

Although our proposed best-correlation-based temperature conditions performed well in both MSE and MAPE, the stability of the forecasting was somewhat inferior compared to the use of pre-24-hour mean temperature conditions. One possible explanation discussed is that the load from large-scale off-peak electrical hot water systems might affect the best-correlation temperature conditions, i.e., it is hard for the previous five-hour instantaneous temperature condition to capture the relationship between load and temperature during the midnight off-peak heating period. Future work is needed to explore methods to enhance forecast stability by considering additional factors, such as large-scale, manually controlled load signals.

While we assessed five annual wildfire seasons in our paper, a more extended period of data sets can be utilised to enhance forecast robustness and reliability. When employing data sets spanning longer than a decade, certain long-term effects, such as the impact of El Niño-Southern Oscillation (ENSO) on load forecasting, should be further considered to fortify model robustness [61].

In conclusion, our paper proposes a deep learning-based multi-factor WRLFM to address the challenges of predicting load under extreme wildfire risks. A comparative analysis was conducted to determine the most effective deep learning forecast structures among recently trending methods. Relationships between load behaviours and multiple factors were discussed to identify the optimal way to incorporate variables into the input of the WRLFM. High-resolution FWI was generated and applied in load forecasting for the first time, improving overall forecast accuracy during wildfire seasons and significantly enhancing performance stability. Compared to original load forecasting without multi-factor application, the MSEs were reduced by 19.04% and 26.45%, and the MAPEs were reduced by 14.37% and 20.86%, for the Bi-GRU and ViT-based models, respectively, after using the proposed WRLFM in our paper. Though the ViT-based model demonstrated a better forecast improvement rate through four main steps, the Bi-GRU-based WRLFM gave the best overall forecast performance with an MAPE of 3% and an MSE of 0.46 MW<sup>2</sup>. While the distribution network load has been assessed, the WRLFM can be further applied to residential load forecasting. Given the widespread usage of FWI in climate literature, our paper begins by exploring high-resolution categorical FWI to bolster the accuracy of energy demand forecasts. In future research, other fire risk indicators, such as the Fire Severity Index and Fire Danger Index, could be compared with FWI [62,63]. Additionally, the generalised WRLFM can be adjusted to incorporate other extreme weather risk indicators, contributing to a more resilient and accurate load forecasting performance in regions facing various types of extreme weather events.

#### CRedit authorship contribution statement

**Weijia Yang:** Writing – original draft, Software, Methodology, Writing – review & editing. **Sarah N. Sparrow:** Methodology, Supervision, Writing – review & editing. **David C.H. Wallom:** Writing – review & editing, Supervision, Conceptualization.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

All data sources used in this research are open-access from references cited in the paper.

#### Acknowledgments

The author would like to acknowledge that this work is supported by the UK Centre for Greening Finance and Investment (CERAF) (NERC grant NE/V017756/1); the Newton Fund through the Met Office Weather and Climate Science for Service Partnership Brazil (WCSSP Brazil) (NF\_MO\_BRA\_491), and the China Oxford Scholarship Fund (COSF).

#### References

- [1] American Nuclear Society. Report: Extreme weather is affecting nuclear power's reliability. Technical Report, American Nuclear Society; 2021.
- [2] Auffhammer M, Baylis P, Hausman CH. Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the United States. *Proc Natl Acad Sci USA* 2017.
- [3] Australian Government Bureau of Meteorology. Understanding fire weather. 2023. <http://www.bom.gov.au/weather-services/fire-weather-centre/bushfire-weather/index.shtml#:~:text=The%20greatest%20danger%20is%20between,produce%20dangerous%20fire%20weather%20conditions>.
- [4] Filkov AI, Ngo T, Matthews S, Telfer S, Penman TD. Impact of Australia's catastrophic 2019/20 bushfire season on communities and environment. Retrospective analysis and current trends. *J Safety Sci Resilience* 2020.
- [5] Australian Institute for Disaster Resilience. Bushfires - black summer. 2021, <https://knowledge.aidr.org.au/resources/black-summer-bushfires-vic-2019-20/>.
- [6] Yang W, Sparrow SN, Ashtine M, Wallom DC, Morstyn T. Resilient by design: Preventing wildfires and blackouts with microgrids. *Appl Energy* 2022.
- [7] Hong T, Fan S. Probabilistic electric load forecasting: A tutorial review. *Int J Forecast* 2016.
- [8] Lee J, Cho Y. National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model? *Energy* 2022.
- [9] Sarker I. Machine learning: Algorithms, real-world applications and research directions. *SN Comput Sci* 2021.
- [10] Xiao H, Pei W, Wu L, Ma L, Ma T, Hua W. A novel deep learning based probabilistic power flow method for multi-microgrids distribution system with incomplete network information. *Appl Energy* 2023.
- [11] Eskandari H, Imani M, Moghaddam MP. Convolutional and recurrent neural network based model for short-term load forecasting. *Electr Power Syst Res* 2021.
- [12] Zhang H, Chen B, Lei N, Li B, Li R, Wang Z. Integrated thermal and energy management of connected hybrid electric vehicles using deep reinforcement learning. *IEEE Trans Transp Electr* 2023.
- [13] Zhang H, Chen B, Lei N, Li B, Chen C, Wang Z. Coupled velocity and energy management optimization of connected hybrid electric vehicles for maximum collective efficiency. *Appl Energy* 2024.
- [14] Dong X, Yang W, Wu L, Tian X, Xie X, Chu W, Yang J, et al. Subsynchronous oscillation characteristic study of wind-thermal power bundling and EHV AC-dc hybrid transmission system. In: 2018 international conference on power system technology. 2018, p. 1995–2000. <http://dx.doi.org/10.1109/POWERCON.2018.8602352>.
- [15] Kumar S, Hussain L, Banarjee S, Reza M. Energy load forecasting using deep learning approach-LSTM and GRU in spark cluster. In: 2018 fifth international conference on emerging applications of information technology. 2018, p. 1–4. <http://dx.doi.org/10.1109/EAIT.2018.8470406>.
- [16] Yu B, Li J, Liu C, Sun B. A novel short-term electrical load forecasting framework with intelligent feature engineering. *Appl Energy* 2022.
- [17] Hu X, Li B, Shi J, Li H, Liu G. A novel forecasting method for short-term load based on TCN-gru model. In: 2021 IEEE international conference on energy internet. 2021, p. 79–83. <http://dx.doi.org/10.1109/ICEI52466.2021.00020>.
- [18] Emre TB, Demren D. Electrical load forecasting using support vector machines. In: 2011 7th international conference on electrical and electronics engineering. 2011, p. 1–49–1–53, URL <https://ieeexplore.ieee.org/document/6140142>.
- [19] Peng J, Kimmig A, Wang J, Liu X, Niu Z, Ovtcharova J. Dual-stage attention-based long-short-term memory neural networks for energy demand prediction. *Energy Build* 2021.
- [20] Maldonado S, González A, Crone S. Automatic time series analysis for electric load forecasting via support vector regression. *Appl Soft Comput* 2019.
- [21] Wang Y, Chen J, Chen X, Zeng X, Kong Y, Sun S, et al. Short-term load forecasting for industrial customers based on TCN-lightgbm. *IEEE Trans Power Syst* 2021.
- [22] Natural Resources Canada. Canadian forest fire weather index (FWI) system. 2018, <https://cwfis.cfs.nrcan.gc.ca/background/summary/fwi>.
- [23] Hancock JT, Khoshgoftaar TM. Survey on categorical data for neural networks. *J Big Data* 2020.
- [24] Yang W, Sparrow SN, Wallom DCH. Optimising multi-factor assistance in a deep learning-based electricity forecasting model with climate resilience: an Australian case study. In: 2023 IEEE PES innovative smart grid technologies europe (ISGT EUROPE). 2023, p. 1–5. <http://dx.doi.org/10.1109/ISGTUROPE56780.2023.10407524>.

- [25] Giannakopoulou C. Fire weather index (FWI) – dataset description. 2023, <https://confluence.ecmwf.int/pages/viewpage.action?pageId=283569774>.
- [26] Max Planck Institute. Climate data operators - tutorials. 2021, <https://code.mpimet.mpg.de/projects/cdo/wiki/Tutorial#Interpolation>.
- [27] Natural Resource Canada. Canadian Fire Weather Index Calculator. 2012, <https://code.google.com/archive/p/pyfwfi/>.
- [28] Van Wagner CE. Development and structure of the Canadian Forest Fire Weather Index System. Technical Report, Petawawa National Forestry Institute; 1987.
- [29] de Groot W. Interpreting the Canadian Forest Fire Weather Index (FWI) System. Technical Report, Northern Forestry Centre; 1987.
- [30] Vapnik VN. The Nature of Statistical Learning Theory. New York: Springer; 1999.
- [31] Fan S. Understanding the mathematics behind support vector machines. 2018, <https://shuzhanfan.github.io/2018/05/understanding-mathematics-behind-support-vector-machines/>.
- [32] Cheng Y-Y, Chan PPK, Qiu Z-W. Random forest based ensemble system for short term load forecasting. In: International conference on machine learning and cybernetics. 2012, p. 52–6. <http://dx.doi.org/10.1109/ICMLC.2012.6358885>.
- [33] Rafi SH, Nahid-AI-Masood, Mahdi MM. A short-term load forecasting technique using extreme gradient boosting algorithm. In: IEEE PES innovative smart grid technologies - Asia. 2021, p. 1–5. <http://dx.doi.org/10.1109/ISGTAsia49270.2021.9715272>.
- [34] Kim J, Lee B, Kim C. A study on the development of long-term hybrid electrical load forecasting model based on MLP and statistics using massive actual data considering field applications. *Electr Power Syst Res* 2023.
- [35] Wang O, Zhou S, Li GY. Few-shot learning for new environment adaptation. In: GLOBECOM 2023 - 2023 IEEE global communications conference. 2023, p. 351–6. <http://dx.doi.org/10.1109/GLOBECOM54140.2023.10437273>.
- [36] Shi H, Xu M, Li R. Deep learning for household load forecasting—A novel pooling deep RNN. *IEEE Trans Smart Grid* 2018.
- [37] Hou X, Bergmann JH. Hinet: Inertial navigation with head-mounted sensors using a neural network. *Eng Appl Artif Intell* 2023.
- [38] Zhang H, Lei N, Liu S, Fan Q, Wang Z. Data-driven predictive energy consumption minimization strategy for connected plug-in hybrid electric vehicles. *Energy* 2023.
- [39] Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). Doha, Qatar: Association for Computational Linguistics; 2014, p. 1724–34. <http://dx.doi.org/10.3115/v1/D14-1179>, URL <https://aclanthology.org/D14-1179>.
- [40] Versloot C. Relu, sigmoid and tanh: today's most used activation functions. 2019, <https://github.com/christianversloot/machine-learning-articles/blob/main/relu-sigmoid-and-tanh-todays-most-used-activation-functions.md>.
- [41] Mateus BC, Mendes M, Farinha JT, Assis R, Cardoso AM. Comparing LSTM and GRU models to predict the condition of a pulp paper press. *Energies* 2021.
- [42] Cahuantzi R, Chen X, Güttel S. A comparison of LSTM and GRU networks for learning symbolic sequences. 2021, arXiv.
- [43] Mounir N, Ouadi H, Jrhilifa I. Short-term electric load forecasting using an EMD-BI-LSTM approach for smart grid energy management system. *Energy Build* 2023.
- [44] Bu X, Wu Q, Zhou B, Li C. Hybrid short-term load forecasting using CGAN with CNN and semi-supervised regression. *Appl Energy* 2023.
- [45] Wei X, Juang B-HF, Wang O, Zhou S, Li GY. Accretionary learning with deep neural networks with applications. *IEEE Trans Cogn Commun Netw* 2023;1.
- [46] Shaikh AK, Nazir A, Khalique N, Shah AS, Adhikari N. A new approach to seasonal energy consumption forecasting using temporal convolutional networks. *Results Eng* 2023.
- [47] Nie Y, Nguyen NH, Sinthong P, Kalagnanam J. A time series is worth 64 words: Long-term forecasting with transformers. In: The eleventh international conference on learning representations. 2023, p. posters, URL <https://openreview.net/forum?id=Jbdc0vTOcol>.
- [48] Wang O, Zhou S, Li GY. Effective adaptation into new environment with few shots: Applications to OFDM receiver design. In: 2023 IEEE 33rd international workshop on machine learning for signal processing. MLSP, 2023, p. 1–6. <http://dx.doi.org/10.1109/MLSP55844.2023.10285904>.
- [49] Wang O, Zhou S, Li GY. New environment adaptation with few shots for OFDM receiver and mmwave beamforming. 2023, arXiv.
- [50] Zhang Q, Chen J, Xiao G, He S, Deng K. TransformGraph: A novel short-term electricity net load forecasting model. *Energy Rep* 2023.
- [51] Wang O, Gao J, Li GY. Learn to adapt to new environments from past experience and few pilot blocks. *IEEE Trans Cogn Commun Netw* 2023;9(2):373–85.
- [52] Rubasinghe O, Zhang T, Zhang X, Choi SS, Chau TK, Chow Y, et al. Highly accurate peak and valley prediction short-term net load forecasting approach based on decomposition for power systems with high pv penetration. *Appl Energy* 2023.
- [53] Zheng D, Eseye AT, Zhang J, Li H. Short-term wind power forecasting using a double-stage hierarchical ANFIS approach for energy management in microgrids. *Protect Control Modern Power Syst* 2017. <http://dx.doi.org/10.1186/s41601-017-0041-5>.
- [54] Australian Bureau of Statistics. 2021 Census all persons QuickStats. 2022, <https://abs.gov.au/census/find-census-data/quickstats/2021/LGA23190>.
- [55] Regional Development Victoria. Regional city of horsham. 2023, <https://www.rdv.vic.gov.au/victorias-regions/horsham>.
- [56] POWERCOR. Powercor zone substation reports. 2022, <https://www.powercor.com.au/industry-partners/renewable-generation/powercor-zone-substation-reports/>.
- [57] ECMWF. ERA5 hourly data on single levels from 1940 to present. 2023, <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview>.
- [58] ECMWF. ERA5 hourly data on pressure levels from 1940 to present. 2023, <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview>.
- [59] Service CEM. Fire danger indices historical data from the copernicus emergency management service. 2023, <https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-fire-historical-v1?tab=overview>.
- [60] Gov V. Electric hot water systems. 2022, <https://www.sustainability.vic.gov.au/energy-efficiency-and-reducing-emissions/save-energy-in-the-home/water-heating/choose-the-right-hot-water-system/electric-hot-water-systems>.
- [61] Goddard C. ENSO impacts on europe. 2023, <https://confluence.ecmwf.int/display/COPSRV/ENSO+impacts+on+Europe>.
- [62] Met Office. England and Wales fire severity index. 2024, <https://www.metoffice.gov.uk/public/weather/fire-severity-index/#?tab=map&fcTime=1710244800&zoom=8&lon=-0.37&lat=51.34>.
- [63] ECMWF. User guide for fire danger indices historical data from the copernicus emergency management service. 2024, [https://confluence.ecmwf.int/display/CEMS/User+Guide+for++Fire+danger+indices+historical+data+from+the+Copernicus+Emergency+Management+Service#:~:text=Fire%20Danger%20Index%20\(FDI\)%3A&text=The%20FDI%20ranges%20from%201,conditions%20and%20increased%20fire%20danger](https://confluence.ecmwf.int/display/CEMS/User+Guide+for++Fire+danger+indices+historical+data+from+the+Copernicus+Emergency+Management+Service#:~:text=Fire%20Danger%20Index%20(FDI)%3A&text=The%20FDI%20ranges%20from%201,conditions%20and%20increased%20fire%20danger).