Applied Energy 325 (2022) 119803

Contents lists available at ScienceDirect

Applied Energy

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

Human-in-the-loop methods for occupant-centric building design and operation

Matteo Favero [a,*], Jan Kloppenborg Møller [b], Davide Calì [b], Salvatore Carlucci [c]

[a] Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, Trondheim, Norway
[b] Department of Applied Mathematics and Computer Science, Technical University of Denmark, DTU, Lyngby, Denmark
[c] Energy, Environment and Water Research Center, The Cyprus Institute, Nicosia, Cyprus

HIGHLIGHTS

- Multilevel modelling was applied to predict subjects' thermal preference vote in a dynamic thermal environment.
- The beta and ordinal mixed-effects models are both valid alternatives for modelling subjects' thermal preference votes.
- Two procedures were used to implement subjects' feedback within the occupant-centric building design and operation paradigm.
- The population-averaged procedure is suitable for the building design phase.
- The cluster-specific procedure is appropriate for the building operation phase.

ARTICLE INFO

Keywords:
Subjective thermal preference
Occupant-centric design and operation
Human-in-the-loop
Multilevel modelling
Statistical thinking

ABSTRACT

A comfortable indoor environment should be one of the main services buildings provide. However, technical building systems are typically designed and operated according to fixed set-point temperatures determined by the 'one-size-fits-all' principle assuming universal thermal comfort requirements, which has been questioned in the last fifty years. Designing and implementing comfortable set-point modulations that consider occupant feedback would be beneficial in terms of increasing comfort, potentially reduce energy consumption and significantly support the clean energy transition. An exploratory study aimed at predicting the thermal preferences of human subjects exposed to a dynamic thermal environment is presented. Using data acquired from a laboratory experiment where subjects were exposed to precisely controlled thermal ramps in an 'office-like' climatic chamber, cluster-specific and population-averaged methods are designed to handle the group-level residual during the prediction of the thermal preference votes. The results show that both approaches are valid strategies for modelling thermal preference votes and are effective in supporting a concrete occupant-centric building design and the building's operation. Furthermore, the population-averaged approach is suitable for the occupant-centric building design phase, where the target is an 'average' occupant. The cluster-specific method is best suited to meet the needs of a specific occupant and is suitable for implementation in the operational phase of the building.

1. Introduction

A comfortable indoor environment should be one of the primary services buildings provide. Nowadays, all thermal comfort standards include recommendations concerning the indoor thermal conditions for both the design and operation phases of buildings. Currently, the most frequently cited thermal comfort standards, namely ASHRAE 55:2020 [1], ISO 7730:2005 [2] and EN 16798–1:2019 [3], which was formerly EN 15251:2007 [4], propose requirements based on Fanger model (beyond also including other approaches), which solves the heat balance equations between the human body and its surroundings, represented as a uniform environment. Fanger defined the 'Predicted Mean Vote' (PMV) as an index that predicts the mean thermal sensation vote on a standard scale for a large group of persons exposed to a given combination of metabolic activity level, clothing insulation and four thermal environmental variables characterising the indoor space: dry-bulb air temperature, mean radiant temperature, air velocity and relative

* Corresponding author at: 7a Høgskoleringen, 7034 Trondheim, Norway.
E-mail address: matteo.favero@ntnu.no (M. Favero).

https://doi.org/10.1016/j.apenergy.2022.119803
Received 29 April 2022; Received in revised form 3 August 2022; Accepted 5 August 2022
Available online 20 August 2022

Nomenclature

$\boldsymbol{I}$	The identity matrix
k	The category of the dependent variable
M	The total number of the simulated random effect
n	The number of events
$\boldsymbol{u}$	The vector of the random effects
$\mathbf{X}$	The design matrix of the fixed effects
$\mathbf{x}$	The vector of the fixed effects
x	Indicates a generic variable
$\mathbf{Y}$	The vector or matrix of the response variable
Y	Indicates a random value of the response variable (usually accompanied by a subscript)
$\boldsymbol{Z}$	The design matrix of the random effects
$\boldsymbol{\beta}$	The vector of parameters of the fixed effects
β	A scalar indicating a parameter of the fixed effects (usually accompanied by a subscript)
$\boldsymbol{\gamma}$	The vector of the cumulative probabilities
γ	A scalar indicating a cumalive probability (usually accompanied by a subscript)
$\boldsymbol{\eta}$	The vector of the linear predictor term
η	A scalar indicating the linear predictor term (usually accompanied by a subscript)
$\boldsymbol{\mu}$	The vector of the expected values
μ	A scalar indicating the expectd value (usually accompanied by a subscript)
$\boldsymbol{\pi}$	The vector of probabilities
π	A scalar indicating a probability (usually accompanied by a subscript)
σ^2	The variance
σ	The standard deviation
Σ	The variance–covariance matrix
τ	The latent threshold parameter
ϕ	The precision parameter

Subscripts

d	Indicates the d^{th} day
i	Indicates the i^{th} observation
k	Indicates the k^{th} category of the dependent variable
n	Indicates the dimension of a square matrix
p	Indicates the p^{th} participant
r	Indicates the r^{th} thermal ramp

humidity [5]. The PMV model is generally considered a static model because it is only suited for predicting thermal sensation under a steady state or slowly changing indoor conditions (i.e., rate of change lower than 2.0 K/h) [2]. Based on the PMV, Fanger introduced another index called the 'Predicted Percentage of Dissatisfied' (PPD) to establish a quantitative prediction of the percentage of thermally dissatisfied people. Additionally, thermal dissatisfaction can be caused by other local factors (e.g., drafts) and is known as local discomfort. For naturally conditioned spaces, ASHRAE 55:2020 [1] prescribes the use of the adaptive model, while EN 16798–1:2019 [3] suggests it only as a possible alternative to the Fanger approach. In 1973, in the first adaptive comfort paper published, Nicol and Humphreys [6] hypothesised the presence of 'control mechanisms' (feedback loops) between the occupants' thermal comfort perception and their behaviour in buildings. After this, research activity on the topic remained muted until the turn of the century, when intensification of research interest occurred, and several papers were published (e.g., [7,8]). The hypothesis of adaptive thermal comfort predicts that contextual factors and past thermal history modify occupant's thermal expectations and preferences [9]. As a result, people in warm climate zones would prefer higher indoor temperatures than people living in cold climate zones, which contrasts with the assumptions underlying comfort standards based on the PMV/PPD model [9]. Before inclusion in the standard EN 15251, the adaptive approach was also used by McCartney and Nicol [10] to develop an adaptive control algorithm (ACA) that was intended to be 'an alternative to fixed temperature setpoint controls within buildings' and 'was also tested in two air-conditioned buildings as part of the SCATs project' with promising results consisting in energy saving without compromising occupants' perceived thermal comfort [10]. In current standards, Fanger's PMV/PPD model is the prerogative of mechanically heated and/or cooled buildings, while the adaptive thermal comfort model is reserved for free-running buildings. EN 16798–1:2019 [3], citing ISO 7730:2005 [2], defines different categories of indoor environments for mechanically heated and cooled buildings, namely I, II, III and IV, with category I being the most stringent in terms of the management of interior conditions. An upper PPD bound is associated with each of the four PMV ranges (and therefore each category level), varying from 6 % to 25 % (see Table 1). A similar schema is present in ASHRAE 55:2020 [1], where the 'acceptable thermal environment for general comfort' is defined as $-0.5 < \mathrm{PMV} < +0.5$, corresponding to category II in Table 1.

Table 1
Default design categories for mechanically heated and cooled buildings.

Category	PMV	PPD (%)
I	–0.2 < PMV < +0.2	<6
II	–0.5 < PMV < +0.5	<10
III	–0.7 < PMV < +0.7	<15
IV	–1.0 < PMV < +1.0	<25

The categories described in Table 1 are recommended for designing mechanically heated and cooled buildings. In practice, assuming the occupants' clothing insulation and metabolic activity levels and the relative humidity and air velocity of the environment, the PMV ranges can be represented in terms of acceptable operative temperature ranges. Maintaining a tight PMV or temperature range demands more energy than allowing a wider operative temperature range. A large increase in energy consumption could only be justified if a tightly controlled thermal environment were to be more comfortable than one under less control. Arens et al. [11] investigated this specific aspect by examining the acceptability of the temperature ranges associated with categories I, II and III of the EN 15251:2007 [4] standard via three databases on occupant satisfaction (specifically, the ASHRAE RP 884 [12], SCATs [10] and Berkeley City Center Project [13] databases). The authors found that in terms of satisfaction, building occupants do not benefit from an indoor environment that is tightly controlled (i.e., a category I environment). Furthermore, they identified only a small difference in satisfaction between categories II and III. Consequently, designing and controlling indoor environments, such as office buildings, following the strict specifications suggested, for example, for category I of EN 15251:2007 [4], is unreasonable [11]. However, the real issue can be traced back to using the PMV/PPD indexes as the theoretical basis for building control and operation in the first place. In a review paper, de Dear et al. [14] state that many rigorous field studies (e.g., [15–17]), founded by ASHRAE in the 1980 s and 90 s, have clearly found the 'one-size-fits-all' approach to achieving a universally comfortable environment 'to be a failure'. The main issue is that the PMV index represents a steady-state thermal comfort model that predicts the mean thermal sensation for a large group of people. Therefore, it fails to account for dynamic and non-uniform thermal environments as well as individual differences. In practice, the indoor environment frequently changes abruptly across buildings or between various parts within a single building. For instance, manually- operated thermostats, windows and window shades can result in considerable and non-systematic changes across the indoor environment. Automatic controllers exhibit, to a lesser degree, a similar behaviour. Moreover, activity modifies an individual's basal metabolic rate over time, and the addition or removal of clothes affects their heat balance. In other words, the steady-state assumption at the root of the Fanger comfort model is very often violated (Ref [18] citing [19]). Building temperature ranges should therefore be based on real-time empirical evidence regarding the needs of the occupants. Measures to improve occupant feedback capabilities should be included in the routine control and operation of the building as well as specified in building designs. For example, Park et al. [20] analyse occupant-centric control (OCC) research, focusing on field-implementation case studies in buildings under realistic conditions. The authors offer a methodological analysis focusing on the various strategies utilised to integrate OCC into existing building systems. Another example can be found in Jung and Jazizadeh [21]. In this review, the authors, distinguishing between simulations and field evaluations, proposed a taxonomy for human-in-the-loop HVAC operations and reviewed methods for integrating human dynamics to control HAVC.

Furthermore, implementing dynamic modulations of the set-point temperature might help time-shift and/or reduce peak space heating and cooling needs, improving the energy flexibility of buildings.

In this context, individual differences between people play an essential role. When it comes to thermal comfort, individual differences result in situations where distinct people perceive the same thermal environment in different ways (i.e., they have inter-individual differences) and/or when the same individual assesses the same environment differently at different times or in different situations (i.e., this individual presents intra-individual differences). Humphreys and Nicol [22] suggested that inter-individual differences encompass both temperature differences to be considered neutral and differences in the interpretation of the semantic scale categories. In contrast, intra-individual differences refer to personal judgments that differ from time to time. Machine learning/data-driven algorithms used for predicting individual comfort responses have exploded in popularity recently and include the classification tree (e.g., [23]), random forest (e.g., [24]), gradient boosting method (e.g., [25]), support vector machine (e.g., [26]), Gaussian process classification (e.g., [27]) and artificial neural networks (e.g., [28]). Although these techniques appear to have the potential to improve prediction ability at the level of a single building occupant, their inherent character as 'black box' models renders them fundamentally unfit to explain their outputs. In predictive modelling, direct interpretability regarding the relationship between the predictors (*X*s) and the outcome of interest (*Y*) is not required; however, transparency is desirable.

In summary, there is a need for a reassessment of how buildings are designed and operated. Implementing comfortable set-point modulations in buildings that consider occupant feedback would be beneficial to comfort, potentially reduce energy consumption and significantly support the clean energy transition. As a consequence, HVAC design and operation should consider both the inter- and intra-individual differences among and within the occupants, respectively.

1.1. Research aim

A paradigm shift from 'set-point-based' control to 'perception-based' human-in-the-loop control of buildings is necessary to increase comfort, reduce energy consumption, and support the transition to clean energy. However, considering these aspects in the building design phase would also be beneficial.

The present research is an exploratory study aimed at predicting the thermal preference vote of human subjects exposed to a dynamic thermal environment. Therefore, the objective of this work is to develop a model for prediction (i.e., forecasting new data points), not for inference (i.e., testing theoretical hypotheses). Here, the data-generation process is viewed as a 'transparent' tool for developing good predictions. However, the modelling strategy does not aim to model the effect of temporal patterns directly but rather to account for them (i.e., account for the lack of independence associated with temporal data). The model is developed using data acquired from a laboratory experiment, where subjects were exposed to precisely-controlled thermal ramps in an 'office-like' climatic chamber.

2. Methodology

2.1. Data acquisition

The dataset used in this study comes from an experimental study conducted by Favero et al. [29] in the ZEB Test Cell Laboratory on the Norwegian University of Science and Technology (NTNU) premises (Trondheim campus) between September 2019 and January 2020. Thirty-eight participants (29 females and 9 males) were recruited from the university campus to participate in a randomised crossover trial, that is, a longitudinal study, in which they were subjected to a randomised sequence of thermal exposures (i.e., thermal ramps). Two identical climatic chambers, furnished like typical single offices, were used to recreate the changes in the environment induced by thermal ramps. Space heating and cooling were provided by a constant air-volume system that supplied 100 % fresh air from outside that was distributed by a 2 m-long perforated fabric tube installed at the ceiling. The operative temperature set-point of $22.0 \pm 1.0\ ^{\circ}\mathrm{C}$ was determined using the thermal comfort limit for winter established for Category A of ISO 7730–2005 [2]. The rates of the temperature changes were: (i) $\pm$ 4.4 K/h, (ii) $\pm$ 3.4 K/h, (iii) $\pm$ 2.2 K/h and (iv) $\pm$ 1.4 K/h, as recommended by ASHRAE 55:2017 [30].

During the experiment, participants were not asked to perform any specific tasks and were allowed to carry out their typical office activities. Nevertheless, the subjects were required to fill out computer-based questionnaires at scheduled intervals. By means of graphic categorical scales (see Fig. 1), these questionnaires were used to assess perception, evaluation, preference, and acceptability of the thermal environment. Further details on the experimental set-up, as well as the experimental conditions and procedure, can be found in Favero et al. [29].

2.2. Statical modelling

Multilevel models (also commonly referred to as mixed or hierarchical models) are a regression-based approach to dealing with clustered and nested data [31]. When individuals form groups or clusters, it is reasonable to expect that two randomly selected individuals from the same group will tend to be more alike than two individuals selected from different groups. Following similar reasoning, measurements taken on the same individual on different occasions will be more highly correlated than measurements taken from different individuals. Therefore, whenever data are clustered and/or nested, the assumption of independent errors is violated.

This experimental study examines a mixture of hierarchical and crossed relationships. As shown in Fig. 2, different measurements on the thermal environment (level 1) are nested within experiment conditions (level 2), which, in turn, are cross-classified by participant and day (level 3). It is essential to mention that the multilevel structure defined here is not the property of a model but rather the property of the experimental/study design, which is then reflected in the data, which the model then encapsulates.

Within the multilevel framework, there are different modelling strategies that can be used. In this study, two different modelling strategies were applied, namely the beta mixed-effects model (a beta model including random effects) and the ordinal mixed-effects model (an

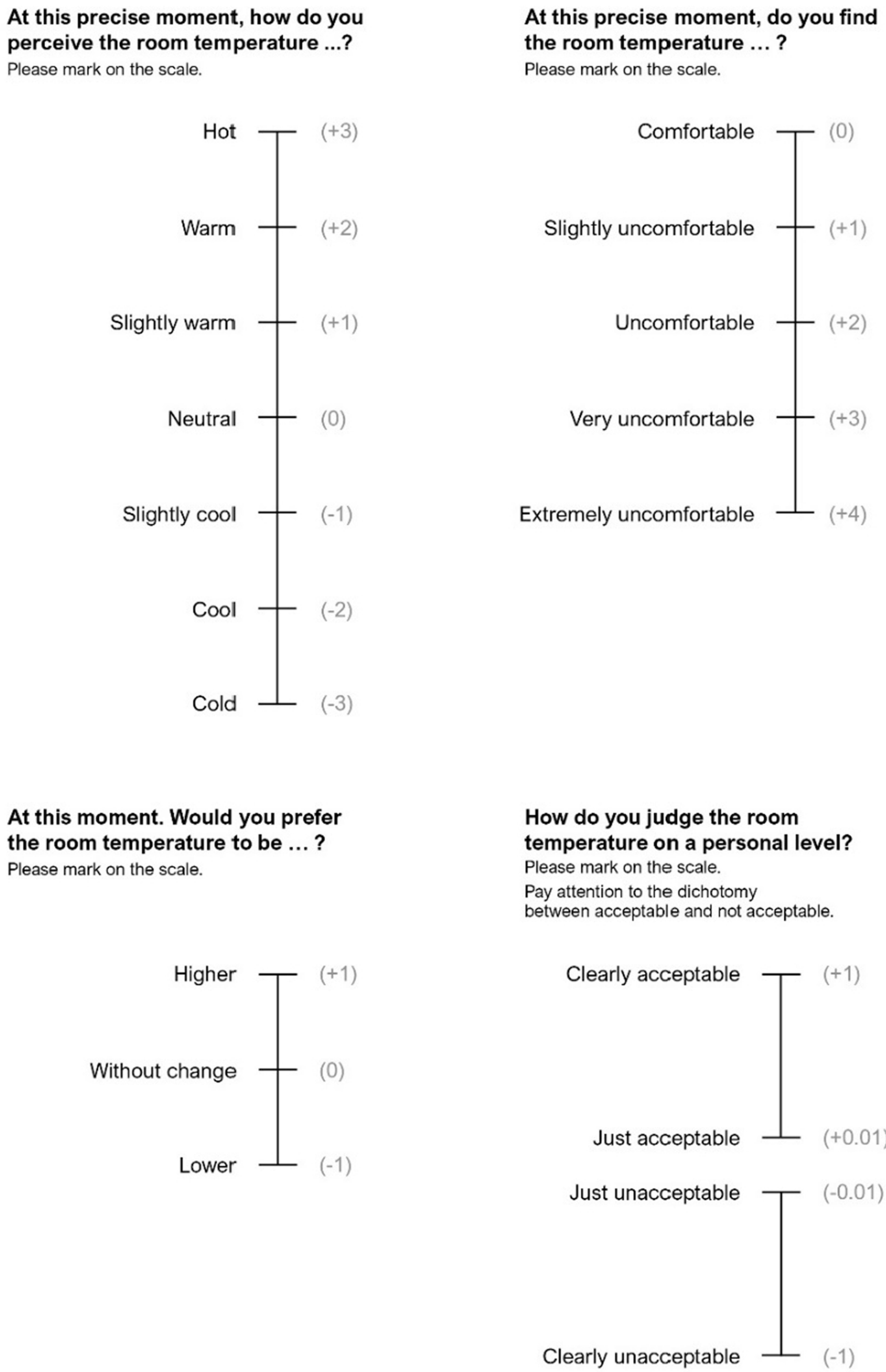

Fig. 1. Subjective scales used to assess perception, evaluation, preference, and acceptability of the thermal environment. Note. The numerical values of the scale were not shown to the participants during data collection.

ordinal model including random effects). These two approaches are described in the following sections.

2.2.1. *Beta mixed-effects model*

Generalised linear models[1] (GLMs) constitute a large class of models where the conditional distribution of the response variable Y_i is assumed to follow an exponential family distribution with mean μ_i. The latter is assumed to be some function of $\eta_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$, where $\mathbf{x}_i$ is the vector of covariates for the i^{th} observation and $\boldsymbol{\beta}$ is the respective vector of parameters to be estimated. However, one of the assumptions behind the model is the independence of the errors, which cannot be assumed whenever data are clustered and/or nested (see Section 2.2). To deal with dependent errors, GLMs can be extended to generalised linear mixed models, in which the linear predictor $\boldsymbol{\eta}$ contains random effects (i.e., $\boldsymbol{Zu}$) in addition to the fixed effects (i.e., $\boldsymbol{X\beta}$).

[1] This class of models is not to be confused with general linear model which usually refers to linear regression models – generally assuming a normal conditional distribution of the response.

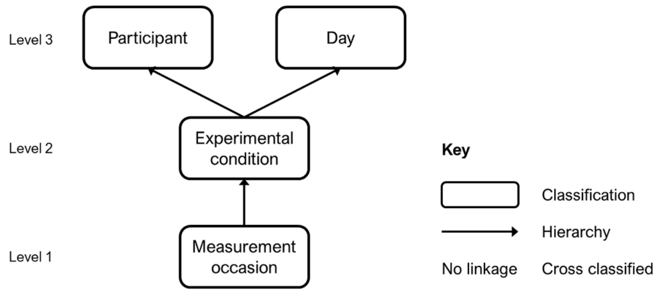

Fig. 2. Schematic of the three-level hierarchical study: repeated measures within experimental conditions cross-classified by participant and day.

In this study, the conditional distribution of the response variable $\boldsymbol{Y}$ is assumed to follow a beta distribution, where its mean $\boldsymbol{\mu}$ is linked to linear predictor $\boldsymbol{\eta}$ through a logit function:

$$\begin{aligned}&\boldsymbol{Y} \sim \text{Beta}(\boldsymbol{\mu}, \phi)\\&\text{Logit}(\boldsymbol{\mu}) = \boldsymbol{\eta}\\&\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Zu}\\&\boldsymbol{u} \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma})\end{aligned} \tag{1}$$

where ϕ is the precision parameter, and $\boldsymbol{X}$ (whose row i is $\mathbf{x}_i^{\mathrm{T}}$ and contains the i^{th} observation of the covariates) and $\boldsymbol{Z}$ are the design matrices for the fixed and random effects, respectively. The reader is referred to Appendix A for more details about the mathematical notation and a practical example. For the sake of clarity and brevity, the beta mixed-effects model with the logit link will hereafter be referred to as simply the beta model.

2.2.2. Ordinal mixed-effects model

Cumulative link models (CLMs) belong to the ordinal regression model class and can be performed using GLMs. A cumulative model is used when latent variable representation is desired. Here, the dependent variable Y is the categorisation of a latent (unobservable) continuous variable $\widetilde{Y}$. Therefore, there are some latent thresholds parameters τ_k, with $k \in \{1, ..., K\}$, that divide the values of $\widetilde{Y}$ into $K+1$ bins, that is, the observable ordered categories of Y. CLMs assume independence of errors and are not suited for modelling clustered and/or nested data. Their extensions for dealing with the dependent errors are the cumulative link mixed models (CLMMs).

In a CLMM, the conditional distribution of the response variable Y_i for the i^{th} observation is assumed to follow a multinomial distribution with probability vector $\boldsymbol{\pi}_i = \{\pi_{i1}, ..., \pi_{ik}\}$, where $\pi_{ik} = \Pr(Y_i = k)$. The cumulative probability corresponding to π_{ik} is $\gamma_{ik} = \Pr(Y_i \leq k)$; hence, $\gamma_{ik} = \pi_{i1} + ... + \pi_{ik}$. The cumulative probabilities are then mapped to the real numbers through a link function. In this study, the logit function was chosen as that link function. The mathematical formulation of the model can be written as:

$$\begin{aligned}&\boldsymbol{Y} \sim \text{Multinomial}(n, \boldsymbol{\pi})\\&\text{Logit}(\gamma_k) = \mathbf{1}\tau_k - \boldsymbol{\eta}\\&\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Zu}\\&\boldsymbol{u} \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma})\end{aligned} \tag{2}$$

where the τ_k are the thresholds parameters and $\boldsymbol{\eta}$ is the linear predictor term with a fixed effect component (i.e., $\boldsymbol{X\beta}$) without an intercept[2] and a random effect component (i.e., $\boldsymbol{Zu}$). The reader is referred to Appendix A for more details about the mathematical notation and a practical example. For the sake of clarity and brevity, the ordinal mixed-effects model with the logit link will hereafter be referred to as simply the ordinal model.

[2] Omitting the intercept term allows the full set of thresholds $\tau_1, ..., \tau_k$ to be identified.

2.2.3. Computing predictions using a multilevel model

Research setting aims to make predictions for certain values of x (e.g., adjusting the values of one x at a time or for combinations of x-values that reflect 'typical' persons) rather than calculating a probability for each individual in the sample. However, for a multilevel model, the treatment of the group-level residual $\boldsymbol{u}$ (i.e., group random effect) for these 'out-of-sample'[3] predictions must be considered.

In this study, two different procedures were used to handle the group-level residual during prediction. For the ordinal model, the first procedure consisted of holding the group-level residual at its mean of zero and calculating the probabilities for some specific x-values. It should be noted that the calculated predictions are not the mean response probabilities for the specific x-value because γ_{ik} is a nonlinear function of $\boldsymbol{u}$(as is $\boldsymbol{\pi}_i$). However, since $\boldsymbol{u}$ is assumed to be normally distributed with mean = median = 0, the $\text{Logit}(\gamma_{ik})$ for $x = x^*$ and $\boldsymbol{u} = 0$ is equal to the median γ_{ik} for $x = x^*$ across groups. This is the case because the logit transformation does not affect the rank order of the observations. The response probabilities thus calculated have a cluster-specific interpretation.

The second procedure outlined a simulation-based approach, which consisted of the following steps:

i. Generate M values for the random effect $\boldsymbol{u}$ from the $\text{Normal}(\mathbf{0}, \boldsymbol{\Sigma})$ distribution;
ii. For each simulated value $(m = 1, ..., M)$ calculate, for the given x-value, the cumulative response probabilities for each $K+1$ ordered categories of Y;
iii. Compute the mean of the M cumulative response probabilities calculated in (ii) for each of the $K+1$ ordered categories of Y;
iv. Repeat steps (i) – (iii) for a different x-value.

The generated M values for the random effect should be a large number, here fixed at $1 \cdot 10^4$. This approach results in probabilities with a population-averaged interpretation (i.e., averaged across experimental conditions, participants and days). The same two procedures were applied to the beta model, with the difference that the prediction was not a vector (i.e., probabilities of voting in each category) but rather a single number (i.e., predicted mean).

2.3. Data pre-processing and analysis

A total of 314 thermal ramps were performed, for a total of 1522 votes. There were three missing values for thermal perception and evaluation, six for thermal preference and 14 for thermal acceptability. However, since only thermal preference was of interest in this study, only the missing values for the latter were eliminated. As a result, the final sample size was reduced to 1516 observations. Fig. 3(a) illustrates the distribution of the thermal preference votes.

As shown in Fig. 1, thermal preference ratings were measured using a graphical categorical scale. Participants could cast their vote by placing a diagonal line anywhere within the limits of the scale (i.e., within 'lower' and 'higher'). Consequently, the resulting distribution of votes is on a continuous, but bounded, scale (Fig. 3(a)). In this study, the conditional distribution of the response (i.e., $Y|\mu$) is assumed to follow a beta distribution. However, since any beta distribution's probability density function (pdf) is defined only on the interval (0,1), the dependent variable needs to be rescaled. Therefore, the thermal preference votes were scaled so that the values at the boundary of the scale, -1 (i.e., 'lower') and $+1$ (i.e., 'higher'), were mapped to $+0.001$ and $+0.999$, respectively.

As an alternative approach to the beta model, the ordinal model was

[3] The term 'out-of-sample' is used here to highlight that fixing variables at some values (e.g., their mean) may not reflect any actual person in the sample.

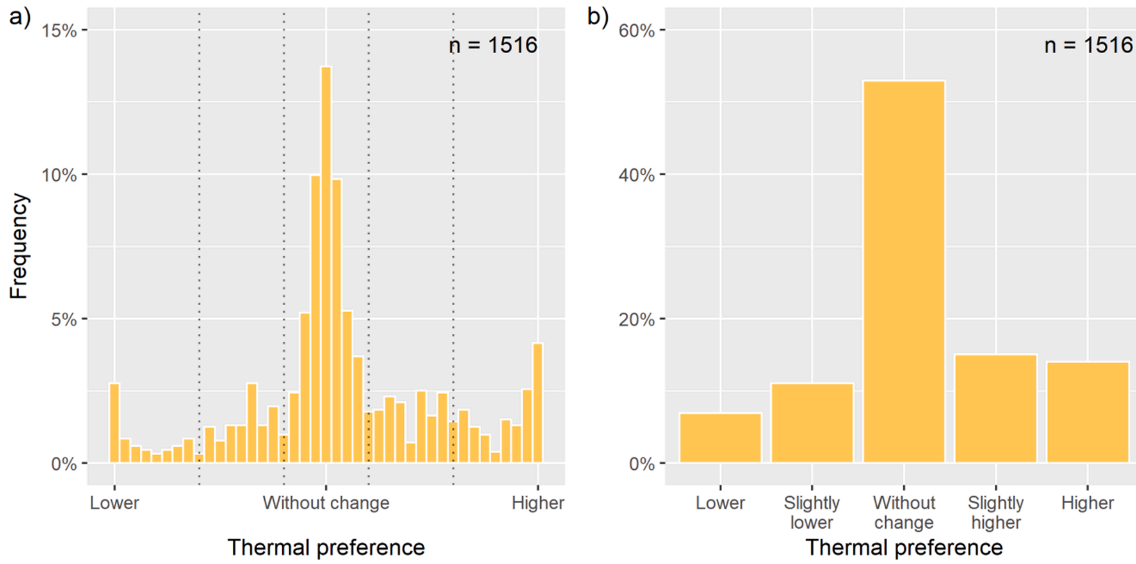

Fig. 3. Frequency distributions of (a) the thermal preference votes and (b) its categorisation. Note. The dotted lines represent the thresholds used for the categorisation, that is, −0.6, −0.2, + 0.2, and + 0.6.

chosen. However, the ordinal model requires the dependent variable to be categorical, which entails categorising the thermal preference votes. Therefore, the votes were binned into five categories according to the thresholds −0.6, −0.2, + 0.2, and + 0.6:

a) thermal preference votes < -0.6 were defined as 'lower';
b) $-0.6 \leq$ thermal preference votes < -0.2 were defined as 'slightly lower';
c) $-0.2 \leq$ thermal preference votes $\leq +0.2$ were defined as 'without change';
d) $+0.2 <$ thermal preference votes $\leq +0.6$ were defined as 'slightly higher';
e) thermal preference votes $> +0.6$ were defined as 'higher'.

The frequency distribution of the resulting bins can be observed in Fig. 3(b).

In a regression-type model, the shape of the distribution of a predictor has no direct impact on the model itself. Therefore, there is no real *a priori* need to transform or categorise a predictor based on its distribution. Of greater importance is the correlation between predictors (i.e., whether or not there is collinearity[4]). There are two types of collinearity: structural and data-based collinearity. The former is a mathematical artefact originating from composing new predictors from other predictors, such as powers (higher-order terms) or products (interaction terms) of predictors. The latter is a 'property' of the data itself, which can be the result of, for example, a poorly designed experiment. To manage the first type of collinearity, predictors lacking a meaningful zero were centred by their grand mean; it should be noted that this standardisation procedure can facilitate the interpretation of the model [32]. Data-based collinearity is more challenging and regrettably, is the most common form of the two. It is typically dealt with via the removal of one or more of the collinear predictors from the regression model. Variable selection was performed with an automated backward elimination employing the Akaike information criterion (AIC) as the selection criterion.

Table 2 presents the descriptive statistics of all the dependent variables used to infer the models. Detailed information concerning these variables can be found in Appendix B, while the instruments' accuracy can be found in Favero et al. [29].

All statistical analyses were performed using R [33] with the RStudio integrated development environment [34]. The beta model and the ordinal model were determined with the *glmmTMB* package [35] and *ordinal* package [36], respectively. Automated backward elimination was performed with the *buildmer* package [37] and all the graphs were

Table 2
Descriptive statistics of the variables used in the models.

Variable	Code	Unit	Mean*	Frequency*	Median (1st, 25th, 75th, 99th)**
Thermal resistance of clothing	*Clothing*	clo	0.86	–	0.87 (0.54, 0.78, 0.97, 1.11)
Gender	*Gender*	female	–	0.78	–
		male	–	0.22	–
Age	*Age*	years	27.11	–	25 (20, 22, 30, 49)
Body Mass Index	*BMI*	kg/m^2	22.09	–	21.67 (17.42, 20.69, 23.94, 29.24)
Time lived in Norway	*Time. Norway*	≤ 3 years	–	0.53	–
		greater than 3 years	–	0.47	–
Air velocity	*Air.vel*	m/s	<0.10	–	0.00 (0.00, 0.00, 0.00, 0.06)
Time of day	*Time. day*	morning	–	0.47	–
		afternoon	–	0.53	–
Vapour pressure	*Vap.pre*	kPa	0.70	–	0.70 (0.39, 0.58, 0.82, 1.06)
Operative temperature	*Top*	°C	22.39	–	22.10 (18.79, 21.08, 23.72, 27.33)

* the mean refers to continuous variables, whereas the frequency refers to categorical variables
** where 1st, 25th, 75th and 99th represent percentiles

[4] Collinearity is semantically equivalent to multicollinearity. In a general sense, collinearity refers to 'the condition of being collinear' and is a property of a set of explanatory variables, not just pairs of them.

created with the *ggplot2* package [38]. The significance level for all analyses was set at 0.05.

3. Results

In this section, the main results of the statical analysis are presented. Table 3 list all the variables used in the two models (i.e., the beta and ordinal models).

3.1. Testing for cluster effects

In Section 2.2, the experimental study was described as a mixture of hierarchical and crossed relationships. Nevertheless, before proceeding with the analysis, it was essential to establish that the three-level cross-classified model fit the data significantly better than the simpler three-levels models and the two-level model nested within it (see Fig. 4). The single-level model (i.e., the model without random effects) was also checked. The likelihood ratio (LR) test was used to perform this initial check.

This preliminary analysis was carried out for both the beta and ordinal models, and its results are presented in Table 4. Here, the three-level cross-classified model is compared with the nested models. For both the beta and ordinal models, the three-level cross-classified model offers a better fit to the data.

3.2. Initial model

In this section, the initial full model is presented. The formulation of the linear predictor η_i is the same for the beta and ordinal models and can be written as:

$$\eta_i = \beta_1 \text{clothing\_c}_i + \beta_2(\text{gender}_i) + \beta_3 \text{age\_c}_i + \beta_4 \text{BMI\_c}_i + \beta_5(\text{time.norway}_i) + \beta_6 \text{air.vel}_i + \beta_7(\text{time.day}_i) + \beta_8 \text{vap.pre\_c}_i + \beta_9 \text{top\_c}_i + u^{(2)}_{\text{ramp\_ID}(i)} + u^{(3)}_{\text{participant\_ID}(i)} + u^{(3)}_{\text{day\_ID}(i)} \quad (3)$$

where the subscript i is used to stress dependence on the i^{th} observation. Since the three-level cross-classified model fits the data better (see Section 3.1), all three random components were added to the initial model. Only the main results of applying automated backward elimination are illustrated in the following sections.

Table 3
Covariates used in the models.

Classification (level)	Code	Variable	Type	Unit
Days (level 3)	*Day_ID*			
Participants (level 3)	*Participant_ID*			
	Gender	Gender	Categorical, time-independent	Female (reference)/ Male
	Age_c	Age (centred)	Continuous, time-independent	Years
	BMI_c	Body Mass Index (centred)	Continuous, time-independent	kg/m^2
	Time.Norway	Time lived in Norway	Categorical, time-independent	Less than or equal to 3 years (reference)/ More than 3 years
Experimental conditions (level 2)	*Ramp_ID*			
	Time.day	Time of day	Categorical, time-independent	Morning (reference)/ Afternoon
	Clothing_c	Thermal resistance of clothing	Continuous, time-independent	clo
Measurement occasions (level 1)	*Timepoint*			
	Top_c	Operative temperature (centred)	Continuous, time-dependent	°C
	Vap.pre_c	Vapour pressure (centred)	Continuous, time-dependent	kPa
	Air.vel	Air velocity	Continuous, time-dependent	m/s
	Therm.pref	Thermal preference	Continuous/ Categorical, time-dependent	–

3.3. Ordinal model

Table 5 summarises the results of the ordinal model after automated backward elimination has been applied. Here, the coefficient estimates are given in units of ordered logits (or ordered log-odds). Five significant predictors were identified – thermal resistance of clothing, Body Mass Index, air velocity, time of day and operative temperature – all negatively associated with $\text{Logit}(\gamma_k)$. The *ordinal* package [36] parametrises the model as:

$$\text{Logit}(\gamma_k) = \mathbf{1}\tau_k - \boldsymbol{\eta} = \mathbf{1}\tau_k - \boldsymbol{X\beta} - \boldsymbol{Zu} \quad (4)$$

so a negative coefficient for β indicates that an increase of the associated variable x_i decreases the thermal preference vote. Stated analogously, votes for higher categories (e.g., prefer 'higher') are less likely. Nevertheless, the aim of this study is not inference but prediction; therefore, the specific values of the model's coefficients are not of interest. Furthermore, utilising any automated model selection procedure (e.g., automated forward selection, backward elimination or stepwise selection) should be avoided for inferential purposes. The parameter estimates are biased away from zero, the standard error and p values are too low and the confidence intervals are too narrow (page 68 of [39]), leading to misleading results. For prediction purposes, model selection can indeed provide a better bias-variance trade-off and improve the out-of-sample error [40,41].

The estimated coefficients for a multilevel model are referred to as cluster-specific effects. For instance, the coefficient of *Top_c* in Table 5 is interpreted as the effect of a one-unit change in *Top_c* on the log-odds that $\Pr(Y \le k)$ for a given cluster (i.e., while the unobserved characteristics captured by the random effects are held constant). However, considering the effects in this manner implies that individuals are compared with the exact same value for fixed and random effects. For some variables (e.g., gender) or other specific purposes, a comparison averaging across unobserved characteristics in the population is often of interest. In such a situation, population-averaged probabilities should be derived (see Section 2.2.3).

Fig. 5 shows the predicted probabilities as functions of the operative temperature for the cluster-specific and population-averaged procedures. It can be seen that the probabilities calculated with the two methods are dissimilar. For example, the maximum predictive probability for 'without change' is about 91 % for the cluster-specific approach, while it is only 55 % for the population-averaged one. Fig. 6(a) shows the probability mass for the ordinal model and cluster-specific procedure. These probabilities are plotted as a function of three different operative temperatures while holding the other covariates constant at their centred values and fixing the random effects at zero. Fig. 6(b) shows the population-averaged procedure's results.

3.4. Beta model

Table 6 summarises the results for the beta model after automated backward elimination has been applied. Here, the estimated coefficients

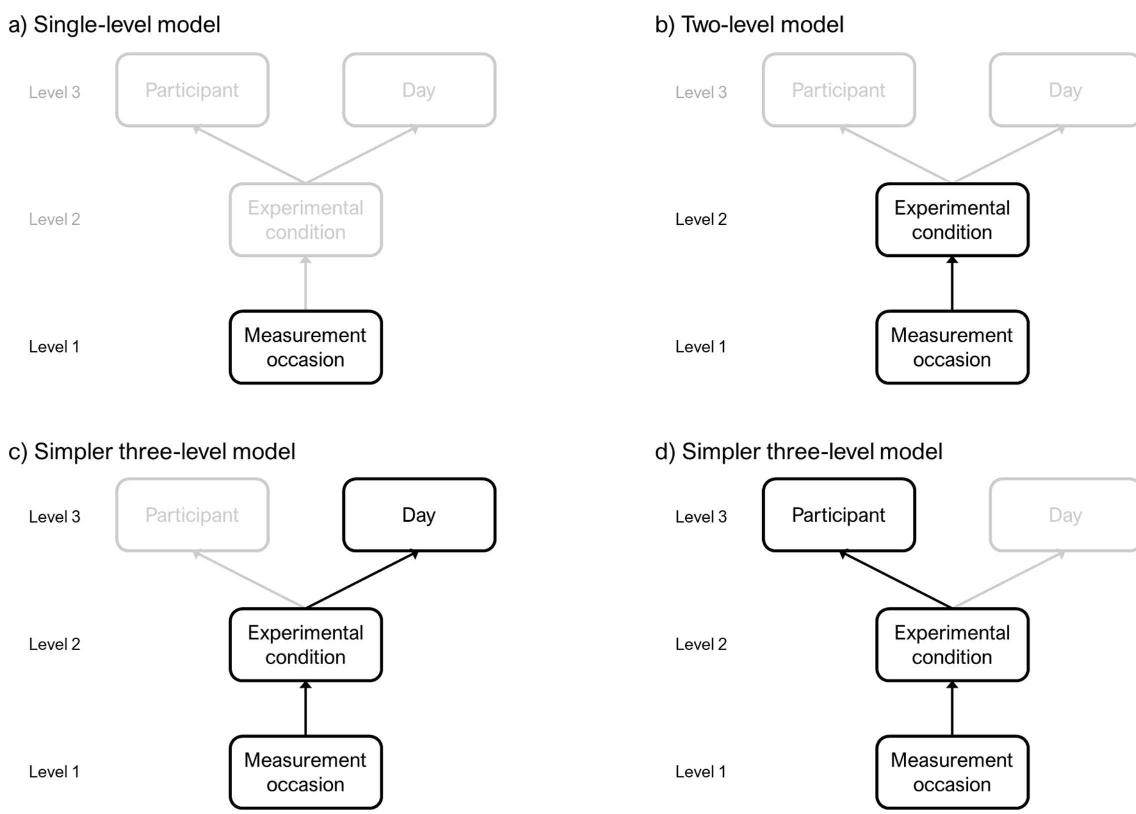

Fig. 4. Schematics of the (a) single-level model, (b) two-level model and (c and d) two simpler three-level models nested within the three-level cross-classified model.

Table 4
Preliminary check.

Modelling strategy	Model comparison	LR test statistic
Ordinal model		
	Testing for multilevel model (see Fig. 4(a))	$\chi^2(3) = 534.4, p < 0.001$
	Testing for participants and days (see Fig. 4(b))	$\chi^2(2) = 59.271, p < 0.001$
	Testing for participants (see Fig. 4(c))	$\chi^2(1) = 12.854, p < 0.001$
	Testing for days (see Fig. 4(d))	$\chi^2(1) = 55.981, p < 0.001$
Beta model		
	Testing for multilevel model (see Fig. 4(a))	$\chi^2(3) = 675.77, p < 0.001$
	Testing for participants and days (see Fig. 4(b))	$\chi^2(2) = 49.328, p < 0.001$
	Testing for participants (see Fig. 4(c))	$\chi^2(1) = 10.946, p < 0.001$
	Testing for days (see Fig. 4(d))	$\chi^2(1) = 46.525, p < 0.001$

are given in units of logits. Four significant predictors were identified – thermal resistance of clothing, Body Mass Index, time of day and operative temperature – all negatively associated with Logit(μ) (the intercept was not considered).

Fig. 7 shows the predicted responses as functions of the operative temperature using the cluster-specific and population-averaged procedures. The predicted response μ is the inverse of the link function, which in this study corresponds to the inverse of the logit:

$$\mu = \text{Logit}^{-1}(\eta) = \text{Logistic}(\eta) = \frac{1}{1+e^{-\eta}} = \frac{1}{1+e^{-(X\beta+Zu)}} \quad (5)$$

More details about the mathematical formulation can be found in the

Table 5
Regression coefficients for the predictors in the ordinal model (after applying automated backward elimination).

Fixed Effects		coeff	se (coeff)	z	p value
Threshold 1, τ_1		−6.325	0.383	−16.519	—
Threshold 2, τ_2		−4.245	0.345	−12.307	—
Threshold 3, τ_3		2.065	0.306	6.761	—
Threshold 4, τ_4		4.070	0.327	12.441	—
Clothing_c		−3.483	1.375	−2.532	0.011*
BMI_c		−0.218	0.093	−2.338	0.019*
Air.vel		−12.584	5.503	−2.287	0.022*
Time.day	morning	reference			
	afternoon	−0.474	0.238	−1.988	0.047*
Top_c		−1.519	0.077	−19.719	< 0.001*
Random effects		sd	var		
Ramp_ID	(Intercept)	1.585	2.512		
Day_ID	(Intercept)	1.008	1.017		
Participant_ID	(Intercept)	1.158	1.340		

Number of groups: *Ramp_ID* = 314, *Day_ID* = 68, *Participant_ID* = 38
* indicates a significant term.

Appendix A.

All the lines in Fig. 7 are plotted as a function of the operative temperature while the other covariates (i.e., the fixed effects) are held constant at their centred values. However, these lines differ in the random effects, specifically:

- The solid black line (cluster-specific procedure) has the random effect fixed at zero;
- The dashed black lines (cluster-specific procedure) have the random effect fixed at the 16^{th} and 84^{th} percentiles (which correspond roughly to ± 1 standard deviation above and below the mean);
- The solid red line (population-averaged procedure) has the random effect derived from simulation (see Section 2.2.3).

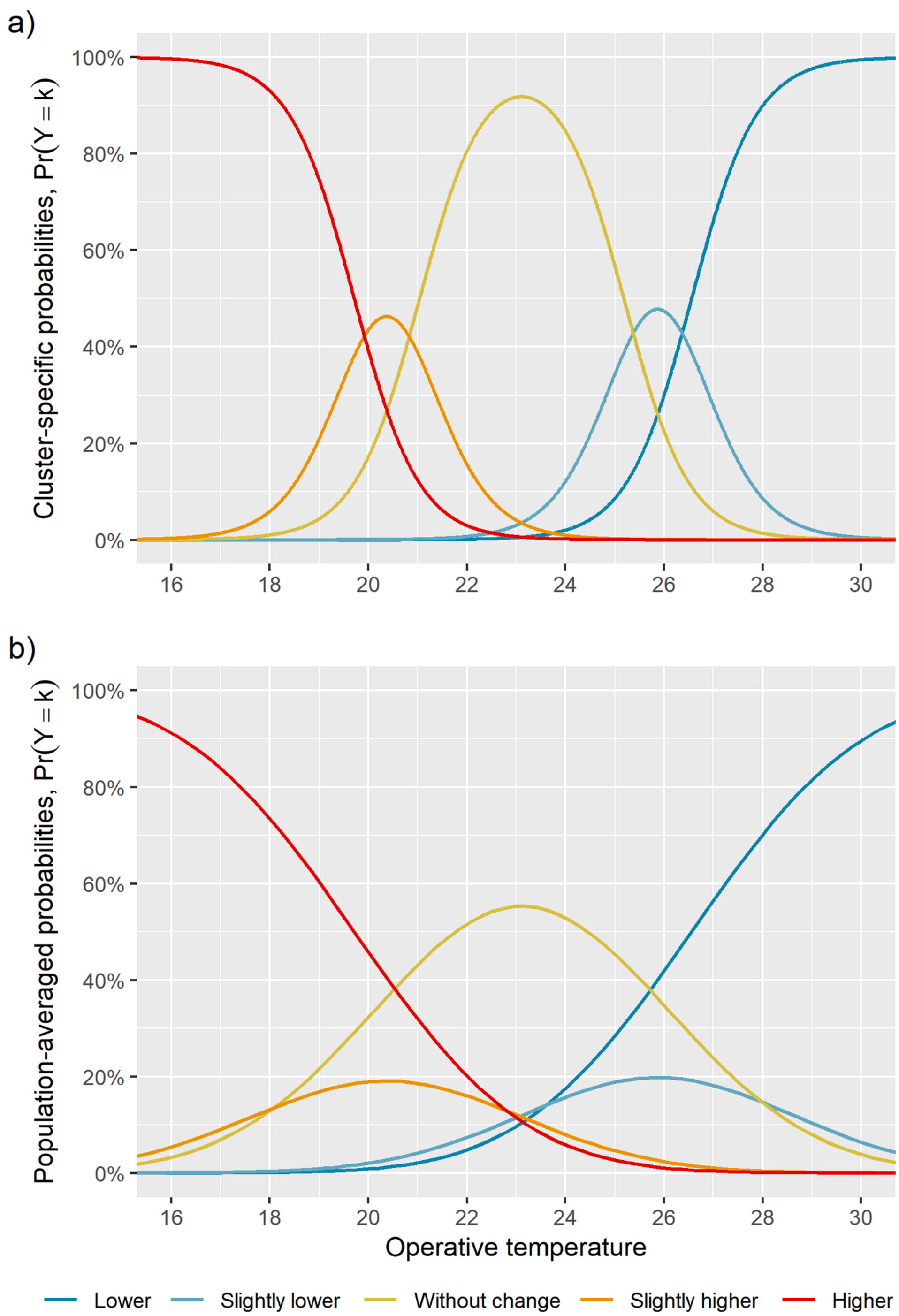

Fig. 5. Predicted probabilities of a thermal preference vote using the (a) cluster-specific and (b) population-averaged procedures.

The points in Fig. 7 are the observed thermal preference votes. While the predicted central tendency follows the general trend of the data, the predictions do not agree well with the observations, particularly close to the upper (i.e., prefer 'higher') and lower (i.e., prefer 'lower') boundaries.

Fig. 8(a) shows the pdfs generated from the beta model's estimated parameters (i.e., μ and ϕ) using the cluster-specific procedure. Each pdf is plotted as a function of three different operative temperatures while the other covariates are held constant at their centred values and the random effects are fixed at zero. It can be observed that the dispersion of the probability densities is relatively high. For instance, for an operative temperature of 26 °C, the probability of voting equal or lower 0.50 (i.e., from 'lower' to 'without change' on the continuous scale) is about 93 %, implying a 7 % probability of voting higher than that. Fig. 8(b) shows the categorised probabilities of the predicted thermal preference votes. Fig. 9 presents the pdfs generated from the beta model with the population-averaged procedure. Here, as in Fig. 8, each pdf is plotted as a function of three different operative temperatures while the other covariates are held constant at their centred values, but the random effects are the results of simulations (see Section 2.2.3).

4. Discussion

Different approaches can be found in the literature for OCC for building (e.g., [42–44]). Futhrtmore, diverse modelling strategies have been developed to predict occupant' thermal preferences, many of

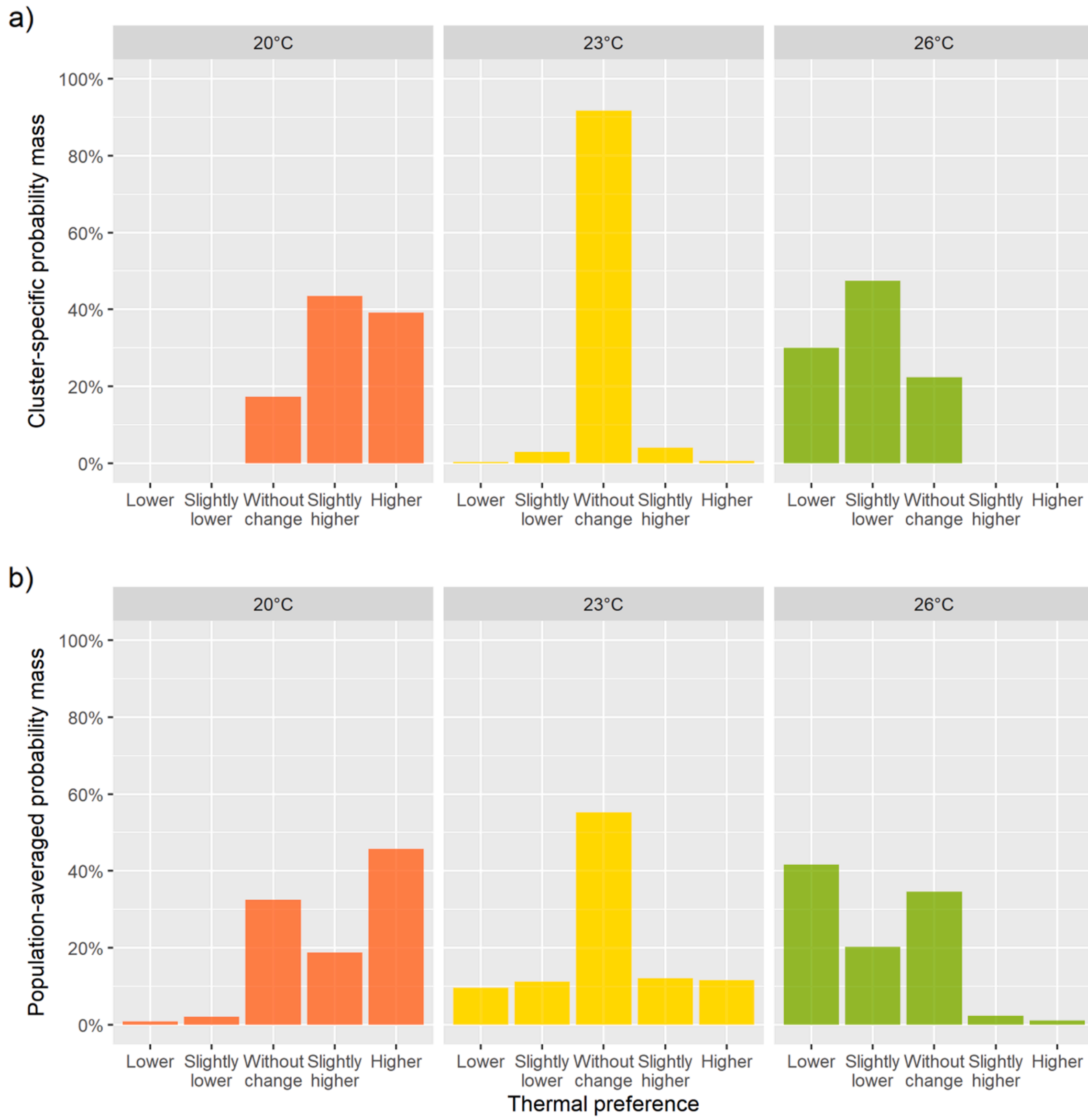

Fig. 6. Predicted probabilities of a thermal preference vote using the (a) cluster-specific and (b) population-averaged procedures for three different operative temperatures.

Table 6
Regression coefficients for the predictors in the beta model (after applying automated backward elimination).

Fixed Effects		coeff	se (coeff)	z	*p* value
(Intercept)		0.301	0.091	3.319	< 0.001*
Clothing_c		−1.291	0.456	−2.829	0.005*
BMI_c		−0.077	0.029	−2.634	0.008*
Time.day	morning	Reference			
	afternoon	−0.177	0.083	−2.136	0.033*
Vap.pre_c		−0.627	0.340	−1.845	0.065
Top_c		−0.413	0.018	−23.121	<0.001*
Random effects		sd	var		
Ramp_ID	(Intercept)	0.565	0.320		
Day_ID	(Intercept)	0.311	0.096		
Participant_ID	(Intercept)	0.343	0.177		

Number of groups: *Ramp_ID* = 314, *Day_ID* = 68, *Participant_ID* = 38
Dispersion parameter: 6.38
* indicates a significant term.

which can be found in the review of Park et al. [20], Jung and Jazizadeh [21] and Ngarambe et al. [45]. Among them a large portion are machine learning/data-driven algorithms. Even though there is sometimes overlap in goals and algorithms, statistical modelling and machine learning are based on two different concepts. The basic goal of stochastic modelling is to understand which probabilistic model could have generated the data observed. The usual procedure can be synthesised in the following steps: (i) choose a potential model from a plausible model family, (ii) fit the model to the data (i.e., estimate its parameters), and (iii) contrast the fitted model with other models. After selecting a model, this is used to conduct investigations, such as hypothesis testing and predicting new values. The estimated model becomes the lens used to interpret the data. Usually, a model that reasonably approximates the underlying stochastic process that has generated the data predicts well. On the contrary, machine learning is a data-driven application that is inspired by pattern recognition and focuses on regression, classification, and clustering techniques. The underlying stochastic process is frequently of secondary importance. Of course, stochastic models and procedures may be used to frame many machine learning approaches. However, the data are not regarded as having been created by that model. Instead, the main objective is figuring out which method or

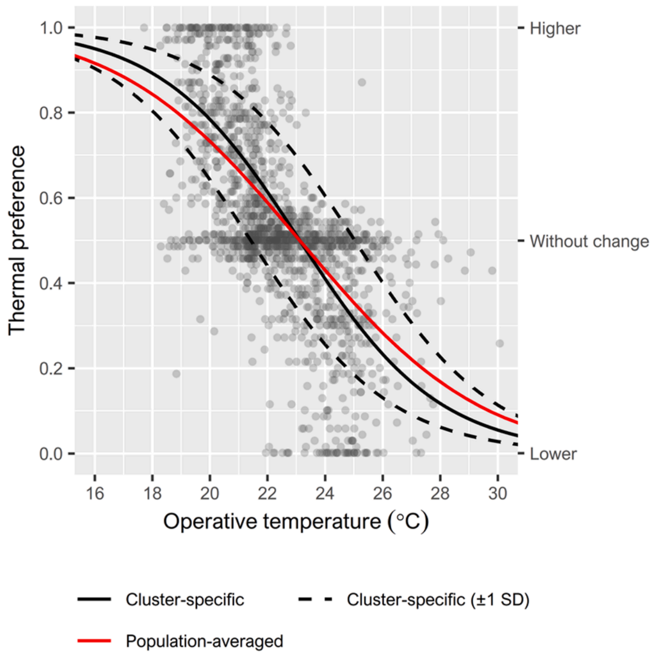

Fig. 7. Predicted responses using the cluster-specific (black line) and population-averaged (red line) procedures. Note. The points are the observed thermal preference votes.

approach performs the specific task. Although these techniques appear to have the potential to improve prediction ability at the level of a single building occupant, their inherent character as 'black box' models renders them fundamentally unfit to explain their outputs. Interpretability is one of the primary problems with machine learning and can be an issue in a specific setting. For instance, understanding why a model reached a particular conclusion is fundamental in a building design setting. In this study, the statistical modelling strategies applied are a transparent tool and, as such, can be easily used in contexts where interpretability is required. Moreover, while the data used to develop the model in this study derives from a laboratory experiment with a mechanically conditioned environment, the modelling strategies are independent of this aspect. As such, they could also be applied to a naturally conditioned space.

In the literature, examples of statistical modelling can be found in the study of Daum et al. [46]. The authors create personalised thermal comfort profiles using multinomial logistic regression, a regression technique used to analyse a dependent variable measured on a categorical scale. The main difference with ordinal regression is that the categorical data are assumed to have no intrinsic ordering. Not taking an inherent order of the dependent variable makes the model more flexible than ordinal regression. However, it is essential to mention that this flexibility comes at a price. The number of parameters to estimate will drastically increase because $k-1$ different linear predictor term (η_k) are needed for the k category of the dependent variable (only $k-1$ because one category of the dependet variable is used as reference). In our study, since the dependent variable has five categories, this would have led to having, for example, four parameters for the operative temperature to estimate instead of only one. More parameters to estimate would require a larger sample size. The other difference compared with the study of Daum et al. [46] is that the diversity between subjects is directly accounted for in the model through the random effects term. Doing so makes it possible to model and predict the thermal preference of a specific (using the cluster-specific procedure) and a 'general' occupant (using the population-averaged procedure). In addition, the beta model allows doing the same when the thermal preference votes are measured on a continuous, but bounded, scale.

In the next sections, the results previously illustrated are examined and interpreted. To begin with, the variables selected by the beta and ordinal model are contrasted and discussed. Subsequently, the beta and ordinal models are compared based on their predictive capabilities and the cluster-specific and population-averaged approaches are analysed. Finally, the limitations of the study are provided.

4.1. Variables selection

As explain in Section 1.1, the focus of the study is prediction and not inference; therefore, the specific value of the models' coefficients is not of interest. However, it is useful to compare variables selection across the models and contrast the relative importance of these variables. Table 5 and Table 6 show that automated backward elimination selected different sets of predictors for the two models. Four out of five predictors are shared by the two models (*Clothing_c*, *BMI_c*, *Time.day* and *Top_c*), while the fifth variables differ. For the ordinal model, automated backward elimination selected *Air.vel,* whereas for the beta model, *Vap.pre_c* was selected. In any attempt to understand the relative importance of the parameters estimated for the models, a direct comparison between their absolute values would be meaningless because the variables are measured using different units. Furthermore, several units could be used to measure the same variable. For example, if the operative temperature had been measured in degrees Fahrenheit instead of degrees Celsius, its estimated regression coefficient would have been different. However, the importance of the variable would not have changed. The relative importance of the predictors could be obtained via standardisation (i.e., subtracting the mean from each observed variable and dividing by its standard deviation) before conducting the statistical analysis. The resulting parameters estimated by the model are on the same scale and can be directly compared. The results of this procedure are show in Table 7. Here, even though the two models have different predictors, the order of relative importance of the common predictors is the same. The variables that differ between the two models are of minor relative importance. However, this importance is purely statistical. To determine the practical importance of the variables, subject-area expertise is required. Note that p values cannot be used directly to assess the importance of the predictors. A predictor can have a small p value when it has a very precise estimate, low variability, or a large sample size. As a result, even effect sizes that are small in practice might have extremely low p values. Understanding the practical importance of the predictors is beyond the scope of this study and is not pursued further. However, for inferential purposes, it is of the utmost importance.

As mentioned in Section 2.3, the Akaike information criterion (AIC) was used for variable selection. This metric is based on the maximised log-likelihood value with a penalty for including more parameters; it is a trade-off between goodness of fit (assessed by the likelihood function) and parsimony (the smaller the number of parameters, the lower the penalty). However, the AIC tends to over-parameterised, thus selecting models with a higher number of predictors, which could explain why the first four relatively important predictors were common to the two models, while their least relatively important predictors differed.

4.2. Models' comparison

The Akaike information criterion is generally used to compare different possible models and determine which one best fits the data. However, it cannot be used to compare models with different likelihood functions.[5] For example, for a discrete distribution (e.g., ordinal response), the likelihood refers to the joint probability mass of the data, whereas for a continuous distribution (e.g., continuous response), the likelihood refers to the joint probability density of the data. Therefore, models based on continuous and ordinal responses cannot be compared directly. For this reason, the two models are compered graphically in

[5] This is generally true for all probability-based statistics.

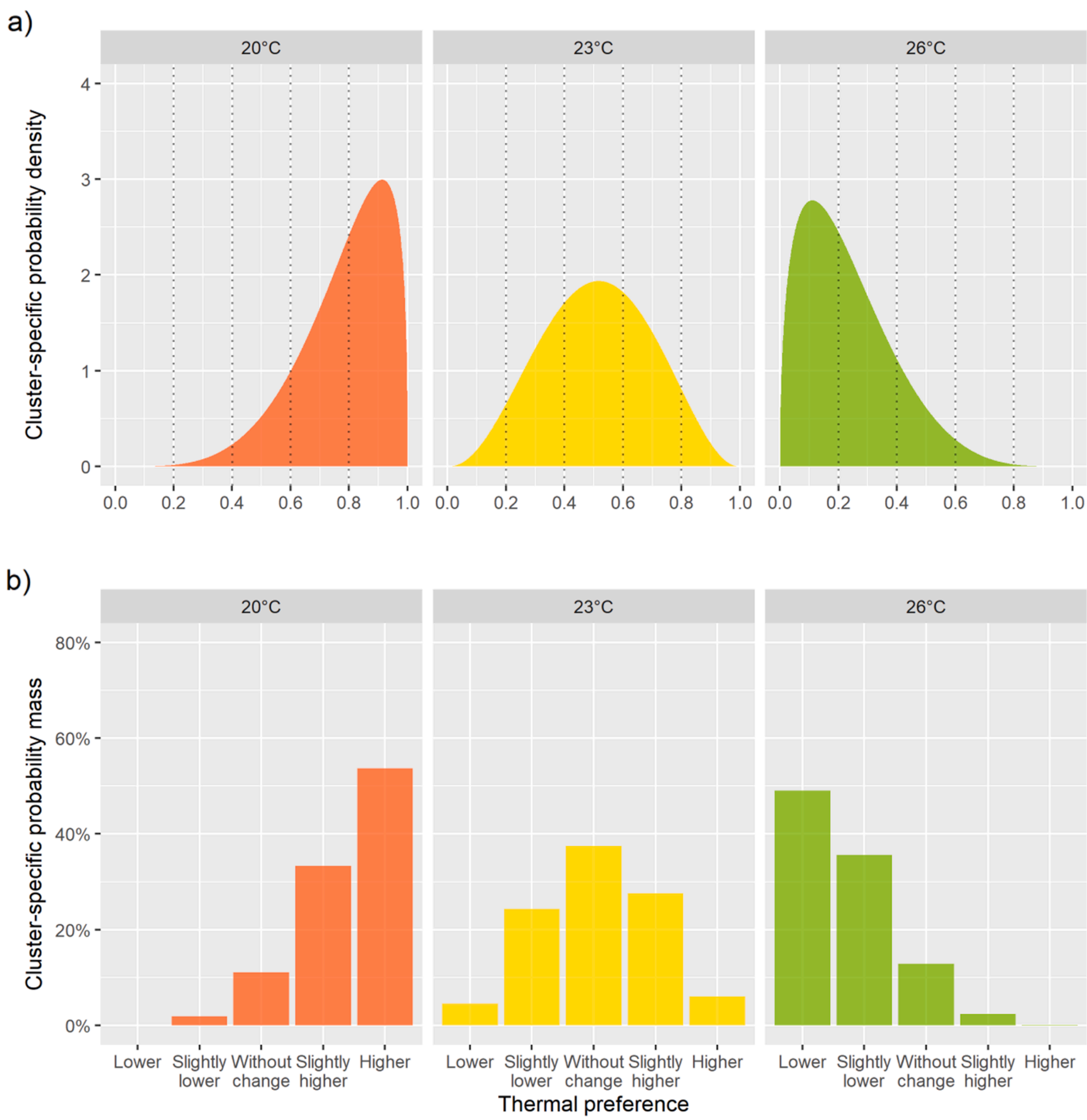

Fig. 8. (a) Probability densities and (b) categorised probabilities of the predicted response using the cluster-specific procedure for three different operative temperatures. Note. The dotted lines in (a) represent the thresholds used for categorisation.

terms of predicted probabilities (see Fig. 6, Fig. 8 and Fig. 9). However, it is important to point out that this method poses a limitation: a different categorisation of the beta distribution would lead to different probabilities. The same applies for the categorisation of the thermal preference vote used to estimate the ordinal model (see Fig. 3). Nevertheless, by comparing the probabilities estimated by the two models, the following general observations can be made. On the one hand, the ordinal model is more flexible in the sense that it can handle different probability distributions (virtually any probability distribution). For example, in Fig. 6, it can handle the spike in the probabilities for the 'without change' category for an operative temperature of 23 °C. On the other hand, the beta model is more detailed since it provides a pdf. For example, in Fig. 8(a), the predicted probability of observing a thermal preference vote between 0.45 and 0.55 for an operative temperature of 23 °C is 19.2 %. An alternative approach to comparing the two models would be to calculate the mean of the estimated probabilities for the ordinal model and contrast it with the predicted mean response of the beta model. The mean of the probabilities can be written as:

$$\text{Mean Pr} = \sum_{1}^{K} \pi_k k \tag{6}$$

where π_k is the probability of a specific category k, $k \in \{1, \dots, K\}$. Here, the category prefer 'lower' was mapped to 1 and the category prefer 'higher' was mapped to 5. The resulting mean probabilities were then rescaled between + 0.001 and + 0.999 to match the predicted mean response of the beta model. Fig. 10 shows this comparison. For the ordinal model, the cumulative probability γ_k is the inverse of the link function, which in this study corresponds to the inverse of the logit:

$$\gamma_k = \text{Logit}^{-1}(\mathbf{1}\tau_k - \boldsymbol{\eta}) = \text{Logistic}(\mathbf{1}\tau_k - \boldsymbol{\eta}) = \frac{1}{1 + e^{-(\mathbf{1}\tau_k - \boldsymbol{\eta})}} = \frac{1}{1 + e^{-(\mathbf{1}\tau_k - X\beta - Zu)}} \tag{7}$$

The probability of a specific category k is calculated as $\pi_k = \gamma_k - \gamma_{k-1}$. The probabilities of all the categories are then used to calculate the mean as in Eq. (6). More details about the mathematical formulation can be found in the Appendix A.

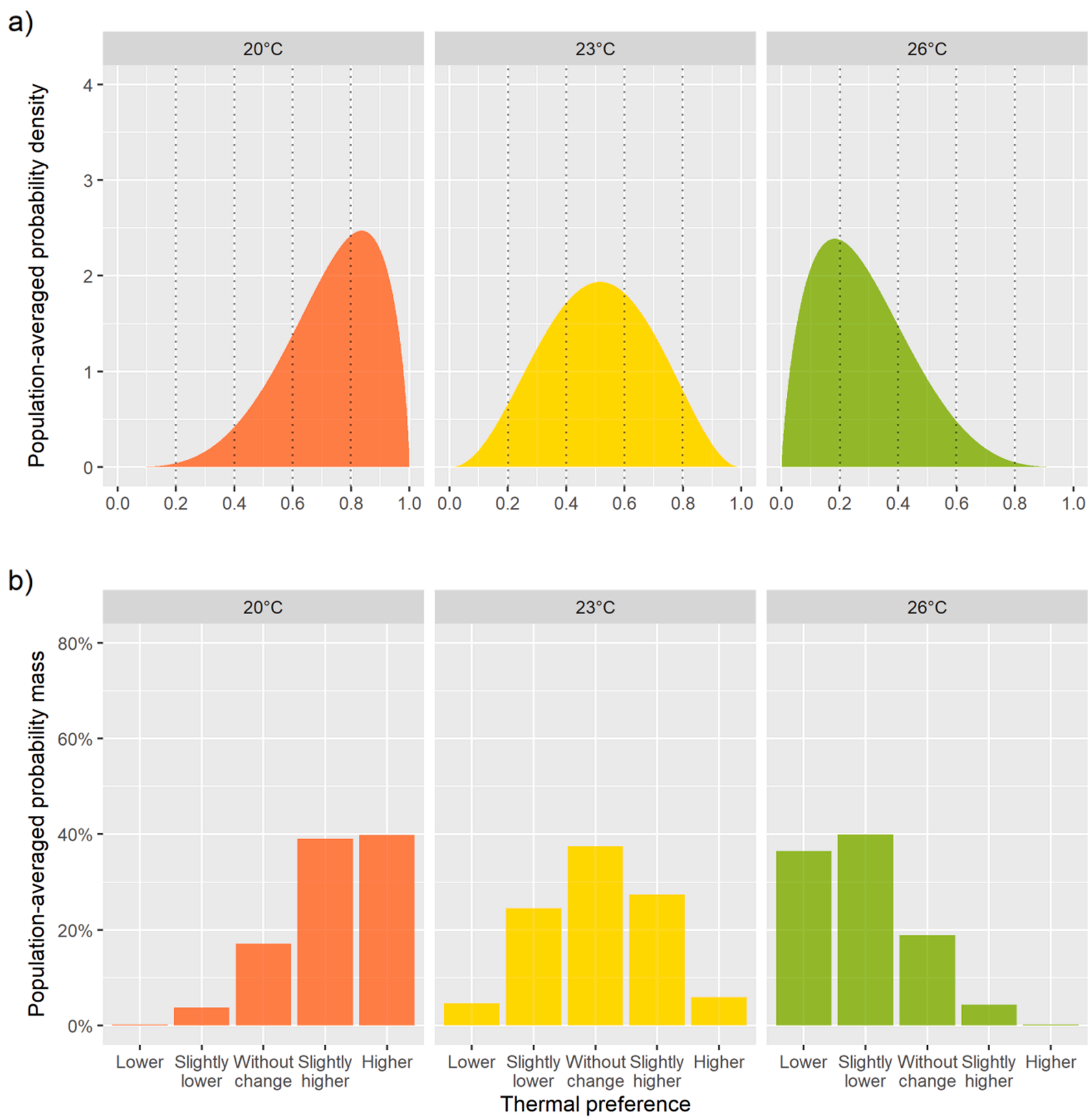

Fig. 9. (a) Probability densities and (b) categorised probabilities of the predicted response using the population-averaged procedure for three different operative temperatures. Note. The dotted lines in (a) represent the thresholds used for the categorisation.

Table 7
Predictors' relative importance for both the beta and ordinal models.

Modelling strategy	Predictor		Standardise coeff	Rank*
Ordinal model	*Clothing*		−0.464	4
	BMI		−0.567	2
	Air.vel		−0.163	5
	Time.day	morning	Reference	
		afternoon	−0.474	3
	Top		−2.917	1
Beta model	*Clothing*		−0.172	4
	BMI		−0.199	2
	Time.day	morning	Reference	
		afternoon	−0.177	3
	Vap.pre		−0.102	5
	Top		−0.793	1

* the higher the absolute value of the standardise coefficient, the higher the rank.

As in Fig. 7, all the lines in Fig. 10 are plotted as a function of the operative temperature while the other covariates (i.e., the fixed effects) are held constant at their centred values, but they differ in the random effects. Specifically:

- The solid black line (beta model, cluster-specific procedure) has the random effect fixed at zero (as in Fig. 7);
- The solid red line (beta model, population-averaged procedure) has the random effect derived from simulation (as in Fig. 7);
- The dashed black line (ordinal model, cluster-specific procedure) has the random effect fixed at zero;
- The dashed red line (ordinal model, population-averaged procedure) has the random effect derived from simulation.

It can be seen that the curve produced by using the cluster-specific procedure for the ordinal model has three inflexion points. This particular behaviour can be explained by looking at the predicted probabilities in Fig. 5(a). Between the operative temperatures of 22–24 °C, the predicted probabilities for 'without change' were much greater than all the others (from 80 % up to more than 90 %). Consequently, within this range, the calculated mean was greatly affected by these probabilities. The same behaviour can be observed for the population-averaged curve

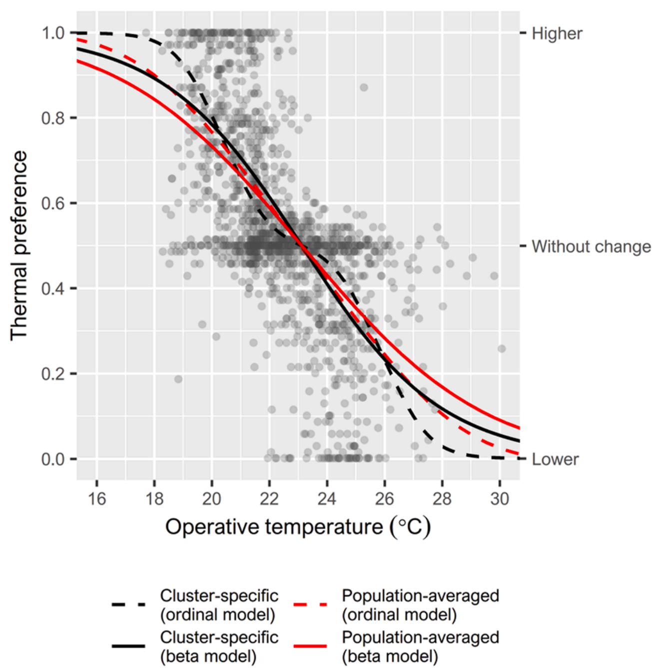

Fig. 10. Predicted responses using the cluster-specific (black solid and dashed lines) and population-averaged (red solid and dashed lines) procedures for the beta and ordinal models, respectively. Note. The points are the observed thermal preference votes.

to a lesser extent. However, more considerable differences are visible at the tails of the curves, that is, the two extremities. Here, the beta model's mean response curve has tails that are heavier than the mean of the estimated probabilities for the ordinal model. Despite these differences, both the beta and ordinal models are both valid strategies for modelling thermal preference votes. However, the choice between the two models should be made based on how the response variable is measured. The beta model is a suitable choice when the thermal preference votes are measured on a continuous, but bounded, scale. In contrast, the ordinal model is appropriate when a categorical scale is used. Unfortunately, in the thermal comfort field, it is common practice to analyse subjective human thermal responses independently of how they were measured [47].

Furthermore, both models identified the largest random-effect component at the 'experimental condition' level, indicating that differences in conditions were the primary source of variability (see Table 5 and Table 6). Therefore, the variability in how individuals react to different dynamic conditions is higher than the variability between individuals, which may indicate that there could be unmodeled information in the variance at this level (i.e., the 'experimental condition'). For instance, the different rates of temperature change could play a role both in terms of absolute magnitude (i.e., the specific value of the rate of change) and sign (i.e., heating or cooling). To this aim, modelling strategies that model the effect of temporal patterns directly should be used (e.g., time series). Furthermore, if the variability of the environmental condition were significantly reduced (i.e., a static environment created), the inter-individual differences (i.e., difference between individuals) would be dominant.

4.3. Approaches' comparison

The predicted thermal preference votes were calculated from the two models using two different approaches: fixing the random effects at their mean of zero (cluster-specific procedure) and using a simulation approach with $M = 1{\cdot}10^4$ (population-averaged procedure). Regarding the ordinal model, from Fig. 6 it can be seen that the most evident difference between the cluster-specific and the population-averaged procedures are the predicted probabilities for an operative temperature of 23 °C. Here, the predictive probability for 'without change' is about 91 % for the cluster-specific approach, while it is only 55 % for the population-averaged one. The reason for the discrepancy lies in the fact that the level 2 variance ($\widehat{\sigma}^2_{\text{ramp\_ID}} = 2.512$) and the level 3 variances ($\widehat{\sigma}^2_{\text{day\_ID}} = 1.017$ and $\widehat{\sigma}^2_{\text{participant\_ID}} = 1.340$) are not close to zero. As the between-cluster variances $\widehat{\sigma}^2_{\text{ramp\_ID}}$, $\widehat{\sigma}^2_{\text{day\_ID}}$ and $\widehat{\sigma}^2_{\text{participant\_ID}}$ in the random-intercepts model increase, the curves will be further apart. For an example the reader is referred to Appendix C. The advantage of having predictive probabilities as outcomes is that they are their own error measures. In Fig. 6, the predicted probability of 'without change' for the cluster-specific approach is 91 %; if one decided not to choose this category as the expected outcome, the probability of this being an error is, by definition, 91 %. Following the same reasoning, for the population-averaged approach, not selecting 'without change' as the expected outcome has a 55 % probability of being an error. As a standard practice, the ordinal model regards the category with the highest probability as the predicted outcome (i.e., thermal preference vote). However, utilising a hard threshold, such as the automatic selection of the category with the higher probability, does not fully use the information contained in the probabilities. For example, in Fig. 5(b), such a threshold would lead to 'slightly lower' and 'slightly higher' never being selected. Here the necessity of defining a utility/cost function that, for example, maximises the expected utility or minimises the expected cost. With regard to the beta model, from Fig. 8 and Fig. 9 it can be seen that, for both the cluster-specific and population-averaged procedures, the distributions of the probability densities (and analogously, the categorised probabilities) for an operative temperature equal to 23 °C are the same. The predicted mean response of the beta model intersects the thermal preference vote at 0.5, at which the prediction at $u_{\text{ramp\_ID}} = u_{\text{day\_ID}} = u_{\text{participant\_ID}} = 0$ (the median) equals the mean prediction (see Fig. 7). The median (i.e., cluster-specific) curve is lower than the population-averaged curve for a predicted thermal preference vote lower than 0.5 but is higher for a predicted thermal preference vote higher than 0.5. Consequently, the cluster-specific probability densities (i.e., the median probabilities) become skewed faster than the population-averaged ones (i.e., the mean probabilities) at operative temperatures higher or lower than 23 °C (see Fig. 8 and Fig. 9 for the operative temperatures of 20 °C and 26 °C).

4.4. Limitations

This study's limitations arise from some simplifications introduced during the statistical modelling. For both models, the functional form was assumed to be linear for simplicity (see Eq. (3)). As a consequence, the models do not account for potential nonlinearities. However, nonlinearities could be considered, for example, by using smoothing splines. Another simplification derives from assuming that all the independent variables were measured exactly, that is, 'error-free'. When covariates are measured with errors, the parameter estimates do not tend to the true values, even in extensive samples. For simple linear regression, this effect is known as the attenuation bias and leads to an underestimation of the coefficient. For more complex methods, such as multilevel models, this issue deserves a proper treatise and is beyond the scope of this study.

For a beta model, the conditional variance is $var(Y_i|\boldsymbol{U} = \boldsymbol{u}) = \mu_i(1-\mu_i)/(1+\phi)$, where μ_i is the mean and ϕ is the precision parameter. The parameter ϕ is known as the precision parameter because for fixed μ_i, the larger the ϕ, the smaller the variance of Y_i. Therefore, the variance is not constant but rather a function of the mean and the precision parameter, here assumed to be constant. However, the precision parameter can be modelled as a function of some predictors, for example, the operative temperature. In this study, this possibility was not explored and should examined in future studies.

To apply an ordinal model, the dependent variable must be categorical. For this reason, the dependent variable was binned into five categories according to the thresholds -0.6, -0.2, $+0.2$, and $+0.6$. However, these cut-off points were arbitrary and indirectly assumed to be the same for all participants. When a categorical scale is used to measure the dependent variable, this choice is made directly by single responders. Consequently, it is unlikely to be the same for all responders (see, for example, [48]). While this study used categorisation to apply the ordinal model, we do not encourage this practice in 'normal' circumstances. It is more appropriate to measure the variable directly with a categorical scale. As stated earlier, cut-off points are arbitrary and generally do not have practical/scientific meaning. Furthermore, the ordinal model has an additional assumption called 'proportional odds' (or equal slope assumptions). This assumption implies that the threshold parameters τ_k are independent of the regression variables or, equivalently, that the regression parameters are not allowed to vary with k, a specific category of Y. This restriction derives from the fact that the thresholds are theoretically linked with the response measure (and therefore assumed to be part of the measurement procedure), not to the predictor's value. However, the *ordinal* package [36], does not yet implement this feature when there is more than one random effect. Therefore, testing for partial and non-proportional odds (called 'nominal effects' in [36]) was not possible.

Moreover, multilevel models offer additional modelling possibilities that this paper has not discussed. For instance, in both the models used in this study, the slope coefficients of the predictors added to the models were assumed to be fixed across higher-level units. However, it is possible that the relationships between the responses and the predictors vary across these higher classification units (e.g., between participants). In multilevel models, it is possible to allow the effects of the predictors to vary randomly across higher classification units by adding a random slope to the model, which can be translated into checking whether, for example, the effect of the temperature differs across different occupants and to what degree of magnitude.

5. Conclusions and future perspectives

This study aimed to predict the thermal preference votes of human subjects exposed to a dynamic thermal environment. To this aim, two different statistical models were proposed, namely the beta and ordinal models. Based on the analyses carried out, the following points can be made concerning the two models:

- A three-level cross-classified model fit the data significantly better than the simpler three-levels models, the two-level model and the single-level model nested within it. Nevertheless, it is important to note that the multilevel structure described here is not a characteristic of the model. It is a feature of the experimental/study design represented by the data and encapsulated in the model. Therefore, it is essential to know how, where and when the data were collected (i.e., the metadata) to model them appropriately.
- In predictive modelling, direct interpretability regarding the model is not required; however, transparency is desirable. Multilevel models can model complex structures and at the same time, offer the advantage of having transparent outputs and modelling steps in contrast to machine learning/data-driven algorithms, which are basically 'black box' models.
- While likelihood-based statistics (e.g., Akaike information criterion, Bayesian information criterion) cannot be used to contrast the two models' performances directly, some qualitative observations can be made by comparing the probabilities estimated by the two models. On the one hand, the ordinal model is more flexible in the sense that it can handle different probability distributions (virtually any probability distribution). On the other hand, the beta model is more detailed because it provides a pdf. The two models used in this study are both valid strategies for modelling thermal preference votes. However, the choice between the ordinal and the beta models should be made based on how the response variable is measured.

Furthermore, two distinct procedures were used in this study, namely the cluster-specific and the population-averaged procedures, to predict the thermal preference votes. These two methods apply directly to the concept of occupant-centric building design and operation. The population-averaged approach is suitable for the occupant-centric building design phase, where the target is the 'general' occupant. On the other hand, during the building operation phase, the notion of a 'general' occupant is pointless, and the focus should be on satisfying the needs of the specific occupant. In this case, a cluster-specific procedure is appropriate. This procedure can be carried out by measuring the specific occupant response to the environment and consequently updating the probabilities of the population-averaged procedure. These procedures could be used to design more energy-efficient and satisfying control strategies according to occupants' feedback(e.g., Ref [49]) in an occupant-centric [20] or human-in-the-loop [21] approach.

CRediT authorship contribution statement

Matteo Favero: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Jan Kloppenborg Møller:** Conceptualization, Formal analysis, Methodology, Validation, Writing – review & editing. **Davide Calì:** Conceptualization, Methodology, Writing – review & editing. **Salvatore Carlucci:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

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

The authors do not have permission to share data.

Acknowledgements

Matteo Favero would like to thank the Research Centre on Zero Emission Neighbourhoods in Smart Cities (FME ZEN, Grant no. 257660) and the Research Council of Norway (Norges Forskingsrådet) for their support.

Davide Calì and Jan Kloppenburg Møller received funding from the project "Human-in-the-Loop Digitalisation and Energy Management of buildings, HuiL-DEMand" (Program: EUDP, demonstration project, Journal number 64021-1021). Davide Calì also received funding from the project "Sustainable plus energy neighbourhoods, syn.ikia" (H2020 No. 869918).

This work has been performed within the framework of the International Energy Agency – Energy in Buildings and Communities Program (IEA-EBC) Annex 79 'Occupant-centric building design and operation'.

Open Access Funding by NTNU (Norges teknisk-naturvitenskapelige universitet).

Appendix A. Mathematical background and examples

In this study-two modelling strategies are used, specifically the beta mixed-effects model with the logit link (simply referred as beta model) and the ordinal mixed-effects model with the logit link (simply referred as ordinal model). Both these models belong to a broad class of models called generalized linear mixed models.

There are four components that are common to any generalised linear mixed models:

- Random component (of the response variable): Specifies the probability distribution of the response variable, such as the normal distribution for $\boldsymbol{Y}$ (i.e., $\boldsymbol{Y} \sim \text{Normal}(\boldsymbol{\mu}, \sigma^2)$) in the (classical) linear regression model. In general, there is no separate error term. Classical linear regression is a special case in which the error term can be extract from the distributional assumption (i.e., $\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\mu} = \boldsymbol{X\beta}$ and $\boldsymbol{\varepsilon} \sim \text{Normal}(\boldsymbol{0}, \sigma^2)$);
- Link function: Specifies the link between the random and the systematic components. It denotes the relationship between the predicted response value (e.g., the mean) and the covariates;
- Systematic component: Specifies the covariates in the model, more specifically, how they are combined (usually through a linear combination);
- Random component (of the random effect): Specifies the probability distribution of the random effects, usually assuming a normal distribution with zero mean (i.e., $\boldsymbol{u} \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{\Sigma})$).

Below, for both beta and ordinal model, the mathematical formulation and examples are provided.

Beta model

The beta model, as specified in Eq. (1), is:

$$\begin{aligned} &\boldsymbol{Y} \sim \text{Beta}(\boldsymbol{\mu}, \phi) \\ &\text{Logit}(\boldsymbol{\mu}) = \boldsymbol{\eta} \\ &\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Zu} \\ &\boldsymbol{u} \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{\Sigma}) \end{aligned} \tag{A1}$$

where the conditional distribution of the response variable $\boldsymbol{Y}$ is assumed to follow a beta distribution. Here, its expected value (i.e., its mean $\boldsymbol{\mu}$) is linked to linear predictor $\boldsymbol{\eta}$ through a logit function. The logit function is defined as the inverse of the cumulative distribution function (cdf) of the standard[6] logistic distribution:

$$\text{Logit}^{-1}(\boldsymbol{\eta}) = \text{Logistic}(\boldsymbol{\eta}) = \frac{1}{1 + e^{-\boldsymbol{\eta}}} \tag{A2}$$

Eq. (A1) is expressed in matrix notation. For a specific case it can be written as:

$$\begin{aligned} &Y_{pdri} \sim \text{Beta}(\mu_{pdri}, \phi) \\ &\text{Logit}(\mu_{pdri}) = \eta_{pdri} \\ &\eta_{pdri} = \mathbf{x}_{pdri}^{\mathrm{T}}\boldsymbol{\beta} + u_p + u_d + u_r \\ &u_p \sim \text{Normal}\left(0, \sigma_p^2\right);\ \ \text{iid} \\ &u_d \sim \text{Normal}\left(0, \sigma_d^2\right);\ \ \text{iid} \\ &u_r \sim \text{Normal}\left(0, \sigma_r^2\right);\ \ \text{iid} \end{aligned} \tag{A3}$$

where the subscript p indicates a participant, d a day, r a thermal ramp, and i is the i^{th} obsevation.

If for example, we consider:

- two participants (no. 1 and 2), where participant 1 visited the lab on days 1 and 2 while participant 2 visited the lab on days 2 and 3,
- each day has two ramps (named 1–6) and only two observation per ramp.

We obtain:

the 16×1 vector of the response variable

$$\boldsymbol{Y} = [Y_{1111}, Y_{1112}, Y_{1121}, Y_{1122}, Y_{1231}, Y_{1232}, Y_{1241}, Y_{1242}, Y_{2231}, Y_{2232}, Y_{2241}, Y_{2242}, Y_{2351}, Y_{2352}, Y_{2361}, Y_{2362},]^{\mathrm{T}}$$

the 11×1 vector of the random effects

$$\boldsymbol{u} = \left[u_{p(1)}, u_{p(2)}, u_{d(1)}, u_{d(2)}, u_{d(3)}, u_{r(1)}, u_{r(2)}, u_{r(3)}, u_{r(4)}, u_{r(5)}, u_{r(6)}\right]^{\mathrm{T}}$$

with variance–covariance matrix equal to

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_p^2\boldsymbol{I}_2 & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_d^2\boldsymbol{I}_3 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \sigma_r^2\boldsymbol{I}_6 \end{bmatrix}$$ where $\boldsymbol{I}_n$ is the identity matrix of dimension n

[6] The logistic distribution is defined by two parametes: a location parameter μ and a scale parameter s. When $\mu = 0$ and $s = 1$, the logistic distribution is called standard logistic distribution, that is: $\text{Logistic}(x; \mu = 0, s = 1) = \frac{1}{1+e^{-(x-\mu)/s}} = \frac{1}{1+e^{-x}}$

the 16 × 11 design matrix of the random effects

$$\boldsymbol{Z} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Ordinal model

The ordinal model, as specified in Eq. (2), is:

$$\begin{aligned} &\boldsymbol{Y} \sim \text{Multinomial}(n, \boldsymbol{\pi}) \\ &\text{Logit}(\boldsymbol{\gamma}_k) = \mathbf{1}\tau_k - \boldsymbol{\eta} \\ &\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Zu} \\ &\boldsymbol{u} \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}) \end{aligned} \tag{A4}$$

where the conditional distribution of the response variable $\boldsymbol{Y}$ is assumed to follow a multinomial distribution with probability vector $\boldsymbol{\pi} = \{\pi_1, ..., \pi_k\}$, where $\pi_k = \Pr(Y = k)$. The cumulative probability corresponding to π_k is $\gamma_k = \Pr(Y \leq k)$; hence, $\gamma_k = \pi_1 + ... + \pi_k$. Here the cumulative probabilities are then mapped to the real numbers through a logit function. In the ordinal model, the logit is function of the linear predictor $\boldsymbol{\eta}$ and $\mathbf{1}\tau_k$, the vector of laten thresholds parameters τ_k, with $k \in \{1, ..., K\}$. That is:

$$\boldsymbol{\gamma}_k = \text{Logit}^{-1}(\mathbf{1}\tau_k - \boldsymbol{\eta}) = \text{Logistic}(\mathbf{1}\tau_k - \boldsymbol{\eta}) = \frac{1}{1 + e^{-(\mathbf{1}\tau_k - \boldsymbol{\eta})}} \tag{A5}$$

Eq. (A4) is expressed in matrix notation. For a specific case it can be written as:

$$\begin{aligned} &\Pr\left(Y_{pdri} = k\right) = \pi_k \\ &\gamma_{pdri,k} = \Pr\left(Y_{pdri} \leq k\right) = \pi_1 + ... + \pi_k \\ &\text{Logit}\left(\gamma_{pdri,k}\right) = \tau_k - \eta_{pdri} \\ &\eta_{pdri} = \mathbf{x}_{pdri}^{\mathrm{T}}\boldsymbol{\beta} + u_p + u_d + u_r \\ &u_p \sim \text{Normal}\left(0, \sigma_p^2\right); \;\; \text{iid} \\ &u_d \sim \text{Normal}\left(0, \sigma_d^2\right); \;\; \text{iid} \\ &u_r \sim \text{Normal}\left(0, \sigma_r^2\right); \;\; \text{iid} \end{aligned} \tag{A6}$$

where the subscript p indicates a participant, d a day, r a thermal ramp, i is the i^{th} observation, and k is a category of the dependent variable.

If for example, we consider:

- two participants (no. 1 and 2), where participant 1 visited the lab on days 1 and 2 while participant 2 visited the lab on days 2 and 3,
- each day has two ramps (named 1–6) and only two observation per ramp.

We obtain, for a specific category k:

the 16 × 1 vector of the response variable

$$\boldsymbol{Y} = [Y_{1111}, Y_{1112}, Y_{1121}, Y_{1122}, Y_{1231}, Y_{1232}, Y_{1241}, Y_{1242}, Y_{2231}, Y_{2232}, Y_{2241}, Y_{2242}, Y_{2351}, Y_{2352}, Y_{2361}, Y_{2362},]^{\mathrm{T}}$$

the 11 × 1 vector of the random effects

$$\boldsymbol{u} = \left[u_{p(1)}, u_{p(2)}, u_{d(1)}, u_{d(2)}, u_{d(3)}, u_{r(1)}, u_{r(2)}, u_{r(3)}, u_{r(4)}, u_{r(5)}, u_{r(6)}\right]^{\mathrm{T}}$$

with variance–covariance matrix equal to

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_p^2 \boldsymbol{I}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_d^2 \boldsymbol{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_r^2 \boldsymbol{I}_6 \end{bmatrix}$$ where $\boldsymbol{I}_n$ is the identity matrix of dimension n

the 16 × 11 design matrix of the random effects

$$
\mathbf{Z} = \begin{bmatrix}
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

Appendix B. Frequency distributions for the independent variables

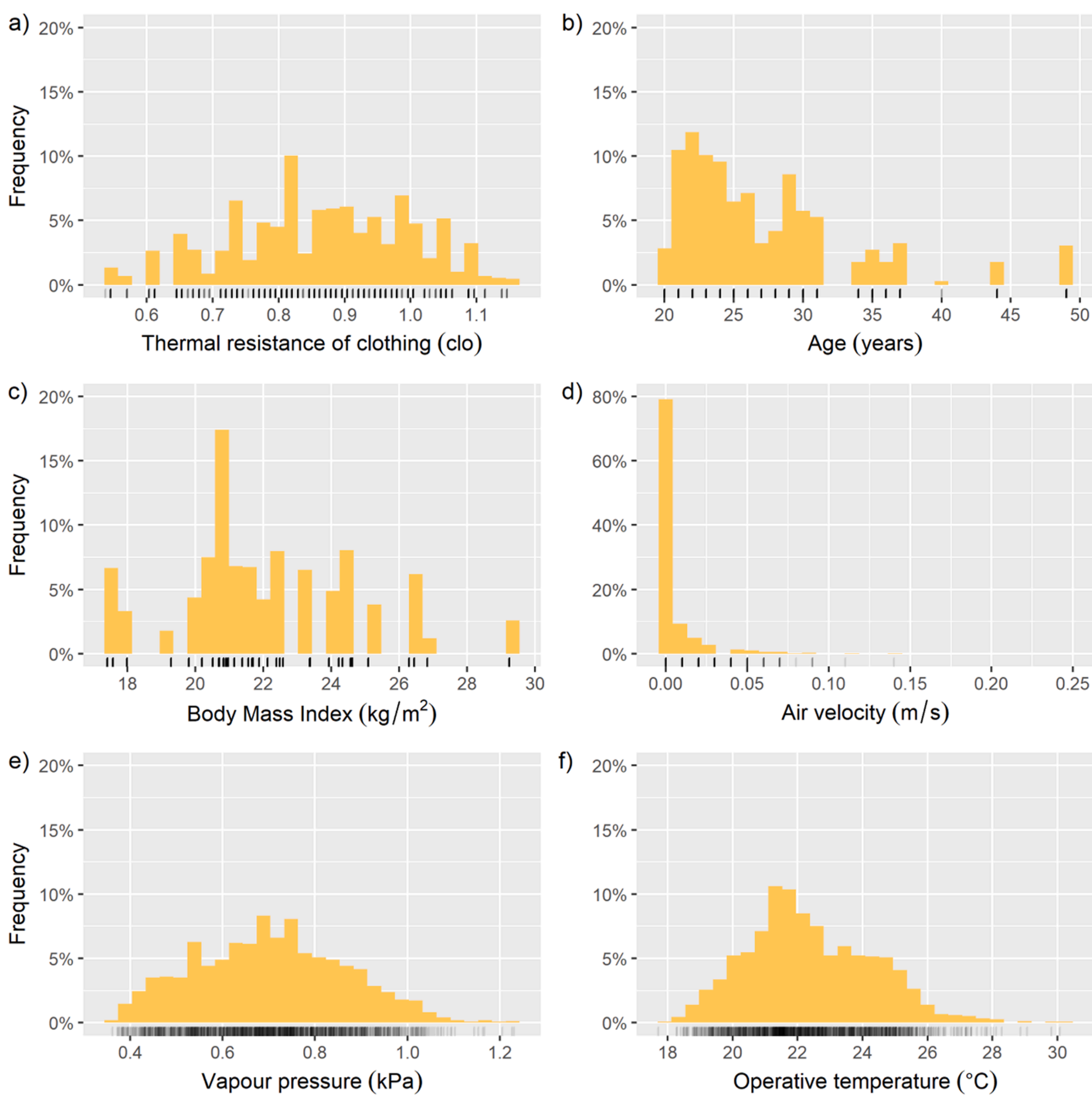

Fig. B1. Frequency distributions for the continuous variables: (a) thermal resistance of clothing, (b) age, (c) Body Mass Index, (d) air velocity, (e) vapour pressure and (f) operative temperature. Note. The vertical black marks at the bottom of each figure are the rug plots.

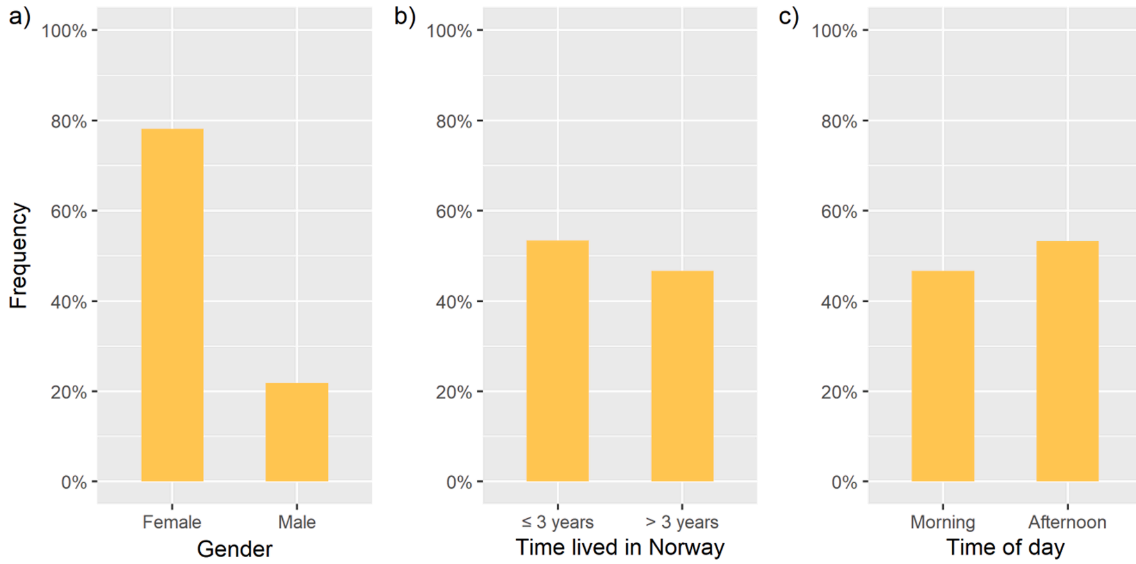

Fig. B2. Frequency distributions for the categorical variables: (a) gender, (b) time lived in Norway and (c) time of day.

Appendix C. Population-averaged and cluster-specific predictions

In Fig. 6, the differences between the predicted probabilities for the cluster-specific and the population-averaged procedures depend on the between-cluster variances. As the between-cluster variances $\hat{\sigma}^2_{\text{ramp\_ID}}$, $\hat{\sigma}^2_{\text{day\_ID}}$ and $\hat{\sigma}^2_{\text{participant\_ID}}$ in the random-intercepts model increase, the curves will be further apart. This situation can be observed in Fig. C1 for category $k \leq 3$ (i.e., from the 'lower' to 'without change' category).

The predicted cumulative response probabilities always intersect at $p = 50\%$, the point at which the prediction at $u_{\text{ramp\_ID}} = u_{\text{day\_ID}} = u_{\text{participant\_ID}} = 0$ (the median) equals the mean prediction. The median (i.e., cluster-specific) curve is lower than the population-averaged curve for predicted probabilities lower than 50 %, while the median curve is higher for probabilities greater than 50 %. As a result, for a given range of *Top* values, the cluster-specific predicted probabilities will be greater than the population-averaged ones. In this case, for the category $k = 3$ (i.e., 'without change'), for *Top* values between 20.8 °C and 25.4 °C, the cluster-specific predicted probabilities are higher than the population-averaged ones (see Fig. C2). The dashed black lines in both Fig. C1 and Fig. C2 are cluster-specific effects, where the random effects $u_{\text{ramp\_ID}}$, $u_{\text{day\_ID}}$ and $u_{\text{participant\_ID}}$ were set to their 16th and 84th percentiles (i.e., $u_{\text{ramp\_ID}} = \{-1.11, 1.19\}$, $u_{\text{day\_ID}} = \{-0.56, 0.69\}$ and $u_{\text{participant\_ID}} = \{-1.05, 0.93\}$) and then added (i.e., $\{-2.71, 2.80\}$). The 16th and 84th percentiles were chosen because they correspond roughly to ± 1 standard deviation above and below the mean and encompass about 68 % of the observed random effects for *Ramp_ID*, *Day_ID* and *Participant_ID*.

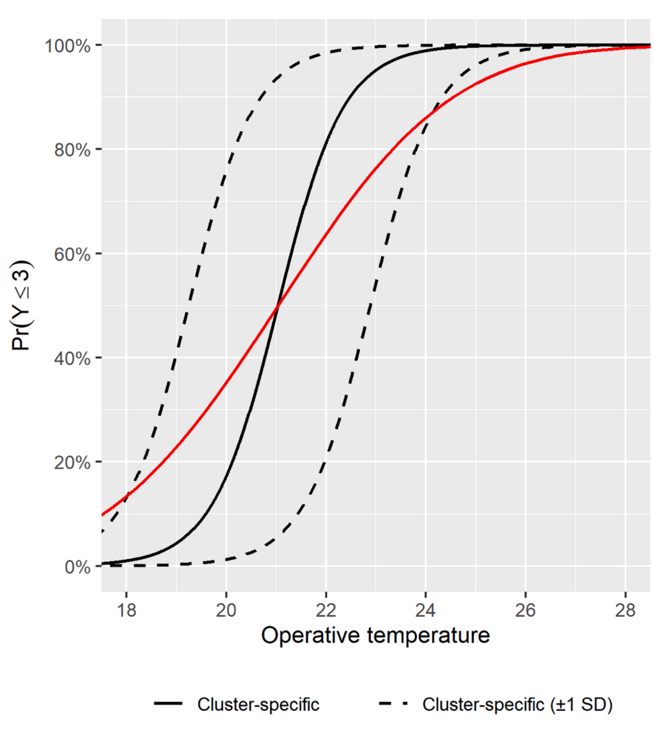

Fig. C1. Predicted cumulative response probabilities for category $k \leq 3$ (i.e., from the 'lower' to 'without change' category) using the cluster-specific (black line) and population-averaged (red solid line) procedures.

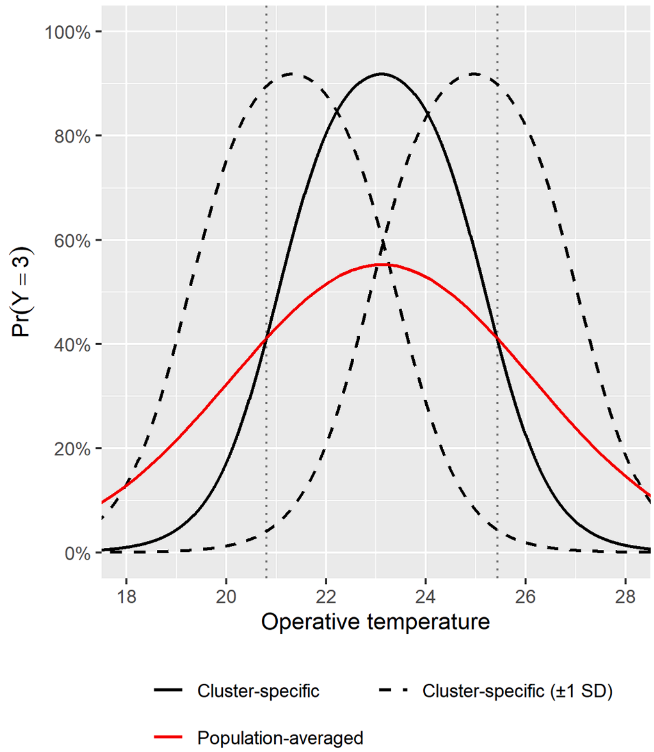

Fig. C2. Predicted response probabilities for category $k = 3$ (i.e., 'without change') using the cluster-specific (black line) and population-averaged (red line) procedures. Note. The dotted black lines represent the intersections between the cluster-specific and population-averaged probabilities.

References

[1] ANSI/ASHRAE Standard 55, Thermal environmental conditions for human occupancy, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, 2020.

[2] 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, International Organization for Standardization, Geneva, Switzerland, 20s05.

[3] CEN 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 - Module Ml-6, European Committee for Standardization, Brussels, Belgium, 2019.

[4] CEN EN 15251, Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics, European Committee for Standardization, Brussels, Belgium, 2007.

[5] Fanger PO. Thermal comfort: analysis and applications in environmental engineering. Copenhagen, Denmark: Danish Technical Press; 1970.

[6] Nicol JF, Humphreys MA. Thermal comfort as part of a self-regulating system. Build Res Pract 1973;1(3):174–9. https://doi.org/10.1080/09613217308550237.

[7] de Dear RJ, Brager RJ. Developing an adaptive model of thermal comfort and preference, ASHRAE Transactions, Amer Soc Heating Ref Air-Conditoning Eng Inc, San Francisco, CA, 1998, pp. 27–49.

[8] Humphreys MA, Nicol JF. Understanding the adaptive approach to thermal comfort, ASHRAE Transactions, Amer Soc Heating, Ref Air-Conditoning Eng. Inc., San Francisco, CA, 1998, pp. 991–1004.

[9] de Dear RJ, Brager GS, Cooper D. Developing an adaptive model of thermal comfort and preference -. Final Rep ASHRAE RP-884 1997.

[10] McCartney KJ, Fergus Nicol J. Developing an adaptive control algorithm for Europe. Energy Build 2002;34(6):623–35. https://doi.org/10.1016/S0378-7788(02)00013-0.

[11] Arens E, Humphreys MA, de Dear RJ, Zhang H. Are 'class A' temperature requirements realistic or desirable? Build Environ 2010;45(1):4–10. https://doi.org/10.1016/j.buildenv.2009.03.014.

[12] de Dear RJ. A global database of thermal comfort field experiments, ASHRAE Transactions, Amer. Soc. Heating, Ref. Air-Conditoning Eng. Inc., San Francisco, CA, 1998, pp. 1141–1152.

[13] Brager GS, Paliaga G, de Dear RJ. Operable windows, personal control, and occupant comfort, ASHRAE Transactions, Am Soc Heating, Ref Air-Conditoning Eng. Inc., Nashville, TN, 2004, pp. 17–35.

[14] de Dear RJ, Xiong J, Kim J, Cao B. A review of adaptive thermal comfort research since 1998. Energy Build 2020;214:109893. https://doi.org/10.1016/j.enbuild.2020.109893.

[15] Cena K, de Dear RJ. Field study of occupant comfort and office thermal environments in a hot, arid climate, ASHRAE Trans Amer Soc Heating, Ref Air-Conditoning Eng. Inc., Seattle, WA, 1999, pp. 204–217.

[16] de Dear RJ, Fountain ME. Field experiments on occupant comfort and office thermal environments in a hot-humid climate, ASHRAE Trans, Amer Soc Heating Ref Air-Conditoning Eng. Inc., Orlando, FL, 1994, pp. 457–474.

[17] Schiller G, Arens EA, Bauman F, Benton C, Fountain M, Doherty T. A field study of thermal environments and comfort in office buildings, ASHRAE Trans Am Soc Heating, Ref Air-Conditoning Eng Inc., Ottawa, ON, 1988, pp. 280–308.

[18] Yao R, Zhang S, Du C, Schweiker M, Hodder S, Olesen BW, Toftum J, Romana d'Ambrosio F, Gebhardt H, Zhou S, Yuan F, Li B. Evolution and performance analysis of adaptive thermal comfort models – A comprehensive literature review. Build Environ 2022;217. https://doi.org/10.1016/j.buildenv.2022.109020.

[19] Hensen JLM. Literature review on thermal comfort in transient conditions. Build Environ 1990;25(4):309–16. https://doi.org/10.1016/0360-1323(90)90004-B.

[20] Park JY, Ouf MM, Gunay B, Peng Y, O'Brien W, Kjærgaard MB, et al. A critical review of field implementations of occupant-centric building controls. Build Environ 2019;165:106351. https://doi.org/10.1016/j.buildenv.2019.106351.

[21] Jung W, Jazizadeh F. Human-in-the-loop HVAC operations: A quantitative review on occupancy, comfort, and energy-efficiency dimensions. Appl Energy 2019;239:1471–508. https://doi.org/10.1016/j.apenergy.2019.01.070.

[22] Humphreys MA, Fergus Nicol J. The validity of ISO-PMV for predicting comfort votes in every-day thermal environments. Energy Build 2002;34(6):667–84. https://doi.org/10.1016/S0378-7788(02)00018-X.

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

[24] Aryal A, Becerik-Gerber B, Lucas GM, Roll SC. Intelligent agents to improve thermal satisfaction by controlling personal comfort systems under different levels of automation. IEEE Internet Things J 2021;8(8):7089–100. https://doi.org/10.1109/JIOT.2020.3038378.

[25] Lee J, Ham Y. Physiological sensing-driven personal thermal comfort modelling in consideration of human activity variations. Build Res Information: Int J Res Dev Demonstration 2021;49(5):512–24. https://doi.org/10.1080/09613218.2020.1840328.

[26] Katić K, Li R, Zeiler W. Machine learning algorithms applied to a prediction of personal overall thermal comfort using skin temperatures and occupants' heating behavior. Appl Ergon 2020;85:103078. https://doi.org/10.1016/j.apergo.2020.103078.

[27] Guenther J, Sawodny O. Feature selection and Gaussian Process regression for personalized thermal comfort prediction. Build Environ 2019;148:448–58. https://doi.org/10.1016/j.buildenv.2018.11.019.

[28] Shan C, Hu J, Wu J, Zhang A, Ding G, Xu LX. Towards non-intrusive and high accuracy prediction of personal thermal comfort using a few sensitive physiological

parameters. Energy Build 2020;207:109594. https://doi.org/10.1016/j.enbuild.2019.109594.
[29] Favero M, Sartori I, Carlucci S. Human thermal comfort under dynamic conditions: an experimental study. Build Environ 2021;204:108144. https://doi.org/10.1016/j.buildenv.2021.108144.
[30] ANSI/ASHRAE Standard 55, Thermal environmental conditions for human occupancy, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, 2017.
[31] Salkind NJ. Encyclopedia of Research Design. California: Thousand Oaks; 2010.
[32] Tabachnick BG, Fidell LS. Using multivariate statistics. 6th ed., new international ed.,. Harlow: Pearson; 2014.
[33] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2022. https://www.R-project.org/.
[34] RStudio Team, RStudio: Integrated Development Environment for R, RStudio, PBC, Boston, MA, 2022. http://www.rstudio.com/.
[35] Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, et al. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. R J 2017;9(2):378–400. https://doi.org/10.32614/RJ-2017-066.
[36] Christensen RHB, ordinal—Regression Models for Ordinal Data, 2019. https://CRAN.R-project.org/package=ordinal.
[37] Voeten CC, buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression, 2022. https://CRAN.R-project.org/package=buildmer.
[38] Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.
[39] Harrell FEJ, Regression modeling strategies, with applications to linear models, logistic and ordinal regression, and Survival Analysis, 2015.
[40] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, New York, NY: Springer New York, New York, NY, 2017.
[41] James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: With applications in R. New York: Springer; 2013.
[42] Baldi S, Korkas CD, Lv M, Kosmatopoulos EB. Automating occupant-building interaction via smart zoning of thermostatic loads: A switched self-tuning approach. Appl Energy 2018;231:1246–58. https://doi.org/10.1016/j.apenergy.2018.09.188.
[43] Korkas CD, Baldi S, Kosmatopoulos EB. 9 - Grid-Connected Microgrids: Demand Management via Distributed Control and Human-in-the-Loop Optimization. In: I. Yahyaoui (Ed.), Advances in Renewable Energies and Power Technologies, Elsevier 2018, pp. 315–344.
[44] Korkas CD, Baldi S, Michailidis I, Kosmatopoulos EB. Occupancy-based demand response and thermal comfort optimization in microgrids with renewable energy sources and energy storage. Appl Energy 2016;163:93–104. https://doi.org/10.1016/j.apenergy.2015.10.140.
[45] Ngarambe J, Yun GY, Santamouris M. The use of artificial intelligence (AI) methods in the prediction of thermal comfort in buildings: energy implications of AI-based thermal comfort controls. Energy Build 2020;211:109807. https://doi.org/10.1016/j.enbuild.2020.109807.
[46] Daum D, Haldi F, Morel N. A personalized measure of thermal comfort for building controls. Build Environ 2011;46(1):3–11. https://doi.org/10.1016/j.buildenv.2010.06.011.
[47] Favero M, Luparelli A, Carlucci S. Analysis of subjective thermal comfort data: a statistical point of view [Manuscript submitted for publication], (2022).
[48] Schweiker M, André M, Al-Atrash F, Al-Khatri H, Alprianti RR, Alsaad H, Amin R, Ampatzi E, Arsano AY, Azadeh M, Azar E, Bahareh B, Batagarawa A, Becker S, Buonocore C, Cao B, Choi J-H, Chun C, Daanen H, Damiati SA, Daniel L, Vecchi RD, Dhaka S, Domínguez-Amarillo S, Dudkiewicz E, Edappilly LP, Fernández-Agüera J, Folkerts M, Frijns A, Gaona G, Garg V, Gauthier S, Jabbari SG, Harimi D, Hellwig RT, Huebner GM, Jin Q, Jowkar M, Kim J, King N, Kingma B, Koerniawan MD, Kolarik J, Kumar S, Kwok A, Lamberts R, Laska M, Lee MCJ, Lee Y, Lindermayr V, Mahaki M, Marcel-Okafor U, Marín-Restrepo L, Marquardsen A, Martellotta F, Mathur J, Mino-Rodriguez I, Mou D, Moujalled B, Nakajima M, Ng E, Okafor M, Olweny M, Ouyang W, Abreu ALPd, Pérez-Fargallo A, Rajapaksha I, Ramos G, Rashid S, Reinhart CF, Rivera MI, Salmanzadeh M, Schakib-Ekbatan K, Schiavon S, Shooshtarian S, Shukuya M, Soebarto V, Suhendri S, Tahsildoost M, Tartarini F, Teli D, Tewari P, Thapa S, Trebilcock M, Trojan J, Tukur RB, Voelker C, Yam Y, Yang L, Zapata-Lancaster G, Zhai Y, Zhu Y, Zomorodian Z. Evaluating assumptions of scales for subjective assessment of thermal environments – do laypersons perceive them the way, we researchers believe? Energy Build 2020;211. https://doi.org/10.1016/j.enbuild.2020.109761.
[49] Cali D, Thilker CA, Specht SA, Real JP, Madsen H, Olesen BW. Human in the Loop: perceived-based control as the key to enhance buildings' performance. Build Simul 2021 Conf 2022.