CityU
香港城市大學
City University of Hong Kong
專業 創新 胸懷全球
Professional·Creative
For The World

CityU Scholars

Universities power energy management
A novel hybrid model based on iCEEMDAN and Bayesian optimized LSTM

He, Yaqing; Tsang, Kim Fung

Published in:
Energy Reports

Published: 01/11/2021

Document Version:
Final Published version, also known as Publisher's PDF, Publisher's Final version or Version of Record

License:
CC BY

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.1016/j.egyr.2021.09.115

Publication details:
He, Y., & Tsang, K. F. (2021). Universities power energy management: A novel hybrid model based on iCEEMDAN and Bayesian optimized LSTM. *Energy Reports*, *7*, 6473-6488. https://doi.org/10.1016/j.egyr.2021.09.115

Citing this paper
Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights
Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission
Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.

Download date: 18/05/2025

Energy Reports 7 (2021) 6473–6488

Contents lists available at ScienceDirect

Energy Reports

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

Universities power energy management: A novel hybrid model based on iCEEMDAN and Bayesian optimized LSTM

Yaqing He *, Kim Fung Tsang

Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR 999077, China

ARTICLE INFO

Article history:
Received 10 August 2021
Received in revised form 22 September 2021
Accepted 26 September 2021
Available online 8 October 2021

Keywords:
iCEEMDAN
Long short-term memory (LSTM)
Bayesian optimizer
Short-term load fore-casting
University power consumption
Deep learning

ABSTRACT

Rapid growth and development around the world will lead to a gradual increase in electricity consumption. At present, colleges and universities have become the primary unit of daily electricity consumption. Therefore, accurately predicting the power consumption of colleges and universities is of great significance to the energy conservation and emission reduction of colleges and universities. Taking the actual power consumption of colleges and universities as an example, this article first analyzes its power consumption data characteristics. Based on the analysis and "decomposition and integration" concept, this paper proposes a hybrid network based on the improved complete ensemble empirical mode decomposition with adaptive noise (iCEEMDAN) and long short-term memory (LSTM) to achieve accurate colleges and universities Short-term load forecasting. First, the original power consumption data is decomposed into a series of patterns with noticeable differences by iCEEMDAN. Then use Bayesian-optimized LSTM to predict each mode individually. Finally, the prediction results of each mode are superimposed and reconstructed to form an overall prediction result. In each training, the Bayesian optimization algorithm is used to select the most suitable LSTM hyperparameter values to match the data characteristics of each model. At the same time, the structure of the LSTM prediction large data set is discussed. The results show that, compared with the prediction errors of other models, the proposed hybrid model can accurately predict university power consumption and provide the highest prediction ability among all survey models.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The Global Energy Outlook 2020 (Terreson et al., 2020) has pointed out that continuous population growth and economic prosperity lead to the rapid increasing of the demand for global energy; thus, electricity is expected to play a significant role in the development of society, economy, and technology (He, 2017). Maintaining the essential dynamic balance between power supply and power consumption has become a necessary guarantee for industrial production, economic development, and people's daily lives. Therefore, accurate prediction of power load is of great significance for effectively reducing costs (Guo et al., 2018), improving power system efficiency (Ružić et al., 2003), and ensuring the stability of power supply and demand (Jin et al., 2021).

As a basic unit of human education and academics and economic and political life, colleges and universities play an important role in the daily energy consumption of a country or region. On the one hand, there are a large number of universities in most countries and serve a large number of people in the daily life of a nation. There are currently 2688 universities or colleges in mainland China according to the 2019 National Education Development Statistical Bulletin recently released by the Ministry of Education of the People's Republic of China (M. of E. of the P.R. of China, 2020). The U.S., the world's largest higher education country, will have nearly 4000 universities or colleges in 2020, according to data from the National Center for Education Statistics of the U.S. (2020). On the other hand, universities and colleges consume enormous energy due to the comprehensive responsibility of education, teaching, scientific research, and other complex tasks. The data from Nationalgrid (2012) has shown that, as early as in 2012, colleges and universities in the U.S. have consumed average electricity of 18.9 kW/h per square foot of floorspace. Based on 2012 data alone, it could be estimated that a typical American higher-educational building with an area of about 50,000 square feet consumes a huge amount of energy each year; accordingly, the total energy consumption of colleges and universities in the United States is an astronomical figure. Besides, the annual energy consumption amount of the universities and colleges will continue to grow. Therefore, energy conservation and emission reduction work in universities is imperative, and accurate prediction of energy consumption in universities is of great significance to the energy management of universities and

* Corresponding author.
E-mail addresses: yaqinghe2-c@my.cityu.edu.hk (Y. He), ee330015@cityu.edu.hk (K.F. Tsang).

https://doi.org/10.1016/j.egyr.2021.09.115
2352-4847/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

even the energy conservation and emission reduction work in the country where they are located.

Power load forecasting can be divided into short-term load forecasting (STLF), medium-term load forecasting, and long-term load forecasting (Zhou et al., 2021) according to the different periods. STLF predicts power demand hours, days, or weeks in advance, which is the most relevant to power load forecasting, and has higher accuracy and reliability for effectively handling daily operations, procurement planning and evaluation, and reducing power waste (Friedrich and Afshari, 2015). Therefore, this study will also be conducted from the perspective of STLF.

Numerous approaches have been proposed in abundant research works power load prediction. Early engineering methods use very complex physical models, which require detailed information on walls, roofs, windows, HVAC systems, building geometry, thermal settings, internal occupancy loads, lighting, equipment, and weather conditions to calculate energy consumption (Zhao and Magoulès, 2012). Representatives of this type of method are DOE-2 (York et al., 1984), EnergyPlus (Crawley et al., 2001), ESP-r (Strachan et al., 2008). Some of them have gained accurate predictions for building load (Thevenard and Haddad, 2006; Lam et al., 2008; Eskin and Türkmen, 2008) and have contributed to the standardization for calculating the energy consumption of buildings and their components (Standard et al., 2008; ISO, 2017). However, tedious professional work is required when using these tools, making it difficult to perform and cost-inefficient (Zhao and Magoulès, 2012). Besides, the lack of detailed information when using these tools will lead to a low accuracy (Deb et al., 2017; Amasyali and El-Gohary, 2018). Therefore, the early engineering methods are complicated and costly in both categories and preparation, which is unsatisfactory.

Then the statistical prediction methods were proposed to solve the forecasting problem of power load time-series data—electricity consumption data could be regarded as time-series data observed at equal time intervals (Box et al., 2013). The operation of these methods, such as multiple regression analysis (Fumo and Rafe Biswas, 2015; Amral et al., 2007), Kalman filtering (Paliwal and Basu, 1987; Yi et al., 2017), exponential smoothing (Liu et al., 2019; Christiaanse, 1971), weighted moving average (Holt, 2004), auto-regressive (AR) (Fan et al., 2019a), autoregressive moving average (ARMA) (Liao et al., 2020), and autoregressive integrated moving average model (ARIMA) (Mohamed and Bodger, 2005; Sen et al., 2016; Box and Pierce, 1970), have received delighted forecasting performance. However, these forecasting performances result heavily from the stationarity, regularity, and linearity of the historical data. As a result, although it reflects a certain periodicity, the time series of power load is more likely to reflect complexity, irregularity, non-stationarity, and nonlinearity. Therefore these methods are not as ideal as possible for the power load forecasting effect.

The statistical methods cannot deal with the nonlinear data with hidden features, and the effect is not ideal for the nonlinear unsteady data. By contrast, the machine learning algorithms, such as artificial neural network (ANN) (Ahmad et al., 2014), support vector machine (SVM) (Chen et al., 2016), decision trees (Yu et al., 2010), random forest (Pham et al., 2020). Ahmad et al. gave a review of the application of ANN and SVM in electrical energy consumption forecasting (Ahmad et al., 2014). By projecting data into higher dimensions using the kernel function, SVM could extract the data features in the more elevated dimensional spaces, strengthening the ability to extract features and build models (Chen et al., 2017, 2016). The building energy demand predictive model based on the decision tree method can also help users quickly extract helpful information except for accurate predictions (Yu et al., 2010). Prediction models based on random forest algorithms are also proposed by Pham et al. (2020) and Wang et al. (2018). These methods, however, have the limited ability to deal with nonlinear and complex relationships.

In recent years, deep learning neural networks have performed better on this issue. Based on the recursive feedback network structure, recurrent neural network (RNN) could consider the time correlation of the time series to carry out more comprehensive and complete modeling of the time series (Zou et al., 2005). However, the structure of RNN will cause gradient explosion or gradient vanishing (Hochreiter and Schmidhuber, 1997) and cannot maintain the long-term dependence well (Chen et al., 2001). The improved deep neural network performs well on this problem. Derived from RNN, the long-short term memory (LSTM) introduces gate control to ensure the ability to remember long-term data, achieving better performance than other RNNs on practical problems (Yu et al., 2017). Thus it has been used in the solution to the electricity consumption prediction problem. LSTM achieves better performance than other RNNs on practical problems (Amarasinghe et al., 2017) and has been used in the solution to the electricity consumption prediction problem (Rahman et al., 2018; Bedi and Toshniwal, 2019; Zheng et al., 2017; Ugurlu et al., 2018; Han et al., 2019; Kong et al., 2019; Wang et al., 2020a). Gated recurrent unit (GRU) is regarded as an improved and optimized model of LSTM. With a similar internal structural unit to LSTM, GRU also effectively solves the gradient disappearance and explosion problem. However, the prediction effect of GRU is not much different from LSTM, although fewer parameters are used (Liu et al., 2017; Yang et al., 2019). In addition, some recent developments in power load forecasting are summarized as follows in Table 1.

When using deep learning neural networks to predict energy consumption, two points need to be carefully considered. One is the hyperparameter selection, and the other one is prediction strategy selection. The predictive output of deep learning neural networks is heavily dependent on the learning rate and other hyperparameter optimization issues. These hyperparameter values are cumbersome, and manual calculation will inevitably harm the accuracy of prediction (Zheng et al., 2017; Han et al., 2019; Kong et al., 2019). The cost of the grid search increases exponentially with the increase of the number of parameters, so this method faces performance problems for the case of more hyperparameters (Shahriari et al., 2016). Random search is another often preferred method (Bergstra and Bengio, 2012). However, a new point sampled in the hyperparameter space using a random search algorithm may be close to the previously evaluated point, resulting in limited information. Besides, random sampling is as likely to sample in a promising area as in a hyperspace area dominated by poor-performing models. All of them result in a decrease in optimization efficiency (Zegers and Van Hamme, 2019). Particle swarm optimization (PSO) algorithm (Kennedy and Eberhart, 2006) is also a commonly used optimization algorithm, which has the characteristics of wide adaptability, fast convergence speed, and simple operation. However, PSO easily falls into the local extreme points for functions with multiple local extreme points and cannot guarantee convergence to the global optimal point (Li et al., 2020b), so sometimes, the optimal result cannot be obtained. In addition, LSTM can achieve different prediction effects based on different hierarchical structures, and the number of layers is not strictly positively correlated with prediction accuracy. Therefore, for the power consumption prediction, the hierarchical structure of the LSTM model also needs to be seriously considered. However, the selection and optimization of LSTM structure through a manual adjustment will inevitably decrease prediction efficiency. Therefore, adjusting the appropriate hyperparameters and hierarchical structure simultaneously is the key that the appropriate optimization algorithm enables the power consumption model to quickly and

Table 1
Summary of the recent work on STLF.

Author and the work	Technique	Dataset	Metric
Zhang et al. (2018)	• Improved empirical mode decomposition • Autoregressive integrated moving average • Wavelet neural network • Fruit fly optimization	Australia and New York City (Energy consumption and temperature data)	MAE, MAPE, MPE, and RMSE
He et al. (2019)	• VMD • LSTM • BO algorithm	Hubei province, China (Energy consumption, temporal and climatic data)	RMSE, MAE, MAPE, and R^2
Liang et al. (2019)	• Empirical mode decomposition (EMD) • Minimal redundancy maximal relevance • General regression neural network with fruit fly optimization algorithm	Energy consumption data in Lang Fang, China	Relative Error (RE), MAE, RMSE, MAPE and Theil's inequality coefficient (TIC)
Fan et al. (2019b)	• Deep learning techniques for feature engineering	Building operation data from an educational building in Hong Kong	RMSE, MAE, and CV-RMSE
Somu et al. (2020)	• Haar wavelet-based mutation operator • Sine cosine optimization algorithm • LSTM	Kanwal Rekhi building energy consumption data from the Indian Institute of Technology	MAE, MAPE, MSE, RMSE, Theil U1 and Theil U2 statistics.
Jin et al. (2021)	• Encoder–decoder architecture with the GRU recurrent neural network • BO method	Actual power load data from American Electric Power	RMSE, MAE, Pearson correlation coefficient R, normalized mean square error (NRMSE), and symmetric mean absolute percentage error (SMAPE).
Han et al. (2019)	• Time-dependency convolutional neural network • Cycle-based LSTM	Daily load of Hangzhou from January 2014 to March 2017 daily load of Toronto from May 2002 to July 2016	Training time and the mean relative error (MRE),
Xia et al. (2021)	• Improved stacked GRU-RNN • SGD algorithm with momentum	Hourly actual wind energy generation dataset in Germany with hourly weather conditions in 2016; hourly load three-year data in Spain	RMSE and MAE.
Wang et al. (2020a)	• Autocorrelation graph • LSTM	Actual power consumption data of a certain cooling system	MSE, RMSE, MAPE, MAE and Theil U statistic; Pearson correlation coefficient; Kendall rank correlation coefficient; Spearman rank correlation coefficient
Wang et al. (2021)	• Deep convolutional neural network based on ResNet • Convolutional neural network	Two laboratory buildings and an office building from the genome project building data	MAPE, MAE, RMSE; pinball score
Bedi and Toshniwal (2018)	• EMD • K-means • LSTM	The electricity consumption data of the city Chandigarh, India	RMSE and Absolute Percentage Error
Nie et al. (2020)	• Complementary ensemble empirical mode decomposition (CEEMD) and singular spectrum analysis (SSA) • Gray wolf algorithm • A combined forecasting mechanism composed of RBF, GRNN, and extreme learning machine (ELM)	Three datasets of half-hour power load of Queensland, South Australia, and Victoria in Australia	MAE, RMSE, MAPE, MSE, the sum of square error (SSE), and directional change (DC)
Shang et al. (2020)	• Whale optimization algorithm (WOA) • Least squares SVM, ELM, and generalized regression neural network	10-week electric load time series from New South Wales, Australia	Average error, MAE, MAE, RMSE, a normalized average of the squares of the errors NMSE, MAPE
Li et al. (2021)	• EEMD • Multivariable linear regression • LSTM	Data from the west area of China, Uzbekistan, and PJM	MAPE, RMSE
Liu and Gao (2019)	• EEMD • WOA • SVM	30-min of electric load data from power stations in New South Wales and Queensland	MAE,MAPE, RMSE, Willmott's Index, the Nash–Sutcliffe coefficient, the Legates and McCabe Index

effectively achieve the expected prediction effect. The Bayesian optimization (BO) method happens to be able to meet these two requirements at the same time, if the number of structural layers of the LSTM is regarded as one of its hyperparameters. BO is a powerful method for extrema searching of a given objective function which estimates the objective function as a Gaussian process and regards it as a proxy function (Aoki, 1965). When some observation could be derived from the objective function, the true evaluating direction could be determined by optimizing the proxy function (Snoek et al., 2012). Therefore, BO is believed has the great potential in the optimization of LSTM for STLF.

On the other hand, the prediction strategy for STLF needs to been considered as well. Due to the complex time-series prediction problem, it is not always possible to find the inherent regulation directly using a single forecasting model, which likely affects forecast accuracy. The signal decomposition method is introduced into the power prediction solution. The decomposition operation and machine learning method are combined for the concept of "decomposition and integration" (Bedi and Toshniwal,

2018). Various hybrid models have also improved the overall prediction accuracy and formed more effective forecasting solutions. A commonly used signal decomposition method is variational mode decomposition (Dragomiretskiy and Zosso, 2014), which is effective and has been widely adopted (Liu et al., 2014; Sun and Zhao, 2020; Wang et al., 2020b). However, the number of decomposition signals obtained by this method needs to be manually selected, increasing the decomposition results' subjectivity and uncertainty. By contrast, the empirical mode decomposition (EMD) method provides an alternative selection (Huang et al., 1998). The technique has creatively offered a practical, fully adaptive solution both for nonlinear and nonstationary signals. In recent years, a variety of modifications of EMD has been proposed to enhance the performance of the method. Ensemble empirical mode decomposition (EEMD) was then proposed (Wu and Huang, 2014) to solve mode mixing to get a more accurate estimation. The method of complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) (Torres et al., 2011; Colominas et al., 2014b) was then proposed. Compared with EEMD, CEEMDAN improves the reconstruction accuracy by adding white noise while controlling the number of IMFs. The emergence of iCEEMDAN reached the pinnacle of the empirical mode decomposition method. It almost perfectly solved the problem that decomposed IMFs will contain some residual noise, thus achieving almost perfect signal decomposition (Colominas et al., 2014a). This method has been initially used for wind speed forecasting (Duan et al., 2021; Zhang et al., 2021), streamflow forecasting (Sibtain et al., 2020; Zhao et al., 2021), air quality forecasting (Li and Zhu, 2018), and price forecasting (Niu et al., 2021), even for some STLF problems (Li et al., 2020a). Still, there are few reports of the method combining with LSTM used for power load consumption forecasting.

In short, establishing a proper LSTM model with an appropriate prediction strategy, advanced technique, and reasonable optimal hyperparameters to obtain the best STLF prediction performance is still a problem that needs to be unscrambled in the current power consumption prediction. In response to this problem, this paper proposes a new hybrid model of iCEEMDAN-BO-LSTM for universities' energy consumption prediction and management. Based on the concept of "decomposition and integration", this solution combines iCEEMDAN, the latest decomposition application, and LSTM neural network with powerful predictive ability. BO is used to optimize the hyperparameters and the structure of LSTM adaptively, aiming to improve prediction efficiency. The contributions of this paper could be concluded as follow:

- Based on the importance and characteristics of university energy consumption, the university campus is innovatively used as the research object of energy consumption, and for which a feasible and efficient forecasting scheme iCEEMDAN-BO-LSTM based on "decomposition-prediction-ensemble";
- The proposed hybrid model is automatic, adaptive, and posteriori. iCEEMDAN could adaptively decompose predicting time series with a certain number of IMFs that is definitely decided by its length, which is different from the method that requires priori knowledge such as VMD. The Bayesian optimizer enables LSTM to explore the optimal hyperparameters automatically for the hybrid model to reduce the predicting errors and increase the predicting efficiency. All the above processes are entirely based on the predicting dataset.
- The structure of LSTM for this kind of prediction problem is verified and confirmed. In the study, BO was used to ascertain the contribution of different input layer structures to the prediction accuracy. The final results determined the characteristics of the LSTM structure of similar complex time series, which provided a reference result for this less concerning issue.
- A step-by-step comparison is used to logically verify the superior performance of the iCEEMDAN-BO-LSTM proposed in this paper. According to the characteristics of the targeted data, the first-step comparison is performed between the basic prediction models, which verifies the superiority of the BO-LSTM, and then the comparison is conducted among models combining empirical model decomposing methods. In this way, the superiority of the algorithm selection of each step of the hybrid model proposed in this study is tested, and the rationality of the satisfactory predictive effect of the whole hybrid model is verified.

The remainder of this paper is organized as follows. The fundamental introduction and characteristic analysis of the adopted yearly power load data are investigated in Section 2. The theoretical introduction, including iCEEMDAN, LSTM, BO, and the proposed model structure and performance assessment indices for the hybrid model, are elaborated in Section 3. Experimental design and computational results are presented in Section 4, and corresponding discussions are demonstrated in Section 5. In the last, the conclusion of this work is summarized in Section 6.

2. Data introduction and analysis

Because of the establishment of an entire Internet of Things information collection system based on the concept of sustainable development, it is easier to collect power consumption data on campuses, such as Arizona State University (ASU). Therefore, this study selects actual power consumption data from ASU (Arizona State University, 2021) to verify the model. ASU is a famous public university responsible for higher education and academic research. In this study, the electrical power consumption data of the polytechnic campus (PTC) of ASU are used. The PTC campus has 701 buildings, providing 245,000 square feet of classroom and lab space, and serves more than 4000 students and hundreds of faculties (Arizona State University, 2021). The total electrical power consumption of PTC in the past 2020 is more than 13 million kWh, as illustrated in Table 2, and will inevitably rise year by year in the future.

The electrical consumption datasets of PTC recorded the power load from 00:00 on January 1 to 23:00 on December 31 of all 2020. The electrical consumption values are updated hourly; thus there are 24 data points recorded daily and 8784 data points in total. Table 2 provides an elucidated perception of statistical descriptions, including the number of data points, the summation, minima, and the maxima, the mean, the standard deviations, the medians, and the quartiles of the dataset. Benefited from the stable operation of the Campus Metabolism system (Arizona State University, 2021), no data points are lost in this data set, providing true and reliable complete data for electricity consumption forecasting.

The diagram of the dataset provides a clear demonstration through Fig. 1. Fig. 1(a) plots the electrical consumption curves of the PTC datasets throughout 2020, Fig. 1(b) shows the hourly power consumption data of PTC in the first 4 weeks of 2020. As these two figures illustrate, this data has a regular fluctuation within 24 h a day. It will reach the highest peak of electricity consumption throughout the day in the morning, and it will reach the second peak of electricity consumption in the afternoon. The difference in electricity consumption between weekdays and weekends is only the magnitude, but the pattern is similar. In addition, this trend repeats by week. PTC has the least daily electricity load in summer, whether it is daily consumption or seasonal consumption. There will be a rapid increase in electricity consumption in autumn, but the daily electricity load in winter and spring is still the highest throughout the year. Fig. 1(c) plots the first differential result of the dataset. As illustrates, the average value of PTC 2020 power

Table 2
The statistical description of the power consumption dataset of PTC of ASU in 2020.

Dataset	Number	Electrical consumption values (kW)						
		Sum	*Min*	*Max*	*Mean*	*SD*	*Median*	*Qtl*
PTC	8784	13 588 419.99	705.25	3405.15	1546.95	413.52	1464.88	546.10

Note: 'Min' refers to the minimum; 'Max' refers to the maximum; 'SD' refers to the standard deviation; 'Qtl' refers to the quartile.

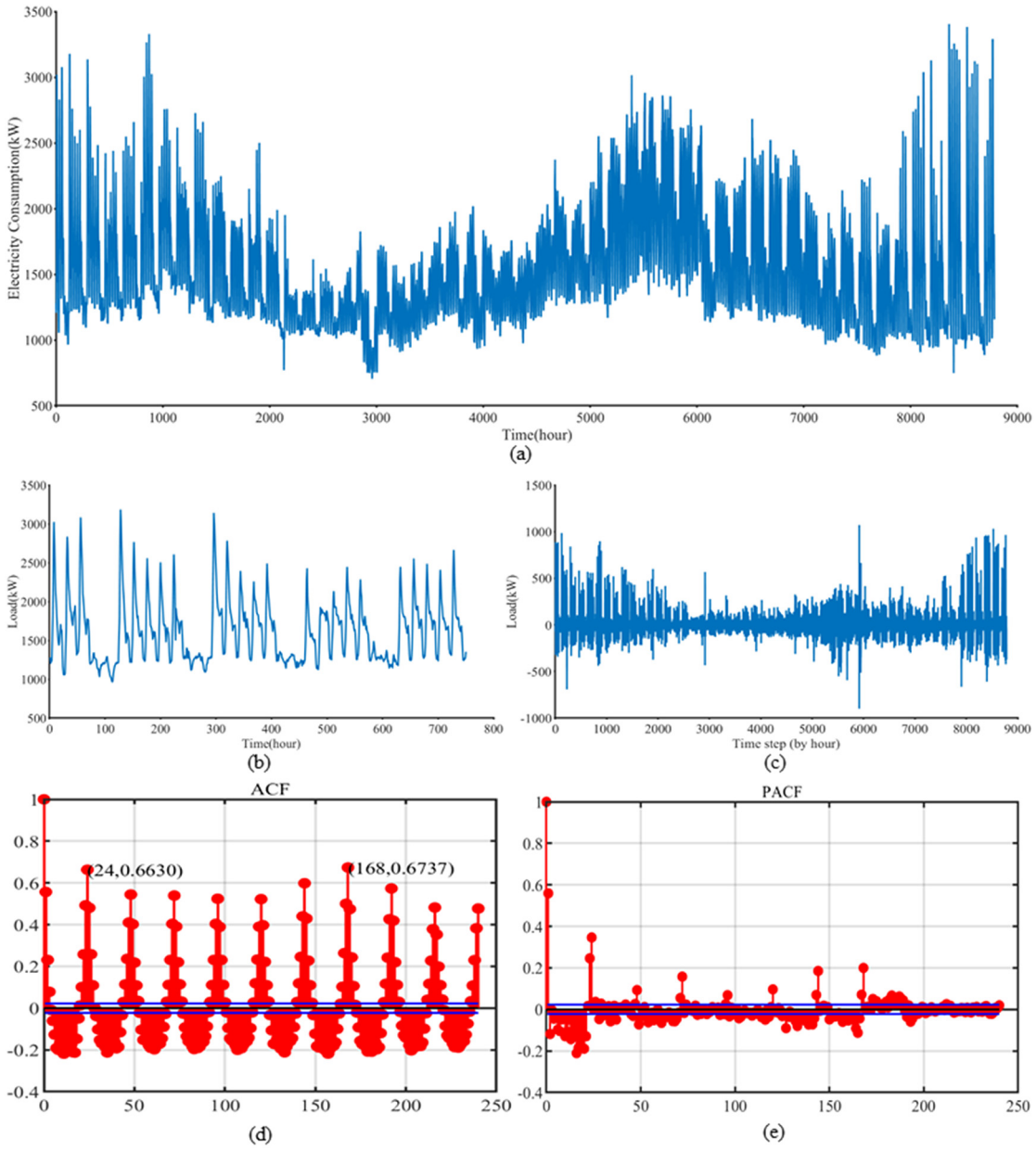

Fig. 1. Diagram exhibitions of hourly power consumption data of PTC of ASU. (a) illustrates the full view of the whole dataset; (b) gives a detailed periodic observation of power consumption data of PTC first 4 weeks of 2020; (c) shows the first differ result of throughout the year in 2020. (d) and (e) plot the results of ACF and PACF of PTC data, respectively.

consumption data returns to 0 after the first-order difference, showing an inevitable stability trend. In addition, the results of the first-order difference passed the Augmented Dickey–Fuller test (Said and Dicky, 1984) and Kwiatkowski–Phillips–Schmidt–Shint test (Kwiatkowski et al., 1992), also indicating the stationary of data after the first-order difference. Autocorrelation functions (ACF) and partial autocorrelation functions (PACF) of PTC data are plotted in Fig. 1(d) and (e), respectively. As illustrates, both a short significant daily periodicity with 24 points and a longer high weekly periodicity with 168 points could be directly observed, as marked in Fig. 1(d), showing significant periodic trends of the PTC dataset. All observations and analyzes show that PTC power load consumption has an intense cycle of 24 h (1 day) and a period of 168 h (1 week).

3. Methodology

3.1. iCEEMDAN

Empirical mode decomposition (EMD) was firstly proposed by Huang et al. (1998). Although the original EMD has creatively provided a practical, fully adaptive solution both for nonlinear and nonstationary signals, the method has a significant weakness of "mode mixing" that leads the decomposed modes contain to extensively diverse scales or contain a similar scale. This shortcoming easily leads IMFs to lose their physical meaning and reduces the robustness and sensitivity of the EMD method to minor disturbances in the dataset. Thus, an improved method named ensemble empirical mode decomposition (EEMD) was then proposed (Wu and Huang, 2014) to overcome the drawback of EMD. EEMD solved the problem of "mode mixing" to a certain extent. Still, the reconstructed signal includes residual noise, and different implementations of adding noise to the signal may result in a different number of IMFs, making the final averaging difficult in some cases. A new improvement of complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) (Torres et al., 2011; Colominas et al., 2014b) was then proposed. Differently, CEEMDAN adds a particular noise to the resulting residue from the former iteration at each decomposing stage, achieving an absolute decomposition with imperceptible reconstruction error and enables a smaller ensemble size. However, some IMFs resulting from CEEMDAN still include some residual noise. The decomposed IMFs will contain a certain number of noise and similar signal scales in the early decomposition stage. Therefore as the improved CEEMDAN, iCEEMDAN (Colominas et al., 2014a) estimates the local average means of the realizations instead of modes themselves using the modes of the additive white noise instead of the white noise itself. Defining $M(\cdot)$ for operating the local mean, and $E_k(\cdot)$ is the kth operation by EMD ($k = 1, 2, \ldots, K$, the same below), hence $X^i[t]$ could be represented as Eq. (1):

$$X^i[t] = X[t] + \beta_0 E_1\left(\omega^i[t]\right), i = 1, 2, \ldots, I \tag{1}$$

where $\omega^i[t]$ is the ith Gaussian white noise, i is the number of realizations (the same below)). The first residue could be obtained by Eq. (2):

$$r_1 = \frac{1}{I}\sum_{i=1}^{I} M(X^i[t]) \tag{2}$$

Thus, the first mode could be calculated by Eq. (3):

$$\widetilde{d_1} = X[t] - r_1 = X[t] - \frac{1}{I}\sum_{i=1}^{I} M(X^i[t]) \tag{3}$$

Similarly, the second residue could be estimated as the local means of the realizations $r_1 + \beta_1 E_2(\omega^i[t])$ and the second mode is calculated by Eq. (4):

$$\widetilde{d_2} = r_1 - r_2 = r_1 - \frac{1}{I}\sum_{i=1}^{I} M(r_1 + \beta_1 E_2(\omega^i[t])) \tag{4}$$

Similarly, the kth residue and the kth mode could be calculated by Eqs. (5) and (6), respectively:

$$r_k = \frac{1}{I}\sum_{i=1}^{I} M(r_{k-1} + \beta_{k-1} E_k(\omega^i[t])) \tag{5}$$

$$\begin{aligned}\widetilde{d_k} &= r_{k-1} - r_k \\ &= r_{k-1} - \frac{1}{I}\sum_{i=1}^{I} M\left(r_{k-1} + \beta_{k-1} E_k\left(\omega^i[t]\right)\right)\end{aligned} \tag{6}$$

The method iterates Eqs. (5)–(6) for all IMFs' calculation and stops until the obtained residue r_k *that* cannot be further decomposed either it satisfies IMF criteria or the residue has less than three local extrema. Given $\beta_k = \varepsilon_k std(r_k)$ to obtain the desired SNR selection at each stage.

3.2. Long short-term memory network (LSTM)

LSTM (Hochreiter and Schmidhuber, 1997), launched by Sepp Hochreiter and Juergen Schmidhuber in 1997, is an improved application of recurrent neural networks. LSTM allows the constant error flow in the network by implementing constant error carrousel in a complex unit called memory cells, with which the past information could be kept. Three multiplicative units—input gate (i_t), forget gate (f_t) and output gate (o_t), are designed to help the information be stored, written, or read from cells, and together with the update gate (u_t) update the current status of the cell. In this way, the structural design of LSTM not only solves the problems of gradient disappearance and explosion in traditional RNNs but also can handle longer time series compared with traditional RNNs. LSTM, thus, is more suitable for processing and predicting time series with time lags of unknown duration.

The structure of LSTM is illustrated in Fig. 2, and below give the implementation of LSTM (Hochreiter and Schmidhuber, 1997). Assuming there is an input series $X = (x_1, x_2, x_3, \ldots, x_{t-1}, x_t)$, and an output series $Y = (y_1, y_2, y_3, \ldots, y_{t-1}, y_t)$, where $t = 1, 2, 3, \ldots, T$ denoting the time step. At time lag t, the input gate i_t filters input information from x_t at current time t (Eq. (7)), and then creates an update implementation (u_t) for updating the information status (Eq. (8)):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{7}$$

$$u_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_c) \tag{8}$$

Then the forget gate f_t filters input historical information, passing the long-term trend information to $t + 1$ time and discarding unimportant information (Eq. (9)).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{9}$$

The old cell status C_{t-1} is upgraded to the new status C_t by removing partial unimportant information from historical and adding the extracted valuable information (Eq. (10)). Here, $*$ indicates the dot multiplication between matrices.

$$C_t = f_t * C_{t-1} + i_t * u_t \tag{10}$$

Finally, the output gate o_t passes updated status C_t to next cell (Eq. (11)), and meanwhile, the hidden layer status h_t of the memory cell is updated as Eq. (12):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{11}$$

$$h_t = o_t * \tanh(C_t) \tag{12}$$

Thus, the output of the final prediction result y_t is obtained by the following Eq. (13):

$$y_t = \sigma(W_y \cdot h_t + b_y) \tag{13}$$

In above, C_{t-1} and C_t denote the cell status at $t - 1$ and t time, respectively. h_{t-1} and h_t denote the hidden layer status, respectively. W_f, W_i, W_C, and W_o are the weight matrix of each layer. b_f, b_i, b_C, and b_o are the bias of each gate, and b_y is the output bias. $\sigma(\cdot)$ and $tanh(\cdot)$ are the sigmoid function and hyperbolic tangent function, separately.

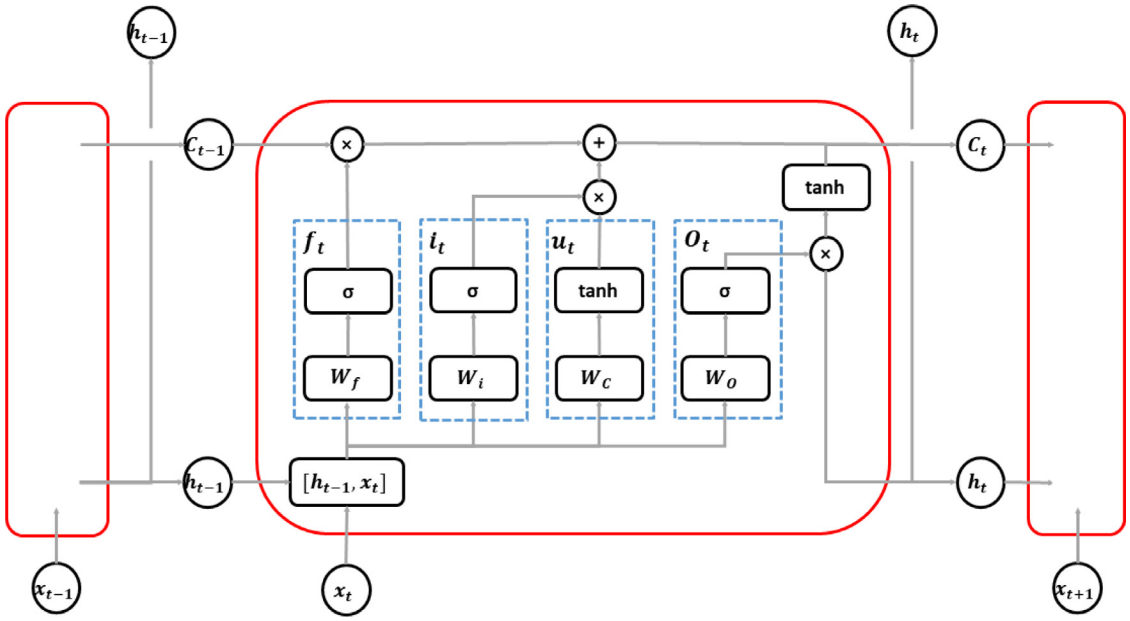

Fig. 2. Structure illustration of LSTM. ⊗ represents element-wise multiplication and ⊕ represents element-wise summation. The blue dash boxes noted with f_t, i_t, u_t, o_t denote input gate (i_t), forget gate (f_t) output gate (o_t) and update gate (u_t) respectively.

3.3. Bayesian optimizer

The hyperparameter optimization of neural networks is usually regarded as a black box problem. Although hyperparameter optimization results are helpful to evaluate the various hyperparameter configurations and improve the generalization performance of the model, the increase in algorithm complexity makes the cost of evaluation face serious challenges. In this case, selecting network hyper-parameters for the algorithm based on experience and numerous attempts is time-consuming and computationally expensive and does not always optimize the algorithm's performance (Shahriari et al., 2016). Past power consumption forecasting work has also fully exposed this problem, as mentioned in section 1 (Zheng et al., 2017; Han et al., 2019; Kong et al., 2019). Bayesian optimizer (BO) (Williams and Rasmussen, 2006) performs particularly satisfying compared to the traditional grid search method. Therefore, only a few iterations are needed to achieve an ideal effect, especially when the parameter dimension is high. More importantly, BO is more robust to non-convex problems and is less likely to fall into a local optimum. Bayesian optimization aims to find an input x^* that satisfies (Eq. (14)):

$$x^* = \arg\max_{x \in \chi} f(x) \tag{14}$$

where $\chi \in \mathbb{R}$. This process requires a prior function $p(f)$ for the objective function to be optimized, for which the Gaussian prior process (GPs) have become a standard surrogate in BO. For $p(f)$, the GPs are specified by its mean function μ and covariance function, k. Given the observation dataset $\mathcal{D} = (x_n, y_n)\,(n = 1, 2, 3, \ldots, N)$, where x_n is the nth set of hyperparameters, and y_n is the corresponding output. With the likelihood $p(\mathcal{D}|f)$, its posterior $p(f|\mathcal{D})$ thus also follows GP with the mean function μ and is a covariance function k. The kernel of GPs determines the impacts of observations on the prediction, Matérn kernel (Matérn, 1960) is the popular choice for the kernel of GPs of BO, as shown in Eq. (15):

$$k(x_i, x_j) = \frac{1}{2^{v-1}\Gamma(v)}(\sqrt{2v}\,\|x_i, x_j\|)^v K_v(\sqrt{2v}\,\|x_i, x_j\|) \tag{15}$$

where $\Gamma(v)$ and K_v are the Gamma function and the Bessel function of order v (Abramowitz, 1965). Specifically, the automatic relevance determination 5/2 Matérn kernel, recommended in Snoek et al. (2012), is adopted in this study, as shown in Eq. (16):

$$k(x_i, x_j) = (1 + \sqrt{5}\,\|x_i, x_j\| + \frac{5}{3}\|x_i, x_j\|^2)exp(-\sqrt{5}\,\|x_i, x_j\|) \tag{16}$$

Then, an acquisition function is required to evaluate the utility from the model posterior to determine the next point of input. The usual selection of the acquisition function include Expected Improvement (EI), which is selected in this work as the acquisition function, as shown in Eq. (17):

$$x_{n+1} = \underset{x}{argmax}\; a_{EI}(x|\mathcal{D}) \tag{17}$$

In summary, the BO method builds a model based on historical data to evaluate the performance of the hyperparameter sets. It then selects the new set of hyperparameters to test based on the model. This process is repeated continuously to obtain the global optimal hyper-parameter (Zhang et al., 2020).

3.4. Parametric selection, model structure, and operating steps

In this study, a hybrid model based on iCEEMDAN and LSTM optimized with BO is proposed for the STLF of a university campus.

Suppose there is an original time series $X = [x_1 \quad x_2 \quad \ldots \quad x_n]$ containing n time steps data. If one wants to determine the possible periodic impact of how many previous values affect the current predicting output value, a new hyperparameter named 'delay' is adopted and denoted as d. Therefore, another set $L = [D_1 \quad D_2 \quad \ldots \quad D_d]$, $d < n$ containing the last d step data 'borrowed' from the last year's power consumption data is used here to reconstruct the feature set. By adding the set L, the original time series set X could be transferred to the following matrix X_{re} with the dimension of $d \times n$, as formula (18)

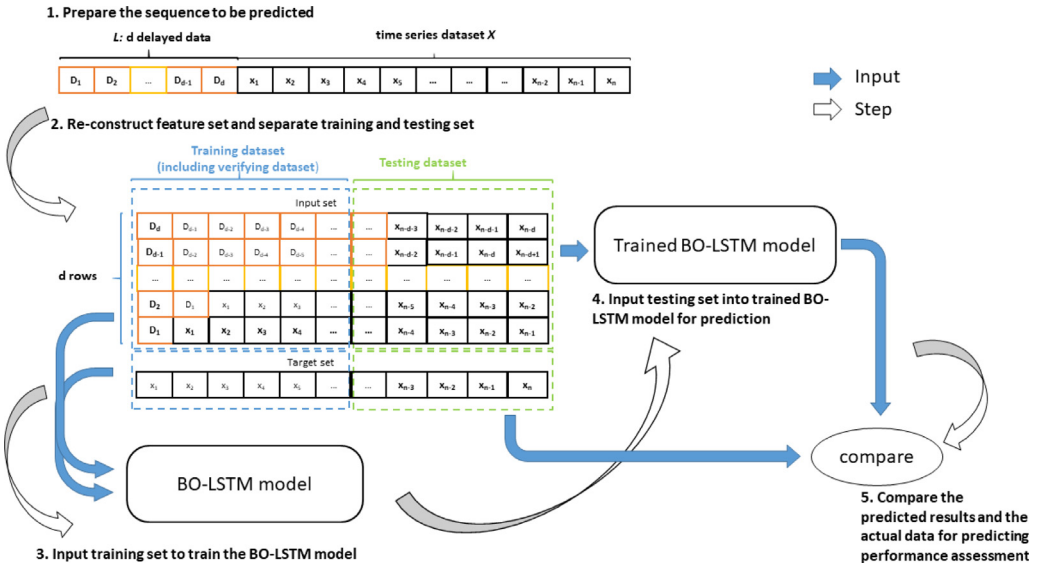

Fig. 3. The diagram of reconstructing training and testing set.

shows:

$$X_{re} = \begin{bmatrix} D_1 & D_2 & D_3 & \dots & D_d & x_1 & \dots & x_{n-d} \\ D_2 & D_3 & D_4 & \dots & x_1 & x_2 & \dots & x_{n-d+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ D_{d-1} & D_d & x_1 & \dots & x_{d-2} & x_{d-1} & \dots & x_{n-2} \\ D_d & x_1 & x_2 & \dots & x_{d-1} & x_d & \dots & x_{n-1} \end{bmatrix} \quad (18)$$

In this way, X_{re} could correspond to the original time series X in dimensionality. Thus by combining X_{re} and, a new matrix X_{pre} with $(d+1) \times n$ could be constructed as the following matrix (formula (19)) for the BO-LSTM model:

$$X_{pre} = \begin{bmatrix} D_1 & D_2 & D_3 & \dots & D_d & x_1 & \dots & x_{n-d} \\ D_2 & D_3 & D_4 & \dots & x_1 & x_2 & \dots & x_{n-d+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ D_{d-1} & D_d & x_1 & \dots & x_{d-2} & x_{d-1} & \dots & x_{n-2} \\ D_d & x_1 & x_2 & \dots & x_{d-1} & x_d & \dots & x_{n-1} \\ x_1 & x_2 & x_3 & \dots & x_d & x_{d+1} & \dots & x_n \end{bmatrix} \quad (19)$$

where the first d rows are the input to the predictor and the last row is the target time series that need to predict. In this way, the data of the first d time steps could be used to predict the current data for each time step prediction. Therefore, by using X_{re} set to predict X set after separating training and testing set for BO-LSTM model, the relationship between the previous values and the predicting value could be detected. It should be noted that for the IMFs and residual decomposed from the original time-series power consumption data, their L_{IMF} should be obtained by decomposing the combination of the original power consumption data X and its delay dataset L. Fig. 3 demonstrates the above progress, together with processes in the second step later in detail.

The parametric settings of the proposed hybrid model are elaborated as follows.

Firstly, the standard deviations of the noise set for EEMD, CEEMDAN, and iCEEMDAN are 0.2. The numbers of realization in the above decomposing methods are set as 1000 equally, and the iteration number is designated as default as infinite, following the current research outputs (Wu and Huang, 2014; Torres et al., 2011; Colominas et al., 2014a).

Secondly, the basic structure of the LSTM model in this study was composed of a hidden layer, a fully connected layer, and an output layer, and the dropout rate was defined as 0.5 to prevent overfitting that is commonly used and received satisfying performance. ADAM (Kingma and Ba, 2015) was adopted in this model to optimize the weight and bias in the LSTM network.

Considering that changes in the number of input layers in LSTM may lead to different prediction performances, in this study, the number of input layers is also set as a hyperparameter, which BO optimizes. Besides, the hyperparameters of the LSTM, including learning rate, batch size, number of hidden layer units, and L2 regularization coefficient, also need to be optimized. In addition, the above-defined hyperparameter 'delay' and denoted as d, of which the range is decided by the weekly period analyzed in Section 2. Accordingly, the hyperparameters set up in this study and their range are given in Table 3.

Table 3
Hyperparameters needs to be optimized and their given range.

Optimizing parameters	Range
Batch size	[32 1024]
Number of input layers	[1 2]
Hidden unit number	[50 200]
Delay	[24 168]
Learning rate	[1e−3 1]
L2 regularization coefficient	[1e−10 1e−2]

Fig. 4 displays the flowchart of the proposed hybrid model, the flow chart of BO, and the processing steps of the hybrid model. The operation steps of the hybrid model are introduced as follows.

1. Preprocessing. The raw data should be preprocessed for this model by the z-score normalization method. The normalized data is divided into the training set, validation set, and test set according to the ratio of 8:1:1. The training progress is set as 60 epochs, which is referred to the existing studies (Jin et al., 2021; Shang et al., 2021; Munem et al., 2020) and verified through actual calculations during the preparation work.

2. Decomposition. The preprocessed data is decomposed into n IMFs by iCEEMDAN. As an adaptive decomposition method, the number of IMFs n here is determined by iCEEMDAN directly and automatically.

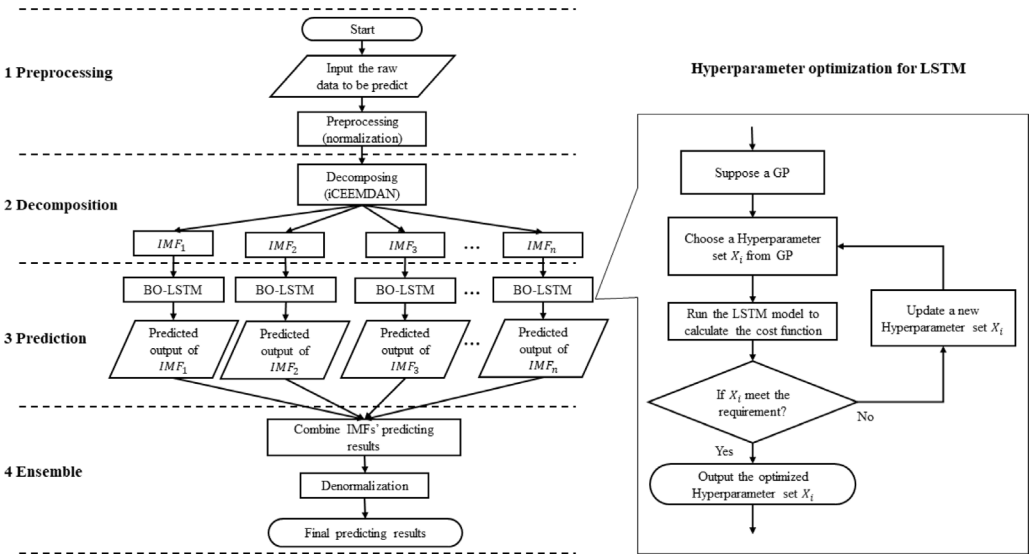

Fig. 4. The flow chart of the use of the hybrid model and the flow chart of using BO for hyperparameter optimization.

3. Prediction. In the prediction step, BO-LSTM is used to predict each IMF separately. Here, BO is used to optimize the hyperparameter selection in the prediction process of each IMFs to find the best prediction results.

4. Ensemble. In this step, all IMF prediction results are added to obtain the final prediction result of the hybrid model.

4. Experiment and result

All programming works were implemented in the Matlab 2020b environment. LSTM and BO were established using Matlab 2020b neural network toolbox deep learning library. The computational experiment is a Window workstation equipped with two Intel(R) Xeon(R) E5-2676 v3 @2.4 GHz processors, an NVIDIA GeForce GTX 1060 6 GB GPU, and 80 G RAM.

4.1. Experimental scenarios

Autoregressive methods are usually used to predict data with periodic characteristics, such as the PTC data analyzed in Section 2. These methods can consider the internal periodicity of the data and reflect it in the forecasting operation. Besides, LSTM also fits the prediction problem for periodicity, nonlinear, and nonstationary data-series data (Wang et al., 2020a). In consequence, in this study, a two-step comparison will be conducted. Firstly, the predicting performance of five basic models, including SARIMA, RNN, LSTM, GRU, and proposed BO-LSTM, are estimated. The performance comparison of the hybrid models, namely EMD-BO-LSTM, EEMD-BO-LSTM, CEEMDAN-BO-LSTM, and iCEEMDAN-BO-LSTM, are then evaluated.

4.2. Prediction performance assessment

In this study, five frequently-used statistical metrics are adopted for a comprehensive assessment of the related models, namely mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2) (Eqs. (20)–(23)). In principle, the smaller the MAE, RMSE, and MAPE, the smaller the actual and predicted values. And the closer R^2 is to 1, the more accurate the prediction result. Here, y_i and $\widehat{y_i}$ denote the actual and predicted values, and $\overline{y_i}$ denotes the average of y_i, respectively. N is the number of actual data points.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \widehat{y_i}| \tag{20}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2} \tag{21}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \widehat{y_i}}{y_i}\right| \tag{22}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{N}(y_i - \overline{y_i})^2} \tag{23}$$

The dependency of time series prediction on the actual data should also be considered, except for the prediction errors. In this study, the accuracy of time series prediction is estimated by the population correlation coefficient using Spearman rank correlation coefficient (*rho*), which is defined as the Pearson correlation coefficient between the rank variables (Myers et al., 2013). The coefficient is ranged between −1 and 1. That is, the closer X is to Y, the closer the Spearman correlation is to 1, in simple. For the predicted result $\widehat{y_i}$ and actual values y_i, each single vale is converted to ranks $rg_{\widehat{y_i}}$ and rg_{y_i}, thus ρ could be computed as (Eq. (24)):

$$rho = \frac{cov(rg_{\widehat{y_i}}, rg_{y_i})}{\sigma_{rg_{\widehat{y_i}}}\sigma_{rg_{y_i}}} \tag{24}$$

where $cov(rg_{\widehat{y_i}}, rg_{y_i})$ is the covariance of the rank variables, $\sigma_{rg_{\widehat{y_i}}}$ and $\sigma_{rg_{y_i}}$ are the standard deviations of the rank variables.

4.3. Result

4.3.1. Decomposing result of iCEEMDAN

The decomposition result of PTC power load data using iCEEMDAN is demonstrated in Fig. 5. The time series of PTC power load data is decomposed into 13 modes—12 IMFs and 1 residual. No obvious "mode mixing" phenomenon is observed in each mode. The decomposition results also show that the PTC data has apparent nonlinearity and non-stationarity. In contrast, the number of IMFs decomposed using EMD, EEMD, and CEEMDAN are 11, 13, and 14, respectively.

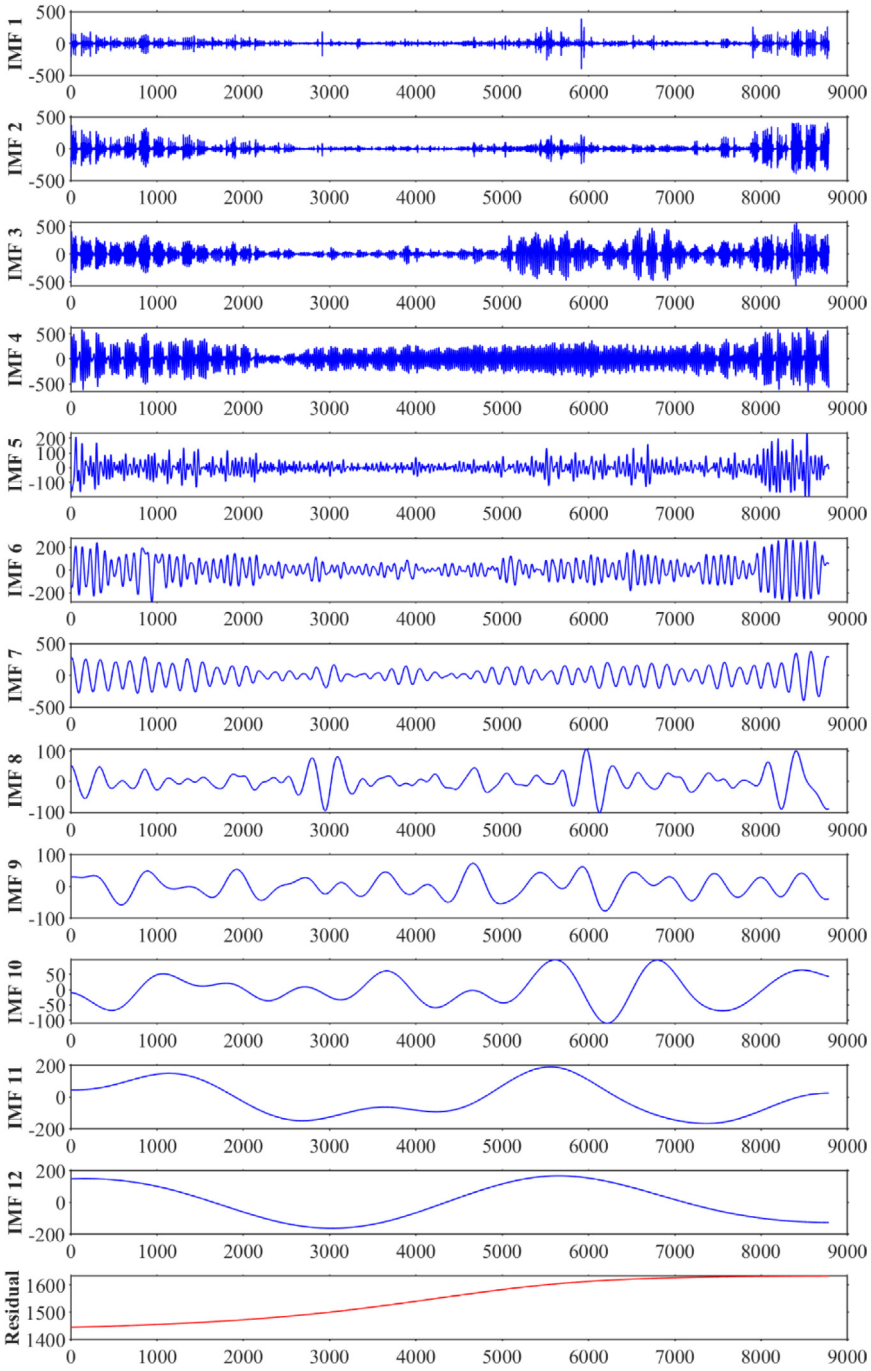

Fig. 5. Decomposition results of PTC electrical power consumption data in 2020 by iCEEMDAN.

4.3.2. Basic model performance comparison

Five basic models are firstly selected to predict the electrical power load of PTC. Here, SARIMA is optimized relying on its structural characteristics ($SARIMA(23, 1, 23)(1, 1, 1)^{168}$). RNN is optimized based on its parametric characteristics because tests have shown that the prediction effect of RNN is unsatisfactory if RNN uses the hyperparameters from LSTM optimized by BO. LSTM and GRU use part of the optimization results of BO-LSTM, and their learning rate and L2 coefficients are set according to the general experience. Noted that here the suggested number of the input layer of BO-LSTM by BO is 1.

The prediction results of each basic model on the test dataset are shown in Fig. 6. The prediction performance assessment results of the basic models are demonstrated in Table 4.

As shown in Fig. 6, each prediction result shows the overall trend of actual power consumption data. It can be observed in Fig. 6(a) that the prediction result of the BO-LSTM model is closer to the actual curve of the PTC data, so the overall prediction effect of the BO-LSTM model is better than that of the GRU, LSTM, RNN, and SARIMA models. GRU-predicted result (Fig. 6(b)) and LSTM-predicted result (Fig. 6(c)) are also closer to the actual power consumption data, but they are not as good as BO-LSTM on weekend power consumption forecasting. The prediction of RNN received a significant impact in the initial stage, as shown in Fig. 6(d). The possible reason may be that the relationship between the trend at the end of the testing set and the change at the beginning of the test set is not appropriately handled. The

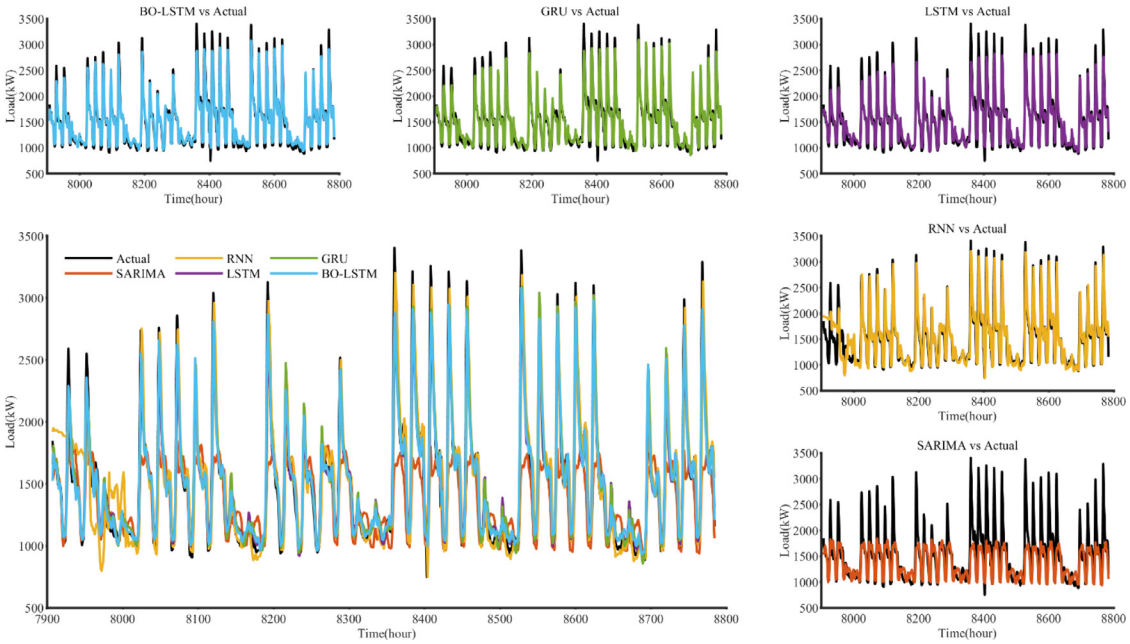

Fig. 6. Comparison of prediction results of basic models of PTC power load. (a) is the combination of the prediction results of basic models and the actual values; (b), (c), (d), (e), (f) are the comparison of the predicted result of each basic model and the actual values.

Table 4
Prediction performance assessment results of basic models.

Basic models	MAE	RMSE	MAPE (%)	R^2	rho
SARIMA	235.9974	393.6564	12.7770	0.4301	0.8086
RNN	155.8342	239.8093	9.9572	0.7885	0.8892
LSTM	93.1897	140.0133	6.6107	0.9046	0.9708
GRU	82.1126	113.4905	5.5063	0.9437	0.9698
BO-LSTM	**68.2425**	**95.1010**	**4.6458**	**0.9604**	**0.9799**

Table 5
Prediction performance assessment of the hybrid models.

Hybrid models	MAE	RMSE	MAPE (%)	R^2	rho
EMD-BO-LSTM	155.7735	203.4263	10.4137	0.8478	0.9113
EEMD-BO-LSTM	52.4634	74.7751	3.3824	0.9794	0.9832
CEEMDAN-BO-LSTM	60.5518	86.1011	3.8808	0.9727	0.9780
iCEEMDAN-BO-LSTM	**40.8411**	**59.6807**	**2.5563**	**0.9869**	**0.9905**

result predicted by SARIMA could reflect the periodic characteristics of power consumption data, as shown in Fig. 6(e), but it has limited ability to characterize the magnitude of daily power consumption changes. Prediction performance assessments listed in Table 4 also demonstrate this observation. From bold content in Table 4, it could be observed that, for single prediction models, BO-LSTM obtained the lowest MAE, RMSE, and MAPE of 68.2425, 95.1010, 4.6458%, and the highest R^2 of 0.9604, indicating the best prediction performance. For other RNN methods, GRU obtained the second-best performance with MAE of 82.1126, RMSE of 113.4905, MAPE of 5.5063%, and the second-highest R^2 of 0.9698. LSTM gets the MAE of 93.1897, RMSE of 140.0133, MAPE of 6.6107%, and R^2 of 0.9708, slightly lower than the performance of GRU. RNN got the MAE of 155.8342, RMSE of 239.8093, MAPE of 9.9572%, and R^2 of 0.7885, whose performance is much unpleasant than GRU and LSTM. SARIMA gets the unsatisfied performance with MAE of 235.9974, RMSE of 393.6564, MAPE of 12.7770%, and R^2 of 0.4301.

As for the results of rho values, BO-LSTM received the highest values of 0.9799, indicating the best prediction efficiency by the highest dependence between the predicted values and actual values. GRU and LSTM got 0.9698 and 0.9708, respectively, suggesting good prediction performances. Here the reliance of LSTM predicted data is better than that of GRU predicted, with the rho value of 0.9708 versus 0.9698. By contrast, RNN received a lower rho value of 0.8892, indicating an unsatisfied dependence between the predicted and actual values.

4.3.3. Hybrid model performance comparison

The proposed and hybrid models based on other empirical decomposition are suitable for power consumption data prediction. By using the testset to check the training model, the comparison results of the evaluation criteria predicted by each mixed model are shown in Table 5 and Fig. 7.

By combining the contents of Fig. 7 and Table 5, a consistent conclusion can be obtained. iCEEMDAN-BO-LSTM obtains the fantastic forecast performance, the parametric indices of which have been bolded in Table 5. Specifically, the MAE, RMSE, and MAPE values of iCEEMDAN-BO-LSTM are 40.8411, 59.6807, and 2.5563%, and 0.5714%, significantly lower than those of other models. R^2 of iCEEMDAN-BO-LSTM is 0.9869, higher than that of other models. This performance indicates that the level and direction accuracy of this prediction is significantly higher. EEMD-BO-LSTM also obtained good prediction results, with MAE of 52.4634, RMSE of 74.7751, MAPE of 3.3824%, and the R^2 of 0.9794. These results are close to the prediction performance of CEEMDAN-BO-LSTM, which got the MAE of 60.5518, RMSE of 86.1011, MAPE of 3.8808%, and the R^2 of 0.9727. EEMD-BO-LSTM obtained the unexpected prediction results, namely MAE of 155.7735, RMSE of 203.4263, MAPE of 10.4137%, and R^2 of 0.8478. Besides, the results of rho values could also support the above observations. iCEEMDAN-BO-LSTM obtained the highest rho value, indicating the best dependence between the actual and predicted data. In contrast, the EMD-BO-LSTM got the lowest rho, referring to the weakest dependence between the actual and predicted data. These results are also exhibited straightforwardly in Fig. 7. The above results show that the iCEEMDAN-BO-LSTM model can significantly improve the model's prediction accuracy and is more effective than other methods. The hyperparameter optimization results when using BO to predict each IMF are also given in Table 6 so that the process can be better understood. From

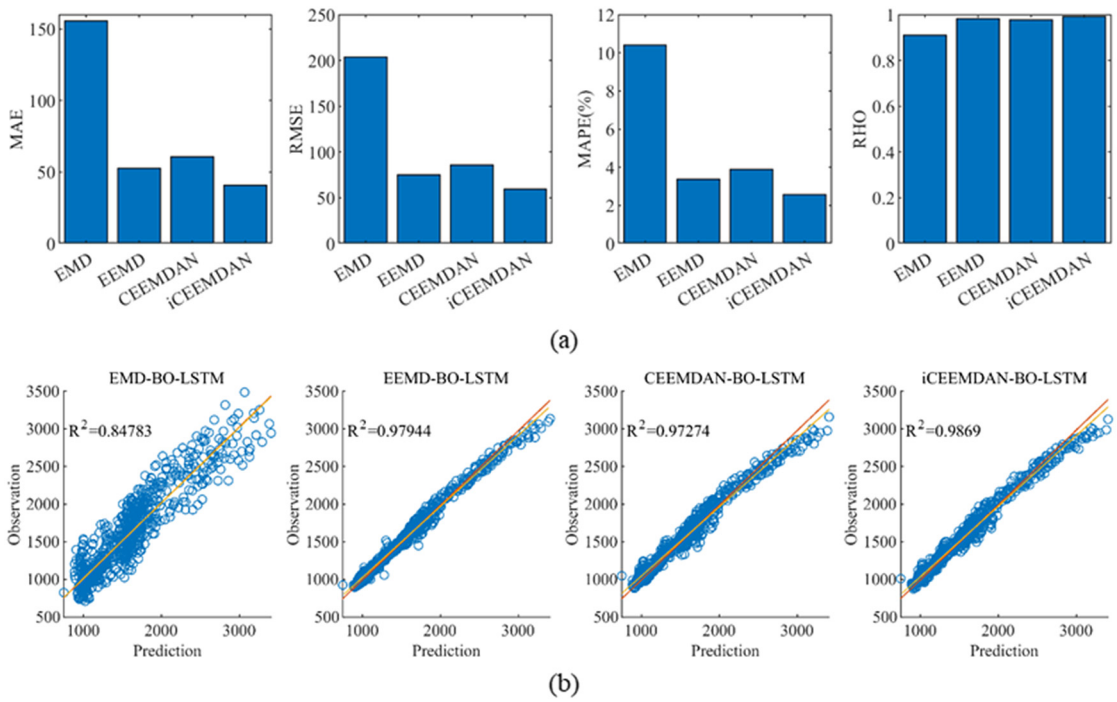

Fig. 7. Prediction results of the proposed hybrid method iCEEMDAN-BO-LSTM and other hybrid models EMD-BO-LSTM, EEMD-BO-LSTM, and CEEMDAN-BO-LSTM. (a) is the bar plots of the comparison of MAE, RMSE, MAPE, and rank correlation rho of each hybrid model; (b) is the fitted line plot of the four hybrid models where the corresponding R^2 is marked.

Table 6
The optimized hyperparameter result for the LSTM prediction of each IMF using BO.

Optimizing parameters	Batch size	Number of input layers	Hidden unit number	Delay	Learning rate	L2 regularization coefficient
IMF1	35	1	188	167	0.0118	$7.3129e^{-10}$
IMF2	156	1	151	162	0.0074	$1.4634e^{-10}$
IMF3	68	2	194	27	0.0054	$3.1055e^{-06}$
IMF4	37	1	188	39	0.0039	$4.1251e^{-09}$
IMF5	41	2	84	28	0.0010	$6.0145e^{-10}$
IMF6	36	1	113	29	0.0026	$3.0323e^{-10}$
IMF7	62	2	165	65	0.0010	$3.6869e^{-07}$
IMF8	907	1	180	24	0.0063	$1.2290e^{-10}$
IMF9	51	1	197	35	0.0013	$1.4824e^{-08}$
IMF10	56	1	197	33	0.0010	$6.1240e^{-08}$
IMF11	125	1	199	47	0.0010	$1.1693e^{-08}$
IMF12	67	1	191	55	0.0177	$1.0204e^{-10}$
Residual	479	1	193	116	0.0022	$1.1965e^{-10}$

Table 6, it could be observed that the hyperparameters optimized by BO in the process of each mode are different. No significant regulations or characteristics could be summarized. Nevertheless, the optimization using the following optimized hyperparameters leads to the satisfying final result of the PTC power consumption prediction.

5. Discussion

This study proposes a hybrid model based on iCEEMDAN and BO-LSTM for universities' short-term power consumption prediction and evaluates the prediction effect in two steps on the power consumption data of PTC throughout the year. This study first estimated the prediction performance comparison of five basic models (SARIMA, RNN, LSTM, GRU, and the proposed BO-LSTM). Secondly, the hybrid models based on different empirical mode decomposition methods, including EMD-BO-LSTM, EEMD-BO-LSTM, CEEMDAN-BO-LSTM, and iCEEMDAN-BO-LSTM, are compared with their predicting performance. Combining all the prediction results, the hybrid model of iCEEMDAN-BO-LSTM proposed in this paper can effectively perform STLF on the real power consumption data of a university campus and has achieved effective results.

First, the statistical analysis indicates that the energy consumption data from the real system has obvious daily periodicity. The correlation coefficient between measured variables is used to find strongly correlated variables with predictive output. Comparing to the objects of power forecasting in other studies, the scale of the power system of the building could be regarded as smaller than that of the university, and the scale of the city's power system is larger than that of the university. Due to the uncertain size of universities, small-scale universities can be regarded as building groups, while large-scale universities have a capacity comparable to towns. The characteristics of power consumption data of buildings and cities are first introduced. The attributes of building electricity consumption data include the following three points. First, the total annual power consumption of a single building is relatively small, ranging from a few kW to a few MW (Pham et al., 2020). Second, the "weekday–weekend" regulation of a single building is very intuitive and obvious (Lusis et al., 2017). Daily power consumption is high on weekdays, and consumption on weekends is meager (Pham et al., 2020). Third,

the seasonal weather effects can be directly observed from the average daily power consumption and the yearly power consumption trend (Somu et al., 2020; Lusis et al., 2017). When it is hot, people need to use refrigeration equipment such as air conditioners to cool down, and when it is cold, people need to use heating equipment for heating, and these are inseparable from electricity.

Electricity consumption in cities has three different characteristics. First, the power consumption in a single day is enormous, usually several hundred (Liang et al., 2019) to thousand MWh (He, 2017; Bedi and Toshniwal, 2018; Li et al., 2020b), or above (Han et al., 2019), depending on the scale of the city. Second, the "weekday–weekend" could be observed, but the separation of daily consumption is not that obvious, which may be due to the comprehensive results of the electricity consumption patterns of multiple industries in the city (He, 2017; Bedi and Toshniwal, 2018). Third, the city's electricity consumption is not only significantly affected by seasonal changes (Friedrich and Afshari, 2015; Koprinska et al., 2015), which are similar to the characteristics of a single building but also has a pronounced "calendar effect" (Guo et al., 2018). The city's total electricity consumption before and after public holidays will be very different, significantly increasing or decreasing.

The power consumption characteristics of universities have commonalities with those of the two research mentioned above objects but also have their own features. Taking the selected PTC as an example, firstly, the daily power consumption of the university is between the average daily consumption of a single building and the single-day consumption of the city, which is between a few hundred to several thousand kW. Secondly, the rule of "weekday–weekend" is evident as well. Compared with the weekdays, there are very few teaching activities on weekends, so much electricity consumption is reduced. Besides, it is also the superposition and strengthening of the apparent "weekday–weekend" rule of multiple buildings for a university campus. Thirdly, the university's power consumption is very obviously affected by the season, similar to the characteristics of a single building and a city. However, the "calendar effect" of the university is more affected by the university's teaching calendar than public holidays. Due to the above reasons, a university with a group of buildings could be used as a new study object for the power consumption problem, with new characteristic regulations of power consumption between the city and building levels.

Secondly, an appropriate sequence decomposition method is selected to decompose the original power consumption sequence into a series of simple feature sequences, which reduces the complexity of each time series and is helpful for accurate prediction. Then build a prediction model based on the LSTM network. It judges what should be retained or forgotten by introducing the concept of gates, demonstrating its excellent predictive ability for time series.

5.1. Model comparison

5.1.1. Basic model performance

In energy consumption forecasting, data usually exhibits strong periodic characteristics, which cannot be ignored. Periodicity is the fluctuation or oscillating change in a time series surrounding a long-term trend, a pattern of alternating fluctuations over time. For example, PTC energy consumption data has daily and weekly cycles or trends. Therefore, obtaining higher prediction accuracy requires the used algorithm has sufficient ability to capture the internal characteristics of the data.

The predicted performances of the five basic models are given in Table 3 and Fig. 6. The SARIMA model can effectively capture the periodic changes reflected in the data. However, as shown in Table 3 and Fig. 6, it is more difficult for SARIMA to capture and map the fluctuation range differences reflected by frequent changes in PTC data. In particular, as shown in the picture, the daily power consumption during the predicted period has suddenly increased by a significant amount. However, from the prediction results, SARIMA is not good at responding to sudden and drastic fluctuation changes. Recurrent neural network models are considered better than SARIMA at processing nonlinear and irregular data sequences. RNN shows good predictive ability in this study, but there is still a significant gap between the prediction results of LSTM and the prediction results of GRU. The possible reason is that RNN cannot deal with internal gradient changes sufficiently and has poor memory ability for information that is far from the predicted data. Therefore, the power of RNN to learn previous information is reduced, resulting in a significant decrease in predictive ability.

From the comparison results in Table 3, the prediction performance of LSTM and GRU are relatively good, both have achieved better prediction accuracy, and the prediction performance of the two is close under the same hyperparameter settings. The results show that both LSTM and GRU have the powerful ability to deal with longer dependencies in time series to achieve higher accuracy. LSTM includes some key hyperparameters, such as the number of hidden neurons, the number of hidden layer iterations, and the learning rate. The number of neurons is related to the fitting ability of LSTM, the number of iterations is associated with the training effect, and the learning rate is related to the convergence performance. This research also considers the input layer of the neural network. If one sets these parameters empirically, it will inevitably lead to the inability to effectively predict the results, as shown in the results in Table 3. Manually adjusting these parameters to find the optimal results will inevitably lead to a considerable workload. BO can make LSTM automatically find the hyperparameter combination that achieves the best prediction effect in the given hyperparameter space, reducing the time and computing cost of the best prediction performance. Such a parameter selection process is hard to be conducted by manual debugging. The prediction results in this article show that the prediction results of LSTM are slightly worse than those of GRU. As shown in Table 3, the values of MAE, RMSE, MAPE, and R^2 of the LSTM prediction results are slightly lower than those of the GRU prediction results, and only the rho is marginally higher than that of GRU. However, in principle, the two structures are very similar, so it is difficult to assert whether LSTM or GRU has the more vital predictive ability. The final prediction performance must be combined with the characteristics of the prediction data and the selection of hyperparameters, and the improvement of computing power. As shown in this study, The MAE of BO-LSTM is 16.89% smaller than that of GRU and 26.77% smaller than that of LSTM; the RMSE of BO-LSTM is 16.20% smaller than that of GRU and 32.08% smaller than that of LSTM; meanwhile, the MAPE of BO-LSTM is 15.63% smaller than that of GRU, and 29.72% smaller than that of LSTM. On the other hand, the R^2 of BO-LSTM is 1.77% higher than that of GRU and 6.17% higher than that of LSTM; the Spearman rank correlation coefficients of BO-LSTM is also 1.04% higher than that of GRU and 0.94% higher than that of LSTM. These significant prediction performance improvements have validated that the BO could enable LSTM to perform a strong predictive ability.

5.1.2. Hybrid model performance comparison

It is inefficient to use basic models to analyze and identify complex data's inherent nonlinear and nonstationary characteristics. Compared with a single prediction model, the prediction accuracy of this model combined with decomposition technology shows a considerable advantage. By comparing the results of

Table 7
Numbers of input layers of EMD-BO-LSTM, EEMD-BO-LSTM, CEEMDAN-BO-LSTM, and iCEEMDAN-BO-LSTM.

Hybrid model	Mode prediction process using 2-input layer LSTM
EMD-BO-LSTM	IMF 1
EEMD-BO-LSTM	IMF 8, 12, 13
CEEMDAN-BO-LSTM	No
iCEEMDAN-BO-LSTM	IMF 3, 5, 7

Tables 4 and 5, it can be found that, except for the EMD-BO-LSTM hybrid model, the prediction accuracy of the hybrid models is generally higher than that of the basic model of a single method. Except for the EMD-BO-LSTM hybrid model, the MAE, RMSE, and MAPE of the hybrid models are all lower than those of the basic models. The R^2 values are generally higher than those of the basic models, and the Spearman rank correlation coefficients are the same or higher than those of the rho values of basic models. In general, the decomposition technology effectively reduces the complexity of each decomposed empirical mode component of the whole power consumption data, thereby transforming the original prediction problem into predicting many modes with significantly reduced nonlinearity and complexity. This process reduces the difficulty of each prediction. In addition, each decomposed empirical mode retains some features of the original data, so under the high-performance prediction of BO-LSTM, each prediction result also contains the actual data features as much as possible. Therefore, the sum of the prediction results of all modes is closer to the original data. Thus, compared with the basic model using a single model method, the proposed model can effectively capture university power consumption data internal characteristics. Besides, it can also improve the prediction accuracy of university power consumption data with complex nonlinear and nonstationary features. The prediction performance of the hybrid model proposed in this paper is verified.

The increase in the number of LSTM layers will bring a moderate improvement in the prediction performance of the neural network to a certain extent, but at the same time, it will also bring about an exponential increase in time and memory overhead. When the number of layers is too large, the apparent disappearance of the gradient between layers will cause the update iteration of the LSTM layer close to the input layer to slow down, and the convergence effect and efficiency will drop sharply. For the above reasons, optimizing the number of input layers in this article is also carried out in a small range, and a cautious exploratory attempt is made. From the results of this study, the LSTM with two input layers was only used in the prediction processes for the following modes, as illustrated in Table 7. Only a few predictions of modes decomposed by different hybrid models used 2-input layer LSTM; thus, it shows that the LSTM with a single input layer has the solid predictive ability and can cope with most data complexity.

5.2. Comparison of iCEEMDAN and other empirical mode decomposing methods

The prediction accuracy of the hybrid model based on iCEEMDAN is significantly higher than the corresponding single model. In addition, the comparison of the prediction results of the mixed model based on different decomposition methods shows that the evaluation index of the hybrid model based on the iCEEMDAN decomposition method is significantly better than the hybrid models based on other empirical decomposing methods.

The unsatisfactory prediction effect of EMD-BO-LSTM may be due to the "mode mixing" among the multiple modes of EMD decomposition. During the training process, even if "mode mixing" occurs in a distant place, under the powerful ability of BO-LSTM, the influence of "mode mixing" will still be reflected in the prediction results, leading to the highest error in the prediction results aggregated by all modes. Therefore, in short, the EMD method is not suitable for the decomposition of the data used in this study. iCEEMDAN almost perfectly solves the problems of 'mode mixing' in EMD and EEMD. The iCEEMDAN method uses additive white noise mode estimation to achieve the local average mean value (Colominas et al., 2014a) to obtain the most satisfactory prediction data decomposition result, which helps to get the most accurate prediction result. Therefore, iCEEMDAN is considered a more valuable and practical power consumption data decomposing method. Thus, the hybrid prediction model based on the iCEEMDAN can effectively decompose the PTC power consumption data into a series of modal components with different bandwidths, which simplifies the data complexity while almost entirely retaining all the signal characteristics of the original data. The decomposition leads to the error of the hybrid prediction model based on the iCEEMDAN decomposition method is smaller, and the overall prediction performance is better. Therefore, the final prediction result synthesized from the accurate prediction results of each simulation component significantly improves the prediction accuracy.

In addition, the number of IMFs decomposed from a given time series data is only determined by the length of the time series itself and the empirical mode decomposition method used. Therefore, a fixed number of IMFs can be decomposed by a specific empirical model decomposition method for a given time series. The adaptive determination of the number of IMF helps reduce the error caused by the different number of components to be decomposed. On the other hand, due to the large capacity of the dataset used in this study, a larger number of IMFs can be decomposed, such as 14 of CEEMDAN. The large number of IMF also brings new challenges to the predicted workload and possible error increase, which will also be further optimized in future work.

6. Conclusion

Accurately predicting the university's electricity consumption is of great significance to the university's energy-saving and emission reduction. Based on the concept of "decomposition and integration", combined with the university's actual power consumption data analysis, this research proposes a new iCEEMDAN-BO-LSTM STLF hybrid method. The hybrid method consists of LSTM neural network optimized by iCEEMDAN and BO. The hybrid model first uses iCEEMDAN to decompose the target data and then predicts each decomposition mode through LSTM. BO can automatically optimize the hyperparameters of LSTM during the prediction process and consider possible LSTM structure changes to achieve better prediction accuracy. The prediction results of each mode are superimposed to reconstruct the final prediction result. The proposed hybrid model is first compared with the other four basic models using a single prediction method. Then the proposed hybrid model is compared with the other three hybrid models based on other empirical model decomposition methods, using five parameter indicators, including MAE, RMSE, MAPE, R^2, and rho. The prediction results show the effectiveness of our proposed hybrid model. BO-LSTM can effectively improve the short-term power consumption prediction accuracy of universities. Then, combined with iCEEMDAN, the proposed iCEEMDAN-BO-LSTM achieved the lowest MAE, RMSE, MAPE, and the highest R^2 and Spearman rank correlation coefficient, indicating that the hybrid model proposed in this study can further improve the prediction accuracy. The hybrid model proposed in this paper can be effectively used in the STLF of the university, providing strong support for the energy-saving and emission reduction of the university.

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.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Abramowitz, M., 1965. Handbook of mathematical functions with formulas, graphs, and mathematical tables.
Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F., Abdullah, H., Saidur, R., 2014. A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renew. Sustain. Energy Rev. 33, 102–109.
Amarasinghe, K., Marino, D.L., Manic, M., 2017. Deep neural networks for energy load forecasting. In: 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE). pp. 1483–1488.
Amasyali, K., El-Gohary, N.M., 2018. A review of data-driven building energy consumption prediction studies. Renew. Sustain. Energy Rev. 81, 1192–1205.
Amral, N., Özveren, C.S., King, D., 2007. Short term load forecasting using multiple linear regression. In: Proceedings of the Universities Power Engineering Conference, pp. 1192–1198.
Aoki, M., 1965. On some convergence questions in bayesian optimization problems. IEEE Trans. Automat. Control 10 (2), 180–182.
Arizona State University, 2021. ASU Campus Metabolism. Arizona State University. [Online]. Available: https://cm.asu.edu/ [Accessed: 25-Nov-2021].
Bedi, J., Toshniwal, D., 2018. Empirical mode decomposition based deep learning for electricity demand forecasting. IEEE Access 6, 49144–49156.
Bedi, J., Toshniwal, D., 2019. Deep learning framework to forecast electricity demand. Appl. Energy 238, 1312–1326.
Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, 281–305.
Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 2013. Time Series Analysis: Forecasting and Control, fourth ed. Wiley.
Box, G.E.P., Pierce, D.A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J. Amer. Statist. Assoc. 65 (332), 1509–1526.
Chen, H., Cañizares, C.A., Singh, A., 2001. ANN-based short-term load forecasting in electricity markets. In: Proc. IEEE Power Eng. Soc. Transm. Distrib. Conf. Vol. 2, no. WINTER MEETING, pp. 411–415.
Chen, Y.H., Hong, W.C., Shen, W., Huang, N.N., 2016. Electric load forecasting based on a least squares support vector machine with fuzzy time series and global harmony search algorithm. Energies 9 (2), 1–13.
Chen, Y., et al., 2017. Short-term electrical load forecasting using the support vector regression (SVR) model to calculate the demand response baseline for office buildings. Appl. Energy 195, 659–670.
Christiaanse, W.R., 1971. Short-term load forecasting using general exponential smoothing. IEEE Trans. Power Appar. Syst. PAS-90 (2), 900–911.
Colominas, M.A., Schlotthauer, G., Torres, M.E., 2014a. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. Biomed. Signal Process. Control 14, 19–29.
Colominas, M.A., Schlotthauer, G., Torres, M.E., Flandrin, P., 2014b. Noise-assisted EMD methods in action. Adv. Adapt. Data Anal. 04 (04), 1250025.
Crawley, D.B., et al., 2001. Energyplus: creating a new-generation building energy simulation program. Energy Build. 33 (4), 319–331.
Deb, C., Zhang, F., Yang, J., Lee, S.E., Shah, K.W., 2017. A review on time series forecasting techniques for building energy consumption. Renew. Sustain. Energy Rev. 74, 902–924.
Dragomiretskiy, K., Zosso, D., 2014. Variational mode decomposition. IEEE Trans. Signal Process. 62 (3), 531–544.
Duan, J., Zuo, H., Bai, Y., Duan, J., Chang, M., Chen, B., 2021. Short-term wind speed forecasting using recurrent neural networks with error correction. Energy 217, 119397.
Eskin, N., Türkmen, H., 2008. Analysis of annual heating and cooling energy requirements for office buildings in different climates in Turkey. Energy Build. 40 (5), 763–773.
Fan, C., Ding, Y., Liao, Y., 2019a. Analysis of hourly cooling load prediction accuracy with data-mining approaches on different training time scales. Sustain. Cities Soc. 51.
Fan, C., Wang, J., Gang, W., Li, S., 2019b. Assessment of deep recurrent neural network-based strategies for short-term building energy predictions. Appl. Energy 236, 700–710.
Friedrich, L., Afshari, A., 2015. Short-term forecasting of the abu dhabi electricity load using multiple weather variables. Energy Procedia 75, 3014–3026.
Fumo, N., Rafe Biswas, M.A., 2015. Regression analysis for prediction of residential energy consumption. Renew. Sustain. Energy Rev. 47, 332–343.
Guo, Z., Zhou, K., Zhang, X., Yang, S., 2018. A deep learning model for short-term power load and probability density forecasting. Energy 160, 1186–1200.
Han, L., Peng, Y., Li, Y., Yong, B., Zhou, Q., Shu, L., 2019. Enhanced deep networks for short-term and medium-term load forecasting. IEEE Access 7, 4045–4055.
He, W., 2017. Load forecasting via deep neural networks. Procedia Comput. Sci. 122, 308–314.
He, F., Zhou, J., Feng, Z., Liu, G., Yang, Y., 2019. 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 237, 103–116.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Holt, C.C., 2004. Forecasting seasonals and trends by exponentially weighted moving averages. J. Econ. Soc. Meas. 29 (1–3), 123–125.
Huang, N.E., et al., 1998. The empirical mode decomposition and the hubert spectrum for nonlinear and nonstationary time series analysis. Proc. R. Soc. A Math. Phys. Eng. Sci. 454 (1971), 903–995.
ISO, B., 2017. 52016-1: 2017 Energy Performance of Buildings–Energy Needs for Heating and Cooling, Internal Temperatures and Sensible and Latent Heat Loads–Part 1: Calculation Procedures. BSI (British Stand. Institution).
Jin, X.-B., et al., 2021. Deep-learning forecasting method for electric power load via attention-based encoder–decoder with bayesian optimization. Energies 14 (6), 1596.
Kennedy, J., Eberhart, R., 2006. Particle swarm optimization. In: Proceedings of ICNN'95 - International Conference on Neural Networks, Vol. 4, pp. 1942–1948.
Kingma, D.P., Ba, J.L., 2015. Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings.
Kong, W., Dong, Z.Y., Jia, Y., Hill, D.J., Xu, Y., Zhang, Y., 2019. Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Trans. Smart Grid 10 (1), 841–851.
Koprinska, I., Rana, M., Agelidis, V.G., 2015. Correlation and instance based feature selection for electricity load forecasting. Knowl.-Based Syst. 82, 29–40.
Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root? J. Econom. 54 (1–3), 159–178.
Lam, J.C., Wan, K.K.W., Tsang, C.L., Yang, L., 2008. Building energy efficiency in different climates. Energy Convers. Manag. 49 (8), 2354–2366.
Li, T., Qian, Z., He, T., 2020a. Short-term load forecasting with improved CEEMDAN and GWO-based multiple kernel ELM. Complexity 2020.
Li, W., Shi, Q., Sibtain, M., Li, D., Mbanze, D.E., 2020b. A hybrid forecasting model for short-term power load based on sample entropy, two-phase decomposition and whale algorithm optimized support vector regression. IEEE Access 8, 166907–166921.
Li, C., Zhu, Z., 2018. Research and application of a novel hybrid air quality early-warning system: A case study in China. Sci. Total Environ. 626, 1421–1438.
Li, J., et al., 2021. A novel hybrid short-term load forecasting method of smart grid using MLR and LSTM neural network. IEEE Trans. Ind. Inform. 17 (4), 2443–2452.
Liang, Y., Niu, D., Hong, W.C., 2019. Short term load forecasting based on feature extraction and improved general regression neural network model. Energy 166, 653–663.
Liao, Z., Gai, N., Stansby, P., Li, G., 2020. Linear non-causal optimal control of an attenuator type wave energy converter M4. IEEE Trans. Sustain. Energy 11 (3), 1278–1286.
Liu, B., Fu, C., Bielefield, A., Liu, Y.Q., 2017. Forecasting of Chinese primary energy consumption in 2021 with GRU artificial neural network. Energies 10 (10).
Liu, Jin, Gao, ., 2019. A new hybrid approach for short-term electric load forecasting applying support vector machine with ensemble empirical mode decomposition and whale optimization. Energies 12 (8), 1520.
Liu, Z., Wang, X., Zhang, Q., Huang, C., 2019. Empirical mode decomposition based hybrid ensemble model for electrical energy consumption forecasting of the cement grinding process. Meas. J. Int. Meas. Confed. 138, 314–324.
Liu, Y., Yang, C., Huang, K., Gui, W., 2014. Non-ferrous metals price forecasting based on variational mode decomposition and LSTM network. Knowl.-Based Syst. 188, 105006.
Lusis, P., Khalilpour, K.R., Andrew, L., Liebman, A., 2017. Short-term residential load forecasting: Impact of calendar effects and forecast granularity. Appl. Energy 205, 654–669.
M. of E. of the P.R. of China, 2020. 2019 national education development statistical bulletin.
Matérn, B., 1960. Spatial variation. Rep. For. Res. Inst. Sweden 49 (5).

Mohamed, Z., Bodger, P., 2005. Forecasting electricity consumption in New Zealand using economic and demographic variables. Energy 30 (10), 1833–1843.
Munem, M., Rubaith Bashar, T.M., Roni, M.H., Shahriar, M., Shawkat, T.B., Rahaman, H., 2020. Electric power load forecasting based on multivariate LSTM neural network using Bayesian optimization. In: 2020 IEEE Electric Power and Energy Conference (EPEC). pp. 1–6.
Myers, J.L., Well, A.D., Lorch, R.F., 2013. Research Design and Statistical Analysis: Third Edition, Res. Des. Stat. Anal, third ed. pp. 1–809.
Nationalgrid, 2012. Managing energy costs in colleges and universities. [Online]. Available: https://www9.nationalgridus.com/non_html/shared_energyeff_college.pdf [Accessed: 03-Aug-2021].
Nie, Y., Jiang, P., Zhang, H., 2020. A novel hybrid model based on combined preprocessing method and advanced optimization algorithm for power load forecasting. Appl. Soft Comput. J. 97.
Niu, H., Xu, K., Liu, C., 2021. A decomposition-ensemble model with regrouping method and attention-based gated recurrent unit network for energy price prediction. Energy 231, 120941.
Paliwal, K., Basu, A., 1987. A speech enhancement method based on Kalman filtering. In: ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing. 12, pp. 177–180.
Pham, A.D., Ngo, N.T., Ha Truong, T.T., Huynh, N.T., Truong, N.S., 2020. Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability. J. Clean. Prod. 260.
Rahman, A., Srikumar, V., Smith, A.D., 2018. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Appl. Energy 212, 372–385.
Ružić, S., Vučković, A., Nikolić, N., 2003. Weather sensitive method for short term load forecasting in electric power utility of Serbia. IEEE Trans. Power Syst. 18 (4), 1581–1586.
Said, S.E., Dicky, D.A., 1984. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71 (3), 599–607.
Sen, P., Roy, M., Pal, P., 2016. Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization. Energy 116, 1031–1038.
Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., De Freitas, N., 2016. Taking the human out of the loop: A review of Bayesian optimization. Proc. IEEE 104 (1)), 148–175, Institute of Electrical and Electronics Engineers Inc..
Shang, Q., Feng, L., Gao, S., 2021. A hybrid method for traffic incident detection using random forest-recursive feature elimination and long short-term memory network with Bayesian optimization algorithm. IEEE Access 9, 1219–1232.
Shang, Z., He, Z., Song, Y., Yang, Y., Li, L., Chen, Y., 2020. A novel combined model for short-term electric load forecasting based on whale optimization algorithm. Neural Process. Lett. 52 (2), 1207–1232.
Sibtain, M., Li, X., Saleem, S., 2020. A multivariate and multistage medium- and long-term streamflow prediction based on an ensemble of signal decomposition techniques with a deep learning network. Adv. Meteorol. 2020.
Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical Bayesian optimization of machine learning algorithms. Adv. Neural Inf. Process. Syst. 4, 2951–2959.
Somu, N., Mr, G.R., Ramamritham, K., 2020. A hybrid model for building energy consumption forecasting using long short term memory networks. Appl. Energy 261.
Standard, E.N., et al., 2008. Energy performance of buildings, calculation of energy use for space heating and cooling.
Strachan, P.A., Kokogiannakis, G., Macdonald, I.A., 2008. History and development of validation with the ESP-r simulation program. Build. Environ. 43 (4), 601–609.
Sun, Z., Zhao, M., 2020. Short-term wind power forecasting based on VMD decomposition, convlstm networks and error analysis. IEEE Access 8, 134422–134434.
Terreson, D., et al., 2020. Global energy outlook 2020: energy transition or energy addition? Resour. Future 1–6.
the National Center for Education Statistics of the U.S., 2020. Digest of education statistics. [Online]. Available: https://nces.ed.gov/programs/digest/2019menu_tables.asp.
Thevenard, D., Haddad, K., 2006. Ground reflectivity in the context of building energy simulation. Energy Build. 38 (8), 972–980.
Torres, M.E., Colominas, M.A., Schlotthauer, G., Flandrin, P., 2011. A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 4144–4147.
Ugurlu, U., Oksuz, I., Tas, O., 2018. Electricity price forecasting using recurrent neural networks. Energies 11 (5).
Wang, J., Chen, X., Zhang, F., Chen, F., Xin, Y., 2021. Building load forecasting using deep neural network with efficient feature fusion. J. Mod. Power Syst. Clean Energy 9 (1), 160–169.
Wang, J.Q., Du, Y., Wang, J., 2020a. Lstm based long-term energy consumption prediction with periodicity. Energy 197, 117197.
Wang, W., Tong, M., Yu, M., 2020b. Blood glucose prediction with VMD and LSTM optimized by improved particle swarm optimization. IEEE Access 8, 217908–217916.
Wang, Z., Wang, Y., Zeng, R., Srinivasan, R.S., Ahrentzen, S., 2018. Random forest based hourly building energy prediction. Energy Build. 171, 11–25.
Williams, C.K., Rasmussen, C.E., 2006. Gaussian Processes for Machine Learning, Vol. 2 (3). MIT Press, Cambridge.
Wu, Z., Huang, N.E., 2014. Ensemble empirical mode decomposition: A noise-assisted data analysis method. Adv. Adapt. Data Anal. 01 (01), 1–41.
Xia, M., Shao, H., Ma, X., De silva, C.W., 2021. A stacked GRU-rnn-based approach for predicting renewable energy and electricity load for smart grid operation. IEEE Trans. Ind. Inform..
Yang, F., Li, W., Li, C., Miao, Q., 2019. State-of-charge estimation of lithium-ion batteries based on gated recurrent neural network. Energy 175, 66–75.
Yi, S.L., et al., 2017. Online denoising based on the second-order adaptive statistics model. Sensors (Switzerland) 17 (7).
York, D.A., Cappiello, C.C., Olson, K.H., L.A.N.L. and L.B.L.A.S., 1984. DOE-2 Reference Manual: Version 2.1C. Los Alamos National Laboratory, Solar Energy Group.
Yu, Z., Haghighat, F., Fung, B.C.M., Yoshino, H., 2010. A decision tree method for building energy demand modeling. Energy Build. 42 (10), 1637–1646.
Yu, Z., Moirangthem, D.S., Lee, M., 2017. Continuous timescale long-short term memory neural network for human intent understanding. Front. Neurorobot. 11 (AUG).
Zegers, J., Van Hamme, H., 2019. Cnn-LSTM models for multi-speaker source separation using Bayesian hyper parameter optimization. In: Interspeech 2019, Vol. 2019-Septe. pp. 4589–4593.
Zhang, L., Wang, J., Niu, X., Liu, Z., 2021. Ensemble wind speed forecasting with multi-objective archimedes optimization algorithm and sub-model selection. Appl. Energy 301 (April), 117449.
Zhang, J., Wei, Y.-M., Li, D., Tan, Z., Zhou, J., 2018. Short term electricity load forecasting using a hybrid model. Energy 158, 774–781.
Zhang, K., Zheng, L., Liu, Z., Jia, N., 2020. A deep learning based multitask model for network-wide traffic speed prediction. Neurocomputing 396, 438–450.
Zhao, X., Lv, H., Lv, S., Sang, Y., Wei, Y., Zhu, X., 2021. Enhancing robustness of monthly streamflow forecasting model using gated recurrent unit based on improved grey wolf optimizer. J. Hydrol. 601 (2020), 126607.
Zhao, H., Magoulès, F., 2012. A review on the prediction of building energy consumption. Renew. Sustain. Energy Rev. 16 (6), 3586–3592.
Zheng, H., Yuan, J., Chen, L., 2017. Short-term load forecasting using EMD-LSTM neural networks with a xgboost algorithm for feature importance evaluation. Energies 10 (8).
Zhou, M., et al., 2021. Short-term electric load forecasting based on variational mode decomposition and grey wolf optimization. Energies 14 (16), 4890.
Zou, Z.D., Sun, Y.M., Zhang, Z.S., 2005. Short-term load forecasting based on recurrent neural network using ant colony optimization algorithm. Power Syst. Technol. 29 (3), 59–63.