



KNN and adaptive comfort applied in decision making for HVAC systems

Pablo Aparicio-Ruiz¹ · Elena Barbadilla-Martín¹ · José Guadix¹ · Pablo Cortés¹

Published online: 4 December 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

The decision making of a suitable heating, ventilating and air conditioning system's set-point temperature is an energy and environmental challenge in our society. In the present paper, a general framework to define such temperature based on a dynamic adaptive comfort algorithm is proposed. Due to the fact that the thermal comfort of the occupants of a building has different ranges of acceptability, this method is applied to learn such comfort temperature with respect to the running mean temperature and therefore to decide the suitable range of indoor temperature. It is demonstrated that this solution allows to dynamically build an adaptive comfort algorithm, an algorithm based on the human being's thermal adaptability, without applying the traditional theory. The proposed methodology based on the K-Nearest-Neighbour algorithm was tested and compared with data from an experimental thermal comfort field study carried out in a mixed mode building in the south-western area of Spain and with the Support Vector Machine method. The results show that K-Nearest-Neighbour algorithm represents the pattern of thermal comfort data better than the traditional solution and that it is a suitable method to learn the thermal comfort area of a building and to define the set-point temperature for a heating, ventilating and air-conditioning system.

Keywords Adaptive comfort · K-Nearest Neighbour · Algorithm · Buildings · HVAC · SVM

1 Introduction

Systems based on a fixed set-point temperature is the simplest way to control indoor temperature. The major limitation in those systems is that the user has to manually readjust the set-point depending on the variations in external temperature conditions. The thermostat compares the temperature set by the user with the temperature of the room. If the temperature in the room achieves the temperature set by the user, it disconnects the power to the actuator which closes the circuit(s). The first solution to face such limitation is the installation of thermostatically-controlled electrothermic actuators, where the system would automatically make a decision about the set-point (Aparicio-Ruiz et al. 2018). In heating, ventilating and

✉ Pablo Aparicio-Ruiz
pabloaparicio@us.es

¹ Grupo de Ingeniería de Organización, Escuela Técnica Superior de Ingeniería, Universidad de Sevilla, Camino de los Descubrimientos S/N, 41092 Seville, Spain

air conditioning systems, energy consumption is directly related to high (low) set-point temperatures when the heating, ventilating and air conditioning system (HVAC) is working in the heating mode (or in the cooling mode). Due to this fact, the decision making regarding the mode of operation of HVAC systems is a research area with an important growth during the last years (Rupp et al. 2015). So, currently, one of the main missions or objectives that has been launched around the world is optimising such set-point temperatures (Tuohy et al. 2010; Si et al. 2015).

Particularly in Spain, the existing regulation (Royal Decree 486/1997) proposes an indoor temperature which ranges from 17 to 27 °C for buildings where sedentary work takes place and from 14 to 25 °C for buildings where light physical work takes place. It also establishes that this criteria should be applied taking into account the limitations or conditions resulting from the particular characteristics of the workplace, the processes that take place and the climate where the building is located. Moreover, the Royal Decree 1826/2009 proposes a maximum and a minimum temperature for administrative, commercial and public buildings: 26 °C maximum for refrigerated spaces and 21 °C minimum for heated enclosures ones. Generally speaking, the recent standards ISO 15251 (ISO Standard 2008) and other references, e.g., the guide CIBSE (2015) and ASHRAE (2013) remark that the users' behaviour regarding the comfort temperatures is dynamic and the concept of comfort varies over the year because users adapt to fluctuations in outside temperatures. Therefore the energy consumption is influenced by meteorological conditions. The temperature pattern is one of the main factors of energy usage, so electric companies analyse the weather conditions in their models, which quantified the relationship between temperature and energy sales. In (Valor et al. 2001) the relationship between electricity load and daily air temperature in Spain is analysed.

Many HVAC systems are designed based on a fixed set-point temperature although it is not the most efficient energy policy. The main objective in the area of thermal comfort and energy consumption is finding a solution which balances the energy consumption without compromising the comforts of the buildings' occupants. Initially, most authors based their studies on the predicted mean vote (PMV) index and predicted percentage dissatisfied (PPD) index (Hussein and Rahman 2009) for doing that. Nevertheless, during the last decades there has been a growing number of thermal comfort studies based on the adaptive comfort theory, which highlights that the thermal history and other contextual factors change the occupant's thermal expectations and preferences (de Dear et al. 1998). The adaptive comfort tries to soften the barriers between indoor and outdoor comfort through the definition of an adaptive control algorithm (ACA) which relates the comfort temperature of the occupants of a building with the daily outdoor temperature. Based on the adaptive comfort theory, differences in weather conditions, seasons and cultural aspects have been taken into account (Barbadilla-Martín et al. 2017) as well as types of buildings: naturally ventilated (NV), fully air-conditioned (AC) or hybrid (mixed mode, NV + AC) (de Dear et al. 1998).

In environmental control systems the building's occupants play little or no role regarding the indoor comfort conditions, but the green building movement tries to reverse this tendency with an improvement of comfort based on designing priorities in orientation, building mass, flues, operable windows, etc. There has also been a tendency to deploy mixed-mode approaches where the occupants use the passive solution to obtain their comfort and when these solutions are not enough, they use automated technologies like HVAC due to the fact that the interaction between the inhabitants and architecture can lead to important energy and carbon reductions (Cole et al. 2010). In hybrid or mixed mode buildings, occupants are more directly involved with building systems and operations through closing/opening doors or windows, turning on-off lights or thermostats, shading devices, vents and other man-

ual controls. Regarding thermal comfort, the PLEA manifesto (Cole et al. 2010) highlights two essential directives: “Buildings should provide their inhabitants with multiple adaptive opportunities optimizing health, satisfaction and productivity” and “Inhabitants should be responsible to take an ‘active’ role for the provision of relative comfort using robust ‘passive and low-energy’ strategies”. Although the adaptive thermal comfort theory proposes the definition of an adaptive control algorithm which depends on the outdoor temperature, it does not depend dynamically on the occupants nor learns their opinion continuously. The present paper exposes a methodology to develop a dynamic adaptive comfort system (DACS) for defining the comfort temperature based on the outdoor temperature and the occupants’ thermal sensation. This DACS would have influence on the energy consumption and it is a feasible solution to define a specific adaptive comfort zone for each building. In Ahmad et al. (2016), a complete review of computational intelligence techniques for HVAC systems is shown. The review reveals that the minimisation of energy consumption, the optimisation of thermal comfort, indoor air quality and occupant preferences are the main research areas, and a classification of computational techniques is shown, where different techniques are used for prediction, optimization, control and diagnosis. In the case of prediction, the principal techniques are artificial neural networks and Support Vector Machine (SVM) but the adaptive comfort is not considered in such review. Others studies in the field of class predictions dynamically, it have focused on finding the neighborhood of a new data instance and choosing the most competent classifier(s) to make the prediction (Yaşar Sağlam and Street 2018). So, in this paper, *K*-Nearest Neighbour is applied with the aim of defining a set-point temperature in order to obtain thermal comfort of the occupants of buildings based on the principles of the adaptive comfort theory.

In Bermejo et al. (2012), a system capable of adapting to the user’s thermal preferences based on the PMV index and the application of the Fuzzy Logic is shown. The limitations of this model are based on the definition of the PMV itself, which considers six parameters: clothing insulation and metabolic rate (generally fixed values based on an assumption), the air temperature, the mean radiant temperature, the air speed and the relative humidity. Moreover, fuzzy logic based HVAC control and fault detection systems have been proposed in Ahmad et al. (2016) and in Kang et al. (2015), where the application of fuzzy logic based on an on-off controller to maintain a set-point temperature taking into account a solar energy compensation method is exposed. Regarding machine learning methods and thermal comfort, they have been applied to learn occupants’ behaviour in order to infer real-time set-point temperature for a room (Peng et al. 2018), and to reduce the need of human intervention into the control system. In Kim et al. 2018 machine learning in personal comfort models (predicting individuals’ thermal preference using occupant behaviour) is analysed, but there are no previous studies in the current literature which consider machine learning as well as thermal comfort. So, the present paper proposes a methodology in order to analyse and predict dynamically the comfort temperature of the occupants of buildings, based on the principles of the adaptive comfort theory but carrying out the *K*-Nearest Neighbour (KNN) method. Due to the fact that our current modeling approach is based on the thermal preferences of the occupants of a building collected through a survey, the size of this dataset, that is, the number of responses to the survey would limit the performance of the model. In order to avoid it, the proposed methodology is based on previous studies (Kim et al. 2018) which suggest online machine learning to dynamically adapt new patterns in the data and automatically update a model as needed. The paper is structured as follows. In Sect. 2, the traditional adaptive thermal comfort theory is shown. In Sect. 3, the proposed dynamic adaptive comfort system, based on the *K*-Nearest Neighbour algorithm as pattern recognition to predict the comfort,

is exposed. Section 4 focuses on the KNN method and finally, the experiments and results are shown in Sects. 5 and 6 respectively and conclusions in Sect. 7.

2 Traditional methodology

To determine the indoor comfort temperature of a building, the traditional adaptive comfort methodology assumes an adaptive approach of their occupants. In Nicol (2011), is highlighted that “if a change occurs, such as to produce discomfort, people react in ways which tend to restore their comfort”. Particularly, for mixed mode buildings the adaptive comfort opportunities range from opening or closing windows and turning on the HVAC system to changes in clothing. The traditional methodology is based on a field study, for collecting the thermal sensation of the occupants of a building and the environmental variables through sensors, and on a linear regression analysis in which the mean comfort temperature for a group of occupants (or single) is adjusted contrasted with an outdoor environment variable. For that, the subjects’ thermal sensation is analysed by the question “How do you feel just now?” whose answer is declared in the form of a comfort vote based on a 7-point thermal sensation scale. This comfort vote is called TSV (Thermal Sensation Vote) and it is based on the ASHRAE scale (hot, warm, slightly warm, neutral, slightly cool, cool and cold). To predict the comfort temperature ($T_{comfort}$) from the mean thermal sensation vote (TSV) and the globe temperature (T_{globe}) the equation defined in Formula 1 is used (Griffiths 1990).

$$T_{comfort} = T_{globe} - \left(\frac{TSV}{G} \right) \quad (1)$$

where G is the Griffiths constant, a single standard value.

The dynamism of the concept of comfort and the adaptability of users is expressed based on the average external temperatures of the last days. The influence of the days has different weights, where the weight ($alpha, \alpha$) of the nearest days is greater. Weights are reduced as a function of time, thus reflecting the concept of “forget”. A weighted running mean outdoor temperature T_{rm} is calculated as follows:

$$T_{rm} = \alpha \cdot T_{rm-1} + (1 - \alpha) \cdot T_{od-1} \quad (2)$$

α is a constant between 0 and 1 which defines the speed at which the running mean responds to outdoor temperature, T_{od-1} is the daily mean outdoor temperature (°C) for the previous day and T_{rm-1} is the running mean outdoor temperature (°C) for the day before.

After defining the T_{rm} the relationship between the comfort temperature and the outdoor temperature is calculated as follows:

$$T_{comfort} = m \cdot T_{rm} + c \quad (3)$$

where the slope (m) and the constant (c) of this Formula are obtained by a linear regression analysis. Such expression is the so called adaptive control algorithm (ACA) and allows the definition of the daily indoor comfort temperature.

Figure 1 shows the process to define and implement an experimental ACA for a sample building provided with an HVAC system. Such process is defined as follows: firstly, the information obtained from the building’s sensors (every 15 min) is stored in a server (in this case) and at the same time, the data from a weather station installed in the building (external sensors) is stored while the information regarding the occupants of the building is obtained from surveys. The building behaviour data should be stored for at least one year in order to include a winter season and a summer season.

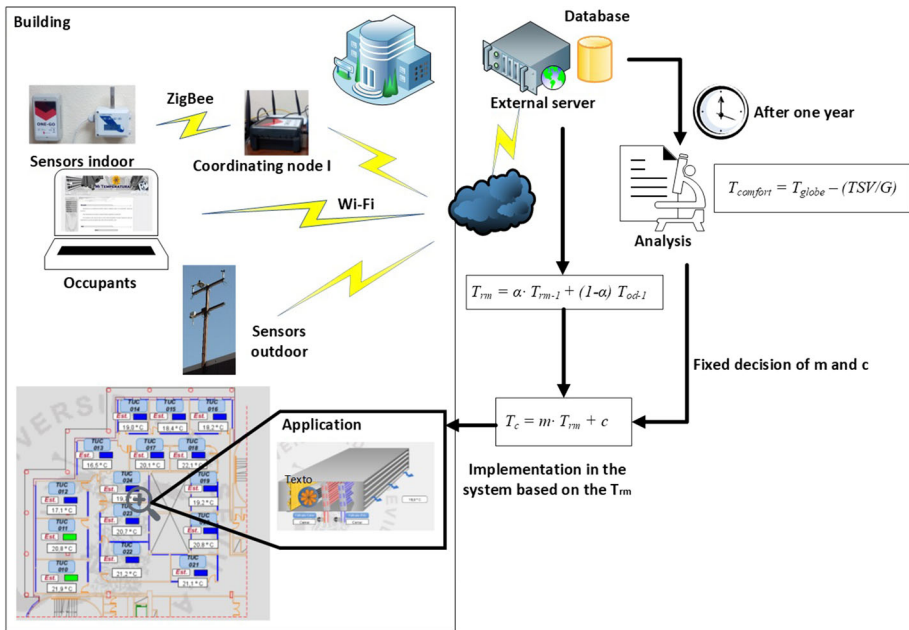


Fig. 1 A traditional system to define and implement an ACA for a sample building

Secondly, a regression analysis is performed with the main objective to obtain the m and c constants (function 3) and therefore an experimental ACA.

Thirdly, such ACA is implemented into the control system of the sample building in order to define the set-point temperature, but taking into account the occupants’ thermal comfort.

3 Dynamic adaptive comfort system (DACS)

The traditional adaptive comfort methodology explained in Sect. 2 is based on an ACA, which is obtained after an experimentation field study and on m and c values which don’t change. Figure 2 shows the alternative methodology proposed in the present paper in order to define an ACA but based on the participation of the occupants and a dynamic and continuous learning of the building, the climate and the comfort algorithm itself.

The dynamic adaptive comfort system proposed, as the traditional system, stores on a server the information obtained from the sensors of the sample building, the outdoor temperature and the answers of the occupants to the question “How do you feel just now?”. Based on such information and the K -NN method (which will be explained in Sect. 4) the proposed system would dynamically and continuously learn the comfort of the building, that is, every moment the system is learning a comfort temperature.

So, although a regression analysis for defining the comfort temperature (as the adaptive thermal comfort theory proposes) is not required, the learning process carried out based on the KNN method allows knowing such temperature as well as the comfort ranges for a certain T_{rm} . Therefore, the proposed methodology could be included in an operating Building Automation and Control Networks (BACnet) based HVAC control system (Song et al. 2007) in order to define the set point temperature and moreover, additional decision criteria, such

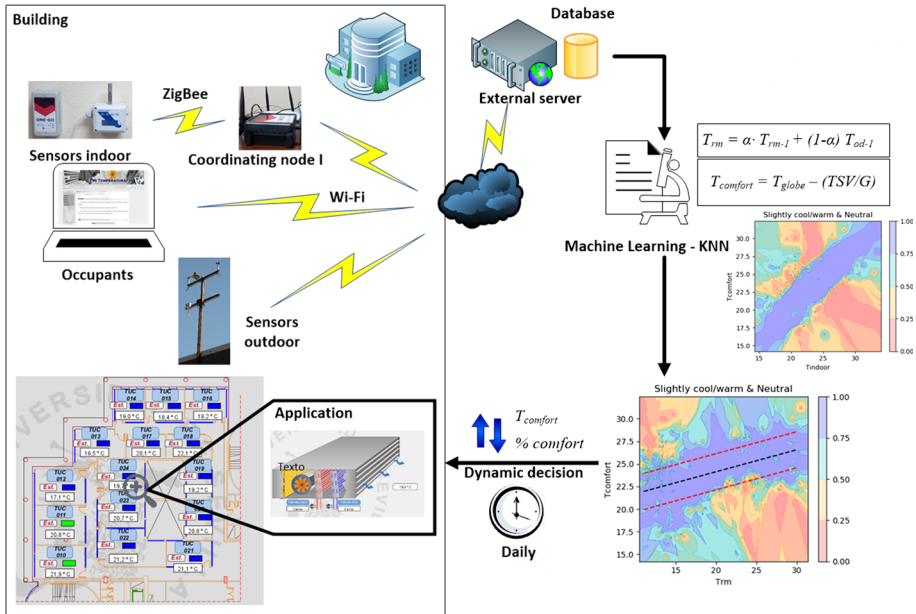


Fig. 2 The proposed system for a sample building: DACS

as consumption per temperature degree, the time of day, trends in the interior, peak hours in consumption, etc. could also be taken into account to define it.

4 K-Nearest Neighbour

K-Nearest Neighbour (KNN) proposes an attractively straightforward classification solution. It is based on classifying a certain point of data taking into account its nearest neighbours and achieves competitive results when enough information is available to carry out the learning process.

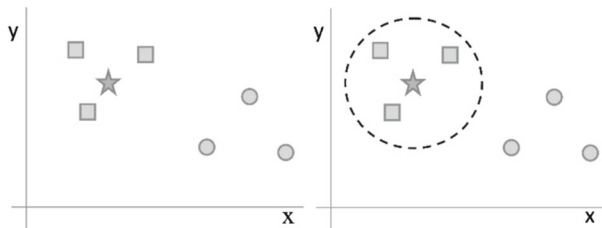
The classical KNN method is an algorithm based on the following idea: firstly, a necessary set of learning data is defined where the points (reference of knowledge) describe each class. The set of numerical features and the selection of a metric (e.g., Euclidean distance) is used to measure the similarity of data points based on all features. After that, in order to analyse a target point, the *K* closest points in training samples (based on the similarity metric) are found and the target point is assigned to a class by majority vote of its neighbours.

The pseudocode of such algorithm is presented in Table 1. Figure 3 shows a simple case to understand it in which two classes are defined (circles and squares) in order to analyse the scenario and to find out the class of the element represented by a star (either a circle or a square). The *K* value in the KNN algorithm is the number of the nearest neighbours to be analysed by the system to define the class of the element. In Fig. 3, *K* = 3, hence, a circle around the star, centred on this point, is defined just as big as to enclose only three data-points on the plane. In this case, the choice is obvious as all three classes from the closest neighbour were squares. The choice of the parameter *K* has a high importance or influence in the algorithm and their effects. All the training observation remains constant, with a given

Table 1 Pseudocode for KNN algorithm

The KNN algorithm

Suppose

 $D =$ training set consisting of samples $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ $Z =$ unknown set consisting in incoming test example $(x'_1, y'_1), \dots, (x'_m, y'_m)$ $k =$ number of the nearest neighbourhood**for** all i in Z (unknown samples): **for** all j in D (known samples): compute the distance between sample i and sample j **end for** find the k smallest distances (select the k closest training examples) locate the corresponding samples j_1, \dots, j_k assign i to the class which appears more frequently**end for****Fig. 3** KNN example

K value, and it defines the boundaries of each class. These boundaries will segregate in different classes. Organizing data into sensible groupings arises naturally in many scientific fields, therefore, it is not surprising the increasing popularity of data clustering during the last years (Jain 2010).

K -Nearest-Neighbour method is a non-parametric method used for classification and regression (Altman 1992) and it is especially successful when decision boundaries are irregular, so in the present study such method is the proposed one as an alternative to analyse the comfort zone, despite its simplicity. The limits of the thermal comfort area have a difficult classification, due to the fact that they are inherently vague and cannot be categorized by a simple model, so the KNN method represents a feasible solution to analyse them. Different metrics exist to develop this analysis, the distance function for completing the KNN is closely related to the scale of the values of characteristics presented. To avoid characteristics with a large range of outweighing characteristics of relatively lower magnitude and thus causing unnecessary weight bias (Shalabi et al. 2006), all the features are in the same range. The distance metric in the KNN classification has a significant impact on the performance, both in accuracy as in time and space complexity. In this study, the Minkowski distance function is selected to evaluate the similarity between the data sets (training and testing). In the analysis, the test data has certain closeness or similarity with the criteria is computed from the training data used to learn the pattern (Cover and Hart 1967). The decision is made to the test data by gauging the output of these functions, based on the class of the training data of the highest

proximity. The Minkowski distance is defined by two points:

$$X = (x_1, x_2, \dots, x_n) \text{ and } Y = (y_1, y_2, \dots, y_n) \in \mathfrak{R}$$

where

$$\left(\sum_{i=1}^n |x_i - y_i| \right)^{\frac{1}{p}} \quad (4)$$

p is usually equal to 1 (Manhattan distance) or to 2 (Euclidean distance). And in the limit case of p reaching infinity the Chebyshev distance is obtained (Cunningham and Delany 2007).

5 Experiment

The following subsections expose the way the experimental data were obtained and the methodology proposed was carried out in order to obtain a dynamic adaptive comfort system (DACS) based on KNN methods.

5.1 Data acquisition

It is necessary to have an appropriate data base in order to test the adaptive comfort system and to obtain more accuracy and efficient solutions.

Particularly, the methodology proposed in the present paper was based on real data obtained from a field study which was carried out in three buildings of the tertiary sector in the city of Seville, south of Spain. For that, 11 offices' rooms were monitored and 59 occupants (all adults) were asked about their thermal sensation votes through a web-survey and using the thermal sensation scale proposed by the EN-15251 standard and translated into Spanish. In Barbadilla-Martín et al. (2017), the comfort field study is shown based on environmental measurements (obtained by sensors) and thermal sensation votes (obtained from more than 4.000 longitudinal questionnaires) collected during the one-year study. This is the data set used for the evaluation, in training set and testing in the present methodology. The training data set is shown in Figs. 4d and 5d, where green colour is the comfort vote of the occupants and the rest of the data (blue and red) define situations in which the users feel cool and warm temperature.

5.2 Application of the KNN methodology

The application of the proposed methodology is mainly based on knowing and checking which is the $T_{comfort}$ for a certain T_{indoor} and T_{rm} . It also allows the definition of the comfort temperature in which the occupants would feel comfortable and the classification of the discomfort areas, so an adaptive comfort area could be built, similar to the traditional one defined by the adaptive comfort theory.

In the KNN method implemented in the present paper, the distance function Minkowski with $p = 2$ (Euclidean distance) was considered. The number of neighbours to use was analysed and it is shown in Figs. 4f and 5f, considering a uniform weight function (where all points in each neighbourhood are weighted equally). Analysing the relation between comfort temperature and indoor temperature (Fig. 4f) the best value is $K = 10$ and analysing the

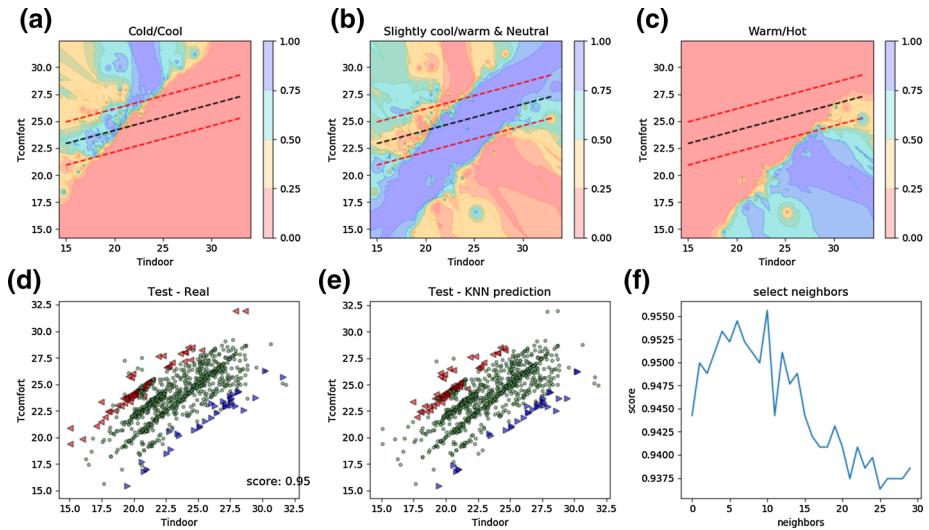


Fig. 4 a–c The diagrams represent the relation between comfort temperature and indoor temperature. **d** Training and **e** Test data set which represents the relation between comfort temperature and indoor temperature. **f** Scores based on the K neighbours in the analysis of the relation between comfort temperature and indoor temperature

relation between comfort temperature and running mean temperature (Fig. 5f) the best value is $K = 8$.

Apart from the uniform weight function, the distance weight function (where the points have different weight by the inverse of their distance) was also analysed and in both cases, the solution had high quality (Figs. 4d and 5d show the score with uniform weight). It has been decided to choose a uniform weight, due to the subjective aspect of the human response to comfort. In this case, if there is an isolated situation, it should not prevail only the closest, but for situations or areas with little data, proximity can be important to choose distance weight. However, it is expected that, as in this case, there will be enough data to delimit the comfort frontier.

The DACS is based on a non-generalizing machine learning method, so the whole training data is remembered and transformed into a fast indexing structure such as a Ball-Tree or KD-Tree (spatial methods have been proposed and tested as a way of alleviating the computational cost of such statistics without resorting to approximate answers).

KD-Tree (Freidman et al. 1977) and the Ball-Tree (Liu et al. 2006) are the most commonly algorithm for KNN methods and they are based on the questions: “What are the K Nearest Neighbours?”, “How many of the K Nearest Neighbours are from the positive class?” and “Are at least t of the K Nearest Neighbours from the positive class?”. Both of them were tested in the present paper and the one with the best fitness and the best performance was selected.

In the analysis, the real data obtained is represented in the Figs. 4d and 5d. This is the real data, and it is compared to the Test solution after the application of KNN prediction. The quality of the solution obtained is represented by the score variable, and in this case, the values 0.96 and 0.89 show that in 96% and 89% of the classifications of the data used for the test, the solution obtained by the proposed methodology is the same as the real classification. If the value of the R^2 in the regression analysis of the comfort temperature based on the

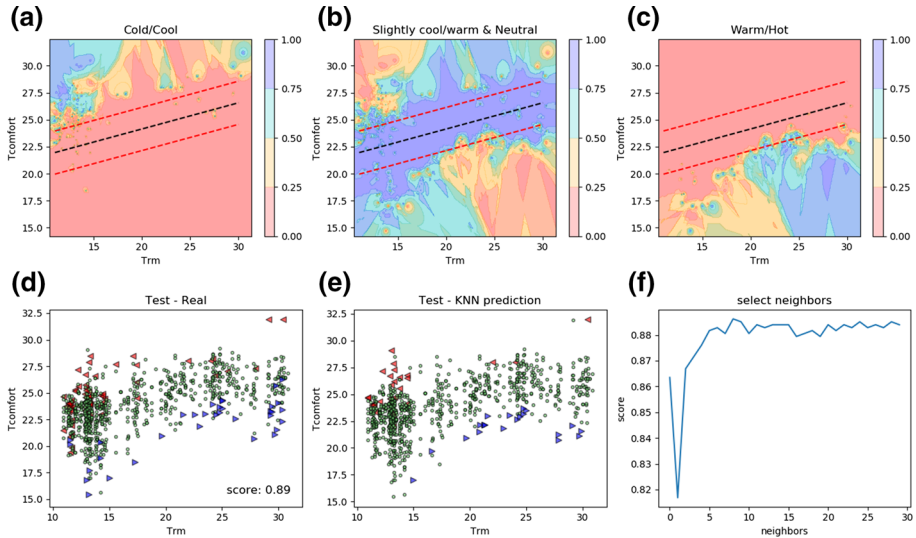


Fig. 5 a–c The diagrams represent the relation between comfort temperature and running mean. **d** Training and **e** Test data set which represents the relation between comfort temperature and running mean temperature. **f** Scores based on the K neighbours in the analysis of the relation between comfort temperature and running mean temperature

traditional method is observed [the values are around 0.41 (Barbadilla-Martín et al. 2017)], it could be concluded that the application of the proposed methodology would be very positive.

6 Results

6.1 The results of KNN in the field study data

The answer to the question “How do you feel just now?” is considered the expression of the subjects’ thermal sensation and for field studies in the area of thermal comfort it is usually represented through a Thermal Sensation Vote (TSV) based on a seven-point scale (hot, warm, slightly warm, neutral, slightly cool, cool and cold), the ASHRAE scale.

In the present studio, three classes were defined for the KNN method based on such thermal sensation scale: warm (including warm and hot), comfort (including slightly warm, neutral and slightly cool) and cool (including cool and cold). Previous studies have already considered similar classifications, demonstrating that, “slightly warm” and “slightly cool” could be considered as “comfort” without bias in the results (Humphreys et al. 2015). Moreover, to predict the comfort temperature $T_{comfort}$ based on the thermal sensation votes (TSV) and the globe temperature T_{globe} , the equation defined in Formula 1 was used with a Griffith constant G equal to 0.5.

Figure 4d shows the training data set, based on the answers of the occupants of the sample building collected in the field study through a thermal comfort survey and Fig. 4a–c show the comfort temperature compared to the indoor temperature, classified into the three categories previously exposed (warm, comfort and cool). In Fig. 4a, the temperature space for which the indoor temperature is below the comfort range for the users of the building is defined. Figure 4b shows the area in which the users felt “comfort” (central purple band in the

figure) and the membership relation is also represented with different colours (range of color defined in the right of the figure). The purple area defines spaces where the occupants could also feel “comfort” (Fig. 4b), feel “cold or cool” (Fig. 4a) or feel “warm or hot” (Fig. 4c). As the proposed methodology is based on data from a survey to train the system, it must be considered that the complete area of discomfort cannot be analysed. That is the reason why a space considered as neutral could be identified in the upper left and this is because of the lack of data that generates learning in such area. This occurs because such situation would never take place, and if it did, the discomfort responses of the users would repair the learning area.

Finally, Fig. 4c shows situations in which the indoor temperature is above the comfort temperature, and therefore users would feel “heat”. Whereas Fig. 4a–c represent the basic adaptive relationship between the comfort temperature and the indoor temperature. Figure 5a–c show the relationship between the comfort temperature and the running mean temperature, that is, the relationship between the adaptive comfort temperature and the outdoor temperature taking into account the previous days (the running mean temperature). Such running mean temperature is based on Formula 2 with a recommended value of α equal to 0.8, which defines higher weights or influence to the outdoor temperatures in recent days. Considering the comfort temperature and the running mean temperature, the same three categories previously explained are defined. Figure 5a represents comfort temperatures for occupants who felt “cold”, which would generate a situation of overheating in the building. Figure 5c shows just the opposite situation, a comfortable temperature that would generate an overcooling situation is represented. Figure 5b shows the comfort temperature range regarding the running mean temperature based on the proposed methodology. Such range temperature is defined in a dynamic way and using a KNN method. Moreover, in Fig. 5a–c the relationship between daily comfort temperature and outdoor temperature based on the thermal adaptive theory (Barbadilla-Martín et al. 2017) (dash black line), is also represented (Formula 5).

$$T_{comfort} = 0.24 \cdot T_{rm} + 19.3 \quad (5)$$

The slope (m) and the constant (c) of this Formula are obtained by a linear regression analysis (traditional methodology). The adaptive thermal comfort assumes that people would feel comfortable in a range of temperature no more than 2° from the comfort temperature (red dash lines), and if the difference between the indoor temperature and the comfort temperature was more than 2° , the discomfort would increase (Nicol and Humphreys 2007).

$$T_{comfort-max} = 0.24 \cdot T_{rm} + 21.3 \quad (6)$$

$$T_{comfort-min} = 0.24 \cdot T_{rm} + 17.3 \quad (7)$$

Comparing the comfort range based on the principles of the adaptive comfort theory (red dash lines) with the comfort range based on the proposed methodology, it could be observed that the solution exposed in the present paper defines a more flexible and wider comfort area and that a range of acceptance could be identified. It would allow defining different indoor temperatures depending on the moment of the day (for example, one temperature for noon and another one for night).

In Fig. 6a is shown the real data used for training, inputs of regression, KNN and SVM methods. In Fig. 6c is shown the real data used for test in the regression, KNN and SVM methods. In the case of regression, the solution of the traditional regression model is applied to this data in the Fig. 6b, d. The real data used for test may be compared with the solution of KNN method in Fig. 5e and the solution of SVM method in Fig. 7e.

The solution highlights that the proposed methodology generates its own knowledge of the building and its response is an indicator of the quality of comfort. The system, once the

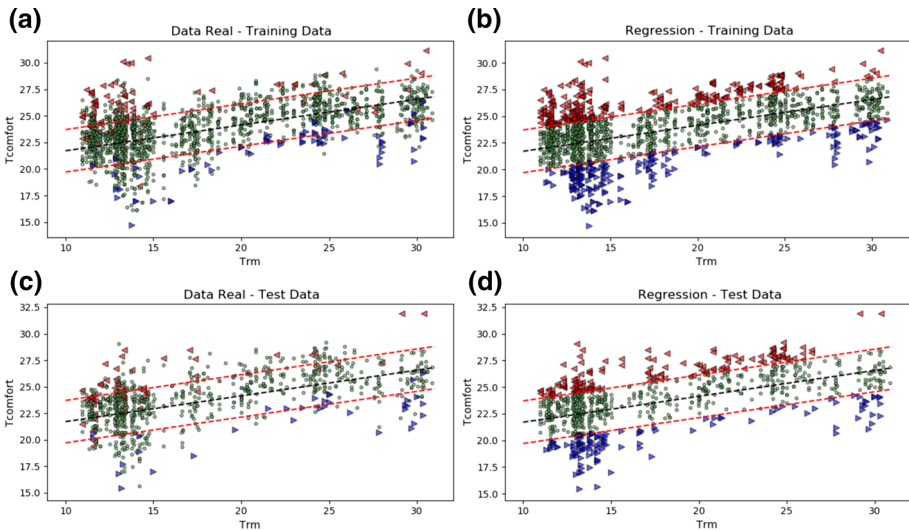


Fig. 6 **a** Training data, **b** traditional solution with ACA applied to training data, **c** test data, **d** traditional solution with ACA applied to test data

Table 2 Comparative table

	Regression	KNN	SVM-RBF
95% CI	(0.683, 0.743)	(0.874, 0.915)	(0.863, 0.905)
Kappa	0.197	0.338	0.161

T_{rm} was known, would obtain a comfort probability space (Fig. 5a–c) and moreover, for a certain indoor temperature, the level of comfort could also be known (Fig. 4a–c). These values would allow to define a set-point temperature for buildings in a dynamic way based on changes in the weather, the buildings and occupants’ behaviour, unlike the classic range.

6.2 Comparative results

The traditional regression method based on the adaptive comfort theory was exposed in Sect. 2 and its results were shown in Fig. 6. The methodology proposed in the present paper based on the K -Nearest Neighbour method was explained in Sect. 4 and its results were shown in Fig. 5. In this section, an alternative method, the Support Vector Machine (SVM) (Vapnik and Vapnik 1998; Wang et al. 2010; Camelo et al. 2015) is considered and compared to the techniques previously exposed: the regression analysis defined in Formula 5 and the KNN method with $K = 8$. Regarding the SVM method, linear SVMs is generally used for linear problems but in the present paper it is not shown because of the solution was poor. Otherwise, in non-linear problems the Radial Basis Function kernel (RBF-kernel) is used and its solution is shown in Fig. 7. Table 2 shows the analysis of the test data considering a 95% confidence interval, in which KNN and SVM methods identify the best solutions.

As the objective is finding a model to minimise the number of errors, the Kappa index test, or statistical Kappa, was used to evaluate the responses to comfort. Kappa statistic is a metric based on a threshold and a qualitative understanding of error and it is commonly

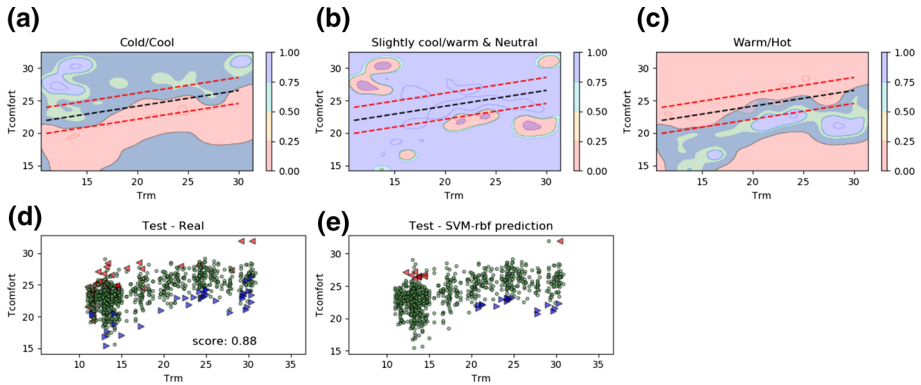


Fig. 7 SVM with RBF kernel function

Table 3 Confusion matrix of the techniques

	Regression			KNN			SVM-RBF		
	C	N	W	C	N	W	C	N	W
Actual									
C	30	33	2	16	49	0	4	61	0
N	112	575	93	14	758	8	6	765	9
W	0	12	22	0	22	12	0	26	8
ACC	0.83	0.72	0.88	0.93	0.89	0.97	0.93	0.89	0.96
AUC	0.66	0.64	0.77	0.67	0.55	0.61	0.5	0.5	0.5
PLR	3.35	1.62	5.76	14.31	1.36	37.28	8.35	1.12	22.0915

C cool, N neutral, W warm

used in direct applications of classifiers. Moreover, Cohen’s kappa is generally thought to be a more robust measure than a simple assessment (Cohen 1960).

Regarding the Kappa measurement, Landis and Koch (1977) and Altman (1990) suggested considering a value 0.01–0.20 as slight or poor and 0.21–0.40 as fair. Based on that and according to the data shown in Table 2, Cohen’s kappa measure shows that the KNN method has fair agreement with the test results.

In Table 3 the confusion matrix is shown. It can be observed that the accuracy (ACC) is good, specially for the KNN method. In the analysis of the area under the roc curve (AUC) the solution is fair in the case of the regression and the KNN method, but poor in the SVM. And in the positive likelihood ratio (PLR), the result is poor for the regression technique, good for the KNN method considering the analysis of ‘cool’ and ‘warm’ and good considering the analysis of ‘warm’ for the SVM method.

7 Conclusions

A methodological analysis based on the *K*-Nearest Neighbour algorithm was proposed, in order to verify if a certain temperature is in the comfort area of a building and if the decision on a change of temperature is acceptable. To validate such methodology, data set from a

thermal comfort field study carried out in three buildings in the south-western area of Spain was considered. Moreover, the relationship between indoor comfort temperature and outdoor temperature (ACA) identified in such field study was compared to the results obtained with the dynamic adaptive comfort system exposed in the present paper and based on the KNN method.

The DACS system and its results represent an alternative solution to define a comfort range for a building and its results also further question the validity of applying fixed set-point temperatures to real working situations and the overuse of heating, ventilating and air-conditioning systems.

It could also be highlighted that the proposed methodology is a flexible and dynamic solution to make decisions as an alternative to the traditional solution one. So, it is concluded that many buildings could apply the proposed methodology (or similar machine learning methodologies) to learn the comfort area of their occupants and to use such information in a decision support system that would define a suitable set point temperature and the limits of the comfort range.

Acknowledgements The authors wish to acknowledge the financial support of project DACAR (Ref. BIA2016-77431-C2-1-R) funded by the *Programa Estatal de Investigación, Desarrollo e Innovación Orientada a los Retos de la Sociedad* (MINECO).

References

- Ahmad, M. W., Mourshed, M., Yuce, B., & Rezgui, Y. (2016). Computational intelligence techniques for hvac systems: A review. *Building Simulation*, 9(4), 359–398. <https://doi.org/10.1007/s12273-016-0285-4>.
- Altman, D. G. (1990). *Practical statistics for medical research*. Boca Raton: CRC Press.
- Altman, N. S. (1992). An introduction to kernel and nearest-neighbour nonparametric regression. *The American Statistician*, 46(3), 175–185. <https://doi.org/10.1198/016214507000000932>.
- Aparicio-Ruiz, P., Barbadilla-Martín, E., Salmerón-Lissén, J. M., & Guadix-Martín, J. (2018). Building automation system with adaptive comfort in mixed mode buildings. *Sustainable Cities and Society*, 43, 77–85. <https://doi.org/10.1016/j.scs.2018.07.028>.
- ASHRAE. (2013). *Handbook fundamentals*. Atlanta: ASHRAE.
- Barbadilla-Martín, E., Salmerón Lissén, J., Guadix Martín, J., Aparicio-Ruiz, P., & Brotas, L. (2017). Field study on adaptive thermal comfort in mixed mode office buildings in southwestern area of Spain. *Building and Environment*, <https://doi.org/10.1016/j.buildenv.2017.06.042>.
- Bermejo, P., Redondo, L., de la Ossa, L., Rodríguez, D., Flores, J., Urea, C., et al. (2012). Design and simulation of a thermal comfort adaptive system based on fuzzy logic and on-line learning. *Energy and Buildings*, 49, 367–379. <https://doi.org/10.1016/j.enbuild.2012.02.032>.
- Camelo, S. A., González-Lima, M. D., & Quiroz, A. J. (2015). Nearest neighbors methods for support vector machines. *Annals of Operations Research*, 235(1), 85–101. <https://doi.org/10.1007/s10479-015-1956-8>.
- CIBSE. (2015). Guide A: Environmental design.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- Cole, R. J., Brown, Z., & McKay, S. (2010). Building human agency: A timely manifesto. *Building Research and Information*, 38(3), 339–350. <https://doi.org/10.1080/09613211003747071>.
- Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>.
- Cunningham, P., & Delany, S. J. (2007). K-nearest neighbour classifiers. *Multiple Classifier Systems* (pp. 1–17). [https://doi.org/10.1016/S0031-3203\(00\)00099-6](https://doi.org/10.1016/S0031-3203(00)00099-6).
- de Dear, R. J., Brager, G. S., Reardon, J., Nicol, F., et al. (1998). Developing an adaptive model of thermal comfort and preference/discussion. *ASHRAE Transactions*, 104, 145.
- Freidman, J. H., Bentley, J. L., & Finkel, R. A. (1977). An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software*, 3(3), 209–226. <https://doi.org/10.1145/355744.355745>.

- Griffiths, I. (1990). *Thermal comfort studies in buildings with passive solar features, field studies*. Report to the Commission of the European Community 35.
- Humphreys, M., Nicol, F., & Roaf, S. (2015). *Adaptive thermal comfort: Foundations and analysis*. London: Routledge.
- Hussein, I., & Rahman, M. H. A. (2009). Field study on thermal comfort in malaysia. *European Journal of Scientific Research*, 37(1), 134–152.
- ISO Standard. (2008). ISO 15251: Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics. EN 15251.
- Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>.
- Kang, C. S., Hyun, C. H., & Park, M. (2015). Fuzzy logic-based advanced on-off control for thermal comfort in residential buildings. *Applied Energy*, 155, 270–283. <https://doi.org/10.1016/j.apenergy.2015.05.119>.
- Kim, J., Zhou, Y., Schiavon, S., Raftery, P., & Brager, G. (2018). Personal comfort models: Predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning. *Building and Environment*, 129, 96–106. <https://doi.org/10.1016/j.buildenv.2017.12.011>.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Liu, T., Moore, A. W., & Gray, A. (2006). New algorithms for efficient high dimensional non-parametric classification. *Journal for Machine Learning Research*, 7(c), 1135–1158.
- Nicol, F., & Humphreys, M. (2007). Maximum temperatures in european office buildings to avoid heat discomfort. *Solar Energy*, 81(3), 295–304. <https://doi.org/10.1016/j.solener.2006.07.007>.
- Nicol, J. F. (2011). Adaptive comfort. *Building Research and Information*, 39(2), 105–107. <https://doi.org/10.1080/09613218.2011.558690>.
- Peng, Y., Rysanek, A., Nagy, Z., & Schlüter, A. (2018). Using machine learning techniques for occupancy-prediction-based cooling control in office buildings. *Applied Energy*, 211, 1343–1358. <https://doi.org/10.1016/j.apenergy.2017.12.002>.
- Rupp, R. F., Vásquez, N. G., & Lamberts, R. (2015). A review of human thermal comfort in the built environment. *Energy and Buildings*, 105, 178–205. <https://doi.org/10.1016/j.enbuild.2015.07.047>.
- Shalabi, L. A., Shaaban, Z., & Kasasbeh, B. (2006). Data mining: A preprocessing engine. *Journal of Computer Science*, 2(9), 735–739. <https://doi.org/10.3844/jcssp.2006.735.739>.
- Song, W. S., Hong, S. H., & Park, T. J. (2007). The effects of service delays on a BACnet-based HVAC control system. *Control Engineering Practice*, 15(2), 209–217.
- Si, Tanabe, Haneda, M., & Nishihara, N. (2015). Workplace productivity and individual thermal satisfaction. *Building and Environment*, 91, 42–50. <https://doi.org/10.1016/j.buildenv.2015.02.032>.
- Tuohy, P., Roaf, S., Nicol, F., Humphreys, M., & Boerstra, A. (2010). 21st century standards for thermal comfort fostering low carbon building design. *Architectural Science Review*, 53(1), 78–86. <https://doi.org/10.3763/asre.2009.0112>.
- Valor, E., Meneu, V., Caselles, V., Valor, E., Meneu, V., & Caselles, V. (2001). Daily air temperature and electricity load in Spain. *Journal of Applied Meteorology*, 40(8), 1413–1421. [https://doi.org/10.1175/1520-0450\(2001\)040<1413:DATAEL>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<1413:DATAEL>2.0.CO;2).
- Vapnik, V., & Vapnik, V. (1998). *Statistical learning theory* (pp. 156–160). New York: Wiley.
- Wang, S., Jiang, W., & Tsui, K. L. (2010). Adjusted support vector machines based on a new loss function. *Annals of Operations Research*, 174(1), 83–101. <https://doi.org/10.1007/s10479-008-0495-y>.
- Yaşar Sağlam, Ş., & Street, W. N. (2018). Distant diversity in dynamic class prediction. *Annals of Operations Research*, 263(1), 5–19. <https://doi.org/10.1007/s10479-016-2328-8>.