

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Cohort comfort models — Using occupant's similarity to predict personal thermal preference with less data

### Permalink

<https://escholarship.org/uc/item/9vk9c79b>

### Authors

Quintana, Matias  
Schiavon, Stefano  
Tartarini, Federico  
[et al.](#)

### Publication Date

2023

### DOI

10.1016/j.buildenv.2022.109685

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# Cohort comfort models — Using occupants’ similarity to predict personal thermal preference with less data

Matias Quintana<sup>1</sup>, Stefano Schiavon<sup>2</sup>, Federico Tartarini<sup>2</sup>, Joyce Kim<sup>3</sup>, Clayton Miller<sup>1,\*</sup>

<sup>1</sup>College of Design and Engineering, National University of Singapore (NUS), Singapore

<sup>2</sup>Center for the Built Environment, University of California Berkeley, U.S.

<sup>3</sup>Department of Civil and Environmental Engineering, University of Waterloo, Canada

\*Corresponding Author: [clayton@nus.edu.sg](mailto:clayton@nus.edu.sg), +65 81602452

---

## Abstract

Cohort Comfort Models (CCM) are introduced as a technique for creating a personalized thermal prediction for a new building occupant without the need to collect large amounts of individual comfort-related data. This approach leverages historical data collected from a sample population, who have some underlying preference similarity to the new occupant. The method uses background information such as physical and demographic characteristics and one-time onboarding surveys (satisfaction with life scale, highly sensitive person scale, personality traits) from the new occupant, as well as physiological and environmental sensor measurements paired with a few thermal preference responses. The framework was implemented using two personal comfort datasets containing longitudinal data from 55 people. The datasets comprise more than 6,000 unique right-here-right-now thermal comfort surveys. The results show that a CCM that uses only the one-time onboarding survey information of an individual occupant has generally as good or better performance as compared to conventional general-purpose models, but uses no historical longitudinal data as compared to personalized models. If up to ten historical personal preference data points are used, CCM increased the thermal preference prediction by 8% on average and up to 36% for half of the occupants in the first of the tested datasets. In the second dataset, one-third of the occupants increased their thermal preference prediction by 5% on average and up to 46%. CCM can be an important step toward the development of personalized thermal comfort models without the need to collect a large number of datapoints per person.

## Keywords:

Thermal comfort, Clustering, Personalized environments, Cold start, Warm start, Recommender systems

---

## Nomenclature

PMV	Predicted Mean Vote
PCM	Personal Comfort Model
PCS	Personal Comfort System
CCM	Cohort Comfort Model
RHRH	Right-Here-Right-Now
HSPS	Highly Sensitive Person Scale
SWLS	Satisfaction With Life Scale
RF	Random Forest
JS	Jensen-Shannon
RBF	Radial Basis Function
B5P	Big Five Personality

## 1. Introduction

In the built environment, indoor thermal comfort conditions may influence health [1, 2], office work performance [3, 4, 5, 6], learning performance [7, 8, 9], well-being [10, 11], and the overall satisfaction of occupants [12, 6, 13]. An analysis done by Graham et al. (2021) based on approximately 90,000 occupant satisfaction survey responses found that roughly 40% of them are dissatisfied with their thermal environment [14].

The Predicted Mean Vote (PMV) [15] thermal comfort index is widely used in industry and research. The PMV model aims to predict the average thermal sensation of a group of people sharing the same thermal environment. Although Fanger [15] was aware that people have different thermal sensations, he developed the PMV model since many buildings are not designed to provide personalized cooling and heating to individual occupants. Hence, the PMV aims to find the optimal indoor conditions to ensure that most occupants are comfortable. Another popular thermal comfort model included in the ASHRAE55 is the Adaptive model [16]. This model is based on the prevailing mean outside air temperature and only applies to naturally conditioned buildings. Both models consider environmental and personal conditions affecting the occupant rather than specific temperature set-points as fixed comfort thresholds, but they average the individual occupants’ responses. A study on the ASHRAE Global Thermal Comfort Database II [17] provides evidence of the low accuracy of the PMV model, i.e., 33% [18] in predicting the thermal sensation of individual occupants. This, combined with the proliferation of Internet-of-Things (IoT) sensors and wearable devices, has pushed researchers to look into

alternative solutions to predict how occupants perceive their thermal environment. Data-driven approaches leverage direct feedback from occupants paired with environmental and physiological data to develop personalized thermal comfort models [19]. These models can then be used to optimize the operation of Heating, Ventilation, and Air-conditioning (HVAC) systems or to quantify the thermal environment’s quality in buildings. With Personal Comfort Model [19] researchers can use environmental, physiological, and behavioral data to predict thermal sensation or preference. To investigate relationships between the variables mentioned above, field experiments offer a more realistic context of a person’s experiences compared to most climate chamber studies [20]. Collecting field data poses several challenges since people must complete surveys while performing day-to-day tasks. This may disrupt their activities, and accurate monitoring and logging of environmental variables with a high spatial and temporal resolution are expensive and nontrivial. Moreover, even when there is enough available data to develop a PCM for all occupants, it is still problematic from the control strategy perspective to ensure that all occupants will find the thermal environment comfortable [21]. Another common approach is to develop a general-purpose model to make the most of the available data. This model is trained on the whole population of occupants so that the available data for training is maximized at the expense of personalization. In open shared spaces where the HVAC system cannot be designed to tailor to individual needs, one option is to control environmental parameters to maximize the number of comfortable occupants and, at the same time, provide Personal Comfort Systems (PCS). Each occupant can use the PCS (e.g., heated and cooled chairs, desk fans, and foot warmers) to adjust the environment based on their needs and preferences.

In this paper, we seek to investigate whether we can leverage historical information gathered from building occupants to predict the thermal comfort preferences of a new person from which little individual data is available. By individual data, we refer to longitudinal subjective responses to Right-Here-Right-Now (RHRN) surveys about thermal preference. In particular, our research hypothesis is that the thermal comfort preferences of building occupants can be estimated by analyzing their psychological and behavioral traits and comparing them to other occupants from which subjective feedback was previously obtained. This way, we leverage the personalization aspect of PCM with less data.

In previous studies, the principal value of subjective responses alone was used to group occupants (i.e., the proportion of subjective feedback that is *prefer cooler*, *no change*, or *prefer warmer*) [22] whereas in [23] a *comfort profile* was first determined based on some historical data. The new occupants were compared against the available group profiles. In both works, once similarities were found between occupants, the group’s data was used for the prediction model for new occupants. Environmental measurements (e.g., air temperature and relative humidity) coupled with physiological measurements (e.g., height and weight) were used as features in the model [24]. However, in both approaches, some data is still required from the new occupant to correctly *assign* it to a respective group.

In this work we introduce Cohort Comfort Models (CCMs). This framework uses similarities between occupants to predict how new occupants with little or no historical data would perceive their environment. More details are explained in Section 2. With this approach we tried to answer the following questions:

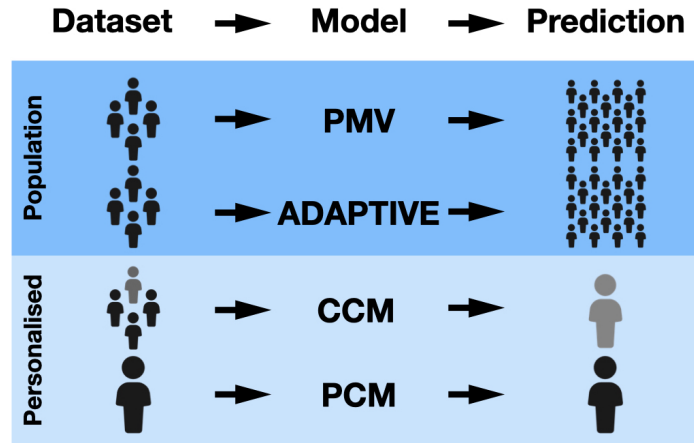
1. Do CCM increase the prediction performance for new occupants compared to a data-driven general-purpose model, i.e., all data are merged to develop a single comfort model?
2. Can we pinpoint which occupants benefited from using CCM, and if so, which features mainly contributed to improving the prediction accuracy?

## 2. Cohort Comfort Models

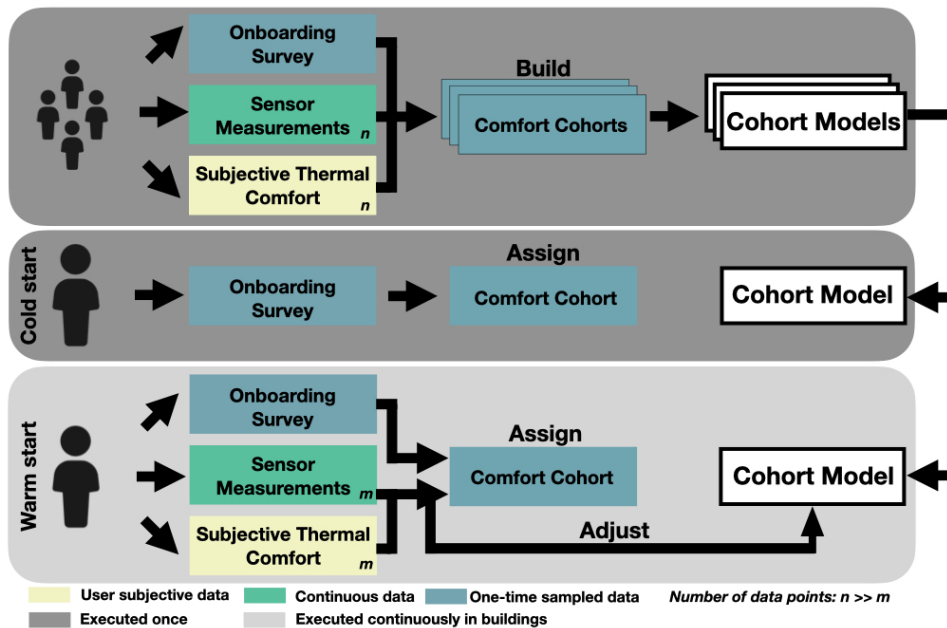
### 2.1. Definition

Cohort Comfort Models (CCMs) is a new framework to predict the thermal preference responses of a new person. CCM builds on Personal Comfort Models (PCMs) and leverages historical data collected from a sample population, who have some underlying preference similarity with the new person. By doing this, CCM leverage the power of both aggregate models and PCMs. The key characteristic of a CCMs is that it is trained on a group of occupants who share features and subjective preferences with the new person we are trying to predict. The comparison of data-driven thermal comfort modeling alternatives against more established ones like Predicted Mean Vote (PMV) and the adaptive model is shown in Figure 1a. An overview of the proposed framework is shown in Figure 1b.

One of the main contributions of this work is to determine ways in which characteristics of occupants, relevant to thermal comfort, can be compared. Firstly, a longitudinal data collection experiment of various buildings occupants is required, since a high number of participants allows the framework to find both occupants with similar and different characteristics. As for the required number of labeled data points per occupant ( $n$  in Figure 1b), previous relevant works suggests that 50 [25], 60 [19], 90 [26], and 200 [27] data points per occupant are needed for acceptable accuracy and stable prediction. However, the current landscape on thermal comfort longitudinal field data collection experiments is very limited in terms of datasets [28] and only a handful of them contain a high number of data points ( $n$ ) per participant. It should also be noted that the number of data points needed for each participant varies as a function of the range of environmental conditions a participant may experience. While a high  $n$  would be ideal, our proposed framework is not restricted to a threshold of available data. Our framework then utilizes the different data streams from the dataset collected (top gray segment, Figure 1b): onboarding survey, sensor data, Right-Here-Right-Now (RHRH). *Onboarding survey* refers to a one-time background survey that each occupant fills out only once before the data collection experiment starts. This survey must gather personal data (e.g., sex, height, weight) and preferably should also include psychological tests, such as the Big Five personality test [29]. This test is also used for occupant similarity, in the healthcare monitoring domain, for well-being label prediction [30]. *Sensor Measurements* encompasses all environmental and physiological data collected when the participant



(a) Diagram of thermal comfort modeling alternatives. The darker blue region highlights thermal comfort models at the population level (i.e., Predicted Mean Vote (PMV) and adaptive). The light blue area comprises personalized thermal comfort models trained on the person itself (Personal Comfort Model (PCM)) or on a group of similar people (Cohort Comfort Model (CCM)).



(b) Cohort comfort modeling framework, built on Personal Comfort Models (PCMs), leverages historical data collected from a sample population, who have some underlying preference similarity, to predict thermal preference responses of new occupants. Firstly the comfort cohorts are created as shown in the top segment following a one-time data collection effort. Comfort cohorts are determined using data from the onboarding surveys, sensor measurements, and subjective thermal comfort data. The middle dark gray segment shows a new occupant from which only the onboarding survey is acquired and is used to assign the occupant to one of the previously created cohorts; this is also known as “cold start”. Finally, the bottom segment shows another new occupant from which  $m$  data points are collected in order to be assigned to an existing cohort with the possibility of using these  $m$  data points to fine-tune the existing cohort model; this is also known as “warm start” when  $n \gg m$ . For both cold and warm start, the model corresponding to the assigned cohort is used to predict the occupant’s thermal comfort with the addition that the latter can use the already collected information from the occupant to fine-tune the cohort model.

Figure 1: Comparison of current thermal comfort modeling alternatives (1a) and overview of our work *Cohort Comfort Model (CCM)* (1b).

completed the RHRH. *Subjective Thermal Comfort* comprises the questions asked in the RHRH survey. These three main data streams are used to define the CCM. These CCMs are groups or clusters capable of reflecting preference patterns extrapolated from the relationship between *Subjective Thermal Comfort* and *Sensor Measurements*, which are longitudinal time-series data we will refer to as *warm start*. Then, CCM obtained from *Subjective Thermal Comfort* and *onboarding Survey responses* will be referred to as *cold start*. The naming convention is the same used

in the field of Recommender Systems. A *cold start* approach means starting off from scratch with no previous “warm up”. In our case this means we start without asking the participant to complete any right-here-right-now surveys. This is a harder start. On the other hand, a *warm start* scenario assumes some initial initialization or “warm up” has started, thus making the subsequent efforts easier. In our case, this means we started with historical longitudinal data.

Once the cohorts are determined and existing occupants are

allocated to one, new occupants can also be assigned to a cohort. The cohort assignment is coupled with how the cohort was created and will vary depending on the data availability for the new occupant. When a new occupant has not been through a longitudinal data collection phase, i.e., only onboarding survey data is available, they can be assigned to a cohort using a “cold start” assignment procedure (middle dark gray section in Figure 1b). On the other hand, when a new occupant has been through a longitudinal data collection phase and a subset of RHRH data points is available (bottom light gray section in 1b), the assignment to a cohort can be done using a “warm start” approach. It should be noted, that in the latter case, the new occupant can still be assigned to its respective cohort without leveraging the collected labeled data points using a “cold start” assignment procedure with the onboarding survey responses. In addition, the availability of the RHRH data for this occupant makes it possible to fine-tune the respective cohort model for the new occupant if required. The details on *cold start* and *warm start* cohort creation and assignments are further explained in Section 2.3.

## 2.2. Current Related Work

### 2.2.1. Data-driven thermal comfort prediction

Data-driven thermal comfort models rely on a handful of the measured features and often outperform the PMV and adaptive comfort models [31], specifically when trying to predict the thermal preferences of individuals [19]. Recent literature employs machine learning models in order to contextualize environmental data [32, 33, 34, 35] as well as thermoregulation from human skin through video [36], and physiological data [27, 37, 38]. In addition, some work has also analyzed the transition time it takes for an occupant to change its thermal preferences in the same indoor environment [39]. At the individual level, [19] has shown that data-driven PCM perform better than conventional models, such as the PMV, on a cohort of 38 occupants. These PCMs are tailored to an occupant’s specific needs and can be considered as one of the best approaches to achieving high-performing models for a particular individual. As for the predicting variable, the 3-point Thermal Preference scale is a well-established scale used in thermal comfort research as the label or target variable for data-driven thermal comfort prediction modeling [19]. Aggregated data-driven models, meaning models that use data from a large group of people to predict another person’s thermal preference, face the challenge of multiple occupants having different thermal comfort preference from each other [40, 41, 12, 42].

### 2.2.2. Cold start prediction

The built environment research is moving towards using a small amount of data from new occupants in order to successfully predict their thermal comfort [23]. This scenario is a well-known one when dealing with personalization, particularly in the field of *Recommender Systems* and is known as “cold start prediction”. The cold start personalization problem refers to starting providing relevant information to the occupant, or a system, as fast and with as little effort as possible [43]. This is a common practice in smart agents (e.g., Cortana, Google Assistant, Siri) or services (e.g., Amazon, Netflix) where oc-

cupants can be clustered based on limited knowledge about them [44, 43]. Alternatively, information about the occupants can be obtained by analyzing historical data and inferring other relevant features [45, 46]. These two techniques can be used together to improve the model prediction performance.

In the context of health monitoring and well-being label (i.e., mood, health, stress) prediction, [47] provided a framework to quantify occupant similarity based on individual behavioral patterns and then cluster them in order to have groups from which group models can be used for new occupants. [30] used the Big Five personality trait survey as the one-time general preference survey to determine groups of occupants in a data-driven manner to then use group-level information for occupant personalization. Both works rely on physiological, environmental, and behavioral data paired up with mood, health, and well-being labels from more than 200 different participants. However, to the best of our knowledge, this is the first work that explores these techniques in the context of thermal comfort prediction.

## 2.3. Cohort creation framework

### 2.3.1. Cold start

Personal characteristics captured during the onboarding one-time survey, include categorical, ordinal, and scalar data that can directly be used to group occupants. In addition, some standardized surveys allow for their questions scores to be aggregated, i.e., Highly Sensitive Person Scale (HSPS) [48] and the Satisfaction With Life Scale (SWLS) [49].

The simple approach to creating cohorts is to use individual one-time survey responses. Cohorts are created in a straightforward manner by dividing occupants based on a simple statistic, e.g., median, or creating as many cohorts as unique response values. Multiple one-time survey responses can also be combined in a data-driven setup where each survey response is a feature and a clustering algorithm can automatically group occupants into cohorts based on the discovered patterns from the questions responses. We determined the optimal number of cohorts,  $k$ , by using the clustering metric known as Silhouette score [50] which ranges from 1 (best score) to -1 (worst score). The appropriate  $k$  is determined by the highest mean value of the Silhouette score.

Once the cohorts exist, new occupants can be directly assigned to one. If the cohorts were created in a data-driven manner, the trained clustering algorithm can then predict the corresponding cohort using the required occupant’s one-time survey responses. On the other hand, new occupants can be assigned to cohorts created based on specific values by simply matching its respective one-time survey response: e.g., if two cohorts are created based on sex, a new person is either assigned to the Female or Male cohort.

### 2.3.2. Warm start

**Responses distribution similarity:** Previous work has looked into grouping occupants solely based on the trend of their thermal comfort responses [22], for the case of thermal preference, occupants who predominantly have more *prefer cooler* responses are grouped together, and so on. Figure 3b illustrates the comparison of two occupants’ thermal preference responses. Unlike work done in [22], where grouping was done based on the ma-

jority response, the probability distribution of responses is first calculated for each occupant, on this example, the responses are thermal preference votes where the values are *prefer warmer*: -1, *no change*: 0, *prefer cooler*: 1. Then, the occupant’s response distributions are compared using the Jensen-Shannon (JS) divergence, bounded in [0, 1] and symmetric, which calculates the similarity between two probability distributions. A lower distance value between occupants means the occupants are more similar.

**Cross-model performance:** We propose to group people into cohorts based on the relationship between their responses in the RHRH survey and the different available environmental and physiological sensor data. [47] suggest that these relationships can be captured by evaluating the model of one occupant, trained with its own data, using labeled sensor data from another participant, this is referred as cross-model performance evaluation. In the context of thermal comfort, a PCM is a good representation of an occupant’s individual comfort requirements based on the data of its environment [31]. Hence, using PCMs for cross-model performance evaluation is an adequate similarity metric. If the PCM of *occupant<sub>A</sub>* has a high performance on data from *occupant<sub>B</sub>*, it suggests that both occupants have similarities in how they perceive their thermal environment. Figure 3a shows an example of evaluating the cross-model performance of two occupants. The data used for the PCM on each occupant comprise only sensor measurements since any personal characteristic data collected via a one-time preference survey would have a repeating value for each data point. For a consistent comparison, each PCM must be trained on the respective occupant’s data, and any hyper-parameter must be fine-tuned using the same evaluation metric. Based on the current literature on PCMs where the target variable comprises 3 ordinal values (i.e., multi-class), one of the most widely used models is Random Forests (RFs) [22, 19, 24] and the evaluation metric is F1-micro score [22, 24] which can be seen as accuracy for the multi-class classification scenario. F1-micro scores range from 0 to 1, with 1 being a better prediction accuracy which in turn indicates two occupants are more similar.

Finally, both metrics of similarity between occupants, *responses distribution similarity* and *cross-model performance*, can be rearranged into what is called a similarity or *affinity* matrix. An affinity matrix is a squared symmetric matrix where each element to compare, in this case occupants, is both the rows and columns indices. Each value of the matrix represents the similarity of its row and column; these values are bounded in [0, 1] where 1 means identical elements. Affinity matrices have shown to be a useful way of comparing occupants when the similarity calculations can be transformed into the [0, 1] range [47]. Both similarity metrics already match the desired range but the responses distributions metric have an inverse interpretation where a value of 0 means identical occupants. Thus, these values are normalized using the Radial Basis Function (RBF) kernel (Equation 1) such that now a value of 1 translates to identical occupants.

$$\text{RBF Kernel}(Z_i, c) = e^{-\frac{(z_{i,j}-c)^2}{2\mu^2}} \quad (1)$$

Where  $Z_i$  is the similarity metric (Jensen-Shannon (JS) divergence in this scenario),  $\mu$  is the standard deviation of  $Z_i$ ,  $z_{i,j} \in Z_i$  is the JS divergence between occupants  $i$  and  $j$ , and  $c$  is the normalization center. In order to consider both similarity metrics, both metrics can be added together with a weight coefficient  $\alpha$  and  $\beta$  (such that  $\alpha + \beta = 1$ ), respectively, to indicate their respective contribution in the final affinity matrix. At this point, Spectral clustering [51] is applied to the affinity matrix and, similar to Section 2.3.1, the adequate number of cohorts,  $k$ , is determined by the highest mean value of the Silhouette score. Figure 3 illustrates both similarity metrics calculation, their rearrangement into an affinity matrix, and cohort calculation.

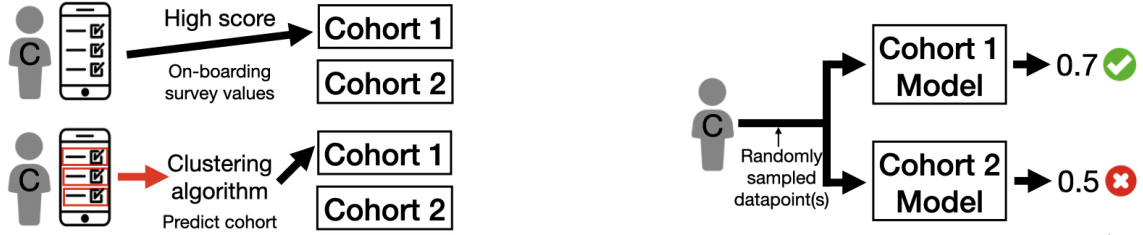
New occupants can be assigned to a cohort only after they complete some RHRH surveys. However, as depicted in Figure 1b, collecting the same number of labeled datapoints,  $n$ , for each new occupant as it was done with the existing occupants can be impractical and cost prohibitive. Hence, we propose to use a “warm start” approach that leverages a much smaller number of data points  $m$  from the new participant, such that  $m \ll n$ . The  $m$  datapoints from the new occupant are used to evaluate the performance of each cohort we previously created. The performance values are then compared and the occupant is assigned to the cohort with the highest performance score overall. The metric used in this last step must be the same as the metric used to determine the initial cross-model performance evaluation, which for the case of a multi-class thermal comfort variable is the F1-micro score. Figure 2b illustrates this warm start cohort assignment.

#### 2.4. Evaluation

Experiments and modeling parameters are detailed in Table Appendix A.1. We used a 0.8 train-test split, participant-wise, on each dataset. Hence, 80% randomly selected occupants and their data are used to build the cohorts following the approaches mentioned in the previous section. The remaining 20% of the occupants are used as test and assigned to a newly created cohort. This entire process was repeated 100 times for each cohort approach on each dataset used. Then, once a test occupant is assigned to a cohort, the cohort thermal preference model is evaluated on the new occupant’s labeled data.

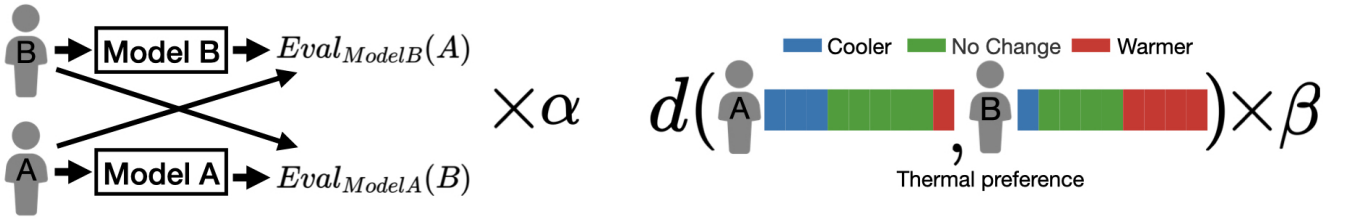
Personal Comfort Models (PCMs) models were computed for cross-model performance and served as an upper-bound baseline. PCMs are trained with RF as the classification model following a 5-fold cross-validation. We also used the entirety of the available data from the train participants to create an aggregate model where everyone belongs to one big cohort from which a *general-purpose* model can be trained on. This model will serve as a lower bound baseline. All processes are repeated 100 times in order to minimize biases.

Additionally, the prediction results of PMV at the personal level are included. The Python package *pythermalcomfort* [52] is used to compute the PMV as per the calculations methods in ISO 7730 [53], and the package *Scikit-learn: Machine Learning in Python* [54] is used for all the remaining data-driven calculations. Similar to how it was done in [19], we used the longitudinal field data (i.e., air temperature, humidity, self-reported clothing insulation) and the static values (i.e., air velocity=0.1



(a) **Cold start:** **Top** The occupant is directly assigned to the cohort with the matching value (b) **Warm start:** Data points collected from the new participant are evaluated on each existing of the onboarding survey used for cohort creation. **Bottom** The clustering algorithm used *Cohort model*. The participant is then assigned to the cohort on which it had the best-averaged for cohort creation is used on the participant’s onboarding survey responses to predict the performance based on F1-micro score.

Figure 2: Participant assignment to the existing cohorts depending if the cohorts were cold start (2a) or warm start (2b). Only 2 cohorts are shown for illustration purposes.



(a) Cross-model performance evaluation calculated with F1-micro score, bounded in  $[0, 1]$ , using each occupants’ PCM with another occupant’s data. The higher the F1-micro score the more similar the two occupants are. (b) Responses distribution similarity based on Jensen-Shannon divergence, bounded in  $[0, 1]$ . After applying an Radial Basis Function (RBF), the higher the value the more similar the two occupants are.

Figure 3: **Warm start** cohort creation approach. 3a calculates the cross-model performance and 3b calculates the distance between occupants’ responses distribution. Each approach is weighted before their summation into an affinity matrix with  $\alpha$  and  $\beta$  respectively, where  $\alpha + \beta = 1$ . Spectral clustering is then run on the affinity matrix on a multiple number of cohorts (e.g.,  $k = [2, 10]$ ). The best  $k$  is the one with the highest Silhouette Score.

m/s, metabolic rate=1.1met) for the PMV calculation unless the values are present in the dataset. To compare the results on the same 3-point scale, the PMV is converted into thermal preference classes based on the following assumptions:  $|PMV| \leq 1.5$  is “no change”;  $PMV > 1.5$  is “prefer cooler”; and  $PMV < -1.5$  is “prefer warmer”. Nevertheless, we acknowledge that by making these many assumptions and simplifications on the values themselves, the accuracy of the PMV calculation will be low. We share our code base for the reproducibility of our analysis on a public GitHub repository: <https://github.com/buds-lab/ccm>.

### 2.5. Datasets

In order to assess the proposed framework, two publicly available longitudinal thermal comfort datasets were used to validate our methodology. One dataset, named Dorn [55], was collected in Singapore where 20 occupants were asked to complete approximately 1000 RHRH surveys over a period of 6 months. Participants were asked to wear a Fitbit smartwatch with the Cozie [56]<sup>1</sup> application installed. This application has been used to collect RHRH surveys [22] and is publicly available for both Android<sup>2</sup> and Apple<sup>3</sup> platforms. The second dataset we used was collected from 37 participants over a period of 12 weeks in an office building located in Redwood City, California, US [57]. Participants were asked to complete three RHRH online surveys

Dataset	# Occupants (Sex)	Age range	#Responses per participant	Duration
Dorn [55]	20 (10 M, 10 F)	20 to 55	872(min), 1332(max)	6 months
SMC [57]	37 (17 M, 20 F)	25 to 37	33(min), 218(max)	12 weeks

Table 1: Overview of the two datasets used in this work.

per day while using a Personal Comfort System (PCS) in the form of an instrumented chair. In this work we will be referring to the form dataset as the “Dorn” and the latter as the “SMC” dataset. Both surveys comprised the following overlapping questions: thermal preferences (using a 3-point scale), clothing, and activity. Environmental measurements (e.g., dry-bulb air temperature, relative humidity) were taken using data loggers installed in near proximity (i.e., within a 5-m radius) from where the participant completed the survey. Additionally, [55] used Fitbit Versa smartwatches to measure and log physiological data and iButtons for near-body temperature and humidity. Table 1 shows a brief overview of both datasets, and Table ?? shows a summary of the sensors used and their accuracy. More details regarding the recruitment criteria and methodology details can be found in the respective publications.

Personal information about the participants (e.g., sex, height, weight) were obtained asking them to complete an on-boarding

<sup>1</sup><https://cozie.app/>

<sup>2</sup><https://github.com/cozie-app/cozie>

<sup>3</sup><https://github.com/cozie-app/cozie-apple>

survey when joining the experiment. Additionally, the Dorn dataset included three standardized surveys: Highly Sensitive Person Scale (HSPS) [48], Satisfaction With Life Scale (SWLS) [49], and a short version of the Big Five Personality (B5P) test known as Ten-Item Personality Inventory [29]. These surveys were used to investigate potential relationships between survey responses from occupants and their thermal preferences.

## 2.6. Data pre-processing

We time-aligned sensor measurements with the RHRH responses. Thermal preference response provided by the occupants are used as the ground truth label for each respective dataset. Feature selection scrutiny was done based on the latest efforts in data-driven personal thermal comfort prediction [58, 22, 19, 27, 59] and based on what the original dataset papers suggest [57, 55]. A summary of the features used in this paper for model training is shown in Table 2 for the present study, and Table 2a and Table 2b for the Dorn [55] and SMC [57] dataset respectively.

To have comparable results between occupants we decided to use a fixed number of responses per occupant on each dataset. While this approach does not address the inherent class balancing issue in thermal preference datasets, i.e., having a disproportionate number of data points for each thermal preference type [60], we decided to prioritize the quantity of available data points per participants. The Dorn dataset contains a minimum of 872 responses per participant (Table 1); however, after removing surveys completed outdoors, in non-transition periods and while participants were exercising, the minimum number of responses per occupant was 231. [19] found that when the dataset comes from an environment that all occupants shared during the experiment, the required number of responses for a stable thermal preference prediction is at least 60. These results were obtained on the SMC [57] dataset, hence, we opted to also use the first 60 responses from each occupant as it also covers most of its participants. This threshold of minimum responses resulted in 35 participants (18 females, 16 males). Figure Appendix A.1 in the Appendix shows the distribution of the filtered thermal preference responses for each participant on both datasets.

## 2.7. Cohort creation

Following the framework presented in Section 2.3, we created two different sets of cohorts one for each datasets. Table 3 shows the *cold start* cohorts (upper row) and *warm start* cohorts (lower row) that are constructed based on the data available in each dataset. As mentioned in Section 2.3.1 the *cold start* cohorts are created using only one-time survey responses. In the Dorn dataset, the onboarding surveys were also used to create other sets of cohorts. The three surveys (HSPS, SWLS, and B5P) were used together (Surveys in Table 3) where each survey score is treated as a feature and the number of cohorts was determined in a data-driven manner via Spectral clustering as detailed in Section 2.3.1. Moreover, each survey was also used individually. The HSPS and SWLS surveys, named *Sensitive* and *Life Satisfaction* respectively in Table 3, have an aggregated numerical score which was used to create two cohorts based on the *median* value of the occupants' scores. All participants were

tested but only participants with extreme values on these survey responses, with a low and high aggregated score, are shown since their results are higher. Occupants with an aggregated score between the 25<sup>th</sup> and the 75<sup>th</sup> interquartile range were filtered out, which meant only 50% of the total occupants in the Dorn dataset were used. Since the responses of the B5P cannot be aggregated into one final score, the number of cohorts based on this survey was determined in a data-driven manner using each personality score as a feature. This cohort approach is named *Personality* in Table 3.

Finally, *warm start* cohorts are created following the two approaches mentioned in Section 2.3.2, using both the responses distribution similarity and cross-model performance combined, and cross-model performance alone (*Dist-Cross* and *Cross* respectively in Table 3). Table Appendix A.1 list the different experimental and modeling parameters. The best performing model, in terms of average expected F1-micro score, is found using a grid search on the list of hyperparameters. In the Dorn dataset we determined that the optimal number of cohorts was 2 for all data-driven approaches. On the other hand, a value of  $k$  equal to 2 and 3 was chosen for the *Dist-Cross* and *Cross* approaches in the SMC dataset, respectively. More details on different numbers of cohorts  $k$  are shown in Figure Appendix A.2 in the Appendix.

In order to corroborate the usefulness of each cohort approach we also purposely assigned each occupant to the opposite cluster (if  $k = 2$  and were created by *cold start* approaches) or assigned the occupant to the worst performing cohort (*warm start* approaches). This will serve as an ablation analysis for the cohort approach and, by extension, the data used for its creation and assignment.

## 3. Results

### 3.1. Overall prediction performance

The performance results, in terms of F1-micro score, of all cohort approaches from Table 3 in new occupants from their respective datasets are summarized in Figure 4. Results sets with a tilde ("~") as a prefix denote those approaches where test users were purposely assigned to an incorrect cluster as detailed in Section 2.7.

As mentioned in Section 2.3.2, *Warm start* cohort approaches like *Dist-Cross* and *Cross* need that new participants complete a few RHRH surveys prior to assigning them to a given cohort. Different number of labeled data points, or RHRH, were tested, i.e., 1, 3, 5, and 7, and all of them fall within the baselines. Figure Appendix A.3 shows the results for the Dorn dataset, similar results were obtained with the SMC dataset. For the purpose of evaluating the minimum required number of labeled data points, all results in Figure 4 used one single, randomly selected, data point from each test occupant for assignment. Compared to using more labeled data points, the performance of using a single data point still lead to improved results, on average 10% compared to the general-purpose model.



Feature	Source
Air temperature Relative humidity	Fixed sensor (indoor)
Near-body temperature Heart rate	Wearable sensor
Clothing level	RHRN survey
Sex Height Weight	One-time survey

(a) Subset of features chosen for data-driven modeling in the Dorn dataset based on the original work [55]. Near-body (wrist level) temperature measurements are found to contain more information than skin temperature [27, 59].

Feature	Source
Dry-bulb air temperature Operative temperature Relative humidity Slope in air temperature	Fixed sensor (indoor)
Control location Control intensity Control frequency in the past x (1h, 4h, 1d, 1wk) Occupancy status Occupancy frequency in the past x (1h, 4h, 1d, 1wk) Ratio of control duration over occupancy duration in the past x (x=1 h, 4 h, 1 d, 1 wk)	PCS control behavior
Outdoor air temperature Sky cover Weighted mean monthly temperature Precipitation	Outdoor Environment
Clothing level Hour of the day Day of the week	RHRN survey
Sex Height Weight	One-time survey

(b) Subset of features recommended for data-driven modeling [19] in the SMC dataset [57].

Table 2: Features chosen for data-driven modeling: 2a Dorn [55], 2b SMC [57] dataset. Features obtained through the one-time survey are dropped for PCMs due to their constant value within each participant.

Type	Cohorts approaches	Data used
Cold start	Sex†	Self-reported value
	Surveys	All onboarding surveys (HSPS, SWLS, B5P)
	Sensitive	HSPS scores
	Life Satisfaction	SWLS scores
	Personality	B5P scores
	Agreeableness	Best performing trait from B5P
Warm start	Responses <b>Distribution</b> and <b>Cross-model performance (Dist-Cross)†</b>	Thermal preference responses and occupants' PCM
	<b>Cross-model performance (Cross)†</b>	Occupants' PCM

Table 3: Cohort approaches chosen. The upper rows are cold start cohorts and the lower rows are warm start. † were cohorts approaches used on both datasets.

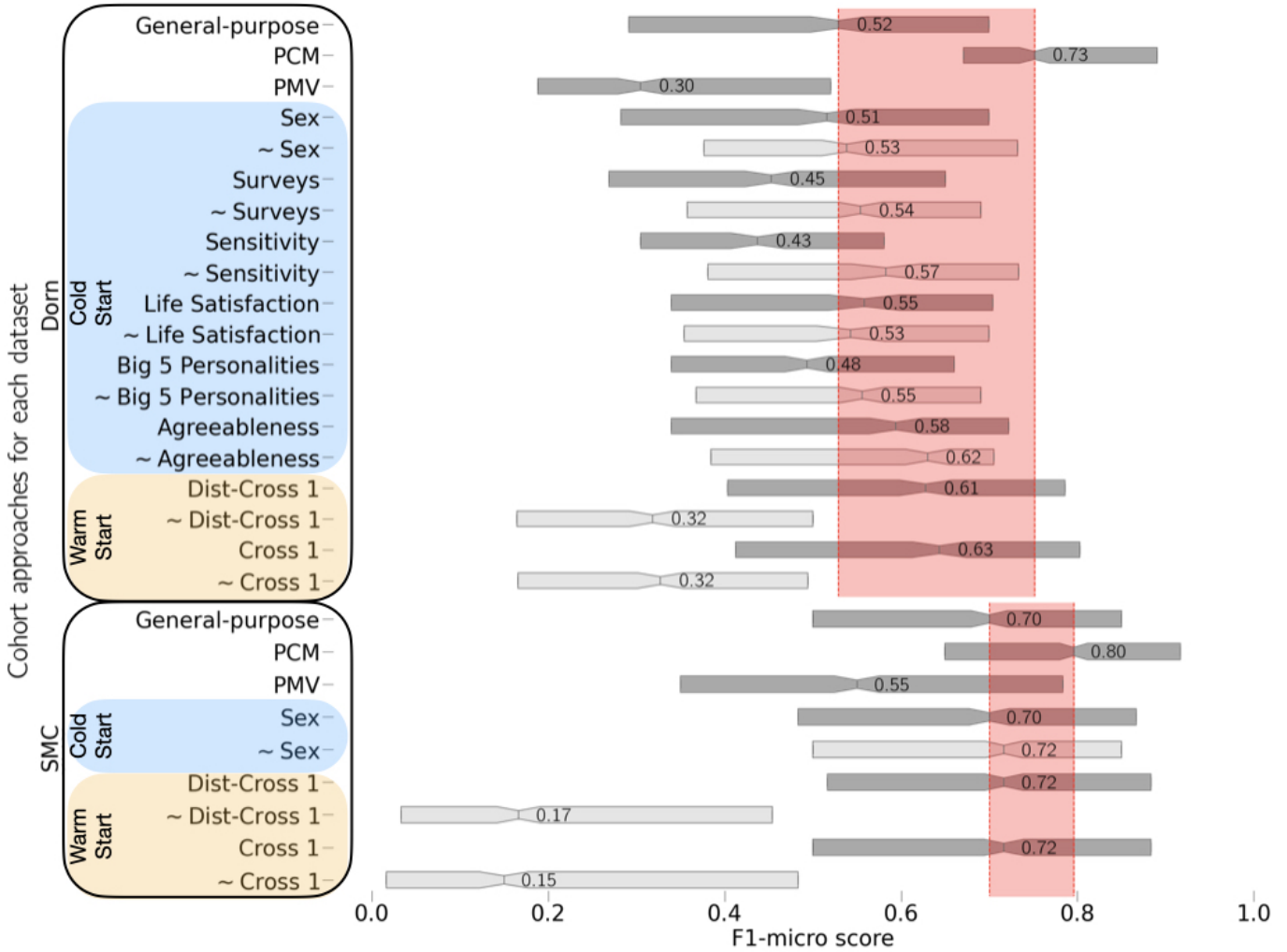


Figure 4: Performance results in F1-micro score of 100 iterations for each cohort approach on each dataset Dorn and SMC (y-axis) and their baselines after correct assignment (dark gray) and incorrect assignment (light gray). Approaches with occupants incorrectly assigned to cohorts have the “tilde” symbol (~) before their name. *Cold start* approaches are highlighted by a light-blue region and *Warm start* approaches by a yellow region. The general-purpose model and PCM median value are highlighted with red dashed lines on each of their boxplot and filled with red. Additionally, PMV results, following the calculation in Section 2.4 are shown. *Warm start* approaches (Dist-Cross and Cross) have used only one randomly selected data-point. Dorn dataset: *General-purpose* boxplot contains 20 values; *Sensitive* and *Life Satisfaction* contain 200 values (2 test occupants at a time  $\times$  100 iterations) and all remaining box plots have 400 values (4 test occupants at a time  $\times$  100 iterations). SMC dataset: *PCM* boxplot, contains 35 values and all remaining boxplots have 700 values (7 test uses at a time  $\times$  100 iterations). The cohort approaches are taken from Table 3.

### 3.1.1. Under performing cohorts

Cohort approaches such as *Sex*, *Surveys*, *Sensitivity*, and *Big 5 Personalities* have a median performance below the *General-purpose* in the Dorn dataset (upper plot in Figure 4). Although “sex” has been found to have some influence in thermal preference, dividing occupants based on it does not provide significant changes, on average, on thermal preference prediction when compared to a *general-purpose* model on both datasets. This is supported by the similar performance of occupants being assigned to their same sex cohort and opposite sex cohorts (*Sex* and *~Sex* in Figure 4). The lower performance of two of the *Surveys* approach components (*Sensitivity* and *Big 5 Personalities*, when used individually, might be the reason for this approach’s lower than *General-purpose*

performance. We hypothesize the HSPS survey, used for the *Sensitivity* was not able to capture enough information due to the geographical location and participants’ background of the data collection experiment. Participants from Singapore (Dorn dataset) may have provided a non-accurate response to this survey since they mostly experience a hot climate, compared to a full range of cold and hot climates, which could translate into the low performances of this cohort approach. Each personality trait from the B5P survey was also tested individually and *Agreeableness* performed above the *General-purpose* value (Figure Appendix A.3). This suggests that looking at all personality traits together (*Big 5 Personality*) is not as beneficial as looking into each personality trait individually. Moreover, it is possible that the small sample size of participants

on the Dorn dataset provided little variability on these survey responses or that the three onboarding surveys chosen, HSPS, SWLS, and B5P, are simply not relevant for thermal preference. Overall, the efficacy and usability of these warm start approaches remains an open question.

### 3.1.2. Above-baseline performing cohorts

The Life Satisfaction is the only survey that by itself has an above General-purpose performance. This indicates that treating each survey individually instead of combining them (Surveys) is a better approach for cohort prediction performance. In fact, only the HSPS and B5P surveys lead to a cohort approach with favorable performances above the general-purpose models. Since both of these cohort approaches are related to satisfaction and optimism, we hypothesize occupants within the resulting cohorts are less prone to preference changes due to small variations, meaning they have a wider acceptability range.

On the other hand, cohort approaches Dist-Cross and Cross achieved the highest median performance on both datasets (top and bottom plots in Figure 4 for the Dorn and SMC dataset, respectively). While one single labeled data point used for cohort assignment may be prone to bias, the consistent F1-micro scores after 100 iterations still positioned both cohort approaches as the top-performing ones, with a much clearer difference in the Dorn dataset than in the SMC dataset. Moreover, when each of this cohort approach is compared to their respective worst assigned cohort, the median performance is reduced by almost half for Dist-Cross (from 0.61 to 0.32) and Cross (from 0.63 to 0.32) in the Dorn dataset; and the performance is reduced 4 times for Dist-Cross (from 0.72 to 0.17) and Cross (from 0.72 to 0.15) in the SMC dataset. The median performance for Life Satisfaction is also reduced when occupants are incorrectly assigned (from 0.56 to 0.51). The reduced number of occupants used for this approach, 10 instead of 20, might be one of the reasons why the difference in performance is less pronounced when compared to Dist-Cross and Cross. Nevertheless, the Agreeableness cohort approach exhibits different behavior. Among all 20 occupants in the Dorn dataset, the lowest score obtained was 4 out of 7 which indicates the occupants rank highly on this personality trait. As previously mentioned, this personality trait might influence the acceptable thermal preference range of an occupant and, with only two cohorts created, both created cohorts could improve an occupant's thermal preference prediction. These results suggest that when the cohort modeling framework is followed correctly, it leads to an overall increase in median F1-micro performance. This increase is shown to be higher with *warm start* approaches (Dist-Cross and Cross) than with *cold start* approaches (Life Satisfaction and Agreeableness).

### 3.2. Occupant-specific improvement

Figure 5 shows a scatter plot of the average percentage changes in occupant-specific F1-micro score of the *above-baseline performing* cohort approaches mentioned in the previous subsection for *warm start* and *cold start* types of cohorts on both datasets.

For each occupant, a positive average percentage change in F1-micro score is desirable since it indicates an increase in thermal preference prediction performance when using a cohort model instead of a general-purpose model. Conversely, a negative value indicates the cohort model fails to provide useful personalization, and a generic model, which dismisses occupants' similarity entirely, performs better on average since it was trained with more data. Overall, in the Dorn dataset around half of the population of occupants and one-third of the population of occupants in the SMC dataset saw a boost in prediction accuracy, in terms of F1-micro score, under the cohort modeling framework in their respective cohort approach. For the remaining occupants, there is little change in performance. In Figure 4 the cohort Life Satisfaction scored marginally better than the General-purpose (0.55 to 0.52) and the Agreeableness approached score a slightly higher median performance (0.58). This is highlighted in Figure 5 Dorn:Life Satisfaction where 40% of occupants saw a performance boost of less than 1%, whereas in Dorn:Agreeableness 55% of occupants saw a performance boost of around 2%.

Dist-Cross and Cross approaches show a performance boost of up to 36% and 46%, and an average of 5% and 8%, on the Dorn and SMC dataset respectively. Overall, the number of occupants that benefited from these approaches are more than half (45% and 60%) and one-third (35% and 32%), respectively on each dataset. The number of occupants that are better-off from using Dist-Cross or Cross are different on each dataset. The Dist-Cross approach benefits more occupants than the Cross approach on the SMC dataset, where the data was collected in an office building across many multiple shared spaces on two different floors. Under these circumstances, personal thermal preferences captured by the feedback distribution (the Dist component of Dist-Cross) contribute to more occupants being better-off by Dist-Cross compared to not using this information (Cross approach). On the other hand, in the Dorn dataset all occupants provided their thermal preference response in their own work environments with no guarantees of similar conditions and exposures amongst all of the spaces. Occupants' feedback distribution (the Dist component of Dist-Cross) was not able to capture enough preference characteristics that could be quantified and compared among the occupants, resulting in more occupants being better-off with the Cross approach alone. Nevertheless, it is also important to highlight that some occupants are actually worse-off regardless of the cohort approach used, and their average percentage performance is reduced. The entire distribution of percentage changes across all 100 iterations for these approaches and for each occupant can be found in Figure Appendix A.4 and Appendix A.5 for the Dorn and SMC dataset respectively.

Finally, metadata about the participants who are better-off and worse-off is summarized in Table Appendix A.2. Overall, there are no major differences between the sex and mean and standard deviation of height and weight among the occupants in both groups, in most cohort approaches. The cohort approach of Life Satisfaction on the Dorn dataset and Cross on the SMC dataset show a slight majority of Female and Male better-off occupants, respectively. Nevertheless, due to the small

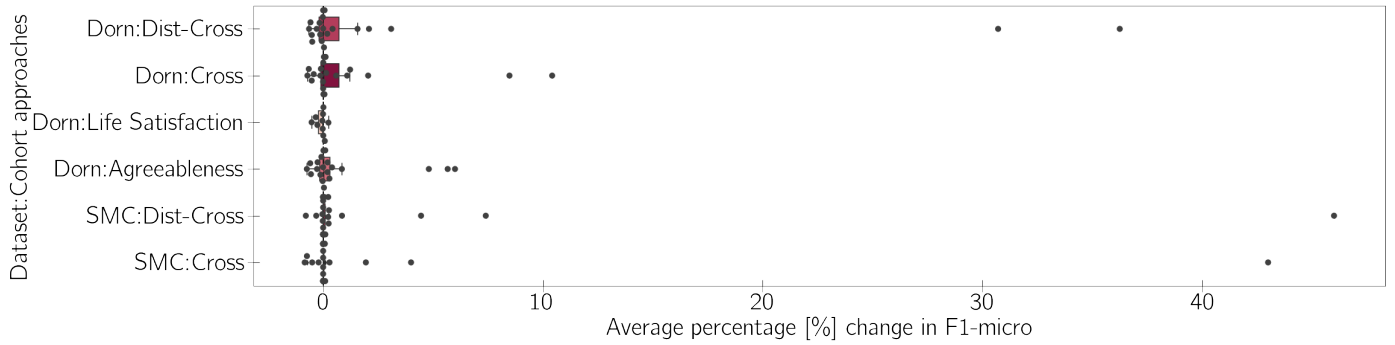


Figure 5: Scatter plot of occupant-specific percentage changes in F1-micro score from the general-purpose performance (`General-purpose`) to the respective cohort approach performance. Each dot represents the average value across all 100 iterations for one occupant. A positive value means the occupant benefited from the cohort approach and a negative one means they are worse-off. For the cohort approaches in their respective datasets (y-axis) 45%, 60%, 40%, 55%, 35% and 32% of the occupants saw an increase in their thermal preference prediction performance by relying on their peers’ data, respectively, although the actual benefit is small.

sample size of occupants on both datasets it is not possible to generalize these findings.

#### 4. Discussion

We showed that CCM can improve the thermal preference prediction of a new occupant by using other occupants’ data who are grouped based on preference similarity. We determined that: i) *Warm start* cohort approaches outperform *cold start* approaches at the cost of requiring labeled data points from each new occupant; ii) while around half and one-third of occupants, in each Dorn and SMC dataset respectively, benefited from being assigned to a cohort via warm start and using its cohort model, the remaining occupants saw a very little variation in their performance for both being better-off or worse-off; and iii) the applicability of cohorts is dependent on the availability of data from the occupants as well as the number of occupants willing to provide it, as we found ourselves able to investigate more cohort approaches on the dataset with more sensed modalities and more one-time onboarding surveys, like in the Dorn dataset.

##### 4.1. Warm vs cold start

*Warm start* cohorts are bottom-up approaches that focus on the granular data with a direct connection with thermal preference, i.e., relying on feedback distribution and cross-model performance, whereas *cold start* cohorts are a top-bottom approach that relies on the set of questions used to showcase the separability of personal thermal preferences. It can be argued that the latter approaches were expected to underperform since by their definition, information regarding the actual thermal preference of occupants, e.g., the labeled data points, was not used to create the cohorts. Nevertheless, recent literature in personalized healthcare monitoring [47, 30] highlights the usability of these sources of information as primary ways to leverage other occupants’ data. We found that the SWLS surveys, or Life Satisfaction, and the personality trait of Agreeableness from the BSP survey lead to cohorts where thermal preferences are shared, leading to slightly improved thermal preference prediction when compared to general-purpose models. One advantage of using this information is that allows a cold start from a new occupant.

However, the notion of “no historical data” is not a harsh

threshold since depending on the amount of data available, some work still considers a cold start when very little historical data is used [46], i.e., less than 10 data points [47]. While work done in [23] explores a Bayesian approach to cluster occupants based on historical measured data and then uses the cluster’s thermal comfort models with as little as 8 data points which also falls under the cold start definition used in [47], to the best of our knowledge our proposed framework is the first to not only look at measured data streams to find occupant similarity but includes the aforementioned mechanisms from other fields. Our framework allows for a full cold start prediction and a warm start prediction with as little as one single historical data point. Both types of cohort approaches have shown that while only some occupants benefited from this framework, the majority experience very little variability in performance, either better-off or worse-off. Nevertheless, this also shows that our framework allows for prediction performance just as good as the general-purpose models but with at least half less training data, for the case of only two cohorts.

##### 4.2. From “how much can you benefit” to “who can benefit”

The *warm start* cohorts at the occupant level produced an overall average increase of up to 8% (`Dorn:Dist-Cross`) and 5% (`SMC:Dist-Cross`) on the Dorn and SMC dataset respectively among the occupants who were indeed better-off. These numbers changed to 4% (`Dorn:Dist-Cross`) and 2% (`SMC:Dist-Cross`) when all participants are considered in the calculation. However, on both datasets there are occupants who could significantly take advantage of these approaches with performance increases as high as 45%, but also occupants who saw a decrease in their prediction performance. While this scenario is not uncommon when dealing with group-level occupant personalization [30], cold start approaches like our framework offers the possibility to identify these better-off and worse-off occupants even before the occupant starts occupying the building.

From a facility manager’s perspective, being able to identify if an occupant will be worse-off, or better-off, on a specific thermal zone with another group of occupants can help prioritize certain occupants, avoid complaints, and quantify the need for PCS.

The trend of hot-desking is not well received among occupants [61] and can show a negative effect on work engagement, job satisfaction, and fatigue [62]. Thus, a more suitable approach would require engaging with specific occupants and adapt their needs. On the other hand, although the results here are limited to two datasets, it is possible that some occupants are inherently more difficult to predict than others, and pinpointing back to these occupants alone is equally important. Be that as it may, cold start approaches that rely on one-time surveys, such as `co1d start` cohort approaches, raise data privacy concerns. Facility managers and building practitioners must take the necessary precautions when dealing with such data from its occupants.

#### 4.3. Limitations and future work

The evaluation of our proposed framework based on one-time onboarding surveys and sensor measurements as features has some limitations. First, the amount of occupants on each dataset is 10 times smaller than those used in related work in the health-care domain [30, 47] and in a related thermal comfort cluster prediction [23]. Additionally, one of the two datasets used (Dorn dataset) lacks seasonal variability because it was collected in Singapore (tropical climate), whereas the other dataset (SMC) which included both seasons and a more diverse group of occupants, lacks the one-time surveys required by the framework. While multiple datasets can be used together as one much bigger dataset, like the ASHRAE Thermal Comfort Database II [17], the methodology and available measurements should overlap, which is not the case for the recent field data collection experiments.

Second, the models and features used were not chosen after an exhaustive feature and model selection pipeline. One could argue that features extracted from the physiological time series measurements from each occupant, matched with their thermal preference responses, could contain useful information which could in turn be used for comparison across occupants and subsequent cohort creation. While extracting those temporal attributes is out of the scope of the present work, the modularity of the framework and its data-agnostic proposition makes the inclusion of these newly learned or discovered features, or any other new modalities, plausible.

Third, in our evaluations, we considered that once an occupant is assigned to a cohort, it remains its membership for the duration of the evaluation experiment. It is likely that influenced by external factors, an occupant might be better-off in terms of thermal preference prediction if it is assigned to a different cohort in certain settings. Future work can build on this cohort framework and insert the respective adaptive mechanisms that allow occupants to change cohorts as time goes by. Particularly, when coupled with the energy repercussions of using cohort models to predict thermal preference, an entire feedback loop from the occupant to the system could enable this online incremental learning.

Aware of the context-dependent nature of our framework and its components, we do not claim generalizable results. Ideally, a dataset with a large enough group of people where all occupants undertake the same one-time onboarding set surveys, with overlapping sensor measurements, could provide more generalizable

insights. Also, if multiple datasets from different contexts are used, features that represent the location, climate, and building type could be considered as part of the modeling portion of the framework. However, the presented framework is modular enough to easily adapt to new sets of features either as part of the cohort models themselves or as part of how the cohorts are created. We plan to investigate the results on more diverse datasets with occupants from different backgrounds and populations.

## 5. Conclusion

Cohort Comfort Models (CCMs) are a new approach to thermal comfort modeling that builds on Personal Comfort Models (PCMs). Cohort models leverage the previously collected data from other occupants to predict an occupant’s thermal preference based on similarities with other occupants. The advantage of this method in two different datasets shows that the proposed framework for cohort creation and occupant assignment lead to half and one-third of the occupants in each dataset experiencing an average increase of 5% and 8%, respectively, in their thermal comfort prediction performance with as little as a single labeled data point required from the new occupant when compared to general-purpose data-driven models.

Unlike related literature focused on occupant segmentation for better thermal preference prediction, our proposed framework offers the ability to identify the occupants who will be better-off from said cohort approach, and those who might not, based on the occupant’s background information or, at most, a single labeled data point from the occupant at the building premises. We provided the framework in a data and site-agnostic manner and described its different implementations depending on what information is available with the scalable potential to various data streams. Cohort comfort models can benefit the building industry by improving the level of thermal comfort among occupants without the need to rely on individual customization which can be unfeasible and too expensive. Further global data collection experiments with multiple occupants are encouraged to investigate the generalization of the potential different cohorts and the potential discovery of thermal preference *signatures* shared across different occupant populations. Additionally, we encourage more research on control strategies that take advantage of the cohort-customization instead of individual-level catering.

### CRediT author statement

**Matias Quintana:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Visualization, Writing - Original Draft. **Stefano Schiavon:** Conceptualization, Visualization, Supervision, Writing - Reviewing and Editing, Funding acquisition. **Federico Tartarini:** Resources, Data Curation, Writing - Reviewing and Editing. **Joyce Kim:** Resources, Data Curation, Writing - Reviewing and Editing. **Clayton Miller:** Conceptualization, Resources, Visualization, Supervision, Project administration, Writing - Reviewing and Editing, Funding acquisition.

### Acknowledgements

This research was funded by the Republic of Singapore’s National Research Foundation through a grant to the Berkeley

Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST2) Program and a Singapore Ministry of Education (MOE) Tier 1 Grant (Ecological Momentary Assessment (EMA) for Built Environment Research). BEARS has been established by the University of California, Berkeley as a center for intellectual excellence in research and education in Singapore. The authors would like to acknowledge the team behind the data collection and processing, including Mario Frei, Yi Ting Teo, and Yun Xuan Chua.

**Appendix A.**

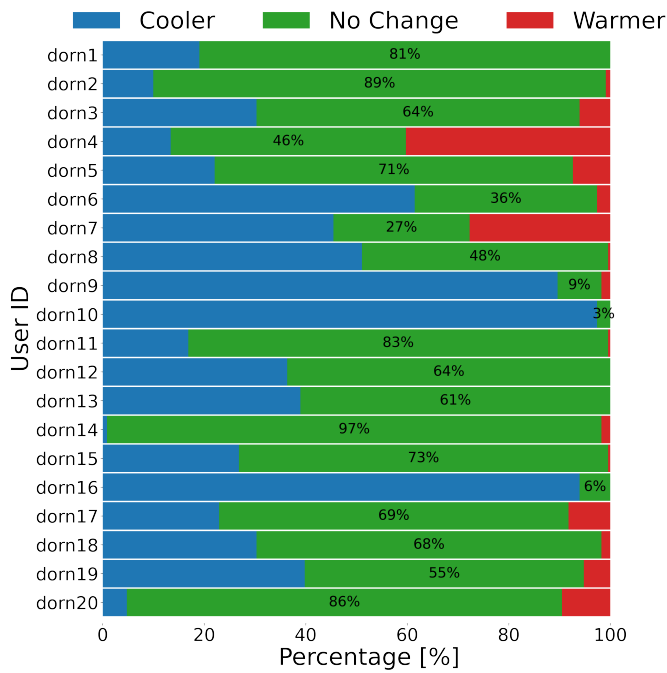
Parameter	Description or value
Train-test ratio	0.8, participant-wise
Model	Random Forest
Model Hyperparameters	Number of trees: 100, 300, 500 Split criterion: Gini index Tree depth: 1, . . . , 10 Min. samples for split: 2, 3, 4 Min. samples on an edge : 1, 2, 3
Model training	Hyperparameter grid search 5-fold Cross-Validation
Model metric	F1-micro score
Clustering	Number of cohorts tested: $k \in [2, 10]$ Metric: Silhouette score
Iterations	Modeling pipeline is repeated 100 times

Table Appendix A.1: Experiments and modeling parameters

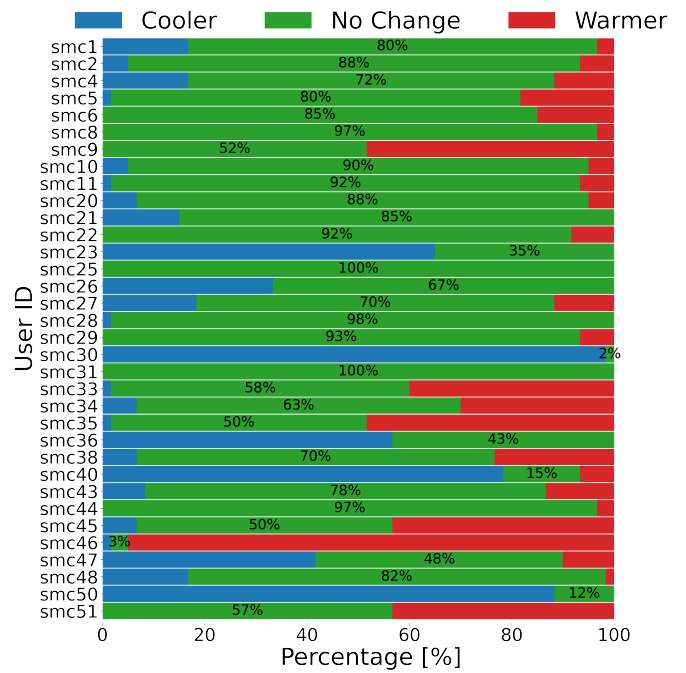
Cohort Approach	Better/Worse	Sex	Height (m)	Weight (kgs)
Dorn:Dist-Cross	Better-off	5M, 4F	168±4.71	64.3±12.04
	Worse-off	5M, 6F	167±10.63	64.55±14
Dorn:Cross	Better-off	6M, 6F	169.08±8.95	64.08±12.99
	Worse-off	4M, 4F	166±7.38	65±13.38
Dorn:Life Satisfaction	Better-off	1M, 3F	165.25±8.95	55.25±5.76
	Worse-off	3M, 3F	170±5.86	70.17±8.53
Dorn:Agreeableness	Better-off	7M, 4F	167.18±8.70	65.82±14.33
	Worse-off	3M, 6F	168.67±8.16	62.78±11.34
SMC:Dist-Cross	Better-off	8M, 4F	169.58±11.50	74.92±15.77
	Worse-off	8M, 14F	166±8.85	73.59±17.02
SMC:Cross	Better-off	9M, 2F	174±7.36	78.36±18.86
	Worse-off	7M, 16F	164.04±9.50	72±14.98

Table Appendix A.2: Breakdown of occupants' metadata based on the cohort approaches they are better or worse-off. Height and Weight columns display the mean ± standard deviation



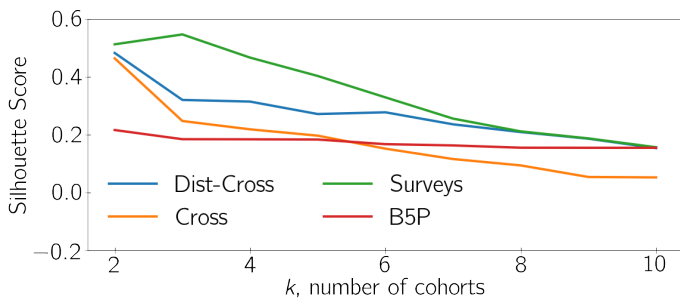


(a) Dorn [55] participants and their first 231 responses per participants.

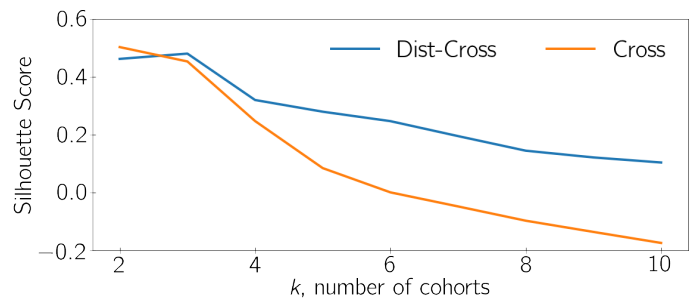


(b) SMC [57] participants and their first 60 responses per participants

Figure Appendix A.1: Thermal preference response distribution for all participants in both datasets. Appendix A.1a shows the Dorn [55] dataset and Appendix A.1b shows the SMC [57] dataset.



(a) Average Silhouette scores for cold start (Surveys and B5P) and warm start (Dist-Cross and Cross) approaches on the Dorn dataset.



(b) Average Silhouette scores for warm start approaches on the SMC dataset.

Figure Appendix A.2: Number of cohorts determined based on the average Silhouette scores at different number of cohorts  $k$  ( $k \in [2, 10]$ ) on both datasets. The higher the Silhouette score the better. Appendix A.2a and Appendix A.2b showcase the values for the Dorn and SMC datasets respectively.

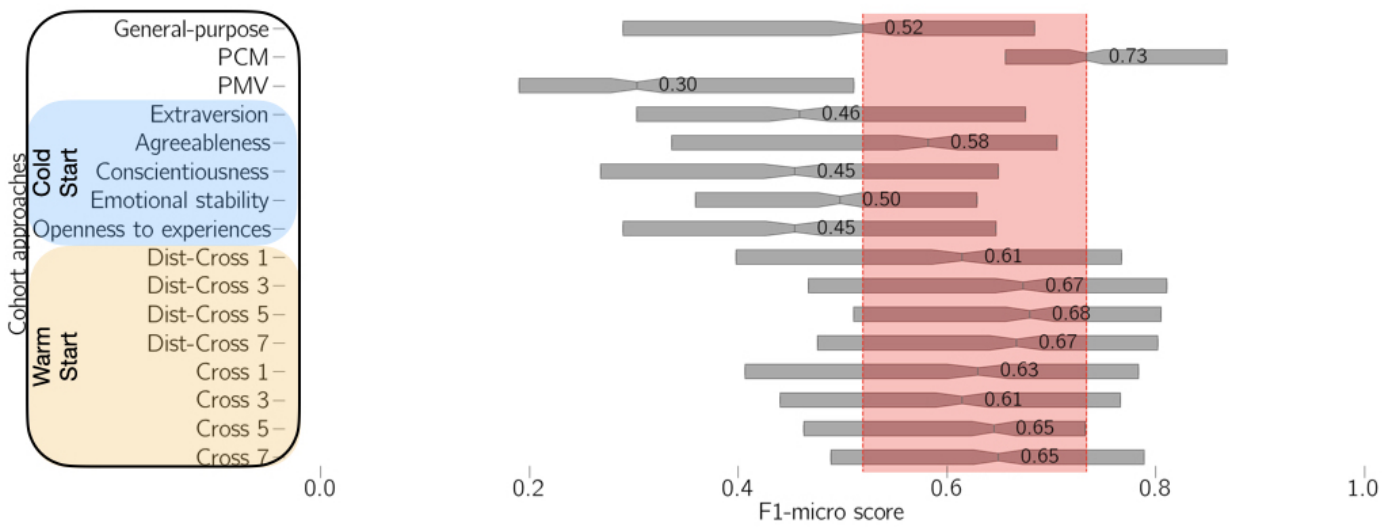


Figure Appendix A.3: Performance results in F1-micro score of 100 iterations for each cohort approach on the Dorn dataset (y-axis). *Cold start* approaches are highlighted by a light-blue region and *Warm start* approaches by a yellow region. These results are complementary to Figure 4. Each personality trait from the B5P (Extraversion, Agreeableness, Conscientiousness, Emotional stability, Openness to experiences) are considered individually as *cold start* approaches. *Warm start* approaches, *Dist-Cross* and *Cross*, have the number of labeled data points used for assignment; i.e., 1, 3, 5, and 7; as suffixes.

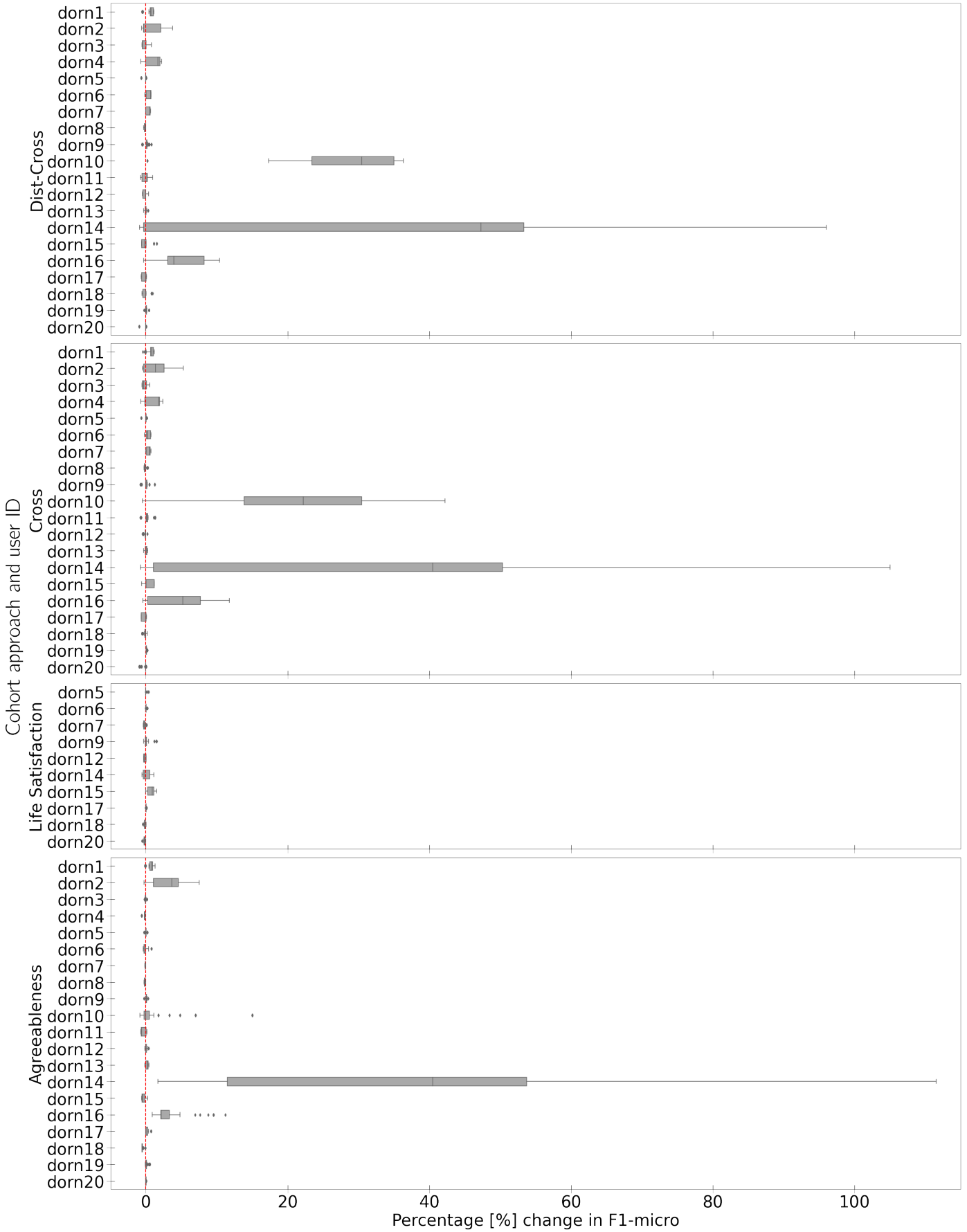


Figure Appendix A.4: Performance percentage change in F1-micro score of 100 iterations for each occupant in each cohort approach (y-axis) for the Dorn dataset. The percentage change is calculated based on the respective cohort approach performance and the **General-purpose** performance. The average value of each occupant is reported in Figure 5.

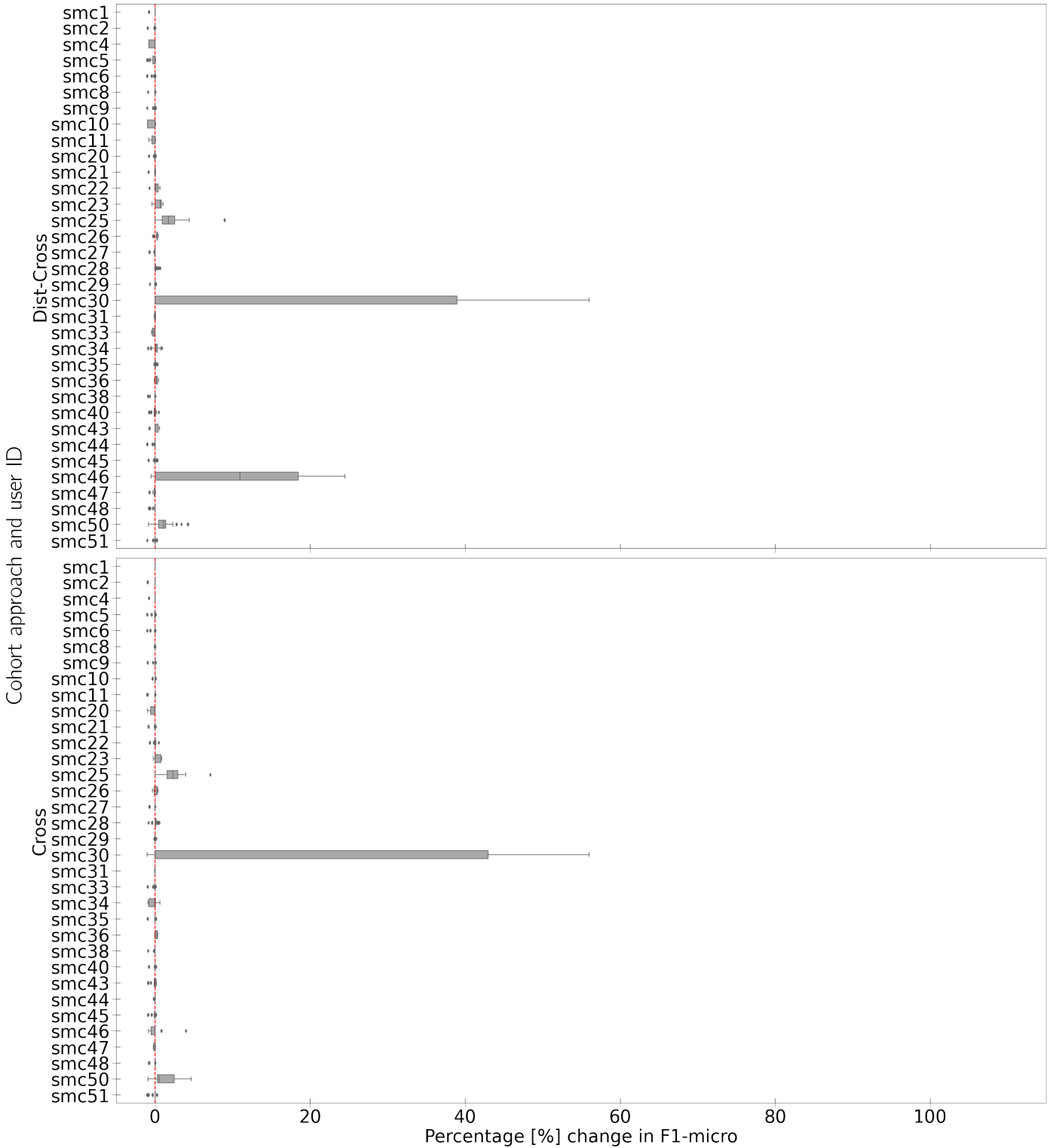


Figure Appendix A.5: Performance percentage change in F1-micro score of 100 iterations for each occupant in each cohort approach (y-axis) for the SMC dataset. The percentage change is calculated based on the respective cohort approach performance and the `General-purpose` performance. The average value of each occupant is reported in Figure 5.

## References

- [1] David Ormandy and Véronique Ezratty. Health and thermal comfort: From WHO guidance to housing strategies. *Energy Policy*, 49:116–121, 2012.
- [2] K Pantavou, G Theoharatos, A Mavrikis, and M Santamouris. Evaluating thermal comfort conditions and health responses during an extremely hot summer in Athens. *Building and Environment*, 46(2):339–344, 2011.
- [3] P. A. Hancock, Jennifer M. Ross, and James L. Szalma. A meta-analysis of performance response under thermal stressors. *Human Factors*, 49(5):851–877, 2007.
- [4] Olli A. Seppänen and William Fisk. Some quantitative relations between indoor environmental quality and work performance or health. *HVAC and R Research*, 12(4):957–973, 2006.
- [5] David P Wyon and Pawel Wargocki. Room temperature effects on office work. In *Creating the productive workplace*, pages 209–220. Taylor & Francis, 2006.
- [6] Fan Zhang, Richard de Dear, and Peter Hancock. Effects of moderate thermal environments on cognitive performance: A multidisciplinary review. *Applied Energy*, 236(July 2018):760–777, 2019.
- [7] M J Mendell and G A Heath. Do indoor pollutants and thermal conditions in schools influence student performance? A critical review of the literature. *Indoor Air*, 15(1):27–52, 2005.
- [8] Pawel Wargocki, Jose Ali Porras-Salazar, and Sergio Contreras-Espinoza. The relationship between classroom temperature and children’s performance in school. *Building and Environment*, 157(February):197–204, 2019.
- [9] Pawel Wargocki and David P. Wyon. Ten questions concerning thermal and indoor air quality effects on the performance of office work and schoolwork. *Building and Environment*, 112:359–366, 2017.
- [10] Li Lan, Zhiwei Lian, and Li Pan. The effects of air temperature on office workers’ well-being, workload and productivity-evaluated with subjective ratings. *Applied Ergonomics*, 42(1):29–36, 2010.
- [11] Sergio Altomonte, Joseph Allen, Philomena Bluysen, Gail Brager, Lisa Hescong, Angela Loder, Stefano Schiavon, Jennifer Veitch, Lily Wang, and Pawel Wargocki. Ten questions concerning well-being in the built environment. *Building and Environment*, page 106949, 2020.
- [12] R. J. De Dear, T. Akimoto, E. A. Arens, G. Brager, C. Candido, K. W.D. Cheong, B. Li, N. Nishihara, S. C. Sekhar, S. Tanabe, J. Toftum, H. Zhang, and Y. Zhu. Progress in thermal comfort research over the last twenty years. *Indoor Air*, 23(6):442–461, 2013.
- [13] Ken Parsons. *Human thermal environments: the effects of hot, moderate, and cold environments on human health, comfort and performance*. CRC press, 2014.
- [14] Lindsay T Graham, Thomas Parkinson, and Stefano Schiavon. Lessons learned from 20 years of CBE’s occupant surveys. *Building & Cities*, 2(1):166–184, 2021.
- [15] P. O. Fanger. *Thermal comfort: analysis and applications in environmental engineering*. McGraw-Hill, New York, 1970.
- [16] Richard J. de Dear and Gail Schiller Brager. Developing an adaptive model of thermal comfort and preference. *ASHRAE Transactions*, 104(1):1–18, 1998.
- [17] Veronika Földváry Ličina, Toby Cheung, Hui Zhang, Richard de Dear, Thomas Parkinson, Edward Arens, Chungyoon Chun, Stefano Schiavon, Maohui Luo, Gail Brager, Peixian Li, and Soazig Kaam. ASHRAE Global Thermal Comfort Database II. *Dataset*, v4:1–4, 2018.
- [18] Toby Cheung, Stefano Schiavon, Thomas Parkinson, Peixian Li, and Gail Brager. Analysis of the accuracy on PMV – PPD model using the ASHRAE Global Thermal Comfort Database II. *Building and Environment*, 153(December 2018):205–217, 2019.
- [19] Joyce Kim, Yuxun Zhou, Stefano Schiavon, Paul Raftery, and Gail Brager. Personal comfort models: Predicting individuals’ thermal preference using occupant heating and cooling behavior and machine learning. *Building and Environment*, 129(December 2017):96–106, 2018.
- [20] Hui Zhang, Edward Arens, Charlie Huizenga, and Taeyoung Han. Thermal sensation and comfort models for non-uniform and transient environments, part II: Local comfort of individual body parts. *Building and Environment*, 45(2):389–398, 2010.
- [21] Eun-jeong Shin and Roberto Yus. Exploring Fairness in Participatory Thermal Comfort Control in Smart Buildings. *BuildSys’17 Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*, 2017.
- [22] Prageeth Jayathissa, Matias Quintana, Mahmoud Abdelrahman, and Clayton Miller. Humans-as-a-sensor for buildings: Intensive longitudinal indoor comfort models. *Buildings*, 10(174):1–23, 2020.
- [23] Seungjae Lee, Ilias Bilionis, Panagiota Karava, and Athanasios Tzempelikos. A Bayesian approach for probabilistic classification and inference of occupant thermal preferences in office buildings. *Building and Environment*, 118:323–343, 2017.
- [24] Maohui Luo, Jiaqing Xie, Yichen Yan, Zhihao Ke, Peiran Yu, Zi Wang, and Jingsi Zhang. Comparing machine learning algorithms in predicting thermal sensation using ASHRAE Comfort Database II. *Energy and Buildings*, 210:109776, 2020.
- [25] Da Li, Carol C. Menassa, and Vineet R. Kamat. Personalized human comfort in indoor building environments under diverse conditioning modes. *Building and Environment*, 126(July):304–317, 2017.
- [26] David Daum, Frédéric Haldi, and Nicolas Morel. A personalized measure of thermal comfort for building controls. *Building and Environment*, 46(1):3–11, 2011.
- [27] Shichao Liu, Stefano Schiavon, Hari Prasanna Das, Costas J Spanos, and Ming Jin. Personal thermal comfort models with wearable sensors. *Building and Environment*, page 106281, 2019.
- [28] Maíra André, Renata De Vecchi, and Roberto Lamberts. User-centered environmental control: a review of current findings on personal conditioning systems and personal comfort models. *Energy and Buildings*, 222, 2020.
- [29] Samuel D. Gosling, Peter J. Rentfrow, and William B. Swann. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6):504–528, 2003.
- [30] Boning Li and Akane Sano. Extraction and Interpretation of Deep Autoencoder-based Temporal Features from Wearables for Forecasting Personalized Mood, Health, and Stress. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(2):1–26, 2020.
- [31] Joyce Kim, Stefano Schiavon, and Gail Brager. Personal comfort models—A new paradigm in thermal comfort for occupant-centric environmental control. *Building and Environment*, 132:114–124, 2018.
- [32] Toby C.T. Cheung, Stefano Schiavon, Elliott T. Gall, Ming Jin, and William W. Nazaroff. Longitudinal assessment of thermal and perceived air quality acceptability in relation to temperature, humidity, and CO2 exposure in Singapore. *Building and Environment*, 115:80–90, 2017.
- [33] Liliana Barrios and Wilhelm Kleiminger. The Comfstat - Automatically sensing thermal comfort for smart thermostats. In *EEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 257–266, mar 2017.
- [34] Peter Xiang Gao and Srinivasan Keshav. Optimal Personal Comfort Management Using SPOT+. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, BuildSys’13, pages 1–8, New York, NY, USA, 2013. ACM, ACM.
- [35] Peter Xiang Gao and Srinivasan Keshav. SPOT: a smart personalized office thermal control system. In *Proceedings of the Fourth International Conference on Future Energy Systems*, e-Energy ’13, pages 237–246, New York, NY, USA, 2013. ACM, ACM.
- [36] Farrokh Jazizadeh and S Pradeep. Can computers visually quantify human thermal comfort?: Short paper. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*, pages 95–98. ACM, 2016.
- [37] Parisa Mansourifard, Farrokh Jazizadeh, Bhaskar Krishnamachari, and Burcin Becerik-Gerber. Online learning for personalized room-level thermal control: A multi-armed bandit framework. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, BuildSys’13, page 1–8, New York, NY, USA, 2013. Association for Computing Machinery.
- [38] Liang Zhang, Abraham Hang-yat Lam, and Dan Wang. Strategy-proof thermal comfort voting in buildings. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, pages 160–163. ACM, 2014.
- [39] Pimpatsohn Sae-Zhang, Matias Quintana, and Clayton Miller. Differences in thermal comfort state transitional time among comfort preference groups. In *16th Conference of the International Society of Indoor Air Quality and Climate: Creative and Smart Solutions for Better Built Environments, Indoor Air 2020*, number November, 2020.
- [40] Panagiota Antoniadou and Agis M. Papadopoulos. Occupants’ thermal comfort: State of the art and the prospects of personalized assessment in office buildings. *Energy and Buildings*, 153:136–149, 2017.

- [41] Gail Brager, Hui Zhang, and Edward Arens. Evolving opportunities for providing thermal comfort. *Building Research and Information*, 43(3):274–287, 2015.
- [42] Joost Van Hoof, Mitja Mazej, and Jan L.M. Hensen. Thermal comfort: Research and practice. *Frontiers in Bioscience*, 15(2):765–788, 2010.
- [43] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. Facing the cold start problem in recommender systems. *Expert Systems with Applications*, 41(4 PART 2):2065–2073, 2014.
- [44] Xuan Nhat Lam, Thuc Vu, Trong Duc Le, and Anh Duc Duong. Addressing cold-start problem in recommendation systems. *Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication, ICUIMC-2008*, pages 208–211, 2008.
- [45] Andrew I Schein, Alexandrin Popescul, Lyle H Ungar, and David M Pennock. Methods and Metrics for Cold-Start Recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '02*, pages 253–260, 2002.
- [46] Nikola Banovic and John Krumm. Warming Up to Cold Start Personalization. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):1–13, 2018.
- [47] Hsien Te Kao, Shen Yan, Homa Hosseinmardi, Shrikanth Narayanan, Kristina Lerman, and Emilio Ferrara. User-Based Collaborative Filtering Mobile Health System. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4), 2020.
- [48] Elaine Nancy Aron and Arthur Aron. Sensory-Processing Sensitivity and Its Relation to Introversion and Emotionality Sensory-Processing Sensitivity and Its Relation to Introversion and Emotionality. *Journal of Personal and Social Psychology*, 73(2):345–368, 1997.
- [49] Ed Diener, Robert A Emmons, Randy J Larsen, and Sharon Griffin. The Satisfaction With Life Scale. *Journal of Personality Assessment*, 49:71–75, 1985.
- [50] Peter J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational North-Holland and Applied Mathematics*, pages 53–65, 1987.
- [51] Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14, 2001.
- [52] Federico Tartarini and Stefano Schiavon. pythermalcomfort: A Python package for thermal comfort research. *SoftwareX*, 12:100578, 2020.
- [53] ISO7730 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. *Management*, 3(605):e615, 2005.
- [54] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Pasos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [55] Federico Tartarini, Stefano Schiavon, Matias Quintana, and Clayton Miller. Personalized thermal comfort models using wearables and IoT devices. *Indoor Air*.
- [56] Prageeth Jayathissa, Matias Quintana, Tapeesh Sood, Negin Narzarian, and Clayton Miller. Is your clock-face cozie? A smartwatch methodology for the in-situ collection of occupant comfort data. In *CISBAT2019 Climate Resilient Buildings - Energy Efficiency & Renewables in the Digital Era*, Lausanne, Switzerland, 2019.
- [57] Joyce Kim, Fred Bauman, Paul Raftery, Edward Arens, Hui Zhang, Gabe Fierro, Michael Andersen, and David Culler. Occupant comfort and behavior: High-resolution data from a 6-month field study of personal comfort systems with 37 real office workers. *Building and Environment*, 148(September 2018):348–360, 2019.
- [58] Joon Ho Choi and Vivian Loftness. Investigation of human body skin temperatures as a bio-signal to indicate overall thermal sensations. *Building and Environment*, 58:258–269, 2012.
- [59] Chengcheng Shan, Jiawen Hu, Jianhong Wu, Aili Zhang, Guoliang Ding, and Lisa X. Xu. Towards non-intrusive and high accuracy prediction of personal thermal comfort using a few sensitive physiological parameters. *Energy and Buildings*, 207:109594, 2020.
- [60] Matias Quintana, Stefano Schiavon, Kwok Wai Tham, and Clayton Miller. Balancing thermal comfort datasets: We GAN, but should we? In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pages 120–129, Virtual Event, Japan, 2020.
- [61] Rachel L. Morrison and Keith A. Macky. The demands and resources arising from shared office spaces. *Applied Ergonomics*, 60:103–115, 2017.
- [62] Sabina Hodzic, Bettina Kubicek, Lars Uhlig, and Christian Korunka. Activity-based flexible offices: effects on work-related outcomes in a longitudinal study. *Ergonomics*, 64(4):455–473, 2021.