The Squeaky Wheel: machine learning for anomaly detection in subjective thermal comfort votes

Zhe Wang[1, 2], Thomas Parkinson[1], Peixian Li [1,3], Borong Lin [4,*], Tianzhen Hong[2,*]

[1] Centre for Built Environment, University of California Berkeley, CA 94720, USA
[2] Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[3] Department of Civil Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[4] Department of Building Science, Tsinghua University, Beijing 100084, China

*Corresponding email: linbr@tsinghua.edu.cn
*Corresponding email: thong@lbl.gov

ABSTRACT

Anomalous patterns in subjective votes can bias thermal comfort models built using data-driven approaches. A stochastic-based two-step framework to detect outliers in subjective thermal comfort data is proposed to address this problem. The anomaly detection technique involves defining similar conditions using a k-Nearest Neighbor (KNN) method and then quantifying the dissimilarity of the occupants' votes from their peers under similar thermal conditions through a Multivariate Gaussian approach. This framework is used to detect outliers in the ASHRAE Global Thermal Comfort Database I & II. The resulting anomaly-free dataset produced more robust comfort models avoiding dubious predictions. The proposed method has been proven to effectively distinguish outliers from inter-individual variabilities in thermal demand. The proposed anomaly detection framework could easily be applied to other applications with different variables or subjective metrics. Such a tool holds great promise for use in the development of occupancy responsive controls for automated building HVAC systems.

Key words:

Thermal comfort, Subjective votes, ASHRAE Global Thermal Comfort Database, Anomaly detection, k-Nearest Neighbors, Multivariate Gaussian, occupancy responsive controls

1. Introduction

The thermal environment is an important component of the Indoor Environmental Quality (IEQ) of a building. It has been shown, along with acoustics, to exert a marked influence on occupants' overall satisfaction [1], [2], [3] in office buildings. Beyond satisfaction, thermal comfort has been linked to self-reported productivity measures in office workers [4], [5]. The importance of the thermal environment on satisfaction and productivity is often cited as justifying the significant energy use associated with the provision of thermal comfort in buildings. Heating, ventilation, and air-conditioning (HVAC) equipment used to deliver comfortable indoor environments accounts for 50% of total building energy consumption in the US [6], 40% in Europe [7], 33% in Hong Kong [8], and more than 70% in Middle East countries [9]. The efficient and effective management of indoor thermal environments is therefore crucial for both the well-being of occupants and reducing the energy usage and carbon footprint of buildings.

The oft-cited dictum by the prominent management thinker Peter Drucker - you cannot manage what you cannot measure [10] - is particularly relevant to building HVAC operation and occupant satisfaction. Metrics used for thermal comfort assessments can be broadly categorized into either physical (or objective) and subjective measures. Physical metrics describe the indoor environmental quality as measured by sensors, which includes air temperature, relative humidity, radiant temperature, and air

https://doi.org/10.1016/j.buildenv.2019.01.050
https://escholarship.org/uc/item/3z1242jb

speed. It is generally acknowledged, however, that physical measurements alone may not reliably determine or predict occupants' thermal comfort due to the significant inter-individual differences in preference [11]. These differences have been observed across gender [12], age [13], physical fitness [14], and activity levels [15], and are partly influenced by clothing decisions (or corporate dress codes), basal metabolic rates, and a person's unique experience and perception of their thermal environment. These types of effects can result in some people feeling comfortable in the same ambient conditions while others expressing dissatisfaction with the environment.

Metrics from subjective evaluations of thermal comfort can effectively address the limitations of physical measurements by directly asking the occupants about their perception of and satisfaction towards the thermal environment. Frequently used psychometrics include Likert-type scales of thermal sensation, thermal comfort, thermal acceptability, thermal preference, and thermal satisfaction [16]. Each of these rating scales have different objectives and are therefore selected based on application scenarios [16]. The utilization of occupants' subjective thermal responses in the context of building controls and operation is referred to as occupancy responsive control. Comparisons with conventional strategies based solely on objective measures has demonstrated the potential of occupancy responsive control to simultaneously enhance thermal comfort and reduce energy consumption [17], [18] in buildings.

Empirical evidence supporting the use of subjective measures in building HVAC control systems is encouraging. However, there are significant differences in both the structure and reliability of objective and subjective data types that require careful consideration before widespread use in automated control algorithms. Unlike instrumental measurements of the physical environment, subjective evaluations of indoor environments relies on voluntary completion of surveys by occupants. The resulting data is intrinsically different from measurements of quantities as it involves evaluations by people that is subject to concerns of reliability and precision, perhaps more so than data from modern sensors. For example, respondents might misunderstand the question, form a response based on tangential factors, or fail to objectively evaluate their thermal environment for a range of reasons. Additional sources of error may occur in the coding and input of paper-based survey data. These biases and errors result in outliers, defined in statistics as an observation that lies an abnormal distance from other values in a random sample from a population [19]. In this study, outliers refer to those thermal comfort votes that are substantially and illegitimately different from others that are comparable. Whilst a seemingly erroneous vote may in fact be a valid response from an occupant towards the extreme end of a given population, such an outlier introduces noise and uncertainty to any type of model being built on subjective data and may result in the specification of suboptimal control strategies for the building management system (BMS). A method of detecting and handling such outliers is therefore critical to the successful utilization of subjective measures of thermal comfort for automated building HVAC controls.

1.1 Anomaly detection in building HVAC systems

Outlier detection techniques, namely anomaly detection, have played an important role in building and HVAC system operations for decades. For example, Fault Diagnosis and Detection (FDD) is a typical application of anomaly detection that monitors building HVAC systems to identify faults [20]. There are two typical methods to detecting outliers: the model-based approach or the stochastic-based approach. A model-based approach aims to build a physical model to predict a reasonable range for a normal observation. If the observed value lies far enough outside the predicted reasonable range then it is flagged as an outlier. This approach was used by Yu et al. to detect faults in building HVAC systems [20]. Other notable examples of applying model-based anomaly detection method in the built environment context include FDD for air-handling units [21] or HVAC compressors [22]. Whilst the model-based approach can result in an effective anomaly detection tool, it does require detailed information and expert knowledge about the particular HVAC system and is therefore suitable for clearly defined and well-established technologies. Some of the barriers to widespread uptake of model-based anomaly detections in buildings [23] may be removed by recent innovations in Building Information Models (BIM).

In contrast to the model-based approach to anomaly detection, a stochastic-based approach assumes and fits a statistical distribution of the target variable from either historical data or a group of peers with similar attributes and under similar conditions. If a new observation has a low probability density given the statistical distribution fitted to earlier data then it is flagged as an anomaly. Examples of stochastic anomaly detection techniques in building HVAC systems can be found in other research literature [24], [25]. Whether a model-based or stochastic-based approach is appropriate largely depends on the availability of input data and the desired application of the model, and is therefore determined on a case-by-case basis. The model-based approach requires a white-box physics model with all the requisite information, which poses significant challenges particularly for complicated systems. Alternatively, the stochastic-based approach is data intensive, requiring a relatively large database with adequate data coverage to fit a robust statistical model for the target variables. The large amount of data collected by increasingly pervasive Building Automation System (BAS) and the impending Internet of Things (IoT) revolution make the stochastic-based anomaly detection methods [26] more attractive for future control strategies.

1.2 Research aims

Although the heat-balance based PMV-PPD model [27] and the adaptive comfort model [28] are available and extensively used in thermal comfort studies, they are concerned with the prediction of human perception and are therefore less resolved than physical models used for air-handling units or compressors. There appear to be few attempts to apply anomaly detection techniques to subjective thermal comfort data within the research literature. This is somewhat surprising in light of the increasing attention given to occupant responsive controls. As such, this study proposes a stochastic-based two-step anomaly detection framework to automatically flag potential outliers in subjective thermal comfort datasets. The proposed method is tested using the recently published ASHRAE Thermal Comfort Database II [29].

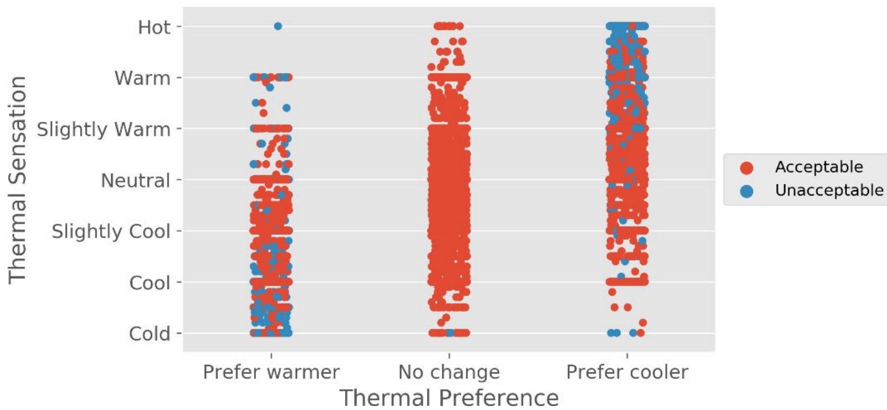
2. Methods

This section summarizes the ASHRAE Global Thermal Comfort Database and visualizes the presence of outliers to demonstrate the need for anomaly detection in subjective thermal comfort data. A detailed procedure of the anomaly detection method is introduced (section 2.2) with an explanation of how the proposed method is able to differentiate individual differences from true outliers (section 2.3).

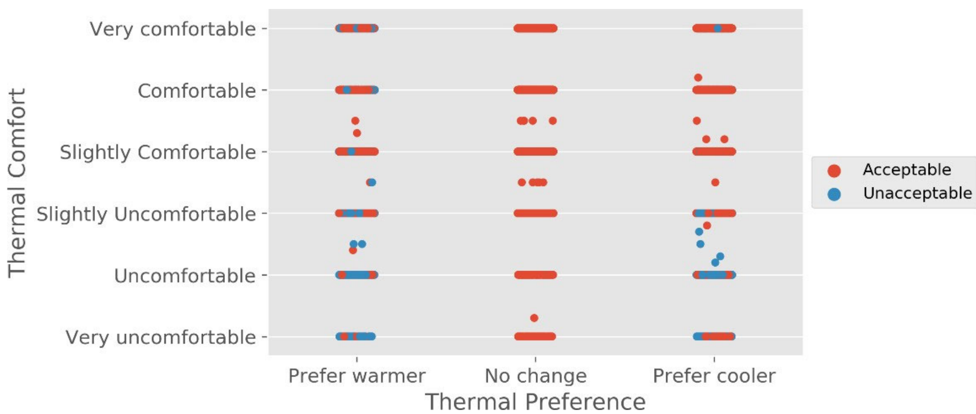
2.1 ASHRAE Global Thermal Comfort Database

A large dataset is required to build a stochastic model for anomaly detection. The ASHRAE Global Thermal Comfort Database II was a key deliverable of an international effort to collate both physical measurements and contemporaneous subjective evaluations from 52 field studies conducted in 160 buildings around the world [29]. Combining the original RP-884 Database (now referred to as Thermal Comfort Database I) [30] with the new ASHRAE Database II resulted in 107,583 records. This combined database is collectively referred to here as the comfort database. A resource of this size and detail is unprecedented in thermal comfort research, and affords an ideal opportunity to develop and test a stochastic-based model for anomaly detection.

The comfort database contains four subjective thermal metrics: thermal sensation (7-point), thermal comfort (6-point), thermal preference (3-point) and thermal acceptability (2-point). Subsetting the database with records containing all four subjective measures resulted in approximately 11,000 rows that were used in this analysis. Figure 1 plots the distribution of thermal preference and thermal acceptability on thermal sensation (Figure 1a) and thermal comfort scales (Figure 1b). These simple visualizations offer some evidence for strange or irrational voting behaviors. For example, the blue dots in the lower right segment of Figure 1(a) illustrate occupants who reported feeling unacceptably cold but still preferred cooler; the blue dots in the upper left segment of Figure 1(b) show occupants who reported feeling very comfortable but still deemed their thermal environment to be unacceptable and wanted warmer.



(a) Thermal Sensation



(b) Thermal Comfort

Figure 1. Jitter plots of thermal preference and thermal acceptability votes on (a) thermal sensation scale and (b) thermal comfort scale. Each dot marks an individual vote and the colors indicate the thermal acceptability as described in the legend. Most studies in the ASHRAE Database used thermal comfort scale allowing integer votes only.

There could be many reasons why such counter-intuitive or irrational votes exist, but it is more likely that these outliers are the result of dubious responses or incorrect data coding and input. If all records in the dataset were assumed to be true and subsequently used to develop and train an occupant responsive HVAC control model, the resulting predictions might be suboptimal by encouraging antagonistic operation or oscillations in equipment activity. This is especially true if the model is built using algorithms sensitive to outliers. It is therefore important to employ a method of flagging potential outliers before they are used to train building HVAC control models.

2.2 Anomaly detection method

The fundamental logic underpinning the development of a stochastic-based model for anomaly detection, as illustrated in Figure 2, is that an outlier is flagged when an occupant's vote is significantly different from his/her peers under similar conditions. After rescaling the inputs (step 1), there are two key steps: defining similar conditions (step 2) and quantifying the difference between a vote and its peers (step 3). The proposed anomaly detection technique uses the k-Nearest Neighbor algorithm (KNN) to define similar conditions (step 2) based on thermal sensation and thermal comfort votes and then uses Multivariate Gaussian methods to quantify the difference in thermal preference or thermal acceptability (step 3) between records with similar thermal sensation and thermal comfort votes. The key assumption behind this logic is that occupants with similar thermal sensation and comfort votes would have similar thermal preference and acceptability.

Pseudocode

```
--------------------------------------------------------------------------------
For each observation in the database:
    Rescaling each dimension to the same range of 0 to 1                          Step1: rescaling
--------------------------------------------------------------------------------
    Find its nearest neighbors based on thermal sensation and thermal comfort by calculating     Step2: defining similar conditions
      its Euclidean Distance with the remaining observations in the database
--------------------------------------------------------------------------------
    Fit the simple multivariate Gaussian distribution on thermal acceptability and thermal      Step3: quantifying dissimilarities
      preference with its neighbors
    Calculate the p-value of the specific observation
--------------------------------------------------------------------------------
    if the p-value is no less than the threshold:                                 Step4: making decisions
        Flagged as a normal observation
    else:
        Flagged as a potential outlier
    end
end
```

Figure 2. Pseudocode for the proposed anomaly detection algorithm

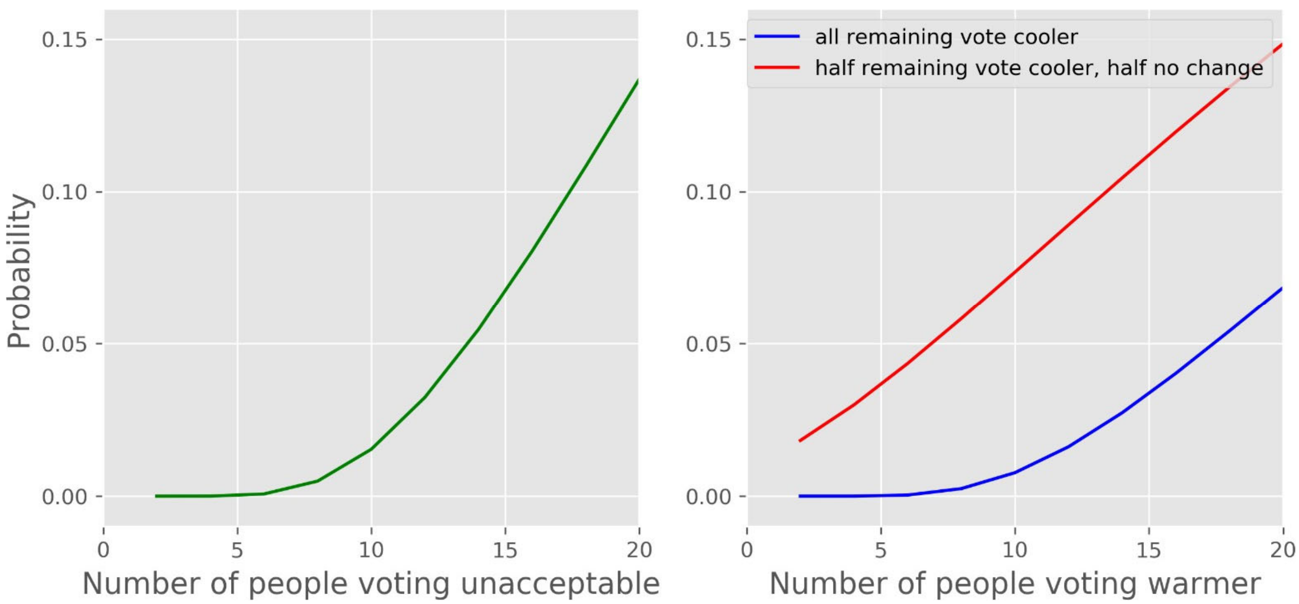
The first step of the anomaly detection method is rescaling the input data to harmonize the dimensions of different scales to the same range, which is 0 to 1 in this case. Data rescaling guarantees different dimensions have the same weight in the Euclidean Distance calculation used by the algorithm in subsequent steps. For example, the 7-point thermal sensation scale (from -3 to 3) is different from the binary option of thermal acceptability (from 0 to 1). Without rescaling, the difference in the Euclidean distance between *unacceptable (0)* and *acceptable (1)* is equivalent to the difference between *neutral (0)* to *slightly warm (1)*. The process of rescaling removes this discrepancy and allows for equal comparison between parameters.

After data rescaling, the second step in the process is to determine similar thermal conditions for each record in the database based on the Euclidean Distance. The KNN is used as a pattern recognition method to identify samples that are similar. Other distance measures such as Manhattan Distance or Chebyshev Distance would also be suitable. The key parameter for KNN is the k value, which was set as one tenth of the total sample size for the present analysis. Considerations for the selection of hyperparameters such as k is discussed in more detail in Section 4.

The third and final step is to quantify the difference in voting between each individual case (i.e. an occupant) and its peers (or neighbors) in a similar condition. This step requires the assumption that subjective thermal votes follow a Multivariate Gaussian distribution in order to fit the model and derive the mean and variance. From this, the probability of a new observation (p-value) being within the expected range of values is determined based on the fitted Gaussian distribution. If the probability is below a predefined threshold value then the observation is classified as an outlier. The choice of the threshold value is another important hyperparameter that significantly influences the performance of anomaly detection method.

2.3 Distinguish individual differences from outliers

Although the method of anomaly detection described here is capable of efficiently identifying erroneous votes, the nature of subjective data means that an occupant, in some cases, will vote differently than their peers under similar thermal comfort conditions due to individual differences and preferences. A robust and reliable anomaly detection method should therefore be able to distinguish individual differences from actual outliers.



(a) thermal acceptability votes (b) thermal preference votes

Figure 3. The calculated probability (p-value) of occupants voting *unacceptable* (left) or for *prefer warmer* (right) in different hypothetical scenarios (assuming a sample size of 100). The calculated probability increases as the total number of occupants (or neighbors) vote in a similar fashion.

Figure 3 illustrates how the p-value is used to determine a true outlier in a hypothetical scenario for a group of 100 occupants in a similar thermal environment. If 99 of the 100 occupants voted *acceptable*, and only one occupant voted *unacceptable*, then that occupant voting *unacceptable* would likely be flagged as an outlier due to the close to zero calculated probability as shown in Figure 3(a). But if 90 occupants voted *acceptable* and 10 voted *unacceptable*, the p-value of this case is approximately 2% which may be deemed likely depending on the choice of threshold value. Thermal preference is more complicated than acceptability as there are three possible choices. If the majority of occupants in a similar thermal environment voted *prefer cooler*, then the probability of someone voting *prefer warmer* is lower, as shown by the blue line in Figure 3(a). An occupant voting *prefer warmer* would likely be flagged as an outlier in this case, since the Euclidean distance of his/her vote is far from the popular vote. However, if clear consensus is not reached, with half of the occupants voting *prefer cooler* and the other half voting *prefer no change*, the fitted variance would be high and the probability of someone voting *prefer warmer* would be higher as shown by the red line in Figure 3(b). In this scenario, the occupant voting *prefer warmer* is less likely to be detected as an outlier because of the difference in the evaluation by their peers. Individual difference in this specific application inflate the variance of the Gaussian distribution providing a robust and reliable method to differentiate between an inter-individual variability and a true outlier.

The proposed algorithm has two mechanisms to avoid incorrectly flagging individual differences as outliers. First, similar thermal conditions are defined using votes of thermal sensation and thermal comfort rather than instrumental measurements of thermal parameters such as air temperature. If an occupants' thermal sensation or comfort vote is markedly different from their peers under similar thermal conditions as defined by air temperature and humidity, it may be due to individual difference and should not be flagged as an outlier. Second, the calculated p-value from the fitted Gaussian distribution quantifies and considers the dissimilarity between an occupant and their peers. If their neighbors in a similar thermal environment have markedly different thermal preference votes, for example, then the fitted variance of the Gaussian distribution would be high. The resulting p-value would be high, as shown by the red line compared with the blue line in Figure 3(b), and therefore less likely to result in an incorrect outlier classification irrespective of how an occupant perceives their immediate thermal environment.

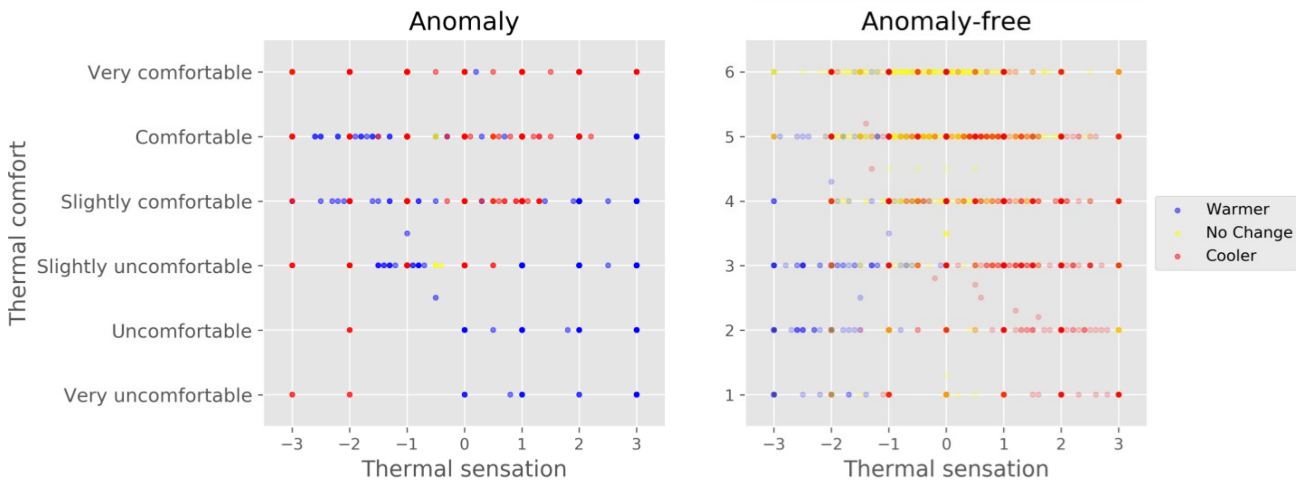
3. Results

The anomaly detection method discussed in the previous section is applied to ASHRAE Database I & II to detect outliers in Section 3.1, before discussing how a dataset free of outliers is able to provide more robust thermal comfort models compared with the original database (Section 3.2).

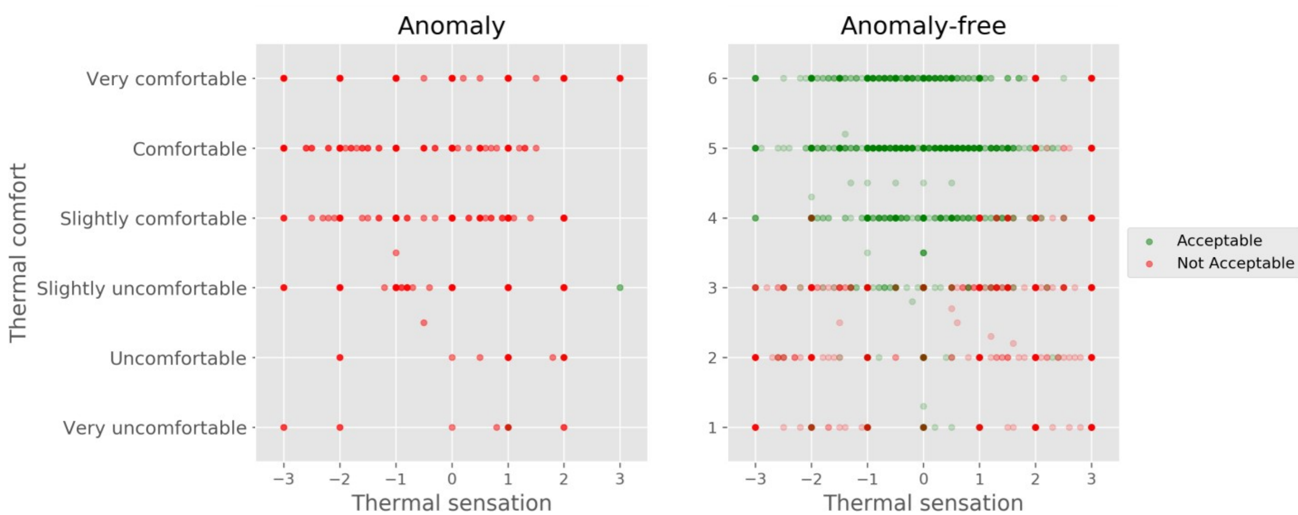
3.1 Anomaly detection

The results of the anomaly detection method applied to subjective votes of thermal comfort in ASHRAE Database I & II are shown in Figure 4. Those votes flagged as outliers are shown on the right plots distinct from the normal or reasonable votes shown on the left plots. Examples of unusual thermal preference voting patterns include occupants reporting feeling *hot* but still *prefer warmer* (blue dots on the right), occupants reporting feeling *cold* but still *prefer cooler* (red dots on the left), and some occupants voting *comfortable* or *very comfortable* but indicating they would *prefer warmer* or *cooler* (red or blue dots on the top). The majority of the flagged outliers for thermal acceptability are occupants who deem the thermal environment as *comfortable* but *unacceptable* (red dots on the top), which is considered unusual since acceptability is often considered to be more easily achieved than comfort.

Some of the anomalous votes shown in Figure 4 are likely to be incorrectly classified and instead may represent unusual but not erroneous voting. For example, occupants who vote *prefer warmer* when they feel *cold* (blue dots on the left of Figure 4a), or those who feel the thermal environment is *unacceptable* but *comfortable* (red dots on the bottom of Figure 4b). These false positives occur because outliers are determined by both the thermal preference and thermal acceptability votes. If either thermal preference or thermal acceptability vote is markedly different from others, then that occupant is more likely to be flagged as an outlier even though their other vote is reasonable.



(a) Thermal preference



(b) Thermal acceptability

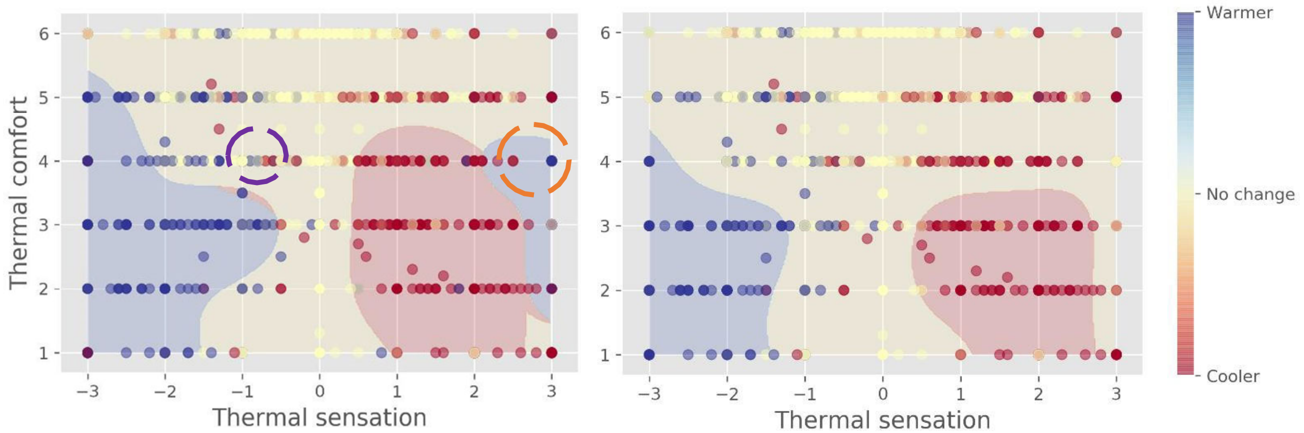
Figure 4. The results of the outlier detection method when used on thermal preference (top) and thermal acceptability (bottom) votes in the ASHRAE Database I & II. The right plots show the observations flagged as anomalous, and the left plots show the dataset with those anomalies removed.

As shown in Figure 4, there are still anomalies observed in the anomaly-free dataset. For example, in Fig 4(a) occupants with slightly cool and cool votes reported to be comfortable and very comfortable and yet preferred to be cooler. Likewise, in Fig 4(b), though understandable but occupants with votes warm and hot on TS reported to be comfortable and uncomfortable. Our justifications are twofold. First, no machine learning algorithm could be 100% accurate, it is possible that some outliers failed to be detected. Second, the boundary between outliers and individual difference are blurry in the thermal comfort field. For instance, if subjects personally prefer a very cool environment, it is possible that they feel slightly cool and comfortable but still preferred to be cooler. Because of this ambiguity, there is a trade-off between a high precision rate and a high recall rate. An aggressive algorithm, which aims to detect as many outliers as possible, would also improve the chance of incorrectly labelling individual difference as outliers. A conservative algorithm, which aims to prevent incorrectly labelling individual difference as outliers, would increase the chance of failing to detect true outliers. This tradeoff could be optimized by tuning hyperparameters of the algorithm, which would be further discussed in Section 4.

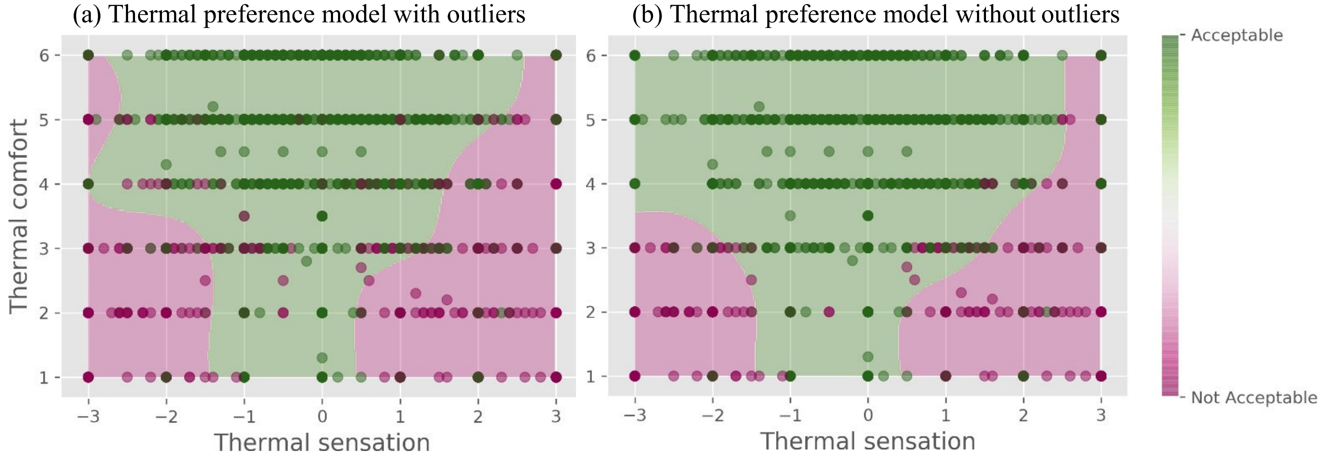
3.2 Influence of outliers on thermal comfort models

A common motivation for collecting subjective thermal comfort data is the development of a model capable of predicting thermal comfort in different contexts. For instance, the PMV-PPD model is a popular tool used to predict the percentage of dissatisfied occupants based on environmental (e.g. air temperature) and personal parameters (e.g. clothing) [27], and the adaptive comfort model is used to derive an acceptable temperature range for buildings in different climates using outdoor temperature [28]. Both of these are data-driven models, where the accuracy and predictive power is largely dependent upon the data used to develop the model. If the training data is noisy or unreliable, the performance of the resulting model will be compromised regardless of which sophisticated statistical tools are used to develop it. This idea is encapsulated in the popular 'garbage in, garbage out' modeling rule that is widely known in the field of data science.

To demonstrate the effect that outliers or anomalies have on the performance of a predictive thermal comfort tool, two models for thermal preference and thermal acceptability were developed using a Support Vector Machine (SVM) algorithm based on thermal sensation and thermal comfort votes. These models are not meaningful beyond this analysis, but rather serve as a useful demonstration of the negative influence of outliers on model development.



(a) Thermal preference model with outliers (b) Thermal preference model without outliers



(c) Thermal acceptability model with outliers (d) Thermal acceptability model without outliers

Figure 5. Models of thermal preference (top) and thermal acceptability based on thermal comfort and thermal sensation votes. An SVM algorithm was used to develop the models based on a dataset with outliers (left) and without outliers (right). The colored shading shows the classification boundaries of the predicted votes. The purple dotted circle marks an example individual difference where occupants feel *slightly comfortable* and *slightly cool*, with some *prefer no change* and others *prefer warmer* or *prefer cooler*. In contrast, the orange dotted oval indicates a likely outlier as the majority of occupants voted *prefer cooler* in similar conditions.

Figure 5(a) offers a visual representation of how the proposed anomaly detection method is able to discern the outliers from individual differences or preferences. Figure 5(a) and 5(c) show the prediction models built using the full dataset with any outliers still present; figure 5(b) and 5(d) show the models built using the dataset with outliers removed. The presence of outliers in the thermal preference model led to some strange outputs indicated in the blue region on the right of Figure 5(a).There it predicts that occupants will *prefer warmer* when they are *slightly comfortably warm* or *slightly uncomfortably warm*. However, after flagging and excluding the potential outliers, the predictive performance of the thermal preference model (Figure 5b) appears to improve. The thermal acceptability model is less affected by the outliers than the thermal preference model. The model trained by the dataset without outliers predicts a higher acceptability rate (Figure 5d)) when thermal comfort votes are high than the model trained with the entire dataset (Figure 5c)).

4. Discussion

This section discusses the impact of hyperparameter tuning on model performance and the trade-off between model robustness and representation of occupants' thermal perception.

4.1 Hyperparameter settings

Developing an anomaly detection algorithm using the proposed method requires the definition of three important hyper-parameters: the choice of distribution family (simple multivariate or covariate Gaussian distribution), the value of k (the number of neighbors required for a condition to be considered similar), and the threshold of the *p-value* set to detect outliers. These hyperparameters are important as they have a significant influence on the overall performance of the anomaly detection method.

Before evaluating the effect of tuning the hyperparameters, it is necessary to define a numerical indicator to determine the performance of the anomaly detection. The accuracy rate, True Positive Rate (TPR), True Negative Rate (TNR), and F1-score are widely used numerical evaluators for classification problems. However, these metrics are not suitable in this particular application because the true outliers need to be known to calculate the accuracy rate. Unfortunately there is no way to determine which observation is true outlier in the ASHRAE Thermal Comfort Database I & II. Instead, the prediction accuracy of the thermal preference/acceptability classifier (section 3.2) was used as an indirect performance indicator. Figure 5 indicated two causes of incorrect predictions in thermal preference or acceptability: individual difference and the existence of outliers. If those outliers are detected and removed then the SVM classifier should in theory be more accurate. Accordingly, the prediction accuracy of the SVM classifier was chosen as an indirect performance metric to evaluate the anomaly detection and test the hyperparameter settings.

Variable Dependence

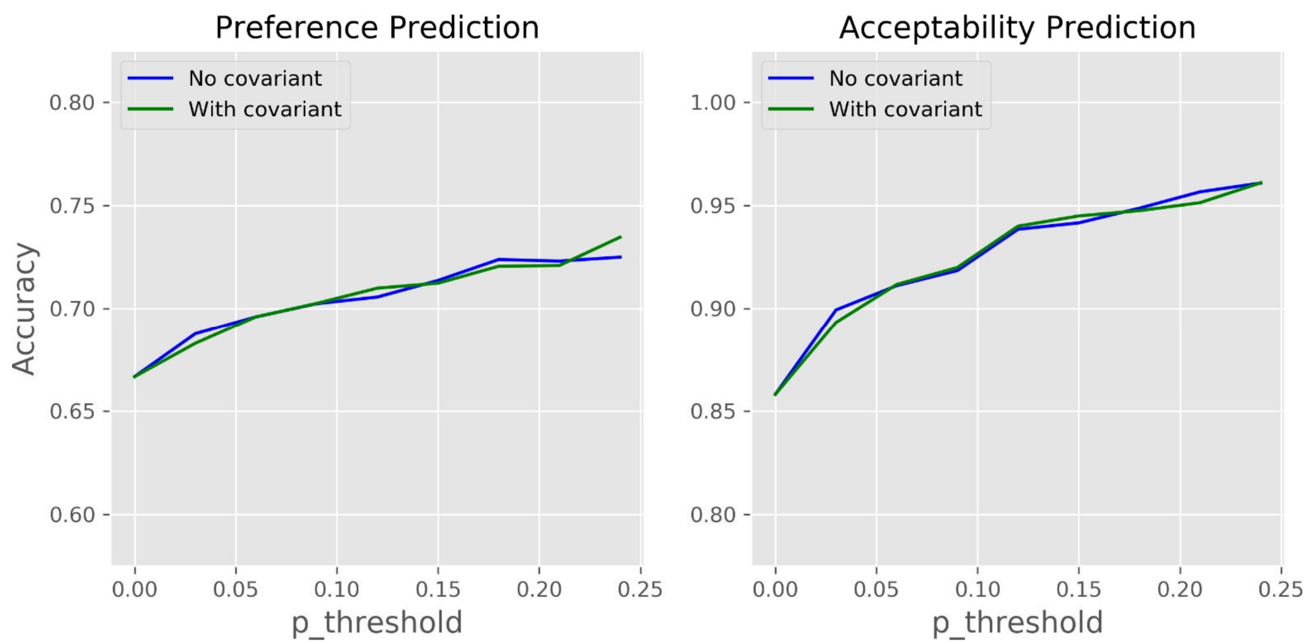
The density function for a 2-dimensional vector is calculated in two ways. If the two attributes are independent, the probability density is the product of the probability densities of each attribute, as shown in Equation 1. The probability density of each attribute is a Univariate Gaussian Distribution and could be calculated from Equation 2, where μ and σ denotes the mean and standard deviation of thermal preference and thermal acceptability respectively. If the two attributes are correlated, then the Multivariate Gaussian Distribution should be used to calculate the probability density, as shown in Equation 3, where μ and Σ denotes the mean vector and the covariance matrix respectively. The first approach to calculate the probability density is a simplified version of the second approach, which ignores the covariant terms in the covariance matrix, assuming the covariance matrix in the form of: $\begin{bmatrix} \sigma_{TP}^2 & 0 \\ 0 & \sigma_{TA}^2 \end{bmatrix}$

$$\mathrm{p}(TP, TA|\mu, \Sigma) = \mathrm{p}(TP|\mu, \Sigma) * \mathrm{p}(TA|\mu, \Sigma) \qquad \text{Equation (1)}$$

$$\mathrm{p}(x|\mu, \Sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) \qquad \text{Equation (2)}$$

$$p(TP, TA|\mu, \Sigma) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} T & P \\ T & A \end{bmatrix} - \mu\right)^T \Sigma^{-1} \left(\begin{bmatrix} T & P \\ T & A \end{bmatrix} - \mu\right)\right) \quad \text{Equation (3)}$$

Both approaches were tested for the anomaly detection method, and Figure 6 shows an insignificant difference between the results. The slight discrepancy may be explained by the very low correlation (Pearson coefficient) between thermal preference and thermal acceptability shown in Table 1. The close-to-zero covariant terms are therefore unsurprising, and explain why including covariant terms did not make any significant difference. Whether the covariance should be considered or not does not make significant difference if the parameters are independent.



(a) Prediction accuracy

Figure 6. A comparison between the performance of anomaly detector of thermal preference and thermal acceptability both with (blue) and without (green) a covariant term.

Table 1 Pearson coefficient between thermal comfort metrics

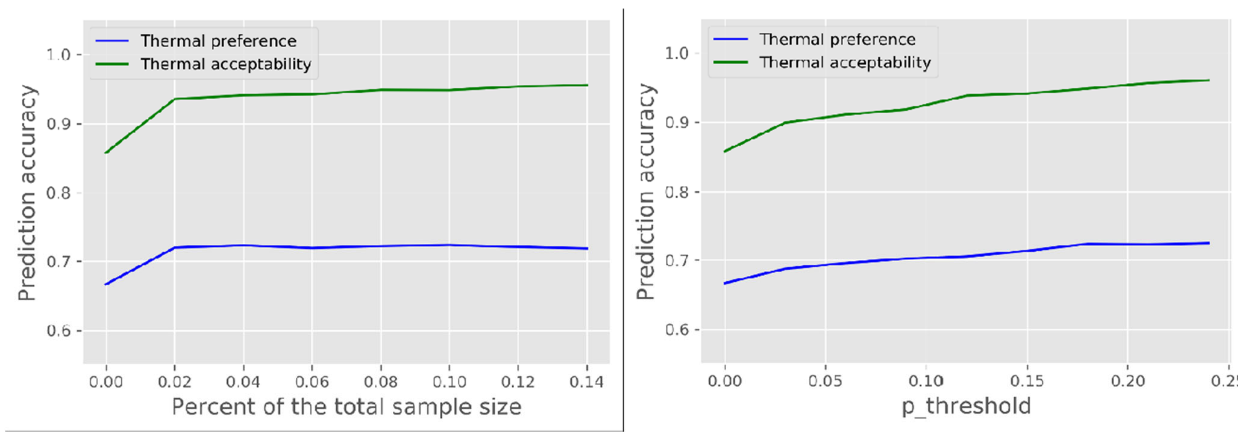
	Thermal Sensation	Thermal Comfort	Thermal Preference	Thermal Acceptability
Thermal Sensation		-0.15	-0.12	-0.24
Thermal Comfort	-0.15		0.04	0.40
Thermal Preference	-0.12	0.04		0.00
Thermal Acceptability	-0.24	0.40	0.00	

Number of neighbors (k)

The second important hyper-parameter for tuning the anomaly detection algorithm is the value of *k*, which sets the number of neighbors required to classify given thermal conditions as similar. If k is set too low it could lead to overfitting and reduce model reliability, and if it is set too high it might underfit. Figure 7(a) show the effect of the *k* value on the prediction accuracy for the thermal preference and thermal acceptability models. Increasing the value of *k* slightly improves the prediction accuracy for thermal acceptability but has marginal influence on the performance of the thermal preference model. This is because the subset of the ASHRAE Thermal Comfort Database I & II used for this analysis is large, so using only 2% of the data is sufficient in detecting outliers. The model presented in Section 3 used 10% of the total sample size as the number of neighbors to define similar thermal conditions. To provide a robust result, a higher proportion of samples might need to be selected to define similar conditions in applications where the sample size is smaller.

P-value Threshold

The influence of the *p-value* threshold on the model prediction accuracy is shown in Figure 7(b). A higher threshold value means more observations are likely to be detected as outliers and removed from the training dataset used to build a thermal comfort model. Since more points with unusual voting patterns are removed by a higher p-value, there is likely to be more unanimity in the remaining data that varies in a more predictable way. Therefore, the prediction accuracy will increase with a higher *p-value* threshold, as shown in Figure 7(b). However, raising *p-value* threshold would also increase the risk of detecting normal or true observations as outliers (false negative). Setting the value of *k* depends on the application and the quality of the available training data. If the desire is to use as many observations as possible in the interest of developing a robust model, then a low threshold may be set to limit the number of observations flagged as outliers. On the other hand, if the training data is known to be noisy, which is often the case for subjective thermal comfort data from surveys, then a less strict threshold might be preferred. The model used in Section 3 had a *p-value* threshold of 0.15. Determining an appropriate *p-value* therefore requires a balance of prediction accuracy, and representation of the available data and future prediction performance.



(a) Value of k (b) Threshold of p

Figure 7. The effect of the k value (left) and the p value (right) hyperparameters on the model prediction accuracy for thermal preference (blue) and thermal acceptability (green). A k-value of 0 indicates that no outliers were removed from the training dataset.

4.2 Limitations of current study and proposed future work

The proposed method to detect outliers, whilst delivering promising results, should not be taken as a suggestion for a wholesale reliance on mathematical or statistical procedures designed to remove different voting behaviors and flatten individual differences in order to develop reliable models for building HVAC control. Indeed, the question on what is the best approach to provide comfort for occupants with unusual or different thermal preferences remains an open question, and one that is likely to be addressed on a case-by-case basis. From the statistical point of view, if the subject's preference is different from the remaining, he/she would be considered as an outlier. But whether this outlier's opinion should be ignored or considered in the practice of building control is beyond the discussion of this paper. Instead, it is the hope of the authors that the procedures introduced in the present work to automate anomaly detection will draw attention to the importance of considering the effect of outliers on thermal comfort model development and performance. By highlighting spurious voting patterns, those tasked with implementing occupancy responsive control solutions can adjust the models to ensure appropriate and efficient system response to a range of different comfort demands.

In this paper, we applied the anomaly detection algorithm to detect erroneous voting pattern among the four common subjective thermal metrics, which does not consider any environmental measures such as temperature or relative humidity.

However, the stochastic-based two-step framework we proposed could be applied to other applications, not only subjective survey but also measured data. The two steps are first to define similar condition, and then to quantify the dissimilarities between individual measurement with its neighbors under similar condition. It is also worthy to point out that, other algorithms could be used for defining similar conditions (e.g. density based clustering) or quantifying dissimilarities (e.g. distance based dissimilarity).

5. Conclusions

Outliers from strange voting behaviors have been found in databases of subjective thermal comfort votes that could bias subsequent models or lead to suboptimal operation of building automation systems. An efficient method of identifying and handling outliers is needed to facilitate the utilization of subjective thermal metrics in automated building HVAC operation. A stochastic-based two-step framework has been proposed to detect outliers in subjective thermal comfort data. The first step is to define similar conditions using a KNN algorithm and then apply the Multivariate Gaussian method to compare an occupant vote to their peers under similar thermal conditions to detect possible outliers. This method was shown to be capable of distinguishing outliers from real but unusual voting patterns arising from personal preferences. The anomaly detection algorithm was used to successfully determine anomalies in the ASHRAE Global Thermal Comfort Database. Using the anomaly-free dataset led to the development and training of more robust thermal comfort models that were less likely to make strange predictions. It would be possible to improve the performance of the anomaly detection process by tuning the hyperparameters further depending on the dataset used. This framework could easily be applied in other applications or settings with different variables or metrics of comfort. We believe the proposed method could help researchers efficiently detect potential outliers in large datasets, and also be a powerful data processing tool for practitioners developing occupant responsive building controls.

Acknowledgements

The second author is supported by the Republic of Singapore's National Research Foundation Singapore through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley as a center for intellectual excellence in research and education in Singapore.

References

[1] Z. Wang, H. Zhao, B. Lin, Y. Zhu, Q. Ouyang, and J. Yu, “Investigation of indoor environment quality of Chinese large-hub airport terminal buildings through longitudinal field measurement and subjective survey,” *Build. Environ.*, vol. 94, pp. 593–605, Dec. 2015.

[2] M. Frontczak, S. Schiavon, J. Goins, E. Arens, H. Zhang, and P. Wargocki, “Quantitative relationships between occupant satisfaction and satisfaction aspects of indoor environmental quality and building design,” *Indoor Air*, vol. 22, no. 2, pp. 119–131, Apr. 2012.

[3] J. Kim and R. de Dear, “Workspace satisfaction: The privacy-communication trade-off in open-plan offices,” *J. Environ. Psychol.*, vol. 36, pp. 18–26, Dec. 2013.

[4] S. Tanabe, Y. Iwahashi, S. Tsushima, and N. Nishihara, “Thermal comfort and productivity in offices under mandatory electricity savings after the Great East Japan earthquake,” *Archit. Sci. Rev.*, vol. 56, no. 1, pp. 4–13, Feb. 2013.

[5] A. Lipczynska, S. Schiavon, and L. T. Graham, “Thermal comfort and self-reported productivity in an office with ceiling fans in the tropics,” *Build. Environ.*, vol. 135, pp. 202–212, May 2018.

[6] L. Pérez-Lombard, J. Ortiz, and C. Pout, “A review on buildings energy consumption information,” *Energy Build.*, vol. 40, no. 3, pp. 394–398, Jan. 2008.

[7] C. A. Balaras *et al.*, “Solar air conditioning in Europe—an overview,” *Renew. Sustain. Energy Rev.*, vol. 11, no. 2, pp. 299–314, Feb. 2007.

[8] K. F. Fong, T. T. Chow, C. K. Lee, Z. Lin, and L. S. Chan, "Comparative study of different solar cooling systems for buildings in subtropical city," *Sol. Energy*, vol. 84, no. 2, pp. 227–244, Feb. 2010.

[9] H. El-Dessouky, H. Ettouney, and A. Al-Zeefari, "Performance analysis of two-stage evaporative coolers," *Chem. Eng. J.*, vol. 102, no. 3, pp. 255–266, Sep. 2004.

[10] P. Drucker, *The practice of management*. New York, NY, USA: Harper & Brothers, 1954.

[11] Z. Wang *et al.*, "Individual difference in thermal comfort: A literature review," *Build. Environ.*, vol. 138, pp. 181–193, Jun. 2018.

[12] L. Schellen, M. G. L. C. Loomans, M. H. de Wit, B. W. Olesen, and W. D. van M. Lichtenbelt, "The influence of local effects on thermal sensation under non-uniform environmental conditions — Gender differences in thermophysiology, thermal comfort and productivity during convective and radiant cooling," *Physiol. Behav.*, vol. 107, no. 2, pp. 252–261, Sep. 2012.

[13] D. W. DeGroot and W. L. Kenney, "Impaired defense of core temperature in aged humans during mild cold stress," *Am. J. Physiol.-Regul. Integr. Comp. Physiol.*, vol. 292, no. 1, pp. R103–R108, Jan. 2007.

[14] G. T. Leites, P. L. Sehl, G. dos S. Cunha, A. Detoni Filho, and F. Meyer, "Responses of Obese and Lean Girls Exercising under Heat and Thermoneutral Conditions," *J. Pediatr.*, vol. 162, no. 5, pp. 1054–1060, May 2013.

[15] K. A. Dougherty, M. Chow, and W. L. Kenney, "Responses of Lean and Obese Boys to Repeated Summer Exercise Heat Bouts," *Med. Sci. Sports Exerc.*, vol. 41, no. 2, pp. 279–289, Feb. 2009.

[16] J. Kim, S. Schiavon, and G. Brager, "Personal comfort models – A new paradigm in thermal comfort for occupant-centric environmental control," *Build. Environ.*, vol. 132, pp. 114–124, Mar. 2018.

[17] Y. Murakami, M. Terano, K. Mizutani, M. Harada, and S. Kuno, "Field experiments on energy consumption and thermal comfort in the office environment controlled by occupants' requirements from PC terminal," *Build. Environ.*, vol. 42, no. 12, pp. 4022–4027, Dec. 2007.

[18] Jazizadeh Farrokh, Ghahramani Ali, Becerik-Gerber Burcin, Kichkaylo Tatiana, and Orosz Michael, "Human-Building Interaction Framework for Personalized Thermal Comfort-Driven Systems in Office Buildings," *J. Comput. Civ. Eng.*, vol. 28, no. 1, pp. 2–16, Jan. 2014.

[19] C. C. Aggarwal, "Outlier Analysis," in *Data Mining: The Textbook*, C. C. Aggarwal, Ed. Cham: Springer International Publishing, 2015, pp. 237–263.

[20] B. Yu, D. H. C. Van Paassen, and S. Riahy, "General modeling for model-based FDD on building HVAC system," *Simul. Pract. Theory*, vol. 9, no. 6, pp. 387–397, May 2002.

[21] J. Schein, S. T. Bushby, N. S. Castro, and J. M. House, "A rule-based fault detection method for air handling units," *Energy Build.*, vol. 38, no. 12, pp. 1485–1492, Dec. 2006.

[22] M. Kim and M. S. Kim, "Performance investigation of a variable speed vapor compression system for fault detection and diagnosis," *Int. J. Refrig.*, vol. 28, no. 4, pp. 481–488, Jun. 2005.

[23] G. A. Benndorf, D. Wystrcil, and N. Réhault, "Energy performance optimization in buildings: A review on semantic interoperability, fault detection, and predictive control," *Appl. Phys. Rev.*, vol. 5, no. 4, p. 041501, Sep. 2018.

[24] Y. Zhao, S. Wang, and F. Xiao, "Pattern recognition-based chillers fault detection method using Support Vector Data Description (SVDD)," *Appl. Energy*, vol. 112, pp. 1041–1048, Dec. 2013.

[25] J. E. Seem, "Using intelligent data analysis to detect abnormal energy consumption in buildings," *Energy Build.*, vol. 39, no. 1, pp. 52–58, Jan. 2007.

[26] A. H. Oti, E. Kurul, F. Cheung, and J. H. M. Tah, "A framework for the utilization of Building Management System data in building information models for building design and operation," *Autom. Constr.*, vol. 72, pp. 195–210, Dec. 2016.

[27] P. O. Fanger, *Thermal comfort. Analysis and applications in environmental engineering*. Danish Technical Press, 1970.

[28] G. S. Brager and R. de Dear, "Thermal adaptation in the built environment: A literature review," *Energy Build.*, vol. 27, no. 1, pp. 83–96, 1998.

[29] V. Földváry Ličina *et al.*, "Development of the ASHRAE Global Thermal Comfort Database II," *Build. Environ.*, vol. 142, pp. 502–512, Sep. 2018.

[30] de Dear, "Global database of thermal comfort field experiments," presented at the ASHRAE Transactions, 1998, vol. 104, pp. 1141–1152.