Total Environment Research Themes 8 (2023) 100083

Contents lists available at ScienceDirect

Total Environment Research Themes

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

Development and comparison of adaptive data-driven models for thermal comfort assessment and control

Giulia Lamberti [a,b,*], Roberto Boghetti [c], Jérôme H. Kämpf [c], Fabio Fantozzi [a], Francesco Leccese [a], Giacomo Salvadori [a]

[a] University of Pisa, School of Engineering, Largo Lucio Lazzarino, 56122 Pisa, Italy
[b] Institut de Recherche en Constructibilité, Université Paris-Est, ESTP, 28, Avenue du Président Wilson, 94230 Cachan, France
[c] Energy Informatics Group, Idiap Research Institute, 1920 Martigny, Switzerland

ARTICLE INFO

Keywords:
Thermal comfort
Data-driven model
Adaptive thermal comfort
PMV
ASHRAE database

ABSTRACT

Thermal comfort prediction is an important issue, as it can largely influence occupants' well-being and buildings' energy consumption. Nowadays, models used to assess thermal comfort have been increasingly discussed, and a growing number of data-driven models with several input parameters developed. Although these models allow reasonably accurate predictions of thermal comfort, using complex algorithms to determine thermal comfort might be unsuitable for some use cases, such as quick estimations or real-time control of Heating, Ventilation, and Air Conditioning (HVAC) systems.

In this paper, a data-driven model was developed based on 61710 samples of subjective responses associated with environmental parameters from field studies available in two ASHRAE databases. Two models resulted from this analysis, one with higher accuracy and one simplified, which improved the prediction in comparison to other regression models and PMV.

However, since thermal comfort cannot be conceived as a punctual condition, comfort areas were derived, i.e., respective comfort ranges at 90%, 80%, and 70% of thermal acceptability. The result is that the error in the prediction of the new models is below the 90% acceptable range, which means that the models' error does not lead to a reduction in the evaluation of occupant comfort.

Built upon influential parameters, these models enable thermal comfort estimates and occupant-centered HVAC control. The notion of comfort as a non-fixed state empowers more flexible building management criteria, reducing energy use while upholding indoor comfort.

List of Symbols

G	Griffiths' constant ($°C^{-1}$)
HVAC	Heating Ventilation and Air Conditioning
I_{cl}	Clothing insulation (clo)
M	Metabolic rate (met)
MAE	Mean Absolute Error
PMV	Predicted Mean Vote
PD	Percentage of Dissatisfied
PPD	Predicted Percentage of Dissatisfied
RH	Relative humidity (%)
RMSE	Root Mean Square Error
T_a	Air temperature (°C)
T_{comf}	Comfort temperature (°C)
TCV	Thermal Comfort Vote
T_g	Globe temperature (°C)
T_{op}	Operative temperature (°C)
T_{out}	Mean monthly outdoor temperature (°C)
TSV	Thermal Sensation Vote
T_r	Mean radiant temperature (°C)
V_a	Air velocity (m/s)

1. Introduction

The importance of healthy indoor conditions in buildings has become more evident due to increased time spent indoors. Thermal comfort plays a crucial role in people's well-being, productivity, and energy consumption (Lamberti, 2020; Zhang et al., 2019). Previous

* Corresponding author.
E-mail address: giulia.lamberti@phd.unipi.it (G. Lamberti).

https://doi.org/10.1016/j.totert.2023.100083
Received 25 July 2023; Received in revised form 31 August 2023; Accepted 5 October 2023
Available online 6 October 2023

studies have extensively examined indoor thermal comfort (Enescu, 2017; Fantozzi and Rocca, 2020), with Fanger's rational and the adaptive being the primary models used (Djongyang et al., 2010).

Fanger's rational model (Fanger, 1970) evaluates occupants' thermal sensations using the Predicted Mean Vote (PMV) index, which is based on four environmental (air temperature, relative humidity, mean radiant temperature, and air velocity) and two individual (metabolic rate and clothing insulation) parameters. The rational model considers the heat balance between the human body and the environment and was developed through subjective climate chamber experiments. This model is widely used for thermal comfort analysis, particularly in the design phase, but its numerous input parameters make it challenging for real-time control of HVAC systems.

To overcome this limitation and account for adaptive processes occurring in buildings, the adaptive model was derived from field studies (Humphreys et al., 2016). Thermal adaptation, encompassing behavioural, physiological, and psychological aspects (Brager and de Dear, 1998), recognizes occupants as active contributors to comfort creation, not merely passive subjects. Factors like building type (Rupp et al., 2019), climate (Lamberti et al., 2023), and ventilation mode (Kumar, 2022) significantly shape occupants' thermal perception and adaptability and should be considered in comfort analysis. The importance of adaptation in different settings has become particularly relevant (Oliveri et al., 2016; Castilla et al., 2018) and passive strategies, based on occupants' adaptive capacities, can also improve thermal comfort conditions (Zinzi et al., 2021). In this case, the only parameters involved in the adaptive relationship are the indoor and outdoor temperatures. Although this greatly simplifies the model and makes it more suitable for quick estimations and real-time control, it reduces thermal comfort to the relationship between indoor and outdoor temperatures, without considering all the other parameters that are involved in the heat balance (Fanger and Toftum, 2002).

To consider the issue that the original adaptive models reduce the prediction of thermal sensation to the relationship between indoor and outdoor temperature only, machine learning algorithms have been recently used, leveraging the increasing availability of field data. Indeed, data-driven models present high flexibility in input parameters (Xie et al., 2020) so that they can consider aspects such as skin temperature (Dai et al., 2017), personal characteristics (Li et al., 2017; Lee et al., 2017; Zhao et al., 2014), or occupants' interaction with control systems (Kim et al., 2018). Several studies comprised the input parameters of PMV including in some cases outdoor temperature or HVAC operation mode (Jiang and Yao, 2016; Lu et al., 2019).

In particular, they have been applied to the prediction of different indices, such as 3-points Thermal Sensation Vote (TSV) (Chaudhuri et al., 2018; Chaudhuri et al., 2019; Wang et al., 2019), 7- points Thermal Sensation Vote (Jiang and Yao, 2016; Lu et al., 2019; Rana et al., 2013; Wang et al., 2014; Wu et al., 2018; Du et al., 2019), Thermal Preference Vote (Dai et al., 2017; Li et al., 2017; Lee et al., 2017; Kim et al., 2018), or Thermal Comfort Vote (TCV) (Farhan et al., 2015; Cosma and Simha, 2019).

To build data-driven models, different algorithms were used, which showed a median predictive accuracy of 84% with a standard deviation of around 15% (Xie et al., 2020). While it is true that this predictive performance is better than that of the PMV (Xie et al., 2020; Cheung et al., 2019) and thus shows how promising the use of these algorithms is for better understanding the parameters that most influence thermal comfort, they often require a greater number of input parameters for predicting thermal comfort indices.

In addition, while more complex algorithms have improved thermal sensation prediction compared to the PMV model (Xie et al., 2020), their effectiveness is limited due to the subjective nature of occupant-perceived thermal sensations. Variability in individual experiences of thermal comfort exists, making precise individual predictions impractical and unnecessary. A pragmatic alternative is to focus on "comfort zones" rather than specific conditions (Humphreys et al., 2016).

With the purpose of developing a model suitable for real-time HVAC system control, two primary challenges were identified. Firstly, the current adaptive models overlook various factors influencing the heat balance between the human body and the environment, thereby impacting thermal sensation. Secondly, these adaptive models were initially designed to simplify comfort prediction, prioritizing accessibility for building professionals.

Although it has been shown in numerous studies that the PMV often fails in predicting comfort (Cheung et al., 2019; Humphreys and Nicol, 2002), there is a need to develop models that can accurately predict occupant comfort using a minimal set of input parameters. These models should be suitable for quick estimations and real-time control.

This paper aims to fill this gap by developing a data-driven model that:

1. Is based on a considerable amount of data, to account for adaptive processes, which are only detectable from field studies.
2. Selects and includes the parameters that are most relevant to thermal perception.
3. Allows real-time control and is suitable for use by building practitioners, in line with the original adaptive model.
4. Presents a satisfying predictive performance, suitable for its use in building management.

To overcome the problem that laboratory studies may not represent accurately occupants' behaviour, ASHRAE's databases including data from real buildings well-spread around the world were used (de Dear, 1998; Földváry Ličina et al., 2018).

2. Methodology

In this section, the methodology used for developing the model is described (Fig. 1). Rows with parameters T_a, T_r, RH, V_a, I_{cl}, M, T_{out}, PMV, and TSV were extracted from ASHRAE databases. PMV was recalculated for comparison with user-perceived TSV. Filtering retained comfort data (excluding thermal stress), resulting in 62,317 to 61,710 records. Then, the filtered dataset was split into 90% training and 10% test sets. Once the baseline was established, a polynomial regression to create two thermal sensation prediction models was performed and the test set was used for validation. Further methodology details are provided in the following paragraphs.

2.1. Data source and preparation

2.1.1. Description of the databases

There are several quality-controlled thermal comfort databases available for scientific studies, such as the ASHRAE RP-884 database developed in the 1990 s, with 25,616 samples (de Dear, 1998), the Smart Control and Thermal Comfort database developed in the 2000 s (McCartney and Nicol, 2002), with 27,284 samples, and the more recent ASHRAE Database II, which has the largest size with 81,968 samples (Földváry Ličina et al., 2018).

Previously, machine learning algorithms predicted thermal comfort using individual databases (Lu et al., 2019; Farhan et al., 2015; Luo et al., 2020; Gao et al., (2019); Zhou et al., 2020). Addressing concerns from researchers encouraging broader samples (McCartney and Nicol, 2002), ASHRAE's Databases Database II and RP-884 were combined, integrating global field studies with a consistent methodology. Utilizing both databases offers a comprehensive, extended data range.

Selected model parameters include air temperature (T_a), mean radiant temperature (T_r), relative humidity (RH), air velocity (V_a), clothing insulation (I_{cl}), metabolic rate (M), and mean monthly outdoor temperature (T_{out}). When the mean radiant temperature was absent but the globe temperature (T_g) was available, T_r was derived using the ISO 7726 standard (ISO 7726, 2001). The six parameters (T_a; T_r, RH, V_a, I_{cl}, M) correspond to those in Fanger's PMV model, while outdoor

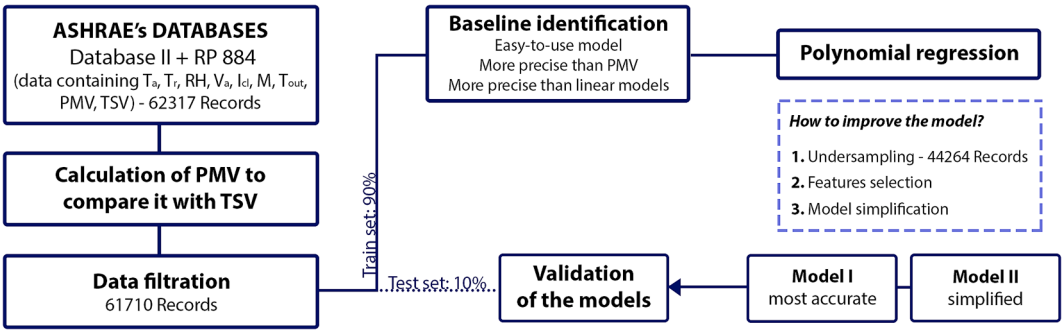

Fig. 1. Flow chart showing the procedures used for developing the thermal comfort models.

temperature (T_{out}) reflects adaptability, foundational to the adaptive model.

Then, Fanger's Predicted Mean Vote (PMV) was calculated. The database already includes the values of PMV for each entry however, as it is possible for a database of this size to contain errors (Cheung et al., 2019; Humphreys and Nicol, 2002), PMV was recalculated using the *pythermalcomfort* Python library (Tartarini and Schiavon, 2020) to ensure consistency. Samples with missing values of T_a; T_r or T_g, RH, V_a, I_{cl}, M, T_{out}, and TSV were excluded from the study. As a result, the considered database had a size of 62,317 observations, obtained from studies carried out from 1982 to 2016.

The Thermal Sensation Vote (TSV), which represents occupants' thermal sensation on a 7-points scale (from −3 cold to + 3 hot), was also extracted from the databases for the comparison between the predicted and real thermal sensation. The data originates from peer-reviewed field studies, capturing effective thermal experiences of occupants in real environments. Thermal Sensation Votes (TSV) were gathered using combined questionnaires and instrumental measurements within the same spatial and temporal context. This approach enabled the correlation of individual thermal responses with concurrent environmental parameters. Rigorous quality checks and validation procedures were applied to all submissions to mitigate potential transmission errors (Földváry Ličina et al., 2018).

2.1.2. Data filtration

After an initial analysis (see Supplementary Material), data was filtered and values beyond practical ranges were excluded. Acceptability ranges were defined based on thermal comfort standards for each input feature (ISO 7726, 2001; ISO 7730, 2006).

For instance, T_a and T_r below 10 °C were omitted due to cold stress risk. No upper limit was set, considering diverse climates. A wide humidity range of 10–90% was accepted since studies were performed in dry and humid climates. Air velocity ranged from 0.0 to 3.0 m/s, typical for indoor settings. Parameters like metabolic rates below 0.8 Met (sleeping condition) and clothing insulation below 0.0 clo were disregarded. PMV and TSV were limited to −3 to + 3 as per ASHRAE. Duplicate samples were removed, resulting in 607 exclusions.

Then, the dataset was split: 90% for training (baseline identification and model building), and 10% for testing model validation.

2.2. Baseline identification

The model aims to predict occupants' thermal sensation (7-point scale) using the identified parameters (T_a, T_r, RH, V_a, I_{cl}, M, T_{out}). After identifying influential parameters through field studies, the model aims for user-friendliness such as adaptive models. It seeks improved predictability compared to PMV and linear models, approaching complex algorithms' performance.

Model performance is assessed using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Bias. Lower values indicate higher precision. Bias detects systematic overestimation/underestimation by the predictive model. These error indices can be calculated as follows.

$$MAE = \frac{\sum_{i=1}^{n} |p_i - a_i|}{n} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (p_i - a_i)^2}{n}} \quad (2)$$

$$Bias = \frac{\sum_{i=1}^{n} (p_i - a_i)}{n} \quad (3)$$

where p_i is the predicted value; a_i is the target value, and n is the number of samples.

For comparison, the predictive performance of PMV was also assessed using these three indices (MAE, RMSE, and Bias).

Then, the performance of a linear least squares regression model, eventually regularized with different values of L1 (Lasso) and L2 (Ridge) penalties, as well as combinations of both (Elastic Net), was calculated and used as a baseline that the proposed model is expected to improve, and a primary comparison between this approach and Fanger's PMV prediction was provided.

Linear regression is one of the simplest data-driven modelling algorithms, and it has the advantage that its results can be easily understood and interpreted. Ridge regression (Hoerl and Kennard, 1970) extends it by introducing the L2 penalty to reduce the variability of the model and to mitigate the problem of multicollinearity, as it penalises higher predictors' coefficients, effectively shrinking them to values that can be close, but not equal, to zero. The Lasso regression (Tibshirani, 1996) uses a similar approach with the introduction of the L1 penalty term, which allows the parameters to be effectively reduced to zero. In this way, redundant or not useful predictors are removed, which should lead to a simpler and potentially better model. Finally, the Elastic Net regression (Zou and Hastie, 2005) is a method that combines both the L1 and L2 penalties used in the Lasso and Ridge regression models.

The models were built and optimized using Python's Scikit-learn library (Pedregosa et al., 2011). A 5x5-fold cross-validated random search was made to find optimal values of the regularization parameters. The Scikit-learn implementations of these models use the hyperparameters alpha to control the strength of the total penalty term, while in Elastic Net the l1_ratio controls the ratio of penalty assigned to the L1 term. The range of values tested was 0.05 – 50 for alphas on a logarithmic scale and 0.05–0.95 on the l1_ratio with steps of 0.05. For the Lasso, Ridge, and Elastic Net regressions, 100 different regularization penalty parameters (alphas) were tested and for the Elastic Net, 19 different mixing parameters between Lasso and Ridge (l1_ratio) were examined.

2.3. Polynomial regression

To enhance model performance, an instance selection algorithm was applied to training data, commonly used in classification but less explored in regression (Song et al., 2017; Arnaiz-González et al., 2016;

Arnaiz-González et al., 2016). This technique undersamples training data, retaining model predictiveness while reducing questionnaire data noise. The instance selection methodology applied was based on clustering, similar to common filter methods used for classification (Olvera-López et al., 2010).

Specifically, complete-linkage hierarchical agglomerative clustering (Nielsen, 2016) was employed, allowing control over same-cluster element distances. This possibility, coupled with a different preliminary scaling of each feature, results in the ability to impose a maximum difference between extreme values of each feature inside a cluster. These constraints were chosen as the limit values beyond which the thermal environment is perceived differently by occupants according to the international standard ISO 7726 (ISO 7726, 2001). The limit values of thermal environment perception, coinciding with the maximum distances of each feature allowed in a single cluster, are reported in Table 1.

Employing this approach, 44,264 clusters emerged from the filtered database. Instance selection during model training retained central samples from each cluster, refining the dataset.

From the filtered database, the parameters were combined in polynomials of degree 5th or lower, resulting in a total of 792 polynomials (also called predictor variables or input features in the context of data-driven modelling). To select the polynomials that influence the most thermal perception of the building's occupants, a Lasso regression was applied to the training set. In fact, Lasso can shrink the independent variable to 0, effectively removing predictors that are redundant or non-correlated with the target variable. Before applying the Lasso regression, each feature was also scaled to a range (0,1) using the MinMaxScaler. While scaling features do not affect the performance of linear least square regression models, this step was necessary to measure the relative importance of the polynomials as a function of the Lasso coefficients. The Lasso regression was then applied using 5x5-fold cross-validation to find the optimal value of alpha.

Thermal sensation evaluation was approached as regression, aligning with global comfort theories, as thermal comfort analysis typically forecasts group thermal perceptions, not individual responses. This is achieved through data binning, averaging sensations under similar conditions. TSV and PMV are expressed as continuous real values within the −3 to + 3 range, facilitating direct comparison.

3. Results and analysis

3.1. Data analysis

ASHRAE's databases include global studies across diverse climatic zones, illustrated in Fig. 2 (see Supplementary Material for data distribution). The dataset contains 29,236 records from naturally ventilated, 23,149 from air-conditioned, and 9325 from mixed-mode buildings. Predominantly, studies were conducted in temperate (Zone C, 38,061 records), tropical (Zone A, 11,015 records), dry (Zone B, 7118 records), and continental (Zone D, 5516 records) climates.

Table 1
Limit values of the input parameters used for clustering data.

	Parameter	Limit value	Unit
Environmental parameters	Air temperature (T_a)	0.5	(°C)
	Relative humidity (RH)	5	(%)
	Mean radiant temperature (T_r)	2	(°C)
	Air velocity (V_a)	0.05	(m/s)
	Mean monthly outdoor temperature (T_{out})	1	(°C)
Individual parameters	Clothing insulation (I_{cl})	0.1	(clo)
	Metabolic rate (M)	0.1	(Met)
	TSV	1	(-)

Observations were primarily gathered in summer (27778 records), followed by winter (22922 records), spring (5054 records), and autumn (2542 records). Data were also available for wet (2300 records) and dry (1114 records) seasons.

Noteworthy building types encompass offices (38567 records), schools (11552 records), residential buildings (7845 records), senior centres (449 records), light industrial factories (118 records), and other categories (3179 records).

Table 2 provides parameter descriptions, including assumed ranges, mean values, and predefined acceptability ranges.

Table 3 presents the parameter correlation matrix. TSV's weak correlations with other inputs reflect personal thermal sensation variability. Notably, TSV displays a notable link with air/mean radiant temperatures (r = 0.37), underscoring their impact on thermal perception. Additionally, TSV exhibits a significant correlation with clothing insulation (r = -0.18), indicating occupants' adaptive responses. This adaptability is reinforced by I_{cl}'s correlations with T_a (r = -0.44), T_r (r = -0.45), and T_{out} (r = -0.37).

However, TSV's correlation with metabolic rate is minimal (r = 0.05) due to limited sedentary activity variation in the sample. TSV's connection with RH is also weak (r = -0.03), despite RH's broad range (10% to 90%), reflecting the complex role of RH in thermal comfort (Djamila, 2017). Similarly, TSV's correlation with air velocity is modest (r = 0.05), influenced by low indoor variability. The effect of outdoor temperature on TSV is low (r = 0.03), potentially due to database-specific building variations.

Noteworthy is the correlation between I_{cl} and T_{out}, underscoring outdoor temperature's significance in thermal sensation, particularly in naturally ventilated settings. The correlation matrix underscores occupants' real-world adaptability, evident through clothing insulation's impact on thermal perception.

3.2. Model's baseline

The model is based on field studies, enabling the evaluation of user comfort within real environments, which allow to consider thermal adaptation.

PMV was used as a baseline, following previous studies on data-driven models (Xie et al., 2020). Simple regression models were also used as a reference for predictive ability. It was not possible to consider as baselines other data-driven models in the literature due to their classification-based approach for TSV evaluation, as discussed earlier.

PMV's predictive performance was evaluated using three error indices. It showed a MAE of 0.991, indicating an average error of approximately one point on ASHRAE's 7-point scale. The RMSE was 1.276, slightly higher than the standard deviation of TSV (SD_{TSV} = 1.268), while the Bias was low at −0.003, indicating no systematic underestimation or overestimation of thermal comfort.

Simple regression models, including linear regression (MAE = 0.873, RMSE = 1.135, and Bias = 0.000), Ridge regression (MAE = 0.873, RMSE = 1.135, and Bias = 0.000), Lasso (MAE = 0.872, RMSE = 1.145, and Bias = 0.000), and Elastic Net (MAE = 0.872, RMSE = 1.145, and Bias = 0.000), showed improved predictive performance compared to PMV across all three evaluation indices.

The predictive model developed in the next section should aim to further enhance the performance achieved by these simple linear models.

3.3. Prediction of the Thermal Sensation Vote

The relevant environmental and individual parameters were combined into polynomials, resulting in 792 features for regression models. Lasso regression was then used to select important parameters and reduce predictors to 37 polynomials, as shown in Table 4.

The MAE, RMSE, and Bias of the cross-validations, calculated as the average value of the validation sets of each fold, were 0.867, 1.121, and

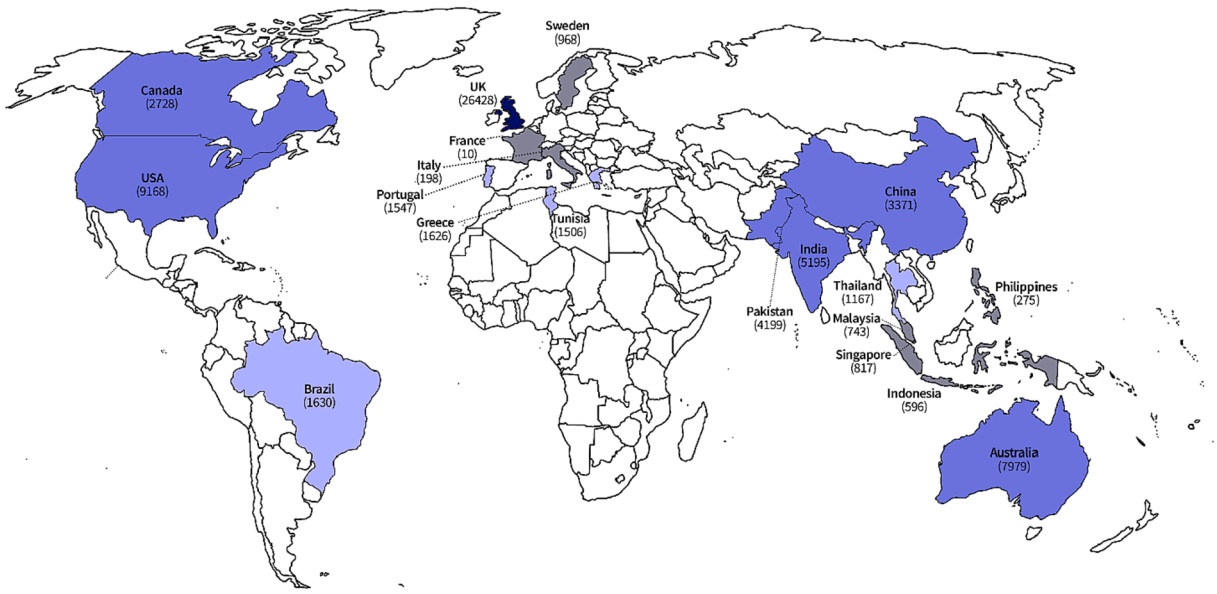

Fig. 2. Geographical distribution of the data contained in ASHRAE's databases.

Table 2
Description of the Databases' parameters.

Parameter	Acceptability range	Range assumed	Mean value
Air temperature T_a (°C)	≥ 10	10.0 – 42.7	24.0
Relative humidity RH (%)	10 – 90	10 – 89	47
Mean radiant temperature T_r (°C)	≥ 10	10.0 – 49.5	24.2
Air velocity V_a (m/s)	0 – 3	0.00 – 2.81	0.13
Clothing insulation I_{cl} (clo)	≥ 0	0.03 – 2.87	0.73
Metabolic rate M (Met)	≥ 0.8	0.8 – 4.5	1.2
Mean monthly outdoor temperature T_{out} (°C)	–	−24.9 – 35.0	17.5
PMV	−3 to + 3	−3 to + 3	0.13
TSV	−3 to + 3	−3 to + 3	0.16

0.001 respectively. The most relevant polynomial is composed of the product of T_a and T_r, revealing the significance that these two parameters have on thermal sensation.

For instance, fitting a linear regression with this polynomial only led to a MAE, RMSE, and Bias of 0.886, 1.165, and 0.003 respectively. Fig. 3 reports the Lasso coefficients of the different features in the subset, which show their relative importance in determining the thermal sensation. In blue, the features used for the model of Eq. (4), while in light blue the features used for the model of Eq. (5) are reported.

The feature comprising T_a and T_r stands out as the most important, confirming their influence on the perception of the thermal environment in the adaptive model. However, the effects of relative humidity, outdoor temperature, and metabolic rate included in Eq. (4) should not be neglected.

To make the utilization of the 37 terms in the polynomial equations from Lasso regression more practical, it was essential to further simplify the model's complexity. The objective was to identify the minimum number of polynomials necessary to attain comparable accuracy. This involved repeating curve fitting with progressively larger predictor subsets until the error approximated that of the full polynomial model. Employing 5x5-fold cross-validation ensured accurate MAE, RMSE, and Bias estimation for each model. The process was halted after the third iteration.

Across the combinations, MAE ranged from 0.866 to 0.947, RMSE spanned from 1.128 to 1.261, and Bias fluctuated between −0.012 and 0.017. Given consistently low Bias, always under the recommended ± 0.25 threshold (Humphreys and Nicol, 2002), and RMSE below TSV's standard deviation (SD = 1.307 for the clustered dataset), MAE was chosen as the criterion for selecting the optimal combination, prioritizing model accuracy.

The best model found, resulting in a MAE of 0.866, a RMSE of 1.128, and a Bias of 0.008 on the cross-validation, was:

$$TSV_I = 0.0039 T_a \cdot T_r - 3.3259 \cdot 10^{-8} T_r^2 \cdot RH \cdot T_{out}^2 + 4.5622 \cdot 10^{-7} \cdot T_a^3 \cdot RH \cdot M - 2.1152 \quad (4)$$

This equation suggests that the combination of the most important features that determine the thermal sensation includes T_a, T_r, RH, T_{out}, and M.

However, this model includes parameters that are difficult to control in real-time, such as mean radiant temperature or metabolic rate, which is typically estimated using tables based on the activity performed for global thermal comfort assessment (Cheung et al., 2019). To simplify further the equation another combination was considered, which had slightly higher MAE and RMSE, but lower Bias (MAE = 0.872, RMSE =

Table 3
Correlation matrix expressing the correlation coefficient (r) between the different parameters.

	TSV	T_a	RH	V_a	T_r	M	I_{cl}	T_{out}
TSV	1	0.37	−0.03	0.05	0.37	0.05	−0.18	0.03
T_a	0.37	1	0.16	0.39	0.93	−0.07	−0.44	0.53
RH	−0.03	0.16	1	0.22	0.19	−0.08	−0.30	0.50
V_a	0.05	0.39	0.22	1	0.36	−0.06	−0.16	0.36
T_r	0.37	0.93	0.19	0.36	1	−0.05	−0.45	0.52
M	0.05	−0.07	−0.08	−0.06	−0.05	1	−0.04	−0.10
I_{cl}	−0.18	−0.44	−0.30	−0.16	−0.45	−0.04	1	−0.37
T_{out}	0.03	0.53	0.50	0.36	0.52	−0.10	−0.37	1

Table 4
Features selected after the application of the Lasso regression.

ID	Features	ID	Features
1	$T_a \bullet T_r$	20	$T_r \bullet RH \bullet T_{out}$
2	$T_r \bullet RH^3 \bullet I_{cl}$	21	$T_a \bullet I_{cl} \bullet T_{out}^2$
3	T_{out}^5	22	$T_a \bullet RH^2 \bullet I_{cl}^2$
4	T_r^2	23	$RH \bullet V_a^2 \bullet M \bullet T_{out}$
5	$T_a^3 \bullet T_r \bullet RH$	24	$RH^2 \bullet M$
6	$RH \bullet I_{cl}^4$	25	$RH^3 \bullet M$
7	T_{out}	26	$V_a \bullet T_{out}^4$
8	$T_r \bullet T_{out}^4$	27	$T_a^2 \bullet T_r \bullet RH \bullet M$
9	$I_{cl} \bullet T_{out}^4$	28	$T_a^2 \bullet RH \bullet V_a \bullet I_{cl}$
10	$T_a^2 \bullet RH \bullet M \bullet T_{out}$	29	$T_r^2 \bullet RH \bullet T_{out}^2$
11	$T_a \bullet M \bullet I_{cl} \bullet T_{out}^2$	30	$M \bullet T_{out}$
12	RH^3	31	$T_a \bullet V_a \bullet I_{cl}$
13	$V_a^2 \bullet T_{out}^3$	32	V_a
14	I_{cl}^2	33	$T_a^2 \bullet RH^2 \bullet T_{out}$
15	$T_a \bullet I_{cl} \bullet T_{out}$	34	$RH \bullet V_a \bullet M^3$
16	$T_a \bullet M$	35	$V_a \bullet I_{cl}^2 \bullet T_{out}^2$
17	$M \bullet I_{cl}$	36	$T_a \bullet V_a \bullet M^2 \bullet I_{cl}$
18	$RH^2 \bullet T_{out}$	37	$T_a^3 \bullet RH \bullet M$
19	T_a		

1.136, Bias = 0.001). With this equation, it is possible to predict thermal sensation with few environmental parameters that are easy to measure, maintaining the predictive performance close to the previous one.

$$TSV_{II} = -0.0228T_{out} - 1.9492 \cdot 10^{-6} RH^2 \cdot T_{out} + 0.1867T_a - 3.8609 \quad (5)$$

Table 5 summarizes the indices of error for MAE, RMSE, and Bias for simple regression models (linear, Lasso, Ridge, and Elastic Net), PMV, and the two developed models. It can be noticed that the best predictive performance is achieved with the Model of Eq. (4).

It can be noticed that the correlation matrix reveals a consistent inverse relationship between TSV and I_{cl}. This means that an adaptive tendency to adjust clothing's thermal insulation based on perceived warmth or coldness, and vice versa to experience warmth or coldness while donning more or less insulated clothing. However, in the derived models, I_{cl} is not included in the equations. This outcome, rigorously derived, aligns with adaptive models, which recognise the adaptive equation in the relationship between outdoor and indoor temperatures.

3.4. Validation and boundaries of the model

The databases employed in this analysis encompass typical environments, which implies that the model's predictions might deviate under extreme conditions. To assess the models' performance, the variations in environmental parameters were investigated, focusing on Bias as the error indicator (Fig. 4). For comparison purposes, PMV was included, aligning with prior research (Xie et al., 2020).

The models' validity was evaluated by establishing accurate prediction ranges for both environmental and individual parameters. Fig. 4 illustrates the difference between predicted and actual thermal sensation values across diverse parameters. This assessment was carried out on the test set, confirming the models' reliability and setting their operational boundaries. The models were calculated based on an acceptable Bias range of ± 0.25 points on the TSV scale (Humphreys and Nicol, 2002). In Fig. 4, it's evident that despite significant response variability, the new models exhibit a tendency to decrease the Bias between predicted and observed thermal sensations compared to PMV.

The new models perform well in predicting sensations within indoor **air temperatures** of 10 °C to 35 °C, albeit overestimating for temperatures beyond 35 °C. Unlike the original PMV model, valid from 10 °C to 30 °C, the new models improve and extend predictive capacity. Regarding models in Eq. (4) and (5), distinctions mainly emerge in 10–15 °C and 35–40 °C ranges, often beyond moderate comfort, requiring specialized indices for heat or cold stress. For accurate thermal sensation forecasts using the new models, maintaining air temperature within 15–35 °C is advised. The range could potentially expand to 10–35 °C with minor precision reduction (Table 6). Beyond these limits, model performance might decrease, elevating heat or cold stress risks.

Bias remained consistent across various **relative humidity** ranges (Fig. 4), indicating occupants' modest sensitivity to humidity modifications. Overall, the proposed models slightly surpass PMV in predictive capability, with median Bias often within acceptable thresholds. Particularly, models Eq. (4) and (5) display greater enhancements over PMV for relative humidity values surpassing 60%. While not immediately perceptible, relative humidity considerably impacts long-term

Table 5
MAE, RMSE and Bias for simple regression models, PMV and the two developed models.

Model	MAE	RMSE	Bias
Linear regression	0.873	1.135	0.000
Lasso regression	0.872	1.135	0.000
Ridge regression	0.873	1.135	0.000
Elastic Net regression	0.872	1.135	0.000
PMV	0.991	1.276	−0.003
Model Eq (4)	0.866	1.128	0.008
Model Eq (5)	0.872	1.136	0.001

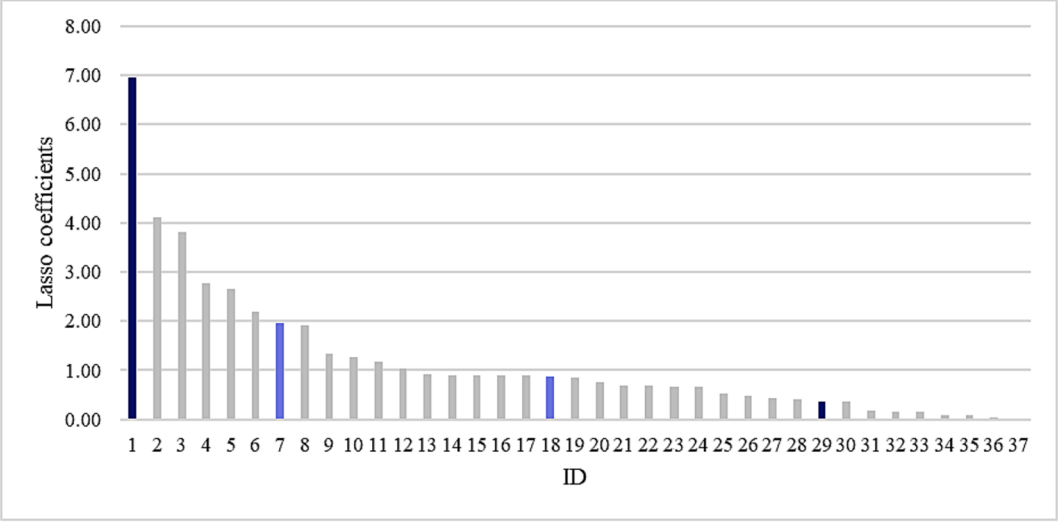

Fig. 3. Lasso coefficients associated to the different features.

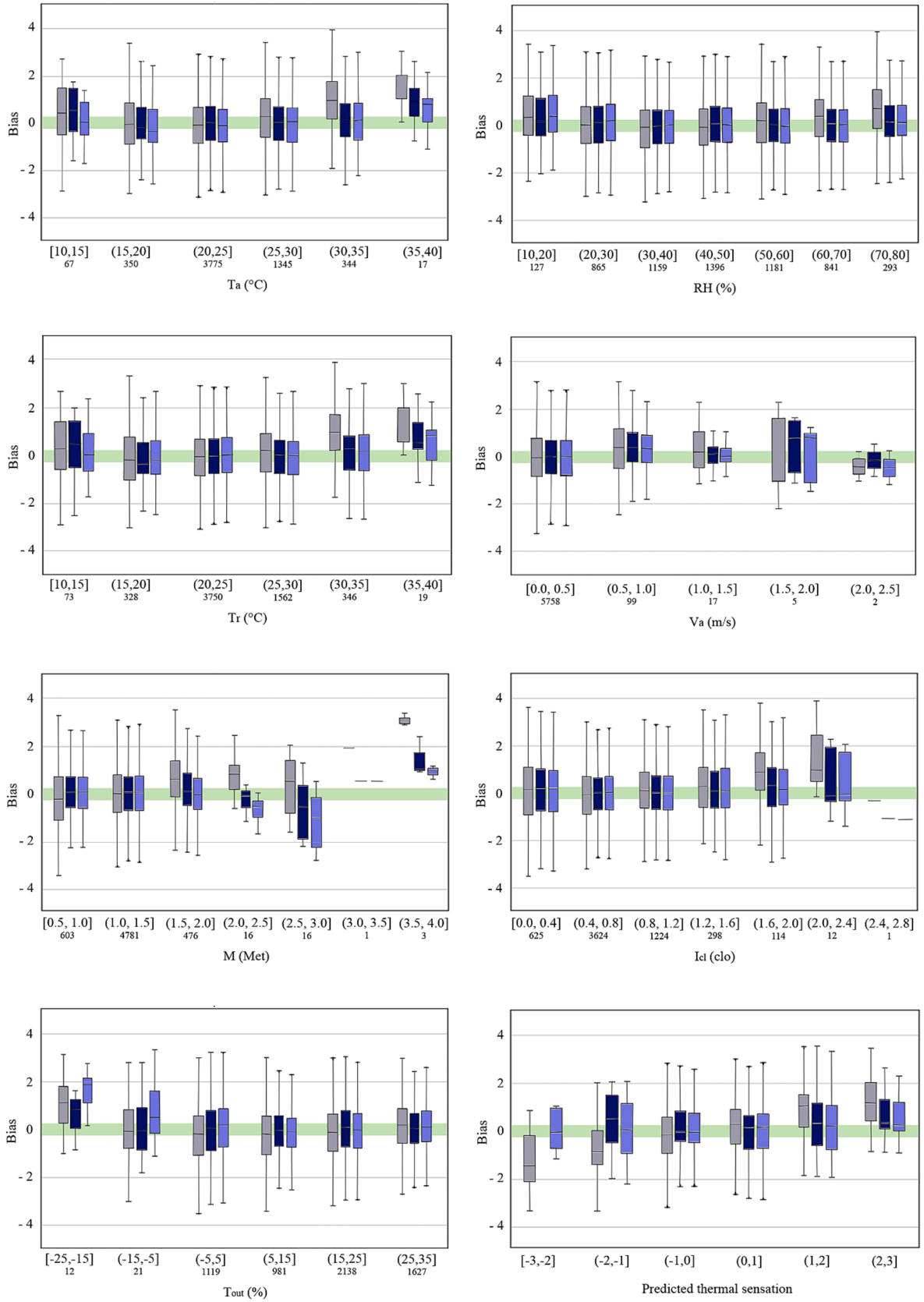

Fig. 4. Bias between the predicted and real thermal sensation against different parameters. On the x-axis the range of the parameters is reported, and the number of samples is reported below them. In grey Bias in PMV, in blue bias in model of Eq. (4), and in light blue Bias in model of Eq. (5). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
Boundaries of PMV, models of Eq. (4) and (5).

Parameter	PMV	Model Eq. (4)	Model Eq. (5)
Air temperature T_a (°C)	10.0 – 30.0	(10*)15.0 – 35.0	(10*)15.0 – 35.0
Relative humidity RH (%)	–	10.0 – 80.0	10.0 – 80.0
Partial vapour pressure (P_a)	0 – 2700	–	–
Mean radiant temperature T_r (°C)	10.0 – 40.0	15.0 – 35.0	15.0 – 35.0
Air velocity V_a (m/s)	0 – 1	0.0 – 0.5 (1.5*)	0.0 – 0.5 (1.5*)
Clothing insulation I_{cl} (clo)	0.0 – 2.0	0.0 – 1.6 (2.4*)	0.0 – 1.6 (2.4*)
Metabolic rate M (Met)	0.8 – 4.0	0.8 – 2.0	0.8 – 2.0
Mean monthly outdoor temperature T_{out} (°C)	–	−15.0 – 35.0	−5.0 – 35.0
Model prediction	−3 to + 3	−3 to + 3	−3 to + 3

*Extended value reduces the precision.

health (Taylor, 2020). The models suggest suitability within 10–80% relative humidity (Table 6), but maintaining values between 40% and 60% is advisable.

Bias trends for **mean radiant temperature** parallel those of air temperature (Fig. 4). Model Eq. (4) is less accurate within 10–15 °C, excelling in higher mean radiant temperatures (>30 °C), where PMV tends to overestimate sensations. However, the new models tend to overestimate thermal sensations above 35 °C. For optimal model performance, use within 15–35 °C or 10–35 °C with slight precision reduction (Table 6) is recommended, still surpassing PMV. Beyond these ranges, model predictive capability wanes, heightening heat or cold stress likelihood.

Indoor environments typically feature limited **air velocity**, with most samples within 0.0 to 0.5 m/s range (Fig. 4). This accounts for Bias distribution's variability in this range, although models exhibit reduced variability. These models often overestimate sensations in 0.5–1.0 m/s and 1.5–2.0 m/s ranges. The 2.0–2.5 m/s range has few samples due to the rarity of such high velocities in daily life. Accuracy within 0.5–1.0 m/s slightly falls below ± 0.25 range. Generally, predictions within 0.0–0.5 m/s can be deemed accurate, extending to 0.0–1.5 m/s with minimal precision reduction (Table 6).

Clothing insulation significantly influences occupants' thermal sensation. The new models notably enhance predictive performance, particularly within the 1.2–2.0 clo range. Higher insulation values (>2 clo) yield relatively fewer samples (Fig. 4) as occupants indoors typically avoid excessively warm clothing. As such, applying the new models within the 0–1.6 clo range is recommended, extendable up to 2.4 clo with slightly increased uncertainty (Table 6).

Metabolic rate strongly impacts thermal sensation, but accurate estimation can be challenging, leading to consistent bias. Bias remains relatively low within 0.8–1.5 Met range (Fig. 4). Model Eq. (4) improves predictions even for higher Met values (1.5–2.5 Met), while model Eq. (5) underestimates sensations in 1.5–2.0 Met range. All models overestimate sensations for high metabolic rates (3.5–4.0 Met). Notably, the database includes limited samples with Met values > 2.0 Met, as typical indoor environments (offices, residential buildings, schools) rarely exhibit such high activity levels. Hence, applying the two models within 0.8–2 Met range is recommended (Table 6).

Mean monthly outdoor temperature is not a PMV input, but its inclusion in models of Eq. (4) and (5) accounts for its significance in thermal adaptation. Very low temperatures (-25 °C to −15 °C) result in poor predictive performance for all models, while accuracy improves for higher temperatures (especially −5°C to 35 °C) (Fig. 4). Limited sample size exists within the −15 °C to −5°C range, with model Eq. (5) exhibiting a tendency to overestimate sensations. Consequently, model Eq. (5) should be restricted to −5°C to 35 °C range, while model Eq. (4) can potentially extend to lower temperatures down to −15 °C (Table 6).

For precise thermal sensation prediction, free from systematic overestimation or underestimation, Bias was calculated for each predicted range. PMV and models Eq. (4) and (5) cover the −3 to + 3 TSV range, shown on the x-axis of Fig. 4. This allows measuring Bias between predicted and observed sensations when the model predicts specific sensations, illustrating model performance on the 7-point ASHRAE scale. Overall, the two new models demonstrate lower errors than PMV in practical conditions, despite being simplified and requiring fewer input parameters. On average, these models tend to reduce PMV's overestimation/underestimation for TSV < -1 or TSV > 1. However, it's important to acknowledge that values close to ± 3 might indicate heat or cold stress situations, demanding careful utilization of thermal comfort indices.

4. Discussion

4.1. Definition of the "comfort areas"

Enhancing predictive accuracy is crucial, yet not the sole determinant of efficacy. Comfort is better understood as a range rather than a fixed state, in line with adaptive principles (Lamberti et al., 2023). This section compares the "comfort zones" established by conventional adaptive models with those originating from the present models.

To derive the thermal comfort range it was assumed a relative humidity of 50%, a metabolic rate of 1.2 met, and a uniform environment ($T_a = T_r = T_{op}$, where T_{op} is the operative temperature), as recommended by international standards (EN 16798-1, 2019). These values were chosen as they are typical of many indoor environments, but the model can accommodate different conditions within the defined boundaries in Table 6.

Using these assumptions, the comfort temperatures were obtained for the models of Eq. (4) and (5), and the relationship between outdoor temperature (T_{out}) and comfort temperature (T_{comf}) was analysed through regression analysis for T_{out} ranging from 10 to 35 °C. This relationship is depicted by the solid line in Fig. 5. There is to notice that, based on adaptive thermal comfort research, achieving comfort for operative temperatures up to 35 °C is possible only in certain types of climates (Rawal et al., 2022), which were included in the development of the current models.

For models of Eq. (4) and (5) the acceptability ranges were calculated for 90% acceptability (Category I, dotted line), with the assumption that TSV=±0.5 and a Percentage of Dissatisfied (PD) equal to 10%, for 80% acceptability (Category II, dashed line) assuming TSV=±0.8 and PD = 20%, and for 70% acceptability (Category III, dashed and dotted line) assuming TSV=±1.1 and PD = 30% (ISO 7730, 2006).

Table 7 shows the ranges that define the comfort areas for Eqs. (4) and (5). It will be necessary to add and subtract respectively Δ_{UP} and Δ_{LOW} to the above equations to obtain the desired comfort ranges. It can be observed that the comfort temperature ranges for Models of Eq. (4) and (5) are larger than those given by the EN 16798–1 standard (EN, 2019).

This means that occupants have a significant adaptive capacity and can be in comfort situations over a wide temperature range. It should be noted, however, that it is appropriate to refer mainly to Categories I and II, as Category III leads to environmental conditions that may be tolerable within certain limits, but plausibly not conditions of comfort for a large proportion of individuals.

The relationship between T_{out} and T_{comf} for the model of Eq. (4) is represented by a curve (Fig. 5). It is possible to notice that the comfort temperature increases significantly with the outdoor temperature, while it remains stable for lower T_{out}. This is because at lower temperatures heating systems are usually provided while cooling systems not always are.

Thus, people present increased adaptation to higher outdoor temperatures, which can have significant implications also on energy consumption. This trend is in line with adaptive thermal comfort, which shows that below a certain outdoor temperature, the comfort

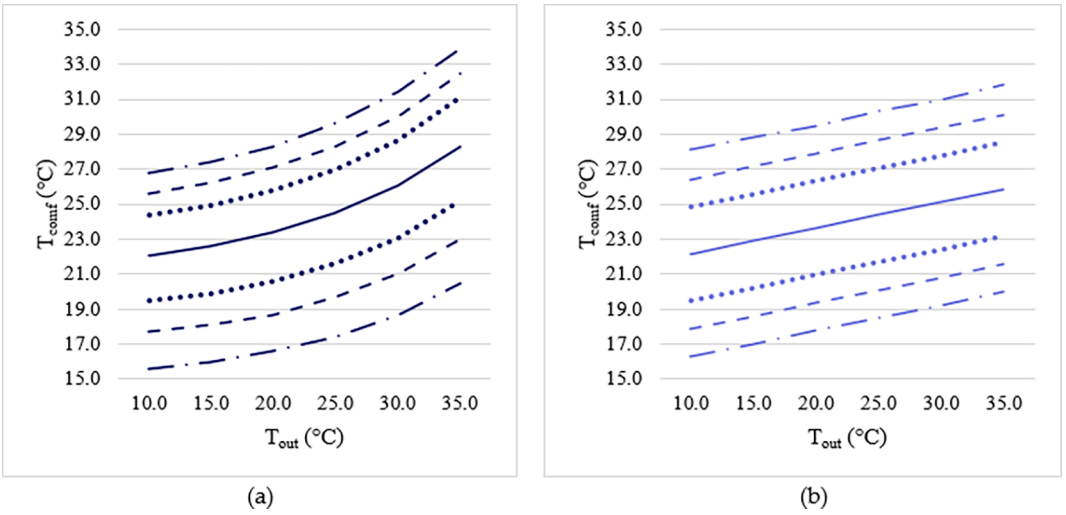

Fig. 5. Relationship between mean T_{out} and T_{comf} for the model of Eq. (4) (a), and model of Eq. (5) (b), assuming RH = 50%, M = 1.2 met, and $T_a = T_r = T_{op}$. Legend: comfort temperature (solid line), 90% acceptability range (dotted line), 80% acceptability range (dashed line), and 70% acceptability range (dashed and dotted line).

Table 7
Upper (Δ_{UP}), lower (Δ_{UP}), and total (Δ_{TOT}) ranges of comfort temperatures for the models of Eq. (4), (5), and EN16798-1 (EN, 2019) standard.

	Model Eq. (4)			**Model Eq. (5)**			**EN 16798–1**		
	Δ_{UP} (°C)	Δ_{LOW} (°C)	Δ_{TOT} (°C)	Δ_{UP} (°C)	Δ_{LOW} (°C)	Δ_{TOT} (°C)	Δ_{UP} (°C)	Δ_{LOW} (°C)	Δ_{TOT} (°C)
Category I	2.3	2.6	4.9	2.6	2.7	5.3	2.0	3.0	5.0
Category II	3.5	4.4	7.9	4.2	4.3	8.5	3.0	4.0	7.0
Category III	4.7	6.5	11.2	5.9	5.9	11.8	4.0	5.0	9.0

temperature remains constant (Humphreys et al., 2016). The comfort temperatures reached with the model of Eq. (4) are similar to the ones given by EN 16798–1 standard (EN, 2019) but with an increased comfort area and a non-linear relationship between T_{out} and T_{comf}.

The model of Eq. (5), like the original adaptive model (EN, 2019) presents instead a linear relationship between T_{out} and T_{comf} (Fig. 5) but the adaptation is lower and the slope is less steep. Concerning the comfort temperatures, they result lower than the ones given by standards (EN, 2019) especially for high T_{out}, while the range of comfort in model of Eq. (5) is again wider than the original adaptive model.

In comparison to the adaptive model of EN 16798–1 (EN, 2019), the two models are favourable as they can be used not only for the fixed environmental and individual parameters (low metabolic rate, fixed range of humidity, etc.) but within the acceptability range discussed in the previous section. Furthermore, the increased range of comfort temperatures has implications on energy consumption, even if the use of the model of Eq. (4) is more appropriate to this aim.

4.2. Practical applications of the new models

Since these models were developed for actual use in building management, it is necessary to understand how much the possible error in the prediction of the thermal sensation of the two models (Table 5) can affect thermal comfort. To this aim, the temperature variation associated with an MAE of 0.866 (Model of Eq. (4)) and 0.872 (Model of Eq. (5)) must be investigated.

To quantify the error in the prediction of TSV in terms of temperature, Griffiths' method was used. (Griffiths, 1990). According to this method, the comfort temperature can be calculated as follows (Eq. (6)):

$$T_{comf} = T_{op} + \frac{TSV}{G} \quad (6)$$

where T_{comf} is the comfort temperature, and G is the Griffiths' constant ($°C^{-1}$), assumed equal to 0.50 (Humphreys et al., 2013). Although recent studies show that the Griffiths constant can also be a variable, this value was chosen because it was estimated from the data available in the ASHRAE databases, which were also used for the development of the new models.

Assuming that T_{comf} is equal to T_{op} (i.e. occupants in comfortable conditions), a MAE of 0.866 means that the model predicts a TSV equal to this value and thus an error in T_{comf} of 1.73 °C. In the case of the model in Eq. (5) (MAE = 0.872), the error in the prediction of T_{comf} is 1.74 °C.

Comparing these results with Table 7, the error of about 1.7 °C is below the 90% acceptable range corresponding to Category I. This means that the error of the new models is, on a practical level, acceptable, as it does not actually lead to a reduction in users' comfort.

In general, these effective models are valuable for thermal comfort prediction. When real-time data on mean radiant temperature and metabolic rate are available, Eq. (4) is recommended. Alternatively, Eq. (5) predicts accurate thermal sensation using only indoor air temperature, relative humidity, and monthly mean outdoor temperature.

4.3. Limitations and future studies

Although the two new models were developed using a comprehensive database, they present some limitations. Indeed, data were mostly collected in the comfort area of Thermal Sensation Vote (TSV) and few cases reported TSV out of the range ± 2. For this reason, models of Eq. (4) and (5) perform better when discomfort is not too high, and in the case of extreme conditions indices for heat or cold stress should be adopted. Future studies should select a wider range of input parameters, especially for the metabolic rate that remained in a restricted range, despite a large amount of data.

Furthermore, for the development of the model, a database comprehensive of a wide range of building types, operation modes, and climates was used. However, the perception of thermal comfort can be

also influenced by the background and the level of expectations of the users, which can be a function of the previously mentioned issues. Although predictive ability has been shown to be satisfactory, future studies should focus on the analysis of different building types, operation modes, expectation levels, etc.

Finally, future research should be focused on assessing not only the thermal sensation, but also preference and acceptability, and to consider possible diversities among them.

Despite these limitations, the proposed models constitute a useful and user-friendly approach that improves the prediction if compared to previous thermal comfort models and can be used by researchers and practitioners to assess thermal comfort and possibly reduce the associated energy consumption.

5. Conclusions

In this study, two original data-driven models to predict the thermal sensation of buildings' occupants were developed, using a wide sample of field studies from ASHRAE databases, which allow to account for thermal adaptation. The purpose was to develop a data-driven model that (i) includes the adaptive processes, (ii) selects and includes the most relevant parameters in thermal perception, (iii) can be practically used for real-time control, and (iv) presents a good predictive performance.

The first point was addressed by developing the models on large databases based on field studies, which allows to account for adaptation.

The second point was fulfilled by having included in the analysis all the parameters that are considered to be involved in the perception of the thermal environment, which are air temperature (T_a), mean radiant temperature (T_r), relative humidity (RH), air velocity (V_a), clothing insulation (I_{cl}), metabolic rate (M) and outdoor monthly air temperature (T_{out}). These parameters were combined in polynomials and used to fit a Lasso regression to define which subset of combinations has the highest predictive power concerning the thermal sensation in real case studies and therefore select the parameters that most influence it.

The third point was considered by providing easy-to-use equations to define the thermal sensation in buildings, namely Eq. (4), and (5).

Finally, the new models improved the predictive performance when compared to simpler regression models or the typically used PMV, as shown in Table 5. In particular, the predictive performance of the model of Eq. (5) is slightly reduced (MAE = 0.872, Bias = 0.001, RMSE = 1.136) if compared to Eq. (4) (MAE = 0.866, RMSE = 1.128, Bias = 0.008), but it is significantly lower than the PMV.

The two new models present an increased performance if compared to simple linear models and to Fanger's PMV, providing a model that can be easily applied by practitioners and that takes into account influential parameters in the perception of comfort. Furthermore, in comparison to the original adaptive model, these new models can also be used for variable environmental and individual parameters.

With these models it was also possible to define "comfort areas", taking into account an occupant acceptability of 90%, 80%, and 70% (Categories I, II, III respectively). In addition, by deriving the acceptability ranges for conditions typically found indoors (RH = 50%, M = 1.2 met, $T_a = T_r = T_{op}$), it was shown that the comfort ranges found for the new models are wider than the adaptive ones, especially for Categories II and III (by about 1 °C for Category II and 2 °C for Category III), resulting in energy savings in their use.

In conclusion, leveraging extensive comfort databases enabled the development of two improved models to predict thermal sensation, accounting for multiple influencing factors, that surpass the accuracy of commonly employed ones. These models could be integrated into an updated database version, offering calculated comfort indices aligned with respondents' subjective thermal perception.

Funding

This research was partially funded by University of Pisa, within the program PRA 2022, competitive call for the financial support to multidisciplinary research projects developed at the university Pisa (ID research project PRA_2022_70).

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.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.totert.2023.100083.

References

Arnaiz-González, Á., Díez-Pastor, J.F., Rodríguez, J.J., García-Osorio, C., 2016. Instance selection for regression: Adapting DROP. Neurocomputing 201, 66–81. https://doi.org/10.1016/j.neucom.2016.04.003.

Arnaiz-González, Á., Díez-Pastor, J.F., Rodríguez, J.J., García-Osorio, C.I., 2016. Instance selection for regression by discretization. Expert Systems with Applications 54, 340–350. https://doi.org/10.1016/j.eswa.2015.12.046.

Brager, G.S., de Dear, R.J., 1998. thermal adaptation in the built environment: a literature review. Energy and Buildings 27, 83–96. https://doi.org/10.1016/S0378-7788(97)00053-4.

Castilla, N., Llinares, C., Bisegna, F., Blanca-Giménez, V., 2018. Affective evaluation of the luminous environment in university classrooms. Journal of Environmental Psychology 58, 52–62. https://doi.org/10.1016/j.jenvp.2018.07.010.

Chaudhuri, T., Zhai, D., Soh, Y.C., Li, H., Xie, L., 2018. Thermal comfort prediction using normalized skin temperature in a uniform built environment. Energy and Buildings 159, 426–440. https://doi.org/10.1016/j.enbuild.2017.10.098.

Chaudhuri, T., Soh, Y.C., Li, H., Xie, L., 2019. A feedforward neural network based indoor-climate control framework for thermal comfort and energy saving in buildings. Applied Energy 248, 44–53. https://doi.org/10.1016/j.apenergy.2019.04.065.

Cheung, T., Schiavon, S., Parkinson, T., Li, P., Brager, G., 2019. Analysis of the accuracy on PMV – PPD model using the ASHRAE global thermal comfort database II. Building and Environment 153, 205–217. https://doi.org/10.1016/j.buildenv.2019.01.055.

Cosma, A.C., Simha, R., 2019. Machine learning method for real-time non-invasive prediction of individual thermal preference in transient conditions. Building and Environment 148, 372–383. https://doi.org/10.1016/j.buildenv.2018.11.017.

Dai, C., Zhang, H., Arens, E., Lian, Z., 2017. Machine learning approaches to predict thermal demands using skin temperatures: Steady-state conditions. Building and Environment 114, 1–10. https://doi.org/10.1016/j.buildenv.2016.12.005.

de Dear, R., 1998. Global database of thermal comfort field experiments. ASHRAE Transactions 104, 1141–1152.

Djamila, H., 2017. Indoor thermal comfort predictions: Selected issues and trends. Renewable and Sustainable Energy Reviews 74, 569–580. https://doi.org/10.1016/j.rser.2017.02.076.

Djongyang, N., Tchinda, R., Njomo, D., 2010. Thermal comfort: A review paper. Renewable and Sustainable Energy Reviews 14, 2626–2640. https://doi.org/10.1016/j.rser.2010.07.040.

Du, C., Li, B., Liu, H., Ji, Y., Yao, R., Yu, W., 2019. Quantification of personal thermal comfort with localized airflow system based on sensitivity analysis and classification tree model. Energy and Buildings 194, 1–11. https://doi.org/10.1016/j.enbuild.2019.04.010.

EN 16798-1, Energy performance of buildings - Ventilation for buildings - Part 1: Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics, 2019.

Enescu, D., 2017. A review of thermal comfort models and indicators for indoor environments. Renewable and Sustainable Energy Reviews 79, 1353–1379. https://doi.org/10.1016/j.rser.2017.05.175.

Fanger, P.O., 1970. Thermal comfort. Analysis and applications in environmental engineering. Danish Technical Press, McGraw-Hill, Copenhagen.

Fanger, P., Toftum, J., 2002. Extension of the PMV model to non-air-conditioned buildings in warm climates. Spec. Issue Therm. Comf. Stand. 34, 533–536. https://doi.org/10.1016/S0378-7788(02)00003-8.

Fantozzi, F., Rocca, M., 2020. An extensive collection of evaluation indicators to assess occupants' health and comfort in indoor environment. Atmosphere. 11, 90. https://doi.org/10.3390/atmos11010090.

Farhan, A.A., Pattipati, K., Wang, B., Luh, P., Predicting individual thermal comfort using machine learning algorithms, in,, 2015. IEEE int. Conf. Autom. Sci. Eng. CASE 2015, 708–713. https://doi.org/10.1109/CoASE.2015.7294164.

Földváry Ličina, V., Cheung, T., Zhang, H., de Dear, R., Parkinson, T., Arens, E., Chun, C., Schiavon, S., Luo, M., Brager, G., Li, P., Kaam, S., Adebamowo, M.A., Andamon, M. M., Babich, F., Bouden, C., Bukovianska, H., Candido, C., Cao, B., Carlucci, S., Cheong, D.K.W., Choi, J.-H., Cook, M., Cropper, P., Deuble, M., Heidari, S., Indraganti, M., Jin, Q., Kim, H., Kim, J., Konis, K., Singh, M.K., Kwok, A., Lamberts, R., Loveday, D., Langevin, J., Manu, S., Moosmann, C., Nicol, F., Ooka, R., Oseland, N.A., Pagliano, L., Petráš, D., Rawal, R., Romero, R., Rijal, H.B., Sekhar, C., Schweiker, M., Tartarini, F., Tanabe, S., Tham, K.W., Teli, D., Toftum, J., Toledo, L., Tsuzuki, K., De Vecchi, R., Wagner, A., Wang, Z., Wallbaum, H., Webb, L., Yang, L., Zhu, Y., Zhai, Y., Zhang, Y., Zhou, X., 2018. Development of the ASHRAE global thermal comfort database II. Building and Environment 142, 502–512. https://doi.org/10.1016/j.buildenv.2018.06.022.

G. Gao, J. Li, Y. Wen, Energy-Efficient Thermal Comfort Control in Smart Buildings via Deep Reinforcement Learning, (2019). arXiv:1901.04693.

Griffiths, I., 1990. Thermal comfort studies in buildings with passive solar features. Field studies.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12, 55–67. https://doi.org/10.1080/00401706.1970.10488634.

M. Humphreys, F. Nicol, R. S., Adaptive Thermal Comfort: Foundations and Analysis, Routledge, London, 2016. https://doi.org/10.4324/9781315765815.

Humphreys, M., Nicol, F., 2002. The validity of ISO-PMV for predicting comfort votes in every-day thermal environments. Energy and Buildings 34, 667–684. https://doi.org/10.1016/S0378-7788(02)00018-X.

Humphreys, M.A., Rijal, H.B., Nicol, J.F., 2013. Updating the adaptive relation between climate and comfort indoors; new insights and an extended database. Building and Environment 63, 40–55. https://doi.org/10.1016/j.buildenv.2013.01.024.

ISO 7726, Ergonomics of the thermal environment – Instruments for measuring physical quantities, 2001.

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, 2006.

Jiang, L., Yao, R., 2016. Modelling personal thermal sensations using c-Support vector classification (C-SVC) algorithm. Building and Environment 99, 98–106. https://doi.org/10.1016/j.buildenv.2016.01.022.

Kim, J., Zhou, Y., Schiavon, S., Raftery, P., Brager, G., 2018. Personal comfort models: Predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning. Building and Environment 129, 96–106. https://doi.org/10.1016/j.buildenv.2017.12.011.

Kumar, S., 2022. Subject's thermal adaptation in different built environments: an analysis of updated metadata-base of thermal comfort data in india. J. Build. Eng. 46, 103844 https://doi.org/10.1016/j.jobe.2021.103844.

Lamberti, G., Leccese, F., Salvadori, G., Contrada, F., Kindinis, A., 2023. investigating the effects of climate on thermal adaptation: A comparative field study in naturally ventilated university classrooms. Energy and Buildings 294, 113227. https://doi.org/10.1016/j.enbuild.2023.113227.

Lamberti, G., Thermal comfort in the built environment: current solutions and future expectations, in,, 2020. IEEE int. Conf. EEEIC ICPS Eur. 2020, 1–6. https://doi.org/10.1109/EEEIC/ICPSEurope49358.2020.9160558.

Lee, S., Bilionis, I., Karava, P., Tzempelikos, A., 2017. A bayesian approach for probabilistic classification and inference of occupant thermal preferences in office buildings. Building and Environment 118, 323–343. https://doi.org/10.1016/j.buildenv.2017.03.009.

Li, D., Menassa, C.C., Kamat, V.R., 2017. Personalized human comfort in indoor building environments under diverse conditioning modes. Building and Environment 126, 304–317. https://doi.org/10.1016/j.buildenv.2017.10.004.

Lu, S., Wang, W., Lin, C., Hameen, E.C., 2019. Data-driven simulation of a thermal comfort-based temperature set-point control with ASHRAE RP884. Building and Environment 156, 137–146. https://doi.org/10.1016/j.buildenv.2019.03.010.

Luo, M., Xie, J., Yan, Y., Ke, Z., Yu, P., Wang, Z., Zhang, J., 2020. Comparing machine learning algorithms in predicting thermal sensation using ASHRAE comfort database II. Energy and Buildings 210, 109776. https://doi.org/10.1016/j.enbuild.2020.109776.

McCartney, K.J., Nicol, J.F., 2002. Developing an adaptive control algorithm for europe. Spec. Issue Therm. Comf. Stand. 34, 623–635. https://doi.org/10.1016/S0378-7788(02)00013-0.

Nielsen, F., 2016. Introduction to HPC with MPI for data science. In: Hierarchical Clust. Springer International Publishing, Cham, pp. 195–211.

Oliveri, M., Pennisi, S., Scaccianoce, G., Vaccaro, V., Bisegna, F., The relavance of indoor comfort in the process of prisoners' rehabilitation: A case study, in,, 2016. IEEE 16th int. Conf. Environ. Electr. Eng. EEEIC 2016, 1–6. https://doi.org/10.1109/EEEIC.2016.7555673.

Olvera-López, J.A., Carrasco-Ochoa, J.A., Martínez-Trinidad, J.F., Kittler, J., 2010. A review of instance selection methods. Artificial Intelligence Review 34, 133–143. https://doi.org/10.1007/s10462-010-9165-y.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J.T., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine learning in python. Journal of Machine Learning Research 12, 2825–2830.

Rana, R., Kusy, B., Jurdak, R., Wall, J., Hu, W., 2013. Feasibility analysis of using humidex as an indoor thermal comfort predictor. Energy and Buildings 64, 17–25. https://doi.org/10.1016/j.enbuild.2013.04.019.

Rawal, R., Shukla, Y., Vardhan, V., Asrani, S., Schweiker, M., de Dear, R., Garg, V., Mathur, J., Prakash, S., Diddi, S., Ranjan, S.V., Siddiqui, A.N., Somani, G., 2022. Adaptive thermal comfort model based on field studies in five climate zones across india. Building and Environment 219, 109187. https://doi.org/10.1016/j.buildenv.2022.109187.

Rupp, R.F., Kim, J., Ghisi, E., de Dear, R., 2019. thermal sensitivity of occupants in different building typologies: the griffiths constant is a variable. Energy and Buildings 200, 11–20. https://doi.org/10.1016/j.enbuild.2019.07.048.

Song, Y., Liang, J., Lu, J., Zhao, X., 2017. An efficient instance selection algorithm for k nearest neighbor regression. Neurocomputing 251, 26–34. https://doi.org/10.1016/j.neucom.2017.04.018.

Tartarini, F., Schiavon, S., 2020. Pythermalcomfort: A python package for thermal comfort research. SoftwareX. 12, 100578 https://doi.org/10.1016/j.softx.2020.100578.

S. Taylor, Putting People First: The Healing Power of Indoor Air, (2020). https://www.ashrae.org/file%20library/technical%20resources/covid-19/seminar-21_presentation-1.pdf.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B-Methodol. 58, 267–288.

Wang, Z., Li, A., Ren, J., He, Y., 2014. thermal adaptation and thermal environment in university classrooms and offices in harbin. Energy and Buildings 77, 192–196. https://doi.org/10.1016/j.enbuild.2014.03.054.

Wang, Z., Yu, H., Luo, M., Wang, Z., Zhang, H., Jiao, Y., 2019. Predicting older people's thermal sensation in building environment through a machine learning approach: Modelling, interpretation, and application. Building and Environment 161, 106231. https://doi.org/10.1016/j.buildenv.2019.106231.

Wu, Z., Li, N., Peng, J., Cui, H., Liu, P., Li, H., Li, X., 2018. Using an ensemble machine learning methodology-Bagging to predict occupants' thermal comfort in buildings. Energy and Buildings 173, 117–127. https://doi.org/10.1016/j.enbuild.2018.05.031.

Xie, J., Li, H., Li, C., Zhang, J., Luo, M., 2020. Review on occupant-centric thermal comfort sensing, predicting, and controlling. Energy and Buildings 226, 110392. https://doi.org/10.1016/j.enbuild.2020.110392.

Zhang, F., de Dear, R., Hancock, P., 2019. Effects of moderate thermal environments on cognitive performance: A multidisciplinary review. Applied Energy 236, 760–777. https://doi.org/10.1016/j.apenergy.2018.12.005.

Zhao, Q., Zhao, Y., Wang, F., Wang, J., Jiang, Y., Zhang, F., 2014. A data-driven method to describe the personalized dynamic thermal comfort in ordinary office environment: from model to application. Building and Environment 72, 309–318. https://doi.org/10.1016/j.buildenv.2013.11.008.

Zhou, X., Xu, L., Zhang, J., Niu, B., Luo, M., Zhou, G., Zhang, X., 2020. Data-driven thermal comfort model via support vector machine algorithms: Insights from ASHRAE RP-884 database. Energy and Buildings 211, 109795. https://doi.org/10.1016/j.enbuild.2020.109795.

Zinzi, M., Pagliaro, F., Agnoli, S., Bisegna, F., Iatauro, D., 2021. On the Built-Environment quality in nearly Zero-Energy renovated schools: Assessment and impact of passive strategies. Energies 14. https://doi.org/10.3390/en14102799.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net (vol B 67, pg 301. Journal of the Royal Statistical Society, Series B 67 (2005), 768. https://doi.org/10.1111/j.1467-9868.2005.00527.x.

DEVELOPMENT AND COMPARISON OF ADAPTIVE DATA-DRIVEN MODELS FOR THERMAL COMFORT ASSESSMENT AND CONTROL

Giulia Lamberti[a,b], Roberto Boghetti[c], Jérôme Kämpf[c], Fabio Fantozzi[a], Francesco Leccese[a], Giacomo Salvadori[a]

a. University of Pisa, School of Engineering, Largo Lucio Lazzarino, 56122 Pisa, Italy
b. Institut de Recherche en Constructibilité, Université Paris-Est, ESTP, 28, Avenue du Président Wilson, 94230 Cachan, France
c. Energy Informatics Group, Idiap Research Institute, 1920 Martigny, Switzerland

Abstract

Thermal comfort prediction is an important issue, as it can largely influence occupants' well-being and buildings' energy consumption. Nowadays, models used to assess thermal comfort have been increasingly discussed, and a growing number of data-driven models with several input parameters developed. Although these models allow reasonably accurate predictions of thermal comfort, using complex algorithms to determine thermal comfort might be unsuitable for some use cases, such as quick estimations or real-time control of Heating Ventilation, and Air Conditioning (HVAC) systems.

In this paper, a data-driven model was developed based on 61710 samples of subjective responses associated with environmental parameters from field studies available in two ASHRAE databases. Two models resulted from this analysis, one with higher accuracy and one simplified, which improved the prediction in comparison to other regression models and PMV.

However, since thermal comfort cannot be conceived as a punctual condition, comfort areas were derived, i.e., respective comfort ranges at 90%, 80%, and 70% of thermal acceptability. The result is that the error in the prediction of the new models is below the 90% acceptable range, which means that the models' error does not lead to a reduction in the evaluation of occupant comfort.

Built upon influential parameters, these models enable thermal comfort estimates and occupant-cantered HVAC control. The notion of comfort as a non-fixed state empowers more flexible building management criteria, reducing energy use while upholding indoor comfort.

Keywords

Thermal comfort – Data-driven model – PMV – Adaptive thermal comfort – ASHRAE database

Supplementary materials

Data analysis

In this section, the data obtained from the filtration to remove the outliers are analysed. The acceptability ranges considered were:

Air temperature:	$T_a \geq 10°C$,
Relative humidity:	$10\% \geq RH \geq 90\%$,
Mean radiant temperature:	$T_r \geq 10°C$,
Air velocity:	0 m/s $\geq V_a \geq 3$ m/s,
Clothing insulation:	$I_{cl} \geq 0$ clo,
Metabolic rate:	$M \geq 0.8$ met,
Predicted Mean Vote:	$-3 \geq PMV \geq +3$,
Thermal Sensation Vote:	$-3 \geq TSV \geq +3$.

Figure S1 shows the number of samples per publication included in the analysis and Figure S2 the year distribution of the studies. Studies were also classified per type of building (Figure S3), climate zone (Figure S3), operation mode (Figure S4), and season (Figure S4), as these aspects can influence the perception of the thermal environment.

To understand the variability of the parameters considered, the distribution of air temperature, globe temperature, mean radiant temperature, mean monthly outdoor temperature (Figure S5), relative humidity (Figure S6), air velocity, metabolic rate, and clothing insulation (Figure S7) is reported. Figure S8 shows the distribution of the calculated PMV, clipped PMV (PMV was clipped for values between -3 and +3 to compare it with the TSV), and for the TSV obtained from the subjective responses. The scatter matrix, showing the correlation between the different parameters is reported in Figure S9.

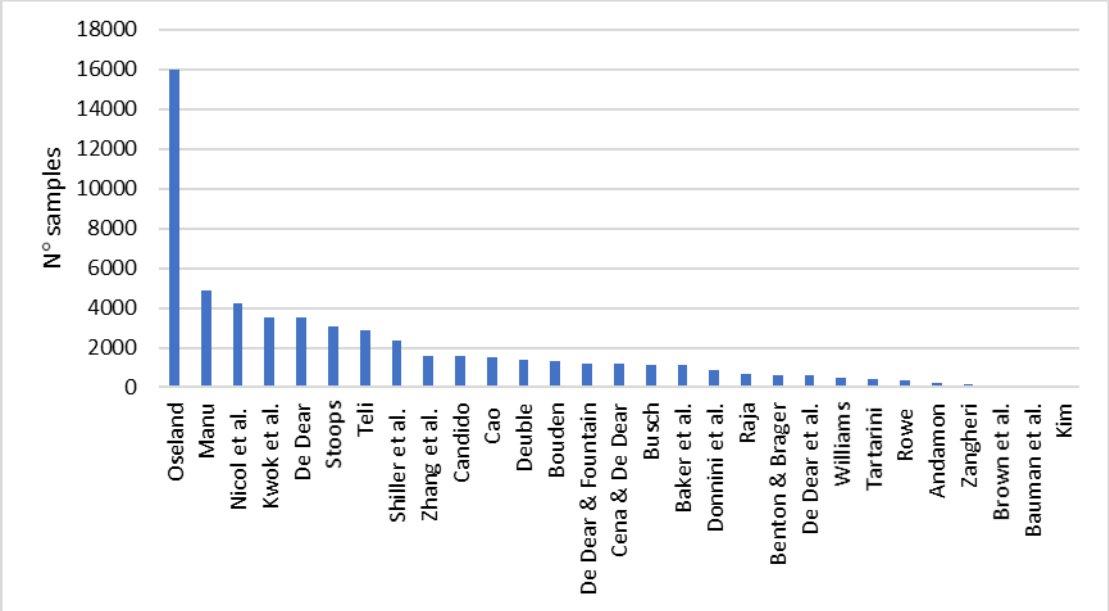

Figure S1. Number of samples per publication.

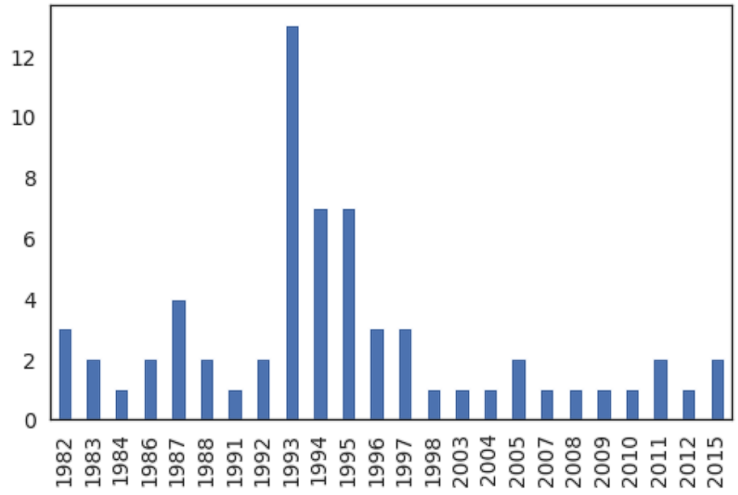

Figure S2. Year distribution of the studies.

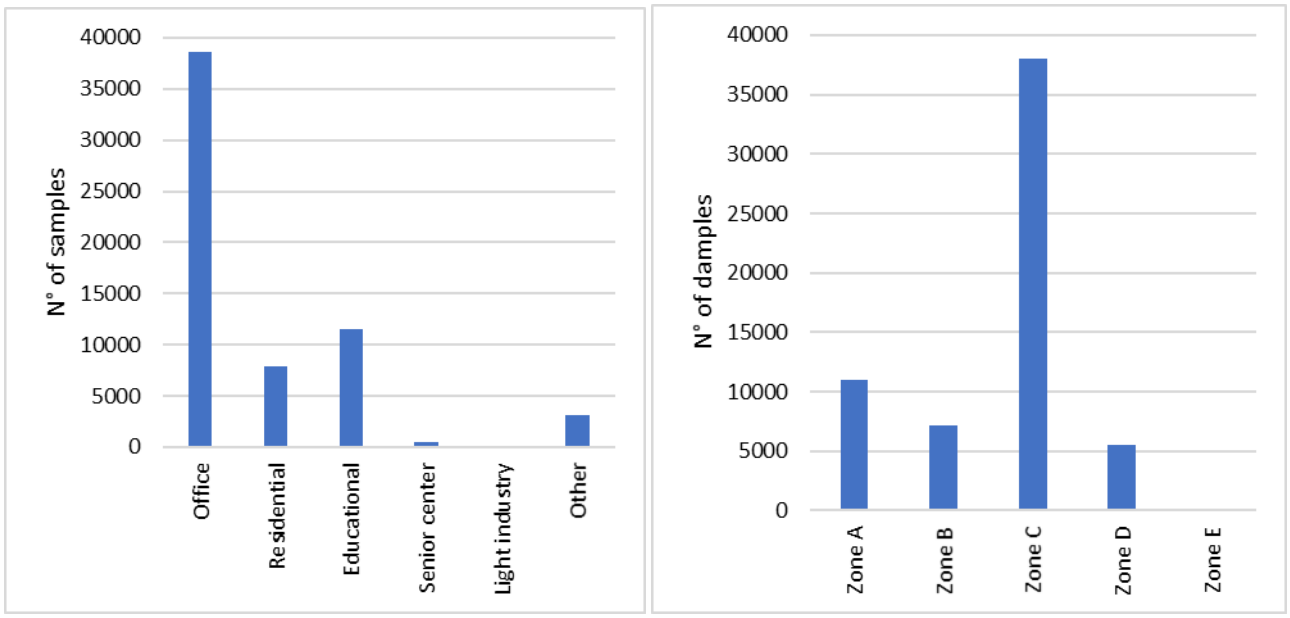

Figure S3. Number of samples for type of building (left) and climate zone (right).

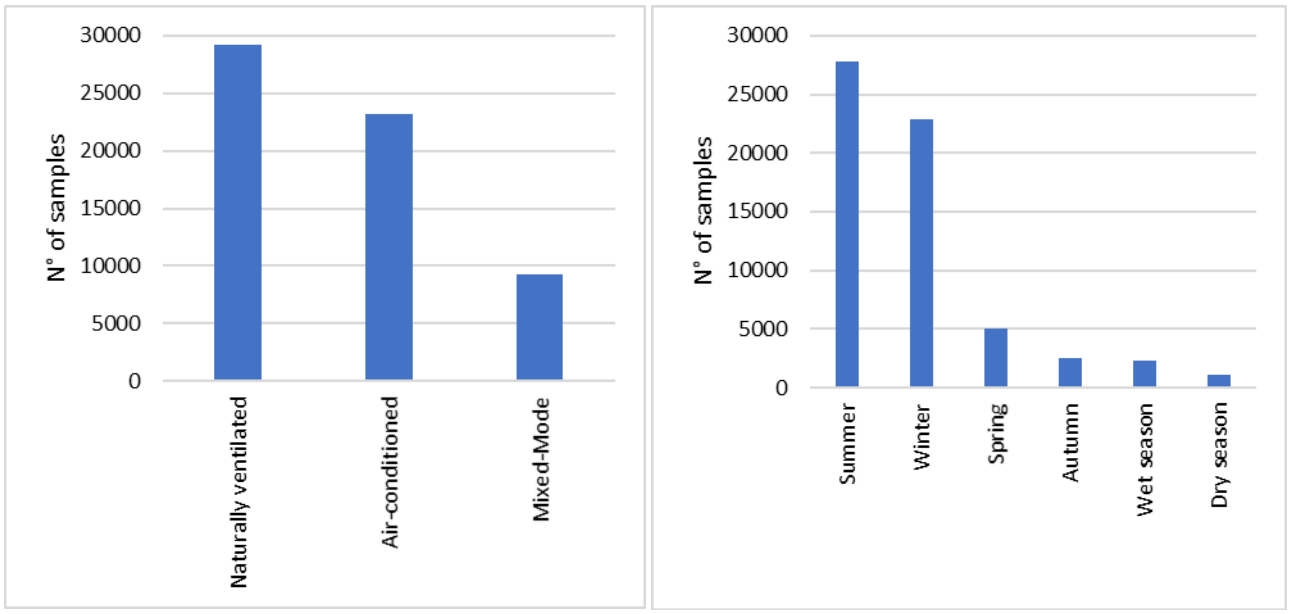

Figure S4. Number of samples for operation mode (left) and season (right).

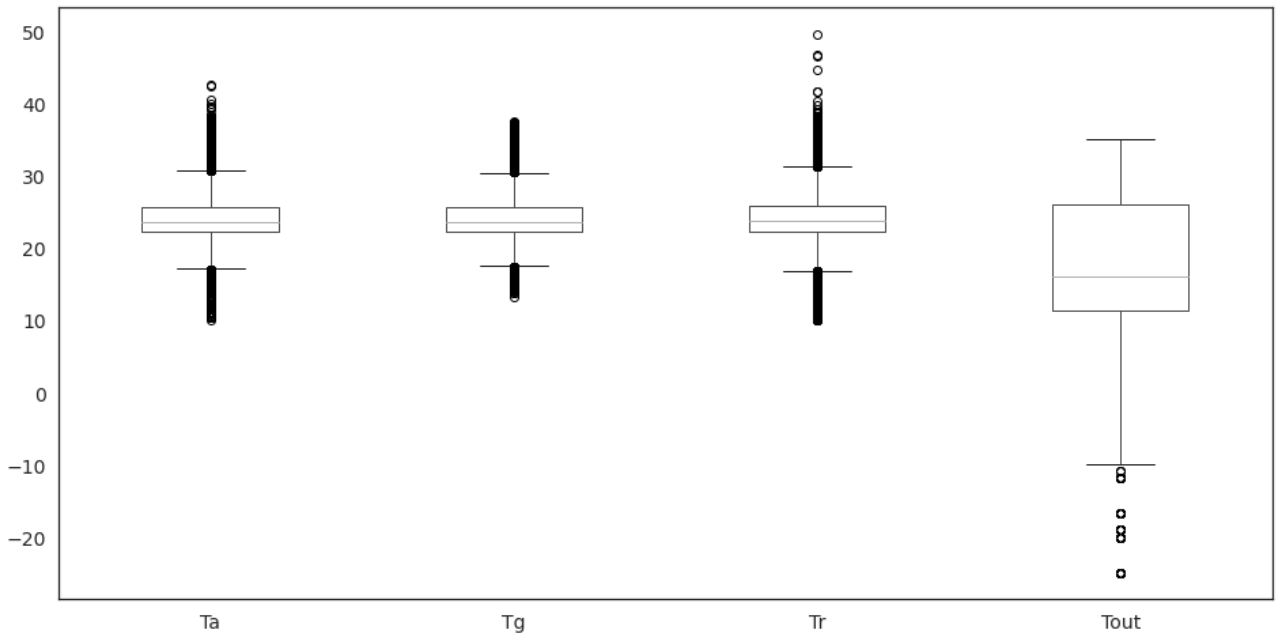

Figure S5. Distribution of the air temperature (T_a), globe temperature (T_g), mean radiant temperature (T_r), and mean monthly outdoor temperature (T_{out}).

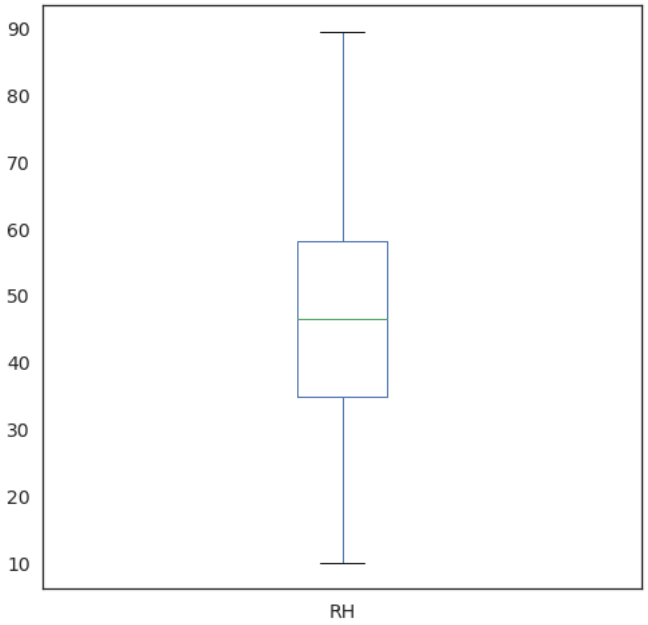

Figure S6. Distribution of the relative humidity (RH).

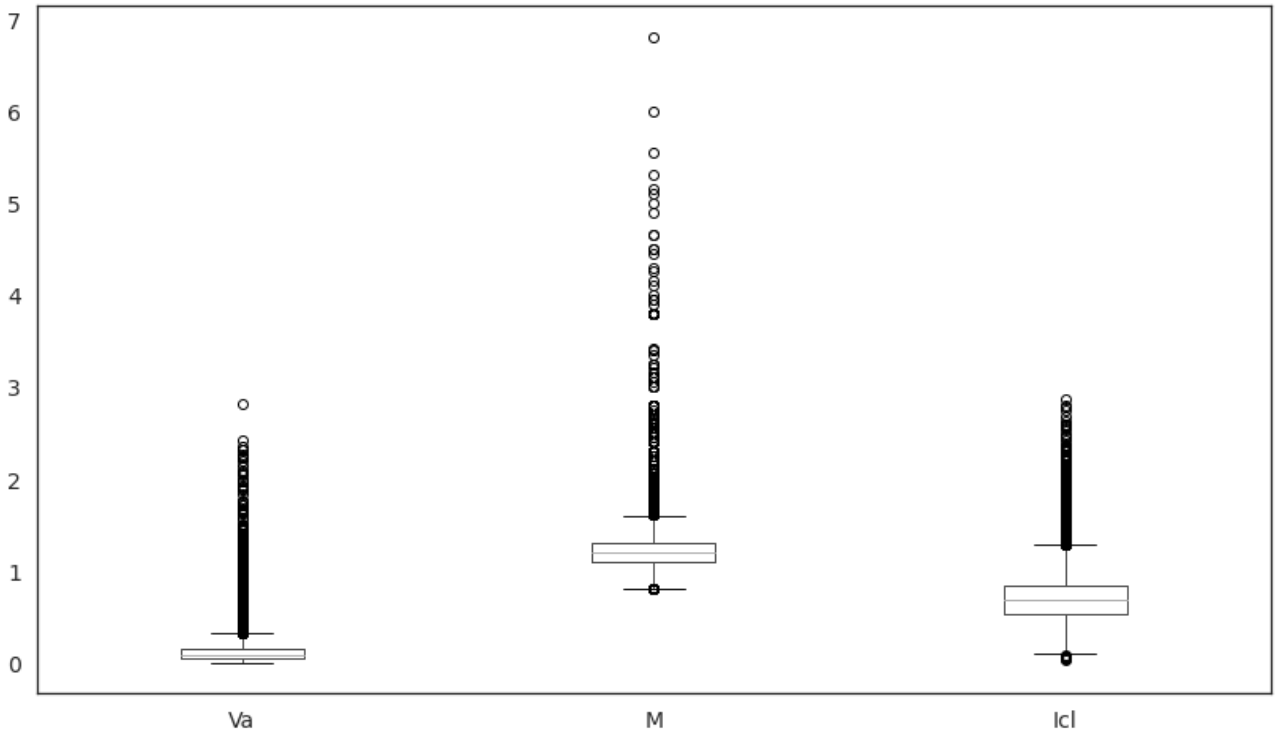

Figure S7. Distribution of the air velocity (V_a), metabolic rate (M), and clothing insulation (I_{cl}).

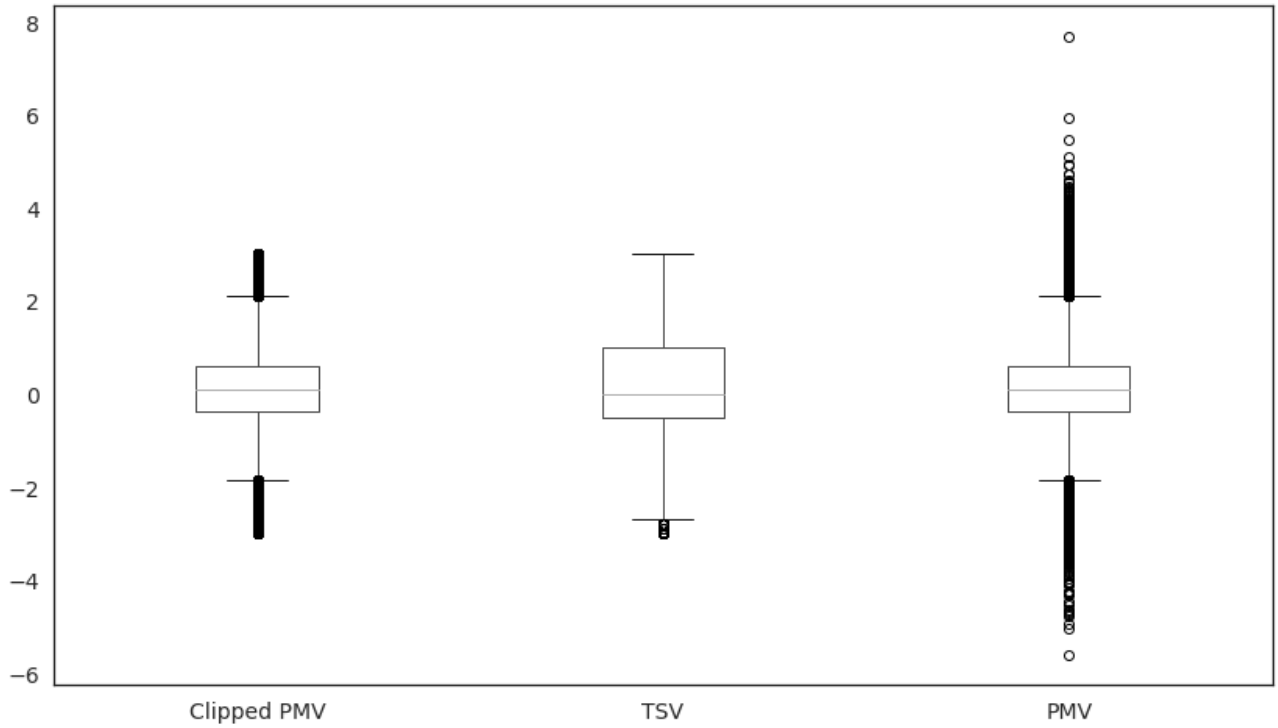

Figure S8. Distribution of the PMV (before and after being clipped) and of the Thermal Sensation Vote (TSV).

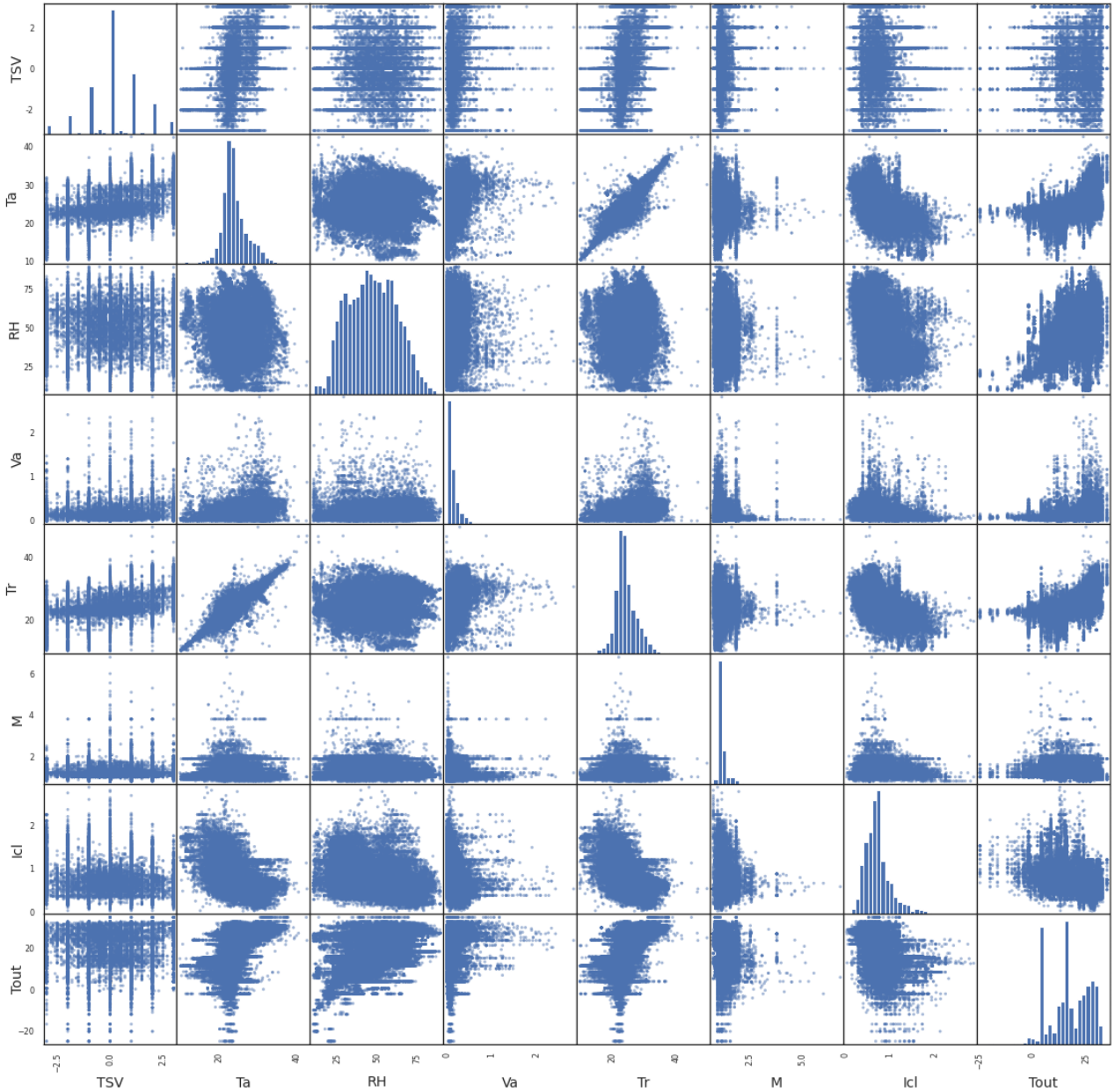

Figure S9. Scatter matrix of the considered parameters