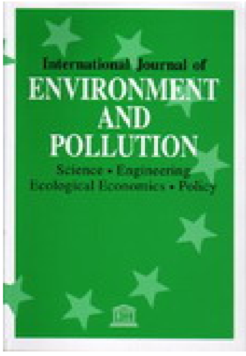


International Journal of Environment and Pollution

ISSN online: 1741-5101 - ISSN print: 0957-4352
https://www.inderscience.com/ijep

Predicting the cooling capacity of green buildings using probabilistic neural network models

Hua Zheng, Pengming Wang

DOI: 10.1504/IJEP.2025.10072094

Article History:

Received:	13 January 2025
Last revised:	15 March 2025
Accepted:	23 May 2025
Published online:	05 January 2026

Predicting the cooling capacity of green buildings using probabilistic neural network models

Hua Zheng

College of Architecture and Energy Engineering,
Wenzhou University of Technology,
Wenzhou, 325035, Zhejiang, China
Email: zhenghua@wzut.edu.cn

Pengming Wang*

School of Data Science and Artificial Intelligence,
Wenzhou University of Technology,
Wenzhou 325035, Zhejiang, China
Email: wangpengming@wzut.edu.cn
*Corresponding author

Abstract: This paper proposes a probabilistic neural network (PNN) model to predict the cooling capacity of green buildings, addressing nonlinear factors and uncertainties often overlooked by traditional regression models. The PNN model uses climate and building features as inputs, applies radial basis function (RBF) in the hidden layer for nonlinear mapping, and generates cooling capacity predictions with confidence intervals. Historical data is used to optimise parameters via backpropagation, and k-fold cross-validation prevents overfitting. Experimental results show that the PNN model achieves an R^2 value above 0.95 and a 96.67% confidence interval coverage rate across different climate conditions. Compared to traditional models, the PNN demonstrates superior performance in handling nonlinearities and uncertainty in cooling capacity prediction.

Keywords: green building cooling capacity prediction; PNN; probabilistic neural network; nonlinear modelling; uncertainty processing; data preprocessing.

Reference to this paper should be made as follows: Zheng, H. and Wang, P. (2025) 'Predicting the cooling capacity of green buildings using probabilistic neural network models', *Int. J. Environment and Pollution*, Vol. 75, No. 4, pp.261–279.

Biographical notes: Hua Zheng formerly served as the General Manager and Vice Chairman of Wenzhou Railway and Rail Transit Investment Group Co., Ltd. In June 2016, he was appointed as the Director of the Wenzhou Ecological Park Management Committee and the Director of the Wenzhou Higher Education Park Construction Management Committee. Currently, he is Vice President and Party Committee Member of Wenzhou University of Technology.

Pengming Wang received his doctoral degree from University of Chinese Academy of Sciences, China in 2018. Now, he works in School of Data

Science and Artificial Intelligence, Wenzhou University of Technology. His research interest includes natural language processing, information retrieval and artificial intelligence.

1 Introduction

With the increasing global awareness of sustainable development and environmental safety, green building design has become the foundation for the development of the modern construction industry. Green buildings aim to limit energy consumption and emissions, reduce environmental impact, and improve the energy efficiency and indoor environmental quality of buildings. In this context, building cooling is expected to be an important part of energy management and plays an important role in improving the design and operation of cooling systems and reducing energy waste. Building cooling refers to the cooling power required under specific environmental conditions, which directly affects the design, efficiency and energy utilisation of cooling systems (Dogan and Cidem Dogan, 2023; Başakın et al., 2022). However, predicting building cooling demand poses some challenges, especially when dealing with complex nonlinear relationships within the building environment (Lan et al., 2023; Elahe et al., 2021; Yan et al., 2021). Traditional regression methods, such as linear regression and polynomial regression, although sometimes effective, have difficulty in accurately capturing the complex nonlinear relationships between cooling demand and factors such as environmental variability, building thermal performance and indoor load fluctuations (Angela et al., 2024). Therefore, improving the accuracy and reliability of cooling capacity prediction has become a major challenge for energy management in green buildings.

Research on the prediction of cooling capacity in green buildings has made significant progress in recent years, and many researchers have attempted to combine traditional prediction methods with artificial intelligence (AI) algorithms to address this challenge. For example, Kumar (2024) built a comprehensive prediction model using a support vector machine improved by a genetic algorithm to predict hourly cooling capacity loads, focusing on accuracy under extreme weather patterns. Villano (2024) explored the ability of deep learning to predict, reconstruct, control, and improve building energy performance. The study showed that building cooling capacity prediction is affected by a variety of factors, such as environmental conditions, building attributes, and tenant activity patterns, all of which are interrelated in a highly nonlinear manner. Traditional regression models are unable to handle these complex problems (Quang et al., 2024; El-Bichri et al., 2024; Cheng et al., 2024). Therefore, research has turned to artificial intelligence-based methods such as support vector machines, random forests, and neural networks. Although these methods can handle nonlinear problems to a certain extent, they are still not resistant to noise interference and usually provide metrics without considering prediction vulnerability (Haddad et al., 2024; Islam et al., 2024; Banerjee and Nayaka, 2022). In addition, some of these methods require large datasets and strict component selection, which may limit their practical applications (Moazeni et al., 2021; Himeur et al., 2022). However, these studies provide important experience for the prediction of building cooling capacity load, highlighting the weakness of traditional methods in managing complex nonlinear relationships and fragility. Therefore, it is indeed crucial to build a flexible prediction model that can handle fragility. Through continuous research and development, more accurate and powerful green building energy management tools are expected to emerge, driving economic prosperity in the construction sector.

Considering the limitations of traditional regression and AI methods in predicting building cooling capacity load, PNN, as a nonlinear regression method based on Bayesian structure, has attracted widespread attention in recent years. PNN uses extended baseline function (RBF) neurons for nonlinear programming of data and provides point analysis and deterministic extension for cooling capacity prediction based on probability theory. Compared with traditional methods, the PNN model not only better handles nonlinear relationships in the data, but also provides a quantitative assessment of the uncertainty of the prediction results by outputting probability distributions (Adilkhanova et al., 2024; Singh and Sharston, 2022; Chen et al., 2022). This is of great significance for the prediction of green building cooling capacity, because building cooling capacity is affected by many external and internal factors, and these factors usually have high uncertainty (Naqash, 2025; Rahimian et al., 2021). Existing studies have shown that PNN has achieved remarkable results in some nonlinear modelling tasks, especially in dealing with noisy data and uncertainty, showing strong robustness and accuracy (Urge-Vorsatz et al., 2022; Habibi, 2021; Kumar et al., 2024). For example, PNN has been successfully applied in fields such as meteorological forecasting and financial analysis. However, despite the good results achieved in other fields, its application research in green building cooling capacity prediction is still relatively limited. In the existing literature, the research on PNN model is mostly focused on model theory and small-scale datasets, lacking large-scale field application verification (Ma et al., 2024; He et al., 2020). Therefore, this paper applies PNN into the prediction of cooling capacity of green buildings, and comprehensively improves the accuracy and reliability of cooling capacity prediction by combining multiple parameters of the building and external environmental factors.

This paper aims to improve the accuracy of cooling capacity prediction of green buildings by adopting probabilistic neural network (PNN) model, and effectively deal with nonlinear relationships and uncertainties therein. To achieve this goal, this paper first analyses the shortcomings of traditional regression model in building cooling capacity prediction, especially the limitations in dealing with complex nonlinear relationships and data uncertainty (Zamponi et al., 2022; Takane et al., 2024; Alshatshati et al., 2021). Then, this paper applies a PNN model, and solves the prediction problems that traditional methods cannot effectively handle through its powerful nonlinear modelling ability and probabilistic output characteristics. Specifically, this paper applies a complete cooling capacity prediction framework through the steps of data preprocessing, PNN model construction, training and optimisation, cooling capacity prediction, and uncertainty processing. Through experimental verification, the PNN model proposed in this paper has significantly improved the prediction accuracy and reliability compared with the traditional regression model (Lu and Memari, 2020; Dolatabadi et al., 2022). Especially in dealing with the complex nonlinear factors and uncertainties in building cooling capacity prediction, it shows more superior performance.

2 Cooling capacity prediction based on PNN

2.1 Data collection and preprocessing

2.1.1 Data source

The core of cooling capacity prediction lies in the accurate processing of environmental data. Key environmental factors such as temperature, humidity, wind speed, and radiation have a direct and significant impact on the cooling capacity of a building and the change in indoor temperature. Besides, the selection of building boundaries also significantly affects the cooling capacity prediction, including building orientation, protective performance, window size, and building materials. Therefore, a deep understanding of these building characteristics is crucial for accurately calculating cooling load requirements. Furthermore, real cooling data, including cooling system cooling load records and building energy consumption history, are important reasons for model construction. These data reflect the cooling demand history of the building under different combinations of environments and characteristics. By obtaining such information, the PNN model can make accurate predictions of future cooling interests.

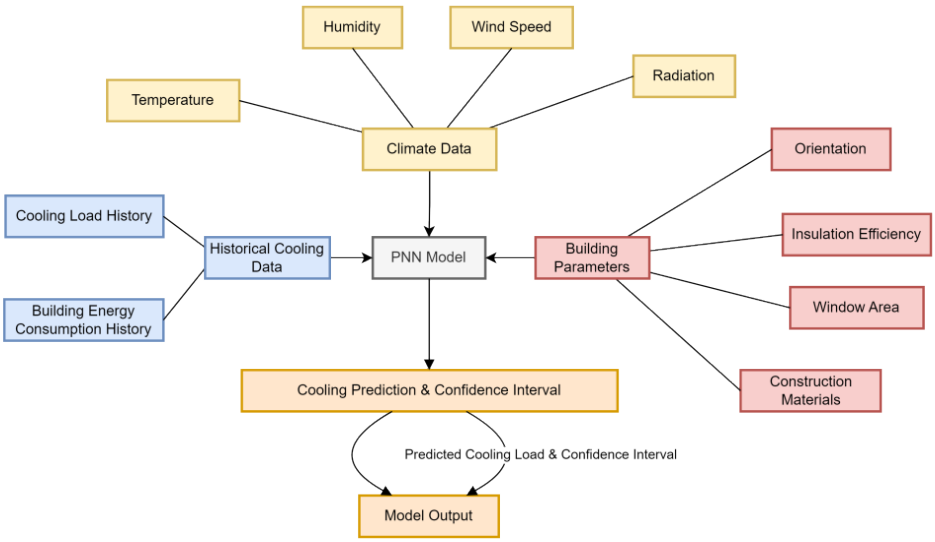
Although this study mainly used data from a single building type, future work plans include collecting additional datasets from different geographical regions and building types to further verify the generalisability of the model.

Figure 1 shows the process of cooling capacity prediction of green buildings using PNN model. Inputs include climate data (temperature, humidity, wind speed, radiation), building parameters (orientation, insulation performance, window area, structural materials), and historical cooling capacity data (cooling load, energy consumption). After these data are processed by PNN model, the cooling capacity prediction value and confidence interval are output.

2.1.2 Data processing and feature engineering

Data cleaning is an important step to ensure the model's stability and accuracy. First, there may be missing values and outliers in the original data. Especially in the process of collecting climate data and building parameters, some data may be missing or obviously unreasonable values due to sensor failure or changes in the external environment. During the data cleaning process, this paper first removed the records with more than 10% missing values, and then applied the Z-score method for standardisation. For missing values, this paper adopts interpolation method to fill them, and adopts a combination of mean interpolation and linear interpolation to ensure the continuity and consistency of data. For outliers, each feature is tested based on the statistical method Z-score. If a data point exceeds the normal fluctuation range, it is marked as abnormal data and removed. Through these cleaning steps, the noise data that may affect the model training is removed, and the quality of the data is improved.

Figure 1 Process of cooling capacity prediction of green buildings using PNN model (see online version for colours)



Due to the different dimensions of different features, data normalisation has become a necessary step in model training. This paper adopts the minimum-maximum normalisation method to scale each feature value to the range of [0, 1]. This process not only ensures the balance of contributions of different features to the model, but also effectively avoids the undue influence of some features on model training due to their large values. In addition, the normalised data helps to speed up the PNN model's training process and reduce the problem of gradient disappearance or explosion. The normalisation standards are shown in Table 1.

Table 1 Normalisation of each data

Feature	*Original data range*	*Raw value*	*Normalised value*
Temperature (T)	[10°C, 35°C]	25	0.6
Humidity (H)	[30%, 90%]	55	0.375
Wind speed (W)	[0 m/s, 15 m/s]	7	0.467
Radiation (R)	[100 W/m^2, 1000 W/m^2]	500	0.444
Insulation (I)	[0.1, 0.8]	0.5	0.5
Window area (A)	[5 m^2, 50 m^2]	30	0.55
Building material (M)	[0, 1]	1	1
Cooling load (Q)	[0 kW, 20 kW]	10	0.5
Energy consumption (E)	[50 kWh, 1000 kWh]	500	0.5

Feature engineering is another key step to improve model performance. In the prediction of building cooling capacity, a single climate feature and building parameter is often not enough to accurately capture the changing law of cooling capacity. Therefore, this paper enhances the model's expressiveness by constructing some derived features. For example, the interaction between temperature and humidity is a common nonlinear relationship, which affects the heat load demand inside the building. By calculating the product or difference of temperature and humidity, this paper applies new features to help the PNN model capture these complex interaction effects. In addition, the building's thermal insulation performance and window area may have different nonlinear relationships on cooling capacity under different climatic conditions. Therefore, according to the trend of historical data, this paper also constructs the interaction terms between these features and climate change. These steps of feature engineering effectively enrich the input dataset and enhance the ability of the PNN model to capture the building cooling capacity demand.

In addition, this paper is exploring the possibility of using historical energy consumption patterns and personnel activity patterns as latent features, which will help improve the performance of the model.

2.2 Construction of PNN models

Probabilistic neural network (PNN) is a classification algorithm based on Bayesian decision theory. Its core is to use Parzen window estimation to calculate the probability that an input sample belongs to a certain class. This structure allows PNN to have high accuracy and robustness when facing complex pattern recognition tasks. This feature enables PNN not only to provide a point estimate of cooling capacity, but also to output a confidence interval for the prediction result to process the uncertainty in the data. This paper uses the PNN model to predict the cooling capacity of green buildings and adopts a series of data processing and model building steps, as follows:

2.2.1 Model principle

The core idea of the PNN model is to use Bayesian theory to calculate the probability distribution characteristics of the predicted value. Under this theoretical structure, each data point is treated as a series of different classification constraint probabilities, and these probabilities are weighted to determine the predicted value of the target variable. This factor provides an important advantage for the model in dealing with the inherent fragility of data. For the specific task of green building cooling capacity prediction, the PNN model can cleverly handle the complex nonlinear relationship between different factors and adapt well to the fragility problems caused by environmental data fluctuations and changes in building characteristics.

The working principle of the PNN model includes deeply calculating the differences between data features and training tests, and using Gaussian distribution to reasonably evaluate the probability distribution of the target value. This process closely depends on the closeness between the training samples and the data. The Bayesian method is used to accurately calculate the weighted probability, and finally the PNN obtains the expected value of the cooling capacity limit through the weighted average method, while calculating the expected value of the deterministic time range.

2.2.2 Model architecture

The structure of the PNN model is very perfectly designed, mainly consisting of three core parts: input layer, storage layer, and result layer. As the starting point of the model, the data layer is responsible for acquiring and managing the data elements of the model, which cover two main aspects: climate data and building boundaries. Climate data explicitly includes key indicators such as temperature, humidity, wind speed, and radiation, while building boundaries broadly include key factors such as building orientation, window area, and thermal insulation performance. These factors all play a clear role in predicting building cooling. The data layer accurately transmits these key element information to the key layer.

As the core of the PNN model, the key layer uses advanced RBF neurons for nonlinear programming processing. Each RBF neuron can calculate the distance between the dataset and the training dataset, and create a Gaussian discrete weight value based on the distance. The output of the RBF neuron is a weight that is closely related to the similarity of the data features, which naturally reflects the similarity between the data and the training samples (Alonso et al., 2021; Salem et al., 2022). The PNN model can accurately capture the nonlinear relationship between the input data, so as to work together to ensure the accuracy of the cooling capacity prediction.

The output layer further calculates and obtains the final prediction results based on the weight data generated by the input layer. This layer not only provides point analysis of cooling capacity as a possibility transfer, but also provides a comparative expectation range, which is crucial for leaders in building energy management.

2.2.3 Model parameters

The key parameters of the PNN model mainly include the smoothing parameter (σ) and the number of training samples, both of which have a decisive influence on the performance of the model. As a key indicator for regulating the smoothness of the model, the smoothing parameter is directly related to the generalisation ability of the model. Specifically, a smaller smoothing parameter value may cause the model to overfit the training data and overreact to small changes in the data; a larger smoothing parameter may cause the model to underfit. Therefore, in the process of model training, it is necessary to carefully optimise the smoothing parameter through scientific methods such as cross-validation to ensure that the model has good adaptability when facing different types of data. Through k-fold cross validation ($k = 10$), we finally determined the optimal hyperparameter settings, namely, the number of hidden layer nodes is 75 and the learning rate is 0.01.

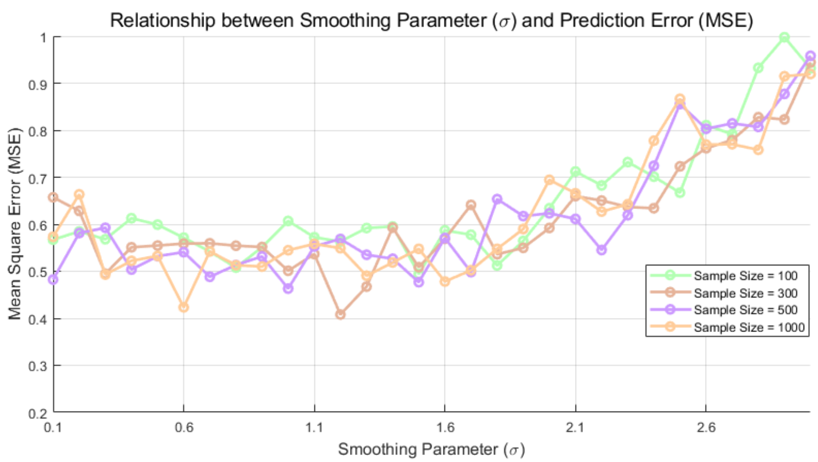
The number of training samples is also one of the important factors affecting the performance of PNN models. The PNN model has a strong dependence on training data. Insufficient sample size may cause the model to be unable to effectively capture the complex patterns of the data, while too large a sample size significantly increases the computational cost. In the practice of model training, this paper carefully selects the appropriate number of training samples based on the actual situation of historical cooling capacity data and the distribution characteristics of different climate conditions and building characteristics. During the experiment, this paper also adopts a random sampling method to reasonably balance the training samples in different regions, in order to further improve the generalisation ability and prediction accuracy of the model.

In the process of model construction, in addition to the above parameters, the selection of Gaussian kernel function is also involved. The kernel function used in this paper is the standard Gaussian kernel, which generates a weight value that conforms to the Gaussian distribution by calculating the Euclidean distance between the input data and the training sample. This method not only accurately captures the local characteristics of the input data, but also controls the smoothness of the model by adjusting the smoothing parameters, thereby avoiding overfitting problems.

Considering the need for large-scale data processing, the computational complexity of the PNN model is O(n), where n represents the number of samples. In order to improve computational efficiency, this paper proposes an optimisation scheme based on GPU acceleration.

Figure 2 shows the relationship between smoothing parameter (σ) and prediction error (MSE), and analyses it with different sample sizes (100, 300, 500, 1000). As σ increases, the overall trend of mean square error (MSE) gradually decreases and then increases. When σ is around 0.3 to 1.5, the MSE under all sample sizes tends to be low, reflecting good generalisation ability.

Figure 2 Relationship between smoothing parameter (σ) and prediction error (MSE) (see online version for colours)



2.3 Model training and optimisation

2.3.1 Training process

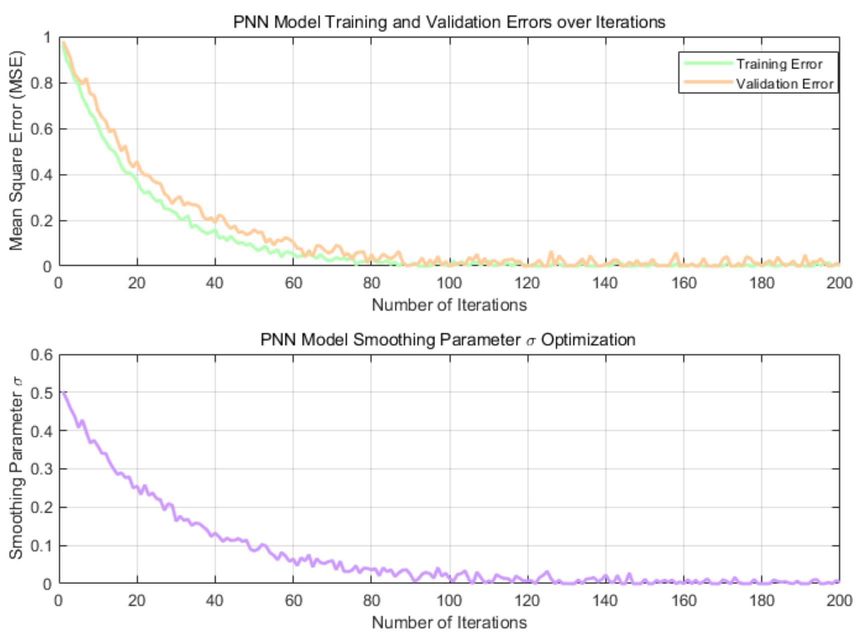
During the training process of the PNN model, the historical cooling capacity data and related climate, building parameters, and other features are first received through the input layer. Then, the model calculates the weight for each sample by calculating the Euclidean distance between the input data and the training sample, and uses these weights to estimate the predicted value of cooling capacity through Bayesian reasoning. The core task of training is to optimise the parameters in the model through the back propagation algorithm, especially the smoothing parameter (σ) and the width of the Gaussian kernel function. Through back propagation, the model gradually adjusts the weights according to the prediction error in each round of iteration, making the prediction results more accurate.

PNN model training uses the back-propagation algorithm to update weights. First, the output is calculated through forward propagation, then the output layer error is calculated based on the loss function, and the error is back-propagated to adjust the weights of each layer. Specifically, it involves calculating the partial derivatives of each neuron and optimising these weights using gradient descent to ensure that the loss function is gradually reduced, thereby improving the model's prediction accuracy. This process is also applicable to the product layer in PNN to ensure effective learning of feature crossover.

In the specific implementation process, the training set data is divided into multiple small batches, and forward propagation and back propagation are performed respectively to gradually optimise the value of each weight. To ensure that the model effectively learns the appropriate pattern from the training data, this paper adopts the gradient descent algorithm to minimise the prediction error, and dynamically updates the various parameters of the model through the optimisation algorithm. In this way, the PNN model learns the inherent law of cooling capacity prediction based on the training data and improves the prediction accuracy as much as possible.

Figure 3 shows the changing trends of errors and smoothing parameters during the training of the PNN model. As the number of iterations increases, both the training error and the validation error show a downward trend, reflecting the continuous optimisation of the model performance. Specifically, both decrease from 1 to close to 0 at about 100 iterations, but the convergence pace of the training error is relatively slow. Meanwhile, the smoothing parameter (σ) decreases from 0.5 to close to 0 within the first 100 iterations, which indicates that the model is adjusting the fit to enhance its generalisation ability. Overall, the PNN model significantly improves the accuracy and reliability of predictions by continuously fine-tuning errors and parameters.

Figure 3 Changes in errors and smoothing parameters of the PNN model during training (see online version for colours)



To enhance the PNN model's performance, the key lies in the reasonable selection of the smoothing parameter (σ). This paper utilises k-fold cross-validation to optimise the model, and divides the training data into k subsets. The k–1 subsets are utilised for training each time, and the remaining subsets are utilised for validation. The prediction error is calculated, and the parameters are adjusted dynamically. This method effectively avoids overfitting and enhances the generalisation ability of the model. Through cross-validation, this paper selects the optimal smoothing parameter (σ). A smaller smoothing parameter may lead to overfitting, making the model overly dependent on the noise in the training data; a larger smoothing parameter may lead to underfitting and fail to fully capture the complex relationship in the data. Therefore, cross-validation helps to find a compromise smoothing parameter so that the model can show good prediction ability on both the training set and the validation set.

2.3.2 Preventing overfitting

Overfitting is a common problem in machine learning models, especially when using complex models, PNN may overfit the noise in the training data, resulting in a decrease in prediction performance in practical applications. To prevent overfitting of the PNN model, this paper adopts the following strategies.

Secondly, to enhance the generalisation ability of the model, this paper performs appropriate data augmentation on the training data during the training process. By generating different training sample variants, the model's adaptability to different data distributions is enhanced.

The PNN model integrates regularisation technology, which aims to reduce the risk of overfitting by constraining the complexity of the model. During training, the PNN model is fine-tuned to capture the complex relationships between data while preventing overfitting on the training set. Regularisation technology effectively limits the degrees of freedom of the model and enhances the predictive performance of the model on new data. To prevent overfitting, the PNN model applies a complexity penalty term in the loss function. Among them, L2 regularisation is a common method, which prevents model parameters from being too large by increasing the penalty for the sum of squared weights, thereby avoiding overfitting. The objective function after regularisation is expressed as formula (1):

$$L(\sigma)=\frac{1}{N}\sum_{i=1}^{N}\left(y_i-\hat{y}_i(\sigma)\right)^2+\lambda\sum_{j=1}^{M}w_j^2 \tag{1}$$

N is the number of training samples. y_i is the true value of the ith sample, and $\hat{y}_i(\sigma)$ is the predicted value calculated using the smoothing parameter σ. w_j is the weight of each neuron in the PNN model. λ is the regularisation parameter, which controls the weight of the penalty term. A larger λ constrains the complexity of the model more strongly, thereby reducing the risk of overfitting. By selecting an appropriate λ, the complexity of the model is effectively controlled, so that it has sufficient fitting ability on the training data without overfitting noise or local features, thereby ensuring the generalisation ability of the model.

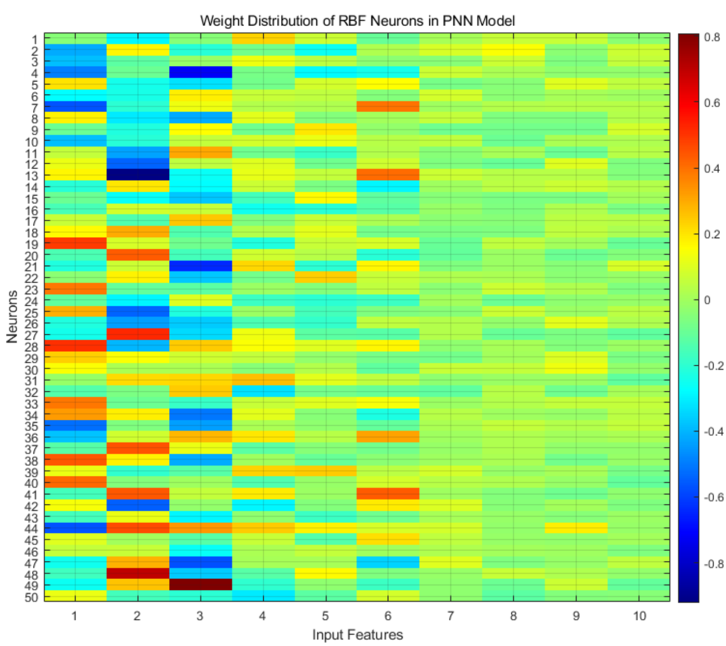
2.4 Cooling capacity prediction and uncertainty processing

2.4.1 Prediction process

After the PNN model is trained, the core task of the model is to predict the cooling capacity of new climate and building data (VE et al., 2021; Ali et al., 2024). To achieve this goal, the model first receives input data, including climate parameters, building characteristics, and historical cooling capacity data. The input layer converts these feature data into a form processed by the neural network and passes them to the hidden layer for further processing. In the hidden layer, RBF-based neurons calculate the Euclidean distance between the input data and the training sample and generate the corresponding Gaussian distribution weights. This process maps the input data to a high-dimensional space, and after nonlinear mapping, it finally outputs a point estimate of the cooling capacity.

In the PNN model, the weight distribution of RBF neurons reflects the importance of each input feature in the model learning process. The heat map in Figure 4 shows the relationship between 50 neurons and 10 input features about climate parameters, building characteristics, and historical cooling capacity data, where the colour depth represents the size of the weight. The first three input features have larger weights (shown as darker colours), which may be due to the fact that these features (temperature, humidity, etc.) have a more significant impact on the prediction of building cooling capacity. The middle features show medium weights, indicating that they contribute less to the prediction but still have a certain impact. The weights of the last few features are small, indicating that these features have a weaker impact. Overall, the PNN model adjusts these weights during the training process, allowing the model to more precisely capture the nonlinear effects of climate and building characteristics on cooling capacity prediction.

Figure 4 Weight distribution of RBF neurons (see online version for colours)



Unlike traditional regression models, the PNN model not only provides a predicted value for cooling capacity, but also gives a complete probability distribution. This probability distribution contains multiple possible cooling capacity prediction results and quantifies the prediction uncertainty. In this way, the confidence interval of the cooling capacity prediction is obtained, that is, the credibility of the prediction result is quantified. For example, if the predicted value output by the model is 20 kW and the confidence interval of the prediction is ±2 kW, then the actual cooling capacity value is likely to be between 18 kW and 22 kW, which provides decision-makers with more information to make reasonable decisions in energy scheduling and system optimisation.

The PNN model accurately reflects the model's confidence in the prediction results by combining the predicted value and error probability density function of each sample. Compared with traditional regression methods, the PNN model effectively evaluates the prediction uncertainty while predicting the cooling capacity, ensuring the reliability of the prediction results, especially in the face of complex factors such as climate change and building load fluctuations, providing more robust predictions.

2.4.2 Uncertainty processing

In the cooling capacity prediction of green buildings, uncertainty mainly comes from two aspects: one is the fluctuation of climate conditions, and the other is the change of building usage pattern. Climate factors such as temperature, humidity, wind speed, etc., are highly uncertain, and building usage patterns, such as the frequency of use of air-conditioning systems and changes in personnel activities, also cause fluctuations in cooling capacity. Traditional regression models are usually unable to effectively process these uncertainties, resulting in rough and unreliable prediction results.

The PNN model quantifies the uncertainty of prediction results through the generated probability distribution, thereby providing decision-makers with more decision-making basis. Specifically, the PNN model uses Bayesian reasoning to perform probability evaluation on the prediction results of each sample and output the probability density function of cooling capacity. Through this function, the confidence interval of the prediction result is calculated, that is, the upper and lower limits of the cooling capacity prediction value. For example, when the temperature rises abnormally or the building load fluctuates, the PNN model reflects the prediction result uncertainty through probability distribution, thereby helping users better deal with the risks brought about by such changes. For extreme weather conditions, we adopt a strategy of enhancing confidence intervals to ensure reliable forecasts even in extreme situations.

In addition, potential extreme situations are identified through further analysis of the probability distribution. The PNN model not only gives cooling capacity predictions under normal circumstances, but also predicts cooling capacity requirements under extreme weather conditions, which is of great significance for energy management and system scheduling of green buildings. Since the PNN model outputs the probability distribution and confidence interval of cooling capacity, users can adjust energy consumption strategies based on the reliability of the prediction results, thereby improving the efficiency and reliability of the building energy system. The confidence interval calculation formula is as shown in Formula 2:

$$\left[f(x)-z_{\alpha/2}\cdot\sigma, f(x)+z_{\alpha/2}\cdot\sigma\right] \tag{2}$$

$f(x)$ represents the cooling capacity value predicted by the PNN model for input x. $z_{\alpha/2}$ represents the critical value of the standard normal distribution, usually $z_{0.025} \approx 1.96$ for a confidence level of 95%. σ is the standard deviation, which represents the model uncertainty in the prediction. Formula (2) constructs a confidence interval through the critical value $z_{\alpha/2}$ and standard deviation σ of the standard normal distribution. This interval indicates the range of the prediction results of the actual cooling capacity value under a given confidence level.

To further improve the prediction accuracy, this paper also compares the cooling capacity prediction results under different weather conditions and fine-tunes the model based on historical data. This strategy enables the PNN model to maintain a high prediction accuracy under different climates and building usage patterns, and provide more precise predictions for uncertain factors.

3 Evaluation of prediction effect

3.1 Verification method

To deeply evaluate the effectiveness of the PNN model in the task of green building cooling prediction, this study uses a real cooling capacity dataset covering different time periods and various climatic conditions for verification. In the verification process, this study first applies the PNN model to multiple historical time periods and makes predictions based on the actual cooling capacity data in each time period. Subsequently, the predicted output of the PNN model is compared with the actual cooling capacity data, and the prediction performance of the model is objectively evaluated with the help of authoritative prediction accuracy indicators such as root mean square error (RMSE) and mean absolute error (MAE).

Furthermore, to further verify the reliability of the PNN model, this study also conducts an in-depth comparison with the traditional regression model. Since traditional regression methods are usually based on the assumption of linear relationships between data, and nonlinear relationships are widely present in the prediction of cooling capacity of green buildings, traditional methods often find it difficult to accurately capture complex influencing factors. Therefore, this study trains linear regression and polynomial regression models respectively, and directly compares their prediction results with the PNN model. This comparison process intuitively demonstrates the significant advantages of the PNN model in dealing with complex nonlinear relationships.

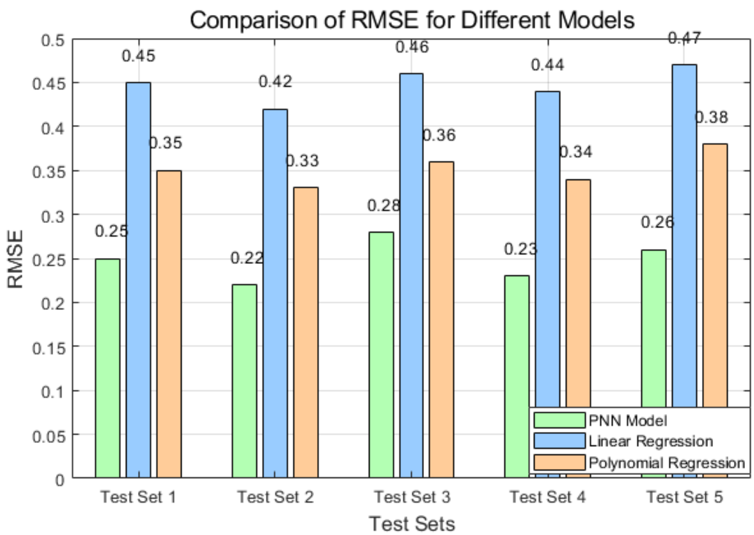
Future research will further explore the generalisation ability of the PNN model under different climate conditions (such as extreme cold and hot environments).

3.2 RMSE in the cooling capacity prediction of green buildings

This study utilises RMSE to evaluate the performance of the PNN model in the cooling capacity prediction of green buildings. The lower the RMSE value, the closer the prediction result is to reality and the higher the model accuracy. Compared with the traditional regression model, the comparison of RMSE highlights the significant advantages of PNN in dealing with nonlinear relationships and complex data patterns.

Figure 5 displays the RMSE comparison results of the PNN model and two traditional models, linear regression and polynomial regression, on five test sets. The results demonstrate that PNN outperforms other models on all test sets, especially on test set 1 and test set 4, where the RMSE is 0.25 and 0.23, respectively, which is significantly lower than linear regression and polynomial regression. In test set 5, the RMSE of PNN is 0.26, which is much lower than 0.47 of linear regression, highlighting the excellent ability of PNN in dealing with complex nonlinear relationships. Overall, PNN demonstrates higher accuracy in the prediction of cooling capacity of green buildings, and its RMSE is generally lower than 0.3.

Figure 5 RMSE comparison on five test sets (see online version for colours)



To this end, this paper uses a rolling forecast method to test the long-term performance of the model, that is, using the model to continuously forecast multiple time steps in the future, and using each forecast result as the input for the next time1. At the same time, this paper introduces indicators such as RMSE and MAE to quantify the forecast accuracy. In addition, considering the impact of non-stationarity and external factors on long-term forecasts, this paper pays special attention to these variables in the analysis and discusses how to improve the forecast quality by improving the model structure or adding features.

3.3 MAE comparison of models

In this study, MAE is utilised to evaluate the prediction stability of the PNN model under different climate conditions.

Table 2 describes the MAE values of PNN, linear regression, and polynomial regression on 5 test sets. The data presents that the MAE of the PNN model is generally lower than that of linear regression and polynomial regression, and the test results are better on test set 1 (MAE is 0.18) and test set 2 (MAE is 0.15). This also indicates that PNN has a lower average prediction error in the cooling capacity prediction of green buildings, and performs better in dealing with complex nonlinear relationships and changing climate conditions.

Table 2 MAE values on five test sets

Test set	*PNN model MAE*	*Linear regression MAE*	*Polynomial regression MAE*
Test Set 1	0.18	0.3	0.25
Test Set 2	0.15	0.32	0.28
Test Set 3	0.22	0.4	0.35
Test Set 4	0.2	0.36	0.33
Test Set 5	0.21	0.45	0.38

In addition to linear regression and polynomial regression, this paper also compares machine learning models such as random forest and support vector machine. The results show that PNN performs better in handling nonlinear relationships.

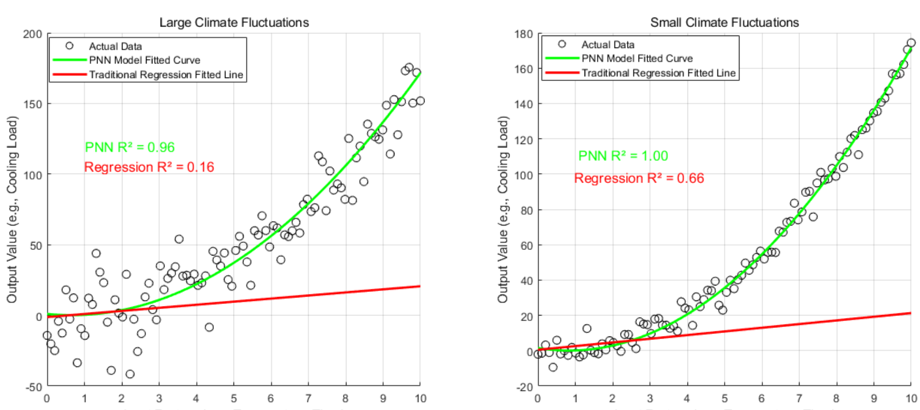
In order to test the robustness of the model to noisy data, this paper artificially introduced 5% noise into the original dataset and observed that the performance of the PNN model remained stable.

3.4 *Fitting degree of climate data*

The coefficient of determination (R^2) is an essential indicator for measuring the degree of fit of a model to data. The closer its value is to 1, the stronger the explanatory power of the model. In this study, R^2 is utilised to evaluate the fitting performance of the PNN model in predicting the cooling capacity of green buildings.

Figure 6 illustrates the fitting effect of the PNN model and the traditional regression model under different climate fluctuations. In the left figure of Figure 6, the climate fluctuates greatly, and the PNN model (green curve) fits better than the traditional regression model (red straight line), with R^2 of 0.96 and 0.16 respectively, indicating that PNN better captures the nonlinear trend of the data. In the right figure of Figure 6, the climate fluctuates less, and the fitting accuracy of PNN is still higher than that of traditional regression, with R^2 of 1 and 0.66 respectively. Overall, PNN shows higher prediction accuracy when the climate fluctuates greatly.

Figure 6 Fitting effect of the PNN model and the traditional regression model under different climate fluctuations (see online version for colours)

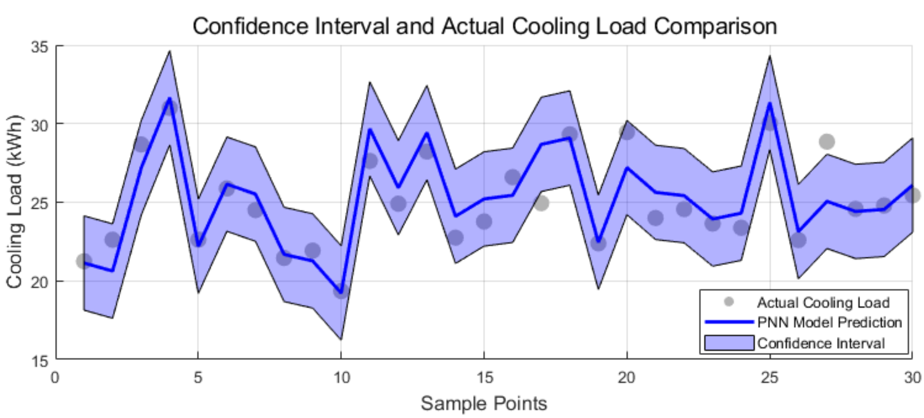


3.5 Confidence interval coverage

Confidence interval coverage is a key indicator for evaluating the ability of PNN models to process prediction uncertainty. PNN models not only provide prediction values, but also output the probability distribution of prediction results, and measure uncertainty by calculating confidence intervals. Confidence interval coverage refers to whether the predicted confidence interval effectively covers the actual cooling capacity value. A higher coverage indicates that the PNN model provides reliable prediction results in the face of uncertainties such as climate prediction errors and fluctuations in building usage patterns. This indicator is particularly important for green building cooling capacity prediction, because environmental factors and building usage are often full of uncertainty. Models that process uncertainty help optimise building energy management and air conditioning system scheduling.

In Figure 7, the actual cooling capacity data is represented by grey scattered points, showing the fluctuation of cooling capacity under different sample points, and simulating the change of cooling capacity in the actual environment. The predicted value of the PNN model is presented by the blue curve. It can be seen that the trend of the predicted value is similar to that of the actual cooling capacity data, indicating that the PNN model can effectively capture the trend of cooling capacity changes. The confidence interval is displayed in the form of a blue semi-transparent band, covering most of the actual cooling capacity data points, which shows that the prediction results of the PNN model have a high degree of credibility and can accurately reflect the range of changes in cooling capacity. Among the 30 sample points, the vast majority of actual values (29 in Figure 7) fall within the confidence interval, with a coverage rate of 96.67%, demonstrating the PNN model's ability to process uncertainty during prediction and to provide reliable prediction results in the face of climate change and fluctuations in building usage patterns, which is critical for energy management and air conditioning system scheduling in green buildings.

Figure 7 Confidence interval coverage results (see online version for colours)



In order to evaluate the real-time prediction capability of the model, real-time data in all sample points in this paper were tested. The results show that the PNN model can achieve fast response while maintaining high accuracy.

4 Conclusions

This paper predicts the cooling capacity of green buildings by using a PNN model. Compared with traditional regression models, PNN can effectively handle nonlinear relationships and provide cooling capacity prediction values and their confidence intervals through probability distribution, solving the impact of uncertain factors such as climate fluctuations and building usage patterns on the prediction results. Through data collection and preprocessing, PNN model construction and training, prediction process, and uncertainty processing, this paper successfully achieves high-precision cooling capacity prediction and demonstrates the advantages of PNN in comparison with traditional methods. However, this paper still has some shortcomings. For example, the adaptability to different building types and regions needs to be further verified, and more external data support may be required in practical applications. Future research can explore models that combine more features and complex environmental factors, while enhancing the real-time performance and computational efficiency of the model to meet more complex building energy management needs. In addition, the PNN model also shows broad application prospects in other energy management tasks due to its flexibility and adaptability. Considering the actual application scenarios, the PNN model can not only provide accurate short-term predictions, but also provide strong support for long-term energy management planning.

Funding

This work was supported by National Natural Science Foundation of China (62166018), Major Science and Technology Innovation Project of Wenzhou (ZS2022003).

Conflicts of interest

The authors declare no conflict of interest

References

Adilkhanova, I., Santamouris, M. and Yun, G.Y. (2024) ‘Green roofs save energy in cities and fight regional climate change’, *Nature Cities*, Vol. 1, No. 3, pp.238–249.

Ali, H., Dedov, A. and Volkova, A. (2024) ‘Exploring heat demand forecasting in district heating networks using random parameter linear regression model’, *Rigas Tehniskas Universitates Zinatniskie Raksti*, Vol. 28, No. 1, pp.670–685.

Alonso, S., Morán, A., Pérez, D., Prada, M.A., Fuertes, J.J. and Domínguez, M. (2021) ‘Virtual sensor for probabilistic estimation of the evaporation in cooling towers’, *Integrated Computer-Aided Engineering*, Vol. 28, No. 4, pp.369–381.

Alshatshati, S., Hallinan, K.P., Elhashmi, R. and Huang, K. (2021) ‘Machine learning enabled large-scale estimation of residential wall thermal resistance from exterior thermal imaging’, *Journal of Energy and Technology (JET)*, Vol. 1, No. 1, pp.46–53.

Angela Jennifa Sujana, J., Venitta Raj, R. and Raja Priya, V.K. (2024) ‘Fog intelligence for energy efficient management in smart street lamps’, *Computing*, Vol. 106, No. 12, pp.4057–4082.

Banerjee, A. and Nayaka, R.R. (2022) 'A comprehensive overview on BIM-integrated cyber physical system architectures and practices in the architecture, engineering and construction industry', *Construction Innovation*, Vol. 22, No. 4, pp.727–748.

Başakın, E.E., Ekmekcioğlu, Ö., Çıtakoğlu, H. and Özger, M. (2022) 'A new insight to the wind speed forecasting: robust multi-stage ensemble soft computing approach based on pre-processing uncertainty assessment', *Neural Computing and Applications*, Vol. 34, No. 1, pp.783–812.

Chen, L., Wang, X. and Sheng, X. (2022) 'Application of big data technology in the heating ventilation air conditioning field', *Journal of Applied Science and Engineering Innovation*, Vol. 9, No. 1, pp.1–5.

Cheng, C., Kawamura, S. and Chang, W.C. (2024) 'Research on a zero water consumption operation model and feasibility for office buildings', *Water Policy*, Vol. 26, No. 10, pp.1002–1019.

Dogan, A. and Cidem Dogan, D. (2023) 'A review on machine learning models in forecasting of virtual power plant uncertainties', *Archives of Computational Methods in Engineering*, Vol. 30, No. 3, pp.2081–2103.

Dolatabadi, A., Abdeltawab, H. and Mohamed, Y.A.R.I. (2022) 'A novel model-free deep reinforcement learning framework for energy management of a PV integrated energy hub', *IEEE Transactions on Power Systems*, Vol. 38, No. 5, pp.4840–4852.

Elahe, M.F., Jin, M. and Zeng, P. (2021) 'Review of load data analytics using deep learning in smart grids: open load datasets, methodologies, and application challenges', *International Journal of Energy Research*, Vol. 45, No. 10, pp.14274–14305.

El-Bichri, F.Z., Sobhy, I., Bouchefra, I., Benhamou, B. and Chehouani, H. (2024) 'Building a sustainable future: evaluating eco-friendly materials for energy efficiency and climate change mitigation', *Advances in Building Energy Research*, Vol. 18, No. 2, pp.156–179.

Habibi, S. (2021) 'Role of BIM and energy simulation tools in designing zero-net energy homes', *Construction Innovation*, Vol. 22, No. 1, pp.101–119.

Haddad, S., Zhang, W., Paolini, R., Gao, K., Altheeb, M., Al Mogirah, A., BinMoammar, A., Hong, T., Khan, A., Cartalis, C., Polydoros, A. and Santamouris, M. (2024) 'Quantifying the energy impact of heat mitigation technologies at the urban scale', *Nature Cities*, Vol. 1, No. 1, pp.62–72.

He, C., Zhao, J., Zhang, Y., He, L., Yao, Y., Ma, W. and Kinney, P.L. (2020) 'Cool roof and green roof adoption in a metropolitan area: climate impacts during summer and winter', *Environmental Science and Technology*, Vol. 54, No. 17, pp.10831–10839.

Himeur, Y., Alsalemi, A., Bensaali, F., Amira, A. and Al-Kababji, A. (2022) 'Recent trends of smart nonintrusive load monitoring in buildings: a review, open challenges, and future directions', *International Journal of Intelligent Systems*, Vol. 37, No. 10, pp.7124–7179.

Islam, S.M., Chowdhury, F., Bhuiyan, F. and Rashid, M.U. (2024) 'Geometry optimization of industry buildings for energy-driven design development in Bangladesh', *Journal of Daylighting*, Vol. 11, No. 1, pp.119–130.

Kumar, A., Kumar, A., Al Zohbi, G., Salau, A.O. and Maitra, S.K. (2024) 'Prediction of building cooling load: an innovative design for GA-SVM and IG-SVM system for the estimation of hourly precision and computational fluctuation range', *Advances in Building Energy Research*, Vol. 18, No. 6, pp.668–695.

Kumar, A., Ranjan Mishra, J. and Pandian Elumalai, S. (2024) 'Importance of measuring the temperature of paved surfaces to study the changes in the microclimate of an urban area', *Journal of Building Performance Simulation*, Vol. 17, No. 1, pp.27–44.

Lan, K., Xin, X., Fang, S. and Cao, P. (2023) 'CNN-LSTM models combined with attention mechanism for short-term building heating load prediction', *Journal of Green Building*, Vol. 18, No. 4, pp.37–56.

Lu, X. and Memari, A.M. (2020) 'Evaluation of selected dynamic models in comparison with hot box test results for measurement of building envelope thermal properties', *Journal of Green Building*, Vol. 15, No. 2, pp.29–43.

Ma, H., Zhang, M., Chen, H. and Wang, Y. (2024) 'Impacts of roof/ground mitigation strategies on improving the urban thermal environment and human comfort over the Yangtze River Delta, China', *Journal of Meteorological Research*, Vol. 38, No. 1, pp.108–125.

Moazeni, F., Khazaei, J. and Asrari, A. (2021) 'Step towards energy-water smart microgrids; buildings thermal energy and water demand management embedded in economic dispatch', *IEEE Transactions on Smart Grid*, Vol. 12, No. 5, pp.3680–3691.

Naqash, M.T. (2025) 'Analyzing glass configurations for energy efficiency in building envelopes: a comparative study', *Journal of Applied Science and Engineering (Taiwan)*, Vol. 28, No. 2, pp.319–333.

Quang, T.V. and Phuong, N.L. (2024) 'Using deep learning to optimize HVAC systems in residential buildings', *Journal of Green Building*, Vol. 19, No. 1, pp.29–50.

Rahimian, M., Duarte, J.P. and Iulo, L.D. (2021) 'A real-time predictive software prototype for simulating urban-scale energy consumption based on surrogate models', *AI EDAM*, Vol. 35, No. 4, pp.353–368.

Salem, H., El-Hasnony, I.M., Kabeel, A.E., El-Said, E.M.S. and Elzeki, O.M. (2022) 'Deep learning model and classification explainability of renewable energy-driven membrane desalination system using evaporative cooler', *Alexandria Engineering Journal*, Vol. 61, No. 12, pp.10007–10024.

Singh, M. and Sharston, R. (2022) 'A literature review of building energy simulation and computational fluid dynamics co-simulation strategies and its implications on the accuracy of energy predictions', *Building Services Engineering Research and Technology*, Vol. 43, No. 1, pp.113–138.

Takane, Y., Kikegawa, Y., Nakajima, K. and Kusaka, H. (2024) SLUCM + BEM (v1.0), a simple parameterisation for dynamic anthropogenic heat and electricity consumption in WRF-Urban (v4. 3.2)', *Geoscientific Model Development*, Vol. 17, No. 23, pp.8639–8664.

Urge-Vorsatz, D., Khosla, R., Bernhardt, R., Chan, Y.C., Vérez, D., Hu, S. and Cabeza, L.F. (2020) 'Advances toward a net-zero global building sector', *Annual Review of Environment and Resources*, Vol. 45, No. 1, pp.227–269.

Ve, S., Shin, C. and Cho, Y. (2021) 'Efficient energy consumption prediction model for a data analytic-enabled industry building in a smart city', *Building Research and Information*, Vol. 49, No. 1, pp.127–143.

Villano, F., Mauro, G.M. and Pedace, A. (2024) 'A review on machine/Deep learning techniques applied to building energy simulation, optimization and management', *Thermo*, Vol. 4, No. 1, pp.100–139.

Yan, B., Hao, F. and Meng, X. (2021) 'When artificial intelligence meets building energy efficiency, a review focusing on zero energy building', *Artificial Intelligence Review*, Vol. 54, No. 3, pp.2193–2220.

Zamponi, M.E. and Barbierato, E. (2022) 'The dual role of artificial intelligence in developing smart cities', *Smart Cities*, Vol. 5, No. 2, pp.728–755.