Designing a short-term load forecasting model in the urban smart grid system
Li, C.

Applied Energy 266 (2020) 114850

Contents lists available at ScienceDirect

Applied Energy

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

Designing a short-term load forecasting model in the urban smart grid system

Chen Li

Institute of Environmental Sciences (CML), Leiden University, P.O. Box 9518, 2300 RA Leiden, the Netherlands

HIGHLIGHTS

- An innovative short-term load forecasting model is developed.
- A train-test ratios determination strategy based on the phase space reconstruction is proposed.
- A multi-objective optimization algorithm is used to optimize the neural network.
- Various measurement methods are conducted to evaluate model performance.

ARTICLE INFO

Keywords:
Smart grid
Short-term load forecasting
Neural networks
Multi-objective optimization algorithm
Urban sustainability

ABSTRACT

The transition of the energy system from fossil fuel towards renewable energy (RE) is rising sharply, which provides a cleaner energy source to the urban smart grid system. However, owing to the volatility and intermittency of RE, it is challenging to design an accurate and reliable short-term load forecasting model. Recently, machine learning (ML) based forecasting models have been applied for short-term load forecasting whereas most of them ignore the importance of characteristics mining, parameters fine-tuning, and forecasting stability. To dissolve the above issues, a short-term load forecasting model is proposed that incorporates thorough data mining and multi-step rolling forecasting. To alleviate the chaos of short-term load, a de-noising method based on decomposition and reconstruction is used. Then, a phase space reconstruction (PSR) method is employed to dynamically determine the train-test ratios and neurons settings of the artificial neural network (ANN). Further, a multi-objective grasshopper optimization algorithm (MOGOA) is applied to optimize the parameters of ANNs. Case studies are conducted in the urban smart grid systems of Victoria and New South Wales in Australia. Simulation results show that the proposed model can forecast short-term load well with various measurement metrics. Multiple criterion and statistical evaluation also show the good performance of the proposed forecasting model in terms of accuracy and stability. To conclude, the proposed model achieves high accuracy and robustness, which will provide references to RE transitions and smart grid optimization, and offer guidance to sustainable city development.

1. Introduction

1.1. Motivation

The use of RE is rising dramatically as technologies have made major advances and policy is pushing for a shift from fossil fuel to clean energies. However, the scaling of RE use to urban smart grid systems introduces big challenges as the volatility and intermittency of RE. To satisfy the continued high urban electricity demand, accurate and persistent short-term load forecasting plays a crucial role in power systems operation and management, especially in power generation expansion, dispatch scheduling of generating production, and sustainable electricity supply [1]. The overestimated forecasting will generate unnecessary electricity and load power storage remains a difficult task nowadays. The continued operation of power generation equipment leads to a large waste of resources, which is also a burden shift to other energy and environmental concerns. Conversely, the underestimated forecasting will cause inevitable damage to industrial production and people's life. A related study has reflected that 10 million operating costs may increase when the forecasting error increases by 1% [2].

In recent years, countries around the world are promoting RE use in the urban smart grid system while current electricity systems are mostly based on traditional fossil energies. Owing to the intermittency of RE,

E-mail address: c.li@cml.leidenuniv.nl.

https://doi.org/10.1016/j.apenergy.2020.114850
Received 14 January 2020; Received in revised form 2 March 2020; Accepted 13 March 2020
Available online 25 March 2020

Nomenclature

y	the original time series
L_1	the initial PF of white noise
ω_0	a constant
S_1	the original value of the first residue
r_1	the final value of the first residue
w	the white noise
x	a time series
N	the length of the time series (sample size)
m	the embedding dimension in the PSR
v	the delay time
d	the Euclidean distance
δ	the spatial distance
C	the correlation integrals
t	the time
l	the step length
n	the number of input layer of the BPNN
n_e	the number of equality constraints
n_v	the number of variables
g_j	the j_{th} inequality
h_j	the j_{th} equality constraints
L_j	the upper limit of j_{th} variable
U_j	the lower limit of j_{th} variable
e	the number of elements in a vector
$P_{optimal}$	Pareto optimal set
P_{front}	Pareto optimal front set
$Iter_{max}$	the maximum number of iterations
y_i	the position of the i_{th} grasshopper
o	the number of objective functions
n_i	the number of inequality constraints
S_i	the social interaction of the i_{th} grasshopper
N_{iter}	the iteration times
G_i	the gravity force for the i_{th} grasshopper
D_i	the wind advection for the i_{th} grasshopper
f	the gravitational constant
u	a constant drift
e_g	the unity vector to the center of the earth
e_w	the direction of the wind
H_d	the higher bounds in the d_{th} dimension
L_d	the lower bounds in the d_{th} dimension
T_d	the best value of d_{th} dimension so for
λ	the reducing factor
k	the iteration counter
A_i	the actual values
F_i	the forecasting values
q	the number of forecasting approaches
e_{it}	the relative error of the i_{th} approach at time t
E	the matrix of the relative error
$E(\cdot)$	the mathematical expectation
$\sigma(\cdot)$	the standard deviation
k	the sample number
R_i	the sum of the ranks
S_i	the i_{th} rank of the second data sample

Abbreviations

RE	renewable energy
ML	machine learning
PSR	phase space reconstruction
ANN	artificial neural network
MOGOA	multi-objective grasshopper optimization algorithm
NSP	numerical simulation prediction
ES	exponential smoothing
GM	gray model
ARMA	autoregressive moving average
MSAR	Markov-switching autoregressive
FLS	fuzzy logic systems
SVM	support vector machine
PSO	particle swarm optimization
FOA	fruit-fly optimization algorithm
CABC	chaotic artificial bee colony
CSA	cuckoo search algorithm
BPNN	Back-Propagation neural network
GPM	Gaussian process mixture
DMD	dynamic mode decomposition
VMD	Variational mode decomposition
LSTM	long short-term memory
GRNN	generalized regression neural network
MLR	multiple-linear regression
GPR	Gaussian process regression
ILMD	improved local mean decomposition
LMD	local mean decomposition
AMFM	amplitude-modulated-frequency-modulated
EMD	empirical mode decomposition
SVD	singular value decomposition
PF	product function
PFO	Pareto-front optimal
AEMO	the Australian Energy Market Operator
FE	forecasting error
TIC	the Theil's inequality coefficient
FVD	forecasting validity degree
AE	average error
MAE	mean absolute error
RMSE	root mean square error
MAPE	mean absolute percentage error
DA	direction accuracy
FB	fractional bias
DM	Diebold Mariano test
K-W	Kruskal-Wallis test
LCA	life cycle assessment

the short-term load series will show more nonlinear characteristics than the traditional ones. Despite the importance and urgency of making a transition from RE to the smart grid, it is still challenging to develop an effective and efficient short-term load forecasting due to this variability, uncertainty, and complexity of the RE resources. Thorough data cleaning and information mining are still insufficient for current forecasting models in modeling future short-term load as noise can be difficult to discard. Moreover, uncertainties still exist and cannot be well explained in ML-based forecasting methods, especially for parameters fine-tuning and determining. Further, besides forecasting accuracy, the robustness of forecasting is always ignored in most current studies. Overall, developing an effective and efficient short-term load forecasting model with high precision and robustness becomes a top priority for urban sustainability development.

1.2. Literature survey

In the past decade, a great number of studies have focused on short-term load forecasting in smart grid systems, therefore, various models have been promoted for the real application. Typically, there are four types of short-term load forecasting models, namely, physical forecasting models, statistical forecasting models, machine learning forecasting models, and hybrid (combined) forecasting models. Physical models, such as numerical simulation prediction (NSP) models [3], take temperature, humidity, and other physical variables into consideration, which are always used for long-term grid system forecasting and management. These models cost much computing resources and perform poorly in short-term forecasting [4]. Statistical models, such as linear regression [5], exponential smoothing (ES) [6], gray models (GMs) [7], Kalman filter [8] and the autoregressive moving average (ARMA) based models [9,10], are widely used for short-term load forecasting in urban smart grid systems. In recent years, the transition of electricity systems is happening as new technologies are maturing, such as the technologies of offshore wind energy, hydrogen energy, and electric vehicles. These large scale development of technologies relies on statistical forecasting models. For example, Pinson et al. used a Markov-switching autoregressive (MSAR) models to forecast wind power at two large offshore wind farms [11]. Amini et al. presented an ARIMA based model for electricity demand and charging demand of electrical vehicles parking lots forecasting simultaneously [12]. Statistical models are straightforward and easy to use but are considered unsuitable to solve nonlinear problems with their statistic hypothesis. Further, they are often limited by assumptions when dealing with different situations.

To overcome the limitations of physical and statistical models, numerous ML-based models have been applied for establishing forecasting models in recent years, namely ANNs [13], fuzzy logic systems (FLS) [14], expert systems [15], feed-forward perceptron [16] and support vector machines (SVMs) [17]. In particular, the discovery and characterization of ML algorithms are making ANNs more technically feasible among these approaches for short-term forecasting [18]. With their good generalization ability, ANNs have received considerable attention in urban smart grid forecasting and management. However, there are still many disadvantages, e.g. easily falling into a local optimum, over-fitting, and exhibiting a relatively low convergence rate. The train-test ratio determination and layers number settings are also big research gaps for ANNs [19]. Fortunately, with the massive development of big data and computational intelligence, heuristic optimization algorithms, such as particle swarm optimization (PSO) [20], fruit-fly optimization algorithm (FOA) [21], chaotic artificial bee colony (CABC) intelligent algorithm [22], and cuckoo search algorithm (CSA) [23] have employed to optimize the parameters of ML models, which too large extent enhance short-term load forecasting accuracy. Other examples can be found in the literature: SVM optimized by GOA is applied for load forecasting under local climatic conditions [24]; culture particle swarm optimization algorithm in combination models is developed to electrical load forecasting [25]; a hybrid GA-PSO algorithm is used in a short-term electrical load forecasting to optimize the parameters of Back-Propagation neural network (BPNN) [26] and; a hard-cut iterative training algorithm to improve a Gaussian process mixture (GPM) model [27]. However, optimization on enhancing robustness of forecasting models is still lacking in the literature. Multi-objective optimization including both forecasting accuracy and stability is another big research gap that needs to be filled in.

There is no cure-all individual or single models that can tackle all the short-term load forecasting problems, thus, research on combined or hybrid models has focused on integrating ANNs with other techniques, such as signal processing methods, optimization algorithms, and statistical leaning. The main novelty was that combined or hybrid models could forecast future values and capture the trend prevailing in the time series with good interpretability and accuracy. In recent years, hybrid models using data mining and data-processing techniques are conducted to extract and detect the inner characteristics of the short-term load series. For example, Mohan et al. employed a dynamic mode decomposition (DMD) to capture the Spatio-temporal dynamics of short-term electrical load [28]; and variational mode decomposition (VMD) method is used to decompose short-term load into a discrete number of modes [29]. Beaufond et al. proposed a combined model based on the Tukey labeling rule and the binary segmentation algorithm, which is verified as a reliable solution to detect and remove outliers [30].

Advanced ML models combined with data cleaning are developing massively as one of the research hotspots. For example, He et al. proposed a hybrid load forecasting model, which takes advantage of VMD, hyper-parameters optimization, and long short-term memory (LSTM) networks [31]. Raza et al. designed a novel hybrid model based on a feed-forward ANN, and a newly global best particle swarm optimization (GPSO) algorithm [32]. Further, research on combination of several ML-based models is employed for short-term forecasting. Bo et al. proposed a combined forecasting mechanism composed of BP, SVMs, ARIMA and generalized regression neural network, which is successfully established using the weight determination theory [33]. These combined models are not only found in the short-term load forecasting models in smart grid systems, but also have large scale use in other related fields. For example, Ahmad et al. designed a combined data-mining method comprising multiple-linear regression (MLR) model, Gaussian process regression (GPR) model and Levenberg Marquardt backpropagation neural network, which has a good performance in cooling load demand prediction [34]. Other cases can be found in [34,35], which indicated that hybrid or combined forecasting models can provide significantly better forecasting persistence, accuracy, and convergence characteristics than single forecasting models. Despite their good forecasting performance in certain areas, improvement for hybrid models is still can be further studied. As mentioned before, the inner mechanism of ANNs still needs to understand since the parameters inside remain large uncertainty. On the other hand, thorough data cleaning strategies and stability optimization also should be improved.

From the above review, the disadvantages of current short-term load forecasting models in smart grid systems can be summarized:

a) Physical models are always employed for long-term forecasting while they are not suitable for short-term forecasting.
b) Statistical models are more applicable to addressing linear trends data but encounter difficulties when dealing with nonlinear data.
c) ANNs are good solutions for nonlinear data analysis but may easily fall into local optimums and obtain a low rate of convergence. Additionally, over-training and poor forecasting are a pair of paradoxes, which makes it challenging to determine the suitable train-test ratios and neural layers.
d) Individual models can cause large forecasting bias so the hybrid (combined) models are the tendency in forecasting areas. However, most current hybrid models are based on single-objective optimization algorithms, which enhances the forecasting accuracy but always ignores the significance of forecasting effectiveness determined by its stability. Moreover, further data mining is insufficiently considered in current research.

1.3. Contributions and innovations

To address the limitations of the abovementioned short-term load forecasting models in urban smart grid systems, an innovative hybrid forecasting model is proposed in this study that is composed of four parts: thorough data cleaning, intelligent forecasting, multi-objective optimization, and comprehensive evaluation. The proposed model successfully achieves desirable and convincing forecasting performance in the urban smart grids, which can guide sustainable city decision-makers.

In the structure, concerning the data cleaning, an improved local mean decomposition (ILMD) method is designed to further mine the uncertain characteristics and discard the high-frequency noise in the original short-term load. Referring to the intelligent forecasting and multi-objective optimization, with the goal of investigation of intrinsic structural features and data mechanisms, the PSR based on the C-C method is designed to determine the suitable train-test ratios and the number of neural layers. The BPNN model optimized by the MOGOA is utilized for forecasting future electricity demand changes of the smart grid to simultaneously accomplish high accuracy and stability. Moreover, the proposed forecasting model is used to implement both one-step and multi-step ahead rolling forecasting for the short-term load. Finally, a variety of evaluation methods are used to comprehensively measure the forecasting performance in the evaluation part. To sum up, the main purpose of this study is to design an effective short-term load forecasting model in the urban smart grid system that makes up the insufficiency of existing research. The simulation results indicate that the proposed model outperforms other comparison models and can be implemented in the urban smart grid system.

The main innovations of this study can be concluded as follows:

a) A thorough data cleaning scheme based on the "decomposed and reconstructed" theory, effectively eliminates the negative influence of noise and mines the inner characteristics of the original short-term load data.
b) An effective train-test ratio determination strategy (the PSR based on the C-C method) is proposed to successfully find the balance of over-fitting and insufficient training, and the settings of parameters in ANNs.
c) A multi-objective optimization algorithm is conducted to optimize the initial weight and threshold of the neural network to simultaneously enhance smart grid forecasting accuracy and stability.
d) A comprehensive experimental analysis, including evaluation of both single and multiple points, forecasting validity degree, and statistical tests are employed to measure the forecasting model from different angles.

1.4. The organization of paper

The remaining part of the paper is organized as follows: the related materials and methods are introduced in Section 2. Section 3 presents the proposed short-term load forecasting model and Section 4 provides the experiments and corresponding analyses. Finally, the discussion and conclusion are respectively shown in Section 5 and Section 6.

2. Methodology

2.1. Data cleaning scheme

In this paper, the ILMD method is employed to mine the uncertain characteristics of the short-term load. The ILMD method is improved by the original LMD and regarded as one of the latest de-noising methods in the LMD family [36]. LMD-based methods consider time series as a superposition of amplitude-modulated-frequency-modulated (AMFM) components, amongst they iterate over the low-frequency components, and then recursively sift out the high-frequency components from the time series. It is proven that LMD-based methods are more suitable than empirical mode decomposition (EMD) based methods for incipient fault detection in nonlinear and nonstationary signal processing [37]. Other data cleaning methods, such as virtual memory device (VMD) and singular value decomposition (SVD), have been applied for short-term load forecasting in the related studies [1,31]. However, when the SVD extracts the watermark in the diagonal direction, the distortion caused by the computing error is inevitable [38]. The VMD adopts default values for both the number of modes and filter frequency bandwidth, but it is not adaptive to the signal being inspected [39]. On the other hand, its reasonable mode number is difficult to pre-set and this would make the loss of useful transient impulses [40]. Compared with these two methods, the ILMD employs the statistical characteristics of white noise, and effectively alleviates the mode mixing problem and the filter bank property.

In the ILMD method, a product function (PF) of white noise with adaptive amplitude is added to the input time series at every decomposition stage for each trial. Then, the LMD is employed to decompose the noise-added time series into one PF and one residue. An average of local means is obtained by taking ensemble mean of the residues, and the obtained ensemble local mean (i.e., the average of residues) is used as the input signal for the next stage. Finally, the corresponding PF is obtained by subtracting the ensemble local mean from the current time series. The detailed mechanism of the ILMD method is described as follows:

- **Step 1:** Add $L_1(w^{(j)})$ to the original time series y to obtain a new time series $y^{(j)} = y + \beta_0 L_1(w^{(j)})$, where $\beta_0 = \omega_0 std(y)$ and ω_0 is a constant. $L_1(w^{(j)})$ is the initial PF of white noise w that the j_{th} decomposed by the LMD method.
- **Step 2:** Decompose the new time series into one PF and one residue, where the original residue is specified as $S_1(y^{(j)})$.
- **Step 3:** Calculate the first residue $r_1 = 1/N\sum_{j=1}^{N} S_1(y^{(j)})$ and obtain the first true PF as $\mathbf{PF}_1 = y - r_1$, where N is the length of y.
- **Step 4:** Calculate the j_{th} residue by $r_j = 1/N\sum_{j=1}^{N} S(r_{j-1} + \omega_{j-1}L_j(w^{(i)}))$ and obtain the j_{th} PF.
- **Step 5:** Repeat Step 3 and Step 4 until the residue has no more oscillation.

2.2. Phase space reconstruction (PSR)

The PSR is a powerful method proposed by Takens et al. [41] to extract valuable features embedded in chaotic time series [42]. After the PSR, the state features of the domain can be displayed in high dimensional space [43]. Given a time series $x = \{x_i, i = 1, 2, \cdots, N\}$, where N is the length of x and x can be reconstructed by the PSR as follows:

$$\mathbf{X} = \{\mathbf{X}_i \mid \mathbf{X}_i = [x_i, x_{i+\upsilon}, \cdots x_{i+(m-1)\upsilon}], i = 1, 2, \cdots, P\} \tag{1}$$

In this formula, X_i is the i_{th} column of the matrix X constructed by the PSR and $P = N - (m - 1) * \nu$. Two parameters m and ν represent the embedding dimension and delay time respectively. In this paper, a C-C method based on two correlation integrals is conducted to obtain suitable input-hidden forms and train-test ratios. The correlation integrals formula can be defined as follows:

$$C(m, N, \delta, t) = \frac{2}{\mathbf{P}'(\mathbf{P}' - 1)} \sum_{1 \leqslant i \leqslant j \leqslant P'} \Theta(\delta - d_{ij}), \delta > 0 \tag{2}$$

where d and δ represent the Euclidean distance and spatial distance, respectively. The values of Θ can be described as follows:

$$\Theta(\delta - d_{ij}) = \begin{cases} 1, \delta - d_{ij} > 0 \\ 0, otherwise \end{cases} \tag{3}$$

A more specific explanation of the PSR based on the C-C method can be found in [44,45].

2.3. Back propagation neural network (BPNN)

The BPNN is regarded as one of the most widely used supervised ANNs [28]. With the error back propagation features, it constantly adjusts the weights and thresholds between layers to achieve the desired output. The forecasting model based on the BPNN mainly comprises three steps: BPNN construction, training, and forecasting. Generally, the signal is delivered from the input layer to the hidden layer, and then the hidden layer processes the information and passes them to the output layer. Based on the simulation output, the errors between the results of the output layer and the target output given in the sample will be back propagated as feedback. According to the feedback, the connection weights of the neutrons among different layers and the threshold values of each neuron can be adjusted. The network will enter the working stage once the training process reaches the stop training requirements. The input information is forward-propagated to obtain the output of the network (the predicted output) [46]. In this paper, the

BPNN is selected as the main predictor for short-term load forecasting, where each BPNN predictor has several input nodes and hidden nodes, and one output node. In this sense, the BPNN has the capability to forecast d-step ahead value $y(t+d)$ using a time series of previous values $y(t)$, $y(t-1)$,…, $y(t-n)$, where n represents the number of input layer of the BPNN.

2.4. Multi-objective optimization algorithm

2.4.1. The basic concepts of multi-objective optimization problems

Comparing solutions of the multiple objectives cannot be employed as the traditional relational operator and a new concept of dominates are therefore proposed by Edgeworth and Pareto [47,48]. The introduction of the optimization problem and Pareto dominance are shown as follows:

Definition of minimization problem: A multi-objective optimization problem can be described as a minimization problem:

$$\begin{aligned} &Minimize: F(\vec{x}) = \{f_1(\vec{x}), f_1(\vec{x}), \cdots, f_o(\vec{x})\} \\ &Subject\ \ to: g_j(\vec{x}) \geqslant 0, j = 1, 2, \cdots, n\_i \\ &\quad h_j(\vec{x}) = 0, j = 1, 2, \cdots, n\_e \\ &\quad L_j \leqslant x_j \leqslant U_j, j = 1, 2, \cdots, n\_v \end{aligned} \tag{4}$$

where o, n_i, n_e and n_v are the number of objective functions, inequality constraints, equality constraints and variables, respectively. g_j and h_j are the j_{th} inequality and equality constraints. And L_j and U_j represent the upper and lower limit of the j_{th} variable.

Definition of Pareto dominance: Considering two vectors $x = (x_1, x_2, \cdots, x_e)$ and $y = (y_1, y_2, \cdots, y_e)$ with e number of elements. Considering $y \prec x$, the Pareto dominance defines that the vector $\boldsymbol{y}$ is dominated by $\boldsymbol{x}$ if and only if:

$$\forall\, t \in [1, d], [f(x_t) \geqslant f(y_t)] \wedge [\ \exists\, t \in [1, d]: f(x_t)] \tag{5}$$

Besides Pareto's dominance, Pareto optimality, Pareto optimal set and Pareto optimal front are three basic definitions of Pareto theory to formulate the solutions, a set of solutions and the values of the objective functions, respectively. Their definitions are presented as follows:

Definition of Pareto optimality: A solution $\vec{x} \in X$ is called the Pareto optimality when and only when:

$$\nexists y \in X,\ s.\ t.\ F(y) \succ F(x) \tag{6}$$

Definition of Pareto optimal set: The set includes all the Pareto optimal solutions is called the Pareto optimal set, which can be expressed as follows:

$$P_{optimal} = \{x, y \in X \mid \exists F(y) \succ F(x)\} \tag{7}$$

Definition of Pareto optimal front: The set contains the values of objective functions for Pareto solutions set:

$$P_{front} = \{F(x) \mid x \in P_{optimal}\} \tag{8}$$

2.4.2. The multi-objective grasshopper optimization algorithm (MOGOA)

Conceptualized by the behavior of grasshopper insects, the GOA was developed by Saremi et al. in 2017 [49]. The grasshoppers always gather together in large swarms and make larger destruction to the agriculture property. The life cycle of grasshoppers can be generalized by three phases: egg, nymph, and adulthood [50]. In the nymph phase, the main characteristics of grasshopper movement can be expressed as jumping and moving in rolling cylinders (with small steps and slow movements), and they eat vegetation found in their paths. Nevertheless, grasshoppers migrate a long distance in swarms with abrupt movements and a large range in the adulthood phase.

The mathematical expression of the behavior of grasshoppers can be described as follows. Considering the position of the grasshopper is y_i:

$$y_i = S_i + G_i + A_i,\ i = 1, 2, \cdots, N_{iter} \tag{9}$$

where S_i represents the social interaction of i_{th} grasshopper and N_{iter} the iteration times.

$G_i = -f\hat{e}_g$ and $D_i = u\hat{e}_w$ represent the gravity force and wind

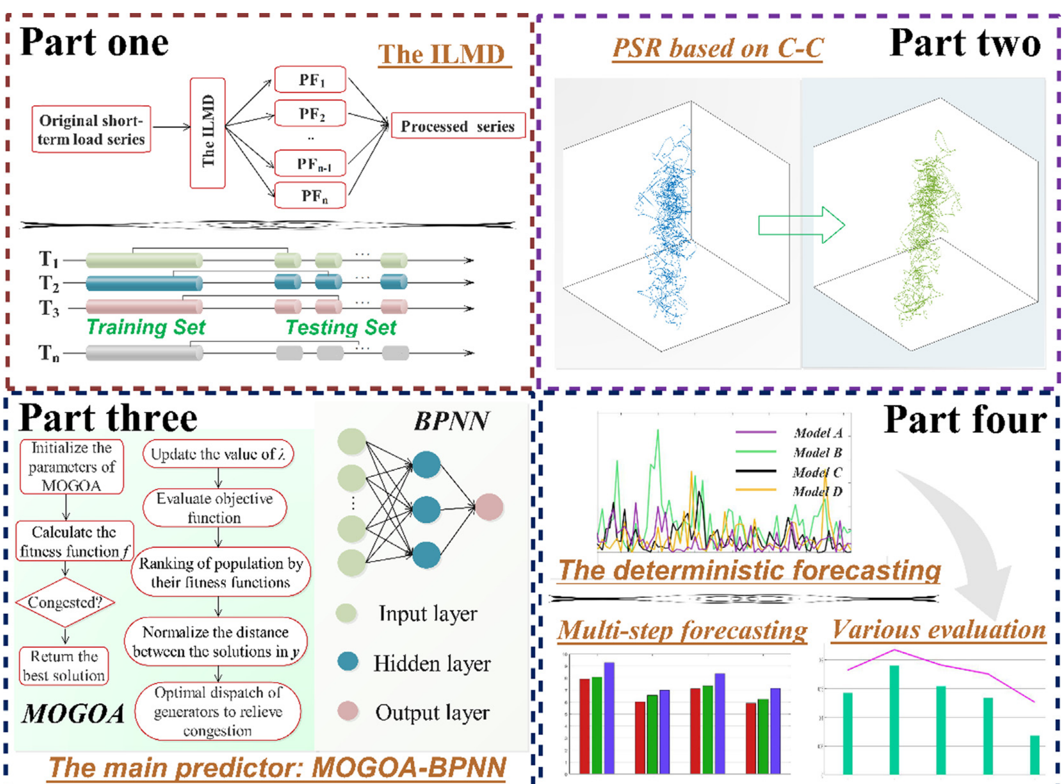

Fig. 1. The framework of the proposed forecasting model.

advection for the i_{th} grasshopper, respectively, where f and u represent the gravitational constant and a constant drift, respectively; e_g and e_w represent the unity vector towards the center of the earth and the direction of the wind, respectively.

It is noteworthy that the exploration and exploitation of the GOA can be adjusted by:

$$y_j^d = \lambda\left\{\sum_{j=1}^{N}\lambda\cdot\frac{H_d - L_d}{2}\cdot s(d_{ji})\cdot\widehat{d}_{ji}\right\} + \widehat{T_d},\ \forall j \neq i \tag{10}$$

where H_d and L_d are the higher and lower bounds in the d_{th} dimension, and T_d is the bast value of d_{th} dimension found so far. The reducing factor λ can be expressed by:

$$\lambda = \lambda_{\max} - k\frac{\lambda_{\max} - \lambda_{\min}}{Iter_{\max}} \tag{11}$$

where k indicates the iteration counter and $Iter_{max}$ represents the maximum number of iterations.

In order to address the multi-objective problem, the MOGOA is employed in this paper. The procedures of the technique for the order of preference by similarity to ideal solution are used to pick up the best compromise solution among the set of Pareto-front optimal (PFO) solutions [51]. The probability of selecting the best solution from the archive and then, a roulette wheel is used to pick up the target from the archive. The detailed Pseudocode of MOGOA is shown in Appendix A.

3. The proposed forecasting model in the urban smart grid system

The proposed forecasting model is described in detail in this section and the corresponding flowchart is shown in Fig. 1. Most current studies pay less attention to the importance of thorough data cleaning and multi-objective optimization, hence they cannot always satisfy the demand for high accuracy and persistence. Further, most ML-based forecasting models often encounter difficulties in determining the input-hidden and train-test ratio. The fine-tuning process of ratios is time-consuming and unstable, which may also cause inevitable losses in smart grid systems. Prior to this study, there is no uniform standard to define the number of training and test samples in ML-based models.

With these factors considered, this paper proposed a hybrid short-term load forecasting model in the smart grid system that comprises a novel data cleaning scheme, an advanced input-hidden and train-test ratio determining strategy, as well as a neural network predictor with multi-objective optimization. In the data cleaning scheme, the ILMD denoising method is applied for eliminating noise in the raw short-term load series and a "decompose and reconstruct" theory is used to discard the negative influence of noise. A detailed description of data cleaning is presented in Fig. 1, **Part one**. Owing to investigate the intrinsic structural features and data mechanisms, the PSR based on the C-C method is designed to determine the suitable input-hidden and train-test ratios, shown in Fig. 1, **Part two**. In the third module (shown in **Part three in** Fig. 1), the BPNN is selected as the main predictor and a multi-optimization algorithm is also developed to improve the accuracy and robustness of forecasting performance simultaneously. Furthermore, a multi-step ahead rolling forecasting framework is established for further short-term forecasting. The schematic diagram is demonstrated in Fig. 1, **Part four**. Finally, a series of evaluation indicators are utilized to comprehensively measure the forecasting performance. To sum up, the proposed hybrid forecasting model takes advantage of the integrity of each approach and ultimately accomplishes applicable, effective and efficient results.

4. Experimental simulations and analysis

4.1. Data description

In this paper, the 30-min load data from urban areas of Victoria and New South Wales in Australia were employed as study samples (detailed description shown in Fig. 2). The data is provided by the Australian Energy Market Operator (AEMO) [52]. Specifically, the data of each city is divided into four datasets, namely spring, summer, autumn, and winter respectively. Meanwhile, each dataset is divided into

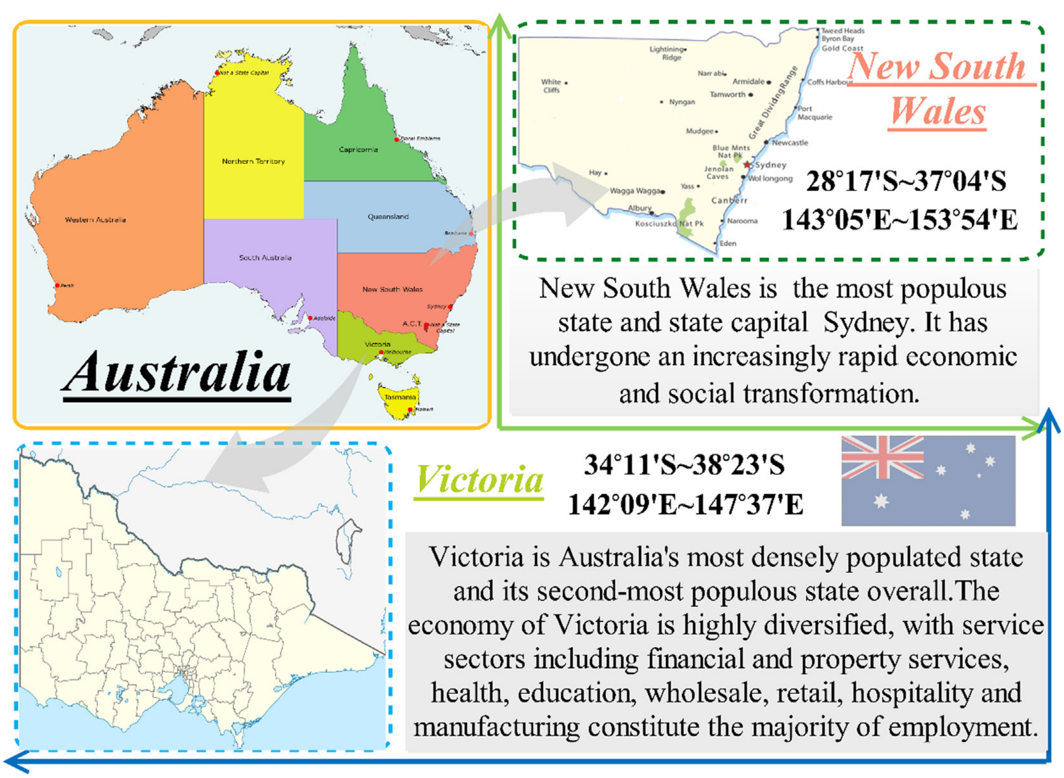

Fig. 2. The description of the two studied cities.

training data and testing data by PSR based on the C-C method, thus the proportion between the training and testing sets is different. As illustrated in Table 2, five statistical indicators *Mean*, *Max*, *Min*, *Median*, and *Std.* were used to perform the descriptive statistical analysis. The basic information of the studied areas is introduced in Fig. 2.

4.2. Simulation environment

All experiments in this paper were carried out in MATLAB R2018a on Windows 10 with 3.40 GHz Intel Core i7-6700, 64 bit having 16 GB of RAM. The parameters of methods mentioned in this paper are based on default values used in other literature [53–56]. Table 1 shows the final parameter setting after fine-tuning.

4.3. Forecasting principle

This paper conducts a multi-step ahead rolling forecasting mechanism, which uses previous forecasting values instead of only historical values to forecast the future values [57]. For example, the forecasting value of *n*-step ahead rolling forecasting $\widehat{y_n}$ is based on the historical data $y_n, y_{n+1}, \cdots y_{m-1}, y_m$ and the previously forecasting values $\widehat{y_1}, \widehat{y_2}, \cdots \widehat{y}_{n-1}$, where *m* is the sample length of the input short-term load series. The detailed rolling forecasting scheme is described in Table 3.

4.4. Evaluation metrics

Evaluation metrics play an important role in model measurement while there is no uniform criterion rule for model comparison and evaluation [58,59]. To systematically and scientifically evaluate the forecasting performance of the proposed short-term load forecasting model, this paper employs various assessment criteria. According to [60], forecasting error (FE) of single points is chosen for point-by-point comparison. Seven multiple-points evaluation metrics and the Theil's inequality coefficient (TIC) are used based on the evaluation system presented in [61,62]. Owing to the different dimensions of each sequence, it is different to measure different forecasting methods in the same validity. In this regard, the forecasting validity degree (FVD) method is also introduced for evaluating the forecasting performance [63]. Further, both parametric and nonparametric tests are conducted in this paper. The detailed introduction of these evaluation metrics are shown below.

4.4.1. Evaluation of single points

The initial level of model evaluation is always to assess the FE of single points. It is calculated the bias between the actual and forecasting values as FE values.

$$FE = \left| \frac{x(t) - \widehat{x}(t)}{x(t)} \right|, t = 1, 2 \cdots, n \tag{12}$$

where *x* represents the forecasting value and $\widehat{x}$ is the actual value of the short-term load time series. Typically, the variation of FE values can be used to assess the overall forecasting persistence.

4.4.2. Evaluation of multiple points

To extensively assess the model performance, seven indexes in multiple points, including average error (AE), mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), direction accuracy (DA), fractional bias (FB), and R^2 are applied to measure the difference between the forecasted and actual values in various perspective. Their equations and criteria are listed in Table 4. Moreover, the TIC is also employed to assess the equality of forecasted results. The formula of TIC is shown as:

$$\mathbf{TIC} = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} (\mathbf{A}_i - \mathbf{F}_i)^2} \Bigg/ \left(\sqrt{\frac{1}{N} \times \sum_{i=1}^{N} \mathbf{A}_i^2} + \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} \mathbf{F}_i^2} \right) \tag{13}$$

where A_i represents the actual values and F_i the forecasted values.

4.4.3. Forecasting validity degree

In the FVD, suppose a time series *x* and there are *q* forecasting approaches. x_{it} represents the forecasting value of i_{th} approach at *t* time, where $i = 1, 2, \ldots, q$, $t = 1, 2, \ldots, n$.

Definition 1. e_{it} is the relative error of the i_{th} approach at time *t* and *E* is the matrix of the relative error.

$$\mathbf{e}_{it} = \begin{cases} -1, & \frac{x_t - x_{it}}{x_t} < -1 \\ \frac{x_t - x_{it}}{x_t}, & -1 < \frac{x_t - x_{it}}{x_t} < 1 \\ 1, & \frac{x_t - x_{it}}{x_t} > 1 \end{cases} \tag{14}$$

Definition 2. The forecasting accuracy of the i_{th} approach at *t* time is $FA_{it} = 1 - |e_{it}|$ and the forecasting validity degree of ith forecasting approach is $FVD_{it} = E(FA_{it})[1 - \sigma(FA_{it})]$, where $E(\cdot)$ is the mathematical expectation and $\sigma(\cdot)$ is the standard deviation [60].

4.4.4. Diebold Mariano (DM) test

The *DM* test is a statistical hypothesis test to assess the difference between two forecasting models [64]. The original hypothesis and the alternative hypothesis are H_0: $E(\varpi_h) = 0, \forall\, n$ and H_1: $E(\varpi_h) \neq 0, \exists\, n$ respectively. Based on the *DM* statistics

$$DM = \frac{\sum_{h=1}^{k} (\mathrm{L}(o_{t+h}^{(A)}) - \mathrm{L}(o_{t+h}^{(B)}))/k}{\sqrt{t^2/k}} t^2 \tag{15}$$

We can make a judgment whether the proposed forecasting model is significantly different from comparison models.

4.4.5. Nonparametric test methods

Nonparametric statistical methods do not have to make assumptions of parameters for the objective we are studying. Due to the volatility and complexity of short-term load series, it is difficult to generalize its characteristics with established distribution . In order to testify whether different data samples obey the same distribution, four nonparametric statistical methods, including the Chi-square test, Kruskal-Wallis (*K-W*) test, Friedman test, and Spearman's rank correlation coefficient, are applied for extensively evaluation in this paper.

A Chi-square test is used to evaluate the significance of various study data samples. In this paper, we employ the Chi-square test based on a classical hypothesis:

Table 1
Parameter setting of the methods.

Models	Experimental parameters	Default values	Ref.
ILMD	Max iterations	50	[53]
	Max number of PFs	10	
	sifting stopping thresholds	[0.0001,0.7,0.05]	
	End extension length to original data	0.2	
PSR	Max delay time	200	[54]
BPNN	Learning velocity	0.1	[55]
	Maximum number of training	100	
	Training requirements precision	0.00001	
MOGOA	Max iterations	100	[56]
	Max size of archive	100	
	The number of grasshopper	200	
	Range of individual size	[-1, 1]	
	Repulsion forces	0.5	
	Attraction forces	0.4	
	Gravitational constant	9.8	

Table 2
The descriptive statistical characteristics of the study samples (MW).

Sites	Seasons	Length	Mean	Max	Min.	Median	Std.
V	Spring	4320	5673.7898	9587.5100	3833.4800	5437.3850	1087.3570
	Summer	4368	5608.6828	7813.3500	3839.8800	5645.4450	821.8795
	Autumn	4416	5561.7136	7699.9300	3705.9300	5619.1550	823.2917
	Winter	4416	5207.4736	9007.5200	3551.6000	5119.8350	743.3472
N	Spring	4320	7873.4508	11186.0000	5449.5900	7911.0200	1134.4818
	Summer	4368	8438.2571	11553.7500	5870.4800	8548.7100	1154.4959
	Autumn	4416	7742.8557	11073.4900	5546.3600	7851.5750	1006.6428
	Winter	4416	7874.3353	13787.8500	5113.0300	7833.7450	1323.2630

*The optimal number of training set and testing set are determined by the PSR based on the C-C.

Table 3
The mechanism of multi-step ahead rolling forecasting.

Multi-step ahead forecasting	Historical variables	Previously forecasting variables
1-step ahead forecasting	$y_1, y_2, \ldots y_{m-1}, y_m$	–
2-step ahead forecasting	$y_2, y_3, \ldots y_{m-1}, y_m$	$\widehat{y_1}$
3-step ahead forecasting	$y_3, y_4, \ldots y_{m-1}, y_m$	$\widehat{y_1}, \widehat{y_2}$
…	…	…
n-step ahead forecasting	$y_n, y_{n+1}, \ldots y_{m-1}, y_m$	$\widehat{y_1}, \widehat{y_2}, \cdots \widehat{y_{n-1}}$

Table 4
The evaluation metrics for multiple points.

Metric	Definition	Equation
AE	The average error of *N* forecasting results	$AE = \frac{1}{N}\sum_{i=1}^{N}(F_i - A_i)$
MAE	The mean absolute error of *N* forecasting results	$\boldsymbol{MAE} = \frac{1}{N}\sum_{i=1}^{N}\lvert F_i - A_i\rvert$
RMSE	The square root of the mean square error	$RMSE = \sqrt{\frac{1}{N}\times\sum_{i=1}^{N}(F_i - A_i)^2}$
MAPE	The mean absolute percent error of *N* forecasting results	$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left\lvert\frac{A_i - F_i}{A_i}\right\rvert\times 100\%$
DA	The direction accuracy of forecasting results	$DA = \begin{cases}1, if\,(A_{i+1} - A_i)(F_{i+1} - A_i) > 0\\ 0, \quad otherwise\end{cases}$
FB	The fractional bias of forecasting results	$FB = 2\times(\bar{\mathbf{A}} - \bar{\mathbf{F}})/(\bar{\mathbf{A}} + \bar{\mathbf{F}})$
R^2	Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=1}^{N}(F_i - A_i)}{\sum_{i=1}^{N}(F_i - \bar{F})}$

$$\mathbf{H}_0: \mu_1 = \mu_2,\ \mathbf{H}_1: \mu_1 \neq \mu_2 \tag{16}$$

The *K-W* test is a goodness of fit test that especially applies for exploring the distribution of continuous random variables [65]. The null hypothesis of the *K-W* test is that the samples to be verified obey the same distribution and conversely, the alternative hypothesis is that the two samples do not obey a distribution. The *K-W* statistic *H* is defined as below:

$$\mathbf{H} = \frac{12}{N(N+1)}\sum_{i=1}^{k} n_i(\bar{\mathbf{R}}_i - \bar{\mathbf{R}})^2 = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{\mathbf{R}_i^2}{n_i} - 3(N+1) \tag{17}$$

where $N = \sum n_i$ is the observation number and k signifies the sample number. R_i and n_i represents the sum of the ranks and the number of observations in the i_{th} sample respectively.

The Friedman rank sum test considers complete block design [66], where the null hypothesis is defined as all the positional parameters are consistent. The Friedman test statistic is described as follows:

$$\mathbf{Q} = \frac{12}{nk(k+1)}\sum_{i=1}^{k}\left(\mathbf{R}_i - \frac{n(k+1)}{2}\right)^2 = \frac{12}{nk(k+1)}\sum_{i=1}^{k}\mathbf{R}_i^2 - 3n(k+1) \tag{18}$$

Theoretically, Q obeys the Chi-square distribution.

The Spearman's rank correlation coefficient is the most far-reaching rank statistic [67] that measures the correlation between two data samples. Similar to the R^2, the Spearman's rank correlation coefficient statistic is defined as follows:

$$\mathbf{S} = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2\sum_{i=1}^{n}(S_i - \bar{S})^2}} = 1 - 6\sum_{i=1}^{n} d_i^2/n(n^2 - 1) \tag{19}$$

where R_i and S_i are the i_{th} rank of the first and second data samples respectively, and $d_i^2 = (R_i - S_i)^2$ measures the distance between two data samples.

4.5. Results and analysis

4.5.1. Case study of Victoria

The 30-min short-term load data from Victoria is used to verify the performance of the proposed model in the urban smart grid system. Multi-step ahead rolling forecasting results, including one-step, two-step and three-step ahead rolling forecasting are shown in Table 5, and further elaboration of the forecasting performance is shown in Table 6, where the values in the bold present the optimal values of each criterion among all the models.

From Table 5 we can conclude that the single BPNN model cannot achieve desirable forecasting results, with the worst performance in comparison with other hybrid BPNN models. For example, the MAE, RMSE, and MAPE values of the BPNN model in one-step ahead forecasting in the V-spring database are 205.5355, 269.0515, and 3.9693 respectively, which are significantly higher than other comparison models listed in the table. Owing to the contributions of the optimization algorithms, hybrid models GOA-BPNN and MOGOA-BPNN accomplish comparatively good forecasting results for different multi-step ahead forecasting. However, it is noteworthy that the forecasting performance is slightly improved by optimization algorithms. Data decomposition methods, conversely, plays a decisive role in enhancing forecasting accuracy and persistence. For example, the ILMD de-noising approach leads to reductions of 140.6381 in MAE, 183.9491 in RMSE, 2.7212 in MAPE for one-step ahead forecasting, 84.3244 in MAE, 112.7299 in RMSE, 1.5699 in MAPE for two-step ahead forecasting, and 27.1475 in MAE, 37.8664 in RMSE, 0.5441 in MAPE for three-step ahead forecasting, respectively. Moreover, Fig. 3 also displays the model comparison results in the case of Victoria. The proposed model is combined with other three hybrid models (e.g. MOGOS-BP, RLMD-BP, and RLMD-GOA-BP), and different models are marked in different colors. It can be found that the proposed forecasting model fits better in the observed values than other models in all the four seasons, with the yellow lines highly consistent with the orange lines. To conclude, the proposed model takes advantage of data preprocessing methods and optimization algorithms, and contribute to forecasting performance.

Table 6 provides an extensive analysis of the proposed forecasting model. Statistical methods (e.g. the Naïve predictor and ARIMA), and five BP-based models are selected as benchmark models. Another five evaluation metrics (e.g. AE, DA, FB, R^2, and TIC) are employed for measuring the forecasting performance. It can be observed from the experimental results that the proposed forecasting model is superior to

Table 5
Performance of multi-step ahead load forecasting models in Victoria (MAE, RMSE, MAPE (%)).

Sample	Horizon	One-step			Two-step			Three-step		
	Indices	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
V-Spring	BP	205.5355	269.0515	3.9693	226.8960	304.5866	4.2770	228.1386	307.5375	4.3464
	GOA-BP	190.7988	243.6851	3.6972	192.3724	244.9566	3.7685	211.6028	289.3475	3.9751
	MOGOA-BP	161.7503	198.3946	3.3796	165.7086	199.1157	3.4677	211.1619	279.1419	4.0030
	ILMD-BP	64.8974	85.1024	1.2481	142.5716	191.8567	2.7071	200.9911	269.6711	3.8023
	ILMD-GOA-BP	63.2538	81.9432	1.2200	136.7703	177.9543	2.6131	193.7965	245.6415	3.7965
	ILMD-MOGOA-BP	**57.0906**	**73.9558**	**1.0936**	**126.8601**	**167.2498**	**2.4117**	**174.0703**	**208.9170**	**3.6392**
V-Summer	BP	282.4125	345.3183	4.9072	294.7932	356.6497	5.1854	321.5768	386.0644	5.6611
	GOA-BP	280.8649	342.5530	4.9539	289.5914	354.7987	5.0284	320.1984	381.2388	5.5726
	MOGOA-BP	279.6738	342.9161	4.8514	289.4781	351.3562	4.9330	313.0681	377.4471	5.4636
	ILMD-BP	58.1187	77.0712	1.0201	146.4057	191.0332	2.5382	269.7049	338.7778	4.6952
	ILMD-GOA-BP	42.5065	53.8342	0.7540	140.5453	178.6732	2.4558	257.4038	320.0347	4.4840
	ILMD-MOGOA-BP	**41.2048**	**52.6109**	**0.7222**	**139.2016**	**170.1014**	**2.4197**	**255.1285**	**318.0413**	**4.4368**
V-Autumn	BP	272.7112	350.5767	5.8019	285.4637	356.4023	6.0025	300.3804	367.5032	6.2375
	GOA-BP	261.7305	333.9519	5.6526	261.2499	326.4393	5.6194	288.1691	348.1933	6.2076
	MOGOA-BP	261.0706	322.9140	5.7019	274.8785	334.9508	5.9629	262.5115	334.8840	5.6653
	ILMD-BP	53.1068	73.1069	1.1158	124.2941	158.7899	2.5660	212.6942	269.6689	4.4369
	ILMD-GOA-BP	45.6097	60.9294	0.9395	123.6944	160.4249	2.5600	201.3773	259.8423	4.1897
	ILMD-MOGOA-BP	**45.1616**	**59.2764**	**0.9268**	**115.0913**	**149.8043**	**2.3828**	**195.0143**	**249.6555**	**4.0531**
V-Winter	BP	215.1376	297.6926	4.8919	220.8910	297.1208	5.0197	229.5910	301.8108	5.2143
	GOA-BP	213.9300	304.7069	4.8444	208.5697	295.0989	4.7299	209.3238	290.0884	4.7450
	MOGOA-BP	174.7423	230.7782	3.9470	179.2815	231.5367	4.0396	**187.2065**	**238.3120**	4.2078
	ILMD-BP	58.7619	81.7712	1.2632	131.5108	175.3115	2.8426	194.6042	258.8000	4.2317
	ILMD-GOA-BP	57.9586	88.1950	1.2445	129.2322	178.6998	2.7839	194.5340	254.1077	4.2467
	ILMD-MOGOA-BP	**56.9982**	**75.4477**	**1.2221**	**128.4235**	**174.2325**	**2.7736**	192.8598	244.4135	**4.1993**

Table 6
Further analysis of one-step ahead load forecasting models in Victoria.

		AE	DA	FB	R^2	TIC
V-Spring	Naïve Predictor	−5.0268	0.6897	0.0009	0.9764	0.0005
	ARIMA	5.7581	0.6520	−0.0011	0.9676	0.0005
	BP	28.8244	0.6461	−0.0054	0.8360	0.0026
	GOA-BP	21.4514	0.6900	−0.0042	0.8876	0.0020
	MOGOA-BP	7.3681	0.7270	−0.0014	0.8938	0.0007
	ILMD-BP	7.2896	0.7238	−0.0015	0.8518	0.0007
	ILMD-GOA-BP	2.2891	0.6959	−0.0004	0.9817	0.0002
	ILMD-MOGOA-BP	**1.5844**	**0.8238**	**−0.0003**	**0.9923**	**0.0001**
V-Summer	Naïve Predictor	**−0.6798**	0.6711	**0.0001**	0.8789	**0.0001**
	ARIMA	22.6145	0.7383	−0.0040	0.9186	0.0020
	BP	33.1321	0.6492	−0.0062	0.8352	0.0030
	GOA-BP	22.4251	0.6731	−0.0043	0.9153	0.0021
	MOGOA-BP	11.5741	0.6740	−0.0022	0.9583	0.0011
	ILMD-BP	11.4120	0.6762	−0.0021	0.9683	0.0011
	ILMD-GOA-BP	10.3961	0.7238	−0.0021	0.9727	0.0011
	ILMD-MOGOA-BP	5.6282	**0.7730**	−0.0011	**0.9890**	0.0005
V-Autumn	Naïve Predictor	6.2209	0.7634	**−0.0013**	0.9648	0.0016
	ARIMA	**1.4884**	0.6989	−0.0043	0.9534	0.0019
	BP	23.0226	0.6424	−0.0045	0.8303	0.0022
	GOA-BP	22.4330	0.6897	−0.0042	0.8958	0.0021
	MOGOA-BP	15.5473	0.6959	−0.0029	0.9185	0.0014
	ILMD-BP	11.7235	0.7082	−0.0022	0.9061	0.0011
	ILMD-GOA-BP	11.4408	0.7017	−0.0024	0.9321	0.0012
	ILMD-MOGOA-BP	9.7636	**0.8000**	−0.0018	**0.9825**	**0.0009**
V-Winter	Naïve Predictor	−1.3532	0.7452	**0.0003**	0.9362	**0.0001**
	ARIMA	1.8191	0.7163	0.0014	0.9151	0.0012
	BP	45.9039	0.6463	−0.0081	0.7523	0.0040
	GOA-BP	42.0556	0.6522	−0.0073	0.7801	0.0036
	MOGOA-BP	35.1496	0.6604	−0.0061	0.7845	0.0030
	ILMD-BP	5.6374	0.8571	−0.0010	0.9939	0.0005
	ILMD-GOA-BP	4.2480	0.8707	−0.0007	0.9942	0.0004
	ILMD-MOGOA-BP	**−2.2078**	**0.8959**	0.0004	**0.9974**	0.0002

other benchmark models in terms of AE, DA, FB, R^2, and TIC. For example, the DA value of the developed forecasting model is 0.8238 for one-step forecasting in database V-spring, whereas, the DA values conducted by the Naïve Predictor, ARIMA, BP, GOA-BP, MOGOA-BP, ILMD-BP, and ILMD-GOA-BP are 0.6897, 0.6520, 0.6461, 0.6900, 0.7270, 0.7238, and 0.6959, respectively. Regardless of one-step, two-step or three-step ahead forecasting, the proposed model always achieves the lowest AE, FB, and TIC values, and the highest DA, and R^2 values. In other words, the proposed forecasting model can effectively and efficiently forecast short-term load with high forecasting accuracy (measured by the error measurement criteria AE and R^2) and the precise direction and equality (measured by DA, FB, and TIC).

Remarks: Based on the above experiments, it is verified that the proposed forecasting model outperforms other comparison models in almost all of the cases. The combination of data-cleaning scheme, multi-objective optimization algorithm, and neural networks capitalizes on the advantages of each part, which results in the good performance of the developed forecasting system in terms of accuracy and stability.

4.5.2. Case study of New South Wales

Another case study is employed to manifest the effectiveness and efficiency of the proposed forecasting model. As mentioned before, 30-min short-term load data in four databases (i.e. *N*-spring, *N*-summer, *N*-autumn, and *N*-winter) are collected from New South Wales. The simulation results for New South Wales are presented in Tables 7 and 8.

According to the experimental results, the proposed model achieves better forecasting results in comparison with the Naïve Predictor, ARIMA, BP, GOA-BP, MOGOA-BP, ILMD-BP, and ILMD-GOA-BP model. For one-step ahead forecasting in the *N*-spring database, the MAE, RMSE, and MAPE values of the proposed model are 83.7981, 54.6077, and 0.7351, respectively. As for *N*-summer, *N*-autumn, and *N*-winter databases, the MAE, RMSE, and MAPE values are 68.4001, 58.6261, and 0.5616, 65.5574, 64.3753, and 0.6816, and 76.6759, 84.2976, and 0.6979, respectively. Similarly, the proposed model is superior to other benchmark models in terms of MAE, RMSE, and MAPE for two-step ahead and three-step ahead forecasting. When it comes to AE, DA, FB, R^2, and TIC, the developed model still performs the best among other comparison models. For example, the R^2 values of the proposed model in *N*-winter is on 0.0823 of increase when compared with ARIMA, 0.2451 of increase with BP, and 0.2129 with MOGOA-BP. Moreover, Fig. 4 shows more model comparison results. Some hybrid models, such

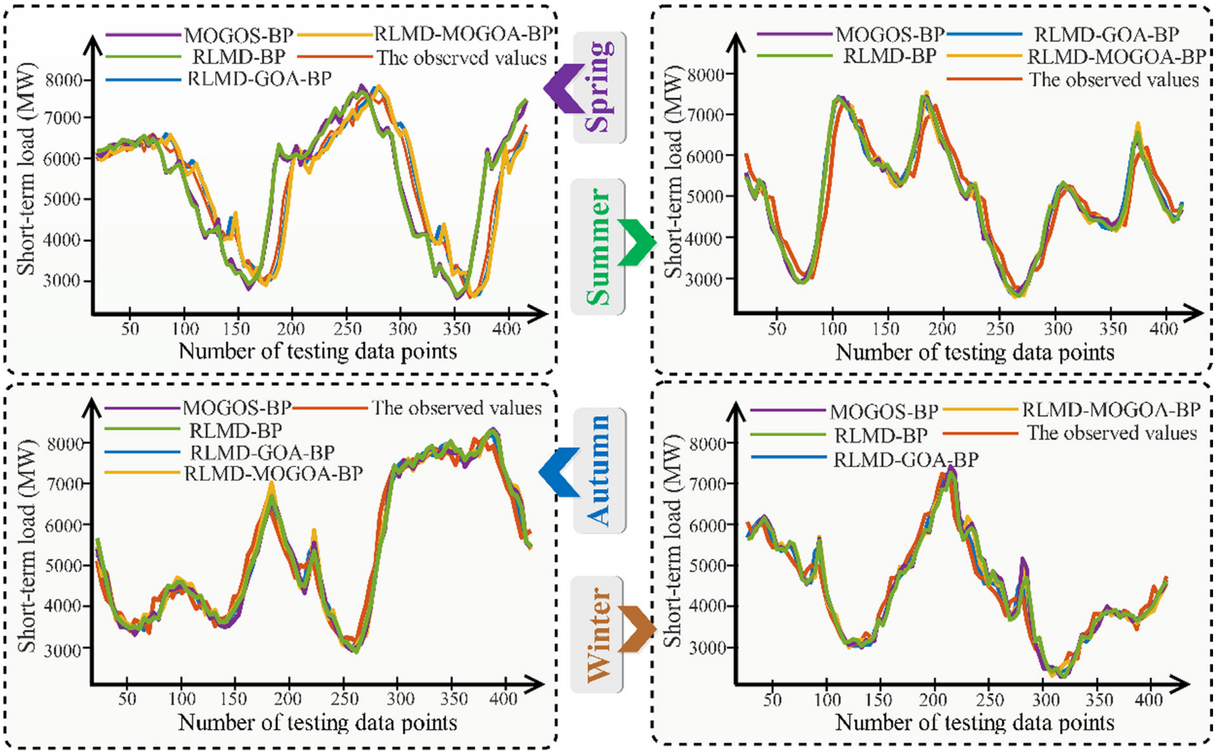

Fig. 3. The experimental results of the proposed model and four hybrid models in the case of Victoria.

as the RLMD-BP model and MOGOS-BP model, have been investigated large forecasting bias in the Autumn case, while the proposed model has stronger persistence and accuracy. Besides, the proposed model shows the optimal trend with the observed values when combined with other comparison models.

Remarks: The case study of New South Wales further demonstrates that the proposed forecasting model is suitable for short-term load forecasting. The simulation results comprehensively verify that the forecasting model can accomplish high forecasting accuracy and robustness, and evidently, it has significant practical application ability.

4.5.3. Forecasting stability assessment

Besides forecasting errors, forecasting stability is another indicator that measures the performance of the proposed forecasting model. It is generally believed that the smaller the variance, the higher the stability. In the proposed model, the MOGOA is employed for enhancing the forecasting accuracy and robustness simultaneously. It can be seen from Table 9 that the proposed forecasting model has the lowest variance in

Table 7
Performance of multi-step-ahead load forecasting models in New South Wales (MAE, RMSE, MAPE (%)).

Sample	Horizon	One-step			Two-step			Three-step		
	Indices	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
N-Spring	BP	529.7089	279.2762	5.7885	502.4200	291.1534	5.0102	528.4126	277.4816	5.7716
	GOA-BP	524.4263	271.6288	5.1432	501.0219	283.0730	5.3556	527.6221	276.0084	5.7649
	MOGOA-BP	488.7897	234.4848	5.0445	500.5400	247.1797	5.3547	496.6040	242.1350	5.1611
	ILMD-BP	110.9198	119.7667	0.9082	249.7321	135.5275	2.3571	408.1608	151.9887	3.9863
	ILMD-GOA-BP	101.8664	98.1533	0.8364	229.8043	93.4284	**2.1840**	391.8702	139.9199	3.8296
	ILMD-MOGOA-BP	**83.7981**	**54.6077**	**0.7351**	**228.8194**	**87.3143**	2.2108	**387.8078**	**136.0640**	**3.7594**
N-Summer	BP	774.4688	643.4429	8.2923	756.5264	696.8721	8.0405	746.5230	837.0403	7.8427
	GOA-BP	715.2809	551.5419	7.3484	720.5982	615.7048	7.3402	731.5901	682.9848	7.4164
	MOGOA-BP	714.9232	424.8949	7.3432	720.5675	443.6386	7.3399	730.2887	568.0597	7.3982
	ILMD-BP	83.1491	198.2655	0.7549	292.3615	381.8710	2.6446	531.4430	406.0237	4.6498
	ILMD-GOA-BP	70.5269	96.8603	0.6158	269.9176	173.7217	2.4032	511.4113	224.4269	4.5332
	ILMD-MOGOA-BP	**68.4001**	**58.6261**	**0.5616**	**267.8809**	**72.9275**	**2.3195**	**507.9023**	**87.1930**	**4.4677**
N-Autumn	BP	646.1327	856.4574	7.1294	649.2271	964.3265	7.2128	686.7251	997.9809	7.3350
	GOA-BP	645.7025	603.3979	6.1266	648.4980	869.2675	7.2096	677.9576	899.7732	7.0399
	MOGOA-BP	652.4728	478.5862	6.0654	644.4271	530.6239	7.0338	656.1369	678.7360	7.3371
	ILMD-BP	83.6346	152.2907	0.8389	200.5854	290.7369	1.9825	336.4515	383.9349	**3.2524**
	ILMD-GOA-BP	68.9596	97.3827	0.7116	**190.9284**	161.0418	**1.8544**	**333.7341**	196.1679	3.2611
	ILMD-MOGOA-BP	**65.5574**	**64.3753**	**0.6816**	194.0096	**76.3231**	1.8979	335.5193	**111.9386**	3.2765
N-Winter	BP	835.5264	956.3082	9.0439	845.2462	988.3999	9.1498	864.4526	989.4014	9.3493
	GOA-BP	834.0804	931.6828	9.0321	844.7169	964.8302	9.1416	864.0943	985.8843	9.3441
	MOGOA-BP	772.8860	614.0613	8.3046	772.4869	653.2632	8.3153	781.5903	691.5067	8.4201
	ILMD-BP	86.3086	460.1059	0.7909	238.2354	526.3725	2.3160	412.2954	650.0954	4.0455
	ILMD-GOA-BP	81.5821	130.8789	0.8252	229.7783	138.1601	2.2438	404.0900	162.6317	3.9835
	ILMD-MOGOA-BP	**76.6759**	**84.2976**	**0.6979**	**215.7254**	**87.9627**	**2.1071**	**348.9468**	**117.3203**	**3.3601**

Table 8
Further analysis of one-step ahead load forecasting in New South Wales.

		AE	DA	FB	R^2	TIC
N-Spring	Naïve Predictor	**2.2075**	0.7665	**−0.0003**	0.9576	**0.0001**
	ARIMA	−11.5563	0.7305	0.0015	0.9355	0.0007
	BP	134.1949	0.6647	−0.0170	0.7689	0.0085
	GOA-BP	114.0209	0.7305	−0.0145	0.8099	0.0072
	MOGOA-BP	109.1966	0.8024	−0.0139	0.7462	0.0069
	ILMD-BP	−5.6630	0.7246	0.0007	0.9888	0.0004
	ILMD-GOA-BP	−4.4796	0.7246	0.0006	0.9868	0.0003
	ILMD-MOGOA-BP	−2.7019	**0.8545**	**0.0003**	**0.9924**	0.0002
N-Summer	Naïve Predictor	22.1400	0.8539	−0.0026	0.9277	0.0013
	ARIMA	−29.3745	0.7528	0.0034	0.9147	0.0017
	BP	108.0240	0.5907	−0.0127	0.4340	0.0063
	GOA-BP	89.4522	0.6292	−0.0103	0.5165	0.0051
	MOGOA-BP	89.0819	0.6292	−0.0103	0.5165	0.0051
	ILMD-BP	−11.7519	**0.8989**	0.0014	0.9853	0.0007
	ILMD-GOA-BP	−9.2414	0.8876	0.0011	0.9855	**0.0005**
	ILMD-MOGOA-BP	**−8.5165**	0.8764	**0.0010**	**0.9934**	**0.0005**
N-Autumn	Naïve Predictor	−6.9694	0.7485	0.0009	0.9568	0.0005
	ARIMA	**1.3405**	0.7665	−0.0028	0.9455	**0.0001**
	BP	124.7928	0.6350	−0.0167	0.4165	0.0083
	GOA-BP	86.7721	0.6826	−0.0113	0.6164	0.0056
	MOGOA-BP	85.8409	0.6766	−0.0111	0.6135	0.0056
	ILMD-BP	9.8916	0.7721	−0.0013	0.9935	0.0006
	ILMD-GOA-BP	6.8555	0.7006	−0.0009	0.9890	0.0004
	ILMD-MOGOA-BP	3.2599	**0.8442**	**−0.0004**	**0.9939**	0.0002
N-Winter	Naïve Predictor	9.6650	0.8333	−0.0012	0.9762	0.0006
	ARIMA	**1.3405**	0.7665	−0.0012	0.9455	**0.0001**
	BP	56.7159	0.7527	−0.0068	0.7181	0.0034
	GOA-BP	55.5225	0.7527	−0.0067	0.7191	0.0033
	MOGOA-BP	54.9644	0.7262	−0.0067	0.6917	0.0033
	ILMD-BP	−16.9945	0.8011	0.0021	0.9457	0.0010
	ILMD-GOA-BP	−6.5315	0.8280	0.0008	0.9565	0.0004
	ILMD-MOGOA-BP	2.8821	**0.8989**	**−0.0004**	**0.9905**	0.0002

Table 9
The variance of the forecasting errors (%).

Site	Models	Spring	Summer	Autumn	Winter	Average
Victoria	BP	0.0377	0.0367	0.0551	0.0504	0.0450
	GOA-BP	0.0311	0.0336	0.0519	0.0499	0.0416
	MOGOA-BP	0.0243	0.0084	0.0464	0.0380	0.0293
	ILMD-BP	0.0109	0.0060	0.0113	0.0140	0.0106
	ILMD-GOABP	0.0103	0.0056	0.0081	0.0136	0.0094
	ILMD-MOGOA-BP	**0.0093**	**0.0050**	**0.0079**	**0.0117**	**0.0085**
New South Wales	BP	0.0616	0.0575	0.0536	0.0551	0.0744
	GOA-BP	0.0533	0.0510	0.0506	0.0497	0.0704
	MOGOA-BP	0.0508	0.0510	0.0504	0.0498	0.0703
	ILMD-BP	0.0097	0.0217	0.0356	0.0219	0.0073
	ILMD-GOABP	0.0087	**0.0199**	0.0343	0.0213	0.0058
	ILMD-MOGOA-BP	**0.0070**	**0.0199**	**0.0341**	**0.0210**	**0.0056**

comparison with other benchmark models. Take Victoria as an example, the variance of the proposed model in spring, summer, autumn, and winter are 0.0093, 0.005, 0.0079, and 0.0117, respectively, and the average variance is 0.0085. Referring to GOA-BP and ILMD-BP, the average variance values are 0.0293 and 0.0106, respectively. Additionally, it is obvious that the proposed model obtains the optimal variance value in all the case studies.

Remarks: The multi-objective optimization algorithm successfully accomplishes both high accuracy and stability in short-term load forecasting. Therefore, the proposed forecasting model in the urban smart grid system can achieve desirable forecasting robustness when compared with other benchmark models.

4.5.4. *Statistical tests*

In this paper, several databases in two case studies (Victoria and New South Wales) are conducted to testify the performance of the proposed forecasting model in the urban smart grid system. To verify the distributions of paired samples are significantly different, two-

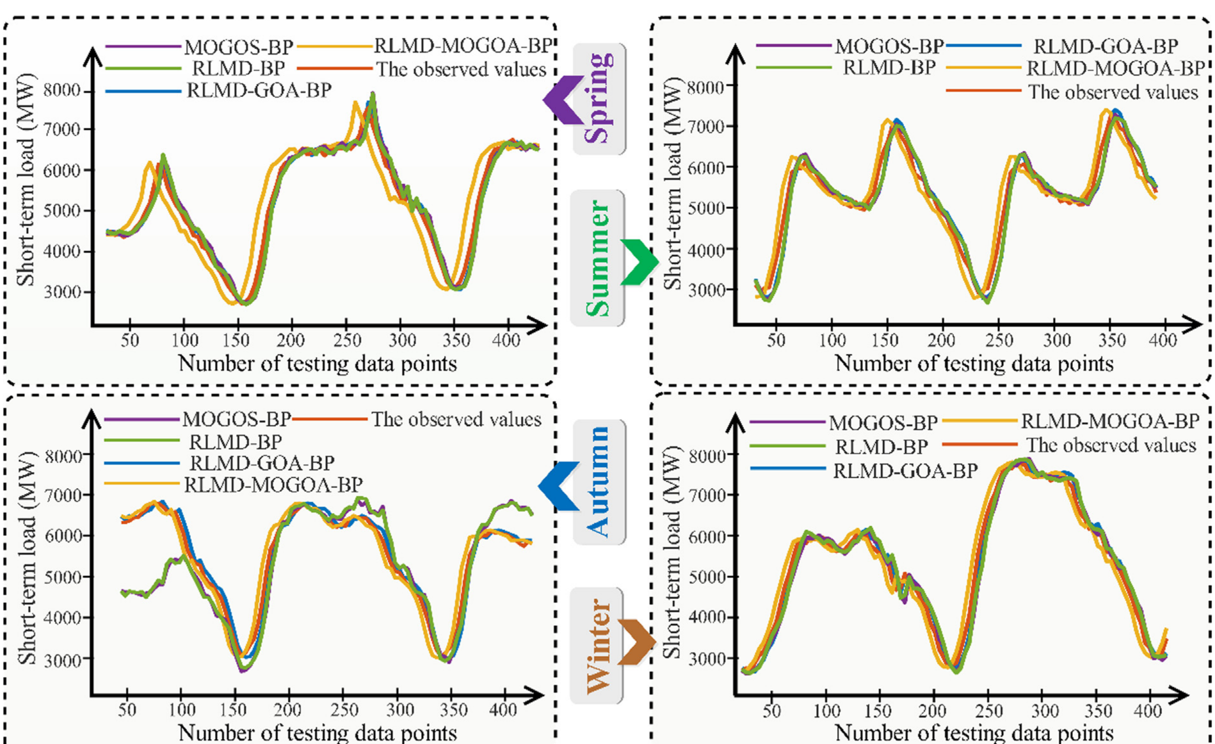

Fig. 4. The experimental results of the proposed model and four hybrid models in the case of New South Wales.

sample statistical nonparametric tests, including the Chi-square test, the *K-W* test, Friedman rank sum test, and Spearman correlation coefficient are implemented. Table 10 shows the tests results and all the databases are verified significantly distinct. The Chi-square test, *K-W* test, and Friedman rank sum test results are statistically significant and the Spearman correlation coefficient results are not highly relative. In other words, all the experiments performed in this study are highly representative because they are based on varied datasets. To conclude, the proposed forecasting model can adapt to different environments and successfully applied in the urban power systems.

The FVD reflects the comprehensive and average accuracy of the forecasting models, and the higher the FVD, the greater the average forecasting accuracy. That is to say, a model achieves a high forecasting validity when the forecasting accuracy is high in all periods. Table 11 presents the FVD values of the proposed model and the benchmark models. In comparison, the proposed model achieves the highest FVD values while single comparison models have comparatively low FVD values. In summary, the proposed model has the global best precision.

Table 11 also shows the DM test results, as well as the responding *p* values. From the experimental results it can be concluded that the developed forecasting model significantly outperforms other comparison models. The DM values of the Naïve Predictor, ARIMA, BP, GOA-BP, MOGOA-BP, ILMD-BP, and ILMD-GOA-BP model in Victoria are 8.9793, 7.0297, 6.8724, 6.2550, 4.6587, 4.0187, and 3.3389, respectively, which are all much larger than $Z_{0.01/2} = 2.58$. Admittedly, it can be observed that there is a 99% probability to accept the alternative hypothesis. That is to say, the proposed forecasting model in the urban smart grid system has a significant difference with the comparison models, where the significance level is 99%. Another case also indicates the significance of the DM test. Besides ILMD-GOA-BP (95% significant), other benchmark models are 99% significantly different from the proposed model.

Remarks: Four statistical nonparametric tests are used to verify the difference of the study datasets, and the test results indicate the high adaptability of the proposed forecasting model. Additionally, various statistical tests manifest that the proposed model achieves the lowest forecasting errors and exhibits a significant improvement in accuracy and stability compared to other models.

5. Discussion

5.1. Comparison of different train-test ratios

There are no standard rules to set up the best train-test ratio, and the PSR based on the C-C method attempts to determine the suitable train-test ratio in this paper. Admittedly, the train-test ratio determining approach cannot make sure the optimal forecasting performance, and parameter fine-tuning is also needed. However, in the past decade, few mechanisms have been studied in this field. Moreover, the fine-tuning process is often time-consuming and based on expert decisions. In this subsection, several different train-test ratios are implemented to verify the good performance of the PSR based determining method. The train-to-verify ratio, percentages and their responding forecasting performance in dataset *V*-spring and *N*-spring are shown in Table 12. In the *V*-spring database, the embedding dimension calculated by the PSR is 15, so the input-output ratio in the different train-test ratios are all set to 14:1. Similarly, the input-output ratio of the *N*-spring are all set to 24:1. It is considered from Table 12 that the PSR train-test ratio determining method can effectively select the optimal number of training and test set. Take New South Wales database as an example, the R^2 values in 2:1, 3:1, 5:1, 10:1, and 20:1 train-test ratio are 0.7660, 0.7561, 0.8230, 0.7448, and 0.7816, respectively, which are inferior to the train-test ratio that determined by the PSR.

5.2. Algorithm tests

This subsection is aimed at comparing three multi-objective algorithms (*i.e.* MODA, MOPSO, and MOBA) with MOGOA to prove the effectiveness and efficiency of the MOGOA. Four test functions (described in Appendix B), two measurement indicators (IGD and SPC) are employed to evaluate the fitting performance of multi-objective algorithms. The formula of IGD and SPC are presented below, where d_i represents the Euclidean distance between the i_{th} true Pareto optimal solution and the nearest ones obtained by algorithms, and N is the number of true Pareto optimal solutions.

$$\mathbf{IGD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} d_i^2} \tag{20}$$

$$\mathbf{SPC} = \sqrt{\sum_{i=1}^{n} (\bar{d} - d_i)^2/n - 1} \tag{21}$$

The experimental parameters setting in the test is noted as follows: The iteration number is 100, and the number of search agents and the archive size are 200 and 100, respectively. The statistic values of IGD and SPC are displayed in Table 13. It can be concluded that the MOGOA accomplishes the better fit performance in terms of all the statistical characteristic of IGD and SPC.

5.3. Comparison of each component of forecasting model

In this subsection, the contributions of each component of the proposed forecasting model are compared and discussed in Table 14. An improvement percentage analysis is used to illustrate the relative improvement of two paired models in *V*-spring and *N*-spring databases. It appears that the considerable improvements by each component. From the comparison of BP and GOA-BP, it is found that the GOA optimization algorithm plays a significant role in enhancing the forecasting accuracy, with the average improvement of 8.3898 in RMSE, 5.3819 in MAPE, 10.3966 in R^2, and 5.2149 in DA, respectively. Whereas the MOGOA is superior to the GOA, with the average improvement of 18.1554 in RMSE, 4.9257 in MAPE, 2.8623 in R^2, and 1.6716 in DA respectively, according to the comparison results of MOGOA and GOA. When referring to the comparison of BP and ILMD-BP, it can be clearly seen that the data cleaning method ILMD makes larger contributions in improving forecasting performance than the MOGOA. In conclusion, the combination of data preprocessing and multi-objective optimization achieves a great improvement in terms of the various measurement matrix.

5.4. Real applications of this study

The balance of power supply and demand is considered as a critical task in power systems. Overload will cause an increase in start-up and

Table 10
Statistical tests for different study samples.

Indices	Chi-sq	K-W	Friedman	Spearman
V-Spring and V-Summer	0.0017	0.0029	0.0000	0.6138
V-Spring and V-Autumn	0.0000	0.0094	0.0000	0.5559
V-Spring and V-Winter	0.0000	0.0000	0.0000	0.7458
V-Summer and V-Autumn	0.0075	0.0234	0.0000	0.7346
V-Summer and V-Winter	0.0000	0.0000	0.0000	0.5567
V-Autumn and V-Winter	0.0000	0.0000	0.0000	0.6112
N-Spring and N-Summer	0.0000	0.0000	0.0000	0.6076
N-Spring and N-Autumn	0.0000	0.0000	0.0000	0.6540
N-Spring and N-Winter	0.0001	0.0019	0.0000	0.7030
N-Summer and N-Autumn	0.0000	0.0000	0.0000	0.6741
N-Summer and N-Winter	0.0000	0.0000	0.0000	0.6423
N-Autumn and N-Winter	0.0000	0.0060	0.0019	0.5637
V and N	0.0000	0.0000	0.0000	0.5247

Table 11
Summary of average DM test values and forecasting validity degrees.

Models	Victoria			New South Wales		
	DM	*p*-value	FVD	DM	*p*-value	FVD
Naïve Predictor	8.9793*	0.0000	88.2645	9.0024*	0.0000	82.6458
ARIMA	7.0297*	0.0000	89.2647	8.5784*	0.0000	86.4447
BP	6.8724*	0.0000	92.0254	5.1016*	0.0000	89.5514
GOA-BP	6.2550*	0.0000	93.0369	4.6051*	0.0000	90.5254
MOGOA-BP	4.6587*	0.0000	93.9584	4.5827*	0.0000	96.6196
ILMD-BP	4.0187*	0.0000	95.3254	3.7874*	0.0000	98.7092
ILMD-GOA-BP	3.3389*	0.0008	95.9685	2.0896**	0.0366	98.1245
ILMD-MOGOA-BP	–	–	98.7000	–	–	99.0122

Note: *represents the1% significance level; **represents the 5% significance level.

Table 12
The forecasting performance of the proposed model in different train-test ratio.

Case study	Train-test ratio	Percentage of training set	MAPE	R^2
Victoria	2:1	66.6667%	4.7155	0.8242
	3:1	75.0000%	3.5675	0.9125
	5:1	83.3333%	4.1795	0.8980
	10:1	90.9091%	4.1158	0.9124
	20:1	95.2381%	1.1371	0.9340
	4123:181	95.7946%	1.0936	0.9923
New South Wales	2:1	66.6667%	5.6331	0.7660
	3:1	75.0000%	4.7181	0.7561
	5:1	83.3333%	5.8369	0.8230
	10:1	90.9091%	2.6125	0.7448
	20:1	95.2381%	1.7983	0.7816
	4027:267	93.7820%	0.7351	0.9924

long-term costs due to the inherent difficulties in electricity storage. Conversely, underload will affect the quality of power supply, rendering it incapable of satisfying regular power demands and potentially compromising the safety and stability of the power system. For short-term load forecasting models, the overestimated forecasting will generate excessive electricity, whereas the underestimated forecasting will cause electrical power shortage, and that means high losses in production and people's life. The proposed forecasting model in urban smart grid

Table 14
Results of the improvement percentages of each component in the proposed model (%).

Improvement percentages	Victoria	New South Wales	Average	Victoria	New South Wales	Average
	BP vs. GOA-BP			GOA-BP vs. MOGOA-BP		
RMSE	2.9892	13.7904	8.3898	10.6045	25.7065	18.1554
MAPE	2.1574	8.6064	5.3819	6.6231	3.2282	4.9257
R^2	6.9150	13.8781	10.3966	2.1933	3.5313	2.8623
DA	4.6827	5.7470	5.2149	1.9335	1.4097	1.6716
Improvement percentages	BP vs. ILMD-BP			ILMD-BP vs. ILMD-GOA-BP		
RMSE	74.8898	65.9867	70.4383	10.1403	54.5075	32.3239
MAPE	76.2538	89.1159	82.6849	10.5268	9.2289	9.8779
R^2	14.3309	67.4139	40.8724	4.3171	0.1150	2.2161
DA	14.7562	20.9451	17.8507	0.9038	1.7487	1.3263
Improvement percentages	BP vs. ILMD-MOGOA-BP			ILMD-GOA-BP vs. ILMD-MOGOA-BP		
RMSE	79.3060	90.4256	84.8658	8.2874	38.1238	23.2056
MAPE	79.7412	91.1543	85.4478	4.6489	10.4650	7.5570
R^2	20.2041	69.8481	45.0261	0.7859	1.3375	1.0617
DA	27.4265	31.4366	29.4316	10.0465	10.6088	10.3277

systems can provide accurate and persistent forecasting results and assist policymakers to conduct effective solutions in a timely manner. Moreover, based on the forecasted short-term load values, a detailed schedule can be made to adjust the energy structure and power dispatch. If the values are larger (or smaller) than the capacity of the powerful motors, the generator sets should be adjusted to avoid damage. The proposed model also can be used in other smart grid systems in different regions around the globe since the transitions of electricity sectors are happening in most of the urban smart grid systems. By showing forecasting results with high accuracy and stability, the proposed forecasting model in this study could provide a powerful basis for international and national policies and have a large influence on transitions of energy and smart grid systems as well.

5.5. Limitations and future work

Driven by the emerging RE technologies development, urban power

Table 13
Statistic values of IGD and SPC for four test functions.

Metrics	Algorithm	ZDT_1					Algorithm	ZDT_2				
		Ave	Std.	Median	Best	Worst		Ave	Std.	Median	Best	Worst
IGD	*MODA*	0.0067	0.0022	0.0070	0.0098	0.0026	*MODA*	0.0250	0.0006	0.0249	0.0262	0.0243
	MOPSO	0.0014	0.0025	0.0014	0.0020	0.0010	*MOPSO*	0.0248	**0.0003**	0.0248	0.0254	0.0243
	MOBA	0.0021	0.0002	0.0021	0.0023	0.0021	*MOBA*	0.0251	0.0005	0.0250	0.0265	0.0242
	MOGOA	**0.0011**	**0.0001**	**0.0012**	**0.0014**	**0.0009**	*MOGOA*	**0.0172**	**0.0003**	**0.0159**	**0.0245**	**0.0141**
	Algorithm	**ZDT_3**					**Algorithm**	**ZDT_4**				
		Ave	Std.	Median	Best	Worst		Ave	Std.	Median	Best	Worst
IGD	*MODA*	0.0012	0.0070	0.0010	0.0020	**0.0003**	*MODA*	0.0018	0.0095	0.0021	0.0027	**0.0004**
	MOPSO	0.0016	0.0005	0.0016	0.0025	0.0008	*MOPSO*	0.0015	0.0003	0.0014	0.0022	0.0010
	MOBA	0.0013	0.0004	0.0012	0.0022	0.0009	*MOBA*	0.0013	0.0003	0.0014	**0.0017**	0.0009
	MOGOA	**0.0011**	**0.0002**	**0.0009**	**0.0017**	0.0008	*MOGOA*	**0.0010**	**0.0001**	**0.0010**	0.0026	0.0008
	Algorithm	**ZDT_1**					**Algorithm**	**ZDT_2**				
		Ave	Std.	Median	Best	Worst		Ave	Std.	Median	Best	Worst
SPC	*MODA*	0.0259	0.0100	0.0261	0.0447	0.0093	*MODA*	0.0336	0.0132	0.0348	0.0576	0.0126
	MOPSO	0.0818	0.0114	0.0813	0.1060	0.0658	*MOPSO*	0.0850	0.0196	0.0832	0.1248	0.0513
	MOBA	0.0117	0.0012	0.0117	0.0141	0.0100	*MOBA*	0.0263	0.0124	0.0238	0.0520	**0.0064**
	MOGOA	**0.0109**	**0.0010**	**0.0107**	**0.0138**	**0.0086**	*MOGOA*	**0.0135**	**0.0018**	**0.0131**	**0.0178**	0.0105
	Algorithm	**ZDT_3**					**Algorithm**	**ZDT_4**				
		Ave	Std.	Median	Best	Worst		Ave	Std.	Median	Best	Worst
SPC	*MODA*	0.0426	0.0653	0.0262	0.3109	0.0069	*MODA*	0.0216	0.0099	0.0212	0.0439	0.0069
	MOPSO	0.0769	0.0127	0.0792	0.0931	0.0582	*MOPSO*	0.0806	0.0196	0.0824	0.1210	0.0324
	MOBA	0.0123	0.0017	0.0120	0.0155	0.0090	*MOBA*	0.0128	0.0024	0.0129	0.0165	0.0091
	MOGOA	**0.0096**	**0.0013**	**0.0086**	**0.0126**	**0.0066**	*MOGOA*	**0.0076**	**0.0023**	**0.0079**	**0.0116**	**0.0013**

grids are anticipated to be more complex in the near future. The uncertainties of smart grid systems are increasing as a great number of factors may affect electricity demand. This paper does not focus on the future load demand in a long-term perspective, but the short-term load fluctuation. The forecasting model proposed in this study does not consider other related factors but is only based on the detailed historical short-term load. Many key impacts may be missing and there are also a big research gaps there. From the perspective of life cycle assessment (LCA), research on the whole system from "cradle-to-grave" is introduced, which can be employed into forecasting models. Moreover, scientific scenarios can be established to combine long-term forecasting with short-term forecasting, and more works also need to be done in related fields. Follow-up studies could be performed in the future work, including but not limited to:

- Additional factors or parameters can be considered in the forecasting model to enhance the short-term load forecasting effectiveness.
- Research on energy systems, especially the use of RE, needs to be studied so that the distribution and structures of future RE can be well known, which is the key factor for short-term load forecasting.
- Paying close attention to the development of data cleaning technologies to deal with irregular and unstable short-term load data so that the adverse impacts of noise can be effectively controlled.
- A dynamic model selection strategy could be considered when determining the weights of hybrid or combined models.
- LCA-based modeling and scenario analysis can be introduced into forecasting models.
- More case studies in different smart grid systems could be done to show the scalability of the proposed forecasting model.

6. Conclusion

The transitions of RE to energy and power systems are developed to achieve urban sustainability goals. However, given the volatility and intermittency of RE resources, it is challenging to propose state-of-art short-term load forecasting models. Accurate and robust short-term load forecasting in the urban grid system can reduce risks and improve the security of power systems, as well as bring more economic and social benefits. In the last decade, numerous studies have focused on improving the short-term load forecasting accuracy but few of them pay attention to forecasting persistence. ML-based forecasting models have been studied but their inner mechanism is not fully considered yet. Besides, most researches ignored in-depth data cleaning and mining, and the intrinsic characteristics of short-term load data are not further analyzed. In this paper, a newly hybrid short-term load forecasting model is developed, which takes advantage of an advanced data cleaning scheme, the PSR based on the C-C data mining method, and a neural network optimized by a multi-objective optimization algorithm. The key findings of this paper are summarized as follows.

- An improved local mean decomposition is added to the data cleaning scheme, which is verified effectively to eliminate the noise and mine the inner characteristics of the short-term load series.
- A scientific parameter determination strategy (the PSR based on the C-C method) is proposed to avoid time-consuming fine-tuning, overfitting and insufficient training problems.
- Multi-objective optimization is conducted to optimize the weight and threshold of the neural network to simultaneously enhance forecasting accuracy and stability.

In the experimental designs, both one-step and multi-step ahead rolling forecasting are implemented, and various evaluation system is conducted, including single and multiple measurements, and parametric and non-parametric statistical tests. Experimental results show that the proposed forecasting model outperforms all the comparison models in terms of almost all measurements and tests. For one-step ahead forecasting, the proposed model leads 3.65%, 3.31%, 2.08%, 0.21% and 0.09% ahead to the BP, GOA-BP, MOGOA-BP, ILMD-BP, ILMD-GOABP model, respectively, in the case of Victoria. To sum up, the proposed model makes full use of the strengths of each component and overcomes the limitations of current forecasting models in smart grid systems, which can guide policymakers for smart grid management and urban sustainability development. Its final success rendering it effective for applications not only in the power systems but also in other fields of engineering in the future.

CRediT authorship contribution statement

Chen Li: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Visualization, Investigation, Validation, Writing - review & editing.

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.

Acknowledgements

China Scholarship Council (CSC) is gratefully acknowledged for its support to Chen Li (file No. 201908210319).

Appendix A. Pseudocode of MOGOA

***Algorithm*: MOGOA**

$$\textbf{\textit{Fitness function:}}\, min \begin{cases} fitness_1 = \frac{1}{N}\sum_i \left|\frac{y_i - \widehat{y_i}}{y_i}\right| \times 100\% \\ fitness_2 = std(y_i - \widehat{y_i}) \end{cases}$$

Output: $\boldsymbol{y}_b$—the value of $\boldsymbol{y}$ with the best fitness value

/*Set the parameters of MOGOA

/*Initialize population $\boldsymbol{y}_i$ $(i = 1, 2, \ldots, N_C)$

/*Set the current iteration $t = 1$

WHILE $(t < Gen_{\text{Max}})$ **DO**

  /*Calculate the fitness function f

  /*Select the best solution $\boldsymbol{y_b}$

  /*Update the value of λ by using **Eq.** (19)

  FOR EACH $j = 1{:}\ N$ **DO**

    /*Normalize the distance between the solutions in $\boldsymbol{y}$ in certain interval

    /*Update $\boldsymbol{y_j}$ by using **Eq.** (18)

  END FOR

END WHILE

RETURN $\boldsymbol{y}_b$

Appendix B. Statistic values of IGD and SPC for four test functions

$\mathbf{ZDT_1}$	$\mathbf{ZDT_2}$
Minimize: $f_1(x) = x_1$	*Minimize*: $f_1(x) = x_1$
Minimize: $f_2(x) = g(x) \times h(f_1(x), g(x))$	*Minimize*: $f_2(x) = g(x) \times h(f_1(x), g(x))$
Where: $G(x) = 1 + \frac{9}{N-1}\sum_{i=2}^{N} x_i$	*Where*: $G(x) = 1 + \frac{9}{N-1}\sum_{i=2}^{N} x_i$
$h(f_1(x), g(x)) = 1 - \sqrt{\frac{f_1(x)}{g(x)}}$	$h(f_1(x), g(x)) = 1 - \frac{f_1(x)}{g(x)}$
$0 \leqslant x_i \leqslant 1, 1 \leqslant i \leqslant n$	$0 \leqslant x_i \leqslant 1{,}1 \leqslant i \leqslant n$
$\mathbf{ZDT_3}$	$\mathbf{ZDT_4}$
Minimize: $f_1(x) = x_1$	*Minimize*: $f_1(x) = x_1$
Minimize: $f_2(x) = g(x) \times h(f_1(x), g(x))$	*Minimize*: $f_2(x) = g(x) \times h(f_1(x), g(x))$
Where: $G(x) = 1 + \frac{9}{N-1}\sum_{i=2}^{N} x_i$	*Where*: $G(x) = 1 + \frac{9}{N-1}\sum_{i=2}^{N} x_i - \left(\frac{f_1(x)}{g(x)}\right)\sin(10\pi f_1(x))$
$h(f_1(x), g(x)) = 1 - \left(\frac{f_1(x)}{g(x)}\right)^2$	$h(f_1(x), g(x)) = 1 - \sqrt{\frac{f_1(x)}{g(x)}}$
$0 \leqslant x_i \leqslant 1{,}1 \leqslant i \leqslant n$	$0 \leqslant x_i \leqslant 1{,}1 \leqslant i \leqslant n$

References

[1] Amjady N. Short-term hourly load forecasting using time-series modeling with peak load estimation capability. IEEE Trans Power Syst 2001. https://doi.org/10.1109/59.962429.

[2] Jiang P, Liu F, Song Y. A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short-term load forecasting. Energy 2017. https://doi.org/10.1016/j.energy.2016.11.034.

[3] Lynch C, Omahony MJ, Scully T. Simplified method to derive the Kalman Filter covariance matrices to predict wind speeds from a NWP model. Energy Procedia 2014. https://doi.org/10.1016/j.egypro.2014.12.431.

[4] Ma X, Jin Y, Dong Q. A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting. Appl Soft Comput J 2017. https://doi.org/10.1016/j.asoc.2017.01.033.

[5] Dudek G. Pattern-based local linear regression models for short-term load forecasting. Electr Power Syst Res 2016. https://doi.org/10.1016/j.epsr.2015.09.001.

[6] Rendon-Sanchez Juan F, de Menezes Lilian M. Structural combination of seasonal exponential smoothing forecasts applied to load forecasting. Eur J Oper Res 2018. https://doi.org/10.1016/j.ejor.2018.12.013.

[7] Bahrami S, Hooshmand RA, Parastegari M. Short term electric load forecasting by wavelet transform and grey model improved by PSO (particle swarm optimization) algorithm. Energy 2014. https://doi.org/10.1016/j.energy.2014.05.065.

[8] Takeda H, Tamura Y, Sato S. Using the ensemble Kalman filter for electricity load forecasting and analysis. Energy 2016. https://doi.org/10.1016/j.energy.2016.03.070.

[9] Lee CM, Ko CN. Short-term load forecasting using lifting scheme and ARIMA models. Expert Syst Appl 2011. https://doi.org/10.1016/j.eswa.2010.11.033.

[10] Nie H, Liu G, Liu X, Wang Y. Hybrid of ARIMA and SVMs for short-term load forecasting. Energy Procedia 2011. https://doi.org/10.1016/j.egypro.2012.01.229.

[11] Pinson P, Madsen H. Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models. J Forecast 2012. https://doi.org/10.1002/for.1194.

[12] Amini MH, Kargarian A, Karabasoglu O. ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation. Electr Power Syst Res 2016. https://doi.org/10.1016/j.epsr.2016.06.003.

[13] Baliyan A, Gaurav K, Kumar Mishra S. A review of short term load forecasting using artificial neural network models. Procedia Comput Sci 2015. https://doi.org/10.1016/j.procs.2015.04.160.

[14] Badri A, Ameli Z, Motie Birjandi A. Application of artificial neural networks and fuzzy logic methods for short term load forecasting. Energy Procedia 2012. https://doi.org/10.1016/j.egypro.2011.12.965.

[15] Duan Q, Liu J, Zhao D. Short term electric load forecasting using an automated system of model choice. Int J Electr Power Energy Syst 2017. https://doi.org/10.1016/j.ijepes.2017.03.006.

[16] Niu M, Sun S, Wu J, Yu L, Wang J. An innovative integrated model using the singular spectrum analysis and nonlinear multi-layer perceptron network optimized by hybrid intelligent algorithm for short-term load forecasting. Appl Math Model 2016. https://doi.org/10.1016/j.apm.2015.11.030.

[17] Li Y, Che J, Yang Y. Subsampled support vector regression ensemble for short term electric load forecasting. Energy 2018. https://doi.org/10.1016/j.energy.2018.08.169.

[18] Tascikaraoglu A. Evaluation of spatio-temporal forecasting methods in various smart city applications. Renew Sustain Energy Rev 2018. https://doi.org/10.1016/j.rser.2017.09.078.

[19] Song J, Wang J, Lu H. A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting. Appl Energy 2018. https://doi.org/10.1016/j.apenergy.2018.02.070.

[20] Raza MQ, Nadarajah M, Hung DQ, Baharudin Z. An intelligent hybrid short-term load forecasting model for smart power grids. Sustain Cities Soc 2017. https://doi.org/10.1016/j.scs.2016.12.006.

[21] Liang Y, Niu D, Hong WC. Short term load forecasting based on feature extraction and improved general regression neural network model. Energy 2019. https://doi.org/10.1016/j.energy.2018.10.119.

[22] Ghadimi N, Akbarimajd A, Shayeghi H, Abedinia O. Two stage forecast engine with feature selection technique and improved meta-heuristic algorithm for electricity load forecasting. Energy 2018. https://doi.org/10.1016/j.energy.2018.07.088.

[23] Zhang X, Wang J, Zhang K. Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. Electr Power Syst Res 2017. https://doi.org/10.1016/j.epsr.2017.01.035.

[24] Barman M, Dev Choudhury NB, Sutradhar S. A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India. Energy 2018. https://doi.org/10.1016/j.energy.2017.12.156.

[25] Li W, Chang L. A combination model with variable weight optimization for short-term electrical load forecasting. Energy 2018. https://doi.org/10.1016/j.energy.2018.09.027.

[26] Hu Y, Li J, Hong M, Ren J, Lin R, Liu Y, et al. Short term electric load forecasting model and its verification for process industrial enterprises based on hybrid GA-PSO-BPNN algorithm—a case study of papermaking process. Energy 2019. https://doi.org/10.1016/j.energy.2018.12.208.

[27] Li LL, Sun J, Wang CH, Zhou YT, Lin KP. Enhanced Gaussian process mixture model for short-term electric load forecasting. Inf Sci (Ny) 2019. https://doi.org/10.1016/j.ins.2018.10.063.

[28] Mohan N, Soman KP, Sachin Kumar S. A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model. Appl Energy 2018. https://doi.org/10.1016/j.apenergy.2018.09.190.

[29] He F, Zhou J, Feng Z, Liu G, Liu G, Yang Y. A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. Appl Energy 2019. https://doi.org/10.1016/j.apenergy.2019.01.055.

[30] Huyghues-Beaufond N, Tindemans S, Falugi P, Sun M, Strbac G. Robust and automatic data cleansing method for short-term load forecasting of distribution feeders. Appl Energy 2020;261. https://doi.org/10.1016/j.apenergy.2019.114405.

[31] Wang Y, He Z, Zi Y. A comparative study on the local mean decomposition and empirical mode decomposition and their applications to rotating machinery health diagnosis. J Vib Acoust 2010. https://doi.org/10.1115/1.4000770.

[32] Zhang C, Li Z, Hu C, Chen S, Wang J, Zhang X. An optimized ensemble local mean decomposition method for fault detection of mechanical components. Meas Sci Technol 2017. https://doi.org/10.1088/1361-6501/aa56d3.

[33] Bo H, Nie Y, Wang J. Electric load forecasting use a novelty hybrid model on the basic of data preprocessing technique and multi-objective optimization algorithm. IEEE Access 2020. https://doi.org/10.1109/access.2020.2966641.

[34] Xiao L, Shao W, Wang C, Zhang K, Lu H. Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting. Appl Energy 2016. https://doi.org/10.1016/j.apenergy.2016.07.113.

[35] Johannesen NJ, Kolhe M, Goodwin M. Relative evaluation of regression tools for urban area electrical energy demand forecasting. J Cleaner Prod 2019;1(218):555–64. https://doi.org/10.1016/j.apenergy.2019.01.055.

[36] Barman M, Choudhury ND, Sutradhar S. A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India. Energy 2018. https://doi.org/10.1016/j.energy.2017.12.156.

[37] Goncalves P, Rilling G, Flandrin P. On empirical mode decomposition and its algorithms. IEEE-EURASIP Work Nonlinear Signal Image Process 2003. https://doi.org/10.1109/ICASSP.2008.4518437.

[38] Lv Y, Yuan R, Song G. Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing. Mech Syst Signal Process 2016. https://doi.org/10.1016/j.ymssp.2016.03.010.
[39] Xinzhong Z, Jianmin Z, Huiying X. A digital watermarking algorithm and implementation based on improved SVD. Proc. - Int. Conf. Pattern Recognit. 2006. https://doi.org/10.1109/ICPR.2006.59.
[40] Shi P, Yang W. Precise feature extraction from wind turbine condition monitoring signals by using optimised variational mode decomposition. IET Renew Power Gener 2017. https://doi.org/10.1049/iet-rpg.2016.0716.
[41] Jiang F, Zhu Z, Li W. An improved VMD with empirical mode decomposition and its application in incipient fault detection of rolling bearing. IEEE Access 2018. https://doi.org/10.1109/ACCESS.2018.2851374.
[42] Tang J, Liu F, Zhang W, Zhang S, Wang Y. Exploring dynamic property of traffic flow time series in multi-states based on complex networks: Phase space reconstruction versus visibility graph. Phys A Stat Mech Its Appl 2016. https://doi.org/10.1016/j.physa.2016.01.012.
[43] Kim HS, Eykholt R, Salas JD. Nonlinear dynamics, delay times, and embedding windows. Phys D Nonlinear Phenom 1999. https://doi.org/10.1016/S0167-2789(98)00240-1.
[44] Kim HS, Kang DS, Kim JH. The BDS statistic and residual test. Stoch Environ Res Risk Assess 2003. https://doi.org/10.1007/s00477-002-0118-0.
[45] Ren C, An N, Wang J, Li L, Hu B, Shang D. Optimal parameters selection for BP neural network based on particle swarm optimization: a case study of wind speed forecasting. Knowledge-Based Syst 2014. https://doi.org/10.1016/j.knosys.2013.11.015.
[46] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks. Adv. Neural Inf. Process. Syst. 2014.
[47] Pareto V. Cours d'Économie Politique; 1897. doi:10.1107/S0567740882007778.
[48] Ngatchou P, Zarei A, El-Sharkawi MA. Pareto multi objective optimization. Proc. 13th Int. Conf. Intell. Syst. Appl. to Power Syst. vol. ISAP'05. 2005. https://doi.org/10.1109/ISAP.2005.1599245.
[49] Saremi S, Mirjalili S, Lewis A. Grasshopper optimisation algorithm: theory and application. Adv Eng Softw 2017. https://doi.org/10.1016/j.advengsoft.2017.01.004.
[50] Rogers SM. Mechanosensory-induced behavioural gregarization in the desert locust Schistocerca gregaria. J Exp Biol 2003. https://doi.org/10.1242/jeb.00648.
[51] Elkholy MM, El-Hameed MA, El-Fergany AA. Harmonic analysis of hybrid renewable microgrids comprising optimal design of passive filters and uncertainties. Electr Power Syst Res 2018. https://doi.org/10.1016/j.epsr.2018.07.023.
[52] https://aemo.com.au/en.
[53] Liu Z, Jin Y, Zuo MJ, Feng Z. Time-frequency representation based on robust local mean decomposition for multicomponent AM-FM signal analysis. Mech Syst Signal Process 2017. https://doi.org/10.1016/j.ymssp.2017.03.035.
[54] Jiang P, Wang B, Li H, Lu H. Modeling for chaotic time series based on linear and nonlinear framework: application to wind speed forecasting. Energy 2019. https://doi.org/10.1016/j.energy.2019.02.080.
[55] Jiang P, Li C. Research and application of an innovative combined model based on a modified optimization algorithm for wind speed forecasting. Meas J Int Meas Confed 2018. https://doi.org/10.1016/j.measurement.2018.04.014.
[56] Hao Yan, Tian Chengshi, Wu Chunying. Modelling of carbon price in two real carbon trading markets. J Cleaner Prod 2020;244:118556. https://doi.org/10.1016/j.jclepro.2019.118556.
[57] Gur Ali O, Pinar E. Multi-period-ahead forecasting with residual extrapolation and information sharing - utilizing a multitude of retail series. Int J Forecast 2016. https://doi.org/10.1016/j.ijforecast.2015.03.011.
[58] Wang Jianzhou, Yang Wendong, Du Pei, Niu Tong. Outlier-robust hybrid electricity price forecasting model for electricity market management. J Cleaner Prod 2020;249:119318. https://doi.org/10.1016/j.jclepro.2019.119318.
[59] Niu T, Wang J, Lu H, Yang W, Du P. Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting. Expert Syst Appl 2020. https://doi.org/10.1016/j.eswa.2020.113237.
[60] Jiang P, Ma X. A hybrid forecasting approach applied in the electrical power system based on data preprocessing, optimization and artificial intelligence algorithms. Appl Math Model 2016. https://doi.org/10.1016/j.apm.2016.08.001.
[61] Yang W, Wang J, Niu T, Du P. A hybrid forecasting system based on a dual decomposition strategy and multi-objective optimization for electricity price forecasting. Appl Energy 2019. https://doi.org/10.1016/j.apenergy.2018.11.034.
[62] Wang J, Du P, Lu H, Yang W, Niu T. An improved grey model optimized by multi-objective ant lion optimization algorithm for annual electricity consumption forecasting. Appl Soft Comput J 2018. https://doi.org/10.1016/j.asoc.2018.07.022.
[63] Tapia Cortez CA, Saydam S, Coulton J, Sammut C. Alternative techniques for forecasting mineral commodity prices. Int J Min Sci Technol 2018. https://doi.org/10.1016/j.ijmst.2017.09.001.
[64] Nowotarski J, Weron R. Recent advances in electricity price forecasting: a review of probabilistic forecasting. Renew Sustain Energy Rev 2018. https://doi.org/10.1016/j.rser.2017.05.234.
[65] Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 1952. https://doi.org/10.1080/01621459.1952.10483441.
[66] Möttönen J, Hüsler J, Oja H. Multivariate nonparametric tests in a randomized complete block design. J Multivar Anal 2003. https://doi.org/10.1016/S0047-259X(02)00068-4.
[67] Puth MT, Neuhäuser M, Ruxton GD. Effective use of Spearman's and Kendall's correlation coefficients forassociation between two measured traits. Anim Behav 2015. https://doi.org/10.1016/j.anbehav.2015.01.010.