Energy Proceedings
Vol 27, 2022
ISSN 2004-2965

A Hybrid Method of Hourly Electricity Consumption Forecasting for Building Cluster Based on PSO-RF

Xiaolin Chu[1,2], Peng Wang[3], Ruijuan Zhao [4*], Dayong Lv[1]

1 School of Financial Technology, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China

2 Shanghai Financial Technology Research Centre, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China

3 Science and Technology on Electromagnetic Scattering Laboratory, Shanghai 200438, China

4 School of Economics and Management, Shanghai University of Political Science and Law, Shanghai 201701, China (Corresponding Author)

ABSTRACT

Building energy consumption prediction is of great significance to realize intelligent decision-making of energy system and improve energy efficiency. A random forest (RF) prediction model optimized via the particle swarm optimization (PSO) algorithm is established to forecast the hourly electricity consumption of the building cluster consisting of interconnected multiple buildings. The accuracy, generalization and robustness are taken as evaluation indexes. In the case study, the building cluster located in Austin is adopted as an example to explore the predicted performance of the proposed PSO-RF model in different seasons. The results show that the hourly electricity consumption PSO-RF model of the building cluster can achieve highest accuracy, strongest generalization, and best robustness, compared with RF, decision tree (DT), XGBoost, and k-Nearest Neighbor (KNN) prediction models. Therefore, the proposed hybrid model can be used as a reliable tool for building cluster electricity consumption prediction and energy management.

Keywords: building cluster, electricity consumption, forecasting method, random forest, particle swarm optimization algorithm

1. INTRODUCTION

Building energy consumption is an important component of total world energy consumption, which accounts for about 33% [1]. With the growing increase in population, the expansion of buildings, and economic growth, building energy consumption is predicted to increase at an average rate of 1.5% per year during the period to 2040 [2]. Therefore, reducing building energy consumption is vital for energy conservation and environmental protection. As the building energy demand growing, reliable and robust energy consumption forecasting is beneficial for planning scientific energy, reducing carbon emissions, and improving energy efficiency [3]. As an important part of building energy planning and management, building energy prediction has received widely attention from academia and industry.

The research about building electricity consumption prediction mainly focuses on physical models and data-driven models [4]. Physical models rely on simulation tools such as EnergyPlus [5], DEST [6], and eQuest [7] to calculate building energy consumption, which are very sensitive to boundary conditions and strongly depend on expertise. Data-driven models are based on training historical data through machine learning algorithms to discover patterns and learn to a prediction model. Compared with physical models, data-driven models are mostly selected to predicting building energy consumption due to short computation time and good performance under sufficient historical data.

Most of the existing data-driven studies focus on exploring the energy consumption prediction problem of individual building by usual machine learning algorithms, such as random forest (RF), XGBoost, and decision tree (DT) [8]. Prediction models of different buildings (e.g., commercial and residential building) are established [9,10] considering the prediction time scale [11], prediction types [12,13], and evaluation indexes [14]. Katsatos et al. [15] use Artificial Neural Networks (ANNs) to forecast the energy consumption in the building of Regulation Authority of Energy in Athens city. And then, the experimental results show that ANNs models present a remarkable prognostic ability to predict the energy consumption. Feng et al. [16] propose a support vector machine (SVM) prediction model forecasting the short-term load of the microgrid in an offshore oil field, and analyze the prediction accuracy under Mean Absolute Percentage Error (MAPE). Zhou et al. [17] present long short term memory integrated with reinforcement learning agents to forecast building next-day electricity consumption, and the strategy can effectively improve the prediction accuracy verified by the experiments. Wang et al. [18] build a monthly RF prediction model of building energy consumption using Coefficient of Determination (R^2), MAPE, and Mean Squared Error (RMSE) as the evaluation indicators. Two academic buildings are taken as examples, and the RF model is verified with best perform in predicting energy consumption compared with regression tree and SVM models.

The building electricity consumption is affected by various parameters (e.g., weather condition) and perturbed by random events (e.g., sudden climate change, monitoring system failure and meter damage), Therefore, the robustness of the prediction model should to be considered. Robustness refers to the ability for a prediction model to resist disturbances in the external environment. Researchers measure the established electricity consumption prediction model for individual building from multiple evaluation perspectives. Zhong et al. [19] and Chu et al. [20] use the support vector regression (SVR) model with high accuracy and generalization as well as robustness for building energy consumption prediction and building carbon emission prediction. Wang et al. [21] propose predicting building energy consumption based on stacking model, and validate it in terms of accuracy, generalization ability, and robustness.

However, in current literature, few research efforts are dedicated to forecasting the building cluster electricity consumption. With the promotion of smart grids, collaborative buildings regarded as a building cluster can maximize the use of on-site energy and increase the flexibility of energy storage systems. Therefore, it is important to move the research boundary from the individual building level to the building cluster level [22]. Walker et al. [23] propose an hourly total energy consumption prediction model for the commercial building cluster based on machine learning algorithms such as RF and ANN, and the accuracy of including R^2, MAPE, and CV-RMSE as the evaluation criteria. The experimental results show that the RF prediction model has high accuracy. Furthermore, RF model can be enhanced by integrating intelligence algorithms and investigated comprehensively with multi evaluation indexes.

To fill this research gap, in this study, the hourly energy consumption prediction model of building cluster with high accuracy, generalization ability, and robustness is propose. The parameters of the RF method are optimized by particle Swarm Optimization (PSO) algorithm, and PSO-RF prediction model is built to forecast the hourly electricity consumption of building cluster. In the case study, the PSO-RF model of building cluster in Austin is investigated to evaluate the accuracy and generalization as well as robustness based on R^2, RMSE, Mean Absolute Error (MAE), compared with several usual machine learning algorithms (i.e., RF, XGBoost, DT, kNN). The experimental results show that the performance of PSO-RF prediction model is superior to other four models in summer, winter, and transition seasons. The building cluster hourly electricity consumption prediction model is beneficial to optimize building cluster energy supply system, which in turn enhance the intelligent decision-making process and improve energy utilization efficiency.

This paper is organized as follows. Section 2 gives the problem description. In Section 3, a PSO-RF prediction model is proposed to forecast building cluster hourly electricity consumption. The numerical results from the case studies for the proposed PSO-RF prediction model compared with four machine learning models under multi evaluation indexes in different seasons are reported in Section 4. Finally, Section 5 provides the conclusions and offers future research directions.

2. PROBLEM DESCRIPTION

The purpose of this paper is to establish a building cluster hourly electricity consumption prediction model and evaluate it with accuracy, generalization capability and robustness. Nine influencing factors are identified by analyzing relationship with building cluster energy consumption, which refers to time series and meteorological parameters (i.e., day, week, hour (h), temperature (℉), humidity (%), wind speed (mph), pressure (in), dew point (℉), and precipitation (in)). At the same time, the building cluster electricity consumption has seasonal characteristics. The PSO-RF method is proposed to forecast building cluster power energy consumption, and is investigated in different seasons with multi perspectives of accuracy and generalization as well as robustness.

3. PSO-RF MODEL FOR BUILDING CLUSTER HOURLY ELECTRICITY CONSUMPTION

3.1 RF algorithm

As a typical supervised machine learning method, Ensemble Learning (EL) algorithm accomplishes the learning task by building and combining multiple learners. RF algorithm is an extended variant of the Bagging regarded as well-known parallel EL algorithm [24]. The RF algorithm adopts Bootstrap resampling technology to randomly draw samples from the training sample set in T times with put-back. k random features are selected independently at each sample set, and T weak learners are trained based on the DT algorithm. In the regression task, T weak learners are usually integrated to a strong learner by averaging, and then prediction result can be obtained.

By stochastic selection of sample sets and attributes in week learner, the RF model has good generalization performance by alleviating the risk of overfitting [25,26]. In this paper, the weak learner in RF is trained using the CART algorithm (a DT algorithm), which uses the Gini index to select features and a top-down greedy algorithm to build a tree structure. The regression task is divided into nodes based on the principle of minimizing R^2, and the CART algorithm is suitable for solving the prediction problem of large sample size of building cluster electricity consumption in this paper. The basic steps of the RF model are given as follows.

(1) The training set $D=\{(\boldsymbol{x_1},y_1),(\boldsymbol{x_2},y_2),\dots,(\boldsymbol{x_m},y_m)\}$ with m samples described by p attributes (e.g, $\boldsymbol{x_1}=\{x_{11};x_{12};\dots;x_{1p}\}$).

(2) Using Bootstrap random sampling to form T training subsets to build weak learner $D_1,D_2,\dots,D_T$ with m samples.

(3) For each weak learner, k features randomly are selected from the original p features.

(4) For the training subset D_t in the t-th weak learner, T decision trees can be obtained by k features based on CART algorithm.

(5) The T results can be aggregated averagely to obtain derive the final prediction results from T weak learners.

Considering fast convergence and easy implementation of PSO algorithm, the parameters T,k are optimized by PSO algorithm for electricity consumption RF prediction model of building cluster in this paper.

3.2 PSO algorithm

The PSO algorithm is an intelligent optimization algorithm based on swarm intelligence. In the PSO algorithm, the candidate solutions of T,k in the RF model are represented by particles. D, I and S represent the particle dimension, the number of iterations and the number of particles, respectively. (vp_{sd}^i,lp_{sd}^i) denotes the spatial position and flight velocity of the s-th particle under the i-th iteration in the d-dimensional search space. In the PSO algorithm, the particle's velocity and position are updated by Eqs. (1) and (2), respectively.

$$vp_{sd}^{i+1}=w\cdot vp_{sd}^i+c_1\cdot rand_1\cdot\left(pbest_{sd}^i-lp_{sd}^i\right)+c_2\cdot rand_2\cdot\left(gbest_d^i-lp_{sd}^i\right)\qquad\forall s,\forall d,\forall i\qquad(1)$$

$$lp_{sd}^{i+1}=lp_{sd}^i+vp_{sd}^{i+1}\qquad\forall s,\forall d,\forall i\qquad(2)$$

w is the inertia weight, and c_1 and c_2 denote the non-negative learning factors. $rand_1$ and $rand_2$ are random numbers varying in the range of 0 to 1. In the i-th iteration, $pbest_{sd}^i$ and $gbest_d^i$ denote the individual extreme point of the s-th particle in the d-th dimension and the population extreme point of the population in the d-th dimension, respectively.

3.3 PSO-RF model

In the PSO-RF model, the parameters T and k of the CART decision tree in weak learners are optimized by the PSO algorithm, and then the prediction results are obtained by a random forest composed of T weak learners. The specific algorithm process is shown in Fig. 1. The sample set of building cluster hourly electricity consumption is divided into training subset and testing subset after data pre-processing. The parameters T,k and particles' velocity as well as position are initialized in the PSO-RF model. Based on the training subset, T weak learners are trained using the CART decision tree algorithm with k features, and corresponding T predicted results are integrated with simple averaging to obtain the predictive results. R^2 in the PSO algorithm is used as the fitness function and the number of iterations is the termination condition. When the termination condition is not reached, the particles' velocity and position in the PSO algorithm are updated and the fitness function R^2 is recalculated. At the same time, the number of iterations is added one, and the next cycle begins. When the termination condition is reached, the T',k' representing optimal T,k can be obtained by PSO algorithm in the PSO-RF model. Furthermore, the prediction results can be got under the training subset. The PSO-RF prediction model of building cluster hourly electricity consumption is learned. The prediction results can be obtained under the testing subset.

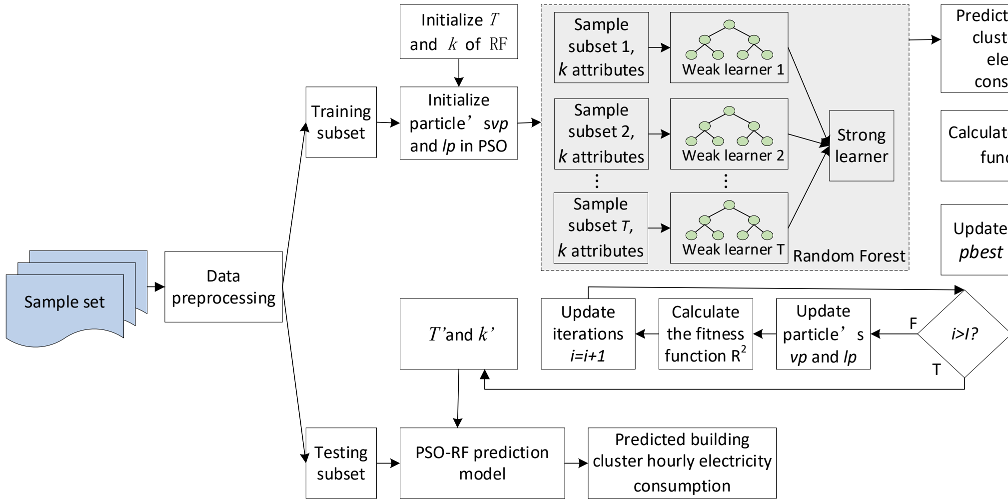


Fig.1 The flow chart of hourly electricity consumption prediction model for the building cluster based on PSO-RF algorithm

3.4 Assessment criteria

R^2, RMSE, and MAE are adopted as evaluation index es in this study. The mean value is taken as the error benchmark in R^2 limited between 0 and 1, which reflects the level of error between the forecast value and the mean benchmark. RMSE mainly reflects the dispersion of the error between the forecast and the actual value. In addition, MAE shows on the mean value of the absolute error between the forecast and the actual value. R^2, RMSE, and MAE can be calculated as Eqs.(3)-(5), respectively.

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(\hat{y}_n - y_n)^2}{\sum_{n=1}^{N}(\bar{y} - y_n)^2} \tag{3}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{y}_n - y_n)^2} \tag{4}$$

$$MAE = \frac{1}{N}\sum_{n=1}^{N}|\hat{y}_n - y_n| \tag{5}$$

N is the total number of samples in sample set. $\hat{y}_n$ and y_n denote the predicted value and actual value of the n-th sample, respectively. $\bar{y}$ is the average of the actual values of all samples. Overall, a larger R^2, a smaller RMSE, and a smaller MAE indicate a better fitting model.

Accuracy, generalization, and robustness are used to evaluate the PSO-RF prediction model of building cluster hourly electricity consumption. Based on the training subset, the PSO-RF prediction model is trained and can be used to predict building cluster hourly electricity consumption. Accuracy can be got under training subset with R^2, RMSE, and MAE. Generalization capability refers to the prediction capability of the model for samples beyond the training range, which can be obtained by testing subset with metrics (i.e., R^2, RMSE, and MAE). In addition, the problem of predicting the building cluster electricity consumption suffers from the phenomenon that the collected data may deviate from the actual data. Testing subset is modified by adding Gaussian white noise to evaluate the prediction model's robustness [27]. R^2 reflects the proportion of the total variation in the dependent variable that can be explained by the independent variable [21]. Therefore, R^2 is selected to evaluate the robustness of the PSO-RF prediction model in this study.

4. CASE STUDY

4.1 Data acquisition

To verify the accuracy, generalization, and robustness of the hourly electricity consumption PSO-RF prediction model of building cluster, cases are investigated for a building cluster consisting of 10 residential buildings located in Austin, Texas, USA. Hourly electricity consumption data in 2018 (i.e., 8760 hours) is collected and shown in Fig.2 [28]. The data of temperature (℉), humidity (%), wind speed (mph), pressure (in), dew point (℉), and precipitation (in) are obtained from relevant meteorological websites [29].

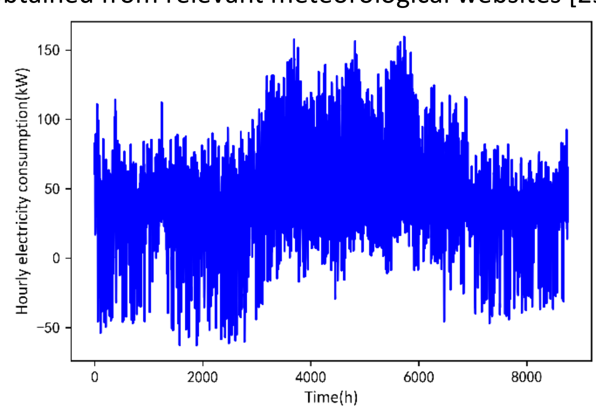


Fig.2 Hourly electricity consumption of the building cluster in Austin, 2018

4.2 Prediction and analysis

To compare the effectiveness of different regression models, four models (i.e., RF, XGBoost, DT, and kNN) are applied to forecast hourly electricity consumption of building cluster. Regression models are programmed with Python programming language on Anaconda platform on a computer equipped with Intel(R) Core (TM) i5-8265U CPU @1.60GHz processor and 8GB memory. Nice influencing factors are used as input feature variables for the prediction models, and the hourly electricity consumption of the building cluster is adopted as the label. In the PSO-RF model, c_1 and c_2 are usually assumed as $c_1 = c_2 = 2$, and I is set as $I = 100$. The number of decision trees T ranges from 0 to 300 [30], and random features k ranges from 1 to 9.

Considering seasonality and cyclicity, the accuracy, generalization, and robustness of the hourly electricity consumption PSO-RF prediction model for building cluster are tested in summer and winter, and transition seasons, respectively. The sample sizes of the training and testing subsets in different seasons, and the evaluation indexes are shown in Table 1.

Table 1 Data sets and evaluation indexes of hourly electricity consumption prediction model for the building cluster in different seasons

		Summer	Winter	Transition season	Assessment criteria
Training subset	Time	Apr,May	Oct,Nov	Jan, Feb, Mar, Jul, Aug	Accuracy
	Sample size	1464	1464	3648	
Testing subset	Time	Jun	Dec	Sep	Generalization, robustness
	Sample size	720	744	720	

4.2.1 Accuracy

Under different seasons, five different prediction models (PSO-RF, RF, XGBoost, DT, and kNN) are trained based on training subset to forecast the building cluster hourly electricity consumption. In the summer, the training subset consists of 1464 samples from April to May. In the winter, the training subset consists of 1464 samples from October to November. In the transition season, the training subset consists of 3648 samples in January-March and July-August. For the PSO-RF prediction model in different seasons, the T' and k' (optimal parameters) are obtained by the PSO algorithm and are given in Table 2.

Table 2 The T' and k' of the PSO-RF models in different seasons

Parameter	Summer		Winter		Transition season	
	T'	k'	T'	k'	T'	k'
Value	256	6	189	5	87	3

In summer, the predicted and actual building cluster hourly electricity consumption under training subset in the five prediction models are displayed in Fig.3. The dashed line denotes that predicted and actual building cluster hourly electricity consumption are equal. Compared with RF, XGBoost, DT, and KNN prediction models, the R^2 of the PSO-RF model can be improved by 2.06% to 12.5% in predicting building cluster hourly energy consumption in summer. Therefore, the predicted results in PSO-RF model are more closely matched to the actual values under training subset in summer.

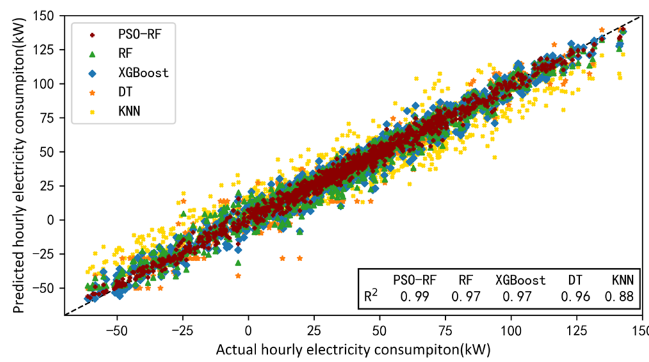


Fig.3 The predicted and actual hourly electricity consumption of the building cluster under training subset in summer

In winter, the predicted and actual building cluster hourly electricity consumption under training subset in the five prediction models are shown in Fig. 4. The dashed line indicates predicted values equaling with actual values. In comparison, the R^2 of the PSO-RF prediction model for forecasting hourly electricity consumption of the building cluster is improved by 3.12% to 15.12% in winter. Combining the scatter distribution and R^2 values in Fig. 4, it is found that the PSO-RF model has the best prediction effect, followed by the RF model, XGBoost model, DT model and kNN models.

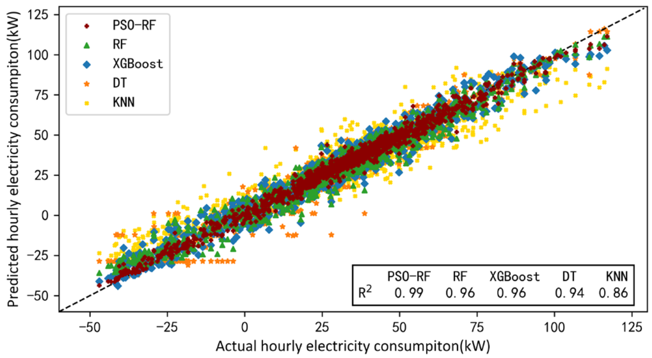
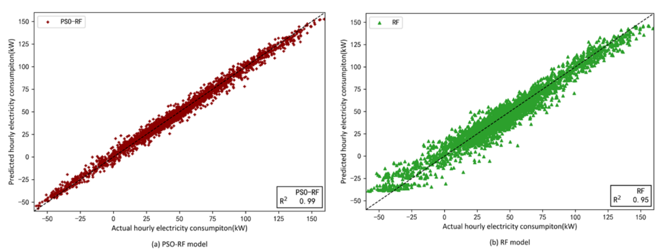


Fig.4 The predicted and actual hourly electricity consumption of the building cluster under training subset in winter

In transition season, the predicted and actual building cluster hourly electricity consumption under training subset in the five prediction models are displayed in Fig. 5(a) to (e), respectively. The dashed lines in the subplots represent the predicted values equaling with actual values. The transition season involves January-March and July-August. From Fig. 5, it can be concluded that the PSO-RF prediction model has the highest R^2 with 0.99. Compared with other four models, the R^2 of the PSO-RF model can be improved by 4.21% to 11.24% in the transition season. Figs 4,5 and 6 indicate that the PSO-RF prediction model has higher R^2 than other four prediction models for predicting building cluster hourly electricity consumption under training subsets in different seasons.



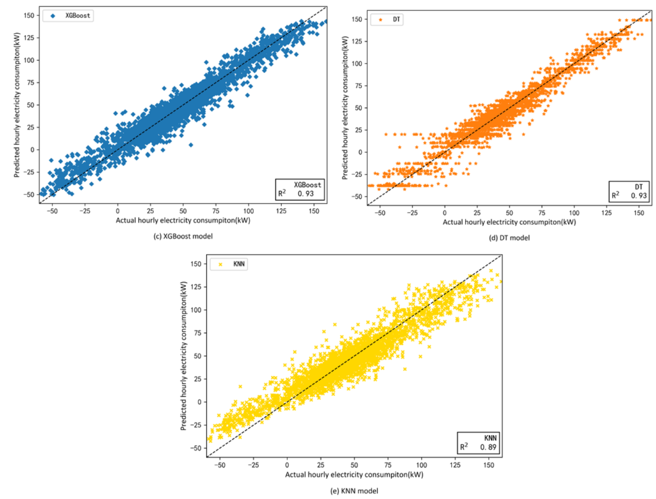


Fig.5 The predicted and actual hourly electricity consumption of the building cluster under training subset in transition season

The accuracy of the five models for predicting the hourly electricity consumption of building cluster in different seasons are shown in Table 3. The PSO-RF prediction model has the largest R^2 and the smallest RMSE and MAE values, which indicates the highest accuracy under training subsets. Furthermore, compared to the RF model, the PSO-RF model can achieve R^2 improved by 2.06% to 4.21%, RMSE reduced by 39.39% to 53.13%, and MAE reduced by 38.41% to 52.41% in different seasons. Compared with KNN model, R^2 is improved by 11.24% to 15.12%, and RMSE is reduced by 67.76% to 70.50% with MAE reduced by 68.33% to 71.16% in different seasons. Therefore, under training subsets, the PSO-RF model for forecasting the building cluster hourly electricity consumption can achieve higher accuracy than all the other four models.

Table 3 Comparison of accuracy indexes of different models of the building cluster hourly electricity consumption in different seasons

Prediction model		PSO-RF	RF	XGBoost	DT	kNN
Summer	R^2	0.99	0.97	0.97	0.96	0.88
	RMSE(kW)	3.57	5.89	6.00	6.18	12.10
	MAE(kW)	2.63	4.27	4.48	3.69	9.12
Winter	R^2	0.99	0.96	0.96	0.94	0.86
	RMSE(kW)	3.16	5.26	5.38	6.02	9.80
	MAE(kW)	2.32	3.99	4.02	3.75	7.44
Transition season	R^2	0.99	0.95	0.93	0.93	0.89
	RMSE(kW)	3.75	8.00	8.71	8.83	11.72
	MAE(kW)	2.80	6.01	6.56	6.09	8.84

4.2.2 Generalization

Under difference seasons, the five different prediction models trained by training subsets are tested by testing subsets with evaluation index of generalization ability. The testing subset in summer and winter as well as transition season consist of 720 samples, 744 samples and 720 samples of the building cluster, respectively.

In summer, the predicted and actual hourly electricity consumption of the building cluster in the five prediction models are displayed in Fig. 6. The dashed line indicates predicted values equaling with actual values. Fig. 6 demonstrates that the PSO-RF prediction model has the highest R^2 (i.e., 0.95) followed by the RF model, XGBoost, kNN, and DT prediction models. The weakest generalization ability is the DT model with an R^2 value of 0.74.

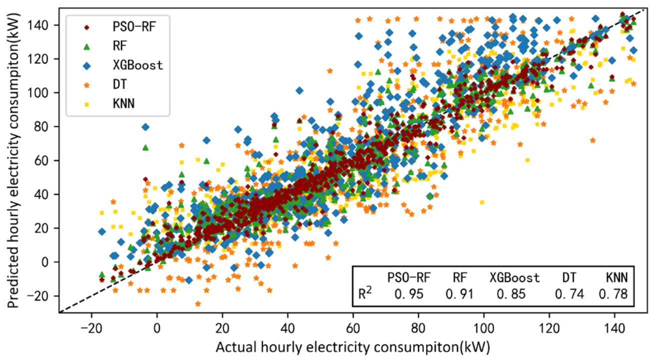


Fig.6 The predicted and actual hourly electricity consumption of the building cluster under testing subset in summer

In winter, the predicted and actual building cluster hourly electricity consumption under testing subsets in the five prediction models are shown in Fig. 7. The horizontal and vertical coordinates indicate the actual and the corresponding predicted the building cluster hourly electricity consumption, respectively. The PSO-RF prediction model with highest R^2 (i.e., 0.90) is suitable for forecasting building cluster hourly electricity consumption in winter.

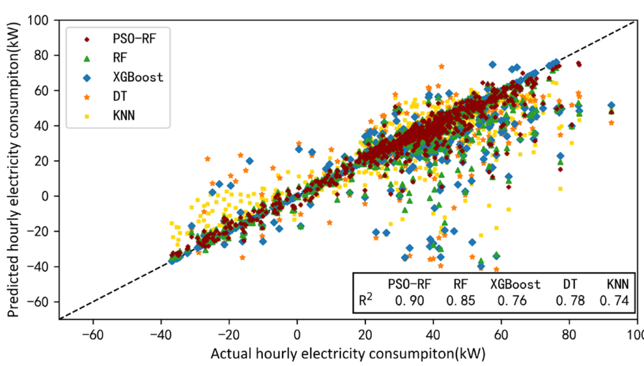


Fig.7 The predicted and actual hourly electricity consumption of the building cluster under testing subset in winter

In transition season, the predicted and actual hourly electricity consumption of the building cluster under testing subset are shown in Fig. 8. Fig.8 demonstrates that the PSO-RF model has the highest R^2 of 0.87. The worst performance is the DT prediction model with R^2 of 0.68. Figs 6,7 and 8 indicate that the PSO-RF prediction model has higher R^2 than other four prediction models under testing subsets in different seasons.

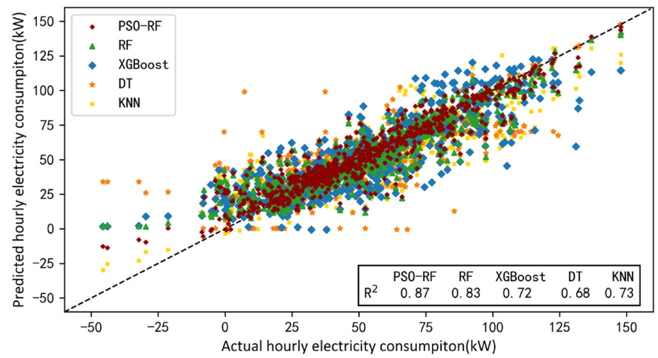


Fig.8 The predicted and actual hourly electricity consumption of the building cluster under testing subset in transition season

The generalization of the five models for forecasting the hourly electricity consumption of the building cluster in different seasons are shown in Table 4. According to largest R^2 and smallest RMSE as well as MAE values, the built PSO-RF prediction model has the strongest generalization ability compared with other four prediction models. Furthermore, the PSO-RF model compared with RF model can realize R^2 improved by 4.40%-5.88%, RMSE reduced by 15.14%-32.85% and MAE reduced by 19.85%-29.82% in different seasons. According to Tables 3 and 4, the proposed PSO-RF prediction model can achieve higher accuracy and generalization than RF, XGBoost, DT, and kNN prediction models for forecasting building cluster hourly electricity consumption in different seasons.

Table 4 Comparison of generalization indexes of the building cluster hourly electricity consumption prediction models in different seasons

Prediction model		PSO-RF	RF	XGBoost	DT	KNN
Summer	R^2	0.95	0.91	0.85	0.74	0.78
	RMSE(kW)	7.88	10.46	17.17	21.20	16.75
	MAE(kW)	4.93	6.84	12.01	15.33	12.99
Winter	R^2	0.90	0.85	0.76	0.78	0.74
	RMSE(kW)	7.81	11.63	13.61	13.31	13.94
	MAE(kW)	4.19	5.97	5.36	5.59	9.33
Transition season	R^2	0.87	0.83	0.72	0.68	0.73

RMSE(kW)	10.93	12.88	16.17	17.19	16.00
MAE(kW)	7.39	9.22	12.25	11.89	12.24

4.2.3 Robustness

Different levels of Gaussian white noise are added to the testing subsets in the summer, winter, and transition seasons to measure the robustness of the prediction models. The noise intensity is in the range of [0,1] and is increased by step size 0.1. With the changing noise intensity, the R^2 are recalculated in different prediction models and shown in Fig. 9. The R^2 is decreasing as the noise intensity increasing, which reflects the decreasing prediction effect. When the noise intensity reaches 100% in the testing subsets, the R^2 values of PSO-RF models decrease to 0.80, 0.57, and 0.71 in summer, winter, and transition seasons, respectively. At the same time, the R^2 of the DT models decrease to 0.43, 0.22, and 0.47 in summer, winter, and transition seasons, respectively. The line trend in Fig.9 indicate that the overall R^2 of the PSO-RF prediction model is higher other four prediction models expressing the stronger robustness. As shown in Tables 3, 4 and Fig. 9, the proposed PSO-RF prediction model can obtain more accuracy, generalization, and robustness than RF, XGBoost, DT, kNN prediction models in forecasting the building cluster hourly electricity consumption.

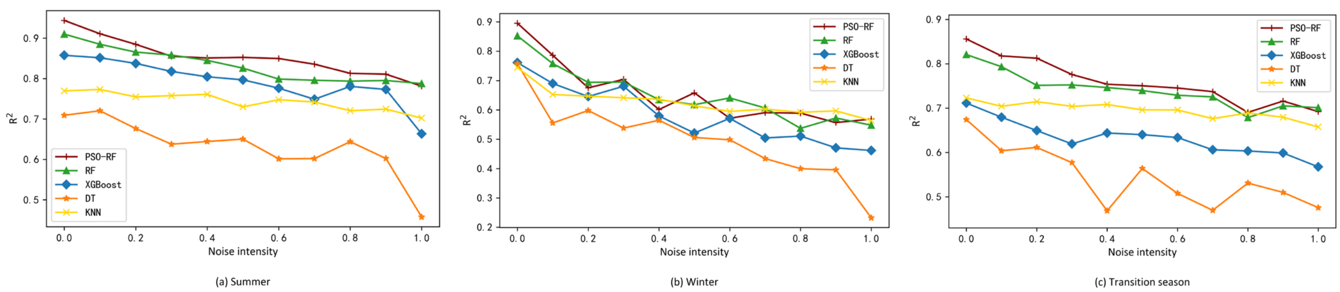


Fig.9 The robustness of the building cluster hourly electricity consumption prediction models with noise intensity in different seasons

5. CONCLUSTION

In this paper, a PSO-RF prediction model for hourly electricity consumption of building cluster is proposed considering 9 influencing factors. Two parameters of the RF model are optimized by PSO algorithm to enhance performance. The case study results show that the presented PSO-RF prediction model applied in Austin's a building cluster can achieve highest R^2, smallest RMSE and MAE compared with RF, XGBoost, DT, kNN prediction models with different evaluation indexes in summer, winter, and transition seasons. For example, the R^2 of the PSO-RF prediction model can be improved by 2.06% ~12.5% compared with RF, XGBoost, DT, and KNN prediction models in summer with the perspective of accuracy. Therefore, the PSO-RF prediction model can be used to forecast building cluster hourly electricity consumption with good performance. The prediction method has the potential to operate stably under different working conditions throughout the year, and can provide a way to optimize the building cluster energy supply system.

For future work, this research can be extended to consider building types. In addition, prediction and optimization can be researched jointly to enhance building cluster efficiency.

ACKNOWLEDGEMENT

We are grateful for the support from Shanghai Financial Technology Research Centre Project (2022-JK03-A), Financial Data Center Innovation Project of Shanghai Lixin University of Accounting and Finance, Young Scholar Training Project of Shanghai Universities, Teacher Profession Development Project of Shanghai, and Golden Course of Shanghai Lixin University of Accounting and Finance.

REFERENCE

[1] Cao X, Dai X, Liu J. Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade. Energy and Buildings 2016; 128: 198-213.
[2] U.S. Energy Information Administration. International Energy Outlook 2016 – EIA[EB/OL], 2016, [2022-03-25]. https://www.eia.gov/outlooks/ieo/pdf/0484(2016). pdf.
[3] Ghedamsi R, Settou N, Gouareh A, Khamouli A, Saifi N, Recioui B, Dokkar B. Modeling and forecasting energy consumption for residential buildings in Algeria using bottom-up approach. Energy and Buildings 2016; 121: 309-317.
[4] Wang Z, Srinivasan RS. A review of artificial intelligence based building energy use prediction:

contrasting the capabilities of single and ensemble prediction models. Renewable and Sustainable Energy Reviews 2017; 75: 796–808.
[5] Ding C, Feng W, Li X, Zhou N. Urban-scale building energy consumption database: a case study for Wuhan, China. Energy Procedia 2019; 158: 6551-6556.
[6] Yan D, Xia J, Tang W, Song F, Zhang X, Jiang Y. DeST—An integrated building simulation toolkit Part I: Fundamentals. Building Simulation 2008; 1(2): 95-110.
[7] Fahmy M, Mahmoud S, Abdelalim M, Mahdy M. Generic energy efficiency assessment for heritage buildings; Wekalat El-Ghouri as a case study, Cairo, Egypt. Energy Procedia 2019; 156: 166-171.
[8] Sun Y, Haghighat F, Fung BCM. A review of the-state-of-the-art in data-driven approaches for building energy prediction. Energy and Buildings 2020; 221:110022.
[9] Elbeltagi E, Wefki H. Predicting energy consumption for residential buildings using ANN through parametric modeling. Energy Reports 2021; 7: 2534-2545.
[10] Dong Z, Liu J, Liu B, Li K, Li X. Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification. Energy and Buildings 2021; 241: 110929.
[11] Raza M Q, Khosravi A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renewable Sustainable Energy Reviews 2015; 50: 1352–1372.
[12] Coelho VN, Cohen MW, Coelho IM, Liu N, Guimarães FG. Multi-agent systems applied for energy systems integration: State-of-the-art applications and trends in microgrids. Applied Energy 2017; 187: 820–832.
[13] Li X, Wen J, Bai E W. Developing a whole building cooling energy forecasting model for on-line operation optimization using proactive system identification. Applied Energy 2016; 164: 69–88.
[14] Ahmad MW, Mourshed M, Rezgui Y. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy and Buildings 2017; 147: 77–89.
[15] Katsatos AL, Moustris KP. Application of artificial neuron networks as energy consumption forecasting tool in the building of regulatory authority of energy, Athens, Greece[J]. Energy Precedia 2019; 157: 851-861.
[16] Xiao R, Wei Z Q, Zhai X Q. Hourly energy consumption forecasting for office buildings based on support vector machine. Journal of Shanghai Jiao Tong University 2021; 55(3): 331-336.
[17] Zhou X, Lin W, Kumar R, Cui P, Ma Z. A data-driven strategy using long short term memory models and reinforcement learning to predict building electricity consumption[J]. Applied Energy 2022; 306: 118078.
[18] Wang Z, Wang Y, Zeng R, Ravi SS, Sherry A. Random Forest based hourly building energy prediction. Energy and Buildings 2018; 171: 11-25.
[19] Zhong H, Wang J, Jia H, Mu Y, Lv S. Vector field-based support vector regression for building energy consumption prediction. Applied Energy 2019; 242: 403-414.
[20] Chu X, Zhao R. A building carbon emission prediction model by PSO-SVR method under multi-criteria evaluation. Journal of Intelligent & Fuzzy Systems 2021; 41: 7473-7484.
[21] Wang R, Lu S, Feng W. A novel improved model for building energy consumption prediction based on model integration. Applied Energy 2020; 262: 114561.
[22] Sun Y, Huang G, Xu X, Lai ACK. Building-group-level performance evaluations of net zero energy buildings with non-collaborative controls. Applied Energy 2018; 212: 565–576.
[23] Walker S, Khan W, Katic K, Maassen W, Zeiler W. Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings. Energy and Buildings 2020; 209: 109705.
[24] Breiman L. Random Forests. Machine Learning, 2001; 45: 5–32.
[25] Yu SK, Song YQ. Research on the application of GRA-RF in optimization and calculation of wage elements of rural credit financial enterprise. Chinese Journal of Management Science 2021; 20(04): 86-95.
[26] Jiang F W, Ma T, Zhang HW. High risk low return?explanation from machine learning based conditional CAPM model. Journal of Management Sciences in China 2021; 24(01): 109-126.
[27] Cai M, Pipattanasomporn M, Rahman S. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. Applied Energy 2019; 236:1078–1088.
[28] PECAN STREET. DATAPORT [EB/OL], 2020, [2022-03-25]. https://dataport.pecanstreet.org/.
[29] Weather Underground. Austin, TX Weather History [EB/OL], 2021, [2022-03-25]. https://www.wunder ground.com/history/monthly/us/tx/austin/ KAUS/date.
[30] Zhang P, Yin Z Y, Jin YF, Chan THT. A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest. Engineering Geology 2020; 265: 105328.