Building and Environment 211 (2022) 108723

Contents lists available at ScienceDirect

Building and Environment

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

Hybrid model combining multivariate regression and machine learning for the rapid prediction of interior temperatures affected by thermal diodes and solar cavities

Yi He [a,b], Fangcheng Kou [a,b], Xin Wang [a,b], Ning Zhu [a,b], Yehao Song [a,b,*], Yingnan Chu [a,b], Shaohang Shi [a,b], Mengjia Liu [a,b], Xinxing Chen [a,b]

[a] School of Architecture, Tsinghua University, Beijing, China
[b] Key Laboratory of Urban–Rural Eco Planning and Green Building, Ministry of Education, Beijing, China

ARTICLE INFO

Keywords:
Hybrid model
Multivariate regression
Machine learning
Heat pipe thermal diode
Solar cavity
Rapid prediction

ABSTRACT

Sustainable design often requires highly efficient building performance evaluations. This study proposed a hybrid model combining multivariate regression modelling (MRM) and machine learning modelling (MLM) for the rapid prediction of interior temperatures affected by heat pipe thermal diodes and solar cavities based on experimental data. A heat pipe thermal diode can promote unidirectional heat transmission from the solar cavity on the south side of our newly built experimental house to the indoor environment to increase the interior temperature and reduce the heating load in cold climates. Experimental data were collected and then imported, cleaned, and split according to MRM and MLM requirements, respectively. In MRM, linear multivariate formulas were generated according to the thermal diode's two different working conditions. In MLM, a machine-learning model was created and trained using the experimental data. The results our hybrid model produced were comprehensively evaluated via R-square, statistical discrepancies, and complex MRM analyses. The similarity between the prediction and experimental results clearly demonstrates our model's accuracy and efficiency. This research was an original attempt to integrate emerging computational tools and provide a means to perform highly efficient quantitative analysis of indoor thermal environments for environmental studies and sustainable designs in the early stages.

1. Introduction

High-performance architectural components and specific building configurations play important roles in sustainable design [1]. Heat pipe thermal diodes and solar cavities are examples. The utilization of these facilities has been recommended in some passive design strategies in response to cold climates, as they have great potential to increase the interior temperature and reduce the heating load in winter. A heat pipe thermal diode is a vital component of thermal management systems. It transfers heat through the pipe in one direction only, while preventing heat conduction in the opposite direction [2]. Thermal conduction is driven by the working fluid's natural circulation through the system. The working fluid can change state in the liquid and vapor phases without any energy input [3,4]. In this study, a solar cavity was defined as an empty space near a building's exterior surface. A solar cavity can absorb sunlight through transparent glass. Radiation energy is then transformed into heat energy and preserved in the solar cavity. The greenhouse effect undergirds this process [5]. An example is the space in a double-skin façade [6,7].

A solar cavity can be used to collect solar heat, and a heat pipe thermal diode can be used to transport heat to indoor environments. Their integration in a building would generate complex influences on thermal performance [8]. Previous studies have rarely considered such complex influences, and the conventional study methodology is relatively time consuming and tedious. Specifically, the essential steps of building modelling, thermal simulation, and results analysis in an iterative study all require significant time commitments. This motivated us to study influences, especially on indoor environmental parameter

* Corresponding author. School of Architecture, Tsinghua University, Beijing, PR China.
E-mail addresses: yihe2021@mail.tsinghua.edu.cn (Y. He), kfc19@mails.tsinghua.edu.cn (F. Kou), wangxinlj@tsinghua.edu.cn (X. Wang), 13717742900@163.com (N. Zhu), ieohsong@mail.tsinghua.edu.cn (Y. Song), zhuyn18@mails.tsinghua.edu.cn (Y. Chu), shish19@mails.tsinghua.edu.cn (S. Shi), liumj17@mails.tsinghua.edu.cn (M. Liu), chenxx20@mails.tsinghua.edu.cn (X. Chen).

https://doi.org/10.1016/j.buildenv.2021.108723
Received 23 September 2021; Received in revised form 20 December 2021; Accepted 22 December 2021
Available online 27 December 2021
0360-1323/

prediction methodology.

1.1. Relevant studies on heat pipes and solar cavities

A heat pipe system usually consists of an evaporation section, an adiabatic section, and a condensation section. It aims to transfer heat from a hot reservoir to a cold reservoir using a working fluid that changes state in vapor and liquid phases. The heat flowing into the evaporation section is transferred to the working fluid, causing the liquid to evaporate. The vapor flows out of the evaporation section through the adiabatic section and into the condensation section. The vapor then condenses, and the heat released in the condensation section passes into the cold reservoir. The liquid returns to the evaporation section via capillary wicking. Heat can be conducted constantly via this process during circulation [9,10].

In 2000, Varga et al. tested thermal diode panels and incorporated heat pipes under cooling season conditions. They measured the temperature distributions in an experimental test facility and used numerical simulations to evaluate the thermophysical properties [2]. Liu et al. (2014) developed a three-dimensional unsteady thermal and hydrodynamic model to analyze fluid flow and heat transfer in a flat-plate oscillating heat pipe. The model's reasonability was experimentally verified. They investigated detailed bubble behavior and size distribution. The study results suggested that the dispersed bubbles in the heat pipe were produced by nucleate boiling in the evaporator and short vapor plug condensation in the condenser [11]. Sun et al. (2018) conducted an experiment to evaluate a prototype radiant heating terminal with a thermoelectric cooling unit and a flat heat pipe. The working characteristics under practical operating conditions were determined. They developed a simplified energy equilibrium model for a heating terminal [12]. In another study, they investigated the thermal performance of a novel radiant heating and cooling terminal integrated with a flat heat pipe. Their results showed differences in the value ranges of the heating and cooling temperature response speeds and vertical temperature [13].

Wang et al. (2020) proposed a closed-cavity radiation system with a built-in finned heat exchanger. The system's heat transfer mechanism, performance, and operational characteristics were analyzed [14]. Girma et al. (2021) carried out an experimental investigation of the cavity air gap depth to enhance ventilated rain screen walls' thermal performance. Their study found that a certain cavity depth had the lowest airflow rate and resulted in the highest heat gain to the building [15]. Motivated by the considerable potential for solar energy utilization, Wang et al. (2021) investigated the concept, design, and performance of a novel solar double-skin façade [16]. Nghana et al. (2021) adopted the computational fluid dynamics (CFD) method to investigate the impact of ventilation cavity design on rain screen wall assemblies' energy performance. The CFD model was validated with experimental data, considering the root mean square error [17].

1.2. Relevant studies using rapid evaluation methods

In a systematic review of the use of machine learning algorithms, machine learning was defined as an approach using computers to simulate human learning in order to identify and acquire knowledge to improve performance of some tasks [18]. There are several frequently used machine learning algorithms, including linear regression, artificial neural networks, decision trees, support vector machines, and support vector regression [19]. The technology can reveal the hidden and complex relationships among correlative parameters in dynamic environments based on sufficient training with numerous samples–and these are near impossible for experiments or simulations to present [20]. Kobayashi et al. (2021) used machine learning to examine prediction of the drag reduction effect caused by pulsating pipe flows. Their research proved that the machine learning model can accurately estimate the drag reduction effect by predicting the nonlinearity of the pressure gradient during acceleration and deceleration [21].

Although the technology has several advantages and powerful functions, few studies have adopted a rapid evaluation method. Wang et al.'s (2021) comprehensive review focused on the application of nanofluids in heat pipes based on machine learning. Their review suggested that the utilization of machine learning technology is still primitive in several aspects, including lack of comprehension of particular properties and predictive models' limitations [23]. Wang et al. (2021) critically reviewed heat pipe technologies, especially the aspect of using advanced big data and machine learning techniques to optimize and enhance the performance of heat pipe technologies. Their review found that the existing computerized analytical and numerical models for heat pipe simulation are extremely time consuming and impractical, as they usually require large parametrical data input. They suggested that the utilization of big data and machine learning technologies could promote prediction efficiency. Relevant experimental data, such as operation and performance parameters, could play a crucial role in training machine learning models [24].

1.3. Summary

In summary, few studies exist on indoor environments affected by heat pipes, thermal diodes, and solar cavities, although these technologies' application has been recommended in some sustainable designs. In particular, (1) few studies have considered the complex influence heat pipe thermal diodes and solar cavities exert; (2) influenced thermal environmental parameters are relatively difficult to predict; and (3) despite their proven effectiveness, the utilization of rapid evaluation methods in relevant studies has been relatively low. Conventional methods, such as computational simulations and mathematical models, are well recognized and widely used. However, the utilization of these methods is often complicated and time consuming. Simulations usually require graphic modelling, simulation settings, and lengthy computing processes. It is necessary to calculate several equations to solve mathematical models. In particular, for comparative investigations, iterative analysis takes a relatively long time. Although multiple computational tools such as multivariate regression and machine learning have powerful predictive functions, the emerging tools have rarely been used together to analyze built environments [25,27], especially regarding the interior temperature influenced by a combined system of heat pipe thermal diodes and solar cavities.

In this study, a hybrid model combining multivariate regression modelling (MRM) and machine learning modelling (MLM) was proposed. It aimed to rapidly predict interior temperatures affected by heat pipe thermal diodes and solar cavity systems based on experimental data. Experimental houses with and without a heat pipe thermal diode system were designed and built. The cavities on the houses' south side functioned as greenhouses to absorb solar heat and increase the interior temperature. The heat pipe thermal diode system can promote the transfer of heat from the cavity to the indoor environment. As the system only enables unidirectional heat transmission, indoor heat could be preserved when the interior temperature was higher than the exterior temperature. This would significantly reduce the heating load in cold climates. After experimental setup, data were measured and collected for the development of MRM and MLM. During development, the experimental data were imported, cleaned, and split according to the modelling requirements. In MRM, linear multivariate regressions were implemented to generate linear multivariate formulas according to the thermal diode's two different working conditions. In MLM, a machine learning model was created and trained using the experimental data. These methods were selected because of their relatively high accuracy and suitability for describing the features of the heat conduction process. The hybrid model immediately predicted the interior temperatures based on a few primary parameters. The results were evaluated comprehensively through R-square, statistical discrepancies, and complex MRM analyses. The research outcomes can provide useful information for built environment analysis and architectural design optimization. There is great potential for universal application of the hybrid model development framework, as presented in Fig. 1.

2. Hybrid model development

The hybrid model was developed using multivariate regression and machine learning methods based on on-site measurements taken from a full-scale experimental house. Experiment setup, MRM, and MLM are explained in this section.

2.1. Experiment setup

Two experimental houses were built for comparison on the Tsinghua University campus in Beijing. A heat pipe thermal diode system was installed in one house but not the other. As shown in Fig. 2, the two houses had the same cuboid shape. They were also the same size: 2.1 m high, 2.1 m wide, and 2.1 m long. Each house featured a cavity on the south side. The experimental houses were built using refabricated thermal insulation materials. The research presented in this paper focused on studying the complex heat conduction process that occurred in each house, influenced by the heat pipe thermal diode system. Temperature was measured using twelve wireless data loggers located in the outdoor space, the solar cavity, the heat pipe's evaporation, adiabatic, and condensation sections, the exterior and interior wall surfaces, and the middle of the indoor space (Fig. 3). They were calibrated by placing their waterproof probes in a bucket containing a mixture of water and ice. The water and ice mixture's stable temperature (0 °C) was used as the benchmark to adjust the measured temperature values [26].

Owing to the greenhouse effect, the solar cavity can generate relatively high temperatures due to the influence of daytime solar radiation [15,28]. The heat pipe thermal diode system exhibited high thermal conduction in only one direction. Its evaporation and condensation sections were in the solar cavity and indoor space, respectively. It was adopted to promote heat conduction from the solar cavity to the indoor space and thermal conservation by preventing heat flow in the opposite direction. The solar cavity and heat pipe thermal diode system worked together to increase the temperature of the indoor environment and reduce the heating load in cold climates [29,30].

After experiment setup, numerous environmental parameters were measured and collected from the house that had been fitted with a heat pipe thermal diode system (Fig. 4). This research focused on the following parameters: temperature of the outdoor space (TO), temperature of the cavity (TCA), temperature of the heat pipe's evaporation section (TES), temperature of the heat pipe's condensation section (TCS), temperature of the exterior wall surface (TEX), temperature of the interior wall surface (TIN), and temperature in the middle of the indoor space (TM). The experimental facilities required adjustments at the beginning of the experiment. The indoor space's isolation from the outdoor environment was compromised by repeated entry to the house to facilitate adjustment. Human disturbance may influence heat conduction and cause fluctuations in the interior temperature. Hence, post-adjustment measurement data were used, and the remaining data were discarded to ensure accurate model development. Therefore, experimental data for the period 11–21 May 2021 were selected for this study in the interest of data stability and integrity (Fig. 5). In its entirety, the abovementioned work took nearly 4 months in total: the construction of the experimental houses took approximately 1 month, the installation of measurement equipment took approximately 2 weeks, and the experiments took approximately 2 months.

2.2. Multivariate regression modelling

MRM aimed to generate linear multivariate formulas with interior temperature prediction functions for relatively convenient utilization in the early stage of the design process [31]. According to Fourier's law of heat conduction, the heat flux along a certain direction should be the product of the temperature gradient in that direction and thermal conductivity. Thermal conductivity is usually a coefficient for homogenous materials [35]. This study used several temperatures taken from the exterior side of the wall as independent variables to predict the dependent variable, interior temperature. Linear multivariate formulas with several coefficients should be appropriate, and MRM could help establish a correlation between the independent and dependent variables [36]. This method is more accurate than simple linear regression. The resultant prediction is relatively convenient for manual calculations. The multivariate regression model was developed in the R Project environment. The software tool provides a wide variety of statistical analysis functions, such as producing high-quality, well-defined plots, including mathematical models.

2.2.1. MRM data preprocessing

Heat conduction in the experimental house was analyzed before MRM. The heat pipe thermal diode system had two major working conditions according to the positive and negative temperature difference between TES and TCS: (1) When TES was higher than TCS, the heat could be conducted from the evaporation section to the condensation section through the heat pipes. (2) When TES was lower than TCS, heat conduction from the evaporation section to the condensation section ceased.

Considering the experimental house's heat conduction features, we decided, in the interest of achieving higher accuracy, to use two multivariate linear equations to predict the interior temperature (TM) for the two working conditions, instead of a single equation. Four parameters,

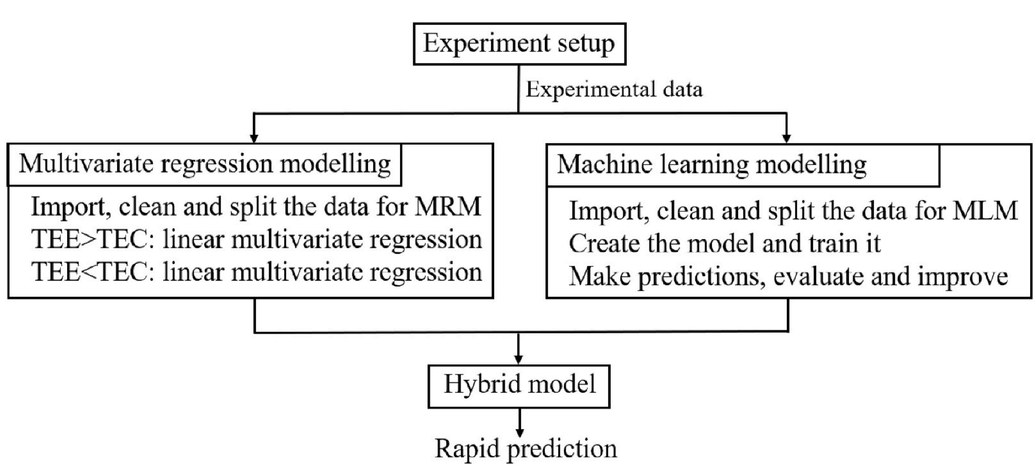

Fig. 1. MRM–MLM hybrid model development framework.

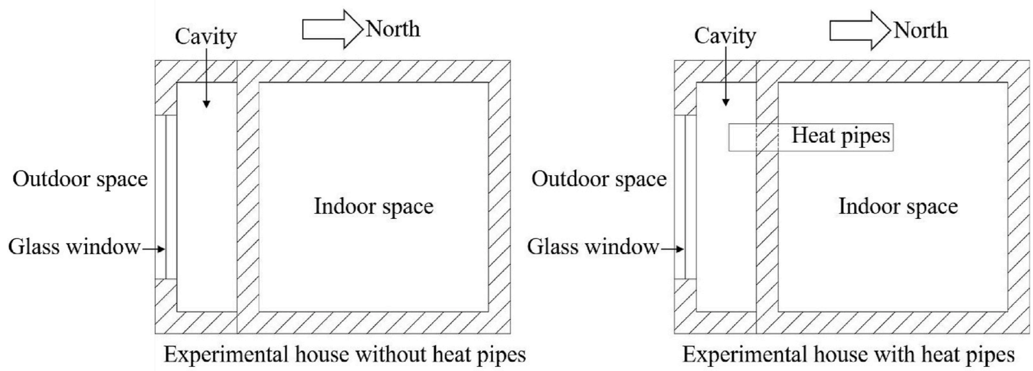

Fig. 2. Configurations of the two experimental houses with and without the heat pipe thermal diode system.

Fig. 3. Left: The experimental houses. Right: The data loggers and their probes located in the heat pipe system and the middle of the indoor space.

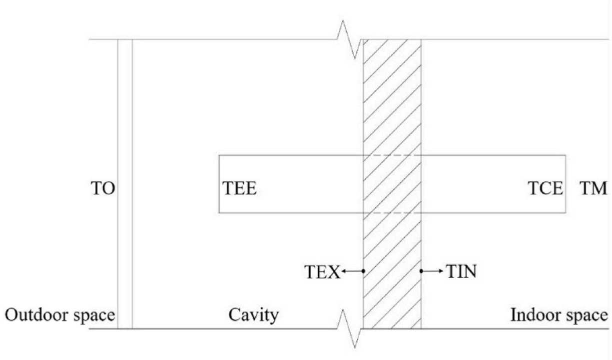

Fig. 4. Temperatures measured in the experimental house fitted with a heat pipe thermal diode system.

TO, TES, TCA, and TEX, were used to calculate TM. The temperature data measured in the experiment were split into two parts: one for the development of the multivariate regression model and the other for model validation. Subsequently, the data were further divided according to the two working conditions. The data were collected and saved in comma-separated values (CSV) files, following the recommendations contained in the R Project manuals and relevant instructions [32–34].

2.2.2. Multivariate regression

Several libraries were imported in R Project, including “ggplot2,” “ggcorplot,” “tidyr,” and “GGally.” To read the CSV files, their file paths were entered, and the data were converted into matrices for further processing. Specific data pertaining to TES, TCA, TEX, and TM were defined and selected according to the matrices’ rows and columns. TM data were used as the regression target. A new matrix containing TES, TCA, TEX, and TM data was established for the multivariate regression. A linear model fit function was used to implement the regression. A typical linear model usually takes the form of “response terms,” where “response” is the response vector and “terms” are a series of terms specifying a linear predictor for the “response.” If the “response” is a matrix, the linear model should be fitted separately to each of the matrix’s columns via least-squares. A summary and analysis of variance table of the regression results could be obtained and printed, including the significance analysis as well as the R-square, coefficients, and intercepts calculations, and so on. Finally, the coefficients of the linear multivariate formulas and MRM were visualized using the generic plotting function.

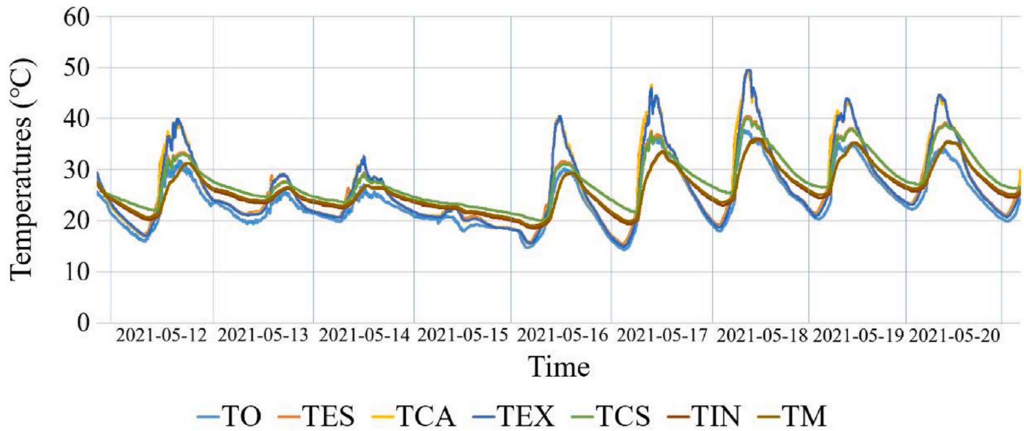

Fig. 5. The experimental house's temperature curves.

2.3. Machine learning modelling

Similar to MRM, MLM aims to generate a machine learning model capable of rapidly predicting the interior temperature in the early stage of the design process. Adequate experimental data were available for model training. Machine learning technology was used for the hybrid model because of its suitability for nonlinear prediction [31,37]. In this study, the heat conduction process was different in different parts of the experimental house. Interior temperature exerts a relatively complicated influence [38,39]. The Anaconda tool was selected because the other editors are not ideal for MLM, given the need to frequently inspect data from a technical aspect.

The "Numpy," "Pandas," and "SciKit-Learn" (SK-Learn) libraries were adopted in MLM to provide algorithms with different functions. A multidimensional array of data, such as data with a two-dimensional structure, can be provided and analyzed. Data were selected and analyzed in rows and columns in a data frame similar to an Excel spreadsheet. Common prediction algorithms, such as decision trees and neural networks, can be obtained from the SK-Learn library. Prediction and prediction evaluation were mainly used in MLM. The modelling process involved a number of steps, as explained in detail in Fig. 6.

2.3.1. MLM data importation, cleaning, and splitting

Given the necessity of ensuring that the input data were in good, clean shape for MLM, the original experimental data were imported, cleaned, and split in the first step. The data type was considered for MLM before importing, cleaning, and splitting. The data were pre-processed and stored in CSV files. Irrelevant, duplicate, and incomplete data were removed or modified. Data with different formats, such as data in a text format, were converted into numerical values. This is because the model would produce incorrect results if it learned bad patterns from inappropriate data containing repetitive or irrelevant information. In this research, the Excel files that contained data reflecting the temperature in different parts of the experimental house were read and standardized first. TES, TCA, and TEX data were dropped as the original feature parameters; TM data were set as the prediction regression target for the original feature parameters. All data were split into segments for model training and testing. Appropriate training and testing can significantly improve the accuracy of the model's results. In this study, an algorithm with the Pandas library's split function was used to process the data, including building information and wind environment analysis results. The parameters were set appropriately in the algorithm to determine the proportion of training and test data.

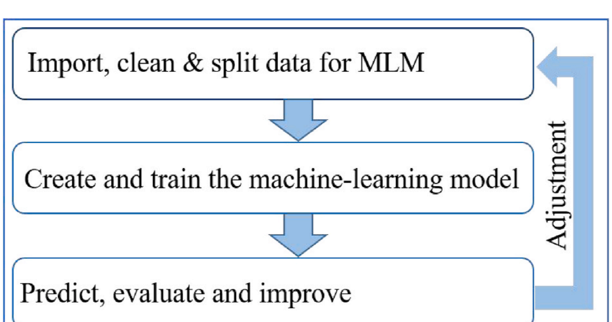

Fig. 6. The machine learning modelling process.

2.3.2. Model creation and training

The model was created and trained at the second step. An algorithm was selected to analyze the data according to the specific problem and input data, given the advantages and disadvantages in terms of accuracy and performance. The extremely randomized trees regressor (ERTR) prediction model was adopted for algorithm development owing to its relatively high efficiency. ERTR is a decision tree ensemble method from the open-source SK-Learn project with multiple functions of classification, regression, and clustering [40]. It combines several base estimators' predictions to reduce variance and improve generalizability and robustness. ERTR's randomness is a further step in computing splits, compared to the similar forests of randomized trees (FRT) averaging algorithm. FRT is usually composed of multiple independent decision trees constructed independently based on a random data subset [41]. The forests' predictions were averaged using bootstrap aggregation and random feature selection, and several individual decision trees were trained in parallel [42]. The ERTR algorithm can generate thresholds for all candidate features and select the best one based on the splitting rule. Therefore, ERTR outperforms FRT in terms of reducing the prediction model's variance [43]. Training data were fed into the prediction model for model training. The prediction model can automatically look for the pattern in the data based on the ERTR algorithm by making the original feature parameters fit the prediction regression target. The initial prediction model was therefore capable of making primary predictions of TM results according to the TES, TCA, and TEX parameters.

2.3.3. Prediction, evaluation, and improvement

At the third step, some primary predictions were made using the machine learning model, the prediction results were evaluated, and the model was improved based on the evaluation. Given that the primary predictions were sometimes inaccurate at the beginning, it was necessary to promote prediction accuracy through evaluation-based improvement. In particular, there was a need to ensure the precision of the TM results for this study. Two major methods were used to improve the prediction model. The first was to find appropriate relatively high-performance algorithms to produce accurate results in order to solve specific problems. The second was to modify the algorithms' particular parameters to optimize prediction accuracy. As explained above, we selected the ERTR algorithm for this study after testing several algorithms. Adoption of the ERTR algorithm could lead to relatively accurate TM predictions in a short time. We adjusted the parameters with particular values of data splitting and ERTR algorithms several times to improve the model.

2.4. Results analysis

2.4.1. R-square analysis

Known as the coefficient of determination, R-square is a statistical measure of fit that can describe the strength of the relationship between the input and output variables (for linear regression). In this research, R-square was adopted as a parameter to determine the goodness of fit of the multivariate regression model and the machine-learning model. As shown in the equation below, is the residual sum of squares and is the total sum of squares [44]. This calculation can indicate how much variation in the target dependent variable, TM, the original parameters in our prediction models predicted [45]. Ranging from 0 to 1, a higher R-square indicated better model fit. Our initial test showed that the R-squared values were generally approximately 80%–95%.

$$R^2 = 1 - \frac{RSS}{TSS} \tag{1}$$

In MRM, we calculated multiple R-square (normal R-square) and adjusted R-square, as there were several parameters. Adjusted R-square, modified from R-square, aimed to account for parameters with relatively low significance. When comparing multiple and adjusted R-square, a lower adjusted R-square indicated that the additional input parameters did not contribute to the multivariate regression model, while a higher adjusted R-square indicated that the additional input parameters could improve the model. The equation for adjusted R-squared is shown below. In the equation, is the number of parameters, and is the number of cases.

$$R^2\left(adj\right) = 1 - \frac{RSS/(n-p-1)}{TSS/(n-1)} \tag{2}$$

For the multivariate regression model, the R-squared value was calculated and presented using the linear modelling regression program's summary script. For the machine learning model, the SK-Learn metric method was used to calculate the R-squared value in the prediction evaluation program.

2.4.2. Analysis of statistical discrepancies for comparison

The statistical discrepancies between the experimental results and the hybrid model prediction results were evaluated in quantitative comparisons [46,47]. The TMs the experiments produced, the MRM predictions, and the MLM predictions were compared and analyzed, respectively, by calculating the parameters of predicted root mean square division (), normalized mean square error (), fractional bias (), and the correlation coefficient () (EM stands for "experiment" and "model"). These parameters were named and , and , and , and and for the experimental results analysis, MRM predictions, and MLM predictions in the following section [48,49]. The following equations show the specific calculations applicable to the parameters. In the equations below, is the number of measurement cases, is the temperature value of a case produced by the experiments, is the temperature value of a case produced by the hybrid model, is the mean temperature value of a series of cases produced by the experiments, and is the mean temperature value of a series of cases produced by the hybrid model [50].

$$PRMSD_{EM} = \frac{1}{E_m} \times \sqrt{\Sigma_{i=1}^{n} \frac{(E_i - M_i)^2}{n}} \tag{3}$$

$$NMSE_{EM} = \frac{\Sigma_{i=1}^{n}(E_i - M_i)^2}{\Sigma_{i=1}^{n}(E_i - M_i)} \tag{4}$$

$$FB_{EM} = \frac{E_m - M_m}{0.5 \times (E_m + M_m)} \tag{5}$$

$$CC_{EM} = \frac{\Sigma_{i=1}^{n}[(E_i - E_m) \times (M_i - M_m)]}{\sqrt{\Sigma_{i=1}^{n}(E_i - E_m)^2} \times \sqrt{\Sigma_{i=1}^{n}(M_i - M_m)^2}} \tag{6}$$

2.4.3. Complex MRM analysis

In the complex analysis of multivariate regression modelling, we visualized residuals versus fitted, normal quantile-quantile (normal Q-Q), scale location, and residuals versus leverage. The residuals versus fitted plots show whether the residuals had nonlinear patterns, the normal Q-Q plots show whether the residuals were normally distributed, the scale location plots show whether the residuals were spread equally along the predictor ranges, and the residuals versus leverage plots help to identify influential cases [51].

3. Results and evaluation

This section presents the results and evaluation of the multivariate regression model and the machine learning model.

3.1. Multivariate regression model results and evaluation

As explained in the methodology section (2.2.1 MRM data pre-processing), the heat pipe thermal diode system's two working conditions were considered in MRM to improve prediction accuracy. The first working condition was TES > TCS; the second working condition was TES < TCS.

3.1.1. Working condition: TES > TCS

When TES was larger than TCS, heat was conducted from the evaporation section to the condensation section. The model fit summary (Table 1) presents the coefficients, significance, multiple and adjusted R-square values, and *p*-value. The analysis shows that the TO, TES, TCA, and TEX parameters are highly significant in model fitting. The *p*-value was relatively low. Both the multiple and adjusted R-square reached 93.4% after several adjustments. They are relatively high, as they are close to 100%. This confirms the multivariate regression model's prediction accuracy. The model is shown in Fig. 7. According to the coefficients, the TM prediction formula can be expressed as follows:

$$TM = (0.48) \times TO + (0.91) \times TES + (-0.92) \times TCA + (0.46) \times TEX + 1.71 \tag{7}$$

Residuals versus fitted, normal Q-Q, scale location, and residuals versus leverage are presented in Fig. 8. The residuals versus fitted plot shows that there are some nonlinear patterns in the residuals. The normal Q-Q plot shows that the residuals follow a straight dashed line quite well. This suggests that the residuals are normally distributed. There is only a small deviation when the theoretical quantile passes 2.5 along the x-axis. The scale location plot shows that there is a horizontal

Table 1
A brief model fit summary (TES > TCS).

	Intercepts	TO	TES	TCA	TEX
Coefficients	1.71	0.48	0.91	−0.92	0.46
Significance	***	***	***	***	***
Multiple R-square	0.9343	Adjusted R-square		0.9342	
p-value	<2.2e-16				

Note Significance codes: 0–0.001: '***'; 0.001–0.01: '**'; 0.01–0.05: '*'; 0.05–0.1: '.'; 0.1–1: ''.

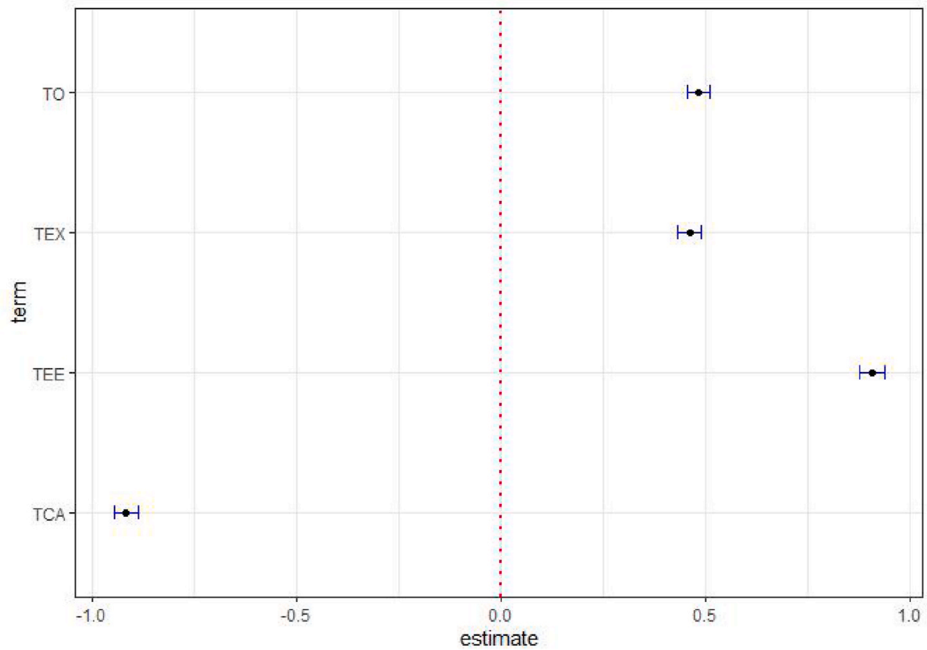

Fig. 7. Visualization of the multivariate regression model's coefficients (TES > TCS).

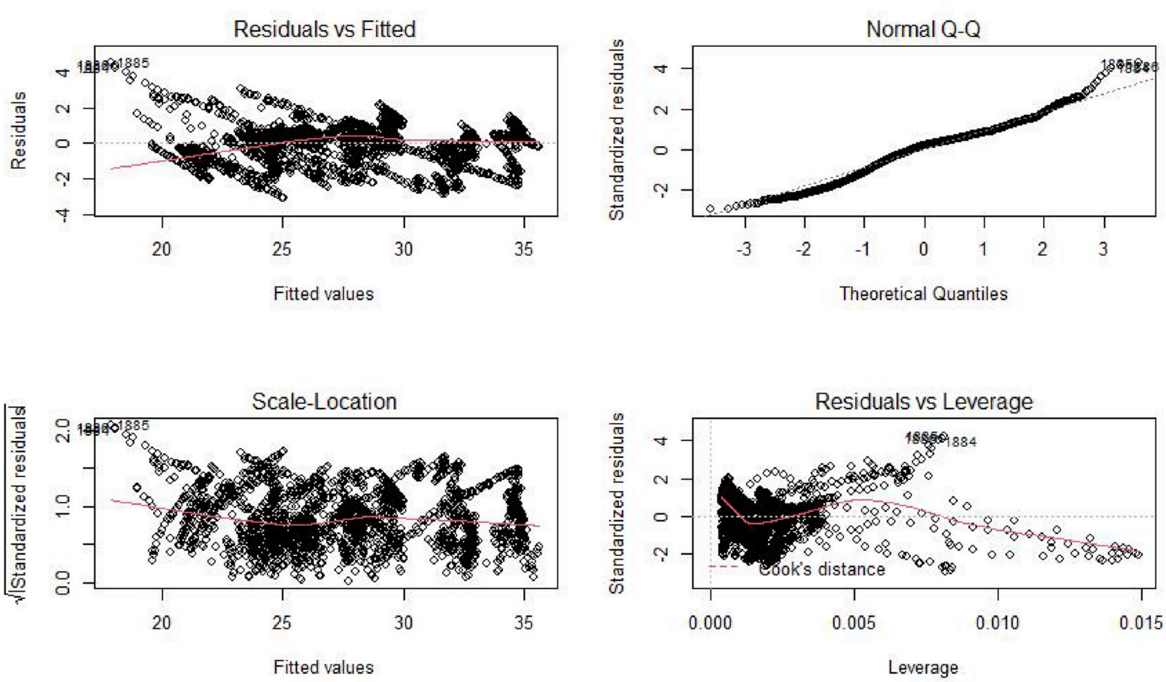

Fig. 8. Visualization of the MRM (TES > TCS).

line with randomly spread points. This indicates good homoscedasticity. The residuals versus leverage plot shows that there are a few outlying values at the right and lower right corners. This suggests that a few cases influence the regression line.

3.1.2. *Working condition: TES < TCS*

When TES was smaller than TCS, heat conduction from the evaporation section to the condensation section ceased. The model fit summary (Table 2) presents the coefficients, significance, multiple and adjusted R-square values, and *p*-value. The analysis shows that all the parameters are highly significant in the model fit, and the *p*-value is

Table 2
A brief model fit summary (TES < TCS).

	Intercepts	TO	TES	TCA	TEX
Coefficients	7.65	0.68	0.48	−1.34	0.95
Significance	***	***	***	***	***
Multiple R-square	0.9177	Adjusted R-square		0.9177	
p-value	<2.2e-16				

Note Significance codes: 0–0.001: '***'; 0.001–0.01: '**'; 0.01–0.05: '*'; 0.05–0.1: '.'; 0.1–1: ''.

quite low. Both the multiple and adjusted R-square reach 91.77% after adjustments. The results are relatively high, confirming the multivariate regression model's predictive accuracy. The model is shown in Fig. 9. According to the coefficients, the TM prediction formula can be expressed as follows:

$$TM = (0.68) \times TO + (0.48) \times TES + (-1.33) \times TCA + (0.95) \times TEX + 7.65 \quad (8)$$

Residuals versus fitted, normal Q-Q, scale location, and residuals versus leverage are presented in Fig. 10. There are some nonlinear patterns in the residuals, according to the residuals versus fitted plot. The normal Q-Q plot shows that the residuals are aligned well on a straight dashed line. This indicates a normal residuals distribution. A small deviation occurs only when the theoretical quantile passes 2.0 along the x-axis. The scale location plot shows a roughly horizontal line with randomly spread points, indicating good homoscedasticity. The residuals versus leverage plot shows that there are very few outlying values at the right corners. This suggests that very few cases influence the regression line.

3.1.2.1. Comparison of experimental and MRM TMs. As shown in Fig. 11, comparison of the TMs produced by the experiments (TM-EXP) and those the multivariate regression model predicted (TM-MRM) shows a relatively small deviation between the curves. The statistical discrepancy mentioned in the methodology section (2.4.2. Analysis of statistical discrepancies for comparison) was analyzed for the two working conditions. As presented in Table 3, the analysis results show that , , and are relatively close to 0, and is very close to 1. They satisfy the recommended standards very well. The analysis results further demonstrate that the TMs produced by the experiments and the multivariate regression model are in good agreement with each other under the two working conditions.

3.2. Machine learning model results and evaluation

After several rounds of adjustments, the R-square of the machine learning model reached 98.8009% when the relevant test size parameter was 0.1 and another random state parameter was 33. The R-square is relatively high, as it is close to 100%, which confirms the accuracy of the prediction.

The experimental results at 13 time points were compared with the results the machine-learning model produced. As presented in Fig. 12, comparison of TM-EXP and TM-MLM shows that the deviation between the points is relatively small. Statistical discrepancy analysis was performed. As presented in Table 4, the analysis results show that , , and are relatively close to 0, and is very close to 1. The results satisfy the recommended standards. The analysis further demonstrates that the TMs the experiments and machine-learning model produced are in good agreement with each other.

4. Discussion

In this section, the advantages of the hybrid model, its limitations, potential, and future research are discussed.

4.1. Advantages of the hybrid model

This research contributes to the original knowledge pertaining to thermal performance evaluation of built environments by presenting the development and analysis of the hybrid model. This demonstrates that the coordinated functions of multivariate regression and machine learning can provide efficient and accurate predictions of interior temperatures influenced by solar cavities and heat pipe thermal diodes. A hybrid model can be a practical alternative in relevant studies, especially for the comparison of iterative indoor environment investigations.

In typical heat conduction involving homogenous materials, thermal conductivity is usually a coefficient for calculating heat flux along a particular direction. The MRM-generated linear multivariate formulas are similar to a simplified model representing the complexity of heat conduction. Although simplification might influence the accuracy of the prediction results, utilization of linear multivariate formulas with several coefficients is relatively reasonable because of their similarity to the heat conduction calculation. The temperatures of different parts of the exterior side of the wall are used as independent variables to predict the dependent variable of interior temperature, as a correlation between the variables was established. The simplification can make the prediction fast and convenient for manual calculations.

Machine learning technology has been adopted in hybrid model development because of its suitability for nonlinear prediction. The influence the solar cavity and heat pipe thermal diode exert on the interior temperature is complicated, causing irregular interior temperature changes. The machine learning model has an advantage in terms of predicting nonlinear temperature curves. In MLM, the model is developed by learning from experience based on particular algorithms and a

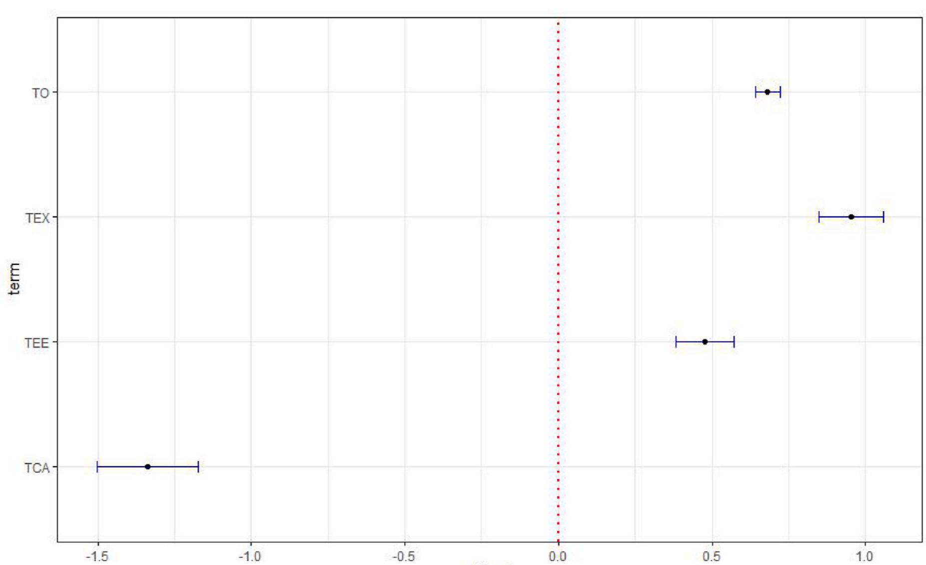

Fig. 9. Visualization of the multivariate regression model's coefficients (TES < TCS).

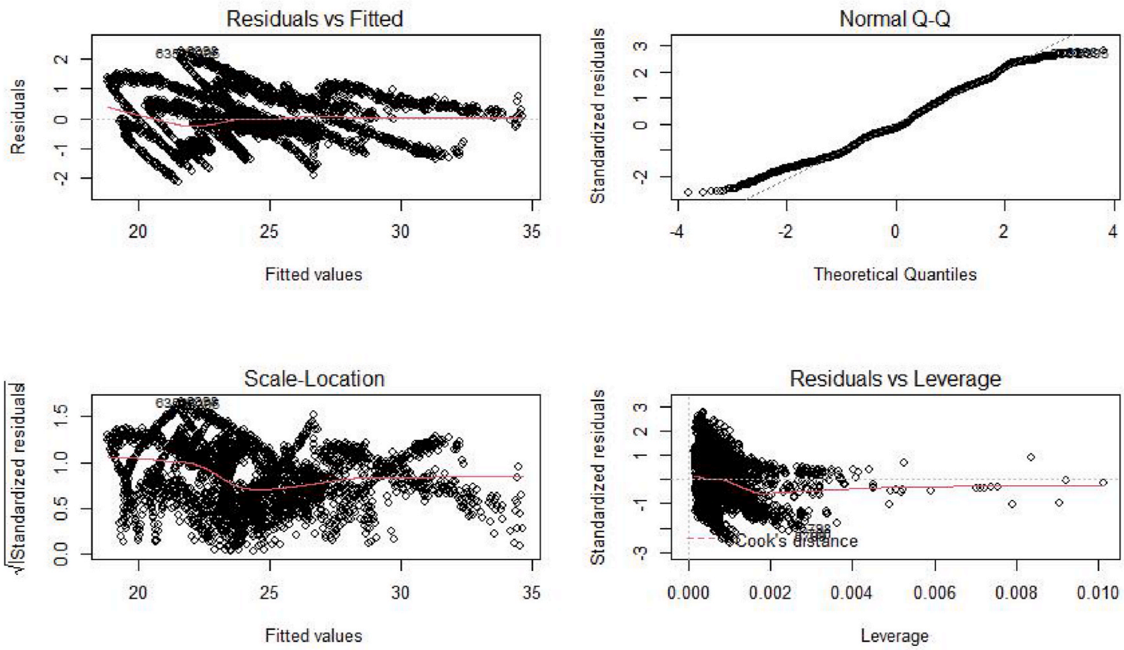

Fig. 10. Visualization of MRM (TES < TCS).

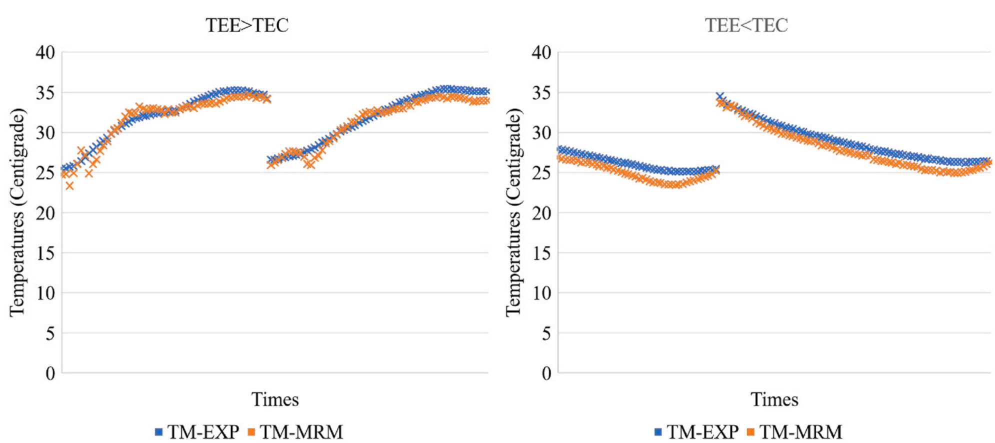

Fig. 11. Comparison of experimental and predictive MRM TMs.

Table 3
Results for $PRMSD_{EMRM}$, $NMSE_{EMRM}$, FB_{EMRM}, and CC_{EMRM}.

	$PRMSD_{EMRM}$	$NMSE_{EMRM}$	FB_{EMRM}	CC_{EMRM}
TES > TCS	0.034995	0.001204	0.00323	0.972104
TES < TCS	0.03834	0.001513	0.036807	0.995171

Note. $PRMSD_{EMRM}$ - predicted root mean square division, $NMSE_{EMRM}$ - normalized mean square error, FB_{EMRM} - fractional bias, CC_{EMRM} - correlation coefficient, EMRM - experiment and multivariate regression model.

large amount of experimental data. Therefore, the machine learning model's prediction is relatively accurate.

Our test showed that predictions can be made very rapidly, and the deviation between the prediction results and experimental data is considered to be more than acceptable. In the R-square analysis, the R-square values reached 93.4% and 91.8% for the multivariate regression model with two working conditions; the R-square value can be as high as 98.8% for the machine learning model. These high R-squared values suggest that the hybrid model can generate relatively accurate predictions. In the statistical discrepancy analysis, the calculations for predicted root mean square division, normalized mean square error, fractional bias, and the correlation coefficient proved that the data the experiments and hybrid model produced were in good agreement. In the complex MRM analysis, the visualization of residuals versus fitted, normal Q-Q, scale location, and residuals versus leverage further demonstrates the efficiency and accuracy of the multivariate regression model.

4.2. Limitations

Although the hybrid model's effectiveness has been proven, the methodology cannot replace conventional physical mathematical models that directly describe the heat conduction process. The hybrid model was developed using MRM and MLM through statistical analysis based on experimental data. This approach should be different from the development of physical mathematical models. The development of a hybrid model is a preliminary attempt to rapidly predict built environment parameters. Many efforts have been made to demonstrate the accuracy and efficiency of this novel methodology. These are key issues related to the effectiveness of rapid prediction. The primary experiment

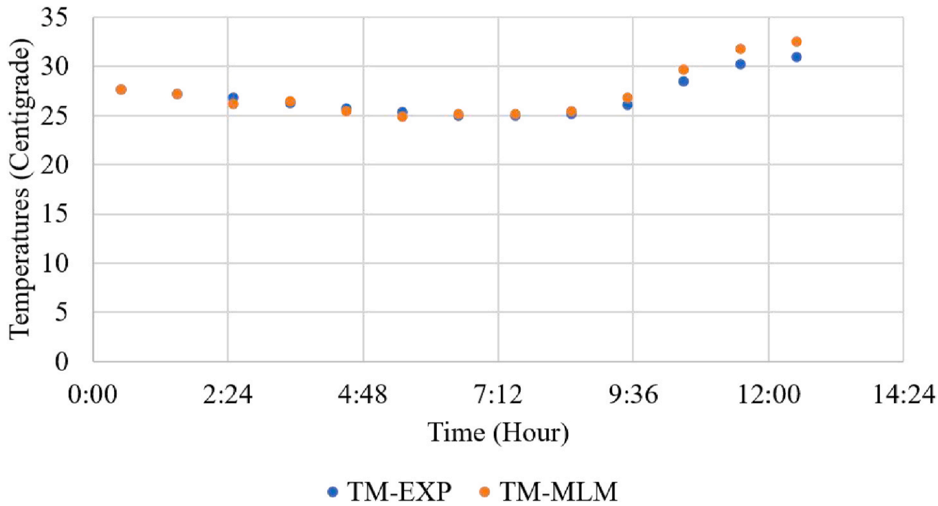

Fig. 12. Comparison of experimental and predictive MLM TMs.

Table 4
Results for $PRMSD_{EMLM}$, $NMSE_{EMLM}$, FB_{EMLM}, and CC_{EMLM}.

$PRMSD_{EMLM}$	$NMSE_{EMLM}$	FB_{EMLM}	CC_{EMLM}
0.028878	0.000819	−0.01229	0.983636

Note. $PRMSD_{EMLM}$ - predicted root mean square division, $NMSE_{EMLM}$ -normalized mean square error, FB_{EMLM} - fractional bias, CC_{EMLM} - correlation coefficient, EMLM - experiment and machine learning model.

provides initial data for the development and validation of the hybrid model at this stage.

Currently, the hybrid model is tailored for predicting the interior temperature of an experimental house with the particular configuration of a solar cavity and a heat pipe thermal diode system. Only some temperature values were used as independent variables. The other parameters related to the configuration of architectural components and building geometries were not considered because of the experimental facilities' limitations. Our methodological framework can be universally applied to the development of new prediction models. For new predictions considering more factors (such as the number of heat pipe units, the size of the solar cavity, and the geometry of the indoor space), new experiments are needed to provide initial data for MRM and MLM.

More experimental data can also help in multivariate regression and machine learning model training to further improve prediction accuracy. In this study, experimental data for only 11 days in May were used for MRM and MLM. The machine learning model's R-squared value is higher than that of the multivariate regression model. The TMs the machine learning model produced are closer to the experimental results than those the multivariate regression model produced. This suggests that the machine learning model's prediction should be more accurate than that of the multivariate regression model. Although our hybrid model's accuracy has been demonstrated in the analysis, the deviation between the predictions and experimental results sometimes reaches approximately 1 cm. This deviation might be relatively large in some cases. This is because heating demand, cooling load, and energy consumption are sensitive to temperature values. The deviation would ideally be under 0.5 °C in those cases. Although tens of thousands of experimental data have been collected in this study, MRM and MLM demand enormous amounts of data over a wide temperature range. More experiments can be implemented in the summer, winter, and transition seasons, or even over an entire year.

4.3. Model potential and future research

The hybrid model and methodological framework have great potential to rapidly predict other sustainable design analyses. They can seamlessly be employed to evaluate different built environment parameters, including the temperatures of different positions in buildings, humidity, predicted mean vote parameters, and the heat transfer coefficient. More experimental data can be used in MRM and MLM.

The hybrid model is flexible and accessible. The machine learning model can make numerous predictions with relatively high accuracy using a computer. The machine learning model program can immediately produce multiple results with the simple entry of several groups of initial parameters. Manual calculations are even available using the multivariate regression model. Interior temperature can be calculated quickly, as multivariate linear formulas are very easy to solve.

The hybrid model can predict interior temperature considering the complex influences the solar cavity and heat pipe system exert. For relevant research, the system's thermal performance can be easily estimated using the hybrid model. Estimation was completed without additional simulations or experiments. The model can provide useful functions, such as the quick verification of results and help adjusting settings for further simulations and experiments. In architectural practice, the hybrid model can immediately analyze indoor thermal performance. The passive design strategy using a solar cavity and heat pipe thermal diode system can be evaluated in the early stages. The hybrid model can provide useful information for sustainable design optimization.

The model and its framework can facilitate the study of more relevant parameters and specific building cases. A large number of design cases can be considered in MRM and MLM, especially for typical building configurations. As mentioned in the methodology section (2. Hybrid model development), algorithms were developed for MRM and MLM. Computational scripts were written to formulate the programs' functional modules. Functional modules and specific scripts can be easily replaced or renewed to create better alternatives or suit new demands. Novel algorithms and computational tools with better performance and new functions will be developed in the future. With iterative improvements, future development of a design optimization tool to consider more environmental and architectural factors is possible. This would make the analysis more accurate and practical for the improvement of indoor thermal environments. The design was optimized by adjusting the configuration of the architectural components to adapt to external environments.

5. Conclusion

This research successfully developed an MRM–MLM hybrid model to rapidly predict interior temperatures affected by heat pipe thermal diodes and solar cavities. This paper presents the development of such a hybrid model and demonstrates its efficiency and accuracy. The model can accurately predict TMs in a very short time. Although the research focuses on a particular thermal parameter and building configuration, the hybrid model and its framework are generally applicable to relevant research and design. As mentioned in Section 4, the hybrid model provides useful information for experimental adjustment, simulation estimation, and design optimization in the early stages.

Experimental houses were built, and numerous environmental parameters were measured and collected from a house fitted with a solar cavity and a heat pipe thermal diode system. Based on the experiment, MLM and MRM featured in the development of the hybrid model. In MRM, multivariate regression was performed after data preprocessing. Linear multivariate formulas were generated considering the heat pipe thermal diode's two working conditions. In MLM, the data were first imported, cleaned, and split; then, the machine learning model was created and trained using the data. The model was evaluated and improved after predictions. The prediction results can be easily obtained via manual calculations or by running the computation program according to different situations. The results were fully evaluated through comprehensive analyses of R-square, predicted root mean square division, normalized mean square error, fractional bias, the correlation coefficient, residuals versus fitted, normal Q-Q, scale location, and residuals versus leverage. Thus, prediction accuracy has been demonstrated from several aspects.

The combination of the two emerging computational methods of MRM and MLM with different technical principles in the development of the hybrid model is an original attempt. The model can provide an efficient iterative analysis of the interior temperatures of an experimental house with a solar cavity and a heat pipe thermal diode system. The hybrid model can make new predictions pertaining to other parameters if it is fed with other initial experimental data. Given that an immediate comparison study can offer helpful information for design, simulation, and experimental optimization and adjustment, architects and researchers can use the model to improve their designs and studies in the early stage. This research has demonstrated the hybrid model's great potential to combine multiple computational methods in sustainable and environmental studies.

CRediT authorship contribution statement

Yi He: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Fangcheng Kou:** Data curation. **Xin Wang:** Resources. **Ning Zhu:** Resources. **Yehao Song:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Yingnan Chu:** Data curation. **Shaohang Shi:** Data curation. **Mengjia Liu:** Data curation. **Xinxing Chen:** Data curation.

Declaration of competing interest

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

Acknowledgements

This research was financially supported by the Natural Science Foundation of China (Grant No. 52078264) and the Key Research and Development Program of the Hebei Province of China (Grant No. 20374507D).

References

[1] S.M. Amin Hosseini, Reza Yazdani, Albert de la Fuente, Multi-objective interior design optimization method based on sustainability concepts for post-disaster temporary housing units, Build. Environ. 173 (2020) 106742, https://doi.org/10.1016/j.buildenv.2020.106742. ISSN 0360-1323.

[2] Szabolcs Varga, Armando C. Oliveira, Clito F. Afonso, Characterisation of thermal diode panels for use in the cooling season in buildings, Energy Build. 34 (Issue 3) (2002) 227–235, https://doi.org/10.1016/S0378-7788(01)00090-1, 0378-7788.

[3] Z.Y. Wang, Y.H. Diao, L. Liang, Y.H. Zhao, T.T. Zhu, F.W. Bai, Experimental study on an integrated collector storage solar air heater based on flat micro-heat pipe arrays, Energy Build. 152 (2017) 615–628, https://doi.org/10.1016/j.enbuild.2017.07.069, 0378-7788.

[4] Sheng Du, Quan Zhang, Peilin Hou, Yue Chang, Sikai Zou, Experimental study and steady-state model of a novel plate loop heat pipe without compensation chamber for CPU cooling, Sustain. Cities Soc. 53 (2020) 101894, https://doi.org/10.1016/j.scs.2019.101894, 2210-6707.

[5] Lisandra Fachinello Krebs, Erik Johansson, Influence of microclimate on the effect of green roofs in Southern Brazil – a study coupling outdoor and indoor thermal simulations, Energy Build. 241 (2021) 110963, https://doi.org/10.1016/j.enbuild.2021.110963, 0378-7788.

[6] Thanyalak Srisamranrungruang, Kyosuke Hiyama, Balancing of natural ventilation, daylight, thermal effect for a building with double-skin perforated facade (DSPF), Energy Build. 210 (2020) 109765, https://doi.org/10.1016/j.enbuild.2020.109765, 0378-7788.

[7] Aleksandar Jankovic, Francesco Goia, Impact of double skin facade constructional features on heat transfer and fluid dynamic behaviour, Build. Environ. 196 (2021) 107796, https://doi.org/10.1016/j.buildenv.2021.107796, 0360-1323.

[8] Shuai Yan, Xianting Li, Baolong Wang, Wenxing Shi, Weihua Lyu, A method to describe the thermal property of pipe-embedded double-skin façade: equivalent glass window, Energy Build. 195 (2019) 33–44, https://doi.org/10.1016/j.enbuild.2019.04.041, 0378-7788.

[9] Patrik Nemec, Čaja Alexander, Milan Malcho, Mathematical model for heat transfer limitations of heat pipe, Math. Comput. Model. 57 (Issues 1–2) (2013) 126–136, https://doi.org/10.1016/j.mcm.2011.06.047, 0895-7177.

[10] A. Shafieian, M. Khiadani, A. Nosrati, A review of latest developments, progress, and applications of heat pipe solar collectors, Renew. Sustain. Energy Rev. 95 (2018) 273–304, https://doi.org/10.1016/j.rser.2018.07.014, 1364-0321.

[11] Xiangdong Liu, Yongping Chen, Fluid flow and heat transfer in flat-plate oscillating heat pipe, Energy Build. 75 (2014) 29–42, https://doi.org/10.1016/j.enbuild.2014.01.041, 0378-7788.

[12] Hongli Sun, Borong Lin, Zhirong Lin, Yingxin Zhu, Hui Li, Xiaoying Wu, Research on a radiant heating terminal integrated with a thermoelectric unit and flat heat pipe, Energy Build. 172 (2018) 209–220, https://doi.org/10.1016/j.enbuild.2018.04.054, 0378-7788.

[13] Hongli Sun, Yifan Wu, Borong Lin, Mengfan Duan, Zhirong Lin, Hui Li, Experimental investigation on the thermal performance of a novel radiant heating and cooling terminal integrated with a flat heat pipe, Energy Build. 208 (2020) 109646, https://doi.org/10.1016/j.enbuild.2019.109646, 0378-7788.

[14] Yan Wang, Yingning Hu, Wenhao Zhao, Wenqi Qin, The performance of a closed cavity radiation system with built-in heat exchanger, Build. Environ. 174 (2020) 106788, https://doi.org/10.1016/j.buildenv.2020.106788, 0360-1323.

[15] Gediyon M. Girma, Fitsum Tariku, Experimental investigation of cavity air gap depth for enhanced thermal performance of ventilated rain-screen walls, Build. Environ. 194 (2021) 107710, https://doi.org/10.1016/j.buildenv.2021.107710, 0360-1323.

[16] Chuyao Wang, Niansi Li, Bendong Yu, Md Muin Uddin, Jie Ji, A novel solar spectrum-splitting utilization photocatalytic CdTe double-skin façade: concept, design and performance investigation, Build. Environ. 195 (2021) 107776, https://doi.org/10.1016/j.buildenv.2021.107776, 0360-1323.

[17] B. Nghana, F. Tariku, G. Bitsuamlak, Assessing ventilation cavity design impact on the energy performance of rainscreen wall assemblies: a CFD study, Build. Environ. 196 (2021) 107789, https://doi.org/10.1016/j.buildenv.2021.107789, 0360-1323.

[18] Ivens Portugal, Paulo Alencar, Donald Cowan, The use of machine learning algorithms in recommender systems: a systematic review, Expert Syst. Appl. 97 (2018) 205–227, https://doi.org/10.1016/j.eswa.2017.12.020, 0957-4174.

[19] J. Jie Wu, Y. Zhao, Machine learning technology in the application of genome analysis: a systematic review, Gene 705 (2019) 149–156, https://doi.org/10.1016/j.gene.2019.04.062, 0378-1119.

[20] T.P. Carvalho, F.A.A.M.N. Soares, R. Vita, R. da P. Francisco, J.P. Basto, S.G. S. Alcalá, A systematic literature review of machine learning methods applied to predictive maintenance, Comput. Ind. Eng. 137 (2019) 106024, https://doi.org/10.1016/j.cie.2019.106024, 0360-8352.

[21] W. Kobayashi, T. Shimura, A. Mitsuishi, K. Iwamoto, A. Murata, Prediction of the drag reduction effect of pulsating pipe flow based on machine learning, Int. J. Heat Fluid Flow 88 (2021) 108783, https://doi.org/10.1016/j.ijheatfluidflow.2021.108783, 0142-727X.

[23] Xianling Wang, Liang Luo, Jinwei Xiang, Senlin Zheng, Samson Shittu, Zhangyuan Wang, Xudong Zhao, A comprehensive review on the application of nanofluid in heat pipe based on the machine learning: theory, application and prediction, Renew. Sustain. Energy Rev. 150 (2021) 111434, https://doi.org/10.1016/j.rser.2021.111434, 1364-0321.

[24] Zhangyuan Wang, Xudong Zhao, Zhonghe Han, Liang Luo, Jinwei Xiang, Senglin Zheng, Guangming Liu, Min Yu, Yu Cui, Samson Shittu, Menglong Hu, Advanced big-data/machine-learning techniques for optimization and performance

enhancement of the heat pipe technology – a review and prospective study, Appl. Energy 294 (2021) 116969, https://doi.org/10.1016/j.apenergy.2021.116969, 0306-2619.

[25] He Yi, Xiao-Hui Liu, Hong-Liang Zhang, Wei Zheng, Fu-Yun Zhao, Marc Aurel Schnabel, Yi Mei, Hybrid framework for rapid evaluation of wind environment around buildings through parametric design, CFD simulation, image processing and machine learning, Sustain. Cities Soc. 73 (2021) 103092, https://doi.org/10.1016/j.scs.2021.103092, 2210-6707.

[26] Magdalena Hajdukiewicz, Marco Geron, M. Marcus, Keane, Calibrated CFD simulation to evaluate thermal comfort in a highly-glazed naturally ventilated room, Build. Environ. 70 (2013) 73–89, https://doi.org/10.1016/j.buildenv.2013.08.020, 0360-1323.

[27] He Yi, Marc Aurel Schnabel, Yi Mei, A novel methodology for architectural wind environment study by integrating CFD simulation, multiple parametric tools and evaluation criteria, Build. Simul. 13 (3) (2019).

[28] Bie Yu, Ming Li, Fei Chen, Grzegorz Królczyk, Zhixiong Li, Heat transfer mathematical model for a novel parabolic trough solar collecting system with V-shaped cavity absorber, Sustain. Cities Soc. 52 (2020) 101837, https://doi.org/10.1016/j.scs.2019.101837, 2210-6707.

[29] Qian Su, Shinan Chang, Chen Yang, Loop heat pipe-based solar thermal façade water heating system: a review of performance evaluation and enhancement, Sol. Energy 226 (2021) 319–347, https://doi.org/10.1016/j.solener.2021.08.019, 0038-092X.

[30] M.Y. Wong, C.Y. Tso, T.C. Ho, H.H. Lee, A review of state of the art thermal diodes and their potential applications, Int. J. Heat Mass Tran. 164 (2021) 120607, https://doi.org/10.1016/j.ijheatmasstransfer.2020.120607, 0017-9310.

[31] Tianhu Zhang, Yuanjun Liu, Yandi Rao, Xiaopeng Li, Qingxin Zhao, Optimal design of building environment with hybrid genetic algorithm, artificial neural network, multivariate regression analysis and fuzzy logic controller, Build. Environ. 175 (2020) 106810, https://doi.org/10.1016/j.buildenv.2020.106810, 0360-1323.

[32] Daniel Wallach, David Makowski, James W. Jones, François Brun, Chapter 2 - the R programming language and software, in: Daniel Wallach, David Makowski, James W. Jones, François Brun (Eds.), Working with Dynamic Crop Models, third ed., Academic Press, 2019, pp. 45–95, https://doi.org/10.1016/B978-0-12-811756-9.00002-2, 9780128117569.

[33] Efthymia Nikita, Chapter 7.2 - an introduction to the R language, in: Zuzana Obertová, Alistair Stewart, Cristina Cattaneo (Eds.), Statistics and Probability in Forensic Anthropology, Academic Press, 2020, pp. 353–372, https://doi.org/10.1016/B978-0-12-815764-0.00028-9, 9780128157640.

[34] Ch. Zisi, A. Pappa-Louisi, P. Nikitas, Separation optimization in HPLC analysis implemented in R programming language, J. Chromatogr. A 1617 (2020) 460823, 0021-9673.

[35] Ibrahim Dincer, Dogan Erdemir, Chapter 4 - system Analysis, in: Ibrahim Dincer, Dogan Erdemir (Eds.), Heat Storage Systems for Buildings, Elsevier, 2021, pp. 115–178, https://doi.org/10.1016/B978-0-12-823572-0.00003-5, 9780128235720.

[36] G. Hanrahan, F. Udeh, D.G. Patil, CHEMOMETRICS and statistics | multivariate calibration techniques, in: Paul Worsfold, Alan Townshend, Colin Poole (Eds.), Encyclopedia of Analytical Science, second ed., Elsevier, 2005, pp. 27–32, https://doi.org/10.1016/B0-12-369397-7/00077-7, 9780123693976.

[37] Changzhi Dai, Hui Zhang, Edward Arens, Zhiwei Lian, Machine learning approaches to predict thermal demands using skin temperatures: steady-state conditions, Build. Environ. 114 (2017) 1–10, https://doi.org/10.1016/j.buildenv.2016.12.005, 0360-1323, 2017.

[38] Souad Amjad, Abdelhalim Abdelbaki, Zaki Zrikem, Transfer functions method and sub-structuration technique for two-dimensional heat conduction problems in high thermal mass systems: application to ground coupling problems, Energy Build. 35 (Issue 6) (2003) 593–604, https://doi.org/10.1016/S0378-7788(02)00169-X, 0378-7788.

[39] Fernando Branco, António Tadeu, Nuno Simões, Heat conduction across double brick walls via BEM, Build. Environ. 39 (Issue 1) (2004) 51–58, https://doi.org/10.1016/j.buildenv.2003.08.005, 0360-1323.

[40] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach. Learn. 63 (1) (2006) 3–42.

[41] Claudia Lindner, Chapter 1 - automated image interpretation using statistical shape models, in: Guoyan Zheng, Shuo Li, Gabor Székely (Eds.), Statistical Shape and Deformation Analysis, Academic Press, 2017, pp. 3–32, https://doi.org/10.1016/B978-0-12-810493-4.00002-X, 9780128104934.

[42] Siddharth Misra, Jiabo He, Chapter 4 - stacked neural network architecture to model the multifrequency conductivity/permittivity responses of subsurface shale formations, in: Siddharth Misra, Hao Li, Jiabo He (Eds.), Machine Learning for Subsurface Characterization, Gulf Professional Publishing, 2020, 2020, pp. 103–127, https://doi.org/10.1016/B978-0-12-817736-5.00004-1, 9780128177365.

[43] S.A. Glantz, B.K. Slinker, Primer of Applied Regression and Analysis of Variance, McGraw-Hill, 1990, 978-0-07-023407-9.

[44] N.R. Draper, H. Smith, Applied Regression Analysis, Wiley-Interscience, 1998, 978-0-471-17082-2.

[45] D. Wilks, in: Statistical Methods in the Atmospheric Sciences, second ed., Academic Press, Cambridge, 2006.

[46] J.C. Chang, S.R. Hanna, Air quality model performance evaluation, Meteorol. Atmos. Phys. 87 (1–3) (2004) 167–196.

[47] ANSI/ASHRAE Standard 140-2001, Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Atlanta, GA, 2001.

[48] S J.M. Chambers, Linear models, in: J.M. Chambers, T.J. Hastie (Eds.), Chapter 4 of Statistical Models, Wadsworth & Brooks/Cole, 1992.

[49] G.N. Wilkinson, C.E. Rogers, Symbolic descriptions of factorial models for analysis of variance, Appl. Stat. 22 (1973) 392–399, https://doi.org/10.2307/2346786.

[50] W.B. Frtcher, Measured and Predicted Root-Mean-Square Errors in Square and Triangular Antenna Mesh Facets, 1989. NASA TP 2896.

[51] Anil K. Bera, Carlos M. Jarque, Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo Evidence, Econ. Lett. 7 (Issue 4) (1981) 313–318, https://doi.org/10.1016/0165-1765(81)90035-5, 0165-1765.