Building and Environment 204 (2021) 108087

Contents lists available at ScienceDirect

Building and Environment

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

Information match between continuous occupant data streams and one-time manual surveys on indoor climate

Niels Lassen [a,b,*], Jan Kloppenborg Møller [c], Francesco Goia [a]

[a] Department of Architecture and Technology, Norwegian University of Science and Technology, NTNU, Trondheim, Norway
[b] Skanska Norge AS, Oslo, Norway
[c] Department of Applied Mathematics and Computer Science, Technical University of Denmark, DTU, Lyngby, Denmark

ARTICLE INFO

Keywords:
Occupant satisfaction
Post-occupancy evaluation
Subjective feedback
Thermal comfort
Logistic regression
Field test

ABSTRACT

Occupant-centric data streams, and more specifically continuous subjective occupant feedback (CSOF) systems, offer the possibility for autonomous collection of occupant feedback in buildings. They are made possible by recent developments in pervasive ICT technology and can enable a continuous flow of information that may enhance human-centric building design and operation. Due to the relative novelty of these systems, no research has been developed so far to systematically evaluate whether information collected by CSOF systems is truly representative of the entire population's opinions and evaluations. In this study, we analyze how information on occupant's opinions on indoor climate collected though a multi-level CSOF system compare to the information obtained though simultaneously performed manual surveys. We used data collected from five field tests in modern office buildings with uninformed occupants, and compare a total of 317 *Satisfaction evaluations,* 124 *Complaints,* and 44 *Control actions* with 546 surveys. Using logistic regression techniques, we investigated the relations between the feedback information and the information from surveys. We found that cumulative link models were suitable for modeling the relationship between feedback and survey data. The Building ID tag was the most important variable for modeling *Occupant satisfaction* and *Occupant complaint feedback*. *Occupant control actions* was best modeled using the Workplace ID. When comparing CSOF with surveys, we found a Mean Absolute Error (MAE) of 16% and of 12%, for *Occupant satisfaction* and for *Occupant complaint feedback*, respectively. We demonstrated that the adopted methods are suitable for understanding the meaning of the collected CSOF data. Further studies based on this methodology and using a larger dataset should be carried out to deepen the understanding of CSOF feedback significance and to increase the soundness of the results obtained in this study.

1. Introduction

Acquiring real occupants' opinions on indoor environmental quality is crucial for improving building performance through both design and operation. Post Occupancy Evaluation (POE) methods that include surveys or interviews are currently the only established tool for collecting subjective evaluations from building users. These methods have been described as the most people-oriented approaches for analyzing architectural spaces [1]. POEs are point in time investigations where, (most often standardised) surveys are distributed to occupants during the use-phase of a building. They are still rarely used as an "everyday" process in either the building commissioning phase or in other phases of the building life with the exception of research activities [2]. POEs are gradually becoming more and more used by the building industry beyond research, as a higher focus is currently being placed on the comfort and health of occupants in the whole value chain, yet we are far from seeing POEs carried out on a regular basis as part of the building's assessment, development and maintenance plans. The most common explanations for this are that POEs are expensive and that conducting POEs may uncover legal liabilities [3]. Other more technical limitations of POEs are that they are usually one-point-in-time sampling and that their response rate is usually low to moderate, hence the ability of a classical POE survey to fully reflect the population's opinion might be questioned. Nonetheless, per today, POEs though survey-collected feedback are among the best procedures we have to systematically and quantitatively know what the occupant's opinions are.

Systems based on pervasive digital technologies have opened up new possibilities, options and approaches to collecting occupant feedback with different information depth. In a previous study [4], we classified

* Corresponding author. Department of Architecture and Technology, Norwegian University of Science and Technology, NTNU, Trondheim, Norway.
E-mail address: niels.lassen@skanska.no (N. Lassen).

https://doi.org/10.1016/j.buildenv.2021.108087
Received 13 January 2021; Received in revised form 20 June 2021; Accepted 22 June 2021
Available online 24 June 2021

Acronyms/nomenclature

CSOF	Continuous Subjective Occupant Feedback
IEQ	Indoor Environmental Quality
MAE	Mean Absolute Error
MLR	Multiple Linear Regression
OCD	Occupant Centric Data
POE	Post Occupancy Evaluation
RMSE	Root Mean Squared Error
SPS	Satisfaction Polling Station
TSV	Thermal Sensation Vote
LOOA	Leave One Out Analysis
clm	Cumulative Link Models (package in R)
AIC	Akaike Information Criterion
PCS	Personal Comfort System

the different common technologies and interfaces into the following types: "Participatory sensing apps" (such as [5–14]), "General feedback apps" (such as [15–18]), "Intelligent thermostats" (such as [19,20]), "Intelligent Personal Comfort Systems" (PCS) (such as [21]), and "Polling stations" (such as [22,23]). A closer description of this classification cannot be given here for the sake of brevity. However, even without entering into the details, it is clear the different types of continuous subjective occupant feedback (CSOF) methods and systems may collect information with distinct nature, as there are variations in theme, sensitivity or threshold for responding, and differences in the psychological origin of different categories of subjective feedback [24]. Hence, CSOF systems target, through their design, different information types, and it may be desirable to combine several designs to obtain a comprehensive set of occupant data.

The above-mentioned studies investigate the use of different types of Continuous Subjective Occupant Feedback (CSOF) systems for various uses, such as indoor climate control applications, research applications, building benchmarking, building operation, or creation of personal preference models. They investigate the functioning of the systems in field or laboratory settings, and often evaluate how feedback is correlated to environmental conditions. At the best of our knowledge, no studies have questioned the validity and nature of the information embedded in the collected data when compared to traditional survey methods such as those used in POEs, or to other feedback methods. Furthermore, none of the studies available in the literature assessed in a quantitative way the quality of the collected information compared to traditional surveys. This knowledge gap is somehow natural, as POEs surveys are, in general, the established benchmark of occupant opinion tracking [1], while CSOF methods are still an extremely new and very unexplored domain.

The aim of this study is therefore to investigate and to evaluate in a quantitative way the representativeness of continuously and automatically collected feedback from occupants. By doing this, we aim at unveiling what are the links between the feedback collected though continuous systems and a correspondent "ground truth" survey sample. Data from five field tests carried out in five case buildings have been used for the study. In each of the case buildings, uninformed occupants were asked to use two to three different feedback solutions over time targeting: occupant's *satisfaction evaluations, complaints, and control actions*. The feedback types are described more in detail in section 2.2. Separate surveys were performed on selected days, making it possible to study the links between feedback and survey responses. A statistical, quantitative assessment of the relations between CSOF and survey data will help the research community understand the nature of CSOF methods and help build trust in CSOF collected data. The research presented in this article builds on two previously published studies [4, 25] that introduced related, but simpler, analyses of the data collected from the same field studies. In this study we aim at "concluding" the investigation on the CSOF systems by assessing the data validity through the application of a more advanced statistical method, and by considering all the data from the different case buildings all together, in a comparative way.

The research questions tackled by this study are:

- RQ1 – What is the most suitable regression logic to predict survey results with feedback data?
- RQ2 – What do the models and modeled coefficients tell us about the nature of the systems tested?
- RQ3 – What prediction accuracy can be achieved for predicting survey results with feedback data using these modeling methods?

The research questions are formulated to assess how information acquired through continuous subjective feedback methods compares to the information collected through surveys, and to identify the defining factors for the relationship between these two methods. By identifying the key variables that impact on the (potential) difference in the information collected by the two methods, we aim at unveiling critical features and pinpointing future research needs in this very recent research area, to support a more robust and informed use of these innovative continuous user-centric information sources.

The research design to find the answers to the RQs has led to the following steps:

1. To deploy a comprehensive CSOF system in different office buildings and collect both feedback and survey responses during the same days.
2. To develop and test several models and modeling techniques for multiple linear regression (MLR) in order to investigate the relations between the feedback and the survey (and through this we answer to RQ1).
3. To compare and discuss modeling coefficients and predictions from the tested models (and though this we answer to RQ2).
4. To perform a so-called *leave-one-out* analysis, assessing the prediction accuracy through the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) (and through this we answer to RQ3). Both metrics are used to provide a more comprehensive assessment of the models since RMSE is sensitive to few large errors (and thereby quantifies the effect of large errors) while MAE counts all errors identically.

2. Methodology

2.1. Overview and assumptions

The main idea behind this study is to assess CSOF systems against a "ground truth" for occupant opinions, which we consider to be the simultaneously performed surveys collected from field studies under "normal" operations. The two data sets (the CSOF data and the "ground truth" survey data) consisted of 5288 feedback instances made through a multi-level CSOF system by 183 occupants in 5 buildings under normal operations, and of 628 survey responses from manually distributed surveys in the same field tests.

While the feedback through the continuous subjective systems was collected continuously, the surveys were performed in the afternoon on selected days. Feedback responses were grouped per day and were then assumed to be comparable to the survey responses for the same building and day. This was found to be the most suitable time resolution in order to have a representative number of responses per time-step, although the climate conditions may have varied throughout a single day. We assumed that voters did not change opinions between the time of entering the feedback in the CSOF system and the time of answering the survey. There is, however, no previous literature reference to back up this assumption, as the activities presented in this manuscript are one of the first of its kind and the literature in this field is therefore very limited at the time being. The two data sets were considered to be comparable as they are taken in the same population on the same day, and they have the same sample (sample here meaning the occupants who *could* respond). Although 70–100% of present occupants responded to the survey, a significantly lower percentage gave feedback to the other solutions.

Automatically collected *feedback data* (continuously collected and assumed to have a lower data quality) and *survey data* (spot measurements made manually by researchers approaching every occupant in the study, assumed to have a high data quality) were consequently compared for each level of feedback. As previously mentioned, survey data were considered the ground truth of the experiment.

The research questions are answered by applying the following methods:

RQ1:

We applied the chosen Cumulative Link Model function (clm) [27] to the data in a variety of ways, and with 1–3 explanatory variables. Variables are modeled as both nominal and ordinal to investigate which approach gives the best combination of accuracy and simplicity. Where appropriate we tested variables as random effects. The best performing models according to likelihood and Akaike Information Criterion (AIC) were selected for further analysis and the differences in performance were discussed.

RQ2:

The model predictions and metrics (coefficients, intercepts) were further discussed and interpreted in view of the practical setting to understand what the models can tell us about the data and the mechanisms affecting the feedback and survey data.

RQ3:

We evaluated the accuracy of the models on a "new" building by removing one building from the dataset and comparing model predictions to actual survey results from the survey day that had the highest number of SPS responses for that building - this type of verification is commonly called leave-one-out-analysis (LOOA). The results were displayed graphically and quantified in the form of Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). This analysis was only done for the buildings where there was enough data to perform the tests.

Multiple Linear Regression (MLR) techniques is a form of the so called "grey box" modeling, and this technique was chosen for the study as they offer a more transparent and practically understandable data driven model as opposed to a large range of other possible unstructured models known with the name of "black box" models. MLR models are more suited for answering the question regarding the nature of the systems tested. They are also better suited for use with small datasets, while "black box" models require a far greater amount of data points to obtain accurate results. The choice of the modeling technique is further described in section 2.3.2.

2.2. Field studies and data collection

The field tests were conducted as longitudinal "blind" tests in five real office environments, where the occupants had not been informed in detail on the actual intent of the study (i.e. to investigate how they used the feedback systems and how the information collected through different methods is correlated), but were only notified about the general goal of assessing their experience with the indoor climate in the building. We believed this approach to be adequate to simultaneously inform the occupants on the overall experiment activity while maintaining the occupants free from any bias that could arise by detailing the specific use of the data about their interactions with the CSOF methods. Two of the field experiments were conducted in the buildings located in Oslo metropolitan area (Norway), between September 2018 and January 2019. Two studies were performed in the San Francisco Bay Area (California, USA), between April and July 2019. The last study was conducted in Oslo (Norway), from January 2020 until it was prematurely stopped due to the Covid-19 pandemic in March 2020.

All the tests were conducted in a similar way – the feedback systems were deployed in the building and used by the occupants for a period of approximately 1 to 3 months. On selected days, a separate "manual" survey was carried out. For this data collection method, two different approaches were used, as a result of refining the methodology during the entire research activity. In Building 1 and Building 2 an electronic survey was distributed to the test subjects via an email link and the respondents were asked to take the last week into consideration when answering all the questions. In Building 3–5 a researcher approached each occupant in the room/office and asked them to fill out a 2-min survey on a tablet computer. Survey questions from all buildings are given in Appendix B.

The feedback systems tested can be divided into 3 levels, where level 1 and 2 were both incorporated into a feedback polling station (named Satisfaction Polling Station, SPS), and level 3 was integrated in a separate system for individual environmental control. The term and idea of describing feedback system on different "levels" is further described in a separate study [24] where we developed a hierarchical structure based on different levels to organize and define subjective occupant feedback data streams and the embedded information for each of these. More in details, the three levels' characteristics were:

- Level 1*: Occupant satisfaction polling stations (SPS)* – A publicly available polling station (Tablet computer on a stand) with the question "How satisfied are you with the indoor climate today?" followed by 5 smiley face buttons.
- Level 2: *Occupant complaints* – Occupants who chose one of the two negative smileys on the SPS were prompted to answer a second question asking "Please can you pinpoint the problem?". They could then choose among seven categories such as "Too cold" or "Lighting issues". This option was only available in Buildings 3–5.
- Level 3: *Occupant control actions* – Occupants were given a 60 W personal electric radiant heater which was attached to the underside of their desk. They could switch on their heater for 30 min by pressing a button lying on their desk. This option was only available in Building 3.

The three components of the feedback system were introduced gradually, one by one, and occupant surveys were performed at certain points in time. So-called *temperature interventions* (deliberate stepwise changes in the ambient room temperature) were carried out on selected days. The temperature interventions were performed in three of the buildings with the intention of provoking feedback and control actions from the occupants. Detailed descriptions of the experiments are given

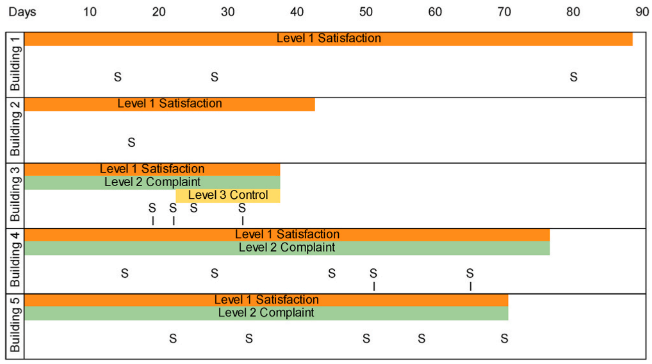

Fig. 1. Experiment procedures with tests of CSOF types (given by rows with color), surveys (where each survey is marked by an S) and interventions (where each intervention is given by an I) for each building. Time is on the X axis. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

in Appendix A. A schematic of the field studies is shown in Fig. 1.

Although similar, the experiments were not all alike due to the development of the CSOF system and of the testing methodologies themselves during the research. It was an aim for the field tests to be as close as possible to realistic conditions, which also limited the control the researchers had over the experiments. The main differences between the tests were:

- All levels were not tested in all buildings
- Building 1 and 2 have surveys where occupants were asked to answer for the previous week. These responses are still compared to feedback for the same day only.
- The indoor and outdoor climate conditions were different.
- The studies were performed on different occupant groups and cultures.
- The study and feedback equipment may have been communicated differently in the different studies because of constraints on the conventional flow of information to the occupants, although it was sought to do this in a similar manner.

We assume that the differences between the studies do not make a comparison nonvalid, since the point here is exactly to understand "on average" how representative CSOF systems are compared to surveys. Hence, a certain degree of different details is healthy to increase the spectrum of investigated conditions and make the studies representative of real office conditions.

The data quantities collected in each building are shown in Table 1. Information embedded in each feedback type is given in Table 2.

Table 1
Collected data quantities for each building and feedback type.

Building	Feedback type	Survey-days	Survey responses	Feedback responses
1	SPS Satisfaction	3	24	16
2	SPS Satisfaction	1	7	5
3	SPS Satisfaction	5	93	19
3	SPS Complaints	5	9	20
3	Heater use	1	18	44
4	SPS Satisfaction	5	336	199
4	SPS Complaints	5	72	97
5	SPS Satisfaction	5	75	66
5	SPS Complaints	5	10	7

Table 2
Embedded information in the datasets used for modeling.

Feedback type	Embedded information
Satisfaction evaluations	Satisfaction score (feedback), satisfaction score (survey), building number, feedback response rate, intervention, survey number
Complaints	Complaints (feedback), complaints (survey), building number, intervention, survey number
Control actions	Number of heater activations per workplace, Thermal Sensation Vote (TSV) per workplace (survey), workplace identifier, building number (only Building 4)

2.3. Modeling techniques and assumptions

2.3.1. Data preparation

We created data sets for each of the three feedback types, where each set contained feedback data, building number, date, and corresponding survey results from the same day. Some sets also contained information about temperature intervention, feedback response rate, or workplace identifier. The datasets were processed using the following steps: 1) feedback data from all field tests were combined with columns for date and building number. 2) Survey data from all experiments were matched and merged to these data using date and building number as identifiers. We now only kept feedback data where survey data were collected from the same day in the same building. 3) Data sets were populated with information and calculations of response rate (number of votes for that day per assumed number of occupants in the room) and temperature interventions (binary 1/0 for whether there had been a temperature intervention on that day). 4) Data sets were transformed to long format using "source" and "count" as key and value columns.

2.3.2. Model construction

Several possible regression techniques exist to model relations between measured data. Lately, machine learning techniques such as random forest or classification tree models or similar have become popular, as they offer a high level of automation and high accuracy. Given the limited amount of data and the few explanatory variables available in this study it was more natural to choose a more classical statistical approach using manual model selection. A clear advantage of the simpler regression models is that the model parameters may be interpreted and give practical meaning. In this case model parameters can give insights into the nature and functioning of the data, users, or building.

Previous studies have questioned the equidistance assumption of thermal sensation and comfort vote scales [26]. The same argument was assumed to apply for the collected subjective data in this study. Therefore, linear regression analysis cannot be applied and models based on logistic regression were chosen. Furthermore, the data sets contained a combination of categorical (both ordinal and nominal) and continuous factors. Nominal data is defined as non-parametric data that is used for naming or labelling variables, without any quantitative value. Ordinal data is a type of categorical data with an order (non-parametric ordered). In some cases it is not incidentally clear whether the data is ordinal or nominal, while in other cases it is obvious that the variables are in practice one of the two. The reason, in our study, to test them in the models as both ordinal and nominal variables was to quantify the effect of changing the way they enter into the model. In order to interpret models, the model complexity must be kept as low as possible and complexity is affected by choice of ordinal or nominal data. Therefore, ordinal, nominal, and mixed effect regression analysis was chosen using function "clm" and "clmm2" from the R package ordinal [27]. "clm" models are in the family of generalized linear models, while "clmm2" is in the family of generalized mixed effect models [28]. Mixed effect models treat some of the variables as random effects representing some population (in our case this was User ID and/or Building ID). An overview of all the models used is given in Table 3.

Table 3
Overview of models tested. Satisfaction score (sat), Feedback type (source), Building ID (building_no), Intervention (int), Response rate (resp), Number of presses per day (usage), Workplace location (location).

Feedback type	Model no.	Syntax	Description
Satisfaction evaluations	M1.1	clm(sat ~ source)	One coef. (ordinal: source)
	M1.2	clm(sat ~ 1, nominal = ~ source)	One coef (nominal: source)
	M2.1	clm(sat ~ 1 + building_no, nominal = ~ source)	Two coef, (ordinal: building, nominal: source)
	M2.2	clm(sat ~ 1 + int, nominal = ~ source)	Two coef, (ordinal: intervention, nominal: source)
	M2.3	clm(sat ~ 1 + respr, nominal = ~ source)	Two coef, (ordinal: response rate, nominal: source)
	M2.4	clm(sat ~ 1 + survey_no, nominal = ~ source)	Two coef, (ordinal: survey no., nominal: source)
	M3.1	clm(sat ~ 1, nominal = ~ source + building_no)	Two coef, (nominal: building & source)
	M3.2	clm(sat ~ 1, nominal = ~ source + int)	Two coef, (nominal: intervention & source)
	M3.3	clm(sat ~ 1, nominal = ~ source + respr)	Two coef, (nominal: response rate & source)
	M4.1	clm(sat ~ 1+respr, nominal = ~ source + building_no)	Three coefs, (ordinal: response rate, nominal: building & source)
	M4.2	clm(sat ~ 1 + int, nominal = ~ source + building_no)	Three coefs, (ordinal: intervention, nominal: building & source)
	M1.1re	clmm(sat ~ source + (1|building_no))	Two coef, (ordinal: source, random effect. building)
Complaints	M1.1	clm(compl ~ source)	One coef. (ordinal: source)
	M1.2	clm(compl ~ 1, nominal = ~ source)	One coef (nominal: source)
	M2.1	clm(compl ~ 1 + building_no, nominal = ~ source)	Two coef, (ordinal: building, nominal: source)
	M2.2	clm(compl ~ 1 nominal = ~ source + building_no)	Two coef, (nominal: building & source)
Control actions	M1	clm(TSV ~ usage)	One coef. (ordinal: TSV)
	M2	clmm(TSV ~ source+ (1|location)	Two coef (ordinal: usage, random effect: location)

Table 4
Model performance by the Analysis of Deviance method where the P column states whether the model performs significantly better than the model listed above.

Model no.	no.par	Log Likelihood	AIC	Pr (>Chisq)
M1.1	5	−1259.349	2528.698	NA
M1.1re	6	−1250.960	2513.921	***
M1.2	8	−1223.371	2462.741	***
M2.2	9	−1222.168	2462.337	ns
M2.3	9	−1223.013	2464.026	NA
M2.4	9	−1221.935	2461.870	NA
M2.1	12	−1211.606	2447.213	***
M3.2	12	−1217.725	2459.451	NA
M3.3	12	−1217.315	2458.630	NA
M3.4	12	−1219.738	2463.476	NA
M3.1	24	−1196.722	2441.444	***
M4.1	25	−1195.572	2441.145	ns
M4.2	25	−1196.718	2443.435	NA

3. Results and discussion

3.1. SPS satisfaction evaluation

3.1.1. Model performance

A number of regression models using different parts of the data as explanatory variables were tested. Model performance was evaluated by comparing the models by likelihood ratio (or Analysis of Deviance) tests [28]. The model information and performance metrics are listed in Table 4, showing the number of parameters in the model (no.par), likelihood ratio (Log Likelihood) performance, Akaike Information Criterion (AIC), and whether the model performs significantly better than the model listed above (Pr). Corresponding to $P > 0.05$ – ns (non-significant), $0.01 < p < 0.05$ *, $0.01 < p < 0.001$ **, $p < 0.0001$ ***

The results show how the cumulative link model that assumes nominal data for the satisfaction scores (M1.2) performs significantly better than the model which assumes ordinal data (M1.1). Furthermore, the models that also include the building number (Building ID) as an explanatory variable (M2.1 and M3.1) perform significantly better than those using response rate, intervention and survey number. They do, however, introduce a higher number of parameters (12 and 24, respectively). Modeling the Building ID as a random effect variable (M1.1re) would be meaningful if the model is to be applied in new buildings where the "building_no" parameter is not already established. This model approach, however, was found to have a low performance as assessed by the AIC and Log Likelihood. This is likely due to the low number of buildings in the data set and the fact that the source variable must be modeled as an ordinal factor when using the "clmm" function. It was in this case not possible to model the source variable with a multinomial distribution while at the same time treating the Building ID as a random variable, as there are no currently available standard tools for this. The models containing information about temperature interventions (M2.2, M3.2 and M4.2) do not perform well. Nor do the models containing information about the SPS response rate (M2.3, M3.3 and M4.1) and the survey number (M2.4 and M3.4). The addition of response rate in combination with the Building ID and source explanatory variables in M4.1 gives an improvement, although not significant.

The following models were selected for further analysis: M1.1 (as it is very simple and hence easy to interpret), M1.2 (as it is also simple but gives a quantification of the difference between surveys), M2.1 (as it also quantifies the difference between buildings), M3.1 (as it separates models for each building and each survey). The mixed effect model M1.1re using the Building ID as a random effect is not selected for further study as it displayed a lower performance.

3.1.2. Determination of important variables

The results in Table 4 indicate that the source and the Building ID are the most influential parameters for predicting the Survey result. Further, it is found that the SPS scores and Building ID should be modeled as nominal factors. The models M1.2, M2.1 and M3.1 are selected as the most important for further analysis.

The probabilities predicted for SPS and survey in each satisfaction level using models M1.1 (continuous line) and M1.2 (dotted line) are plotted in Fig. 2. The slope of the line can be understood as the translation from SPS data to survey data, hence a flat line corresponds to a perfect match between SPS information and survey information. In model M1.1, where the satisfaction levels are modeled as ordinal, the slopes are determined in relation to each other and there are relatively minor differences in the slope from level to level. In model M1.2, where the satisfaction levels are modeled as nominal, each slope is determined individually, and the differences are larger. For level "Very dissatisfied", there is a much higher likelihood of receiving this vote on the SPS than on the survey. The same is evident for levels "Dissatisfied" and "Very satisfied". This means that under (what is assumed to be) the same conditions, with the same population, there is a larger likelihood of receiving these votes on the SPS (where we do not know who is voting and the same voters may vote repeatedly) than on the survey (where each occupant votes one time). The translation for "Neutral" and "Satisfied" are inverse of this, meaning that the probability is lower to receive these votes on the SPS (where some occupants may choose not to vote) than in the survey. The results may indicate that, for the entire dataset, SPS votes need to be corrected for a "multiple-response bias" for the "Very dissatisfied", "Dissatisfied" and "Very satisfied" levels, while they need to be corrected for a "non-response bias" for the levels "Neutral" and "Satisfied". However, the differences could also be caused by the fact that the SPS and survey responses have been entered at slightly different time and place, where a "here-and-now" effect may have affected the results. However, in this case we would expect the differences in votes to be more uniformly distributed across the different satisfaction levels. These findings are in line with the observations from an earlier study where we applied a different method to analyze the data from the same experiments in Buildings 4 and 5 [25]. The most likely lesson-learned from this evidence is that occupants who are neutral or satisfied do not vote at the SPS as often as people who are dissatisfied or very satisfied do. An example of the model coefficients (here from model M1.2) is given below.

Threshold coefficients:

	vdiss|diss	diss|neu	neu|sat	sat|vsat
(Intercept)	−1.1667	−0.3999	0.1138	1.0070
Sourcesurvey	−1.5538	−1.0058	−0.4808	0.4609

The "sourcesurvey" coefficients determine the conversion factor between SPS and survey in the logit domain for the intercepts between each level. Only the intercepts are given since this is a cumulative model.

The Building ID variable represents individual differences between buildings. These differences could involve cultural differences, difference in voting habits (which may be influenced by how the SPS was introduced, how it was placed, occupant opinions, and much more), difference in physical climate, level of personal control, expectations, and much more. In Fig. 3, the y-value of each point represents the probability of receiving votes at that level. For instance, the green dotted (model M3.1) line for Building 2 in the "sat" level is higher than the others, meaning that there is a higher probability of "sat" votes in this building, both on the SPS and survey. This could be explained considering that the occupants may be satisfied with the indoor climate. The red line for Building 1 is much lower, meaning that an opposite behavior is seen in this building. The two lines are however approximately

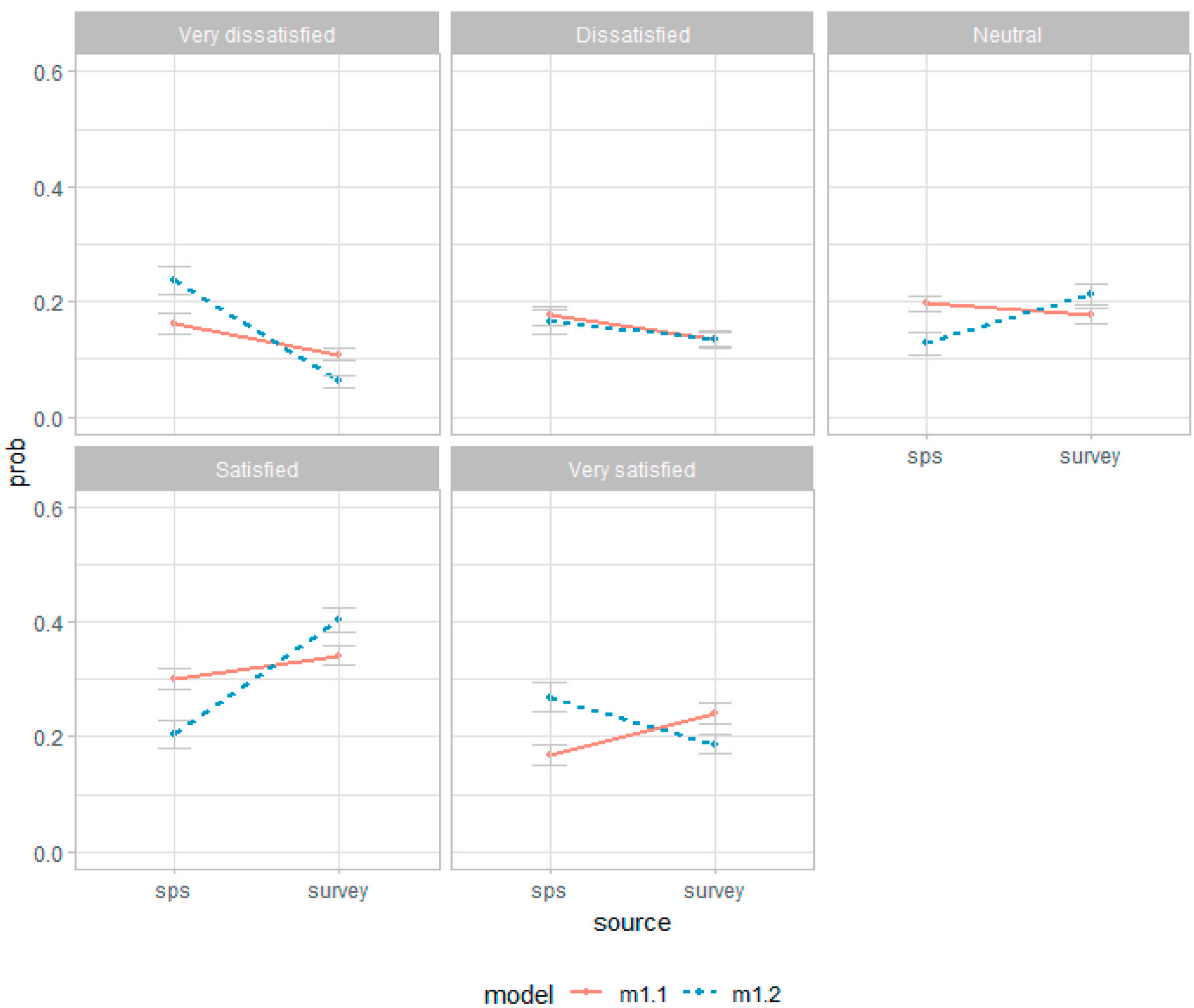

Fig. 2. Probabilities (y-axis) for receiving votes for each of the satisfaction levels on SPS and survey for all buildings combined using models M1.1 (continuous line) and M1.2(dotted line). The standard error is given in error bars (grey). We see how M1.2 results in larger differences from level to level. The conversion factor between SPS and Survey is shown as the slope of the line.

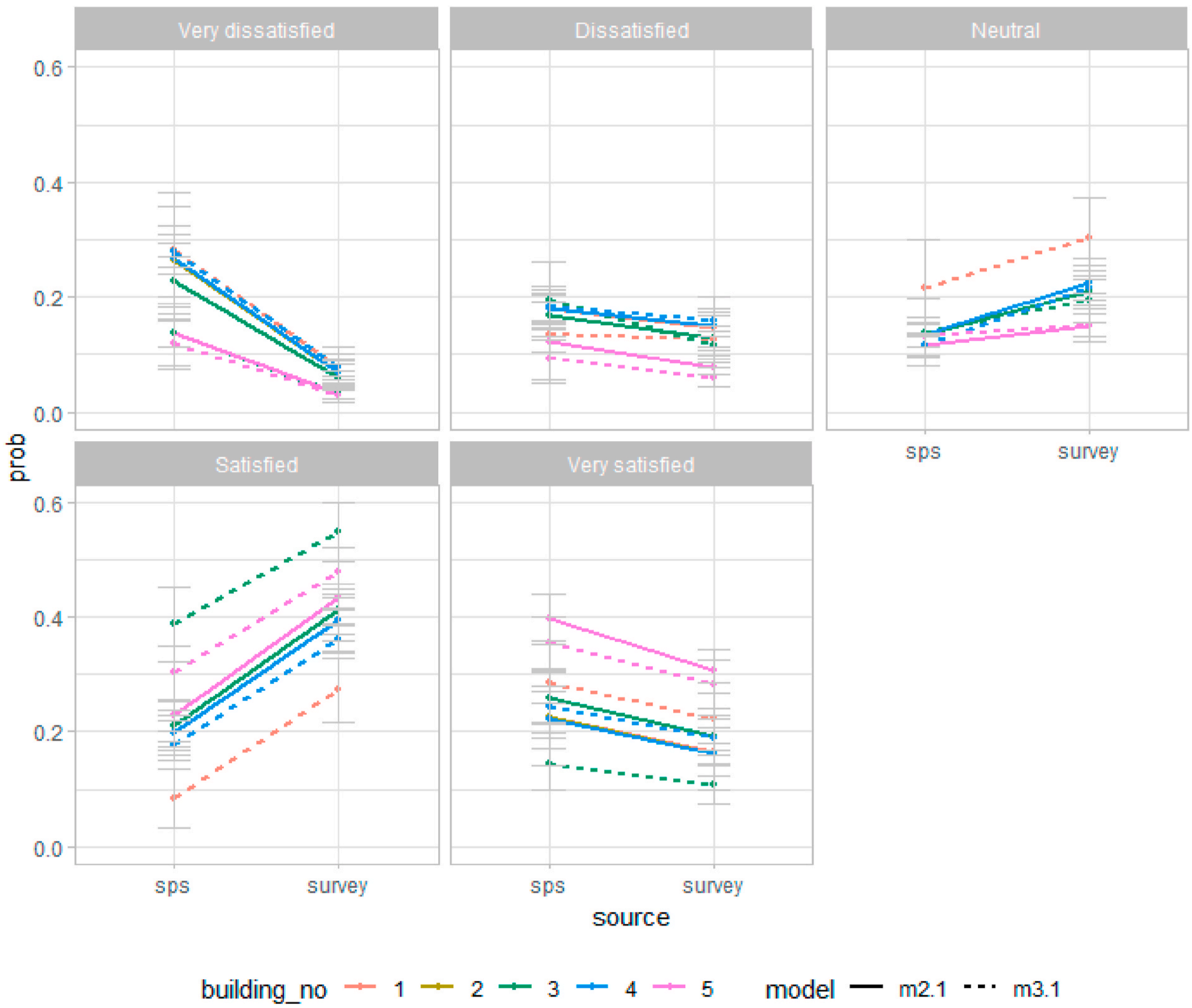

Fig. 3. Probabilities (y-axis) for receiving votes for each of the satisfaction levels on SPS and survey (x-axis) for each building using models M2.1 and M3.1. The standard error is given in error bars (grey). Due to low data quantities, Building 2 was removed from the dataset for model M3.1. We see how model M3.1 estimates larger differences between the individual buildings. The conversion factor between SPS and Survey is shown as the slope of the line.

parallel, meaning the conversion factor is similar in both buildings. This is not the case in level “diss” for the same two buildings, as the lines have different slopes. For some unknown reason, occupants in Building 2 have a higher probability of voting “diss” on the SPS than voters in Building 1, even though the probability is similar on the survey in the same two buildings. There seems to be some cultural, communication related, or other unexplained reason for voters using the SPS differently in the two buildings. This shows that the conversion factors between SPS and survey are not necessarily alike in all buildings, and the reasons for the differences may be difficult to identify and predict without doing several surveys to establish a conversion factor for each specific building.

3.1.3. Prediction accuracy for new buildings

A *leave-one-out-analysis* was performed to identify the potential errors when model M3.1 was used on a “new” set of data. Data from one building at a time were removed from the dataset, and new model coefficients were obtained without the data from the removed building. Predictions were made for the survey result using SPS data on one chosen survey-day. The predictions were compared to the actual survey results from the same day. The results are shown in Fig. 4. It should be noted that the data quantities were low for Building 1 and Building 2, resulting in reduced predicting performance.

The total prediction performance of the models on unknown buildings is expressed as mean absolute error (MAE) for model M3.1 of 15.9%, with a RMSE of 0.229. For model M2.1 the MAE is found to be 15.8% and the RMSE 0.231. Model M2.1 had fewer, more extreme errors than M3.1 (since the RMSE is comparatively higher than the MAE). It should be noted that, in addition to the data volumes being low for Buildings 1–3, the error estimate was sensitive to which survey days were chosen. The sample sizes for this analysis are given in the caption of Fig. 4. Unfortunately, we do not know of any other studies comparing feedback results to survey results and it is therefore difficult to determine a reasonable threshold for what could be considered “sufficient” or acceptable accuracy. In a previous study using the same data, it was found that SPS feedback would converge toward a mean when more than 10 votes had been entered. It is therefore reasonable to assume that samples of less than 10 votes give uncertain results. In this case, only the SPS samples of Building 4 and 5 have more than 10 votes.

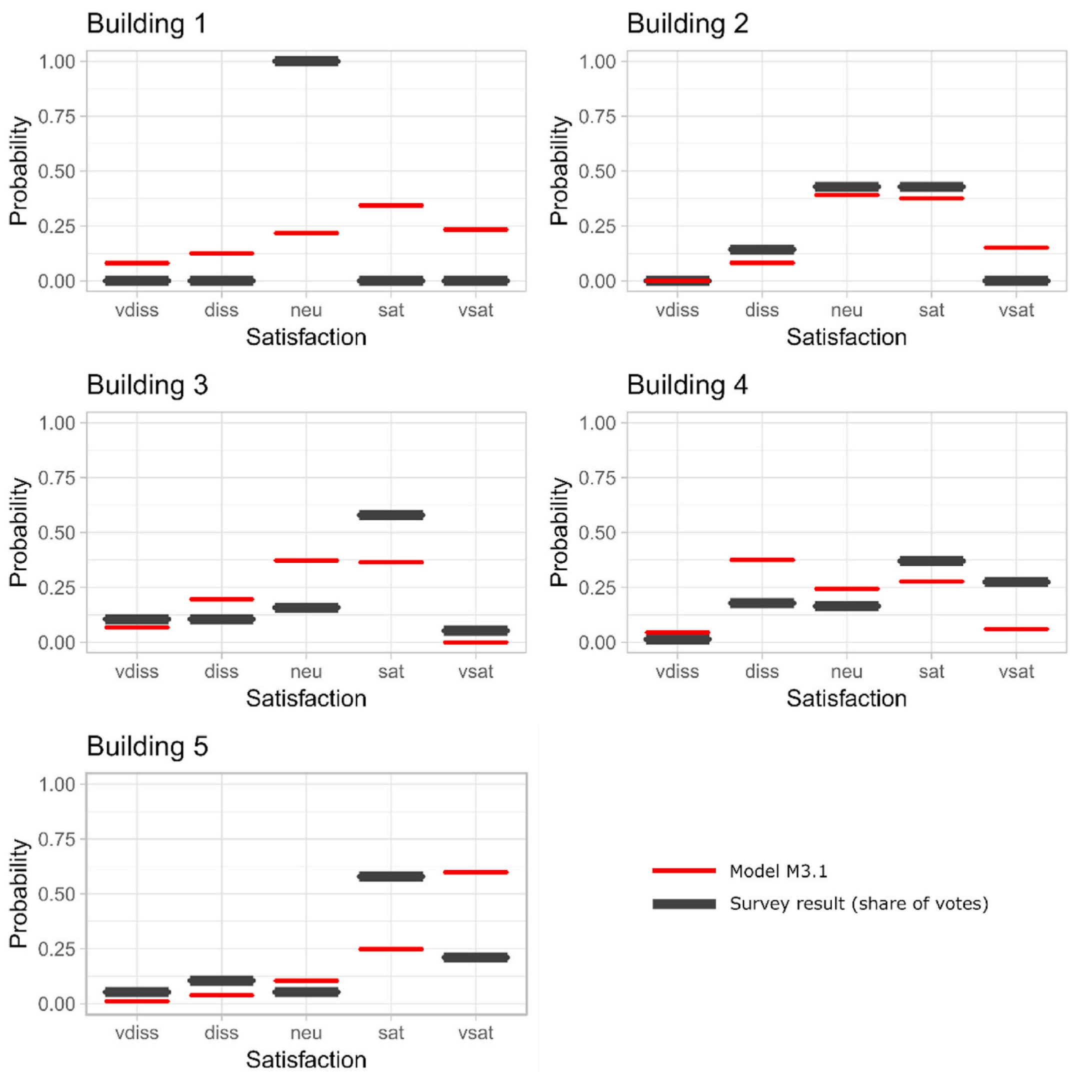

Fig. 4. Leave-one-out analysis for model M1.3. Predictions for model m3.1 (red) are compared to actual survey results (black) for each response alternative. The first survey was chosen for building 1,2 and 3, while the second survey was chosen for building 4 and the third survey for building 5. Sample sizes were: Building 1: SPS:7, Survey:2; Building 2: SPS:5, Survey:7; Building 3: SPS:4, Survey:19; Building 4: SPS:64, Survey:73, Building 5: SPS:20, Survey:19. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.2. Complaints

3.2.1. Model performance

A number of regression models using different parts of the data as explanatory variables were tested. Model performance was evaluated by comparing the models with likelihood ratio [28] tests. The amount of data for complaint "draft" was too low for the models to converge. Therefore complaints because of "draft" were combined with the response category "other" creating a category with sufficient data for modeling. No complaint data were collected for Building 1 and 2, hence this analysis is only performed for Buildings 3–5. The metrics that quantify the models' performance are listed in Table 5, showing the number of parameters in the model (no.par), likelihood ratio (Log Likelihood) performance, Akaike Information Criterion (AIC), and whether the model performs significantly better than the model listed above (Pr). Corresponding to $P > 0.05$ – ns (non-significant), $0.01 < p < 0.05$ *, $0.01 < p < 0.001$ **, $p < 0.0001$ ***

The results show how the model that assumes nominal data for the satisfaction scores (M1.2) performs significantly better than the model that assumes ordinal data (M1.1). This is reasonable as the complaint categories are not ranked or dependent on each other. The model that also includes the Building ID as an ordinal factor performs significantly better than M1.2, while using Building ID as a nominal factor (M2.2) does not lead to a significantly better performance. When modeled as a nominal factor, the effect of the Building ID is considered individually for each complaint level. This indicates, therefore, that the differences between buildings are smaller than with the satisfaction evaluations, and the value of adding the Building ID variable is smaller. Since the performance of the nominal model is not better, the differences between complaint levels shown for the M2.2 model in Fig. 4 can most likely be as due to the low data quantity and cannot probably be considered fully representative of a given phenomenon.

Table 5
Model performance by the Analysis of Deviance method where the P column states whether the model performs significantly better than the model listed above.

Model no.	no.par	Log Likelihood	AIC	Pr (>Chisq)
M1.1	6	−336.1733	684.3465	NA
M1.2	10	−330.7594	681.5189	*
M2.1	12	−322.2268	668.4536	***
M2.2	20	−320.9679	681.9359	ns

3.2.2. Determination of important variables

The results in Table 5 indicate that the Building ID parameter is less influential for predicting the Survey results for complaints than for predicting satisfaction evaluations. Furthermore, it was found that the source parameter should be modeled as a nominal factor. The probabilities predicted for SPS and survey in each satisfaction level using models M1.1 (continuous line) and M1.2 (dotted line) are plotted in Fig. 5. As for the previous figures, the slope of the line can be understood as the translation from SPS data to survey data. In model M1.1, where the complaint levels are modeled as ordinal, there are relatively minor differences in the slope for the different complaints and the conversion line is close to horizontal for all complaints. This indicates that the share of complaints for one complaint category is likely to be equal for both SPS and survey. In model M1.2, where complaints are modeled as nominal, there are larger differences in the slope. The differences are still smaller than for the satisfaction votes. “Too hot” complaints and complains related to indoor air quality (IAQ) seem to appear less frequently on the SPS than in the survey, while the other complaints show a slightly opposite trend. Although M1.2 model performed significantly better than the M1.1 model, it is still likely that these mentioned effects are random. It is clear, however, that the translation coefficients are smaller for complaints than for satisfaction evaluations.

A similar plot of model M2.1 is shown in Fig. 6. Differences between the buildings are relatively small. Many "too cold" complaints were recorded in both SPS and survey of building 5, which also fit the researchers’ opinion who experienced this building as having a lower temperature than the others. The slope of the lines between SPS and survey are, however, similar for many of the complaints. In this case too, the amount of collected data appear to be too low for any sound conclusions to be drawn from the model parameters for each individual building. Nonetheless, the results show a clear trend with the present data.

3.2.3. Prediction accuracy for new buildings

A *leave-one-out analysis* was performed to identify the potential errors when the models were used on a “new” set of data. The procedure for this analysis is identical to that presented in section 3.1.3. The results are shown in Fig. 7, where it is possible to see how the predictions perform compared to the actual survey results for each of the buildings. It should be noted that the data quantities were low for all buildings, resulting in reduced predicting performance (see Fig. 7).

The total performance of the models on unknown buildings is expressed by the MAE and RMSE metrics. The mean absolute error (MAE) for model M2.1 was 12%, while 19% for model M1.2. The RMSE was 0.17 for model M2.1 and 0.25 for model M1.2. This indicates that model M2.1 performs better than M1.2. It also indicates that the errors are smaller for the complaint data than for the satisfaction data. The sample sizes for this analysis are given in the caption of Fig. 7. Building 4 was the only building with more than 10 complaints on the survey day.

3.3. Control actions

3.3.1. Model performance

The "clm" function was also used to model each occupant’s TSV vote based on the frequency of heater use for the same person (work-desk). The amount of data was low as only data from one survey-day in Building 3 could be used. For this reason, only two models were tested when it comes to *Control actions*. The metrics to quantify the models’

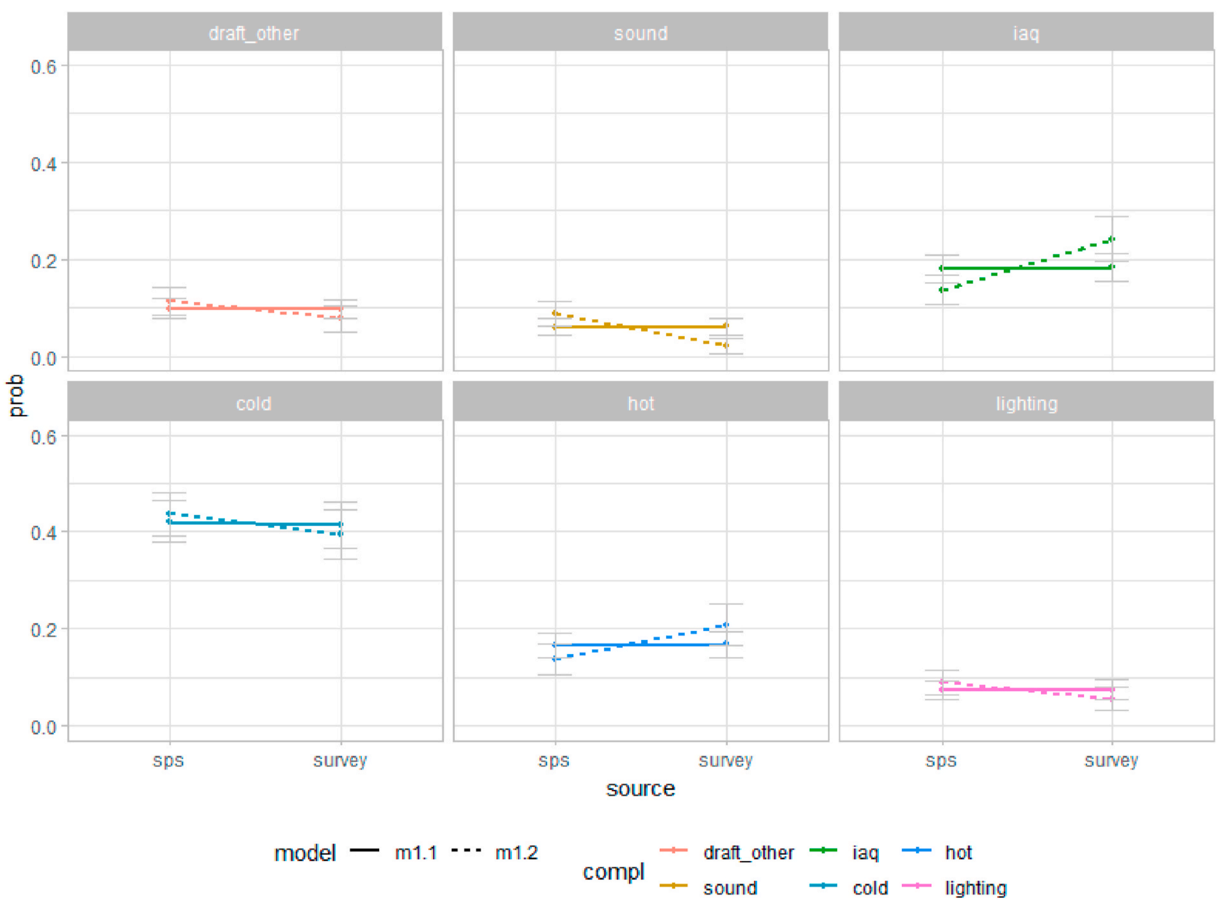

Fig. 5. Probabilities (y-axis) for receiving votes for each of the complaints on SPS and survey (x-axis) for all buildings combined. The standard error is given in error bars (grey). The conversion factor between SPS and Survey is shown as the slope of the line.

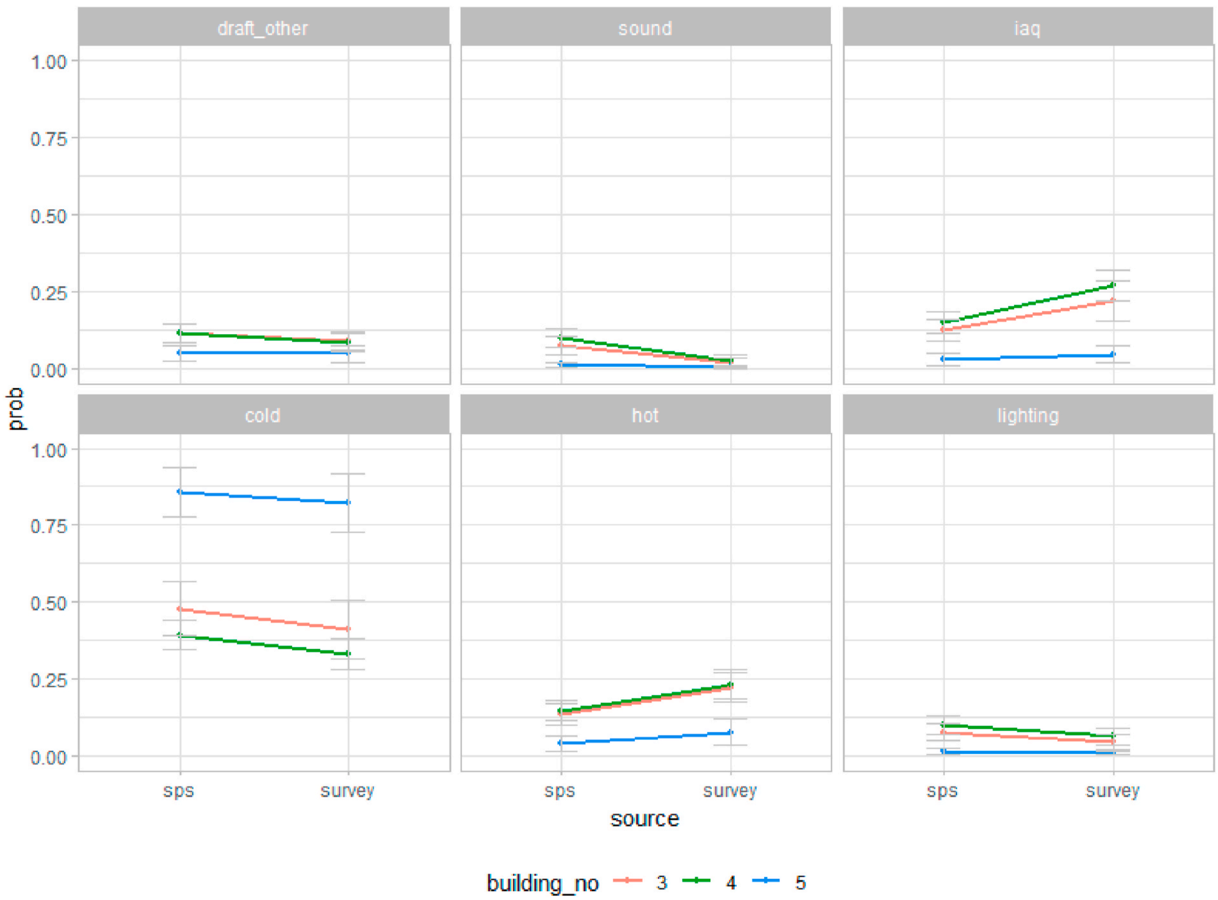

Fig. 6. Probabilities (y-axis) for receiving complaints for each of the complaints on SPS and survey (x-axis) per building. The standard error is given in error bars (grey). The conversion factor between SPS and Survey is shown as the slope of the line.

performance are listed in Table 6, showing the number of parameters in the model (no.par), likelihood ratio (Log Likelihood) performance, Akaike Information Criterion (AIC), and whether the model performs significantly better than the model listed above (Pr).

The results show that the model with location as an explanatory variable performs better. However, in this case we only have one data point per location for combination of TSV and usage. This means that there is no variance in the relationship between heater usage and TSV vote within each location (user), and the true difference between the models is not demonstrated. In a scenario where more data is collected, the model M2 should be preferred over model M1 as it is capable of quantifying the variance separately within one location as well as among all the locations.

3.3.2. Determination of important variables

Only model M1 could be analyzed further as M2 could not reach a prediction due to convergence issues and low data quantity. Fig. 8 displays predicted probability of occupants voting TSV vote −2 ("Cool"), −1 ("Slightly cool"), 0 ("Neutral") and 1 ("Slightly warm") on the Thermal Sensation Vote (TSV) scale (none of the occupants had voted −3 “Cold” or higher than 1 during the survey), given the number of heater control actions ranging from 1 to 15. The standard error is given with thin lines. The results show that the probability of voting the lowest votes (reporting to be "Cold" or "Cool") increases by frequency of heater usage. The probability of voting "Slightly cool" both increases and decreases, while the probability of voting “Neutral” or “Slightly warm” is reduced for occupants who have a high heater usage frequency. It should again be noted that the analysis is based on data from only one survey day consisting of 18 survey responses and a total of 44 heater activations. The results still demonstrate that there is a positive relationship between heater usage and cold thermal sensation votes.

3.4. Contextual reflections and take-home lessons

There are quite a few [9,19,21] studies that tested personal preference models based on occupant feedback and compared them to physical measurements of the indoor climate. However, there are no known studies that compare the collected data to simultaneously collected survey data. We previously investigated the qualitative and, to some extent, the quantitative performance of CSOF in other studies [4,25], but this study is at the best of our knowledge the first that investigates in a systematic way the reliability of the information derived from a CSOF data collection method. One study [30] where polling kiosk results were compared with survey results was conducted in a security checkpoint implementation, and it concluded that the results were highly correlated with results of a traditional usability survey, but that the dispersion of kiosk responses was significantly larger than that of the survey. Passing through a security checkpoint is a one-time event for the user, while experiencing the indoor climate of an office is a continuous or reoccurring event. For instance, the non-response bias was eliminated as each user could only use the kiosk once when passing. Due to this difference, it is assumed that the results of this previous study may not directly be applicable to an indoor climate application. Most building surveys during POE processes are also based on voluntary responses, but assume high response rates for statistical representativeness, thus neglecting the likelihood of the sample being biased. To our knowledge,

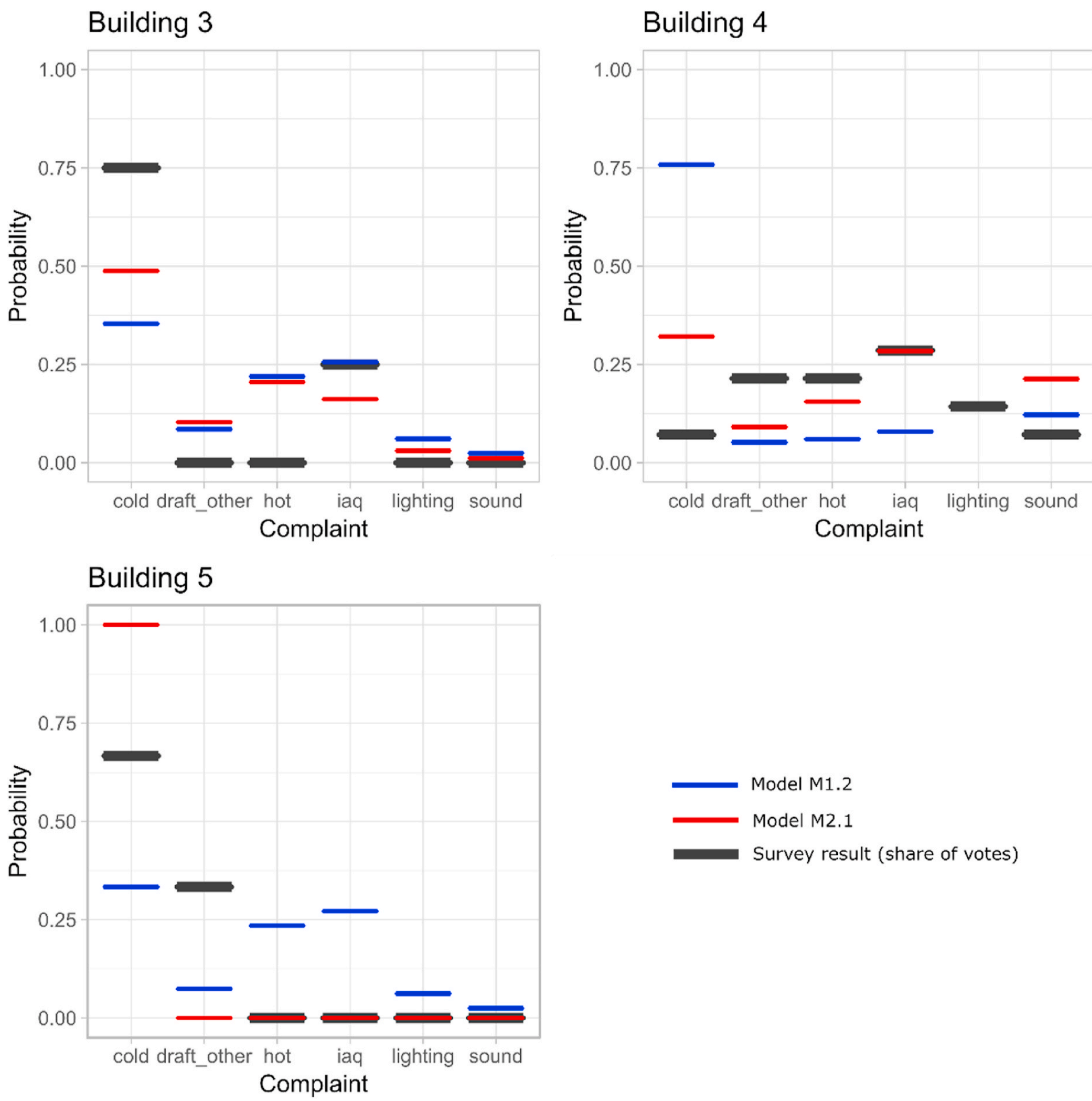

Fig. 7. Leave-one-out analysis for model M1.2 and model M2.1. Predictions based on model M1.2 (in blue) and predictions based on model M2.1 (in red) are compared to actual survey responses (in black) using the following sample sizes: for Building 3, 2 SPS and 4 Survey responses; for Building 4, 26 SPS and 14 Survey responses; for Building 5, 2 SPS and 3 Survey responses. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 6
Model performance by the Analysis of Deviance method where the P column states whether the model performs significantly better than the model listed above. Corresponding to $P > 0.05$ – ns (non-significant), $0.01 < p < 0.05$ *, $0.01 < p < 0.001$ **, $p < 0.0001$ ***.

	no.par	AIC	logLik	Pr(>Chisq)
M1	5	234.6564	−112.3282	ns
M2	40	214.549	−101.2746	***

the potential bias and sample representativeness of these studies has never been investigated closely.

The amount of collected data represents a limitation of the current study, as it is not sufficient to provide sound results for all system levels and the RQs. The results have been presented with the current data material rather than waiting for the possibility to collect more data after the Covid-19 pandemic that spread around the world in 2020. As mentioned, the survey is not performed in the same way in all buildings, which is likely to have resulted in inaccuracies in the data for Buildings 1 and 2. The data quantities of these two buildings do, however, only make up a small part (22%) of the total data, and the effect of this is expected to be small for the data set as a whole. Complaints were made in slightly different ways between Building 3 (several complaints could be made) and Buildings 4 and 5 (only one complaint could be made). The dataset has been investigated to assess in how many instances more than one complaint had been made at a time. It was found that this had happened in only 3–4 instances, so the effect of this difference is expected to be small.

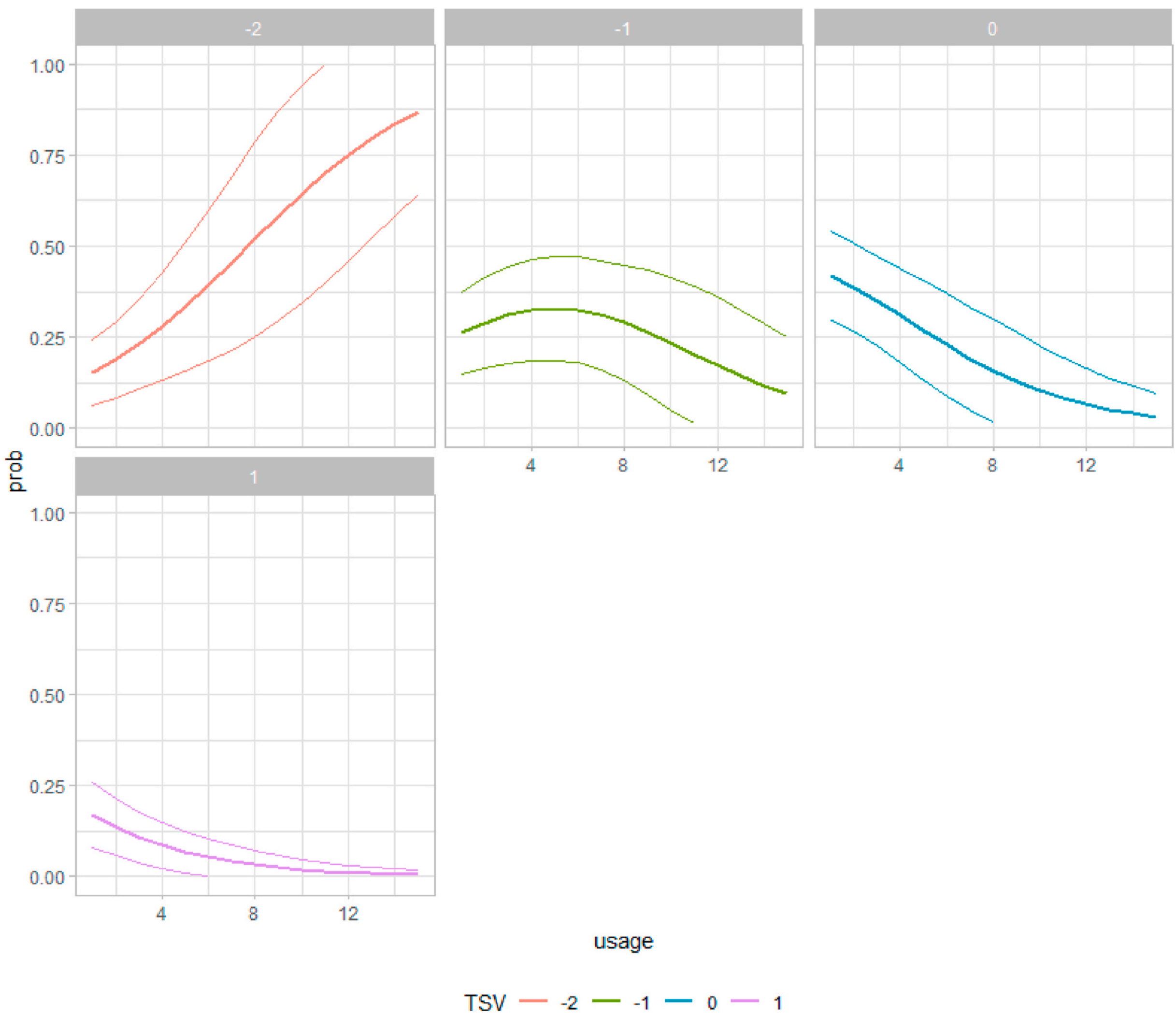

Fig. 8. Probability (y-axis) given by heater control actions per day (x-axis) for each of the 4 TSV votes used. Range of standard error is given by thin lines.

The study compares feedback and survey responses grouped by day, and thereby assumes that the occupants do not change their opinions between the time of the feedback and the survey. All surveys were done in the afternoon and occupants were asked to take the whole day (and in Building 1 and 2, the whole week) into consideration, a procedure that might have led to a certain mismatch between the information sampled through the CSOF and the one collected through the surveys.

For RQ1, (*What is the most suitable regression logic to predict survey results with feedback data?),* the investigation showed that cumulative link models modeling the feedback response (satisfaction or complaint) and Building ID as nominal data perform best for system levels 1 *(Occupant satisfaction)* and 2 *(Occupant complaints)*. The other variables did not have a significant impact on the model performance when included in the model. Level 3 (*Occupant control actions)* was best modeled using the model with location as an explanatory variable. The amount of data was too low in this case to investigate the effect of a random effect model.

For RQ2, (*What do the models and modeled coefficients tell us about the nature of the systems tested?*), we found that, for level 1, there is a significant non-response and multiple response effect where occupants who are neutral or satisfied do not vote at the SPS as often as people who are dissatisfied or very satisfied do. This effect is found for all buildings, although the effect is highly variable between buildings. This is evident by the fact that the Building ID variable was the most influential variable in the models. This variable represents all individual factors for the building, and even more importantly, the occupants. This shows that the subjective feedback devices are used differently in each building, and direct comparison of feedback results between buildings is not advised until more information about the reasons for independent results in different buildings can be unveiled. The bias is found to be much smaller for level 2. For level 3, heater usage frequency was found to be related to TSV vote, where those who made significant use of the heater were more likely to perceive the environment as cold. In this case, the work-desk ID was the most influential factor, and this means that occupants used the heaters differently and feedback data should not be compared directly to feedback from other occupants.

For RQ3 (*What prediction accuracy can be achieved for predicting survey results with SPS smiley, SPS complaint or heater use data using these modeling methods?*) we found a Mean Absolute Error (MAE) of 16% for level 1 and 12% for level 2. There was not enough data to evaluate this for level 3. It should be noted that these values were calculated using only one survey-day per building (the one with most feedback data was chosen) and the results are somewhat dependent on which survey days were chosen for calculation. Data quantities were low for several of the survey-days, and they were not sufficient for providing a sound conclusion for this research question, although some reasonable trends could be seen.

4. Conclusions

This study concludes a wider project that investigated the functioning and validity of CSOF systems, where continuous subjective occupant feedback from field tests is studied by comparison to survey results, measurements of physical indoor climate and results of focus interviews [4,25]. In this last investigation we show that it is possible to model the expected results of satisfaction evaluations in a survey using instead autonomously collected votes from three kinds of feedback.

The conversion factors between SPS and survey were found to be variable between the different levels, and large corrections are needed for some of the satisfaction levels. The predicting accuracy given by the Mean Absolute Error (MAE) for "new" buildings (where the building factor has not been modeled) is found to be 16%. The conversion factors between SPS and survey complaints varied in the same way, but the biases were smaller. The predicting MAE was 12%. The differences between buildings are also small, making predictability higher in new buildings. Data quantities for control actions were very low, but we found a relationship showing that occupants with a high frequency of heater control actions also have a higher probability of responding "cool" or "cold" when asked to give their Thermal Sensation Vote (TSV) in the survey. It was found that the most defining factors for model performance was the Building ID and the work-desk ID (for control actions). The other parameters tested (intervention, response rate, and survey ID) were found not to have a significant impact.

The results presented in this study can be useful in further uncovering the reliability and meaning of data collected in CSOF systems. The method shown for modeling and data treatment in this investigation are recommended for use in other studies to understand the relations between different sources of occupant centric data. Furthermore, we hope that the results herewith presented will be valuable for the practical application of the systems tested. Knowledge of the validity and nature of the collected data, which we have started uncovering in this research, will be crucial input when using the CSOF systems in building design, control, tuning or benchmarking. However, more data should be collected and analyzed in order to strengthen the validity of the models before one with full confidence can use continuously collected occupant feedback instead of surveys.

Future studies that follow up on the data gathering of this study are recommended, using methods that are similar to or refine the ones used in this project. A larger dataset would make it possible to continue and elaborate on the modeling analysis, and would make it possible to reach sound conclusions and useful knowledge for increasing the trustworthiness of CSOF data. Similar studies should also be performed for other occupant feedback systems and collected information to gain a greater understanding of the validity of several types of continuous feedback data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study has been performed with the financial support of the Research Council of Norway, within the project "Methods for real-time user involvement of indoor climate in smart buildings" (project number: 277048). The support provided by the Research Council of Norway, Skanska Norway, DTU Compute and NTNU is gratefully acknowledged.

Appendix A. Description of experiments

4.4.1. Case buildings

A short description of each case building is give below.The buildings have been more closely described in earlier publications [25,31].

- Building 1 is an open plan office outside Oslo, Norway with 26 personal workspaces for occupants performing tasks within engineering consulting. The building had recently been refurbished to a modern passive house and BREEAM Outstanding standard.

- Building 2 consists of two team offices in central Oslo, Norway with 12 workplaces for occupants performing IT consulting tasks. The building is of a modern and energy -effective standard.
- Building 3 is an open plan office in central Oslo, Norway with 50 personal workspaces for occupants performing tasks within engineering consulting. It is estimated that about 25 employees work on this floor on an average day. The building is of a modern and energy -effective standard.
- Building 4 is an open plan office in Berkeley, California with approximately 200 personal cubicles for occupants performing mainly doing administrative tasks related to human resources and accounting. It is estimated that about 95 employees work on this floor on an average day. The building is an old industrial building that has been retrofitted to a low standard office building.
- Building 5 is an open plan office in central San Francisco, California with 50 personal workspaces for occupants performing tasks within engineering. It is estimated that about 25 employees work on this floor on an average day. The building was built in the 1970s, but the floor has been retrofitted in recent years to a high standard.

4.4.2. Experiment procedures

The experiments were all performed in a similar manner, but the procedures were not all identical. The three components of the feedback system were introduced gradually, one by one, and occupant surveys were performed at certain points in time. So-called temperature interventions (deliberate stepwise changes in the ambient room temperature) were carried out at certain points in time. The temperature interventions were performed in three of the buildings with the intention of provoking feedback and control actions from the occupants to see whether they would use the system. The occupants were informed that the study was investigating the indoor climate in their space, but it was not specified that the focus of the tests was about the interactions they made with the feedback system. It was also communicated that no data linked to personal information or identities would be gathered during the experiments, and hence, it was unnecessary to apply for permissions from the Norwegian Centre For Research Data according to the current guidelines.

The experiments are not all performed identically. All levels were not tested in all buildings. Building 1 and 2 have surveys where occupants were asked to answer for the previous week. These responses are still compared to feedback for the same day only. The indoor and outdoor climate conditions were different. The studies were performed on different occupant groups and cultures. The study and feedback equipment may have been communicated differently in the different studies, although it was sought to do this similarly in all buildings.

The field tests were conducted as longitudinal blind tests in five real office environments. Two of the field experiments were conducted in the Oslo-area, Norway during the fall 2018 and winter 2019, two studies were performed in the San Francisco area, California during spring and summer of 2019, and the last study was conducted in Oslo, Norway during late winter 2020, until it was prematurely stopped due to the Covid-19 pandemic in March 2020.

4.4.3. Feedback methods

Occupant satisfaction polling station (SPS).

- The intention of the SPS is to capture voluntary votes related to the occupants' general satisfaction with the indoor climate in the space.
- The data collection method was intended to be non-intrusive and voluntary, but in oder to increase the amount of data the feedback method should not be time consuming for the occupant.
- In the context of this study, where the length and number of questions we can ask at the polling station is constrained, we prioritize asking the occupant for their satisfaction level as it is the "summary state" of subjective evaluation. If the occupant reports to be satisfied with the indoor climate, there is no need for further investigations.
- Although the use of smiley-face polling stations has had a rapid growth for certain applications, we only found one study assessing the accuracy of polling station results, ti was conducted in a security checkpoint implementation and it concluded that the results were highly correlated with results of a traditional usability survey [30]. Several studies have shown how crucial the aspects of usability and adequate interfaces are to collect high frequency occupant data [32,33].
- A Satisfaction Polling Station (SPS) was developed in the form of a webpage displayed on a tablet computer mounted on a stand. See Figure 9.
- As occupants pressed buttons on the touchscreen, the response was saved in a database as integers between −2 and 2 where −2 is "Angry", 0 is "Neutral" and 2 is "Happy". If one of the three right buttons (0–2) were pressed, a *"Thank you for voting!"* screen appeared before the screen returned to the front page. If one of the two left buttons were pressed (−2 or −1),
- There are three known instances where polling stations have been tested for continuous assessment of satisfaction with indoor climate in buildings [22,23,34]. One of the studies investigated personal polling stations, while the other two investigated a public satisfaction polling station (SPS) with smiley face ratings. No studies have assessed the accuracy of SPSs with the feedback from other survey types. To the best of our knowledge, this analysis has not been done with POE surveys either, and the magnitude of non-response bias in these types of surveys is also unknown.

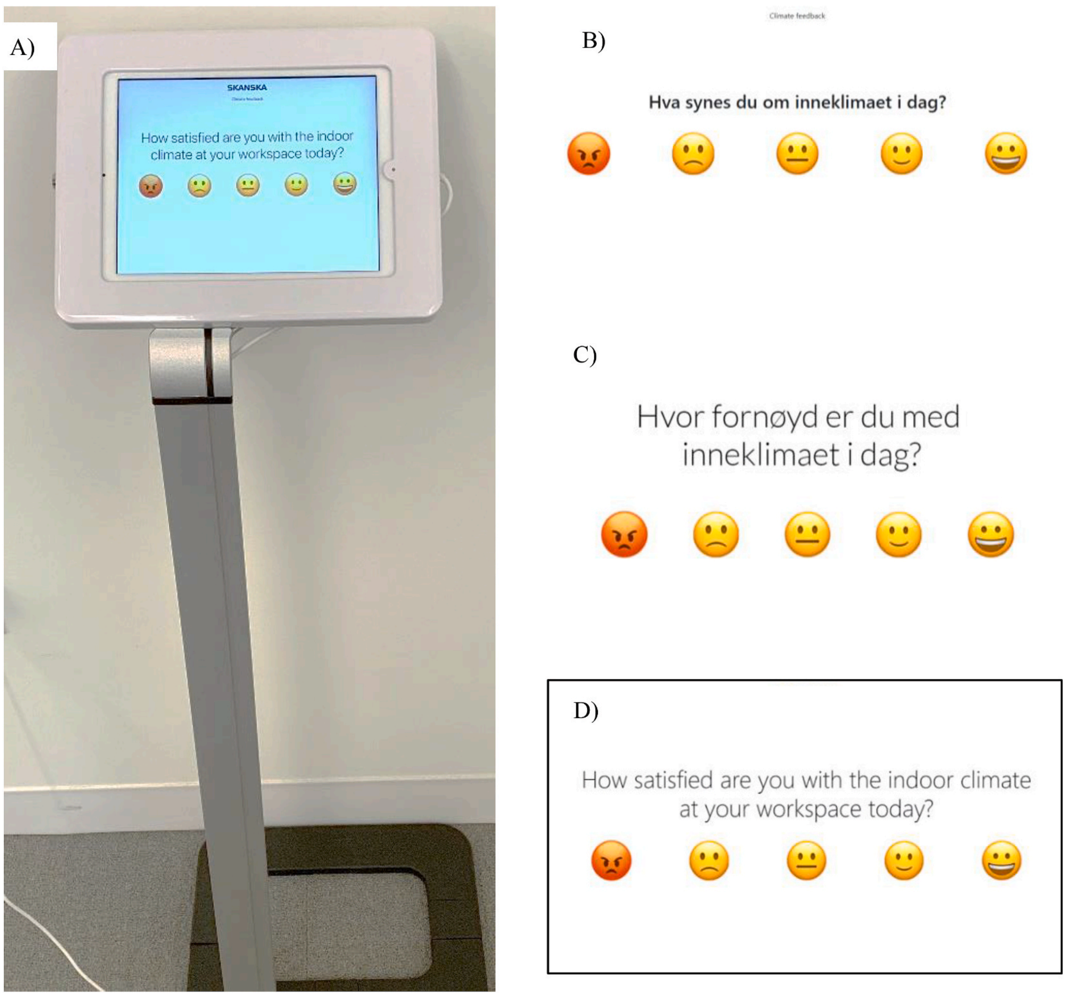

Fig. 9. A) Picture of SPS in Building 4. B) SPS front page in Building 1 and 2 (Norwegian) C) SPS front page in Building 3 (Norwegian) D) SPS front page in building 4 and 5.

Occupant complaints.

- The intention of the complaint feature is to collect complaints from dissatisfied voters.
- Occupant complaints were collected using the second page of SPS in Buildings 3–5. Complaints in buildings 1 and 2 were collected in a different manner and this data is not included in this study.
- If one of the two left satisfaction buttons were pressed (−2 or −1), a second screen appeared with the text *"Please help us pinpoint the problem"* followed by seven buttons with the following statements: *"Too hot"*, *"Too cold"*, *"Draft"*, *"Air quality issues"*, *"Sound issues"*, *"Lighting issues"* and *"Other"*. Only one of the statements could be selected in Building 4 and 5, while more than one could be selected in Building 3. All responses were stored in a database. The available response buttons were selected based on the researchers' experience with known common occupant complaints. It was desirable to have two response alternatives for the thermal to be able to study the sensitivity to environmental changes. After selecting an issue, the *"Thank you"* screen appeared.

A)

B)

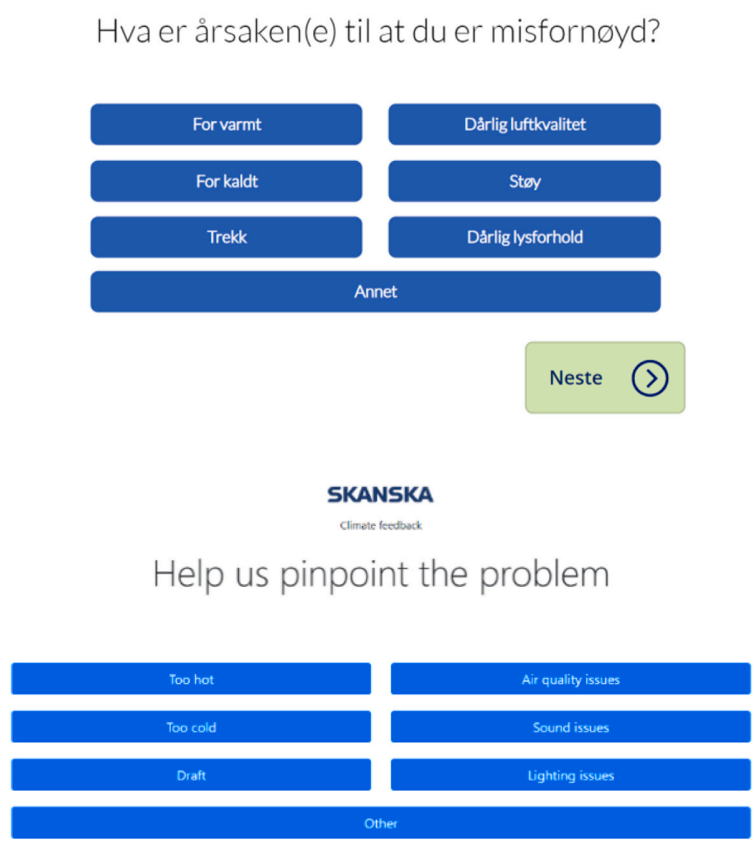

Fig. 10A) SPS Complaint page in Building 3 (Norwegian, traslates to "What is/are the reason(s) for you dissatisfaction? B) SPS Complaint page in Buildings 4 and 5.

Control actions.

- The intention of the *Control action* feature was to collect information of users personal control actions, assuming that a control action can be seen as a wish for change [21].
- Under desk heaters with a wireless switch located on each office desk in Buildnig 3. The buttons had a built-in timer set to 30 min and were paired via the 433 MHz radio band with the Smart Plug at the same desk. When occupants pressed the button on their desk and complained "Too cold", the smart plug under the desk would turn on and start the heater. Users would then have to repeat the procedure to continue receiving heat after the programmed 30 min. This was both a security measure to avoid heaters being left on and a way of increasing the number of responses.
- The heaters are 30 × 60 cm large, attach to the underside of the desk and provide infrared heating with a power range of 40–150 W (see Figure 11 A and B), but were pre-set to approximately 50 W power, producing surface temperatures of 40–50 °C.

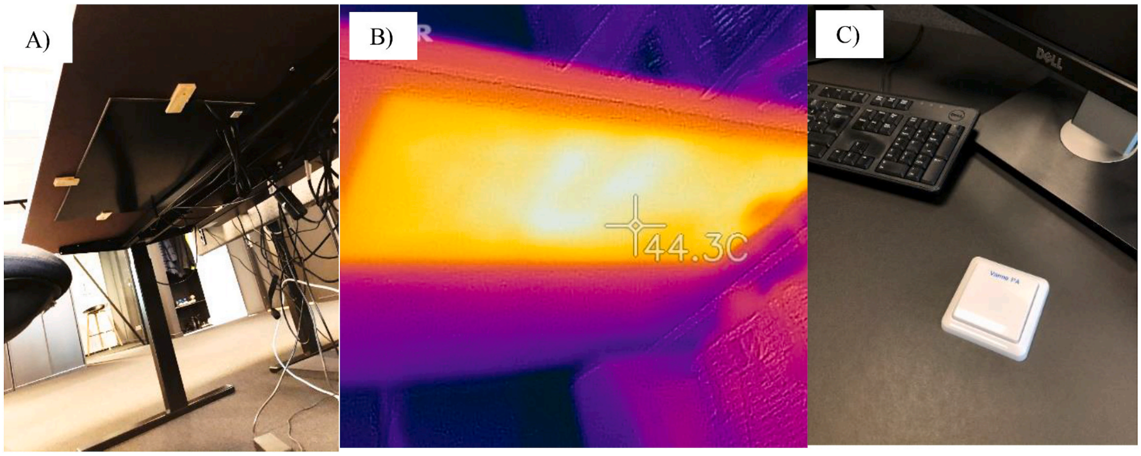

Fig. 11. Control action system A) Under desk heater mounted in Building 3. B) Infrared image of under desk heater in use. C) Wireless "ON" button with 30-min timer on desk.

4.4.4. Survey method. Occupant surveys in Building 1 and 2 were performed differently from in Building 3–5. The survey questions from all buildings are given in Appendix B.

In Buildings 1 and 2, an electronic survey was distributed to the test subjects via an email link. The survey asked the identical same question as was asked on t the *Satisfaction evaluation SPS.* The survey also included several other questions not relevant for this study. The respondents were asked to take the last week into consideration when answering all the questions. The surveys in Building 3–5 were performed differently in order to increase the response rate of the surveys and to make survey answers representative of current day perceptions so they would be directly comparable to the SPS feedback. Each occupant present at the time of survey was personally approached by a researcher and asked to fill out a short survey on a tablet. Occupants filled out the survey themselves while the researcher took a step back. In this case, the occupants were asked to take the current day into consideration when answering. The surveys were always performed in the afternoon, between 2 and 4 p.m. The identical questions as asked on the SPS, both satisfaction evaluation and complaints, were asked in the survey (only those who replied to be dissatisfied were routed to the complaint question). Further, several classic questions commonly found in POE's were asked. Relevant for this study is question 12: "How do you feel about the temperature of your workplace?" where occupants could respond [Hot/Warm/Slightly warm/Neither/Slightly cool/Cool/Cold] on a 7-point scale.

Appendix B. Survey questions

Table 7
Survey questions buildings 1 and 2, translated to English from Norwegian language.

Question	Response alternatives
Q1 – How satisfied are you with the indoor climate at your workplace today?	[5 smiley face buttons]
Q2 – On a 7-point scale, how satisfied are you with the indoor climate at your workplace today?	[7-point slider from "Very dissatisfied" to "Very satisfied"]
Q3 – How acceptable did you find the temperature of your workspace during the period?	[Acceptable/Barely acceptable/Barely unacceptable/Unacceptable]
Q4 – How acceptable did you find the air quality of your workspace during the period?	Acceptable/Barely acceptable/Barely unacceptable/Unacceptable]
Q5 – How do you feel about the temperature of your workspace?	[Hot/Warm/Slightly warm/Neither/Slightly cool/Cool/Cold]
Q6 – Have you experienced it to be colder than what you think is acceptable during the period?	[Yes/No]
Q7 – Have you experienced it to be warmer than what you think is acceptable during the period?	[Yes/No]
Q8 – Have you experienced being so cold it interfered with you work tasks during the period?	[Yes/No]
Q9 – Have you experienced being so warm it interfered with you work tasks during the period?	[Yes/No]
Q10 – Which level of control you perceive to have over your indoor climate?	[No control/Little control/Some control/Much control]
Q11 – Have you used the heater that is mounted under your desk during the period?	[Yes/No]
Q12 – If not, what was the reason for not using the heater?	[I am content – I have no need for extra heat/I was not aware that it was there/I don't understand how it works/I find it too cumbersome to use/It's not working/I don't know/Other]
Q13 – Have you used the smiley-face kiosk located by the entrance during the period?	[Yes/No]
Q14 – If not, why?	[I don't think it works/I find it too time consuming/I don't understand the point of giving feedback/I don't know]
Q15 – Would you, on a regular basis, prefer more information regarding your indoor climate (for instance information about temperature and air quality on a screen by the entrance)?	[Yes/No/Other]
Q16 – Please submit other comments if you wish.	[Text]

Table 8
Survey questions building 3, translated to English from Norwegian language.

Topic	Question	Response alternatives
Metadata (inserted by researcher)	Q1 - Workplace ID	[Text]
	Q2- Approximate age	[Years, binned]
	Q3 - Sex	[Male/Female]
	Q4 – Workplace type	[Open plan, cubicle, single office, Team office]
	Q5 – Workplace comments	[Text]
	Q6 – Other comments	[Text]
SPS questions	Q7 – How satisfied are you with the indoor climate at your workplace today?	[5 smiley face buttons]
	Q8 – Help us pinpoint the problem (if dissatisfied)	[Too hot/Too cold/Draft/Air quality issues/Sound issues/Lighting issues/Other
	Q9 – Please specify the problem(s) (if chosen Other)	[Text]
POE questions	Q10 – How satisfied are you with the temperature of your workspace today?	[Very satisfied/Satisfied/Somewhat satisfied/Neither satisfied nor dissatisfied/Somewhat dissatisfied/Dissatisfied/Very dissatisfied]
	Q11 – How satisfied are you with the air quality of your workspace today?	[Very satisfied/Satisfied/Somewhat satisfied/Neither satisfied nor dissatisfied/Somewhat dissatisfied/Dissatisfied/Very dissatisfied]
	Q12 – How do you feel about the temperature of your workspace?	[Hot/Warm/Slightly warm/Neither/Slightly cool/Cool/Cold]
SPS voting habits	Q13 – How often do you vote at the smiley kiosk?	[Never/Once since it was introduced/A few times sporadically/Regularly each week/Regularly once per day/regularly several times per day]
Perceived control	Q14 – Which level of control you perceive to have over your indoor climate?	[No control/Little control/Some control/Much control]

Table 9
Survey questions building 4 & 5

Topic	Question	Response alternatives
Metadata (inserted by researcher)	Q1 - Workplace ID	[Text]
	Q2- Approximate age	[Years, binned]
	Q3 - Sex	[Male/Female]
	Q4 – Workplace type	[Open plan, cubicle, single office, Team office]
	Q5 – Workplace comments	[Text]
	Q6 – Other comments	[Text]
SPS questions	Q7 – How satisfied are you with the indoor climate at your workplace today?	[5 smiley face buttons]
	Q8 – Help us pinpoint the problem (if dissatisfied)	[Too hot/Too cold/Draft/Air quality issues/Sound issues/Lighting issues/Other
	Q9 – Please specify the problem(s) (if chosen Other)	[Text]
POE questions	Q10 – How satisfied are you with the temperature of your workspace today?	[Very satisfied/Satisfied/Somewhat satisfied/Neither satisfied nor dissatisfied/Somewhat dissatisfied/Dissatisfied/Very dissatisfied]
	Q11 – How satisfied are you with the air quality of your workspace today?	[Very satisfied/Satisfied/Somewhat satisfied/Neither satisfied nor dissatisfied/Somewhat dissatisfied/Dissatisfied/Very dissatisfied]
	Q12 – How do you feel about the temperature of your workspace?	[Hot/Warm/Slightly warm/Neither/Slightly cool/Cool/Cold]
SPS voting habits	Q13 – How often do you vote at the smiley kiosk?	[Never/Once since it was introduced/A few times sporadically/Regularly each week/Regularly once per day/regularly several times per day]

References

[1] P. Li, T.M. Froese, G. Brager, Post-occupancy evaluation: state-of-the-art analysis and state-of-the-practice review, Build. Environ. 133 (2018) 187–202, https://doi.org/10.1016/j.buildenv.2018.02.024.
[2] P. Roberts, Who is post-occupancy evaluation for? Build. Res. Inf. 29 (2001) 463–465, https://doi.org/10.1080/09613210110072674.
[3] K. Hadjri, C. Crozier, Post-occupancy evaluation: purpose, benefits and barriers, Facilities 27 (2009) 21–33, https://doi.org/10.1108/02632770910923063.
[4] N. Lassen, T. Josefsen, F. Goia, Design and in-field testing of a multi-level system for continuous subjective occupant feedback on indoor climate, Build. Environ. 189 (2020) 1–58, https://doi.org/10.1016/j.buildenv.2020.107535.
[5] F. Jazizadeh, A. Ghahramani, B. Becerik-Gerber, T. Kichkaylo, M. Orosz, User-led decentralized thermal comfort driven HVAC operations for improved efficiency in office buildings, Energy Build. 70 (2014) 398–410, https://doi.org/10.1016/j.enbuild.2013.11.066.
[6] A. Hang-yat, D. Wang, Carrying my environment with Me : a participatory-sensing approach to enhance thermal comfort, in: BuildSys '13, 2013, https://doi.org/10.1145/2528282.2528286.
[7] D.A. Winkler, A. Beltran, N.P. Esfahani, P.P. Maglio, A.E. Cerpa, FORCES: feedback and control for occupants to refine comfort and energy savings, in: UbiComp 2016 - Proc. 2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput, Association for Computing Machinery, Inc, 2016, pp. 1188–1199, https://doi.org/10.1145/2971648.2971700.
[8] K. Konis, L. Zhang, Occupant-aware energy management: simulated energy savings acheivable through application of temperature setpoints learned through end user feedback, in: ASHRAE IBPSA-USA SimBuild 2016 Build. Perform. Model. Conf., 2016, pp. 283–289. www.ashrae.org.
[9] F. Jazizadeh, A. Ghahramani, B. Becerik-Gerber, T. Kichkaylo, M. Orosz, Human-building interaction framework for personalized thermal comfort-driven systems in office buildings, J. Comput. Civ. Eng. 28 (2014) 2–16, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000300.
[10] S. Purdon, B. Kusy, R. Jurdak, G. Challen, Model-free HVAC control using occupant feedback, in: Proc. - Conf. Local Comput. Networks, LCN, 2013, https://doi.org/10.1109/LCNW.2013.6758502.
[11] F. Jazizadeh, A. Ghahramani, B. Becerik-Gerber, T. Kichkaylo, M. Orosz, Personalized thermal comfort driven control in HVAC operated office buildings, in: 218 Comput. Civ. Eng., 2013, pp. 218–225.
[12] B. Balaj, H. Teraoka, R. Gupta, A. Yuvraj, ZonePAC: zonal power estimation and control via HVAC metering and occupant feedback bharathan, in: BuildSys '13, ACM, 2013.
[13] V.L. Erickson, A.E. Cerpa, Thermovote: participatory sensing for efficient building HVAC conditioning, in: BuildSys '12 Proc. Fourth ACM Work. Embed. Sens. Syst. Energy-Efficiency Build., 2012, pp. 9–16.
[14] F. Jazizadeh, B. Becerik-Gerber, Toward adaptive comfort management in office buildings using participatory sensing for end user driven control, in: Buildsys'12, ACM, 2012, pp. 1–8.
[15] S.K. Gupta, S. Atkinson, I. O'Boyle, J. Drogo, K. Kar, S. Mishra, J.T. Wen, BEES: real-time occupant feedback and environmental learning framework for collaborative thermal management in multi-zone, multi-occupant buildings, Energy Build. 125 (2016) 1–13, https://doi.org/10.1016/j.enbuild.2016.04.084.

[16] M. Pritoni, K. Salmon, A. Sanguinetti, J. Morejohn, M. Modera, Occupant thermal feedback for improved efficiency in university buildings, Energy Build. 144 (2017) 241–250, https://doi.org/10.1016/j.enbuild.2017.03.048.

[17] Q. Zhao, Y. Zhao, F. Wang, Y. Jiang, F. Zhang, Preliminary study of learning individual thermal complaint behavior using one-class classifier for indoor environment control, Build. Environ. 72 (2014) 201–211, https://doi.org/10.1016/j.buildenv.2013.11.009.

[18] M. Rittenbruch, J. Donovan, Y. Santo, Evaluating the use of ambient and tangible interaction approaches for personal indoor climate preferences, in: UbiComp 2014 - Adjun. Proc. 2014 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput., Association for Computing Machinery, Inc, 2014, pp. 159–162, https://doi.org/10.1145/2638728.2638755.

[19] Y. Peng, Z. Nagy, A. Schlüter, Temperature-preference learning with neural networks for occupant-centric building indoor climate controls, Build. Environ. 154 (2019) 296–308, https://doi.org/10.1016/j.buildenv.2019.01.036.

[20] A.T. Alan, M. Shann, E. Costanza, S.D. Ramchurn, S. Seuken, It is too hot: an in-situ study of three designs for heating, in: Conf. Hum. Factors Comput. Syst. - Proc., Association for Computing Machinery, 2016, pp. 5262–5273, https://doi.org/10.1145/2858036.2858222.

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

[22] J. Berquist, M. Ouf, W. O'Brien, A method to conduct longitudinal studies on indoor environmental quality and perceived occupant comfort, Build. Environ. 150 (2019) 88–98, https://doi.org/10.1016/j.buildenv.2018.12.064.

[23] K.S. Konis, Leveraging ubiquitous computing as a platform for collecting real-time occupant feedback in buildings, Intell. Build. Int. 5 (2013) 150–161, https://doi.org/10.1080/17508975.2013.781499.

[24] N. Lassen, F. Goia, A theoretical framework for classifying occupant-centric data streams on indoor climate using a physiological and cognitive process hierarchy, Energy Build. 241 (2020) 1–36, https://doi.org/10.1016/j.enbuild.2021.110935.

[25] N. Lassen, F. Goia, S. Schiavon, J. Pantelic, Field investigations of a smiley-face polling station for recording occupant satisfaction with indoor climate, Build. Environ. 185 (2020) 107266, https://doi.org/10.1016/j.buildenv.2020.107266.

[26] M. Schweiker, X. Fuchs, S. Becker, M. Shukuya, M. Dovjak, M. Hawighorst, J. Kolarik, Challenging the assumptions for thermal sensation scales, Build. Res. Inf. 45 (2017) 572–589, https://doi.org/10.1080/09613218.2016.1183185.

[27] R.H.B. Christensen, Ordinal—Regression Models for Ordinal Data. R Package Version 2019.12-10, 2019, p. 54. https://cran.r-project.org/package=ordinal.

[28] H. Madsen, P. Thyregod, Introduction to General and Generalized Linear Models, Taylor and Francis Inc., 2011.

[30] L.R. Rabbitt, J.A. Hasselgren, C. Cook, Y.B. Sirotin, Measuring satisfaction with standard survey instruments and single-button responses on kiosks, Proc. Hum. Factors Ergon. Soc. Annu. Meet. (2018), https://doi.org/10.1177/1541931218621325.

[31] N. Lassen, T. Josefsen, F. Goia, Design and In-Field Testing of a Multi-Level System for Continuous Subjective Occupant Feedback on Indoor Climate - Submitted in Revised Version to Building and Environment in October 2020, Ref. Nr, 2020. BAE-S-20-03537.

[32] S. Karjalainen, O. Koistinen, User problems with individual temperature control in offices, Build. Environ. (2007), https://doi.org/10.1016/j.buildenv.2006.10.031.

[33] T. Peffer, M. Pritoni, A. Meier, C. Aragon, D. Perry, How people use thermostats in homes: a review, Build. Environ. 46 (2011) 2529–2541, https://doi.org/10.1016/j.buildenv.2011.06.002.

[34] N. Lassen, T. Josefsen, Experimental setup and testing of an in-field system for real-time occupant feedback, IAQVEC, IOP Conf. Ser. Mater. Sci. Eng. 609 (2019) 1–6, https://doi.org/10.1088/1757-899X/609/4/042045.