Journal of Building Engineering 99 (2025) 111595

Contents lists available at ScienceDirect

Journal of Building Engineering

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

A novel approach to calculate the mean thermal sensation vote for primary and secondary schools using Bayesian inference

Sen Miao *, Marta Gangolells, Blanca Tejedor

Department of Project and Construction Engineering, Group of Construction Research and Innovation (GRIC), Universitat Politècnica de Catalunya, C/ Colom, 11, Ed. TR5, 08222, Terrassa (Barcelona), Spain

ARTICLE INFO

Keywords:
Thermal comfort
Field survey
Educational building
Mean thermal sensation vote
Bayesian inference

ABSTRACT

Existing thermal comfort models defined in relevant standards are often found to be less effective for primary and secondary school students in educational buildings. This is often thought to be primarily due to differences in thermal sensation between children and adults. However, one important factor that is often neglected is the uncertainty associated with thermal comfort survey data. The existing method for calculating the mean thermal sensation vote is oversimplified and does not properly address related uncertainties. As a result, it ultimately affects the performance of the developed thermal comfort models. Hence, this research proposes a novel approach to compute the mean thermal sensation vote data for primary and secondary schools using Bayesian inference. This approach addresses the error caused by the uncertainties associated with the collected thermal sensation vote data in order to improve the effectiveness of the developed thermal comfort models for students. The proposed method was validated through a holistic case study using five thermal comfort models. The results showed that the accuracy of the developed thermal comfort models improved by 10.1 %–30.9 %, and the R^2 improved by 5.3 %–28.8 %. A benchmark for the Bayesian model parameter setting was proposed as the reference for relevant studies. Finally, an open, user-friendly software was developed and is available to relevant users to implement the proposed approach more efficiently. The results of this research have practical implications for the development and optimization of thermal comfort models for students.

1. Introduction

Thermal comfort is one of the most important aspects of indoor environmental quality. The main goal of educational centers is to provide a conducive learning environment for students [1]. Thermal discomfort compromises students' well-being and affects their attention, cognitive abilities, productivity and academic performance [2,3]. The thermal comfort model plays a critical role in the design and control of indoor thermal environments. Considering that heating, ventilation, and air conditioning (HVAC) systems consume up to 40 % of the energy in building operation, an effective thermal comfort model has practical implications for optimizing the building energy efficiency, energy management systems, and power grid operation, which contributes to the energy conservation while responding to the comfort demands of the building occupants [4]; [5–7]. One of the most important and most widely used thermal comfort models is the predicted mean vote (PMV) index, which is adopted in existing thermal comfort standards such as ISO 7730 and ASHRAE 55 [8]. This model predicts the mean thermal sensation of occupants in a space based on indoor thermal

* Corresponding author.
E-mail address: sen.miao@upc.edu (S. Miao).

https://doi.org/10.1016/j.jobe.2024.111595
Received 28 June 2024; Received in revised form 25 November 2024; Accepted 14 December 2024
Available online 18 December 2024

parameters, their clothing and metabolic level. As PMV did not consider the adaptation of occupants in free-running buildings, adaptive models were proposed subsequently and were widely applied in related research. These included adaptive PMV and adaptive regression models [8]; [9]. However, relevant studies found that existing thermal comfort models perform well for adults but are less effective for primary and secondary school students [10]; [1,9]. The reason is generally believed to be that these models were developed using adults as subjects, and children have different thermal sensations due to differences in their physiological conditions and adaptability [10]. Thereby, it is often crucial to conduct field surveys and obtain thermal sensation vote (TSV) data in schools, to validate and develop thermal comfort models for these students.

Previous research by Miao et al. [11] reviewed thermal comfort field studies conducted in primary and secondary schools worldwide over the past 20 years and found considerable variation in the reported accuracy and coefficient of determination (R^2) of validated and developed thermal comfort models across cases. These studies often attribute the results primarily to the distinctive features of the surveyed student samples, such as climate, age and adaptability. However, an almost completely overlooked fact is that, as indicated by Chang et al. [12], the quality of the data has a significant impact on the thermal comfort models that are developed.

Existing studies all adopted a very simple and straightforward approach to calculate the mean thermal sensation vote (MTSV), which is the mean value of all TSVs by students in the classroom in a field survey. The MTSV data obtained by such an oversimplified method actually have very large uncertainties. Firstly, existing studies did not separate students by gender in MTSV calculations. However, there is a difference in thermal sensation between female and male students [13]; [14]. In field surveys, the ratio of male and female students in the classrooms is impossible to control, which leads to uncertainty due to different gender ratios across the classrooms. Secondly, the effect of clothing insulation was not considered in the existing MTSV calculation approach. For classrooms with similar indoor thermal conditions, the difference in students' clothing may lead to a perceived variation in the MTSV that is obtained. Thirdly, in real-world field investigations, the number of students varies across the classrooms, and is often limited by room capacity. The MTSV obtained from a classroom with more students could be more accurate than that obtained from one with fewer

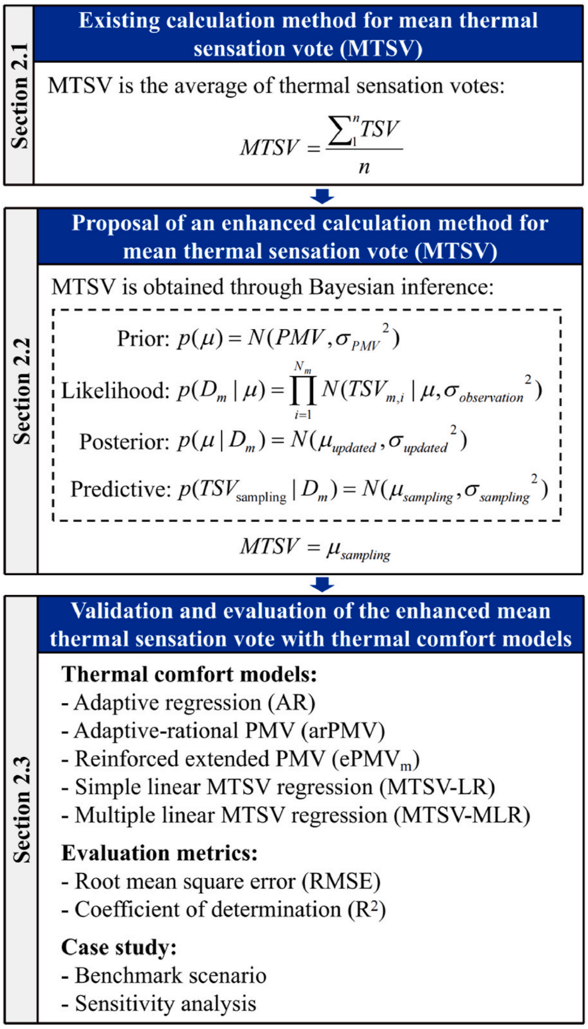

Fig. 1. Brief summary of the methodology of this study.

students. Thus, the uncertainty could be caused by the difference in the sample size of surveyed classrooms. Such uncertainty is particularly associated with the field surveys in primary and secondary schools, because unlike university classrooms that can accommodate more than 100 students, classrooms in primary and secondary schools often accommodate around 20 students or even fewer. Lastly and most importantly, the differences in the ability of students of different ages to perceive and express their thermal sensations could lead to uncertainty. Compared to adults, children have certain difficulties in understanding and expressing their actual thermal sensations [15,16]. This means that the TSV collected from secondary school students could reflect their actual feelings more accurately than those gathered from primary school students. All these uncertainties collectively affect the quality of the calculated MTSV data and ultimately determine the performance of the thermal comfort models that are developed.

Conducting thermal comfort surveys in schools is often a challenging task. It not only requires substantial time, effort and financial cost, but also usually needs the understanding and support of relevant stakeholders as it disturbs the teaching activities. In field studies, the number of surveys that can be carried out is often very limited, which makes the MTSV data that are obtained quite valuable. In this context, developing a method that can address the aforementioned uncertainties to improve the quality of the collected MTSV data is of great significance and contributes to the research on such topics. Bayesian inference is a powerful statistical approach for dealing with uncertainty. In recent years, this technique has been gradually applied to develop thermal comfort models in relevant research [17–19]. However, to the author's knowledge, to date no study has investigated the use of Bayesian inference as an enhancement tool for thermal comfort survey data.

Therefore, this research proposes a novel approach to calculate the mean thermal sensation vote (MTSV) for the field thermal comfort survey in primary and secondary schools, where the uncertainties associated with the survey data are addressed with Bayesian inference. Based on the proposed method, the effect of enhanced MTSV data on thermal comfort model development was evaluated with a case study. A practical software application was also developed to further help relevant stakeholders in the scientific community. The proposed method and the developed software can contribute to the development and optimization of effective thermal comfort models for secondary and primary school students, thereby optimizing the energy use and operational management of educational buildings, while ensuring a comfortable learning environment for students.

Following this introduction, Section 2 describes the proposed methodology, Section 3 presents and discusses the case study results, Section 4 demonstrates and explains the software that was developed, and Section 5 summarizes the conclusions and recommendations.

2. Methodology

This section introduces the methodology of this study step by step. Fig. 1 presents a brief summary.

2.1. Existing calculation method for mean thermal sensation vote

Thermal comfort field surveys in educational buildings are usually hierarchical because repeated surveys are conducted in multiple schools and classrooms in the same geographical location, covering students of different ages. The field survey contains the objective measurement of the indoor thermal condition of the classrooms and the subjective survey of the thermal sensation of the students.

As defined in existing thermal comfort standards such as ISO 7730 [20] and ASHRAE 55 [21], the objective thermal parameters mainly include indoor air temperature (T_a), mean radiant temperature (T_{mrt}), relative humidity (RH), air velocity (V_a) and outdoor air temperature (T_{out}). The subjective survey usually requires students to remain in a sedentary state (1.2 met) for at least 30 min and uses the seven-point thermal sensation scale (Fig. 2) to collect their thermal sensation vote (TSV), based on which the mean thermal sensation vote (MTSV) is obtained. It is also necessary to record the students' clothing and calculate the clothing insulation value (I_{clo}) with reference to the aforementioned thermal comfort standard. The collected data are used to validate and develop thermal comfort models for the students, such as MTSV regression, adaptive regression and an adaptive PMV model.

Equation (1) shows the MTSV calculation approach used in existing studies:

$$MTSV_i = \frac{\sum_1^n TSV}{n} \tag{1}$$

Where i is the number ith field survey and n is the number of students in the classroom in the ith survey. The total amount of MTSV data depends on the number of field surveys conducted.

The existing MTSV calculation method relies entirely on collected data and assumes that the observed data exactly reflect reality. As explained earlier, such a method has the obvious limitation that it does not properly address the uncertainties associated with the collected data. This ultimately affects the performance of the thermal comfort models that are developed.

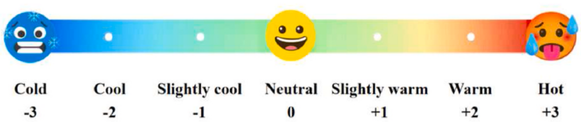

Fig. 2. Seven-point thermal sensation scale.

2.2. Proposal of an enhanced calculation method for mean thermal sensation vote

The proposed method for obtaining MTSV is based on Bayesian inference, which combines both prior knowledge (or beliefs) and observed evidence to estimate the final outcomes [22]. The fundamental principle can be briefly explained as follows.

For a thermal comfort survey conducted in a classroom:

- **Step 1:** based on existing information and knowledge, the thermal sensation of the students is assumed, given the indoor thermal condition of the classroom.
- **Step 2:** the thermal sensation data collected from the students are used as observed evidence to update the assumption.
- **Step 3:** the mean thermal sensation of students is estimated by sampling from the updated information on thermal sensations.

The Bayesian model is developed as follows.

For a given thermal comfort survey *m*:

- **Step 1:** A possible TSV distribution is assumed based on existing information and knowledge, which is known as the prior distribution. There are certain individual differences in students' thermal sensations due to factors such as climate, age, physiological conditions, and adaptability. However, in a field suvery, students in the same classroom share similar characteristics and backgrounds and are exposed to an identical indoor thermal environment [11]. Therefore, in an ideal scenario with a large surveyed sample, the collected TSV should follow a normal distribution with a central tendency represented by the mean value μ as MTSV, reflecting the general sensation of the surveyed sample. This is consistent with the assumption of the predicted mean vote (PMV) index proposed by Fanger [23]. The μ should be assumed using the PMV index based on the students' mean clothing insulation value, metabolic rate and the measured indoor thermal parameters of the classroom (i.e., air temperature, mean radiant temperature, relative humidity and air velocity). Notably, the uncertainty of the PMV should be considered and reflected by setting a standard deviation (σ_{PMV}). Therefore, the assumed MTSV (μ) can be expressed as:

$$\mu \sim N\left(PMV, \sigma_{PMV}{}^2\right) \tag{2}$$

The probability notation of the prior can be expressed as:

$$p(\mu) = N\left(PMV, \sigma_{PMV}{}^2\right) \tag{3}$$

Existing studies have found that the accuracy of PMV increases with the age of the students. PMV was found to be more accurate for secondary school students whose ages are closer to adults, and less accurate for children in primary schools [24]; [11,25]. Therefore, a smaller σ_{PMV} should be defined for older students and a larger value for younger students.

- **Step 2**: A likelihood function for the observed evidence (the TSV collected from the students) should be defined, which is:

$$TSV_{m,i} \sim N\left(\mu, \sigma_{observation}{}^2\right) \tag{4}$$

This means that the individual TSV observation in thermal comfort survey *m* - $TSV_{m,i}$, is drawn from a normal distribution with a mean of μ and a standard deviation of $\sigma_{observation}$. The $\sigma_{observation}$ reflects the uncertainty of the observed TSV data, which is mainly derived from two aspects: (i) the students' ability to perceive and express their thermal sensation, and (ii) the number of surveyed students in the classroom. It can be calculated by:

$$\sigma_{observation}{}^2 = \sigma_{sensation}{}^2 + \sigma_{sample}{}^2 \tag{5}$$

As previously mentioned, children may have difficulties in understanding and expressing their true thermal sensations, which results in uncertainty associated with the collected TSV. Therefore, $\sigma_{sensation}$ should be smaller for older students and larger for younger students. Moreover, the fewer the number of students surveyed in the classroom, the less likely the true MTSV can be observed. Therefore, σ_{sample} should be smaller for the survey with more students in the classroom and larger with fewer students.

The probability notation of the individual likelihood function can be expressed as:

$$p\left(TSV_{m,i} \middle| \mu\right) = N\left(TSV_{m,i} \middle| \mu, \sigma_{observation}{}^2\right) \tag{6}$$

Hence, for the thermal comfort survey *m* with the number of N_m observed TSV data $Dm = \left\{TSV_{m,i}\right\}_{i=1}^{N_m}$, the combined likelihood function for all observations can be expressed as:

$$p(D_m|\mu) = \prod_{i=1}^{N_m} N\left(TSV_{m,i} \middle| \mu, \sigma_{observation}{}^2\right) \tag{7}$$

Combining the prior and likelihood, the posterior distribution (updated TSV distribution) can be calculated by:

$$p(\mu|D_m) \propto p(\mu) \cdot p(D_m|\mu) \tag{8}$$

which is:

$$p(\mu|D_m) = N\left(\mu_{updated}, \sigma_{updated}{}^2\right) \tag{9}$$

where,

$$\sigma_{updated}{}^2 = \left(\frac{1}{\sigma_{PMV}{}^2} + \frac{N_m}{\sigma_{observation}{}^2}\right)^{-1} \tag{10}$$

$$\mu_{updated} = \sigma_{updated}{}^2\left(\frac{PMV}{\sigma_{PMV}{}^2} + \frac{\sum_{i=1}^{N_m} TSV_{m,i}}{\sigma_{observation}{}^2}\right) \tag{11}$$

- **Step 3:** Based on the updated posterior distribution and the observed TSV data, the predictive posterior distribution can be obtained by sampling from the posterior distribution, which can be expressed as:

$$p\left(TSV_{\text{sampling}}|D_m\right) = N\left(\mu_{sampling}, \sigma_{sampling}{}^2\right) \tag{12}$$

where,

$$\mu_{sampling} \cong \mu_{updated} \tag{13}$$

$$\sigma_{sampling}{}^2 \cong \sigma_{updated}{}^2 + \sigma_{obseration}{}^2 \tag{14}$$

The mean of the predictive posterior distribution - $\mu_{sampling}$ is the enhanced MTSV value. The $\mu_{sampling}$ approaches $\mu_{updated}$ as the sampling size increases. The purpose of using $\mu_{sampling}$ instead of directly using $\mu_{updated}$ as the MTSV is to simulate the randomness in reality. When it is not preferred to account for such randomness, the $\mu_{updated}$ can be directly extracted and used for the enhanced MTSV.

In essence, this model is a hierarchical Bayesian model that is dependent on the age of the students, which is particularly useful for thermal comfort field surveys that cover multiple classrooms with students of different ages. Students of the same age share exactly identical prior information, while the difference in the number of students in the surveyed classrooms leads to different levels of uncertainty in the observation. The prior information differs between students of different ages, and the difference in the student's ability to perceive and express their thermal sensations contributes to the observation uncertainty.

Compared to the existing MTSV calculation method, the proposed method involves a more complex process but properly addresses the associated uncertainties. The method can be understood as a tool to calibrate all MTSV data collected throughout the entire thermal comfort field survey by defining a unified rule set that addresses these uncertainties.

2.3. *Validation and evaluation of enhanced mean thermal sensation vote with thermal comfort models*

To evaluate the effect of the enhanced MTSV on thermal comfort model development, a comparative analysis was performed with a case study.

The evaluated thermal comfort models include three types of commonly adopted models:

The first type is the adaptive regression (AR) model. This model is defined in EN 16798-1 [26] and ASHRAE 55 [21] to characterize the relationship between the comfort temperature of the occupants and the outdoor running mean temperature. The AR model is expressed as:

$$T_c = a_r \cdot T_{rm} + b_r \tag{15}$$

where T_c is the comfort temperature, T_{rm} is the running mean temperature, a_r is the slope, and b_r is the intercept.

The comfort temperature can be calculated by Refs. [27]:

$$T_c = T_g - \frac{MTSV}{G} \tag{16}$$

Where T_g is the globe temperature, MTSV is the mean thermal sensation vote, and G is the Griffiths constant of 0.5 as validated by previous studies [28]; [11].

The outdoor running mean temperature can be calculated through the daily mean outdoor air temperatures for the seven days prior to the day of measurement [26,29]:

$$T_{rm} = \frac{(T_{out-1} + 0.8 \cdot T_{out-2} + 0.6 \cdot T_{out-3} + 0.5 \cdot T_{out-4} + 0.4 \cdot T_{out-5} + 0.3 \cdot T_{out-6} + 0.2 \cdot T_{out-7})}{3.8} \tag{17}$$

The second type is the adaptive PMV model, which is a category of thermal comfort models that are improved based on the PMV index. The evaluated adaptive PMV models include the adaptive-rational PMV model (arPMV) and the reinforced extended PMV model ($ePMV_m$), as they were found to be the most effective for primary and secondary school students in the previous study [11].

The arPMV model is introduced by Zhang and Lin [30], based on the aPMV model originally proposed by Yao et al. [31]. The arPMV model is expressed as:

$$arPMV = \frac{PMV'}{1 + \lambda_v \cdot PMV'} - 5 \tag{18}$$

$$\lambda_v = q \cdot \frac{1}{T} + p \tag{19}$$

$$q = \frac{\sum_i^n \frac{MTSV'^2_i}{PMV'_i T_i} \cdot \sum_i^n MTSV'^2_i + \sum_i^n \frac{MTSV'^2_i}{T_i} \cdot \sum_i^n MTSV'_i - \sum_i^n \frac{MTSV'^2_i}{PMV'_i} \cdot \sum_i^n \frac{MTSV'^2_i}{T_i} - \sum_i^n MTSV'^2_i \cdot \sum_i^n \frac{MTSV'_i}{T_i}}{\sum_i^n \frac{MTSV'^2_i}{T_i} \cdot \sum_i^n \frac{MTSV'^2_i}{T_i} - \sum_i^n \frac{MTSV'^2_i}{T_i} \cdot \sum_i^n MTSV'^2_i} \tag{20}$$

$$p = \frac{\sum_i^n MTSV'_i - \sum_i^n \frac{MTSV'^2_i}{PMV'_i} - q \cdot \sum_i^n \frac{MTSV'^2_i}{T_i}}{\sum_i^n MTSV'^2_i} \tag{21}$$

$$PMV' = PMV + 5 \tag{22}$$

$$MTSV' = MTSV + 5 \tag{23}$$

Where T is the operative temperature, PMV is the predicted mean vote index, MTSV is the mean thermal sensation vote, i is the ith thermal comfort survey, and n is the number of total surveys.

The operative temperature can be calculated by Ref. [32]:

$$T_{op} = \frac{(T_{mrt} + T_a)}{2} (0 < V_a < 0.2m/s) \tag{24}$$

$$T_{op} = \frac{\left(T_{mrt} + \left(T_a \times \sqrt{10V_a}\right)\right)}{\left(1 + \sqrt{10V_a}\right)} \quad (V_a > 0.2m/s) \tag{25}$$

Where T_a denotes the air temperature, T_{mrt} is the mean radiant temperature and V_a is the air velocity.

The $ePMV_m$ model is proposed by Zhang and Lin [33] based on the ePMV model originally introduced by Fanger and Tofum [34]. The $ePMV_m$ model is expressed as:

$$ePMV_m = e_m \cdot PMV + c_m \tag{26}$$

$$e_m = a_m \cdot T + b_m \tag{27}$$

$$a_m = \frac{\left(\frac{\sum_i^n PMV_i T_i \cdot \sum_i^n MTSV}{n} - \sum_i^n PMV_i MTSV_i T_i\right)\left(\sum_i^n PMV_i^2 - \frac{(\sum_i^n PMV_i)^2}{n}\right)}{\left(\sum_i^n PMV_i^2 T_i - \frac{\sum_i^n PMV_i T_i \cdot \sum_i^n PMV_i}{n}\right)^2 - \left(\sum_i^n PMV_i^2 T_i^2 - \frac{\sum_i^n PMV_i T_i \cdot \sum_i^n PMV_i T_i}{n}\right)\left(\sum_i^n PMV_i^2 - \frac{(\sum_i^n PMV_i)^2}{n}\right)} - \frac{\left(\frac{\sum_i^n PMV_i \cdot \sum_i^n MTSV_i}{n} - \sum_i^n PMV_i MTSV_i\right)\left(\sum_i^n PMV_i^2 T_i - \frac{\sum_i^n PMV_i T_i \cdot \sum_i^n PMV_i}{n}\right)}{\left(\sum_i^n PMV_i^2 T_i - \frac{\sum_i^n PMV_i T_i \cdot \sum_i^n PMV_i}{n}\right)^2 - \left(\sum_i^n PMV_i^2 T_i^2 - \frac{\sum_i^n PMV_i T_i \cdot \sum_i^n PMV_i T_i}{n}\right)\left(\sum_i^n PMV_i^2 - \frac{(\sum_i^n PMV_i)^2}{n}\right)} \tag{28}$$

$$b_m = \frac{\sum_i^n PMV_i MTSV_i - \frac{\sum_i^n PMV_i \cdot \sum_i^n MTSV_i}{n} - a_m\left(\sum_i^n PMV_i^2 T_i - \frac{\sum_i^n PMV_i T_i \cdot \sum_i^n PMV_i}{n}\right)}{\sum_i^n PMV_i^2 - \frac{(\sum_i^n PMV_i)^2}{n}} \tag{29}$$

$$c_m = \frac{\sum_i^n MTSV_i - a_m \cdot \sum_i^n PMV_i T_i - b_m \cdot \sum_i^n PMV_i}{n} \tag{30}$$

where T is the operative temperature, PMV is the predicted mean vote index, MTSV is the mean thermal sensation vote, i is the ith survey, and n is the number of total surveys.

The third type is the MTSV regression model, which characterizes the relationship between MTSV and relevant factors. In this study, two MTSV regression models were evaluated.

One is the simple linear regression of MTSV based on operative temperature, which is the most commonly used MTSV regression model in relevant studies. This model is named MTSV-LR, and is expressed as:

$$MTSV = \alpha_l \cdot T_{op} + b_l \tag{31}$$

where T_{op} is the operative temperature, a_l is the slope and b_l is the intercept.

The other is the multiple linear regression of MTSV based on operative temperature, running mean temperature and the age of students. It is called MTSV-MLR, and is expressed as:

$$MTSV = \alpha + \beta_1 \cdot T_{op} + \beta_2 \cdot T_{rm} + \beta_3 \cdot Age + \varepsilon \tag{32}$$

where T_{op} is the operative temperature, T_{rm} is the running mean temperature, Age is the age of students, α is the intercept, β is the coefficient for the factor, and ε is the error term.

The performance evaluation metrics for these thermal models are:

- Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(MTSV_i - P_i)^2}{n}} \tag{33}$$

where n is the number of total surveys, $MTSV_i$ is the mean thermal sensation vote for the ith survey, and P_i is the predicted value of MTSV by the model for the ith survey.

The RMSE characterizes the error between the prediction and the true value, which reflects the accuracy of the model. A low RMSE indicates higher accuracy of the model.

- Coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(MTSV_i - P_i)^2}{\sum_{i=1}^{n}(MTSV_i - \overline{MTSV})^2} \tag{34}$$

where $\overline{MTSV}$ is the average of all MTSV values.

R^2 characterizes the proportion of variance in the MTSV that can be explained by the model variables, which reflects the fit of the model to the data. A high R^2 usually suggests a better fit to the data and better performance.

For the case study, the above thermal comfort models were built using MTSV obtained through the existing calculation method and the method proposed by this research. The performance of the thermal comfort models was evaluated and compared to examine the improvement that can be achieved. To further analyze and discuss the proposed method, a benchmark scenario was first established. In the benchmark scenario, appropriate values for the required parameters of the Bayesian model (σ_{PMV}, $\sigma_{sensation}$, σ_{sample}) were defined based on the characteristics of the studied student samples. Then, a sensitivity analysis was performed to assess and discuss the influences of parameter settings on the developed thermal comfort model, to provide recommendations. The details of the case study are presented in the next section.

3. Case study results and discussion

This section analyzes and discusses the results of the case study. Section 3.1 describes the data used for the analysis, Section 3.2 defines the benchmark scenario and evaluates the improvement that can be achieved with the enhanced MTSV data, and Section 3.3 discusses the sensitivity analysis.

The calculation and analysis were performed on the Google Colab platform using Python 3.7.3. The Pythermalcomfort package

Table 1
Descriptive statistics of the female and male student datasets.

Dataset	Female student dataset	Male student dataset
Proportion of students of different ages		
5 year old	20.4 %	18.3 %
9 year old	39.8 %	30.9 %
12 year old	22.2 %	23.4 %
16 year old	17.6 %	27.4 %
Number of students in the classroom in each survey		
Mean	11	11
Standard deviation	4	3
Min	1	3
Median	11	11
Max	19	17

developed by the Center for the Built Environment (CBE Berkeley) was used to calculate the PMV index. The Numpy and Pandas packages were used for data processing. The Pymc and Arviz packages were used for Bayesian inference. The Statsmodels package was used to build the regression model.

3.1. Description of case study data

The case study data were gathered by the IAQ4EDU project [35] from April 2022 to January 2023, which conducted repeated field surveys in 32 classrooms of 9 primary schools and 7 secondary schools in Catalonia, Spain. The thermal comfort surveys involved 5 and 9 year old students in primary schools and covered 12 and 16 year old students in secondary schools. A total of 96 surveys were conducted and 1787 TSV were collected in total. This research is a continuation of the previous study by Miao et al. [11], where details of the field surveys can be referred to.

The field surveys generally obtained relatively balanced student samples. A total of 55 % of the samples were primary school students and 45 % were secondary school students. In terms of gender, 49 % were female students and 51 % were male students. However, the number of students and the gender ratio in each classroom were no different in each field survey. To avoid the uncertainty caused by the gender difference, the original survey data were separated into a female student dataset (with 869 TSV data) and a male student dataset (with 918 TSV data). Table 1 shows the descriptive statistics of the two datasets. For the female student dataset, the number of students in the classroom ranges from 1 to 19 in each survey, with an average of 11. For the male student dataset, the number of students in the classroom ranges from 3 to 17 in each survey, with an average of 11. The proportion of students of different ages in the two datasets is also different. As a result, these datasets are suitable choices for validating the proposed method.

3.2. Enhancement of mean thermal sensation vote and benchmark scenario

The enhanced MTSV was calculated based on the method defined in Section 2.2. Firstly, based on the indoor thermal parameters of the classroom, the metabolic rate, and the clothing insulation value of the students measured in each field survey, the PMV index was calculated for the prior μ in the Bayesian model, and a total of 96 PMV values were obtained. If the personal data of students' height and weight are available, the metabolic rate can be adjusted to obtain a more accurate PMV result [36]. However, due to the government's regulations for protecting the privacy of students, these personal data could not be obtained during the field survey of the IAQ4EDU project. Hence, the default metabolic rate of 1.2 met for the sedentary state was used for the PMV calculation in this case

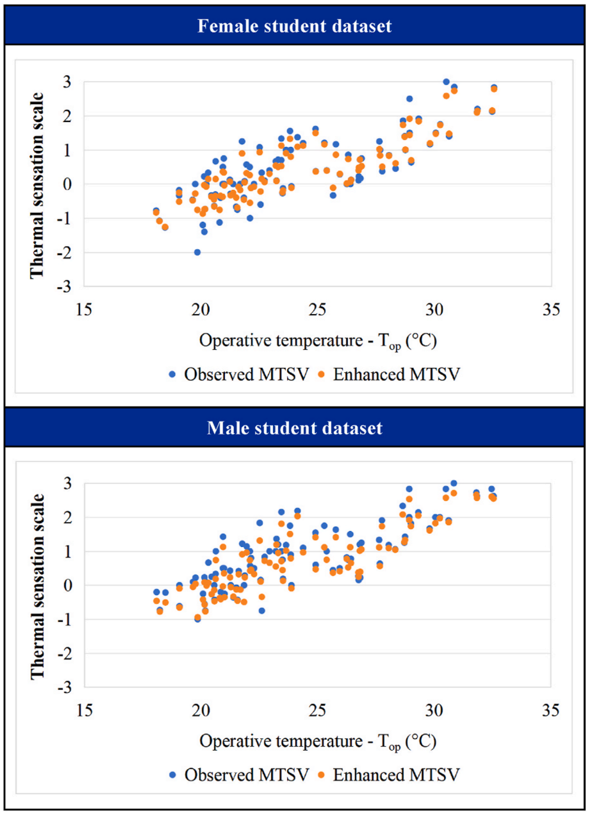

Fig. 3. Observed MTSV and enhanced MTSV.

study. Then, standard deviations were defined for both the prior and observed data. The data used in this case study refer to students in four age groups, namely groups of 5, 9, 12, and 16 years old. Corresponding to these age groups, the standard deviation of prior (σ_{PMV}) was defined as 1.0, 0.75, 0.50 and 0.25, respectively. This means that for 5-year-old students, their actual MTSV should be within ±2 of the calculated PMV index in 95 % of the cases. For 16-year-old students, their actual MTSV should be within ±0.5 of the calculated PMV index. This assumption is made considering the results reported in relevant studies [24]; [11,25]. For the uncertainty associated with the observed data, $\sigma_{sensation}$ was also defined as 1.0, 0.75, 0.50 and 0.25 for each age group, respectively. This means that for 16-year-old students, there is confidence that the difference between their expressed thermal sensation and their true thermal sensation is within ±0.5 in 95 % of the cases, while for 5-year-old students, such uncertainty is greater. Considering the number of students in each survey (Table 1), the σ_{sample} was defined in 4 levels. The value was defined as 0.25 for surveys with more than 15 students in the classroom, 0.50 for surveys with 10–15 students, 0.75 for surveys with 5–10 students, and 1.0 for surveys with less than 5 students. Hence, the enhanced MTSV was calculated with a sampling size of 1,000, and this scenario is regarded as the benchmark for comparison in the subsequent sensitivity analysis.

Fig. 3 presents the MTSV calculated using the existing method (Observed MTSV) and the proposed method (Enhanced MTSV), based on both the female and male student datasets. It can be seen that the enhanced MTSV data points are more compact and less discrete than the observed MTSV. It was also observed that some data points are closer to the observed MTSV while others are not. This is because the uncertainty associated with each survey varies, which results in different enhancement results.

Moreover, Table 2 summarizes the performance of five thermal comfort models built with the observed MTSV and the enhanced MTSV. The parameters of the thermal comfort models that were developed can be found in Appendix A. In general, both the accuracy and R^2 of these thermal comfort models improved substantially with the enhanced MTSV. The achieved improvement in accuracy ranged from 10.1 % to 30.9 %, while the improvement in R^2 ranged from 5.3 % to 28.8 %. The improvement for each model was slightly different. As shown, the adaptive PMV models (arPMV and $ePMV_m$) and the MTSV regression models (MTSV-LR and MTSV-MLR) achieved better improvements with the enhanced MTSV than the AR model. This is because the adaptive PMV and MTSV regression models are developed directly on the MTSV, while the AR model is built on the comfort temperature calculated indirectly by the MTSV. In addition, the comparison showed that the improvements achieved in the female and male student datasets were slightly different. This is mainly due to the characteristics of the two datasets in terms of the number of students in the classroom and the proportion of students of different ages.

3.3. Sensitivity analysis

Based on the benchmark scenario, the σ_{PMV}, $\sigma_{sensation}$ and σ_{sample} were adjusted to observe the influence of the enhanced MTSV on the performance of the five thermal comfort models that were developed, including the following four scenarios summarized in Fig. 4.

Appendix A summarizes the sensitivity analysis results. The parameters of the thermal comfort models can be found in Appendix B. As shown in scenario A (SA-1 to SA-3), when the uncertainty of the observed data ($\sigma_{sensation}$ and σ_{sample}) remains unchanged and the prior uncertainty (σ_{PMV}) increases, the performance of the thermal comfort models gradually decreases by 6.6 %–28.8 % in RMSE and 2.8 %–16.0 % in R^2 compared to the benchmark scenario. In contrast, in scenario B (SB-1 to SB-3) and scenario C (SC-1 to SC-3), when the σ_{PMV} remains unchanged, but $\sigma_{sensation}$ and σ_{sample} increase, the performance of the thermal comfort models also gradually increases. This is because the Bayesian model makes a trade-off between the prior and the observed data, considering the “confidence” (uncertainty) in them. Increasing the uncertainty of one aspect results in giving more weight to the other aspect when the results are calculated. Therefore, increasing confidence in the observed data can affect the enhanced MTSV and developed thermal comfort models when there are more uncertainties or noise in the observed data. By comparing scenario B and scenario C, it was found that the changes in model performance are slightly different. Compared to the benchmark scenario, RMSE increased by 4.5 %–29.7 % and R^2 improved by 1.3 %–18.7 % in scenario B, varying by the thermal comfort models. In scenario C, the values were 4.0 %–28.8 % and 1.0 %–17.9 %, respectively. The reason is that the effect of $\sigma_{sensation}$ on the enhanced MTSV only depends on the proportion of students of different age groups, and the effect of σ_{sample} only depends on the number of surveyed students in the classroom. Furthermore, scenario

Table 2
Performance of thermal comfort models built on the observed MTSV and the enhanced MTSV.

Female student dataset										
Thermal comfort model	AR		arPMV		$ePMV_m$		MTSV-LR		MTSV-MLR	
Metric	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Observed MTSV	1.442	0.676	0.643	0.566	0.627	0.588	0.613	0.606	0.527	0.708
Enhanced MTSV	1.296	0.712	0.456	0.729	0.445	0.741	0.464	0.731	0.396	0.795
Improvement	10.1 %	5.3 %	29.1 %	28.8 %	29.0 %	26.0 %	24.3 %	20.6 %	24.9 %	12.3 %
Male student dataset										
Thermal comfort model	**AR**		**arPMV**		**$ePMV_m$**		**MTSV-LR**		**MTSV-MLR**	
Metric	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**
Observed MTSV	1.52	0.635	0.67	0.504	0.62	0.575	0.708	0.624	0.513	0.709
Enhanced MTSV	1.367	0.676	0.551	0.636	0.515	0.682	0.489	0.713	0.427	0.781
Improvement	10.1 %	6.5 %	17.8 %	26.2 %	16.9 %	18.6 %	30.9 %	14.3 %	16.8 %	10.2 %

Benchmark scenario
σ_{PMV} = 0.25, 0.5, 0.75, 1.0
$\sigma_{sensation}$ = 0.25, 0.5, 0.75, 1.0
σ_{sample} = 0.25, 0.5, 0.75, 1.0
Scenario A (SA): σ_{PMV} increases, $\sigma_{sensation}$ and σ_{sample} remain unchanged
SA-1: σ_{PMV} = 0.5, 0.75, 1.0, 1.25
SA-2: σ_{PMV} = 0.75, 1.0, 1.25, 1.5
SA-3: σ_{PMV} = 1.0, 1.25, 1.5, 1.75
Scenario B (SB): $\sigma_{sensation}$ increases, σ_{PMV} and σ_{sample} remain unchanged
SB-1: $\sigma_{sensation}$ = 0.5, 0.75, 1.0, 1.25
SB-2: $\sigma_{sensation}$ = 0.75, 1.0, 1.25, 1.5
SB-3: $\sigma_{sensation}$ = 1.0, 1.25, 1.5, 1.75
Scenario C (SC): σ_{sample} increases, σ_{PMV} and $\sigma_{sensation}$ remain unchanged
SC-1: σ_{sample} = 0.5, 0.75, 1.0, 1.25
SC-2: σ_{sample} = 0.75, 1.0, 1.25, 1.5
SC-3: σ_{sample} = 1.0, 1.25, 1.5, 1.75
Scenario D (SD): σ_{PMV}, $\sigma_{sensation}$ and σ_{sample} all increases
SD-1: σ_{PMV}, $\sigma_{sensation}$, σ_{sample} = 0.5, 0.75, 1.0, 1.25
SD-2: σ_{PMV}, $\sigma_{sensation}$, σ_{sample} = 0.75, 1.0, 1.25, 1.5
SD-3: σ_{PMV}, $\sigma_{sensation}$, σ_{sample} = 1.0, 1.25, 1.5, 1.75

Fig. 4. Scenarios for the sensitivity analysis.

D (SD-1 to SD-3) shows that the performance of the thermal comfort models is almost the same as in the benchmark scenario, with only minor variations due to sampling randomness. This is because simultaneously increasing the uncertainties of both the prior and the observed data does not change the ratio of uncertainty between the two aspects, and thus does not affect the outcomes. However, it should be noted that an increase in overall uncertainty can lead to a larger standard deviation of the posterior distribution obtained by the Bayesian model. When sampling from the posterior distribution with a small sampling size, the mean of the samples obtained may deviate further from the mean of the posterior distribution. Therefore, σ_{PMV}, $\sigma_{sensation}$ and σ_{sample} should be set with reasonable values and ranges to avoid affecting the enhanced MTSV outcomes. Relevant studies are suggested to refer to the setting of the benchmark scenario of this study and adjust according to the real situation and needs.

4. Development of open software

The MTSV calculation method proposed in this research was implemented with Python language and related packages, which requires the programming skills of users. Therefore, to promote the use of the proposed method and help relevant users to work more efficiently, an user-friendly application software was developed and opened to the public. The software was compiled into an executable (.exe) file with around 460 Megabyte (MB), which encapsulates all required Python packages (Table 3). The users can run this application directly on a 64-bit Windows system without any additional requirements. The software is available on the Github platform: https://github.com/Misaeon/Bayesian-Mean-Thermal-Sensation-Vote-Calculator.

In essence, this software is a calculator that helps users who are unfamiliar with Python programming and Bayesian inference tools to easily run complex computational programs that have been compiled into it. It was designed with a user-friendly, flexible interactive interface. Users only need to refer to the methodology and case study example of this research to define the parameters of the Bayesian model for the entire thermal comfort survey, according to their real situation and needs. After entering the indoor thermal parameters (i.e., indoor air temperature, mean radiant temperature, relative humidity and air velocity), students' metabolic rate, mean clothing insulation value, and the collected TSV data of each field survey into the interface, the application can automatically calculate and display the PMV index following ASHRAE 55 standard [21], the observed MTSV value calculated by the existing method and the enhanced MTSV value calculated through the proposed method in this study, as shown in Fig. 5.

5. Conclusions and recommendations

This research proposed a novel approach to compute the mean thermal sensation vote (MTSV) for primary and secondary schools using Bayesian inference, in order to improve the effectiveness of the developed thermal comfort models for students. Compared to the

Table 3
Python packages, versions and functions for the developed software.

Package	Version	Function
Numpy	1.26.4	Data processing
Pandas	2.1.4	Data processing
Pythermalcomfort	2.10.0	PMV calculation
Pymc	5.15.0	Bayesian inference
Arviz	0.18.0	Bayesian inference

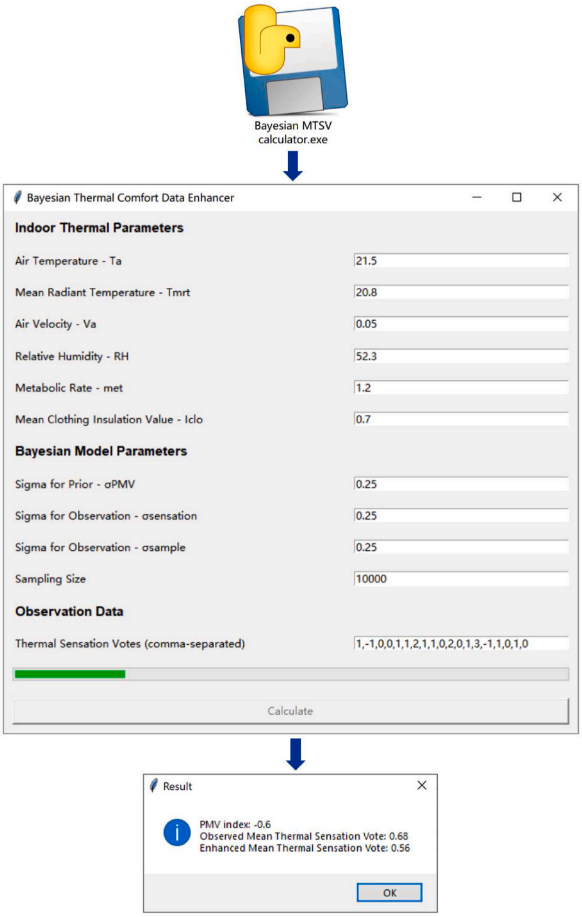

Fig. 5. The developed software application.

existing MTSV calculation method that does not properly address the uncertainties associated with the gathered thermal sensation vote (TSV) data, the proposed method can be considered a unified rule set for addressing uncertainties and calibrating all the MTSV data collected throughout the field survey.

The proposed method is validated through a holistic case study, with data obtained from a total of 96 field surveys repeated in primary and secondary schools, covering student samples of different age groups. Based on the characteristics of the surveyed students' sample in the case study and the existing knowledge reported by related research, the benchmark of the parameter setting for the Bayesian model was proposed and then evaluated with a sensitivity analysis. The results showed that the enhanced MTSV significantly improves all validated thermal comfort models (AR, arPMV, $ePMV_m$, MTSV-LR and MTSV-MLR). The accuracy of these models increased by 10.1 %–30.9 %, and the R^2 improved by 5.3 %–28.8 %. The results of the sensitivity analysis indicated that the parameters need to be set within a reasonable range to avoid affecting the enhanced MTSV outcomes. Relevant studies are suggested to refer to the proposed benchmark for parameter setting in their practical applications.

Finally, an open, user-friendly software was developed to help relevant users to work more efficiently with the proposed method. Future research can refer to this study and use the software to enhance the MTSV data directly with the collected TSV data to develop more effective thermal models for students.

CRediT authorship contribution statement

Sen Miao: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marta Gangolells:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Blanca Tejedor:** Writing – review & editing, Supervision, Conceptualization.

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.

Acknowledgments

This research is part of the R&D project IAQ4EDU, reference no. PID2020-117366RB-I00, funded by MCIN/AEI/10.13039/501100011033. This work was supported by the Catalan Agency AGAUR under their research group support program (2021 SGR 00341). The author Sen Miao is funded by the China Scholarship Council (CSC) as a full-time PhD student, reference no. 202208390065.

Appendix A. Performance of thermal comfort models under different scenarios

AR model								
Scenario	Female student dataset				Male student dataset			
Metric	RMSE	Variation	R^2	Variation	RMSE	Variation	R^2	Variation
Benchmark	1.296	0 %	0.712	0 %	1.367	0 %	0.676	0 %
SA-1	1.334	−2.9 %	0.705	−1.0 %	1.42	−3.9 %	0.663	−1.9 %
SA-2	1.361	−5.0 %	0.698	−2.0 %	1.454	−6.4 %	0.653	−3.4 %
SA-3	1.382	−6.6 %	0.692	−2.8 %	1.468	−7.4 %	0.649	−4.0 %
SB-1	1.275	1.6 %	0.716	0.6 %	1.331	2.6 %	0.686	1.5 %
SB-2	1.251	3.5 %	0.721	1.3 %	1.296	5.2 %	0.696	3.0 %
SB-3	1.238	4.5 %	0.721	1.3 %	1.262	7.7 %	0.706	4.4 %
SC-1	1.279	1.3 %	0.716	0.6 %	1.332	2.6 %	0.686	1.5 %
SC-2	1.251	3.5 %	0.72	1.1 %	1.298	5.0 %	0.695	2.8 %
SC-3	1.244	4.0 %	0.719	1.0 %	1.27	7.1 %	0.703	4.0 %
SD-1	1.293	0.2 %	0.713	0.1 %	1.37	−0.2 %	0.675	−0.1 %
SD-2	1.294	0.2 %	0.713	0.1 %	1.371	−0.3 %	0.675	−0.1 %
SD-3	1.294	0.2 %	0.714	0.3 %	1.372	−0.4 %	0.676	0.0 %
arPMV model								
Scenario	Female student dataset				Male student dataset			
Metric	RMSE	Variation	R^2	Variation	RMSE	Variation	R^2	Variation
Benchmark	0.456	0 %	0.729	0 %	0.67	0 %	0.636	0 %
SA-1	0.526	−15.4 %	0.665	−8.8 %	0.551	17.8 %	0.584	−8.2 %
SA-2	0.56	−22.8 %	0.635	−12.9 %	0.597	10.9 %	0.556	−12.6 %
SA-3	0.587	−28.7 %	0.612	−16.0 %	0.622	7.2 %	0.542	−14.8 %
SB-1	0.424	7.0 %	0.759	4.1 %	0.635	5.2 %	0.674	6.0 %
SB-2	0.386	15.4 %	0.793	8.8 %	0.513	23.4 %	0.715	12.4 %
SB-3	0.35	23.2 %	0.826	13.3 %	0.471	29.7 %	0.755	18.7 %
SC-1	0.422	7.5 %	0.758	4.0 %	0.429	36.0 %	0.672	5.7 %
SC-2	0.388	14.9 %	0.793	8.8 %	0.515	23.1 %	0.711	11.8 %
SC-3	0.352	22.8 %	0.825	13.2 %	0.477	28.8 %	0.75	17.9 %
SD-1	0.47	−3.1 %	0.717	−1.6 %	0.436	34.9 %	0.634	−0.3 %
SD-2	0.478	−4.8 %	0.711	−2.5 %	0.55	17.9 %	0.634	−0.3 %
SD-3	0.47	−3.1 %	0.716	−1.8 %	0.551	17.8 %	0.632	−0.6 %
ePMVm model								
Scenario	Female student dataset				Male student dataset			
Metric	RMSE	Variation	R^2	Variation	RMSE	Variation	R^2	Variation
Benchmark	0.445	0 %	0.741	0 %	0.515	0 %	0.682	0 %
SA-1	0.513	−15.3 %	0.681	−8.1 %	0.556	−8.0 %	0.64	−6.2 %
SA-2	0.547	−22.9 %	0.652	−12.0 %	0.577	−12.0 %	0.617	−9.5 %
SA-3	0.573	−28.8 %	0.631	−14.8 %	0.589	−14.4 %	0.606	−11.1 %
SB-1	0.414	7.0 %	0.769	3.8 %	0.481	6.6 %	0.714	4.7 %
SB-2	0.378	15.1 %	0.802	8.2 %	0.442	14.2 %	0.749	9.8 %
SB-3	0.343	22.9 %	0.833	12.4 %	0.403	21.7 %	0.784	15.0 %
SC-1	0.413	7.2 %	0.769	3.8 %	0.483	6.2 %	0.712	4.4 %
SC-2	0.38	14.6 %	0.801	8.1 %	0.448	13.0 %	0.745	9.2 %
SC-3	0.345	22.5 %	0.831	12.1 %	0.411	20.2 %	0.778	14.1 %
SD-1	0.46	−3.4 %	0.73	−1.5 %	0.513	0.4 %	0.681	−0.1 %
SD-2	0.467	−4.9 %	0.724	−2.3 %	0.515	0.0 %	0.68	−0.3 %
SD-3	0.46	−3.4 %	0.729	−1.6 %	0.512	0.6 %	0.681	−0.1 %
MTSV-LR model								
Scenario	Female student dataset				Male student dataset			
Metric	RMSE	Variation	R^2	Variation	RMSE	Variation	R^2	Variation
Benchmark	0.464	0 %	0.731	0 %	0.489	0 %	0.713	0 %
SA-1	0.511	−10.1 %	0.684	−6.4 %	0.525	−7.4 %	0.679	−4.8 %
SA-2	0.54	−16.4 %	0.66	−9.7 %	0.544	−11.2 %	0.66	−7.4 %

(continued on next page)

(*continued*)

AR model								
Scenario	Female student dataset				Male student dataset			
Metric	RMSE	Variation	R^2	Variation	RMSE	Variation	R^2	Variation
SA-3	0.563	−21.3 %	0.643	−12.0 %	0.555	−13.5 %	0.651	−8.7 %
SB-1	0.427	8.0 %	0.755	3.3 %	0.461	5.7 %	0.737	3.4 %
SB-2	0.395	14.9 %	0.784	7.3 %	0.427	12.7 %	0.766	7.4 %
SB-3	0.367	20.9 %	0.809	10.7 %	0.395	19.2 %	0.793	11.2 %
SC-1	0.425	8.4 %	0.755	3.3 %	0.462	5.5 %	0.736	3.2 %
SC-2	0.399	14.0 %	0.782	7.0 %	0.435	11.0 %	0.76	6.6 %
SC-3	0.371	20.0 %	0.806	10.3 %	0.405	17.2 %	0.785	10.1 %
SD-1	0.464	0.0 %	0.725	−0.8 %	0.487	0.4 %	0.713	0.0 %
SD-2	0.469	−1.1 %	0.722	−1.2 %	0.489	0.0 %	0.712	−0.1 %
SD-3	0.464	0.0 %	0.724	−1.0 %	0.485	0.8 %	0.713	0.0 %
MTSV-MLR model								
Scenario	Female student dataset				Male student dataset			
Metric	RMSE	Variation	R^2	Variation	RMSE	Variation	R^2	Variation
Benchmark	0.396	0 %	0.795	0 %	0.427	0 %	0.781	0 %
SA-1	0.44	−11.1 %	0.765	−3.8 %	0.463	−8.4 %	0.75	−4.0 %
SA-2	0.465	−17.4 %	0.748	−5.9 %	0.482	−12.9 %	0.734	−6.0 %
SA-3	0.484	−22.2 %	0.737	−7.3 %	0.491	−15.0 %	0.727	−6.9 %
SB-1	0.373	5.8 %	0.813	2.3 %	0.401	6.1 %	0.802	2.7 %
SB-2	0.344	13.1 %	0.836	5.2 %	0.367	14.1 %	0.827	5.9 %
SB-3	0.321	18.9 %	0.853	7.3 %	0.336	21.3 %	0.85	8.8 %
SC-1	0.372	6.1 %	0.812	2.1 %	0.401	6.1 %	0.801	2.6 %
SC-2	0.348	12.1 %	0.833	4.8 %	0.372	12.9 %	0.824	5.5 %
SC-3	0.327	17.4 %	0.849	6.8 %	0.344	19.4 %	0.845	8.2 %
SD-1	0.401	−1.3 %	0.794	−0.1 %	0.428	−0.2 %	0.778	−0.4 %
SD-2	0.403	−1.8 %	0.794	−0.1 %	0.429	−0.5 %	0.779	−0.3 %
SD-3	0.401	−1.3 %	0.793	−0.3 %	0.427	0.0 %	0.778	−0.4 %

Appendix B. Thermal comfort models that were developed with original and enhanced MTSV under different scenarios

AR model		
Scenario	Female student dataset	Male student dataset
Original	$T_c = 0.320 \cdot T_{rm} + 18.082$	$T_c = 0.308 \cdot T_{rm} + 17.469$
Benchmark	$T_c = 0.314 \cdot T_{rm} + 18.268$	$T_c = 0.303 \cdot T_{rm} + 17.863$
SA-1	$T_c = 0.317 \cdot T_{rm} + 18.182$	$T_c = 0.306 \cdot T_{rm} + 17.703$
SA-2	$T_c = 0.318 \cdot T_{rm} + 18.152$	$T_c = 0.307 \cdot T_{rm} + 17.625$
SA-3	$T_c = 0.319 \cdot T_{rm} + 18.132$	$T_c = 0.307 \cdot T_{rm} + 17.578$
SB-1	$T_c = 0.311 \cdot T_{rm} + 18.340$	$T_c = 0.303 \cdot T_{rm} + 17.978$
SB-2	$T_c = 0.309 \cdot T_{rm} + 18.420$	$T_c = 0.301 \cdot T_{rm} + 18.096$
SB-3	$T_c = 0.306 \cdot T_{rm} + 18.508$	$T_c = 0.301 \cdot T_{rm} + 18.215$
SC-1	$T_c = 0.312 \cdot T_{rm} + 18.326$	$T_c = 0.303 \cdot T_{rm} + 17.976$
SC-2	$T_c = 0.309 \cdot T_{rm} + 18.407$	$T_c = 0.302 \cdot T_{rm} + 18.083$
SC-3	$T_c = 0.306 \cdot T_{rm} + 18.508$	$T_c = 0.301 \cdot T_{rm} + 18.205$
SD-1	$T_c = 0.314 \cdot T_{rm} + 18.277$	$T_c = 0.304 \cdot T_{rm} + 17.855$
SD-2	$T_c = 0.313 \cdot T_{rm} + 18.289$	$T_c = 0.304 \cdot T_{rm} + 17.860$
SD-3	$T_c = 0.314 \cdot T_{rm} + 18.270$	$T_c = 0.304 \cdot T_{rm} + 17.844$
arPMV model		
Scenario	Female student dataset	Male student dataset
Original	$arPMV = \frac{PMV'}{1 + \left(\frac{-0.843}{T_{op}} + 0.0198\right) \cdot PMV'} - 5$	$arPMV = \frac{PMV'}{1 + \left(\frac{-1.456}{T_{op}} + 0.0303\right) \cdot PMV'} - 5$
Benchmark	$arPMV = \frac{PMV'}{1 + \left(\frac{-0.738}{T_{op}} + 0.0180\right) \cdot PMV'} - 5$	$arPMV = \frac{PMV'}{1 + \left(\frac{-1.147}{T_{op}} + 0.0229\right) \cdot PMV'} - 5$
SA-1	$arPMV = \frac{PMV'}{1 + \left(\frac{-0.773}{T_{op}} + 0.0185\right) \cdot PMV'} - 5$	$arPMV = \frac{PMV'}{1 + \left(\frac{-1.270}{T_{op}} + 0.0259\right) \cdot PMV'} - 5$
SA-2	$arPMV = \frac{PMV'}{1 + \left(\frac{-0.785}{T_{op}} + 0.0186\right) \cdot PMV'} - 5$	$arPMV = \frac{PMV'}{1 + \left(\frac{-1.336}{T_{op}} + 0.0275\right) \cdot PMV'} - 5$

(continued on next page)

(*continued*)

AR model		
Scenario	Female student dataset	Male student dataset
SA-3	$arPMV = \frac{PMV'}{1+\left(\frac{-0.799}{T_{op}}+0.0188\right)\cdot PMV'}-5$	$arPMV = \frac{PMV'}{1+\left(\frac{-1.364}{T_{op}}+0.0281\right)\cdot PMV'}-5$
SB-1	$arPMV = \frac{PMV'}{1+\left(\frac{-0.699}{T_{op}}+0.0172\right)\cdot PMV'}-5$	$arPMV = \frac{PMV'}{1+\left(\frac{-1.069}{T_{op}}+0.0215\right)\cdot PMV'}-5$
SB-2	$arPMV = \frac{PMV'}{1+\left(\frac{-0.650}{T_{op}}+0.0161\right)\cdot PMV'}-5$	$arPMV = \frac{PMV'}{1+\left(\frac{-0.990}{T_{op}}+0.0200\right)\cdot PMV'}-5$
SB-3	$arPMV = \frac{PMV'}{1+\left(\frac{-0.599}{T_{op}}+0.0149\right)\cdot PMV'}-5$	$arPMV = \frac{PMV'}{1+\left(\frac{-0.906}{T_{op}}+0.0184\right)\cdot PMV'}-5$
SC-1	$arPMV = \frac{PMV'}{1+\left(\frac{-0.720}{T_{op}}+0.0180\right)\cdot PMV'}-5$	$arPMV = \frac{PMV'}{1+\left(\frac{-1.065}{T_{op}}+0.0213\right)\cdot PMV'}-5$
SC-2	$arPMV = \frac{PMV'}{1+\left(\frac{-0.651}{T_{op}}+0.0159\right)\cdot PMV'}-5$	$arPMV = \frac{PMV'}{1+\left(\frac{-0.988}{T_{op}}+0.0197\right)\cdot PMV'}-5$
SC-3	$arPMV = \frac{PMV'}{1+\left(\frac{-0.602}{T_{op}}+0.0149\right)\cdot PMV'}-5$	$arPMV = \frac{PMV'}{1+\left(\frac{-0.907}{T_{op}}+0.0183\right)\cdot PMV'}-5$
SD-1	$arPMV = \frac{PMV'}{1+\left(\frac{-0.716}{T_{op}}+0.0173\right)\cdot PMV'}-5$	$arPMV = \frac{PMV'}{1+\left(\frac{-1.164}{T_{op}}+0.0237\right)\cdot PMV'}-5$
SD-2	$arPMV = \frac{PMV'}{1+\left(\frac{-0.699}{T_{op}}+0.0166\right)\cdot PMV'}-5$	$arPMV = \frac{PMV'}{1+\left(\frac{-1.156}{T_{op}}+0.0234\right)\cdot PMV'}-5$
SD-3	$arPMV = \frac{PMV'}{1+\left(\frac{-0.722}{T_{op}}+0.0175\right)\cdot PMV'}-5$	$arPMV = \frac{PMV'}{1+\left(\frac{-1.177}{T_{op}}+0.0241\right)\cdot PMV'}-5$
$ePMV_m$ model		
Scenario	Female student dataset	Male student dataset
Original	$ePMV_m = (0.00099\cdot T_{op} + 0.891)\cdot PMV + 0.345$	$ePMV_m = (-0.0022\cdot T_{op} + 0.932)\cdot PMV + 0.815$
Benchmark	$ePMV_m = (-0.00004\cdot T_{op} + 0.925)\cdot PMV + 0.309$	$ePMV_m = (-0.0055\cdot T_{op} + 1.059)\cdot PMV + 0.664$
SA-1	$ePMV_m = (0.00052\cdot T_{op} + 0.905)\cdot PMV + 0.322$	$ePMV_m = (-0.0027\cdot T_{op} + 0.969)\cdot PMV + 0.718$
SA-2	$ePMV_m = (0.00040\cdot T_{op} + 0.907)\cdot PMV + 0.327$	$ePMV_m = (-0.0020\cdot T_{op} + 0.941)\cdot PMV + 0.749$
SA-3	$ePMV_m = (0.00046\cdot T_{op} + 0.907)\cdot PMV + 0.333$	$ePMV_m = (-0.0023\cdot T_{op} + 0.947)\cdot PMV + 0.769$
SB-1	$ePMV_m = (0.00013\cdot T_{op} + 0.925)\cdot PMV + 0.288$	$ePMV_m = (-0.0052\cdot T_{op} + 1.060)\cdot PMV + 0.612$
SB-2	$ePMV_m = (0.00003\cdot T_{op} + 0.932)\cdot PMV + 0.267$	$ePMV_m = (-0.0059\cdot T_{op} + 1.083)\cdot PMV + 0.564$
SB-3	$ePMV_m = (0.00043\cdot T_{op} + 0.927)\cdot PMV + 0.245$	$ePMV_m = (-0.0068\cdot T_{op} + 1.114)\cdot PMV + 0.514$
SC-1	$ePMV_m = (0.00017\cdot T_{op} + 0.920)\cdot PMV + 0.292$	$ePMV_m = (-0.0063\cdot T_{op} + 1.089)\cdot PMV + 0.616$
SC-2	$ePMV_m = (-0.00075\cdot T_{op} + 0.957)\cdot PMV + 0.277$	$ePMV_m = (-0.0074\cdot T_{op} + 1.127)\cdot PMV + 0.573$
SC-3	$ePMV_m = (0.00046\cdot T_{op} + 0.928)\cdot PMV + 0.248$	$ePMV_m = (-0.0076\cdot T_{op} + 1.139)\cdot PMV + 0.521$
SD-1	$ePMV_m = (0.00034\cdot T_{op} + 0.917)\cdot PMV + 0.301$	$ePMV_m = (-0.0038\cdot T_{op} + 1.011)\cdot PMV + 0.659$
SD-2	$ePMV_m = (-0.00008\cdot T_{op} + 0.930)\cdot PMV + 0.299$	$ePMV_m = (-0.0049\cdot T_{op} + 1.040)\cdot PMV + 0.661$
SD-3	$ePMV_m = (0.00002\cdot T_{op} + 0.924)\cdot PMV + 0.301$	$ePMV_m = (-0.0023\cdot T_{op} + 0.966)\cdot PMV + 0.658$
MTSV-LR model		
Scenario	Female student dataset	Male student dataset
Original	$MTSV = 0.205\cdot T_{op} - 4.504$	$MTSV = 0.203\cdot T_{op} - 4.046$
Benchmark	$MTSV = 0.202\cdot T_{op} - 4.467$	$MTSV = 0.208\cdot T_{op} - 4.327$
SA-1	$MTSV = 0.203\cdot T_{op} - 4.470$	$MTSV = 0.206\cdot T_{op} - 4.219$
SA-2	$MTSV = 0.203\cdot T_{op} - 4.480$	$MTSV = 0.205\cdot T_{op} - 4.157$
SA-3	$MTSV = 0.204\cdot T_{op} - 4.495$	$MTSV = 0.204\cdot T_{op} - 4.134$
SB-1	$MTSV = 0.202\cdot T_{op} - 4.497$	$MTSV = 0.208\cdot T_{op} - 4.387$
SB-2	$MTSV = 0.203\cdot T_{op} - 4.532$	$MTSV = 0.208\cdot T_{op} - 4.439$
SB-3	$MTSV = 0.204\cdot T_{op} - 4.569$	$MTSV = 0.208\cdot T_{op} - 4.494$
SC-1	$MTSV = 0.201\cdot T_{op} - 4.473$	$MTSV = 0.208\cdot T_{op} - 4.389$
SC-2	$MTSV = 0.203\cdot T_{op} - 4.535$	$MTSV = 0.209\cdot T_{op} - 4.445$
SC-3	$MTSV = 0.204\cdot T_{op} - 4.571$	$MTSV = 0.209\cdot T_{op} - 4.499$
SD-1	$MTSV = 0.203\cdot T_{op} - 4.501$	$MTSV = 0.207\cdot T_{op} - 4.307$
SD-2	$MTSV = 0.204\cdot T_{op} - 4.526$	$MTSV = 0.207\cdot T_{op} - 4.315$
SD-3	$MTSV = 0.203\cdot T_{op} - 4.491$	$MTSV = 0.206\cdot T_{op} - 4.291$

(*continued on next page*)

(*continued*)

AR model		
Scenario	Female student dataset	Male student dataset
MTSV-MLR model		
Scenario	Female student dataset	Male student dataset
Original	$MTSV = 0.287 \cdot T_{op} - 0.059 \cdot T_{rm} - 0.051 \cdot Age - 5.03$	$MTSV = 0.255 \cdot T_{op} - 0.038 \cdot T_{rm} - 0.058 \cdot Age - 4.11$
Benchmark	$MTSV = 0.268 \cdot T_{op} - 0.048 \cdot T_{rm} - 0.029 \cdot Age - 5.00$	$MTSV = 0.262 \cdot T_{op} - 0.040 \cdot T_{rm} - 0.044 \cdot Age - 4.57$
SA-1	$MTSV = 0.276 \cdot T_{op} - 0.052 \cdot T_{rm} - 0.038 \cdot Age - 5.01$	$MTSV = 0.261 \cdot T_{op} - 0.040 \cdot T_{rm} - 0.047 \cdot Age - 4.43$
SA-2	$MTSV = 0.280 \cdot T_{op} - 0.055 \cdot T_{rm} - 0.042 \cdot Age - 5.02$	$MTSV = 0.259 \cdot T_{op} - 0.039 \cdot T_{rm} - 0.049 \cdot Age - 4.34$
SA-3	$MTSV = 0.282 \cdot T_{op} - 0.056 \cdot T_{rm} - 0.045 \cdot Age - 5.03$	$MTSV = 0.258 \cdot T_{op} - 0.039 \cdot T_{rm} - 0.051 \cdot Age - 4.28$
SB-1	$MTSV = 0.264 \cdot T_{op} - 0.045 \cdot T_{rm} - 0.026 \cdot Age - 5.01$	$MTSV = 0.260 \cdot T_{op} - 0.038 \cdot T_{rm} - 0.042 \cdot Age - 4.61$
SB-2	$MTSV = 0.262 \cdot T_{op} - 0.042 \cdot T_{rm} - 0.023 \cdot Age - 5.05$	$MTSV = 0.260 \cdot T_{op} - 0.037 \cdot T_{rm} - 0.039 \cdot Age - 4.69$
SB-3	$MTSV = 0.261 \cdot T_{op} - 0.040 \cdot T_{rm} - 0.021 \cdot Age - 5.10$	$MTSV = 0.258 \cdot T_{op} - 0.036 \cdot T_{rm} - 0.037 \cdot Age - 4.74$
SC-1	$MTSV = 0.264 \cdot T_{op} - 0.044 \cdot T_{rm} - 0.026 \cdot Age - 5.01$	$MTSV = 0.262 \cdot T_{op} - 0.039 \cdot T_{rm} - 0.043 \cdot Age - 4.62$
SC-2	$MTSV = 0.261 \cdot T_{op} - 0.042 \cdot T_{rm} - 0.024 \cdot Age - 5.02$	$MTSV = 0.263 \cdot T_{op} - 0.039 \cdot T_{rm} - 0.041 \cdot Age - 4.72$
SC-3	$MTSV = 0.259 \cdot T_{op} - 0.040 \cdot T_{rm} - 0.020 \cdot Age - 5.06$	$MTSV = 0.259 \cdot T_{op} - 0.037 \cdot T_{rm} - 0.040 \cdot Age - 4.73$
SD-1	$MTSV = 0.271 \cdot T_{op} - 0.049 \cdot T_{rm} - 0.032 \cdot Age - 5.05$	$MTSV = 0.261 \cdot T_{op} - 0.039 \cdot T_{rm} - 0.043 \cdot Age - 4.54$
SD-2	$MTSV = 0.273 \cdot T_{op} - 0.049 \cdot T_{rm} - 0.034 \cdot Age - 5.06$	$MTSV = 0.261 \cdot T_{op} - 0.039 \cdot T_{rm} - 0.044 \cdot Age - 4.54$
SD-3	$MTSV = 0.271 \cdot T_{op} - 0.049 \cdot T_{rm} - 0.032 \cdot Age - 5.04$	$MTSV = 0.260 \cdot T_{op} - 0.039 \cdot T_{rm} - 0.043 \cdot Age - 4.53$

Data availability

Data will be made available on request.

References

[1] M.K. Singh, R. Ooka, H.B. Rijal, S. Kumar, A. Kumar, S. Mahapatra, Progress in thermal comfort studies in classrooms over last 50 years and way forward, Energy Build. 188–189 (2019) 149–174, https://doi.org/10.1016/j.enbuild.2019.01.051.
[2] J. Jiang, D. Wang, Y. Liu, Y. Xu, J. Liu, A study on pupils' learning performance and thermal comfort of primary schools in China, Build. Environ. 134 (2018) 102–113, https://doi.org/10.1016/j.buildenv.2018.02.036.
[3] S. Miao, M. Gangolells, B. Tejedor, A comprehensive assessment of indoor air quality and thermal comfort in educational buildings in the Mediterranean climate, Indoor Air 2023 (2023) 6649829, https://doi.org/10.1155/2023/6649829.
[4] C.D. Korkas, S. Baldi, I. Michailidis, E.B. Kosmatopoulos, Intelligent energy and thermal comfort management in grid-connected microgrids with heterogeneous occupancy schedule, Appl. Energy 149 (2015) 194–203, https://doi.org/10.1016/j.apenergy.2015.01.145.
[5] C.D. Korkas, S. Baldi, I. Michailidis, E.B. Kosmatopoulos, Occupancy-based demand response and thermal comfort optimization in microgrids with renewable energy sources and energy storage, Appl. Energy 163 (2016) 93–104, https://doi.org/10.1016/j.apenergy.2015.10.140.
[6] H. Shi, Q. Chen, Building energy management decision-making in the real world: a comparative study of HVAC cooling strategies, J. Build. Eng. 33 (2021) 101869, https://doi.org/10.1016/j.jobe.2020.101869.
[7] N. Upasani, O. Guerra-Santin, M. Mohammadi, Developing building-specific, occupant-centric thermal comfort models: a methodological approach, J. Build. Eng. 95 (2024) 110281, https://doi.org/10.1016/j.jobe.2024.110281.
[8] D. Enescu, A review of thermal comfort models and indicators for indoor environments, Renew. Sustain. Energy Rev. 79 (2017) 1353–1379, https://doi.org/10.1016/j.rser.2017.05.175.
[9] R. Yao, S. Zhang, C. Du, M. Schweiker, S. Hodder, B.W. Olesen, J. Toftum, F.R. d'Ambrosio, H. Gebhardt, S. Zhou, F. Yuan, B. Li, Evolution and performance analysis of adaptive thermal comfort models – a comprehensive literature review, Build. Environ. 217 (2022) 109020, https://doi.org/10.1016/j.buildenv.2022.109020.
[10] Z.S. Zomorodian, M. Tahsildoost, M. Hafezi, Thermal comfort in educational buildings: a review article, Renew. Sustain. Energy Rev. 59 (2016) 895–906, https://doi.org/10.1016/j.rser.2016.01.033.
[11] S. Miao, M. Gangolells, B. Tejedor, Improving the thermal comfort model for students in naturally ventilated schools: insights from a holistic study in the Mediterranean climate, Build. Environ. 258 (2024) 111622, https://doi.org/10.1016/j.buildenv.2024.111622.
[12] C. Chang, X. Li, L. Duanmu, B. Sun, H. Ju, Analysis of the impact of indoor thermal comfort data characteristics on dataset quality, Energy Build. 310 (2024) 114079, https://doi.org/10.1016/j.enbuild.2024.114079.
[13] S. Karjalainen, Thermal comfort and gender: a literature review, Indoor Air 22 (2) (2011) 96–109, https://doi.org/10.1111/j.1600-0668.2011.00747.x.
[14] H. Al-Khatri, M. Alwetaishi, M.B. Gadi, Exploring thermal comfort experience and adaptive opportunities of female and male high school students, J. Build. Eng. 31 (2020) 101365, https://doi.org/10.1016/j.jobe.2020.101365.
[15] A. Martinez-Molina, P. Boarin, I. Tort-Ausina, J.-L. Vivancos, Post-occupancy evaluation of a historic primary school in Spain: comparing PMV, TSV and PD for teachers' and pupils' thermal comfort, Build. Environ. 117 (2017) 248–259, https://doi.org/10.1016/j.buildenv.2017.03.010.
[16] C.M. Rodriguez, M.C. Coronado, J.M. Medina, Classroom-comfort-data: a method to collect comprehensive information on thermal comfort in school classrooms, MethodsX 6 (2019) 2698–2719, https://doi.org/10.1016/j.mex.2019.11.004.
[17] Z. Wang, T. Hong, Learning occupants' indoor comfort temperature through a Bayesian inference approach for office buildings in United States, Renew. Sustain. Energy Rev. 119 (2020) 109593, https://doi.org/10.1016/j.rser.2019.109593.
[18] K.W. Mui, T.W. Tsang, L.T. Wong, Bayesian updates for indoor thermal comfort models, J. Build. Eng. 29 (2020) 101117, https://doi.org/10.1016/j.jobe.2019.101117.
[19] Z.Q. Fard, Z.S. Zomorodian, S.S. Korsavi, Application of machine learning in thermal comfort studies: a review of methods, performance and challenges, Energy Build. 256 (2022) 111771, https://doi.org/10.1016/j.enbuild.2021.111771.
[20] ISO, EN ISO 7730: Ergonomics of the Thermal Environment - Analytical Determination and Interpretation of Thermal Comfort Using Calculation of the PMV and PPD Indices and Local Thermal Comfort Criteria, ISO, Geneva, 2005.
[21] ASHRAE, ASHRAE 55: Thermal Environment Conditions for Human Occupancy, ASHRAE, Atlanta, 2020.
[22] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, Bayesian Data Analysis, third ed., Chapman and Hall/CRC Press, 2013 https://doi.org/10.1201/b16018.
[23] P.O. Fanger, Thermal Comfort, Analysis and Applications in Environmental Engineering, Danish Technical Press, Copenhagen, 1970.

[24] G. Torriani, G. Lamberti, G. Salvadori, F. Fantozzi, F. Babich, Thermal comfort and adaptive capacities: differences among students at various school stages, Build. Environ. 237 (2023) 110340, https://doi.org/10.1016/j.buildenv.2023.110340.
[25] M. Almagro-Lidón, C. Pérez-Carramiñana, A. Galiano-Garrigós, S. Emmitt, Thermal comfort in school children: testing the validity of the Fanger method for a Mediterranean climate, Build. Environ. 253 (2024) 111305, https://doi.org/10.1016/j.buildenv.2024.111305.
[26] CEN, EN 16798-1: indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality. Thermal Environment, Lighting and Acoustics, CEN, Brussels, 2019.
[27] N. Forcada, M. Gangolells, M. Casals, B. Tejedor, M. Macarulla, K. Gaspar, Field study on adaptive thermal comfort models for nursing homes in the Mediterranean climate, Energy Build. 252 (2021) 111475, https://doi.org/10.1016/j.enbuild.2021.111475.
[28] M. Trebilcock, J. Soto-Muñoz, M. Yañez, R. Figueroa-San Martin, The right to comfort: a field study on adaptive thermal comfort in free-running primary schools in Chile, Build. Environ. 114 (2017) 455–469, https://doi.org/10.1016/j.buildenv.2016.12.036.
[29] B. Tejedor, M. Casals, M. Gangolells, M. Macarulla, N. Forcada, Human comfort modelling for elderly people by infrared thermography: evaluating the thermoregulation system responses in an indoor environment during winter, Build. Environ. 186 (2020) 107354, https://doi.org/10.1016/j.buildenv.2020.107354.
[30] S. Zhang, Z. Lin, Adaptive-rational thermal comfort model: adaptive predicted mean vote with variable adaptive coefficient, Indoor Air 30 (5) (2020) 1052–1062, https://doi.org/10.1111/ina.12665.
[31] R. Yao, B. Li, J. Liu, A theoretical adaptive model of thermal comfort – adaptive Predicted Mean Vote (aPMV), Build. Environ. 44 (10) (2009) 2089–2096, https://doi.org/10.1016/j.buildenv.2009.02.014.
[32] S. Kumar, M.K. Singh, A. Mathur, J. Mathur, S. Mathur, Evaluation of comfort preferences and insights into behavioural adaptation of students in naturally ventilated classrooms in a tropical country, India, Build. Environ. 143 (2018) 532–547, https://doi.org/10.1016/j.buildenv.2018.07.035.
[33] S. Zhang, Z. Lin, Extended predicted mean vote of thermal adaptations reinforced around thermal neutrality, Indoor Air 31 (4) (2021) 1227, https://doi.org/10.1111/ina.12792, 1227.
[34] P.O. Fanger, J. Toftum, Extension of the PMV model to non-air-conditioned buildings in warm climates, Energy Build. 34 (6) (2002) 533–536, https://doi.org/10.1016/S0378-7788(02)00003-8.
[35] IAQ4EDU project, Optimal ventilation strategies for balancing indoor air quality, comfort and energy use in educational buildings, in: Funded by State Research Agency of Spain (Agencia Estatal de Investigación), reference number PID2020-117366RB-I00, 2021. https://iaq4edu.upc.edu/en.
[36] R.M.S.F. Almeida, N.M.M. Ramos, V.P. de Freitas, Thermal comfort models and pupils' perception in free-running school buildings of a mild climate country, Energy Build. 111 (2016) 64–75, https://doi.org/10.1016/j.enbuild.2015.09.066.