Title

Causal Thinking: Uncovering Hidden Assumptions and Interpretations ofStatistical Analysis in Building Science

Authors

Sun, Ruiji
Schiavon, Stefano
Brager, Gail
et al.

Publication Date

2024-06-01

Data Availability

The data associated with this publication are available upon request.

Peer reviewed

Causal Thinking: Uncovering Hidden Assumptions and Interpretations of Statistical Analysis in Building Science

Ruiji Sun[1*], Stefano Schiavon[1], Gail Brager[1], Edward Arens[1], Hui Zhang[1], Thomas Parkinson[2], Chenlu Zhang[3]

[1] Center for the Built Environment, University of California, Berkeley, USA
[2] Indoor Environmental Quality Lab, The University of Sydney, Australia
[3] AI Product Development Group, Johnson Controls International, USA
[*] *Corresponding email: ruijis@berkeley.edu*
[*] *Corresponding postal address:390 Wurster Hall, Berkeley, CA 94720-1839, USA*

ABSTRACT

Causal thinking emphasizes the understanding of asymmetric causal relationships between variables, requiring us to specify which variable is the cause (independent variable) and which is the effect (dependent variable). Reversing the causal relationship direction can lead to profoundly different assumptions and interpretations. We demonstrate this by comparing two linear regression approaches used in thermal comfort research: Approach (a), which regresses thermal sensation votes (y-axis) on indoor temperature (x-axis); Approach (b), which does the reverse, regressing indoor temperature (y-axis) on thermal sensation votes (x-axis). From a correlational perspective, they may appear interchangeable, but causal thinking reveals substantial and practical differences between them. Approach (a) represents occupants' thermal sensations as responses to indoor temperature. In contrast, Approach (b), rooted in adaptive comfort theory, suggests that thermal sensations can trigger behavioral changes, which in turn alter indoor temperature. Using the same data, we found that two approaches lead to different neutral temperatures and comfort zones. Approach (b) leads to what we call a 'preferred zone', which is 10 °C narrower than the conventionally derived comfort zone using Approach (a). We hypothesize that the 'preferred zone' might be interpreted as thermal conditions that occupants are likely to choose when they have significant control over their personal and environmental thermal settings. This finding has important implications for occupant comfort and building energy efficiency. We highlight the importance of integrating causal thinking into correlation-based statistical methods, which have been prevalent in building science research, especially given the increasing volume of data in the built environment.

KEYWORDS

Causal inference; Regression analysis; Neutral temperature; Comfort zone; Adaptive thermal comfort; Occupant behavior.

1. INTRODUCTION

Causality describes a directional relationship where making interventions on one entity can lead to changes in another (Imbens & Rubin, 2015). The ability to discern causality is foundational to human intelligence (Pearl & Mackenzie, 2018). For instance, humans understand that the Sun rising causes the rooster to crow, not the other way around, whereas artificial intelligence may only capture an association between the two events, even when trained on millions of data points (Pearl, 2019). Causality is pivotal in scientific inquiry (Pearl & Mackenzie, 2018; Imbens & Rubin, 2015). A scientific inquiry might ask, 'Why does the sun rise in the east and set in the west?' Aristotle proposed a geocentric model, explaining that the sunrise occurs because the Sun orbits the Earth. Later, Copernicus, Galileo, and Kepler introduced the heliocentric model, challenging Aristotle's causal explanation. Over time, with accumulated scientific evidence, we have come to understand that the sunrise is a result of the Earth's rotation on its axis. This pursuit of a deeper and more accurate understanding of causality, known as causal inference, underpins the science- and technology-based civilization in which we live today.

Building scientists have long pursued causal inference studies. Houghten and Yagloglou initially considered temperature, humidity, and air motion as three principal causal factors of human thermal comfort. They proposed an index known as the effective temperature that integrates dry-bulb air temperature and relative humidity (Houghten & Yagloglou, 1923). Subsequently, Bedford expanded this index to include radiant heat and air movement factors (Bedford, 1948). This line of research led to the development of comprehensive thermal comfort models (Gonzalez et al., 1974), such as the Predicted Mean Vote (PMV) model (Fanger, 1970). We consider the PMV model a causal model because all model inputs are causal factors of thermal comfort, and the model development is based on laboratory experiments, where different thermal environments contribute to occupants' comfort variations. This causal model allows us to find combinations of thermal conditions that provide similar comfort but use more energy-efficient systems (Kent et al., 2023; Miller et al., 2021; Schiavon & Melikov, 2008). However, the PMV model has demonstrated limitations in real-world applications (Cheung et al., 2019). Persistent discrepancies between model predictions and actual field study observations suggest there are other potential factors influencing thermal comfort (de Dear et al., 2020; Parkinson et al., 2020; van Hoof, 2008; Humphreys & Nicol, 2002).

Researchers are identifying the additional factors influencing thermal comfort in real-world settings through field studies beyond the investigation in climate chambers through laboratory experiments. Potential influencing factors include outdoor climate, building cooling systems, and aspects related to occupants, such as demographics, expectations, and behaviors (Parkinson et al., 2020; Wang et al., 2020; van Hoof et al., 2017; Fanger & Toftum, 2002; Humphreys & Nicol, 2002; Brager & de Dear, 1998; de Dear & Brager, 1998). While the correlation-based methods commonly used in field studies can reveal associations of interests, they fall short of establishing cause-and-effect relationships. For example, the adaptive comfort model demonstrates a correlation between outdoor conditions and indoor comfort temperature (de Dear & Brager, 1998; M. A. Humphreys, 1978). The model does not explicitly assert a causal relationship between the two variables. However, the applications and interpretations of the adaptive comfort model in building design and simulation imply a causal direction. For naturally ventilated buildings, comfortable indoor temperature ranges are determined by prevailing outdoor air temperatures, but not vice versa. It is not feasible to use comfortable indoor temperatures to predict or determine prevailing outdoor temperatures.

Establishing causality has been recognized as a challenge. In statistics, inferring causation was deemed unscientific and largely discouraged (Ding, 2023; Pearl & Mackenzie, 2018; Pearl, 2009; Wasserman, 2004). As the well-known statistical mantra says, correlation does not imply causation. For instance, while there may be a statistically significant correlation between ice cream sales and shark attacks (given that both increase during the summer), it would be erroneous to infer that one causes the other. Unfortunately, building scientists sometimes rely on correlation-based statistical analysis to frame their questions and interpret results, which may lead to a misalignment with their true interests in understanding causal mechanisms and estimating causal effects. For example, the correlation between the distance from a window and occupant satisfaction leads to an interpretation that the former directly influences the latter (Frontczak et al., 2012), whereas in reality there may be many confounding factors. In this paper, we will focus on the example of neutral temperature, where the true interest lies in finding a temperature that leads to a neutral thermal sensation (Nicol & Humphreys, 2010). However, given the complexity of real buildings, the conventional statistical methods and collected field study data do not allow us to fully understand all the causal relationships or accurately estimate the causal effects of indoor temperature on thermal sensations. Building scientists would benefit from recent advancements in causal inference frameworks that have revolutionized numerous scientific disciplines, ranging from quantum physics to social science (Costa & Shrapnel, 2016; Gelman & Vehtari, 2021; Greenland et al., 1999; Imbens & Rubin, 2015; Pearl, 2009; Rohrer, 2018; Rothman & Greenland, 2005).

In this paper, we focus on causal thinking, which is a preliminary step in causal inference, a rigorous process of identifying causality using data. Causal thinking is a mindset that emphasizes understanding asymmetric causal relationships. It involves categorizing variables not merely as associated with each other, but specifically as causes and effects. We highlight the essential role of causal thinking in building science and advocate for a shift in recent data-centric research, encouraging more rigorous data collection methods that enable analysis to move from a correlational to a causal perspective. To illustrate the importance of causal thinking, we compare the mathematical and theoretical aspects of two regression approaches used for neutral temperature estimation in thermal comfort field studies. The case study results reveal substantial differences between these approaches, with practical implications for future field study methods that could improve our understanding of both building energy efficiency and occupant comfort.

2. BACKGROUND

Neutral temperature is commonly defined from a correlational perspective as the indoor temperature associated with a neutral thermal sensation (Wang et al., 2020; Yan et al., 2017; Cao et al., 2014; Rijal et al., 2010; Schiller et al., 1988). The thermal sensation is evaluated using "right-now and right-here" surveys of Thermal Sensation Votes (TSV) based on the ASHRAE seven-point scale, which ranges from -3 (feeling cold) to 0 (neutral) to +3 (feeling hot). The indoor temperature is a proxy of the indoor thermal environment. It can be simply represented and measured as air temperature. We choose a more comprehensive metric called Standard Effective Temperature (SET), which integrates both physical and personal factors of air temperature, radiative temperature, humidity, air movement, clothing, and metabolic rate (ASHRAE 55, 2023). These variables are all measured simultaneously with TSV. Throughout this paper, the term 'indoor temperature' will sometimes refer to this SET equivalent temperature, representing the indoor thermal condition experienced and, in some cases, partially created by a building occupant.

The neutral temperature is typically estimated through linear regression analysis between building occupants' TSV and corresponding simultaneous indoor temperature. Existing literature presents two approaches for conducting linear regression analysis: ***Approach (a)*** regresses TSV on indoor temperature (de Dear & Brager, 1998; Nicol & Humphreys, 1973), while ***Approach (b)*** regresses indoor temperature on TSV (Wang et al., 2020; Parkinson et al., 2020; Oseland, 1994). In either case, the resulting linear regression model between TSV and indoor temperature is used to calculate the neutral temperature by setting the TSV to zero.

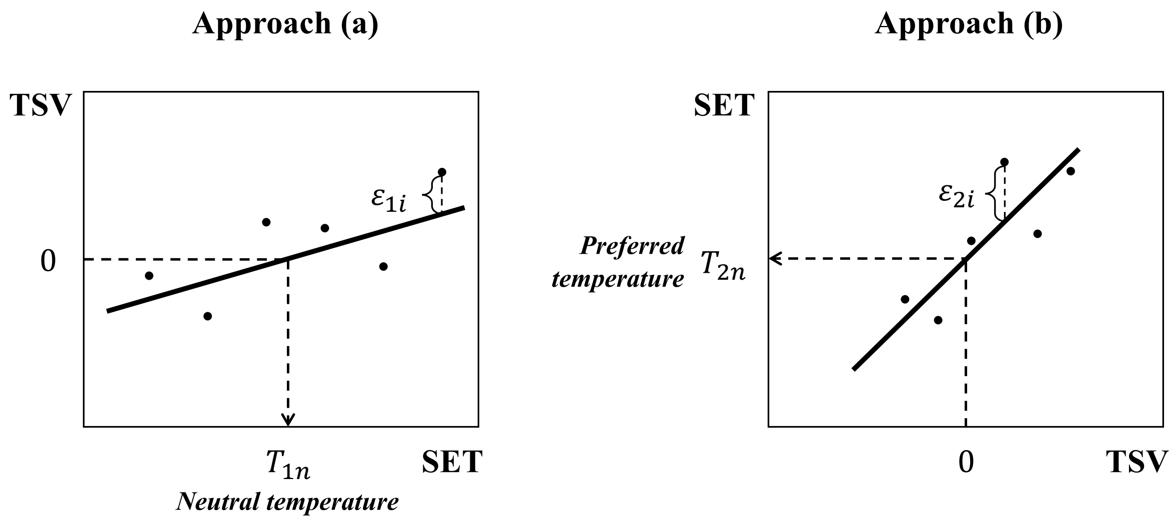

Figure 1. Comparison of two linear regression approaches for estimating 'neutral temperature' by setting the TSV to zero. It is a common misconception that the two approaches are the same and would yield equivalent estimates of the temperature. However, they are fundamentally different both *mathematically* and *theoretically*. We use the term 'preferred temperature' to refer to the 'neutral temperature' determined by Approach (b).

Mathematically, the linear regression models resulting from the two approaches are based on minimizing different *vertical* error terms, visualized as curly brackets in **Figure 1**. Though the two graphs are based on the same set of artificial data points, the error terms represent different factors as the two approaches treat different variables as the y-axis. Therefore, the two regression models, each derived by minimizing the sum of *vertical* errors using the Ordinary Least Squares (OLS) method, are fundamentally different. The mathematical formulas for two OLS estimates of neutral temperature are attached in the Appendix. If the regression models are based on minimizing the sum of *orthogonal* distances between data points and the fitted lines, a method called Total Least Squares (TLS), the two approaches would result in the same estimates of neutral temperature. We acknowledge that the TLS method might be more suitable for neutral temperature estimation because it handles measurement errors in both the conventionally obtained integer values of TSV and continuous values of indoor temperature. However, in this paper, we will focus on the OLS method because it is the standard method used in most thermal comfort data analysis, and the assignment of dependent and independent variables in the OLS method indicates assumptions of causal relationships.

Theoretically, the two approaches also represent different perspectives on which variable influences the other. The conventional Approach (a) represents that people experience a thermal sensation (y-axis) in response to a set of indoor thermal conditions (x-axis). In Approach (b), using the same variables, the relationship is depicted differently, suggesting that indoor thermal conditions (y-axis) are a response to occupants' thermal experiences (x-axis). This perspective offers a different but perhaps equally valid reflection of the complex interplay between the indoor thermal environment and people's response, in terms of both behavior and thermal response. For example, if one is uncomfortable, one might then take actions such as turning up or down the thermostat (often referred to as a "coping mechanism"), making the adjusted indoor temperature the dependent variable. In environments where this is possible, Approach (a) and (b) both co-exist, representing the inherent feedback loops in real buildings.

3. CAUSAL THINKING

From a correlational standpoint, the indoor temperature (represented by SET) and occupants' thermal experience (represented by TSV) are symmetrically associated with each other. In contrast, regression treats one variable as the independent variable and the other as the dependent variable. This is an essential part of causal thinking, which requires us to specify which variable is the cause and which is the effect. Using simplified causal diagrams, we explore two causal relationships between SET and TSV, visualized by the different directional arrows in **Figure 2**. These relationships correspond to the two regression approaches.

Figure 2. Simplified causal diagrams depicting two different causal relationships between indoor thermal conditions and thermal sensations. In the diagrams, the causal factor or *independent variable* is represented as the X node, and the resultant effect or *dependent variable* as the Y node. **Approach (a)** posits that changing in indoor thermal conditions (X) leads to different thermal sensations (Y). Conversely, **Approach (b)** suggests that changing thermal sensations (X) results in different indoor thermal conditions (Y). In this paper, the term 'cause' refers broadly to total causality without distinguishing between direct and indirect causality.

A feedback loop would better reflect the relationship between two variables in any buildings where people have some degree of control over their environment. A comprehensive causal diagram could also show many other factors in between the indoor temperature and thermal sensation relationships. For instance, Approach (a) causal diagram might incorporate skin temperature as a mediating factor, while Approach (b) diagram could encompass the actual control mechanisms affecting SET, like adjusting thermostats, windows, blinds, and fans. Additionally, both graphs could include confounding variables like outdoor climate and occupant demographics that influence both indoor temperature and thermal sensations. Future papers plan to explore these more detailed diagrams. However, in this paper, we focus on simplified two-node diagrams, excluding feedback and mediating factors, to emphasize the importance of causal directions and to discuss their respective implications.

3.1. Approach (a)

Approach (a) reflects a conventional idea that the temperature affects thermal sensations. We use SET to represent an equivalent indoor temperature that integrates both environmental and personal factors of heat transfer. Therefore, the causal relationship is that indoor thermal conditions cause thermal sensations. This relationship has framed many research questions and subsequent data analysis. For example, the design and control of HVAC systems focus on it. The thermal comfort research community has widely studied the effect of indoor thermal conditions on thermal sensation in both laboratory settings and field settings. It is the basis for the definitions and estimations of neutral temperature and the comfort zone.

Approach (a) is clearly applicable under most laboratory conditions, where a person is subject to an imposed indoor temperature, and their thermal sensation is assessed while other thermal factors are maintained at constant values (Fanger, 1970). In this context, participants have little influence over the climate chamber's settings, i.e., temperature, humidity, and air movement. Participants are also not allowed to change their clothing and activities. Approach (a) could also be applicable in certain buildings where subjects do not have the ability, directly or indirectly, to influence their thermal conditions. However, occupants in actual buildings usually have some levels of control over the environment or personal factors. Even in centrally conditioned buildings with inoperable windows, people might be able to adjust clothing, move around, access a portable fan or heater, control a shading element, or complain to a building manager. Such feedback mechanisms mean that the simplified Approach (a) faces limitations in real-world settings by not revealing these control actions.

Approach (a) remains a prevalent method in thermal comfort field studies for determining the neutral temperature in all buildings and the resulting adaptive comfort models for naturally ventilated buildings (Bouden & Ghrab, 2005; de Dear & Brager, 1998; Indraganti, 2010; Nicol & Humphreys, 2002; Singh et al., 2011; Yan et al., 2017; Zaki et al., 2017). However, analyzing the impact of indoor temperature by itself on thermal sensations using Approach (a) doesn't explicitly account for the influences of occupants' psychological and behavioral adaptations, which unfortunately isn't a typical part of the data collection. Moreover, field studies differ from laboratory experiments in that they do not maintain a constant environment. Both indoor thermal environments and occupants' thermal perceptions are dynamic and subject to various confounding factors in real-world settings.

A catalyst for the adaptive comfort model framework was the discrepancy in the effect of indoor temperature on thermal sensation when comparing laboratory experiments to field studies. Many found that when using the PMV model to estimate neutral thermal sensation, the resulting neutral temperatures did not fit various contexts in actual buildings, especially those with natural ventilation. The general conclusion was that while the assumed causal relationship between temperature and thermal sensation as estimated by the PMV may hold true in laboratory experiments, it did not reliably describe what was being experienced by building occupants in real-world settings. In the context of this paper, this long-held thinking is reformulated, both conceptually and mathematically, as a challenge to the limitations of using Approach (a) exclusively. This perspective has been suggested decades ago by other thermal comfort researchers:

"In a climate chamber or fully air-conditioned building the occupant has no or little control over the environmental conditions. If the temperature, humidity, clothing, and activity are fixed, then the reported thermal sensation (i.e., the response) is dependent on the surrounding conditions (i.e., the stimulus); hence most studies regress the thermal sensation on the temperature ***[Approach (a)]****. The adaptive model stresses the importance of the human drive to achieve thermal comfort and argues that thermal comfort is best ensured by giving as much effective control to the occupants as possible, rather than by fixing the room temperature at some theoretically determined optimum. Furthermore, in the home or a "free-running" building, the constraints placed on the occupants are fewer so that they have more individual control over their environment. As the occupants can adapt the environment to their requirements then their thermal sensation is not necessarily the response and the changing environment is not necessarily the stimulus... It may therefore be possible that the neutral temperatures are being miscalculated in other studies"* (Oseland, 1994).

The limitations of implementing Approach (a) in field studies have been highlighted in previous work. Specifically, achieving a statistically significant regression model with indoor temperature regressed against TSV is challenging, often resulting in unreasonable or illogical neutral temperatures (Indraganti, 2010; Rijal et al., 2010; Oseland, 1994). The Griffith method was introduced as a solution to those issues (Nicol & Humphreys, 2010; McCartney & Nicol, 2002). This method sets the regression coefficient, or the slope of the fitted regression line, as a constant value, known as the Griffith constant. Nevertheless, the Griffith constant varies considerably across different contexts (Rupp et al., 2019). The Griffith method has the same causal assumption and faces the same limitations as Approach (a).

3.2. Approach (b)

Approach (b) assumes a different causal relationship between indoor thermal conditions and thermal sensation: the latter *causes* the former. This reversed causal relationship describes how occupants adjust their thermal environmental and personal conditions to achieve their preferred thermal sensations. For example, an occupant feeling slightly warm may decide to modify their environment to achieve a cooler thermal sensation. This may include reducing temperature setpoint on a thermostat, removing layers of clothing, turning on a fan, or opening windows. This causal relationship aligns with the adaptive comfort theory that asserts building occupants are not merely passive recipients of their thermal environment (Brager & de Dear, 1998). Instead, people are active agents who use a variety of adaptive opportunities and coping mechanisms to move themselves from uncomfortable to comfortable conditions.

In contrast to most laboratory settings, this causal relationship likely presents the essential part of feedback loops in occupied buildings. The exclusive one-directional impact of indoor thermal conditions on thermal sensation is diminished in actual buildings because of behavioral adjustment of thermal conditions (such as air movement), psychological adaptation due to experience and expectation (the perception of control can alone increase the temperature range that people find comfortable), and physiological adaptation if exposed for a long time. All these factors effectively *cause* changes in the thermal conditions experienced by that occupant, leading to an amplified impact of thermal sensation on the indoor temperature and other elements of the thermal experience. However, the applicability of Approach (b) in field studies has either been overlooked entirely (in both data collection and subsequent analysis), or has been raised and then considered as the 'wrong' method for neutral temperature estimation for general circumstances:

"It often seems sensible to put the comfort graph the other way around with the temperature on the vertical axis and the comfort on the horizontal ***[Approach (b)]*** *... The problem is that mathematically this means that you are trying to predict the temperature from the comfort vote, which is to assert that the comfort vote is the independent variable that "causes" the change in temperature. There are particular circumstances where this might be defensible but for the normal prediction of comfort from temperature this will give you the wrong answer"* (Nicol et al., 2012).

Adaptive opportunities need to be encouraged in buildings to become the dominant relationship, whether it is simply thermostat control or greater opportunities to have operable windows, fans, and personal comfort systems. Building occupants are commonly dissatisfied with thermal environments (Graham et al., 2021; Karmann et al., 2018), and they should have the ability to exercise thermal adaptations, which in turn affect indoor thermal conditions. In these cases, the observed indoor thermal conditions are more likely to be the result of occupants' effort to create or maintain their preferred thermal sensations at that moment. Initially, people might suffer from hot or cold conditions due to the immediate impact of temperature on thermal sensation. However, in the long term, they will actively adapt to conditions and consistently learn how to make changes to achieve thermal comfort. Assuming adaptive mechanisms are in place in certain buildings, Approach (b) might *conceptually* be more representative than Approach (a), or at least complementary, regardless of whether we have data to quantify the behavioral actions. Again, this has been acknowledged by thermal comfort pioneers:

"When we consider the relation between subjective warmth and room temperature, it is obvious that the room temperature, via the clothing insulation and via the characteristics of the body and the mind, causes the thermal sensation. This is a fairly straightforward example of the thermal environment being the predictor variable (or the 'independent' variable) when regression analysis is used, and therefore the room temperature can be used to predict the thermal sensation. Yet when people have easy control over their room temperature, a reverse path of causation is possible. If they feel too warm, they may turn the heating down, while if they feel too cool they may turn it up a bit. They adjust the room temperature to ensure that their warmth sensation is the one they desire at the time. The subjective warmth could then be regarded as causing the room temperature, and, if so, the subjective warmth becomes the predictor variable and the room temperature the dependent variable" (Nicol et al., 2015).

4. CASE STUDY

We explored different neutral temperature estimation results determined by the two approaches using the ASHRAE Thermal Comfort Database II (version 2.1), which is the largest existing dataset of thermal comfort field studies, containing more than 800 buildings and 10,000 records (Földváry Ličina et al., 2018). In a typical field study, each data collection point includes an individual occupant's right-now-right-here TSV and a full set of indoor thermal condition measurements. A field study could last for several weeks or even months, so multiple collections occur at various times, including both repeated measurements of the same occupant and different measurements across people. Consequently, the impact of earlier collected TSV on subsequent indoor temperature changes, i.e., Approach (b), might be indirectly reflected in the data. In other words, it's *possible* that a building's dataset includes one person's concurrent temperature and too-warm TSV at one time, and the same person or another person's adjusted temperature and neutral TSV at another time, thus reflecting the directional causality implied by the Approach (b). It is important to emphasize that this is an assumption we are making about the relevance of this limited dataset for our analysis to support our conceptual argument, acknowledging that each data point is actually concurrent and without direct recording of the occupants' actions.

We conducted case studies on several randomly selected buildings, and they all resulted in different neutral temperature estimations using Approach (a) and (b). For this paper, we selected one building (ID = 735) from the database that has timestamps for each record and information about occupant control behaviors. The building, an air-conditioned office space in Malaysia, likely provides occupants with access to the thermostat and operable windows, as implied by the original paper (Damiati et al., 2016). The field study was conducted between April 13 and May 5, 2015, during the cooling season. A total of 629 records were collected from 90 occupants across the study period. We used both regression approaches to conduct data analysis for this specific building. Given the different causal assumptions underlying the two approaches, the analysis results obtained from each approach should be interpreted distinctly.

4.1. Neutral Temperature and 'Preferred Temperature'

Using Approach (a), the regression model (TSV $= 0.16\ SET - 4.33$) yields an estimated neutral temperature (SET) of 27 °C (mathematical formulas are provided in the Appendix). This neutral temperature is interpreted as *exposing the average occupant of the example building to an indoor SET of 27 °C would lead to a TSV of zero*. This has been the conventional approach to relating experienced thermal conditions to predicted sensation. Considering the underlying assumption is no behavioral adaptation, and no associated feedback, we believe this interpretation may be limited in thermal comfort field studies, particularly in buildings like this that offer adaptive mechanisms. Similar limitations can also be applied to the interpretation of the slope of the fitted regression line as thermal sensitivity.

In contrast, the regression model using Approach (b) is SET $= 0.65\ TSV + 24.53$, and the differently calculated neutral temperature, which we call preferred temperature, is 25 °C. This result is more than just a mathematical manipulation. For the sake of our conceptual argument presented here, we could perhaps interpret this preferred temperature as suggesting that *if the average occupant can control their environment, the person would adjust the indoor temperature (SET) to 25 °C to achieve a TSV of zero*. This is an entirely different question and answer. Comparing the neutral temperature from Approach (a) and the preferred temperature from Approach (b), we believe, reflects the extent of behavioral actions (i.e., people feel a certain way first, and they then adjust their thermal conditions). If there were no behavioral actions, in theory, these approaches might produce the same result. When there are actions, feedback loops exist, and a combination of Approach (a) and (b) is likely warranted.

We chose this example building because we know, in general, that occupants are making environmental changes, even though it is not reflected in each "right now" concurrent data point. 50% of the building occupants in this field study reported that they had changed the thermostat when they felt hot, and some occupants informed that they opened the windows to let in warmer outdoor air when the indoor temperature was too cold (Damiati et al., 2016). Thus, we interpret our analysis as saying that if each individual could achieve a mean TSV of zero by adjusting the thermostat setting or windows, the mean preferred indoor equivalent temperature (SET) for the group would be 25 °C. In naturally ventilated buildings, occupants' behavioral adaptations, such as changing layers of clothing or opening windows and increasing air movement, could be viewed as adjusting an imagined SET thermostat. The slope of the fitted regression line might indicate occupants' control sensitivity, reflecting the extent to which people would like to adjust an imagined thermostat, given certain thermal sensations.

We acknowledge limitations in our interpretations of the preferred temperature determined by Approach (b), as the concurrent measurement dataset might not perfectly reflect the cause-and-then-effect relationships and inherent feedback loops in buildings. However, our primary intention in this analysis is to highlight the need to apply different interpretations to two regression results, and the idea that they each offer complementary ways to understand occupant response. A correlational perspective doesn't specify the direction of the relationship between two variables, and conventional field study analyses have rarely utilized or discussed the implications of considering Approach (b). Causal thinking has motivated us to explore the difference between the two temperatures and the likelihood they might indicate the degree of behavioral adaptations or interactions that happened in this building.

4.2. Comfort Zone and 'Preferred Zone'

After using the regression model to find two different neutral temperature, we explored comfort zones of this example building, starting with an assumption that TSV being between ±0.85 is considered comfort. It is based on the PMV/PPD relationship that 80% satisfaction corresponds to PMV within ±0.85, a relationship founded on analysis that originally used Approach (a), and was utilized in an earlier version of ASHRAE Standard 55's conventional comfort zone, and also the adaptive comfort model (ASHRAE 55, 2023; de Dear & Brager, 1998). While the adaptive comfort model applied this assumption to regressions from the entire database, we apply it here to the single building just for illustrative comparison, shown in **Figure 3**.

Based on Approach (a) regression model coefficients, the comfort zone is between 21 °C and 32 °C (mathematical formulas are provided in the Appendix). It implies that exposing all building occupants to any indoor temperature between 21 °C and 32 °C would lead to 80% thermal satisfaction rate or a mean TSV within ±0.85. As noted above, this is the conventional way that research has asked and answered questions about thermal experience. But we think the alternative approach might answer different questions.

In comparison, using Approach (b), we estimate a different 'comfort zone', which we term the 'preferred zone'. We want to acknowledge an additional limitation where we are now mixing our quantitative approach—using a relationship derived from Approach (a) and now applying it to Approach (b)—and emphasize that we are doing it to conceptually support our argument that we need to think more holistically about feedback loops in buildings. In spite of these limitations, curiously, this alternatively calculated preferred zone suggests that, in buildings where occupants have adaptive opportunities such as occurred in the case study building, maintaining 80% thermal satisfaction rate among building occupants would result in them controlling the mean indoor SET to range between 24 °C and 25 °C. In other words, if we imagine that the only means for control was through the thermostat (not clothing, air movement, etc.), and that the data actually measured each person's adaptive actions, the indoor temperature would be stabilized at around 24 °C after many back-and-forth thermostat wars. A final mean TSV that would fall within ±0.85 likely represents a compromise that balances different individuals' thermal preferences.

By coloring the data points, we can better understand how these assumed adaptive behaviors might have been happening even just over the three weeks of this field study. For this specific building, both high indoor temperatures and concurrent hot thermal sensations are recorded at the beginning of the field study (early April), which appear as the yellow-green upper-right data points in **Figure 3**. During the field study, occupants might adapt to hot conditions passively or actively, such as lowering metabolic rates, lightening clothing, or changing thermostat settings. Towards the end of the study (early May), lower indoor temperatures and concurrent cooler sensations occurred, appearing as the darker-blue lower-left data points in both graphs. The regression model or the fitted line in this scenario indicates the collective impact of a group of building occupants' thermal sensations on indoor temperature changes. One can imagine that over a month, these changes resulted from a combination of cooler outdoor climates, the HVAC plant responding non-linearly in relation to the outdoor temperature, and also from occupants' directly adjusting the thermostat themselves to create cooler environments and, hence, cooler sensations. Unfortunately, the data doesn't allow us to demonstrate this directly, nor do we know what happened prior to the start of the field study.

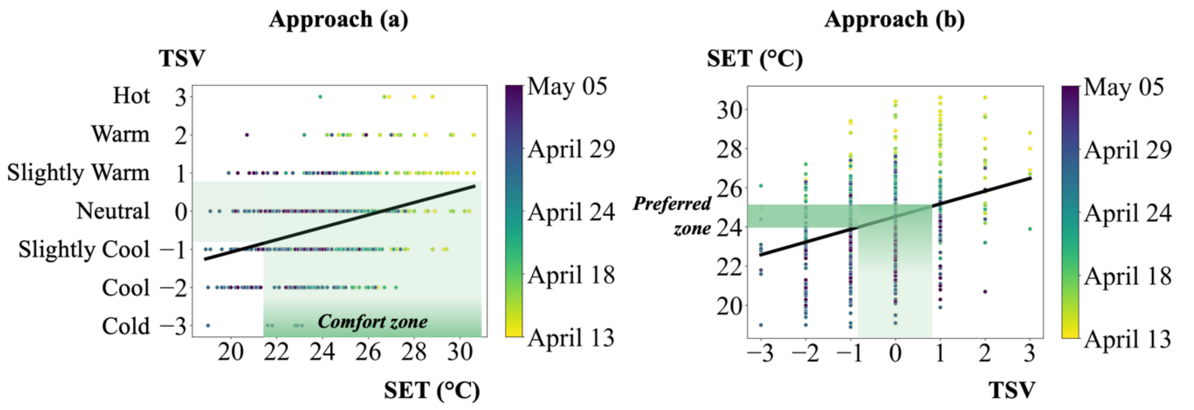

Figure 3. Comparison of two linear regression approaches for estimating 'comfort zone' by setting the TSV to ±0.85. Each data point represents a building occupant's TSV and concurrent indoor temperature, represented by SET. Graph a) shows a fitted regression line and a comfort zone derived using Approach (a). Based on the same dataset and TSV values, Graph b) displays another fitted regression line and a 10 °C narrower zone derived using Approach (b), which we term 'preferred zone'. Though both approaches yield the same R^2 value of 0.11 and the same Pearson correlation coefficient of 0.50, they result in substantially different temperature zones.

The comfort zone has a large implication for building design and building operational energy consumption. Take the operational energy as an example, if the temperature setpoint was set as a lower comfort limit, choosing Approach (b)'s lower limit (24 °C) could save about 30% of building cooling energy compared to Approach (a)'s lower limit (21 °C) (Hoyt et al., 2015). If the setpoint was set as the upper comfort limit, choosing Approach (b)'s upper limit (25 °C) will increase building cooling energy usage, compared to Approach (a)'s upper limit (32 °C).

5. DISCUSSION

We have explored the two approaches using the same dataset, revealing substantial differences in the numerical results. However, interpreting these results is challenging for both approaches. We question the limitations of the prevalent Approach (a) used in field studies, since it does not account for occupant behavior and the inherent feedback available in real buildings. On the other hand, the implications of Approach (b)'s narrower preferred temperature zone remain unclear, and we can only hypothesize what it might mean. Causal thinking points us to examine the data-collection process in thermal comfort field studies as a crucial element.

5.1. Data Collection in Field Studies

Unlike laboratory experiments, typical thermal comfort field studies do not control personal or environmental parameters. This is due to the inability to control such conditions in most situations and the intention of a field study to capture the natural circumstances. The observed subjective and physical measurements don't fully capture how the data influence each other. Consequently, it limits our capacity to sufficiently explore research questions such as adaptive thermal comfort, which involves both directions of causal relationships, formulating feedback loops. In existing field study datasets, it is usually unknown whether variations in indoor temperature could be attributed to several factors - differing outdoor climatic conditions, changes in building façade elements like shades or operable windows, facility manager's operations, or occupant behaviors. A high indoor temperature at a given moment could result from direct solar gain due to occupants raising blinds, opening windows when the outdoor temperatures exceed indoor temperatures, or from earlier adjustments of raising the thermostat setpoint if feeling cold. These are all examples of how one's thermal sensation leads to an *action* that then causes a change in temperature (or air movement, or other factors) but is not typically measured and thus not understood.

These same limitations in the level of detail accompanying physical measurements apply to the subjective measurements from comfort surveys. Field studies of thermal comfort should collect more information about nuanced aspects of human response to better understand the relationship between the physical environment and experience. Causal thinking is an essential starting point of the causal inference framework. It can lead to deeper understanding only if there are datasets to be reliably applied. For example, thermal sensation alone says very little about *perception* or how people understand those sensations (i.e., are they positive or negative, pleasurable, or annoying). It is this *perception* that is going to lead to *actions* that ultimately affect the thermal environment, as understood by Approach (b). In field studies, TSV is often collected as right-now data and matched to the timestamp of coincident environmental measurements. Although this might be similar to what is done in laboratory studies, its meaning might not be the same given the dynamic complexities of real buildings playing out over time. For example, TSV and other perceptions are likely influenced by more than the current moment. It could be partially influenced by the concurrent indoor temperature and also the recent history of thermal conditions, which may or may not have been recorded. It could also be affected by the psychological impact of whether one had personal control over the environmental conditions.

The ASHRAE Global Thermal Comfort Database II we use in this paper does not perfectly match our interests in showing how causal thinking can generate a deeper understanding of how the indoor environments and occupant response can influence each other in both directions. An ideal dataset would be a time-series monitoring of each building occupant's subjective responses (TSV and many more metrics of perception), surrounding thermal environment (temperature and more), thermal control actions, etc. For example, a study might follow a single person over time to record a sequence of perceptions and actions. Feeling hot and *perceiving* adaptive opportunities precedes an *action* of taking off a clothing item or decreasing the thermostat temperature, which leads to new thermal conditions, clothing ensemble measurements, and, subsequently, a new feeling. Some existing data resources might deserve to be explored: the ASHRAE Global Building Occupant Behavior Database (Dong et al., 2022), the SAMBA Continuous IEQ Monitoring Database (Parkinson et al., 2019), six-month field study data of 20 occupants with wearable devices (Tartarini et al., 2022), and six-month field study data of 37 occupants with personal comfort systems (Kim et al., 2019). Perhaps more importantly, new field study data collection protocols should be developed that could standardize these important, more nuanced approaches.

5.2. Causal Thinking in Adaptive Thermal Comfort, Occupant Response, and Building Energy Use

Exploring the application of causal thinking to the role of adaptive actions in influencing one's thermal comfort has raised some discussion points around our understanding of the adaptive framework. One of those discussion points is the difference in the comfort zone resulting from Approach (a) and the preferred zone resulting from Approach (b). Neither of these approaches comprehensively reflect the inherent feedback loops in buildings, which are inadequately represented in typical field study datasets. We believe that, with adequate datasets, new analysis methods under a causal inference framework could potentially address the need to combine the perspectives of both approaches, and generate comfort zones that more accurately represent the implications of occupants' behavioral interactions in buildings.

As noted earlier, the existing adaptive comfort model does not explicitly assert a causal relationship between the prevailing outdoor air temperature and indoor comfort temperature. Adaptive theory hypothesizes that indoor comfort is more related to prevailing outdoor climate patterns in naturally ventilated buildings compared to air-conditioned buildings for several reasons: indoor and outdoor conditions are more closely linked, inhabitants' clothing patterns might vary more (behavioral adaptation), and expectations shift (psychological adaptation) due to a combination of perceived control of operable windows and a greater diversity of thermal experiences. The adaptive comfort model was developed as a simplified tool for practical use, where the outdoor climate is simply used as a proxy and does not explain all these potential causal mechanisms in-between that remain to be studied, measured, and proven.

The applicability of causal thinking can extend to other research questions related to building occupants. Occupant responses and behaviors are not as well understood as building physics, which is a relatively mature field. Building physics has established more accurate cause-and-effect relationships, and the value of causal thinking and causal inference is less critical. Take Fourier's law of heat conduction as an example: $q = -k\nabla T$. In an ideal dataset where there is no measurement error, regressing the heat flux against the temperature gradient, or vice versa, would give us the same estimates of thermal conductivity. In contrast, both the causes and effects of building occupants' responses are very complicated. For instance, research has shown that non-thermal factors, such as access to window views, can influence thermal sensations (Ko et al., 2020). The well-known discrepancy between building energy simulation results and actual energy consumption emphasize the need to include the significant impact of occupant behaviors (Yan et al., 2015). Moreover, data-driven building energy models could lead to biased results (Chen et al., 2024). Therefore, integrating causal thinking is especially critical when addressing research questions related to building occupants and energy use.

6. CONCLUSION

This paper addresses the significance of causal thinking by exploring different causal assumptions and interpretations in two similar regression approaches used in thermal comfort field studies. Our findings reveal that the direction of regression analysis matters, leading to substantial differences in results, and questions about the implications. The disparity between the comfort zone and what we called the preferred zone, calculated from these approaches, remains largely unexplained, primarily due to the limitations in current data collection methods in field studies. While they may mean entirely different things, it is also possible that the implications are complementary and, upon further exploration, may together lead us to a greater understanding. The lack of more detailed and comprehensive data collection within actual buildings hinders our ability to analyze complex causal diagrams or to accurately quantify the causal relationships between building occupants and their physical environment in either direction.

The need for future research to refine and expand the data collection process in field studies is clear, especially in an era where data in the built environment is rapidly expanding. While data-driven statistical methods, ranging from simple regression analysis to other complex machine learning algorithms, are gaining traction, we stress the importance of scrutinizing the causal assumptions underlying these methods, and the relevance of the data used in each. This is particularly important when the statistical model prediction results are interpreted and applied causally. Neglecting to do so can lead to decisions based merely on correlations, potentially compromising occupant comfort, health, and the sustainability of building design and operations.

AUTHOR CONTRIBUTIONS

Ruiji Sun: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Writing - Original Draft. **Stefano Schiavon:** Conceptualization, Methodology, Writing - Review & Editing, Supervision. **Gail Brager:** Interpretation, Writing - Review & Editing, Supervision. **Edward Arens:** Interpretation, Writing - Review & Editing, Methodology Resources. **Hui Zhang:** Conceptualization, Data Resources. **Thomas Parkinson:** Interpretation, Writing - Review & Editing. **Chenlu Zhang:** Methodology Resources.

DATA AVAILABILITY

Data is available on GitHub: https://github.com/CenterForTheBuiltEnvironment/ashrae-db-II.
The code is available on GitHub: https://github.com/ruijis/causal-thinking.git.
Case study results can be reproduced on Binder: https://mybinder.org/v2/gh/ruijis/causal-thinking.git/HEAD.

APPENDIX

Approach (a) presumes a linear relationship between TSV and indoor temperature (SET), as expressed in Equation (a1). The indoor temperature is treated as the cause (independent variable) and the TSV is treated as the effect (dependent variable). The predicted TSV is derived from the regression function in Equation (a2), which contains an error term encapsulating the variance not explained by the SET, which could be unknown causal factors of TSV in addition to random noises (Pearl et al., 2016). For each data point *i*, this error represents the disparity between the predicted value, denoted as $\widehat{TSV_i}$, and the actual value, denoted as TSV_i, as demonstrated in (a3). The model parameters are designated as $\widehat{\alpha_1}$ and $\widehat{\beta_1}$. They are derived by minimizing the sum of squared error terms, depicted in (a4). The prediction function can be found in (a5). By setting the $\widehat{TSV}$ to zero in this function, we can derive the neutral temperature, T_{1n}, as shown in from (a6). By setting the $\widehat{TSV}$ to ±0.85 in this function, we can derive the comfort temperature range, T_{1c}, as shown in (a7). The Pearson correlation coefficient between SET and TSV is calculated as (a8).

$$TSV = \alpha_1 SET + \beta_1 \quad \text{(a1)}$$

$$\widehat{TSV} = \widehat{\alpha_1} SET + \widehat{\beta_1} + \varepsilon_1 \quad \text{(a2)}$$

$$\varepsilon_{1i} = TSV_i - \widehat{TSV_i} = TSV_i - \widehat{\alpha_1} SET_i - \widehat{\beta_1} \quad \text{(a3)}$$

$$\left(\widehat{\alpha_1}, \widehat{\beta_1}\right) = argmin_{(\alpha_1,\beta_1)} \sum_{i=1}^{n} \varepsilon_{1i}{}^2 = argmin_{(\alpha_1,\beta_1)} \sum_{i=1}^{n} (TSV_i - \widehat{\alpha_1} SET_i - \widehat{\beta_1})^2 \quad \text{(a4)}$$

$$\widehat{TSV} = \widehat{\alpha_1} SET + \widehat{\beta_1} \quad \text{(a5)}$$

$$T_{1n} = -\widehat{\beta_1}/\widehat{\alpha_1} \quad \text{(a6)}$$

$$T_{1c} = (0.85 - \widehat{\beta_1})/\widehat{\alpha_1} - (-0.85 - \widehat{\beta_1})/\widehat{\alpha_1} = 1.7/\widehat{\alpha_1} \quad \text{(a7)}$$

$$\rho_{SET,TSV} = \frac{cov(SET,TSV)}{\sigma_{SET}\sigma_{TSV}} = \frac{E[(SET-\mu_{SET})(TSV-\mu_{TSV})]}{\sigma_{SET}\sigma_{TSV}} \quad \text{(a8)}$$

Approach (b) employs an alternate linear model, indicated in (b1). Equations (b1) and (a1) can be interchanged. The model parameters α_2 and β_2 can be represented as $\alpha_2 = 1/\alpha_1$ and $\beta_2 = -\beta_1/\alpha_1$. However, the regression function in (b2) varies from (a2) because of distinct error terms. The error term ε_{2i} in (b3) captures the difference between predicted and actual indoor temperatures, while the ε_{1i} represents the difference between predicted TSV and actual TSV. Since the estimation of model parameters depends on error minimization, the results $\widehat{\alpha_2}$ and $\widehat{\beta_2}$ differ from $\widehat{\alpha_1}$ and $\widehat{\beta_1}$ and are not interconvertible. The prediction function (b5) predicts indoor temperature based on TSV, while (a5) predicts TSV based on indoor temperature. When setting the TSV to zero in (b5), the preferred temperature T_{2n} is equal to $\widehat{\beta_2}$, which varies from $-\widehat{\beta_1}/\widehat{\alpha_1}$. Therefore, T_{2n} is not equal to T_{1n}. When setting the TSV to ±0.85 in (b5), the comfort temperature range, T_{2c} is $1.7\widehat{\alpha_2}$, which varies from $1.7/\widehat{\alpha_1}$. Therefore T_{2c} is not equal to T_{1c}. Interestingly, the Pearson correlation coefficient (b8) is the same as (a8), which might lead to the misunderstanding that the two linear regression approaches are identical.

$$SET = \alpha_2 TSV + \beta_2 \quad \text{(b1)}$$

$$\widehat{SET} = \widehat{\alpha_2} TSV + \widehat{\beta_2} + \varepsilon_2 \quad \text{(b2)}$$

$$\varepsilon_{2i} = SET_i - \widehat{SET}_i = SET_i - \widehat{\alpha_2} TSV_i - \widehat{\beta_2} \quad \text{(b3)}$$

$$\left(\widehat{\alpha_2}, \widehat{\beta_2}\right) = argmin_{(\alpha_2,\beta_2)} \sum_{i=1}^{n} \varepsilon_{2i}{}^2 = argmin_{(\alpha_2,\beta_2)} \sum_{i=1}^{n} (SET_i - \widehat{\alpha_2} TSV_i - \widehat{\beta_2})^2 \quad \text{(b4)}$$

$$\widehat{SET} = \widehat{\alpha_2} TSV + \widehat{\beta_2} \quad \text{(b5)}$$

$$T_{2n} = \widehat{\beta_2} \quad \text{(b6)}$$

$$T_{2c} = 0.85\widehat{\alpha_2} + \widehat{\beta_2} - [-0.85\widehat{\alpha_2} + \widehat{\beta_2}] = 1.7\widehat{\alpha_2} \quad \text{(b7)}$$

$$\rho_{TSV,SET} = \frac{cov(TSV,SET)}{\sigma_{TSV}\sigma_{SET}} = \frac{E[(TSV-\mu_{TSV})(SET-\mu_{SET})]}{\sigma_{TSV}\sigma_{SET}} \quad \text{(b8)}$$

DECLARATION OF GENERATIVE AI IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT only to improve the readability and language of the paper. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

ACKNOWLEDGEMENT

This research was funded by the industry consortium members of the Center for the Built Environment (CBE), University of California, Berkeley.

REFERENCES

ASHRAE 55. (2023). Thermal Environmental Conditions for Human Occupancy. *American Society of Heating, Refrigerating and AirConditioning Engineers*.

Bedford, T. (1948). *Basic Principles of Ventilation and Heating.* https://www.cabidigitallibrary.org/doi/full/10.5555/19642703283

Bouden, C., & Ghrab, N. (2005). An adaptive thermal comfort model for the Tunisian context: A field study results. *Energy and Buildings*, *37*(9), 952–963. https://doi.org/10.1016/j.enbuild.2004.12.003

Brager, G., & de Dear, R. (1998). Thermal adaptation in the built environment: A literature review. *Energy and Buildings*, *27*. https://doi.org/10.1016/S0378-7788(97)00053-4

Cao, B., Zhu, Y., Li, M., & Ouyang, Q. (2014). Individual and district heating: A comparison of residential heating modes with an analysis of adaptive thermal comfort. *Energy and Buildings*, *78*, 17–24. https://doi.org/10.1016/j.enbuild.2014.03.063

Chen, X., Sun, R., Saluz, U., Schiavon, S., & Geyer, P. (2024). Using causal inference to avoid fallouts in data-driven parametric analysis: A case study in the architecture, engineering, and construction industry. *Developments in the Built Environment*, *17*, 100296. https://doi.org/10.1016/j.dibe.2023.100296

Cheung, T., Schiavon, S., Parkinson, T., Li, P., & Brager, G. (2019). Analysis of the accuracy on PMV – PPD model using the ASHRAE Global Thermal Comfort Database II. *Building and Environment*, *153*, 205–217. https://doi.org/10.1016/j.buildenv.2019.01.055

Costa, F., & Shrapnel, S. (2016). Quantum causal modelling. *New Journal of Physics*, *18*(6), 063032. https://doi.org/10.1088/1367-2630/18/6/063032

Damiati, S. A., Zaki, S. A., Rijal, H. B., & Wonorahardjo, S. (2016). Field study on adaptive thermal comfort in office buildings in Malaysia, Indonesia, Singapore, and Japan during hot and humid season. *Building and Environment*, *109*, 208–223. https://doi.org/10.1016/j.buildenv.2016.09.024

de Dear, R., & Brager, G. S. (1998). *Developing an adaptive model of thermal comfort and preference*. https://escholarship.org/uc/item/4qq2p9c6

de Dear, R., Xiong, J., Kim, J., & Cao, B. (2020). A review of adaptive thermal comfort research since 1998. *Energy and Buildings*, *214*, 109893. https://doi.org/10.1016/j.enbuild.2020.109893

Ding, P. (2024). *A First Course in Causal Inference*. CRC Press LLC. https://doi.org/10.48550/arXiv.2305.18793

Dong, B., Liu, Y., Mu, W., Jiang, Z., Pandey, P., Hong, T., Olesen, B., Lawrence, T., O'Neil, Z., Andrews, C., Azar, E., Bandurski, K., Bardhan, R., Bavaresco, M., Berger, C., Burry, J., Carlucci, S., Chvatal, K., De Simone, M., … Zhou, X. (2022). A Global Building Occupant Behavior Database. *Scientific Data*, *9*(1), Article 1. https://doi.org/10.1038/s41597-022-01475-3

Fanger, P. O. (1970). *Thermal comfort: Analysis and applications in environmental engineering*. R.E. Krieger Pub. Co.

Fanger, P. O., & Toftum, J. (2002). Extension of the PMV model to non-air-conditioned buildings in warm climates. *Energy and Buildings*, *34*(6), 533–536. https://doi.org/10.1016/S0378-7788(02)00003-8

Földváry Ličina, V., Cheung, T., Zhang, H., de Dear, R., Parkinson, T., Arens, E., Chun, C., Schiavon, S., Luo, M., Brager, G., Li, P., Kaam, S., Adebamowo, M. A., Andamon, M. M., Babich, F., Bouden, C., Bukovianska, H., Candido, C., Cao, B., … Zhou, X. (2018). Development of the ASHRAE Global Thermal Comfort Database II. *Building and Environment*, *142*, 502–512. https://doi.org/10.1016/j.buildenv.2018.06.022

Frontczak, M., Schiavon, S., Goins, J., Arens, E., Zhang, H., & Wargocki, P. (2012). Quantitative relationships between occupant satisfaction and satisfaction aspects of indoor environmental quality and building design. *Indoor Air*, *22*(2), 119–131. https://doi.org/10.1111/j.1600-0668.2011.00745.x

Gelman, A., & Vehtari, A. (2021). What are the Most Important Statistical Ideas of the Past 50 Years? *Journal of the American Statistical Association*, *116*(536), 2087–2097. https://doi.org/10.1080/01621459.2021.1938081

Gonzalez, R. R., Nishi, Y., & Gagge, A. P. (1974). Experimental evaluation of standard effective temperature a new biometeorological index of man's thermal discomfort. *International Journal of Biometeorology*, *18*(1), 1–15. https://doi.org/10.1007/BF01450660

Graham, L. T., Parkinson, T., & Schiavon, S. (2021). *Lessons learned from 20 years of CBE's occupant surveys* (1). *2*(1), Article 1. https://doi.org/10.5334/bc.76

Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. *Epidemiology*, *10*(1), 37–48.

Houghten, F. C., & Yagloglou, C. P. (1923). Determining Lines of Equal Comfort. *ASHRAE Transactions*, *29*, 163–176.

Hoyt, T., Arens, E., & Zhang, H. (2015). Extending air temperature setpoints: Simulated energy savings and design considerations for new and retrofit buildings. *Building and Environment*, *88*, 89–96. https://doi.org/10.1016/j.buildenv.2014.09.010

Humphreys, M. (1978). Outdoor temperatures and comfort indoors. *Batiment International, Building Research and Practice*, *6*(2), 92–92. https://doi.org/10.1080/09613217808550656

Humphreys, M., & Nicol, F. (2002). The validity of ISO-PMV for predicting comfort votes in every-day thermal environments. *Energy and Buildings*, *34*(6), 667–684. https://doi.org/10.1016/S0378-7788(02)00018-X

Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Indraganti, M. (2010). Using the adaptive model of thermal comfort for obtaining indoor neutral temperature: Findings from a field study in Hyderabad, India. *Building and Environment*, *45*(3), 519–536. https://doi.org/10.1016/j.buildenv.2009.07.006

Karmann, C., Schiavon, S., & Arens, E. (2018). *Percentage of commercial buildings showing at least 80% occupant satisfied with their thermal comfort*. https://escholarship.org/uc/item/89m0z34x

Kent, M. G., Huynh, N. K., Mishra, A. K., Tartarini, F., Lipczynska, A., Li, J., Sultan, Z., Goh, E., Karunagaran, G., Natarajan, A., Indrajith, A., Hendri, I., Narendra, K. I., Wu, V., Chin, N., Gao, C. P., Sapar, M., Seoh, A., Shuhadah, N., … Schiavon, S. (2023). Energy savings and thermal comfort in a zero energy office building with fans in Singapore. *Building and Environment*, *243*, 110674. https://doi.org/10.1016/j.buildenv.2023.110674

Kim, J., Bauman, F., Raftery, P., Arens, E., Zhang, H., Fierro, G., Andersen, M., & Culler, D. (2019). Occupant comfort and behavior: High-resolution data from a 6-month field study of personal comfort systems with 37 real office workers. *Building and Environment*, *148*, 348–360. https://doi.org/10.1016/j.buildenv.2018.11.012

Ko, W. H., Schiavon, S., Zhang, H., Graham, L. T., Brager, G., Mauss, I., & Lin, Y.-W. (2020). The impact of a view from a window on thermal comfort, emotion, and cognitive performance. *Building and Environment*, *175*, 106779. https://doi.org/10.1016/j.buildenv.2020.106779

McCartney, K., & Nicol, F. (2002). Developing an adaptive control algorithm for Europe. *Energy and Buildings*, *34*, 623–635. https://doi.org/10.1016/S0378-7788(02)00013-0

Miller, D., Raftery, P., Nakajima, M., Salo, S., Graham, L. T., Peffer, T., Delgado, M., Zhang, H., Brager, G., Douglass-Jaimes, D., Paliaga, G., Cohn, S., Greene, M., & Brooks, A. (2021). Cooling energy savings and occupant feedback in a two year retrofit evaluation of 99 automated ceiling fans staged with air conditioning. *Energy and Buildings*, *251*, 111319. https://doi.org/10.1016/j.enbuild.2021.111319

Nicol, F., & Humphreys, M. (1973). Thermal comfort as part of a self-regulating system. *Building Research and Practice*, *1*(3), 174–179. https://doi.org/10.1080/09613217308550237

Nicol, F., & Humphreys, M. (2002). Adaptive thermal comfort and sustainable thermal standards for buildings. *Energy and Buildings*, *34*(6), 563–572. https://doi.org/10.1016/S0378-7788(02)00006-3

Nicol, F., & Humphreys, M. (2010). Derivation of the adaptive equations for thermal comfort in free-running buildings in European standard EN15251. *Building and Environment*, *45*(1), 11–17. https://doi.org/10.1016/j.buildenv.2008.12.013

Nicol, F., Humphreys, M., & Roaf, S. (2012). *Adaptive thermal comfort: Principles and practice*. Routledge.

Nicol, F., Humphreys, M., & Roaf, S. (2015). *Adaptive Thermal Comfort: Foundations and Analysis*. Routledge. https://doi.org/10.4324/9781315765815

Oseland, N. (1994). A comparison of the predicted and reported thermal sensation vote in homes during winter and summer. *Energy and Buildings*, *21*(1), 45–54. https://doi.org/10.1016/0378-7788(94)90015-9

Parkinson, T., de Dear, R., & Brager, G. (2020). Nudging the adaptive thermal comfort model. *Energy and Buildings*, *206*, 109559. https://doi.org/10.1016/j.enbuild.2019.109559

Parkinson, T., Parkinson, A., & de Dear, R. (2019). Continuous IEQ monitoring system: Context and development. *Building and Environment*, *149*, 15–25. https://doi.org/10.1016/j.buildenv.2018.12.010

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, *3*(none), 96–146. https://doi.org/10.1214/09-SS057

Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, *62*(3), 54–60. https://doi.org/10.1145/3241036

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics—A Primer* (1st edition). Wiley.

Pearl, J., & Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.

Rijal, H. B., Yoshida, H., & Umemiya, N. (2010). Seasonal and regional differences in neutral temperatures in Nepalese traditional vernacular houses. *Building and Environment*, *45*(12), 2743–2753. https://doi.org/10.1016/j.buildenv.2010.06.002

Rohrer, J. M. (2018). Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42. https://doi.org/10.1177/2515245917745629

Rothman, K. J., & Greenland, S. (2005). Causation and Causal Inference in Epidemiology. *American Journal of Public Health*, *95*(S1), S144–S150. https://doi.org/10.2105/AJPH.2004.059204

Rupp, R. F., Kim, J., Ghisi, E., & de Dear, R. (2019). Thermal sensitivity of occupants in different building typologies: The Griffiths Constant is a Variable. *Energy and Buildings*, *200*, 11–20. https://doi.org/10.1016/j.enbuild.2019.07.048

Schiavon, S., & Melikov, A. K. (2008). Energy saving and improved comfort by increased air movement. *Energy and Buildings*, *40*(10), 1954–1960. https://doi.org/10.1016/j.enbuild.2008.05.001

Schiller, G. E., Arens, E. A., Fountain, M., Doherty, T., Bauman, F. S., & Benton, C. (1988). *A field study of thermal environments and comfort in office bulidings*.

Singh, M. K., Mahapatra, S., & Atreya, S. K. (2011). Adaptive thermal comfort model for different climatic zones of North-East India. *Applied Energy*, *88*(7), 2420–2428. https://doi.org/10.1016/j.apenergy.2011.01.019

Tartarini, F., Schiavon, S., Quintana, M., & Miller, C. (2022). Personal comfort models based on a 6-month experiment using environmental parameters and data from wearables. *Indoor Air*, *32*(11), e13160. https://doi.org/10.1111/ina.13160

Van Hoof, J. (2008). Forty years of Fanger's model of thermal comfort: Comfort for all? *Indoor Air*, *18*(3), 182–201. https://doi.org/10.1111/j.1600-0668.2007.00516.x

Van Hoof, J., Schellen, L., Soebarto, V., Wong, J. K. W., & Kazak, J. K. (2017). Ten questions concerning thermal comfort and ageing. *Building and Environment*, *120*, 123–133. https://doi.org/10.1016/j.buildenv.2017.05.008

Wang, Z., Zhang, H., He, Y., Luo, M., Li, Z., Hong, T., & Lin, B. (2020). Revisiting individual and group differences in thermal comfort based on ASHRAE database. *Energy and Buildings*, *219*, 110017. https://doi.org/10.1016/j.enbuild.2020.110017

Wasserman, L. (2004). 16 Causal Inference. In L. Wasserman (Ed.), *All of Statistics: A Concise Course in Statistical Inference* (pp. 251–262). Springer. https://doi.org/10.1007/978-0-387-21736-9_16

Yan, D., O'Brien, W., Hong, T., Feng, X., Burak Gunay, H., Tahmasebi, F., & Mahdavi, A. (2015). Occupant behavior modeling for building performance simulation: Current state and future challenges. *Energy and Buildings*, *107*, 264–278. https://doi.org/10.1016/j.enbuild.2015.08.032

Yan, H., Mao, Y., & Yang, L. (2017). Thermal adaptive models in the residential buildings in different climate zones of Eastern China. *Energy and Buildings*, *141*, 28–38. https://doi.org/10.1016/j.enbuild.2017.02.016

Zaki, S. A., Damiati, S. A., Rijal, H. B., Hagishima, A., & Abd Razak, A. (2017). Adaptive thermal comfort in university classrooms in Malaysia and Japan. *Building and Environment*, *122*, 294–306. https://doi.org/10.1016/j.buildenv.2017.06.016