Citation for published version:
Vellei, M, Herrera, M, Fosas, D & Natarajan, S 2017, 'The influence of relative humidity on adaptive thermal comfort', Building and Environment, vol. 124, pp. 171-185. https://doi.org/10.1016/j.buildenv.2017.08.005

DOI:
10.1016/j.buildenv.2017.08.005

Publication date:
2017

Document Version
Peer reviewed version

Link to publication

The influence of relative humidity on adaptive thermal comfort

Marika Vellei[a], Manuel Herrera[a], Daniel Fosas[a], Sukumar Natarajan[a,*]

[a]*EDEn - Department of Architecture & Civil Engineering*

University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom

**corresponding author: s.natarajan@bath.ac.uk*

Abstract

Buildings generate nearly 30% of global carbon emissions, primarily due to the need to heat or cool them to meet acceptable indoor temperatures. In the last 20 years, the empirically derived adaptive model of thermal comfort has emerged as a powerful alternative to fixed set point driven design. However, current adaptive standards offer a simple linear relationship between the outdoor temperature and the indoor comfort temperature, assumed to sufficiently explain the effect of all other variables, *e.g.* relative humidity (RH) and air velocity. The lack of a signal for RH, is particularly surprising given its well-known impact on comfort. Attempts in the literature to either explain the lack of such a signal or demonstrate its existence, remain scattered, unsubstantiated and localised. In this paper we demonstrate, for the first time, that a humidity signal exists in adaptive thermal comfort using global data to form two separate lines of evidence: a meta-analysis of summary data from 63 field studies and detailed field data from 39 naturally ventilated buildings over 8 climate types. We implicate *method selection* in previous work as the likely cause of failure to detect this signal, by demonstrating that our chosen method has a 56% lower error rate. We derive a new designer-friendly RH-inclusive adaptive model that significantly extends the range of

acceptable indoor conditions for designing low-energy naturally-conditioned buildings all over the world. This is demonstrated through parametric simulations in 13 global locations, which reveal that the current model overestimates overheating by 30% compared to the new one.

Keywords: adaptive thermal comfort, naturally-conditioned buildings, relative humidity, logistic regressions, tree-based methods

Highlights

- The influence of relative humidity on adaptive thermal comfort explained.
- A new adaptive thermal comfort model which considers the effect of relative humidity introduced.
- The current model is shown to overestimate overheating by 30% over 13 global locations.

1 Introduction

According to the ANSI/ASHRAE Standard 55-2013 [1], thermal comfort is 'that condition of mind that expresses satisfaction with the thermal environment and is assessed by subjective evaluation'. Indoor thermal comfort is among the most important factors affecting occupant well-being, health and productivity in buildings [2]. This is important since people spend up to 90% of their time inside buildings, especially in developed countries [3]. However, typical buildings impose a substantive energy cost to heat or cool them to the desired comfort level. In developed countries, with largely saturated demand, this is estimated to be 20–40% of the total final energy use and nearly 30% of all CO_2 emissions [4, 5]. This makes the building sector the single largest contributor to global CO_2 production and hence climate change. Thermal comfort standards are therefore central to not merely providing comfortable environments but also ensuring a sustainable design through low heating and cooling energy use in buildings.

Two types of comfort standards currently prevail in the literature: *steady-state* and *adaptive*. The steady-state model, pioneered by P.O. Fanger in the late 1960s, is a heat-balance model that defines combinations of a set of six indoor environmental variables that will provide acceptable thermal conditions to the majority of occupants [6]. The six variables are: air temperature, mean radiant temperature, air movement, relative humidity, clothing insulation and metabolic heat generated by human activity. These are folded into an empirical relationship to provide a Predicted Mean Vote (PMV) of thermal comfort, underpinned by the idea of a *neutral temperature* for a given value of the other parameters. In contrast, the relatively recent development of the ASHRAE adaptive model [1] and its European counterpart [7] are based on the idea that the range of acceptable temperatures in naturally ventilated (NV) buildings is larger than in air-conditioned (AC) buildings and dependent *purely* on the prevailing external temperature. Using large scale survey data, such as the ASHRAE RP-884 database [8, 9], from different climatic zones around the world, these models derive a simple linear relationship between indoor comfort temperature and outdoor temperature.

According to Nicol and Humphreys [10], the reason for this extreme simplification is that some of Fanger's conventional thermal comfort factors, *i.e.* clothing insulation and metabolic rate, are significantly correlated to the outdoor air temperature. Interestingly, although relative humidity and air velocity are not shown to strongly depend on the outdoor air temperature [11], their effect is not seen to be large enough to warrant inclusion in the model [12]. However, their importance in determining physiological thermal comfort is well documented

[13]. It is known, for example, that high indoor humidity impairs sweat-induced evaporative cooling, which is the principal physiological mechanism by which the body rejects heat, particularly in warm environments [14-19]. Air movement also influences the evaporative and convective heat exchange to and from the body, affecting its temperature [13, 20].

The absence of a signal for relative humidity (RH) is surprising since outdoor humidity is likely to have a bigger effect on indoor humidity than parameters such as occupant density (which increases indoor moisture production) or window operation (which could decrease indoor humidity if external humidity is lower). This is supported by Figure 1, which shows that the Pearson correlation coefficient between mean daily indoor (RH) and outdoor ($RHout$) relative humidity in the ASHRAE RP-884 database is significantly higher in naturally ventilated (0.52) than in air-conditioned (0.33) buildings. Hence, one might expect that the comfort response in NV buildings is significantly mediated by internal relative humidity, which in turn is a function of the external humidity.

External and internal air velocities, on the other hand, are likely to be decoupled since occupant control of ventilation through window operation and use of fans is likely to have at least as great an influence on the indoor air velocity as the prevailing outdoor weather conditions. Since increased occupant control is now well established as a critical component in increasing occupant satisfaction [21], the absence of an air velocity signal could therefore be hypothesised to be due to the studied buildings having good occupant control of windows and fans [8]. However, unlike RH, the absence of recorded external wind data in the ASHRAE RP-884 database precludes a test of this hypothesis.

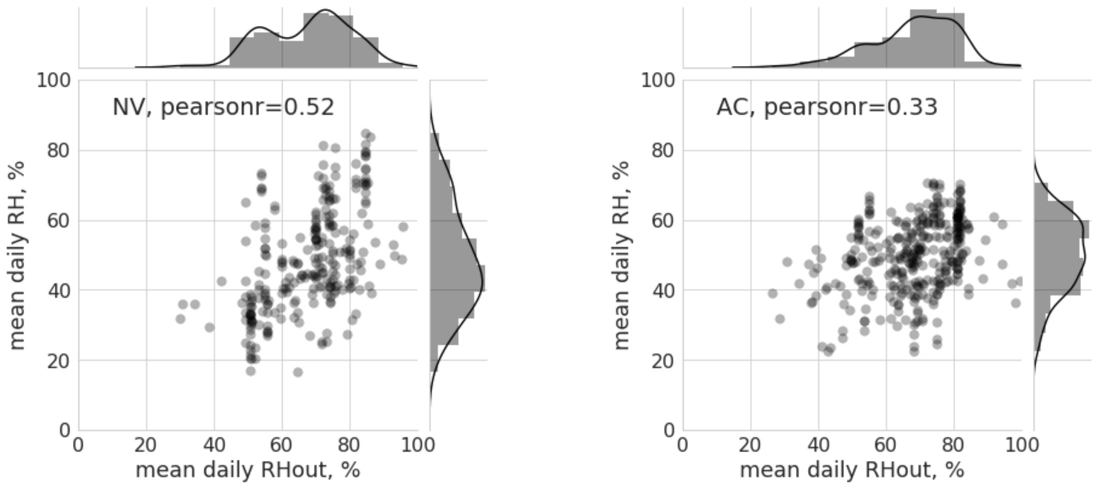

Figure 1. Scatterplot and histograms with kernel density estimates (derived using a Gaussian characteristic function) of mean daily indoor (RH) and outdoor ($RHout$) relative humidity for the ASHRAE RP-884 naturally ventilated (NV, left) and air-conditioned (AC, right) buildings. The number 'pearsonr' is the Pearson correlation coefficient.

The lack of a clear humidity signal, upon which to differentiate adaptive indoor comfort in the present models, is therefore puzzling, and the subject of much previous work in the field [12, 19, 22, 23]. However, no clear explanation for the lack of a humidity signal or a convincing formulation of the effect of humidity on adaptive thermal comfort has hereto emerged.

To address this, we begin by examining the effect of RH on occupant thermal sensitivity through an analysis of the regression gradient in Section 2. This analysis provides the first clear evidence that RH has a measurable impact on occupant thermal sensation. A second independent line of evidence emerges from the analysis in Section 3, which compares the ability of a range of statistical methods already used in the literature against new candidate methods, to explain the data contained in ASHRAE RP-884 database. Although both methods independently verify our hypothesis that RH has an important role to play in adaptive thermal comfort, neither is capable of a practical formulation that can be used by practitioners. Hence, using the knowledge gained in Sections 2 and 3, we cast the RP-884 data within a new formulation, but one that has the strength of being familiar to practitioners. This provides a new adaptive comfort model selectable by different classes of humidity. Finally, Section 5 demonstrates the use of the new model in building performance assessment across a range of global climates.

2 The effect of relative humidity on occupant thermal sensitivity

The current adaptive thermal comfort models are derived using a simple linear regression of neutral temperatures against the corresponding mean outdoor air temperatures, acquired through field studies. The neutral temperature is defined as the indoor temperature which an average occupant finds neither warm nor cool, hence *neutral* [24]. This has historically been determined using two methods:

- By regressing the Thermal Sensation Vote (TSV) against the indoor temperature, with the neutral temperature corresponding to a $TSV = 0$ [25]. Three different types of linear regression are used in the literature: simple, binned (*i.e.* binning the TSV in 0.5°C or 1°C intervals) and weighted binned, where the weights are the number of votes in each interval. The gradient of the linear regression fitted between the TSV and the indoor temperature indicates the temperature perturbation needed for a change of 1 unit in TSV. It is therefore a measure of occupant sensitivity to indoor temperature changes and gives the degree to which a population can adapt to variations in the thermal environment. Lower gradients can be associated with more effectively adapted and less sensitive occupants [26]. A lower slope is also indicative of a larger comfort band which means that occupants can tolerate exposure to a wider range of indoor temperatures [23, 25, 27].
- By using the Griffiths method. Here, the neutral temperature T_n is derived through the following equation:

$$T_n = T_m - TSV_m/G \quad (1)$$

Where TSV_m is the mean Thermal Sensation Vote, T_m is the mean indoor temperature in °C, and G is the assumed regression gradient, also called Griffiths coefficient, in /°C. This method has been used in many field studies all over the world to derive neutral temperatures [28-36], including the derivation of the European adaptive thermal comfort model [37]. This method has been deemed useful when it is difficult to reach statistically significant linear regressions, due to, for example, small sample sizes, low variance of the indoor temperature, or non-linearly dependent data with interaction effects.

Griffiths proposed a gradient equal to 0.33 to use when deriving adaptive models [38], based on Fanger's regression slope [6]. However, there is considerable variation in the actual values of G used in the literature, ranging from 0.25 to 0.50 [25, 28, 30, 39, 40]. The reasons for this vary, but are driven by the need for G to be "fit to purpose". Examples include: $G = 0.50$ to improve the coefficient of determination R^2 of the European adaptive equation [37]; and $G = 0.38$ derived from the weighted mean value of all the regression gradients included in Nguyen's database of field studies in South-East Asia, thus localising its use to hot-humid climates [23].

Nguyen showed that adaptive equations are very sensitive to changes in Griffiths constants [23] (Figure 2), thus suggesting that the choice of the right regression gradient is crucial when deriving an adaptive model.

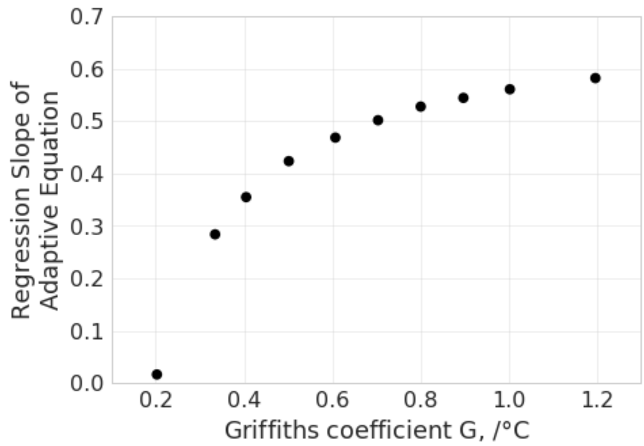

Figure 2. The relationship between the regression slope of the adaptive comfort equation and the value given to the Griffiths coefficient G; adapted from [23].

To put this in context, we reviewed earlier work on the regression gradient and found that:

- The regression gradient decreases as the standard deviation of the indoor temperature ($\sigma(T_i)$) increases, possibly indicating that larger standard deviations of the indoor temperature allow greater opportunities for behavioural and psychological adaptation [26, 41].
- Naturally ventilated buildings have lower gradients than air-conditioned buildings, again indicating greater adaptive opportunities in the former [8, 42-44].
- Occupants are more thermally sensitive to indoor temperature variations during seasonal extremes (*i.e.* summer and winter) than in the intervening milder seasons [14, 32, 45].
- Higher humidity leads to higher gradients and hence to greater occupant sensitivity to temperature variations [46].
- Higher air speed results in lower gradients in warm climates [19].
- Gradients in homes can be significantly lower than those found in offices, again likely due to the larger adaptive opportunities in terms of clothing and air speed adjustments available [8, 47, 48].

While several variables are seen to affect the regression gradient and hence thermal adaptation, the evidence is scattered or localised. For example, only one paper has shown the effect of humidity on the gradient and only based on data from two cities in India [46].

Hence, we examine this further through a meta-analysis of field studies in *naturally-conditioned* buildings. Buildings that are either naturally ventilated or mixed-mode (but operating in free-running mode during the field study) are defined as naturally-conditioned. A total of 63 field studies were thus selected, 18 of which come from the ASHRAE RP-884 database, with the remaining 45 studies from 24 papers published after the release of ASHRAE RP-884. Studies from the standardised ASHRAE RP-884 database [1, 8, 9] were filtered by selecting those achieving statistical significance ($p<0.05$) when linearly regressing TSV against the operative temperature in each study. A majority of the studies are in residential and office buildings, although other building types (educational, museum and cathedral) are present. We include all these building types in the meta-analysis without distinction. This approach is consistent with the ASHRAE standard, which is deemed applicable to all building types. A summary of the selected studies is given in the Appendix.

Our meta-analysis takes the form of a multivariate model derived from the summary statistics of the 63 selected thermal comfort field studies. The response or dependent variable in this model is the regression gradient a and the predictor variables are one or more of the available variables from the selected studies, which were:

- Indoor temperature (Ti, °C) variously measured as:
    - Dry bulb temperature (Tdb, °C),

- Globe temperature (Tg, °C), measured at the centre of a blackened globe with standard diameter of 0.15m,
- Operative temperature (Top, °C), defined as the weighted mean[1] of the air and mean radiant temperatures,

- Mean daily outdoor air temperature on the days of the survey ($Tout$, °C),
- Relative humidity (RH, %),
- Total insulation ($INSUL$, clo),
- Air velocity (VA, m/s),
- Metabolic rate of the subject (MET, met),
- Gender of the subject (SEX, 0=male/1=female).

Here, the indoor variables can be classed into two categories:

CLASS I. Binding environmental variables over which occupants have little control: indoor temperature (Ti) and humidity (RH).

CLASS II. Partially or wholly occupant-mediated variables: air velocity (VA), clothing insulation ($INSUL$) and metabolic rate (MET).

Three observations are pertinent to the selection and use of these variables in our meta-analysis:

- Summary data for CLASS II variables were not always available whereas data for CLASS I variables were available for all studies. Given that CLASS II variables, unlike those of CLASS I, can be directly controlled by the occupants of naturally-

[1] Frequently simplified as the arithmetic mean, an approximation that works well when the difference between the air and mean radiant temperatures is small.

conditioned buildings, and can hence not be viewed as pure predictors, we only consider CLASS I variables in our model.

- The reviewed field studies use different metrics for the indoor temperature, *i.e.* dry-bulb air temperature, globe temperature and operative temperature. In our model, we refer to them under the general term *indoor temperature* (Ti) since several studies have shown that differences between radiant and air temperatures in indoor environments are usually very limited [49], with exceptions in indoor spaces with high thermal mass. Since there are no buildings classed as high mass constructions in our sample, this is not a significant risk.
- Three different methods of linear regression are used in the selected studies: simple, binned and weighted binned. We treat these equally since Djamila has shown that the regression gradients calculated using either methods are very similar [50]. Details of the metrics and methods used for each field study can be found in Appendix.

Hence, the selected predictor variables for our model are the mean and standard deviation of indoor temperature and relative humidity, *i.e.* $\mu(Ti)$, $\mu(RH)$, $\sigma(Ti)$ and $\sigma(RH)$, computed over the total length of each study period. Mean and standard deviation of indoor temperature and relative humidity for all the selected studies are shown in Figure 3. Relative humidity ranges from 24% to 76%, while the temperature spans from 19°C to 35°C; providing a large spread of available mean environmental conditions. There is large variation in the standard deviations of Ti (1°C to 9°C) and RH (3% to 23%) due to the inclusion of field studies from all seasons (see also Appendix).

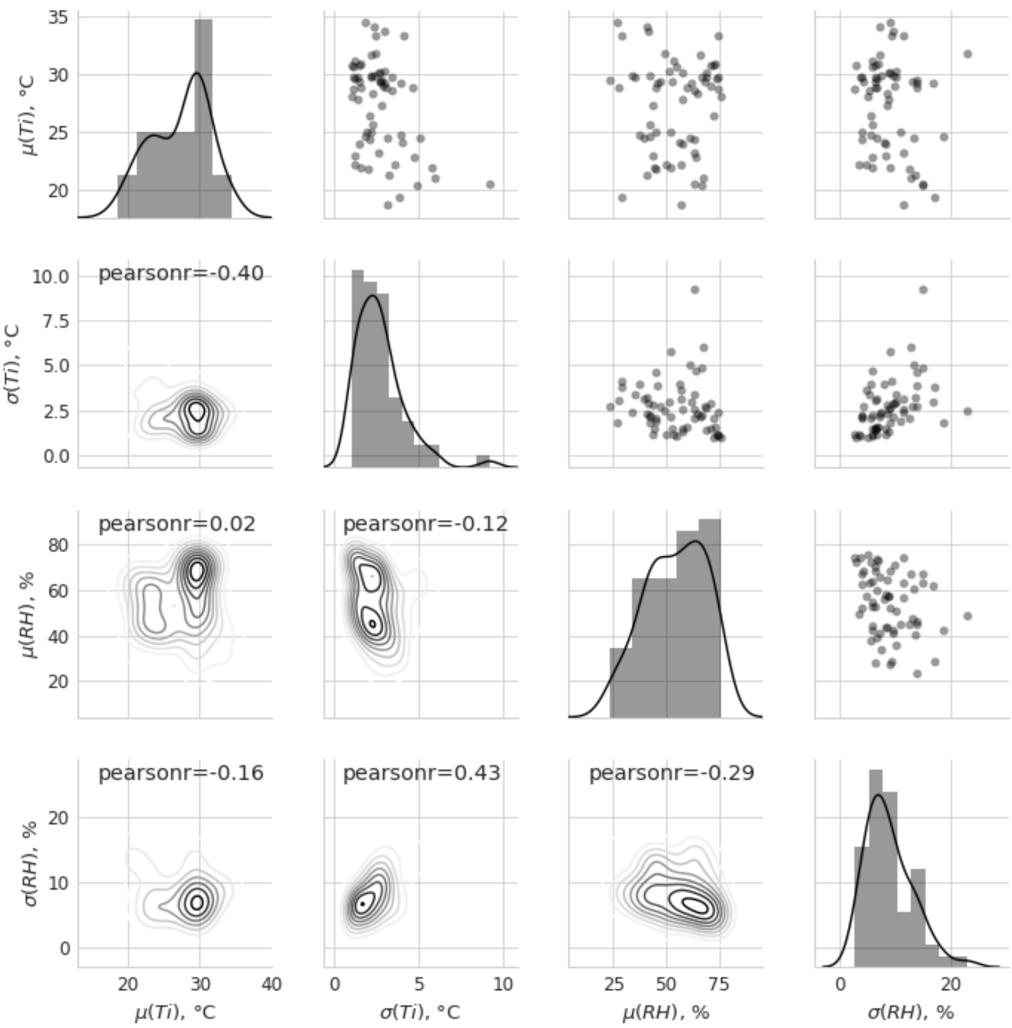

Figure 3. Scatterplot matrix with histograms and kernel density estimates (derived using a Gaussian characteristic function) in diagonal, two-dimensional kernel density plots in the lower half and bivariate scatterplots in the upper half. The number 'pearsonr' is the Pearson correlation coefficient.

2.1 New insights on the regression gradient

Linear regression can be used to describe relationships which are not inherently linear (such as exponential ones) by simply linearizing the data sets. Here, after a log transformation of the dataset, a multivariate linear regression technique is used. All the predictors are regressed collectively against the dependent variable. Then, each predictor is removed from the model to observe the effect on the coefficient of determination (R^2), in a *backward elimination* *process*[2]. If the removal of a given variable does not significantly reduce R^2 ($p<0.05$), then it is eliminated from the model. This process resulted in the rejection of $\mu(Ti)$ and $\sigma(RH)$. Hence, our model suggests that the regression gradient is dependent on $\mu(RH)$ and $\sigma(Ti)$ but independent of $\mu(Ti)$ and $\sigma(RH)$:

$$a = 0.0030 \cdot \mu(RH) + 0.7475 \cdot e^{-0.8 \cdot \sigma(Ti)} \qquad (2)$$

With N = 63, R^2=0.48, $p<0.05$

Humphreys observes that the gradient peaks at a $\sigma(T_i) = 1$ and decreases at lower values of the standard deviation, possibly due to errors in the measurements and in the equation of the operative temperature [41]. In contrast, our model suggests that the regression gradient exponentially increases at decreasing standard deviation. Significantly, a Griffiths coefficient equal to 0.50 – used to derive the European adaptive equation [37] – occurs in only 8% of the sample data.

Additionally, for the first time we observe that the gradient increases at increasing levels of RH (Figure 4). Since the acceptable operative temperature range is inversely proportional to the regression gradient, this also means that the band of acceptable temperature reduces as the RH increases.

---

[2] R^2 measures the proportion of variability in the variable response that can be explained using the predictor variables, and will always fall in the interval [0, 1]. The closer R^2 is to 1, the larger the proportion of the variability in the response variable explained by the regression and the better the model.

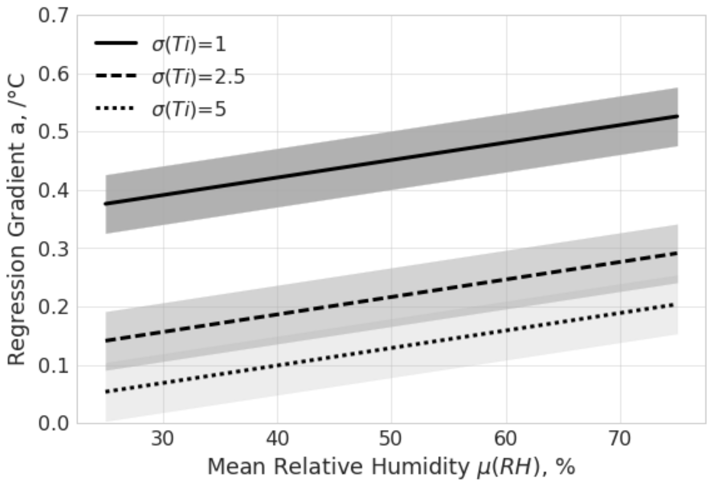

Figure 4. Regression gradient a as a function of the mean relative humidity $\mu(RH)$ for three different values of $\sigma(Ti)$; fitted model with 95% confidence bands.

Our analysis above has provided the clearest evidence thus far for the existence of a humidity signal in adaptive thermal comfort. However, this raises the question of why such a signal was not evident when the adaptive model was being created. After all, if the signal existed it must have been present in the ASHRAE RP-884 data itself, given its detail and geographical spread. One obvious reason is that the adaptive model is derived by regressing the neutral temperature in each location against the corresponding mean outdoor temperature. This process ignores the effect of the gradient, since the neutral temperature is just one point on the gradient line. A subtler reason is to do with the method: perhaps the choice of simple linear regression as a means of analysis did not provide the fidelity needed to demonstrate the presence of a humidity signal. The next section illustrates this by using a *first principles approach*, *i.e.* bringing to bear new statistical techniques that were uncommon when the adaptive model was first proposed.

3 A first principles approach

In the preceding section we were able to demonstrate the presence of a humidity signal in adaptive thermal comfort by undertaking a meta-analysis of summative descriptive statistics from a range of studies. In this section, we take a first principles approach by analysing data from the only publicly available data set that provides complete raw data for a wide range of geographically dispersed NV buildings: the ASHRAE RP-884 data set. Our working hypothesis is that these data will, in principle, be adequate to extract the RH signal (if it exists) provided a method of sufficient power is used.

3.1 Discussion of methods

A review of the literature suggests that simple and multiple linear regressions are the most widely used methods for modelling occupant thermal sensation in thermal comfort research [24]. Simple and weighted linear regressions have been extensively used for calculating occupant neutral temperatures [8, 10, 28, 43, 46, 51-58]; and starting with Bedford's first attempt in the 1930s [49], multivariate linear regression has also been largely used to study the impact of different environmental variables on occupant thermal comfort responses [19, 40, 46, 50, 59-62]. However, if we want to directly model the categorical variable TSV, a model that provides continuous estimates guarantees neither good performance nor proper validation of the linearity hypothesis. Hence, we consider five other methods that either directly improve linear regression or bring new analytical capabilities, described below.

3.1.1 Logistic regression

Logistic regression is a regression specifically designed for binary or dichotomous dependent variables [63]. The logarithm of the odds ratio, *i.e.* $ln\left(\frac{P(Y)}{1-P(Y)}\right)$, of the variable of interest (Y) is modelled based on a combination of values taken by the predictor variables.

Logistic regression can handle all sorts of relationships since it applies a nonlinear log transformation to the predicted odds ratio. Therefore, the key assumptions of linear regression and, in general, of linear models (*i.e.* normality, homoscedasticity and independence of the model residuals) do not need to be met. However, problems could still arise if multicollinearity exists (*i.e.* when two or more predictor variables are highly correlated). Issues in such a model include significant variability in the model coefficients, reducing its utility, or the suggestion of unrealistic relationships between the dependent variable and its predictors. Nonetheless, the thermal comfort literature has recognized logistic regression as a suitable alternative to simple linear regression to deal with discrete dependent variables [64, 65]. When more than one independent variable is hypothesised to affect the dependent variable, *multivariate* logistic regression is used, such as its application in the wider field of indoor environmental quality research [66, 67].

3.1.2 Multinomial logistic regression

When the dependent variable can take a value among C classes or categories with $C > 2$ (*i.e.* a *multiclass* problem), logistic regression can follow an iterative process in which the odds ratio for each category is computed by considering one category at each time and taking the set of remaining categories as a new class. However, a more natural and accurate extension to multiclass problems is done by directly considering a *multinomial* logistic regression. Like the binary logistic regression, the model now aims to approach the posterior probabilities of the C classes via linear functions in the predictors. In such a case, the model parameters are estimated by solving a set of independent binary regressions through variations in the maximum likelihood method [68]. Although not widely used in thermal comfort research, multinomial logistic regression has been used to directly model TSV as function of the indoor air temperature [69].

3.1.3 Ordinal logistic regression

However, when the dependent variable is ordinal - as is the case with TSV in thermal comfort research - *ordinal* logistic regression is needed [70]. This follows the method of multinomial regression, but takes advantage of the additional knowledge contained in the order of the categories. A common technique to undertake ordinal logistic regression is the proportional odds method which works with cumulative probabilities. This method makes the assumption that the relationship measured through the odds between one category and another is the same for any pair of categories of the dependent variable. If this assumption is not met, the straightforward solution is still multinomial logistic regression. An example is the use of ordinal logistic regression to model overall workspace satisfaction as a function of indoor environmental parameters and building characteristics [71].

3.1.4 Decision tree

The preceding three methods are variants on the fundamental idea of regression to create a mapping between dependent and independent variables. A Decision Tree (DT) model, on the other hand, is a method that creates a hierarchical tree graph based on how several independent variables partition a dependent or target variable. This partitioning reveals the strength of relationships in a dataset through the size of the split at each step. DT algorithms, of which there are many, recursively partition the data space into a number of simple regions following an optimal splitting criterion. This way of splitting the data space can be represented by a sequence of nodes and directed edges in a hierarchical structure, forming a tree. The partition algorithm starts at the *root node* of the tree, which will have no incoming edges. Starting from the root node, the data space splits into a number of regions, each one represented as a new node. The process is iterated, generating further new nodes from those previously created, each of which has exactly one incoming edge from its predecessor. Each branch of the tree finishes in a *leaf node*, which provides the category that best represents the corresponding region when the data cannot be split further.

For our analysis we use the most common DT algorithm: the C4.5 algorithm [72], which improves on the earlier ID3 algorithm. Inherent within both ID3 and C4.5 is the idea of information gain to optimise the partition process. This optimisation favours outcomes with higher information gain when undertaking the split, which leads to a division into regions of similar observations (purity per region). The information gain is measured through the difference in *entropy* before and after splitting; where the concept of entropy is related to the misclassification or impurity of a node and takes values in the range [0, 1]. If the elements of a node are equally divided into two or more categories, then the entropy is one. If all the elements in the node belong to the same category then the entropy is zero. So, the decision tree is constructed so that it minimises the entropy at the leaf nodes (ideally reaching the value of zero entropy). Given a categorisation C which divides the dataset S into categories $c_1 \ldots c_n$ and considering the proportion of observations in c_i being p_i, then the entropy of S follows equation (3).

$$Entropy(S) = \sum_{i=1}^{n} -p_i \cdot log(p_i) \tag{3}$$

3.1.5 Random Forest

A random forest (RF) is an ensemble of tree-based models. RF can be used for classification tasks when the base models are classification trees, or regression tasks when the base models are regression trees. For our analysis we use Breiman's RF algorithm, which is based on a bootstrap[3] aggregation (or *bagging*) of tree models [73]. Given the responses $Y = Y_1, \ldots, Y_m$ from the corresponding training set $X = X_1, \ldots, X_m$ a bagging tree is constructed by selecting B samples (sampling with replacement) from (X, Y) and training a decision tree for each sample. Finally, the bagging tree is computed by either averaging all the resulting single trees (if Y is continuous) or taking their majority through a process where each tree is a vote (if Y is discrete).

RFs have several advantages over DTs: they run more efficiently on large data sets, provide more accurate predictions, avoid biases often associated with single DTs, handle missing data well and provide methods for balancing error in unbalanced data sets. For these reasons, RFs have proven to be outstanding predictive models in many classification and regression tasks.

3.2 Data

A description of the RP-884 database can be found in [8, 9], together with a meta-analysis of the data forming the ASHRAE adaptive equation as included in [1]. The data itself is available to download from the University of Sydney[4].

Since the database has been standardized by De dear and Brager allowing consistency of measured and calculated parameters, all the metrics (*e.g.* clothing insulation, operative temperature and metabolic rate) are used as presented in the database. For the analysis, we reduce the seven categories in the standard ASHRAE TSV scale to the following three classes (see Table 1):

- votes in the range of [-3, -1) considered as *cold*,

[3] Bootstrapping is sub-sampling (with replacement) of a sample to infer the characteristic features of the population from which the sample is drawn, but which are fundamentally unknowable. In other words, the sample is treated as if it were the population and the sub-samples are used to measure the quality of inference about the sample (and hence the population), given that the true features of the sample itself are known.
[4] http://sydney.edu.au/architecture/staff/homepage/richard_de_dear/ashrae_rp-884.shtml

- votes in the three central categories, *i.e.* in the range of [-1, 1], regarded as *neutral/comfortable* per the usual definition of thermal comfort [1],
- votes in the range of (1, 3] considered as *hot*.

The reduction to a 3-point scale is supported by the common use of the scale whereby excursions beyond the +1 and -1 limits are considered uncomfortable [69]. The use of three categories instead of seven also has the benefit of improving the explanatory power of the statistical models used, by increasing the number of data points in each group on either side of +1 and -1.

Table 1. The seven-point ASHRAE scale of thermal comfort (top) converted into a simplified scale of thermal comfort (bottom) for the analysis.

How are you feeling right now?

Cold	*Cool*	*Slightly cool*	*Neutral*	*Slightly warm*	*Warm*	*Hot*
-3	-2	-1	0	1	2	3

How are you feeling right now?

Cold	*Neutral*	*Hot*
-1	0	1

A key distinction in method between that used to derive the ASHRAE adaptive model [8, 9] and ours, is the unit of analysis. While the ASHRAE model is derived by aggregating data at the building level, we directly use the raw data from the database and hence operate at the level of an individual occupant. While the building level was considered appropriate due to the similarities between the building contextual factors affecting subjective responses (such as availability and accessibility of personal control, view and connection to the outdoors, interior design, occupancy patterns and social constraints [74]), this approach has the drawback of losing a great quantity of information in the process of aggregation. By using the raw data, we are able to use new techniques to investigate the effect of various predictors on the categorical response variable TSV.

For our analysis, we begin by including all the variables described in Section 2, except dry bulb and globe temperatures whose effect is contained within the operative temperature[5]. Since the adaptive model only applies to NV buildings with adult occupants, only data from these buildings are selected. The ASHRAE RP-884 database provides data from a total of 39 NV buildings over 8 climates (wet equatorial, humid subtropical, temperature marine, Mediterranean, tropical savanna, west coast marine, hot arid desert, semi-arid mid and high altitude). We further restrict the data to only outdoor temperatures within the ASHRAE applicability limits of 10 and 33.5°C, obtaining a total of 9,546 rows of observations. Since the data are already clean and ready to use, the only modification needed is to eliminate 1,289 rows of missing data (14% of the sample) resulting in a total of 8,257 rows available for analysis.

3.2.1 Variable selection

Feature or variable selection is the process of selecting a subset of relevant features/variables from the data [75], in order to:

- improve the interpretability of the data,
- reduce the effect of noise or collinearity,
- increase the predictive ability of the consequent statistical model,
- perform a computationally efficient data analysis.

[5] The operative temperature is defined as the arithmetic mean of the air and mean radiant temperatures in the ASHRAE database.

We perform a correlation analysis to eliminate highly correlated variables from further analysis. Figure 5 shows the correlation matrix for the 7 predictor variables selected. This confirms the results of Brager and de Dear [8] and Nicol & Humphreys [37], where clothing insulation is shown to be strongly inversely correlated with outdoor temperature. As expected in naturally ventilated buildings, Top is strongly correlated with $Tout$. While SEX, MET, VA and RH are not found to be strongly correlated to $Tout$. Hence, we continue our analysis with the following independent variables: Top, RH, VA, MET, SEX and we further include $Tout$ to retain the main assumption of the adaptive hypothesis.

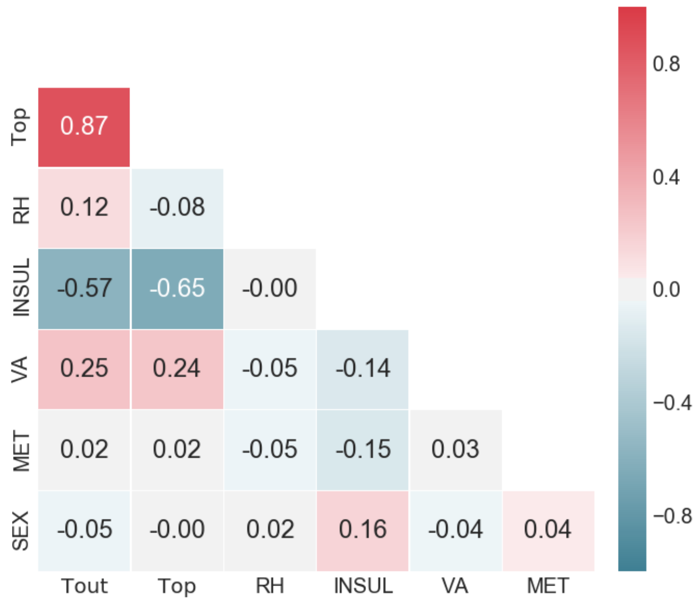

Figure 5. Correlation matrix for the selected variables. Each cell shows the Pearson coefficient, colour coded according to the strength of positive (red) and negative (blue) correlation.

3.3 Experimental study

In this section we report the results of the experiments carried out using the models discussed in Section 3.1 on the ASHRAE RP-884 data.

3.3.1 Model comparison

The aim of the experiments is to find the model that best describes the RP-884 data, *i.e.* the model with the smallest prediction error. We use the Python programming language as a convenient vehicle for comparing the ability of the five models, discussed in Section 3.1. The independent variables are those selected in Section 3.2.1, while the dependent variable to be modelled is TSV as defined in Table 1. Fifty stratified randomized sets of training and test data are created using the function *sklearn.model_selection.StratifiedShuffleSplit()* [76]. By using this function, the test sets preserve the percentage of samples for each class, *i.e.* the test and train sets have the same percentage of data in each of the three classes (*cold*, *neutral*, *hot*). The proportion of the dataset included in the test split is always 20% of the original sample. Model predictions coming from the training data are compared with the test data. Prediction errors for each model and for each set of training/test data are calculated using the F1 score implemented by the Python function *sklearn.metrics.f1_score()* [76]. F1 is a weighted average of the precision and recall scores, reaching its best value at 1 and worst one at 0. Prediction errors are defined as $1 - F1$.

Figure 6 shows a boxplot of the prediction errors associated with the 50 randomized sets of train and test data for each of the 5 models studied, *i.e.* each boxplot contains 50 error scores. Results fall into three clear groups: the logistic and multinomial logistic have the largest errors (mean equal to 0.42 and 0.40, respectively); the random forests classifier has the lowest error (mean error = 0.20); and the ordinal logistic and decision tree are in the middle (mean equal to 0.27 and 0.26, respectively). It is noteworthy that the mean error rate of the RF classifier is 56% lower than that of the simple multivariate logistic regression, the best in class method used in the thermal comfort literature so far.

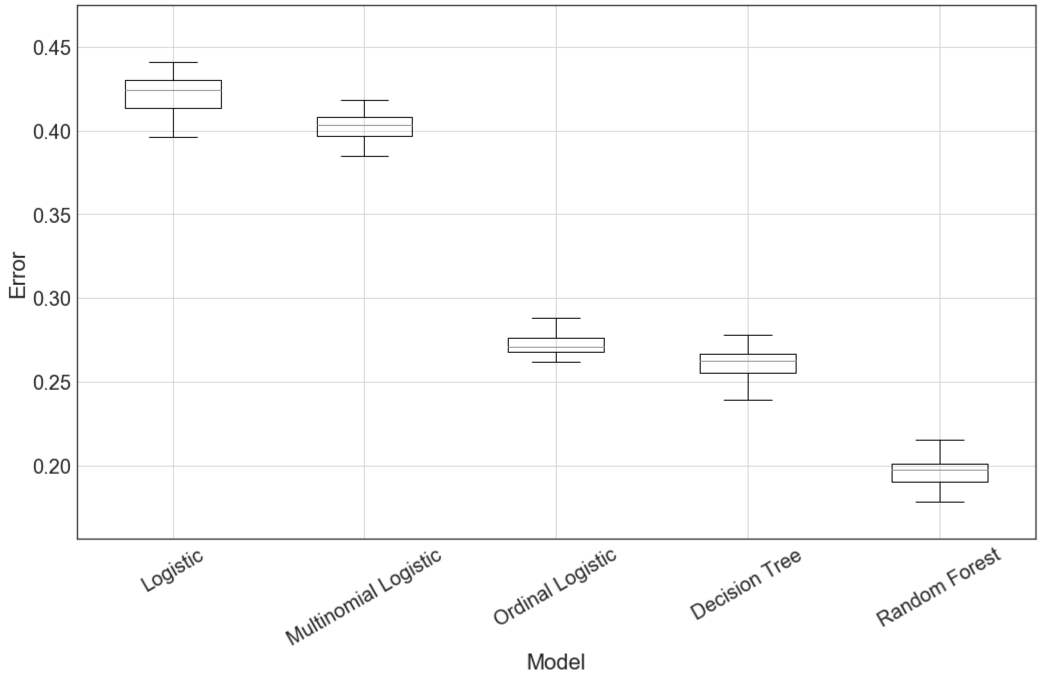

Figure 6. Boxplot of the prediction errors (1 – F1) associated with the different models. The box extends from the lower to upper quartile values of the data, with a line at the median. The whiskers extend from the box to show the range of the data.

3.3.2 Variable importance

Having identified the RF classifier as the method with the least error, Figure 7 shows the relative importance of each studied variable as classified by the RF. This confirms the prevailing adaptive model by demonstrating that Top and $Tout$ are the most influential variables with importance scores of 37% and 23%, respectively.

It also shows that RH follows $Tout$ with an importance score of 14%. This is suggestive of a weaker signal in determining thermal comfort compared to indoor temperature. It is therefore unsurprising that current methods such as multivariate logistic regression were unable to detect such a signal, given their considerably higher error rate in describing the data set.

Interestingly SEX is shown to not be significantly influential in our ranking. Given that VA and MET are factors that can be controlled by the occupants in NV buildings (see Section 1), we reduce the main predictor variables to the following three: Top, $Tout$ and RH.

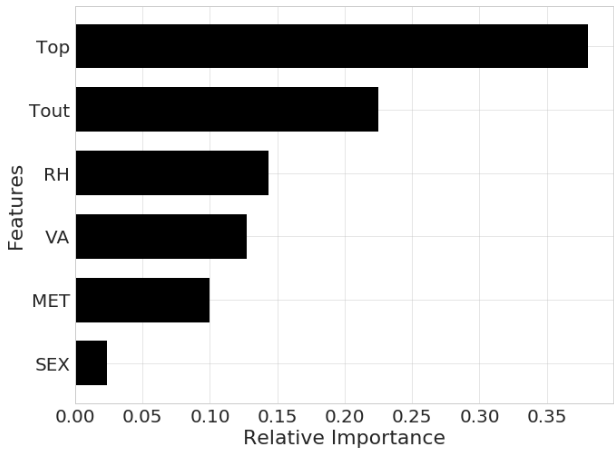

Figure 7. Relative importance of features as given by the RF classifier.

4 A new adaptive thermal comfort model

Sections 2 and 3 provide strong evidence that a humidity signal exists in adaptive thermal comfort. However, neither provides a clear route towards a practical formulation that can be easily interpreted and applied during the design of buildings. Hence, in this section, we derive a new adaptive model that frames the impact of RH within the familiar linear form of the current adaptive model.

To derive our new adaptive model, we use the ASHRAE RP-884 data and consider only neutral votes (as defined in our simplified scale in Table 1). To simplify the continuous nature of the humidity data, we cluster the neutral votes using the widely used *k*-means clustering (as implemented in the Python function *sklearn.cluster.KMeans()* [77]). The *k*-means algorithm clusters data by trying to minimize the distance between data belonging to the same cluster while maximizing the distance between data belonging to different clusters. This leads to a clustering configuration of minimum variance within groups and maximum variance between different groups. This algorithm requires the specification of the number of clusters, which was set to 3. The algorithm was run 10 times, each with different random starting conditions to obtain the clusters. The *k*-means algorithm returns the following 3 clusters:

- *High*: $RH \geq 59\%$
- *Medium*: $37\% < RH < 59\%$
- *Low*: $RH \leq 37\%$

This clustering accords well with *Sterling's criteria* for human exposure to humidity in occupied buildings, which suggests that the optimal conditions to minimize risks to human health occur in the narrow range between 40-60% relative humidity [78]. Hence, the middle range in our clustering is the functional equivalent of Sterling's "Optimum Zone", and the low and high ranges correspond to the non-optimal zones. To improve model readability, we simplify the clusters to convert the middle cluster to the range of 40-60%.

Within each RH cluster, we collate all the TSV votes into *1°C $Tout$ x 1°C Top grid bins*. In order to meet the 80% acceptability criterion incorporated in the current model, we reject any bin with less than 80% of neutral votes, *i.e.* votes falling into the three central categories of the 7-point ASHRAE scale. Finally, we compute mean Top and mean $Tout$ for each grid bin. Now, by applying a simple linear regression to each cluster of RH, three linear models are obtained:

$$TOP_{RH>60\%} = 0.53 \cdot TOUT + 12.85\ (\pm 2.84), with\ R^2 = 0.84\ and\ p < 0.000 \quad (4)$$

$$TOP_{40\% < RH \leq 60\%} = 0.53 \cdot TOUT + 14.16\ (\pm 3.70), with\ R^2 = 0.76\ and\ p < 0.000 \quad (5)$$

$$TOP_{RH \leq 40\%} \quad = 0.52 \cdot TOUT + 15.23\ (\pm 4.40), with\ R^2 = 0.66\ and\ p < 0.000 \qquad (6)$$

The temperature bands in the above equations are given by the prediction intervals. Here, we define a prediction interval as one in which future observations are likely to fall with 0.95 probability. Figure 8 shows the temperature bands for the 3 clusters together with the ASHRAE adaptive model in red.

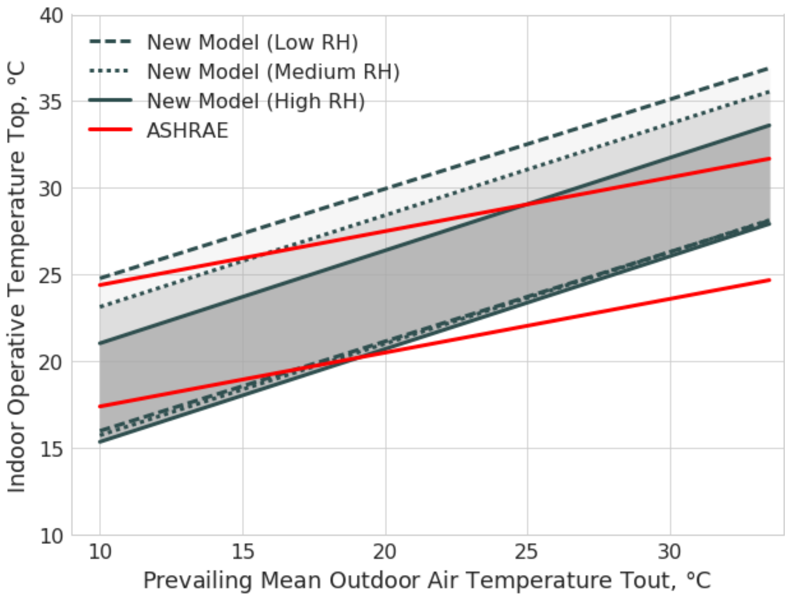

Figure 8. The proposed new, and existing, adaptive models.

The following major outcomes can be observed from the model in Figure 8:

- Comfort temperatures are generally higher and the gradient is much steeper, than those predicted by the current ASHRAE adaptive model.
- Comfort temperatures are lower when humidity is high throughout the range of $Tout$. The difference in comfort temperatures between high and low humidity environments is as high as 4°C.

- The smallest temperature acceptability range corresponds to a high relative humidity, while the acceptability range for a medium humidity is equal to the acceptability range defined in the ASHRAE adaptive model.

It is important to note at this point that while this formulation follows from the relatively simple process of regression, it relies on the evidence uncovered from the RF process demonstrated in Section 3.3 as well as the earlier analysis of thermal sensitivity in Section 2. Without these, the separation by RH would be arbitrary and meaningless. As a corollary, these independent lines of evidence preclude the creation of further "adaptive models" by the application of the method in this section to any of the other variables such as air velocity or gender, even if such models were deemed to be meaningful, without further new evidence.

5 Measuring the impact of the new adaptive comfort model

Section 4 derives a new adaptive comfort model that relates thermal comfort to not just outdoor temperature but also indoor relative humidity. This section considers the potential impact of designing naturally ventilated buildings using this new model by comparing it against the current model. The chosen building type is office since the vast majority of ASHRAE data comes from offices: 57% of the studies are in offices with a further 36% in both (offices + residential) buildings. The chosen performance metric is the widely used "count of overheating hours", measured as the percentage of occupied hours above the maximum operative temperature threshold when using a given comfort model. Overheating is measured by implementing our new thermal comfort model within the well-established EnergyPlus (v8.7) simulation software and applying it to a building simulation case study, together with the current model. An implementation of the new adaptive comfort model is available via the public Python package *vellei_acm*.

The implemented building model represents a NV office based on the Department of Energy reference models for the U.S. [79]. The following adaptations were made to make it suitable for this study:

- Unlike the reference building, the office is set to be naturally ventilated and in free-running mode exclusively. This is needed to allow the application of adaptive models as specified in the ANSI/ASHRAE Standard 55-2013 [1]. To enable this change in operating mode, the following additional changes were made:
    - The original span of the building has been adapted from ≈18m to 12m.
    - Two ventilation schemes were modelled to account for the two most common natural ventilation modes: double sided cross-ventilation and single sided ventilation. For the former, all internal partitions are removed. For the latter, a single partition runs along the length of the building to provide a 6m ventilated depth.
- The model is considered to be located at an intermediate level within a multi-floor office block. Both the ceiling and the floor have been considered adiabatic and no energy transfers are allowed except for heat storage.
- Surrounding buildings are considered at a 20m distance with the same height as the zone under consideration.

Natural ventilation is modelled with an airflow network. Rather than simpler and more traditional methods, airflow networks allow the approximation of pressure-driven air exchanges with the outdoor environment or another zone by modelling the underlying physical laws in greater detail, accounting for wind and stack effects, bidirectional air flows in large openings and cross-ventilation among other phenomena. Windows are sized to a 20%

window-to-wall ratio and the total openable area for natural ventilation is equal to 5% of the total floor area of the office. To compare comfort models, meta-programming of the simulation behaviour through the Energy Management System (EMS) functionality in EnergyPlus was implemented, as follows:

- Windows are opened if the following three conditions are met simultaneously: the zone is occupied, the neutrality temperature is surpassed and the external temperature is below the zone temperature. All temperatures are evaluated as operative temperatures.
- Both comfort models are implemented with two variants (*i.e.* there are a total of 4 variants). The variants are based on the interpretation of outdoor temperature in the models as evidenced in extant practice and ASHRAE recommendations. One variant uses the monthly mean outdoor temperature ('original') and the other an exponentially weighted running mean with $\alpha = 0.8$ ('running mean').

A number of simulation model variants are produced using a scripted *building generator* to cover a wide range of scenarios. These include:

- 13 of the 14 locations where the ASHRAE RP-884 NV buildings were surveyed (a weather file for Saidu in Pakistan could not be obtained),
- 4 different orientations (N/S, E/W, SE/NW and SW/NE - the building is symmetrical),
- 3 levels of shading (low, medium and high, *i.e.* 0, 0.5 and 1 times the required depth to shade the opening at noon during the summer solstice),
- and 3 window openable areas (3.5%, 5% and 6.5% of the office floor area).

Together with the different control algorithms based on the 2 adaptive models with the 2 formulations of the outdoor mean temperature, these result in a total of 1,872 model variants for each ventilation scheme (i.e. a total of 3,744 variants).

5.1 Results

Figure 9 shows a summary of results from the simulations. It is clear that the new model produces considerably lower overheating than the current model and that there is little difference in whether monthly mean or running mean outdoor temperature is used in computing either adaptive model.

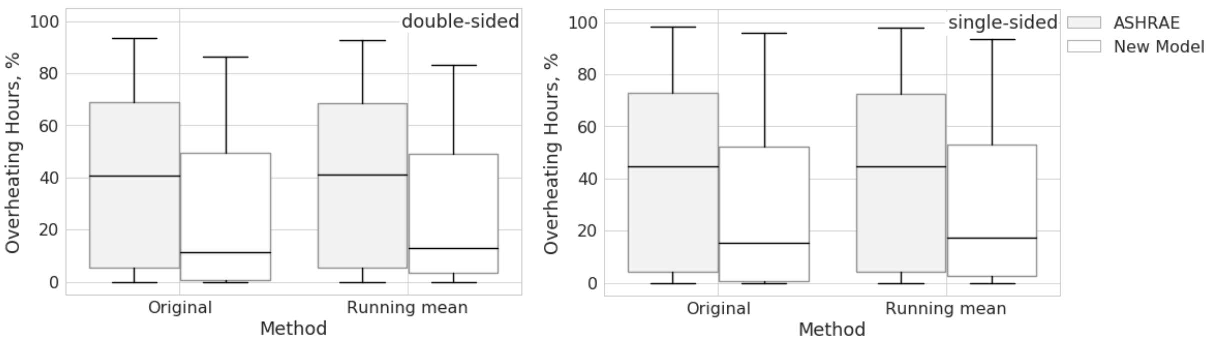

Figure 9. A comparison of overheating hours between the current ASHRAE model and the new model proposed in this paper for double-sided (left) and single-sided (right) offices. Each box-and-whisker plot represents data from 468 variants based on differing location, orientation, shading and window openable areas.

Figure 10 shows that the high levels of overheating observed in Figure 9 are primarily a function of the large diversity of climates represented in the data set. An interesting feature of Figure 10 is that the largest differences between the proposed new model and the current model are observed in climates with low humidity (*e.g.* Quetta, Karachi and Peshawar). Finally, the warmer the climate, the lower the predicted overheating in the new model compared to the current model. In other words, the new model significantly extends the potential range of operation for buildings in all climates, with the most in the warmest and least humid climates.

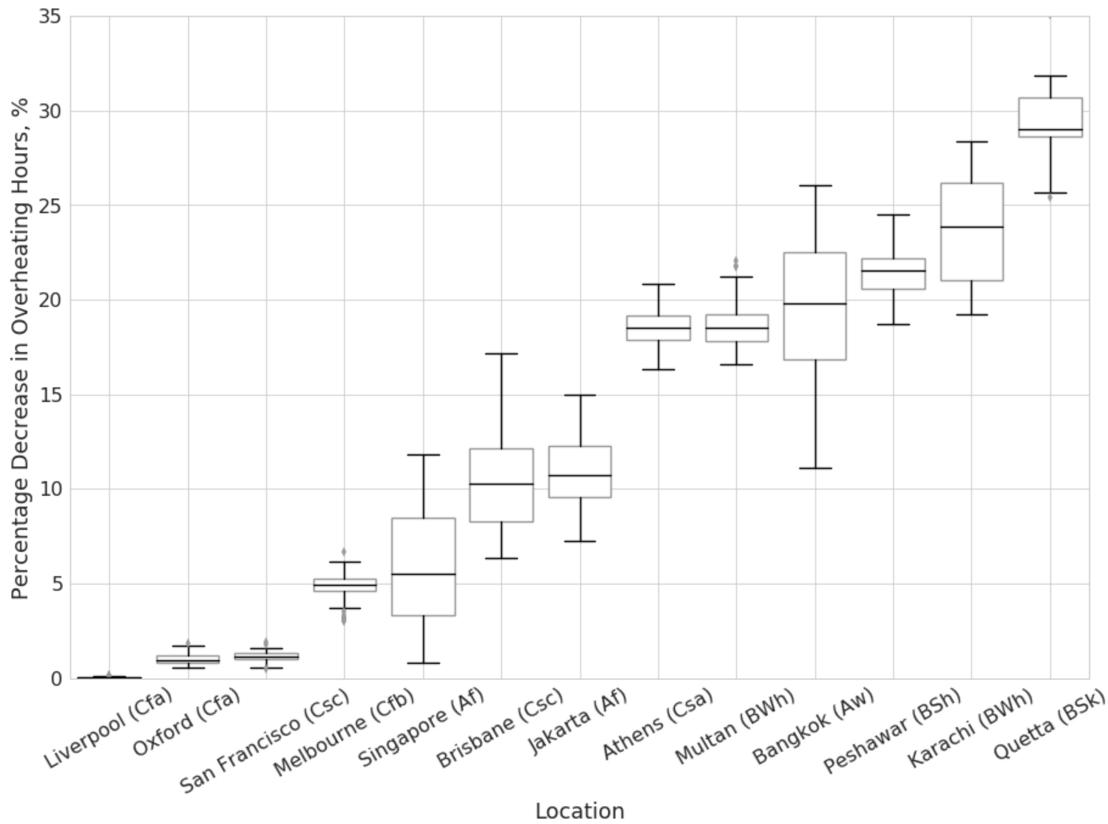

Figure 10. Rank-ordered percentage decrease in overheating hours when using the new adaptive model proposed in this paper in place of the current ASHRAE model. Letters within brackets show the Köppen Geiger climate classification [80] for each location. The box-and-whisker plots represents data from a total of 3,744 simulations.

6 Discussion

Previous attempts to characterise the impact of humidity on the adaptive thermal comfort equation have found limited or no evidence of a change in comfort at varying levels of humidity. These are summarised below:

- A study clustering mean outdoor RH from ASHRAE [8] and other [81] field data into low (<63%), medium (64-75%) and high (>75%) found that neutral temperatures, obtained using Griffiths method, were only about 1°C lower for RH>75% compared to the overall data [12].

- A study deriving an adaptive comfort model for the hot-humid regions of South-East Asia found a similar comfort equation for NV buildings as the ASHRAE adaptive equation [23].
- Another study using the ASHRAE field data found that the regression coefficients of the adaptive equations for hot-humid (0.57) and hot-dry climates (0.58) were nearly double that of the ASHRAE model (0.31), and slightly lower for moderate climates (0.22) [22]. This study did not observe lower comfort limits at higher relative humidity for hot-humid climates. However, hot-dry climates were found to have larger comfortable temperature bands than hot-humid climates. The authors suggest this could be because it is easier to adapt when humidity is low, supported by their observation that in hot-dry climates, a higher indoor RH implies lower comfort temperatures. These results are very interesting and anticipate some of our results, although they do not offer the comprehensive explanation which we provide with our model.

One of the principal reasons suggested in the literature for the lack of a humidity signal in adaptive comfort models is that occupants in humid climates are usually well adapted to high humidity. The use of fans, opening of windows for increasing air movement, and wearing clothing that enhances evaporation of sweat have all been suggested as adaptive actions common in hot and humid climates [19, 82-85].

In contrast, our new model shows that the impact of relative humidly cannot be neglected. This is supported by two independent lines of evidence both of which demonstrate that humidity plays a significant role in mediating adaptive thermal comfort. Although it is possible that the effect of humidity is mitigated by several adaptive actions, it is important to consider that, unlike air velocity, it cannot be directly controlled in NV buildings, as demonstrated in Figure 1. Hence, it is essential that the effect of humidity is explicitly incorporated within the design of such buildings.

7 Conclusions

Adaptive thermal comfort has been a breaking new paradigm which has changed the way of looking at thermal comfort in NV buildings. However, the model has remained essentially the same for the last 20 years and its simplicity, which was its initial strength, now poses some concerns. We highlight the principal concern as the lack of a signal for relative humidity. From a meta-analysis of the regression gradient using summative statistics from a large number of global studies, we demonstrate, for the first time, the clear importance of relative humidity in determining the sensitivity of occupants within the adaptive comfort paradigm. We produce a second, independent, line of evidence using a random forests process on high-resolution thermal comfort data from buildings across the world that strongly supports this initial finding. Finally, we use these data to derive a new adaptive model which incorporates relative humidity in three clusters, obtained via a *k*-means clustering of humidity conditions found within the data. Since the new model is formulated using the familiar linear relationship that designers are already accustomed to, it can be readily used for the design of low-energy naturally ventilated buildings around the world. We demonstrate the use of the new model for the design of a naturally ventilated building in each location from which the empirical data was sourced. Results show that our new model significantly increases the comfort envelope of naturally ventilated buildings since its prediction of overheating is 30% lower than that of the current model. Hence, the use of our model *significantly extends* the current natural adaptive comfort boundary.

Acknowledgments

Daniel Fosas and Marika Vellei appreciate the support of the EPSRC ‘dCarb’ centre (EP/L016869/1). Daniel Fosas is also funded by ‘laCaixa’ Foundation.

Appendix

Table 2, Table 3 and Table 4 show key information about the studies included in the meta-analysis of Section 2.

Table 2. Main information for the field studies, included in the ASHRAE database, surveying naturally ventilated buildings free-running in summer.

ASHRAE Database	Köppen Climate	Location	Building Type	Survey Type	Survey Class	Sample Size
12	Csc	Brisbane, Australia	O	T	II	652
16	Cfb	Melbourne, Australia	O	T	II	582
27	Csa	Athens, Greece	R	L	II	1626
4	Aw	Bangkok, Thailand	O	T	II	392
33	Csc	San Francisco,	O	L and T	I	360
28	Cfa	Oxford, UK	O	L	III	877
18	BWh	Karachi, Pakistan	R and O	L	III	190
23	BSk	Quetta, Pakistan	R and O	L	III	492
20	BWh	Multan, Pakistan	R and O	L	III	437

21	BSh	Peshawar, Pakistan	R and O	L	III	556
25	Cfa	Saidu, Pakistan	R and O	L	III	568
7	Af	Jakarta, Indonesia	O	T	III	97
38	Cfa	Liverpool, UK	O	T	II	167
42	Af	Singapore	O	T	II	583

R=Residential, O=Office, T=Transverse, L=Longitudinal

Table 3. Main information for the newly reviewed field studies included in the meta-analysis of the regression gradient. An empty space means that the information is not available.

Reference	Köppen Climate/	Location	Building Type	Survey Type	Survey Class	Sample Size	Building Operation
Feriadi & Wong, 2004	Am/ dry and rainy	Jogjakarta, Indonesia	R	L	II	525	NV
Karyono, 2008 [87]	Af/ rainy	Bandung, Indonesia	E	L	III	200	MM/free-running
Karyono et al., 2015 [88]	Am/ rainy	Jakarta, Indonesia	Cathedral Museum	T	III	70 77	NV
Ogbonna & Harris, 2008	Aw/ rainy	Jos, Nigeria	R and E	L	II	200	NV
Moujalled et al., 2008 [56]	Cfb/ summer	Lyon, France	O	T	II	221	NV
Yang & Zhang, 2008 [43]	Cfa/ summer	Nanjing, Shanghai,	R and O	L	II	129	NV
	BWh/	Greater Cairo, Egypt	E and O	L		644	NV

Farghal & Wagner, 2010 [52]	autumn and spring					638	NV
						656	NV
						751	NV
Djamila et al., 2013 [50]	Af/ all seasons	Kota Kinabalu, Malaysia	R	L	II	890	NV
Indraganti et al., 2013 [46]	As/ dry and rainy	Chennai, India	O	T	II	207	MM/free-running mode
	BSh/ dry and rainy	Hyderabad, India				352	
Indraganti et al., 2013 [89]	Cfa/ summer	Tokyo, Japan	O	T	II	423	MM/free-running
Gomez-Azpeitia et al., 2014 [53]	BWh/ summer	Hermosillo, Mexico	R	T	II	143	NV
	BWh/ summer	Mexicali, Mexico				174	
	Aw/ dry	Merida, Mexico				150	
	Aw/ dry	Colima, Mexico				196	
Rijal, 2014 [40]	Cfa/ summer	Kanto region, Japan	R	T	III	1915	MM/free-running
Luo et al., 2015 [44]	Cwa/ all seasons	Shenzhen, China	O	T	III	513	MM/free-running
Mustapa et al., 2016 [28]	Cfa/ summer	Fukuoka, Japan	O	T	III	81	MM/free-running
Rijal et al., 2017 [90]	Cfa/ all seasons	Tokyo and Yokohama,	O	T	II	422	MM/free-running
Yan et al., 2017 [91]	Cfa/ summer	Nanjing, Shanghai and	R	L	II		NV

	Dwa/ summer Dfa/ summer Cwa/ summer	Harbin, Changchun and Beijing, Xi'an and Zhengzhou, Guangzhou, Nanning and					
Liu et al., 2017 [32]	Cfa/ spring Cfa/ summer Cfa/ autumn	Chongqing, Chengdu, Wuhan, Nanjing, Hangzhou and Changsha,	R	L	II	2965 2521 3385	NV
Bouden & Ghrab, 2005 [92]	Csa/ all seasons Csa/ all seasons BSh/ all seasons BWh/ all seasons BWh/ all seasons	Kef, Tunisia Tunis, Tunisia Sfax, Tunisia Gabes, Tunisia Gafsa, Tunisia	R and O	T	II		NV
Rijal et al., 2010 [48]	sub-tropical, temperate and	Banke, Bhaktapur,	R	L	III	2180	NV
Dhaka et al., 2015 [57]	BSh/ winter BSh/ moderate season BSh/ summer and	Jaipur, India	R and O		II	610 346 855	NV
Indraganti, 2010 [54]	BSh/ summer	Hyderabad, India	R	T	II	1405	NV

	BSh/ monsoon					1334	
	BSh/ monsoon					1223	
Lachireddi et al., 2017 [93]	Am/ dry	Calicut, India	R	T	III	735	NV
Mishra & Ramgopal, 2014 [94]	Aw/ dry	Kharagpur, India	E	L	III	338	NV
Mishra & Ramgopal, 2015 [35]	Aw/ dry	Kharagpur, India	E	L	III	533	NV

R=Residential, O=Office, E=Educational, T=Transverse, L=Longitudinal, MM=Mixed-Mode

Table 4. Main data used in the meta-analysis of the regression gradient. An empty space means that the information is not available.

Reference	**Indoor Temperature**	**μ(Ti)**	**σ(Ti)**	**μ(RH)**	**σ(RH)**	**Linear Regression**	**a**	**b**	R^2
Feriadi & Wong, 2004	Top	29.8	1.4	68.6	6.6	Simple	0.59	-17.21	0.18
Karyono, 2008 [87]	Top	28.9	1.5	59.8	6.8	Simple	0.31	-7.97	0.68
Karyono et al., 2015 [88]	Tdb	28.8	1.1	74.3	2.8	Simple	1.05	-29.02	0.90
		29.7	1.1	74.1	3.8		0.68	-18.90	0.56
Ogbonna & Harris, 2008	Top	26.5	2.1	72.1	5.6	Weighted Binned	0.36	-9.43	0.32
Moujalled et al., 2008 [56]	Top	27.3	2.8	43.5	8.5	Weighted Binned	0.21	-4.93	0.82

Yang & Zhang, 2008	Top	33.3	2.4	74.0	11.6	Simple	0.25	-7.16	0.47
Farghal & Wagner, 2010 [52]	Tdb	25.6	2.3	42.0	6.1	Simple	0.17	-4.17	0.19
		29.8	3.4	35.5	10.1		0.24	-5.67	0.41
		25.0	2.1	52.0	4.1		0.20	-4.73	0.16
		24.7	3.9	37.5	5.7		0.17	-3.63	0.36
Djamila et al., 2013 [50]	Tdb	30.7	1.5	70.7	6.4	Simple	0.39	-11.87	0.17
Indraganti et al., 2013 [46]	Tg	30.1	2.6	57.2	8.8	Simple	0.31	-8.17	0.29
		29.4	2.7	47.2	13		0.22	-5.68	0.17
Indraganti et al., 2013 [89]	Tg	29.4	1.5	52.6	6.4	Simple	0.31	-7.95	0.36
Gomez-Azpeitia et al., 2014 [53]	Tdb	33.8	2.93	41.3	9.8	Simple	0.18	-4.90	
		33.4	4.1	28.5	9.4		0.13	-3.31	0.23
		34.07	2.3	41.0	7.3		0.17	-3.77	
		29.93	2.1	42.2	9.5		0.29	-7.51	
Rijal, 2014 [40]	Air	28.4	2.3	64.4	8.5	Simple	0.19	-4.81	0.14
Luo et al., 2015 [44]	Top	23.2	2.57	63.1	11.6	Weighted Binned	0.09	-1.97	
Mustapa et al., 2016 [28]	Top	28.1	1	75.9	5.1	Simple	0.49	-13.1	0.21

Rijal et al., 2017 [90]	Tg	25	1.9	45	11	Simple	0.18	-4.6	0.25
Yan et al., 2017 [91]	Top	29.4	2.7	68.3	4.4	Binned	0.36	-9.80	0.96
		24.4	2.1	62.8	4		0.19	-4.93	0.87
		28.6	3.4	63.1	5.3		0.24	-6.55	0.89
		29.7	2.2	66.8	4.2		0.13	-3.68	0.77
Liu et al., 2017 [32]	Tdb	20.4	4.9	66.9	14.9	Binned	0.06	-1.2	0.95
		29.0	2.9	70.8	9.9		0.16	-3.76	0.93
		21.1	6.0	67.1	12.9		0.06	-1.52	0.97
Bouden & Ghrab, 2005 [92]	Tg	20.5	9.2	63	15	Simple	0.17	-3.62	0.84
		22.2	3.6	57	5		0.16	-3.18	0.50
		22.8	4.7	64	6		0.16	-3.27	0.57
		24.1	4	56	8		0.11	-2.31	0.29
		21.9	5.8	52	9		0.17	-3.60	0.74
Rijal et al., 2010 [48]	Tg	24.5	5.1	60.7	13.4	Simple	0.08	-1.95	0.83
Dhaka et al., 2015 [57]	Tdb	21.3	3.2	40.6	13.5	Simple	0.17	-4.39	0.20
		28.9	3.1	27.7	6.4		0.14	-3.89	0.11

		31.8	2.5	49.1	23.0		0.30	-8.79	0.38
Indraganti, 2010 [54]	Tg	34.5	1.8	27.0	9.0	Simple	0.22	-5.93	0.42
		31.2	1.2	53.0	6.0		0.28	-8.30	0.40
		30.7	1.1	55.0	6.0		0.17	-4.67	0.25
Lachireddi et al., 2017 [93]	Top	31.7	2.2	65.7	7.2	Binned	0.56	-16.95	0.90
Mishra & Ramgopal, 2014 [94]	Top	28.9	4.6	44.7	13.9	Weighted Binned	0.18	-4.77	0.86
Mishra & Ramgopal, 2015 [35]	Top	29.3	3.0	61.9	16.9	Weighted Binned	0.22	-6.50	0.73

References

1. ASHRAE, *ANSI/ASHRAE Standard 55-2013, Thermal environmental conditions for human occupancy*. 2013: Atlanta, Ga.
2. Frontczak, M. and P. Wargocki, *Literature survey on how different factors influence human comfort in indoor environments.* Building and Environment, 2011. **46**(4): p. 922-937.
3. Klepeis, N.E., et al., *The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants.* J Expo Anal Environ Epidemiol, 2001. **11**(3): p. 231-52.
4. Pérez-Lombard, L., J. Ortiz, and C. Pout, *A review on buildings energy consumption information.* Energy and Buildings, 2008. **40**(3): p. 394-398.
5. Nejat, P., et al., *A global review of energy consumption, CO2 emissions and policy in the residential sector (with an overview of the top ten CO2 emitting countries).* Renewable and Sustainable Energy Reviews, 2015. **43**: p. 843-862.
6. Fanger, P.O., *Thermal Comfort*. 1970, Copenhagen: Danish Technical Press.
7. EN-15251, *Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics*. 2007.
8. de Dear, R.J. and G.S. Brager. *Developing an adaptive model of thermal comfort and preference*. in *ASHRAE Transactions*. 1998.
9. de Dear, R.J. and G.S. Brager, *Thermal comfort in naturally ventilated buildings: revisions to ASHRAE Standard 55.* Energy and Buildings, 2002. **34**(6): p. 549-561.
10. Nicol, J.F. and M.A. Humphreys, *Adaptive thermal comfort and sustainable thermal standards for buildings.* Energy and Buildings, 2002. **34**(6): p. 563-572.

11. Halawa, E. and J. van Hoof, *The adaptive approach to thermal comfort: A critical overview.* Energy and Buildings, 2012. **51**(0): p. 101-110.
12. Nicol, F., *Adaptive thermal comfort standards in the hot–humid tropics.* Energy and Buildings, 2004. **36**(7): p. 628-637.
13. Parsons, K., *Human Thermal Environments: The Effects of Hot, Moderate, and Cold Environments on Human Health, Comfort and Performance, Second Edition*. 2002: Taylor & Francis.
14. Song, X., et al., *Analysis on Human Adaptive Levels in Different Kinds of Indoor Thermal Environment.* Procedia Engineering, 2015. **121**: p. 151-157.
15. Jing, S., et al., *Impact of Relative Humidity on Thermal Comfort in a Warm Environment.* Indoor and Built Environment, 2013. **22**(4): p. 598-607.
16. Andersen, I., G.R. Lundqvist, and D.F. Proctor, *Human Perception, of Humidity Under Four Controlled Conditions.* Archives of Environmental Health: An International Journal, 1973. **26**(1): p. 22-27.
17. Zhang, Y., et al., *Effects of step changes of temperature and humidity on human responses of people in hot-humid area of China.* Building and Environment, 2014. **80**: p. 174-183.
18. Jin, L., Y. Zhang, and Z. Zhang, *Human responses to high humidity in elevated temperatures for people in hot-humid climates.* Building and Environment, 2017. **114**: p. 257-266.
19. Givoni, B., et al., *Thermal sensation responses in hot, humid climates: Effects of humidity.* Building Research and Information, 2006. **34**(5): p. 496-506.
20. Cândido, C., et al., *Air movement acceptability limits and thermal comfort in Brazil's hot humid climate zone.* Building and Environment, 2010. **45**(1): p. 222-229.
21. Boerstra, A., et al., *Impact of available and perceived control on comfort and health in European offices.* Architectural Science Review, 2013. **56**(1): p. 30-41.
22. Toe, D.H.C. and T. Kubota, *Development of an adaptive thermal comfort equation for naturally ventilated buildings in hot–humid climates using ASHRAE RP-884 database.* Frontiers of Architectural Research, 2013. **2**(3): p. 278-291.
23. Nguyen, A.T., M.K. Singh, and S. Reiter, *An adaptive thermal comfort model for hot humid South-East Asia.* Building and Environment, 2012. **56**: p. 291-300.
24. Mishra, A.K. and M. Ramgopal, *Field studies on human thermal comfort — An overview.* Building and Environment, 2013. **64**: p. 94-106.
25. Humphreys, M.A., H.B. Rijal, and J.F. Nicol, *Updating the adaptive relation between climate and comfort indoors; new insights and an extended database.* Building and Environment, 2013. **63**: p. 40-55.
26. de Dear, R., et al., *Adaptive thermal comfort in Australian school classrooms.* Building Research & Information, 2015. **43**(3): p. 383-398.
27. Indraganti, M., *Using the adaptive model of thermal comfort for obtaining indoor neutral temperature: Findings from a field study in Hyderabad, India.* Building and Environment, 2010. **45**(3): p. 519-536.
28. Mustapa, M.S., et al., *Thermal comfort and occupant adaptive behaviour in Japanese university buildings with free running and cooling mode offices during summer.* Building and Environment, 2016. **105**: p. 332-342.
29. Thapa, S., A.K. Bansal, and G.K. Panda, *Adaptive thermal comfort in the two college campuses of Salesian College, Darjeeling – Effect of difference in altitude.* Building and Environment.
30. Damiati, S.A., et al., *Field study on adaptive thermal comfort in office buildings in Malaysia, Indonesia, Singapore, and Japan during hot and humid season.* Building and Environment, 2016. **109**: p. 208-223.

31. Yang, D., J. Xiong, and W. Liu, *Adjustments of the adaptive thermal comfort model based on the running mean outdoor temperature for Chinese people: A case study in Changsha China.* Building and Environment, 2017. **114**: p. 357-365.
32. Liu, H., et al., *Seasonal variation of thermal sensations in residential buildings in the Hot Summer and Cold Winter zone of China.* Energy and Buildings, 2017. **140**: p. 9-18.
33. Liu, W., et al., *Feedback from human adaptive behavior to neutral temperature in naturally ventilated buildings: Physical and psychological paths.* Building and Environment, 2013. **67**(0): p. 240-249.
34. Mishra, A.K. and M. Ramgopal, *An adaptive thermal comfort model for the tropical climatic regions of India (Köppen climate type A).* Building and Environment, 2015. **85**: p. 134-143.
35. Mishra, A.K. and M. Ramgopal, *A thermal comfort field study of naturally ventilated classrooms in Kharagpur, India.* Building and Environment, 2015. **92**: p. 396-406.
36. Olweny, M.R.O., L.L. Mugagga, and T. Nedala, *A study of thermal comfort and thermal preferences in the upland tropical climate of Uganda*, in *Proceedings of 9th Windsor Conference: Making Comfort Relevant*. 2016: Cumberland Lodge, Windsor, UK.
37. Nicol, F. and M. Humphreys, *Derivation of the adaptive equations for thermal comfort in free-running buildings in European standard EN15251.* Building and Environment, 2010. **45**(1): p. 11-17.
38. GRIFFITHS, I.D., *Thermal comfort in buildings with passive solar features : field studies.* The Publications Office of the European Union 1991. **EUR 13431 EN**.
39. Nicol, F., G.N. Jamy, and O. Sykes, *A Survey of Thermal Comfort in Pakistan: Toward New Indoor Temperature Standards; Final Report*. 1994: School of Architecture, Oxford Brookes Univ.
40. Rijal, H., *Investigation of Comfort Temperature and Occupant Behavior in Japanese Houses during the Hot and Humid Season.* Buildings, 2014. **4**(3): p. 437.
41. Humphreys, M.A., J.F. Nicol, and I.A. Raja, *Field Studies of Indoor Thermal Comfort and the Progress of the Adaptive Approach.* Advances in Building Energy Research, 2007. **1**(1): p. 55-88.
42. Manu, S., et al., *Field studies of thermal comfort across multiple climate zones for the subcontinent: India Model for Adaptive Comfort (IMAC).* Building and Environment, 2016. **98**: p. 55-70.
43. Yang, W. and G. Zhang, *Thermal comfort in naturally ventilated and air-conditioned buildings in humid subtropical climate zone in China.* International Journal of Biometeorology, 2008. **52**(5): p. 385-398.
44. Luo, M., et al., *Evaluating thermal comfort in mixed-mode buildings: A field study in a subtropical climate.* Building and Environment, 2015. **88**: p. 46-54.
45. Dhaka, S., et al., *Evaluation of thermal environmental conditions and thermal perception at naturally ventilated hostels of undergraduate students in composite climate.* Building and Environment, 2013. **66**(0): p. 42-53.
46. Indraganti, M., R. Ooka, and H.B. Rijal, *Field investigation of comfort temperature in Indian office buildings: A case of Chennai and Hyderabad.* Building and Environment, 2013. **65**: p. 195-214.
47. Ye, X.J., et al., *Field study of a thermal environment and adaptive model in Shanghai.* Indoor Air, 2006. **16**(4): p. 320-326.
48. Rijal, H.B., H. Yoshida, and N. Umemiya, *Seasonal and regional differences in neutral temperatures in Nepalese traditional vernacular houses.* Building and Environment, 2010. **45**(12): p. 2743-2753.
49. Nicol, F., M. Humphreys, and S. Roaf, *Adaptive thermal comfort: principles and practice*. 2012.

50. Djamila, H., C.-M. Chu, and S. Kumaresan, *Field study of thermal comfort in residential buildings in the equatorial hot-humid climate of Malaysia.* Building and Environment, 2013. **62**: p. 133-142.
51. Luo, M., et al., *Can personal control influence human thermal comfort? A field study in residential buildings in China in winter.* Energy and Buildings, 2014. **72**(0): p. 411-418.
52. Farghal, A. and A. Wagner. *Studying the adaptive comfort model a case study in a hot dry climate, Cairo, Egypt.* . in *Adapting to Change: New Thinking on Comfort*. 2010. Windsor, UK.
53. Gomez-Azpeitia, L.G., et al., *Extreme Adaptation to Extreme Environments in Hot Dry, Hot Sub-humid and Hot Humid Climates in Mexico.* Journal of Civil Engineering and Architecture, 2014. **8**(8): p. 929-942.
54. Indraganti, M., *Thermal comfort in naturally ventilated apartments in summer: Findings from a field study in Hyderabad, India.* Applied Energy, 2010. **87**(3): p. 866-883.
55. Ogbonna, A.C. and D.J. Harris, *Thermal comfort in sub-Saharan Africa: Field study report in Jos-Nigeria.* Applied Energy, 2008. **85**(1): p. 1-11.
56. Moujalled, B., R. Cantin, and G. Guarracino, *Comparison of thermal comfort algorithms in naturally ventilated office buildings.* Energy and Buildings, 2008. **40**(12): p. 2215-2223.
57. Dhaka, S., et al., *Assessment of thermal environmental conditions and quantification of thermal adaptation in naturally ventilated buildings in composite climate of India.* Building and Environment, 2015. **86**(0): p. 17-28.
58. Karyono, T.H., et al., *Report on Thermal Comfort Study in Bandung, Indonesia*. 2005.
59. Sharma, M.R. and S. Ali, *Tropical summer index—a study of thermal comfort of Indian subjects.* Building and Environment, 1986. **21**(1): p. 11-24.
60. Pellegrino, M., M. Simonetti, and L. Fournier. *A field survey in Calcutta. Architectural issues, thermal comfort and adaptive mechanisms in hot humid climates*. in *Proceedings of 7th Windsor Conference: The Changing Context of Comfort in an Unpredictable World*. 2012.
61. Rangsiraksa, P. *Thermal comfort in Bangkok residential buildings, Thailand*. in *PLEA 2006 - 23rd International Conference on Passive and Low Energy Architecture, Conference Proceedings*. 2006.
62. Erlandson, T., et al., *Environmental and human factors influencing thermal comfort of office occupants in hot - Humid and hot - Arid climates.* Ergonomics, 2003. **46**(6): p. 616-628.
63. Hilbe, J.M., *Logistic Regression Models*. 2009: Taylor & Francis.
64. Webb, C.G., *An Analysis of Some Observations of Thermal Comfort in an Equatorial Climate.* British Journal of Industrial Medicine, 1959. **16**(4): p. 297-310.
65. Indraganti, M., R. Ooka, and H.B. Rijal, *Thermal comfort in offices in India: Behavioral adaptation and the effect of age and gender.* Energy and Buildings, 2015. **103**: p. 284-295.
66. Wong, L.T., K.W. Mui, and P.S. Hui, *A multivariate-logistic model for acceptance of indoor environmental quality (IEQ) in offices.* Building and Environment, 2008. **43**(1): p. 1-6.
67. Lai, A.C.K., et al., *An evaluation model for indoor environmental quality (IEQ) acceptance in residential buildings.* Energy and Buildings, 2009. **41**(9): p. 930-936.
68. Engel, J., *Polytomous logistic regression.* Statistica Neerlandica, 1988. **42**(4): p. 233-252.
69. Haldi, F. and D. Robinson, *On the unification of thermal perception and adaptive actions.* Building and Environment, 2010. **45**(11): p. 2440-2457.

70. McCullagh, P., *Regression Models for Ordinal Data.* Journal of the Royal Statistical Society. Series B (Methodological), 1980. **42**(2): p. 109-142.
71. Frontczak, M., et al., *Quantitative relationships between occupant satisfaction and satisfaction aspects of indoor environmental quality and building design.* Indoor Air, 2012. **22**(2): p. 119-131.
72. Quinlan, J.R., *Induction of decision trees.* Machine Learning, 1986. **1**(1): p. 81-106.
73. Breiman, L., *Random Forests.* Machine Learning, 2001. **45**(1): p. 5-32.
74. O'Brien, W. and H.B. Gunay, *The contextual factors contributing to occupants' adaptive comfort behaviors in offices – A review and proposed modeling framework.* Building and Environment, 2014. **77**(0): p. 77-87.
75. Hastie, T., R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. 2009: Springer New York.
76. Pedregosa, F., et al., *Scikit-learn: Machine learning in Python.* Journal of Machine Learning Research, 2011. **12**(Oct): p. 2825-2830.
77. Sculley, D., *Web-scale k-means clustering*, in *Proceedings of the 19th international conference on World wide web*. 2010, ACM: Raleigh, North Carolina, USA. p. 1177-1178.
78. Sterling, E.M., A. Arundel, and T.D. Sterling. *CRITERIA FOR HUMAN EXPOSURE TO HUMIDITY IN OCCUPIED BUILDINGS*. in *ASHRAE Transactions*. 1985.
79. Deru, M., Field, K., Studer, D., Benne, K., Griffith, B., Torcellini, P., Liu, B., Halverson, M., Winiarski, D., Rosenberg, M., , *U.S. Department of Energy Commercial Reference Building Models of the National Building Stock* 2011.
80. Kottek, M., Grieser, J. , Beck, C., Rudolf, B. and Rubel, F., *World Map of the Köppen-Geiger climate classification updated.* Meteorologische Zeitschrift, 2006. **15**(3): p. 259-263.
81. Humphreys, M.A., B.R. Establishment, and B.R. Station, *Field Studies of Thermal Comfort Compared and Applied*. 1975: Building Research Station.
82. deDear, R.J., K.G. Leow, and A. Ameen, *Thermal comfort in the humid tropics—part 1: climate chamber experiments on temperature preferences in Singapore. part 2: climate chamber experiments on thermal acceptability in Singapore.* ASHRAE Transactions 2004, 1991. **97**: p. 874–879, 880–886.
83. Ballantyne, E.R., R.K. Hill, and J.W. Spencer, *Probit analysis of thermal sensation assessments.* International Journal of Biometeorology, 1977. **21**(1): p. 29-43.
84. Rijal, H., M. Humphreys, and F. Nicol, *Adaptive Thermal Comfort in Japanese Houses during the Summer Season: Behavioral Adaptation and the Effect of Humidity.* Buildings, 2015. **5**(3): p. 1037.
85. Chow, T.T., et al., *Thermal sensation of Hong Kong people with increased air speed, temperature and humidity in air-conditioned environment.* Building and Environment, 2010. **45**(10): p. 2177-2183.
86. Feriadi, H. and N.H. Wong, *Thermal comfort for naturally ventilated houses in Indonesia.* Energy and Buildings, 2004. **36**(7): p. 614-626.
87. Karyono, T.H., *Bandung Thermal Comfort Study: Assessing the Applicability of an Adaptive Model in Indonesia.* Architectural Science Review, 2008. **51**(1): p. 60-65.
88. Karyono, T.H., et al., *Thermal comfort studies in naturally ventilated buildings in Jakarta, Indonesia.* Buildings, 2015. **5**(3): p. 917-932.
89. Indraganti, M., R. Ooka, and H.B. Rijal, *Thermal comfort in offices in summer: Findings from a field study under the 'setsuden' conditions in Tokyo, Japan.* Building and Environment, 2013. **61**: p. 114-132.
90. Rijal, H.B., M.A. Humphreys, and J.F. Nicol, *Towards an adaptive model for thermal comfort in Japanese offices.* Building Research & Information, 2017: p. 1-13.

91. Yan, H., Y. Mao, and L. Yang, *Thermal adaptive models in the residential buildings in different climate zones of Eastern China.* Energy and Buildings, 2017. **141**: p. 28-38.
92. Bouden, C. and N. Ghrab, *An adaptive thermal comfort model for the Tunisian context: a field study results.* Energy and Buildings, 2005. **37**(9): p. 952-963.
93. LACHIREDDI, G.K.K., P. MUTHUKUMAR, and S. SUBUDHI, *Thermal comfort analysis of hostels in National Institute of Technology Calicut, India.* Sādhanā, 2017. **42**(1): p. 63-73.
94. Mishra, A.K. and M. Ramgopal, *Thermal comfort field study in undergraduate laboratories - An analysis of occupant perceptions.* Building and Environment, 2014. **76**: p. 62-72.