Empirical study of massive set-point behavioral data: Towards a cloud-based artificial intelligence that democratizes thermostats

Alfonso P. Ramallo-González
Department of Information and Communication Engineering
Univeristy of Murcia
Murcia, Spain
alfonsop.ramallo@um.es

Aurora González-Vidal
Department of Information and Communication Engineering
Univeristy of Murcia
Murcia, Spain
aurora.gonzalez2@um.es

Antonio Skarmeta
Department of Information and Communication Engineering
Univeristy of Murcia
Murcia, Spain
skarmeta@um.es

***Abstract*—The research showed in this document consisted on monitoring 546 air conditioners of individual offices located in two large buildings for the later evaluation of the behaviors of users with respect to their controllers. Data was collected over 14 months and provided important insights about the phenomenon. It was seen that users can be separated in two groups, one that likes to interact with the controllers often and change the temperature at least once a week and another that interact less. It was seen that the variability of users with respect to the thermostat values they prefer is high, and that this should be taken into account when creating a "one fits all" solution. Also, it appears that adaptive thermal comfort theories that suggest users want lower temperatures in cold months are not reflected on the set-points chosen. In addition, we have seen that people interacting more with the controllers tend to waste less energy, this would be interesting if an app to interact with the user for this purpose is design. More communication with the user may imply less energy wasted.**

Keywords—thermostat, energy, buildings, meters, IoT, BigData

I. Introduction and State of the Art

The new scenario of climate change [1] has motivated governments and organizations to start initiatives with the aim of reducing carbon emissions. The building sector has not been an exception on this; on the contrary, buildings have been seen to be responsible of a large amount of the carbon emissions being near the 20% [2].

Within buildings, it has been seen that occupants have a substantial impact on the energy consumption [3], and it is for that reason that several studies have been carried out to understand the behavior of building occupants [4] or [5]. Also to try to reduce energy use via the change in occupants' habits [6]. It has been shown by the literature that this change in habits can result in some cases to 20% savings [7] [8].

Although the thermostat control is the main behavior to regulate the most energy consuming aspect of the behavior of the building, little research has been found in the modelling of this particular aspect. The work of Shipworth et al. [9] and the work of Andersen et al. [10] are two of the most relevant studies in this concern. However, they only consider set-point for heating, leaving un-studied the cooling part of the thermostatic control. Also, the two previous studies focus on dwellings leaving unstudied non-domestic buildings. In addition to that, the study of the use of the thermostat in time seem to be under looked. This could be the result of not being able to capture such data.

With the new paradigm of *Internet of Things* new initiatives for controllers are appearing in buildings. In the same way that this movement has (and will continue) populating buildings with devices connected to the internet [11], smart meters and other energy-related components will undergo a substantial growth [12].

The new situation has also allowed companies to offer at the domestic level solutions that use smart algorithms to optimize thermostat control such as Google's NEST or Honeywell's smart thermostat. This opens a new avenue of opportunity on how to control conditioning systems, and although the previously mentioned solutions have been designed for dwellings one could see the great potential for non-domestic buildings.

Deep Learning and other techniques have been used in the past for demand forecasting [13], but there is a substantial lack of study on the temporal human preferences in what is respect to controllers. As mentioned before, no works have been found that studied the set-point values of conditioning systems as a time series. In this work we have done so with sensors collecting information from 546 terminal units, used each one of these by one office user.

In this study we do a comprehensive study for the first time, to the authors' knowledge, of the time series produced by monitoring the interactions of users with their controllers of the air conditioning machines. The machines and the data collected

This paper has been made possible thanks to the support of the European Commission through the H2020-ENTROPY-649849, the Spanish National Project CI-CYT EDISON (TIN2014-52099-R) and MINECO TIN2014-52099-R project (grant BES-2015-071956) granted by the Ministry of Economy and Competitiveness of Spain (including ERDF support). Ramallo-González would like to thank the program Saavedra Fajardo (grantnumber220035/SF/16) funded by Consejería de Educación y Universidades of CARM, via Fundación Séneca-Agencia de Ciencia y Tecnología de la Región de Murcia

978-1-5386-4705-9/18/$31.00 ©2018 IEEE
DOI 10.1109/SMARTCOMP.2018.00093

from them covers winter and summer, and has been compared to other relevant variables such as outside temperatures, inside temperatures, operation and so forth.

II. Sensing framework

This study was possible thank to the data collected by 546 sensor sets installed in the terminal units of the conditioning system of several buildings. From this, 460 sets were suitable to be used.

Three different non-domestic buildings of multiple use were monitored in this study. These buildings are conditioned using centralized Variable Refrigerant Flow (VRF) conditioning systems consisting on large chilling/heating units installed in the roof and individual fan coil terminals in each room. The terminal unit can deliver the necessary cold or heat demanded by each space and user.

Each terminal unit (or room units) has a remote controller that facilitates the interaction of the user with the conditioning system. The user can turn on an off the room unit at their will, but cannot program the operation based on a timer. Also, the user can control the set-point temperature. This means that the user can change at any time the thermostat control of the unit to any value between 16 and 29-degree C at their will. In addition, the fan speed can be changed from Automatic to low – medium or high. Each room has also a wall mounted screen that shows the temperature of the room captured by the machine, the set-point (thermostat) temperature and the fan operation mode.

The buildings from which these data are being collected include two multi-use. These two are very different buildings with one built on the 80's and the other built on the 00's. Both have terminal conditioning machines in offices of similar sizes (around 10 m2), occupied by a single person. Laboratories, stores and other spaces for special use are conditioned with independent machines that are not part of the data or the study.

The data is collected every 12 minutes and it has been collected between November 2015 (included) and February 2017 (included). This ensure that we have a full year of data and therefore can capture the winter and summer operation. There exists a small gap of data on August due to maintenance of the equipment, but the rest of the data is complete and has not shown any anomaly after a preliminary sanity check.

It is believed that this data is rather unique; because, although some studies have been done using set-point data (Shipworth), none has been found in the literature that uses the set-point temperature as a time series, therefore allowing the study of the interactions of the user with the controllers.

The data studied represents a total of 1.78 Gb and could be available under request to other researchers or reviewers.

III. Methodology

We performed a detailed analysis of the data previously described with the aim of gaining a better understanding of the needs of an algorithm that could control large office buildings. To perform the studies, the data previously stored on a postreSQL database was extracted isolating each single unit for its posterior analysis. Although the complexity of the code to extract the data from the databases was trivial, it should be noted that this is a process computationally expensive and that requires a substantial amount of time. This is mentioned as it could be of use for researchers or computer scientists trying to implement a solution for thermostat controlling at a large scale.

A. *Interactions with the thermostat*

One of the main things we wanted to study in this work was the interactions with the thermostat in the sense of the number of times that users feel the need to change the set-point values of their conditioning machine. The reason for this, is that after preliminary observation of the data, we noticed very different behaviors between users that pointed us to formulating the hypothesis of having two "types" of users: One that needs an almost constant "dialogue" with the machine, and a second that leaves the controllers alone once the right set-point has been chosen.

To evaluate this, we have checked the number of times that a user interacts with the thermostat to change the temperature. We have considered only changes in temperature, in this case, and no other manipulations of the control such as turning on or off, or checking the current temperature. Only data-sets in which the machines were on for at least 50 hours were used, less than this was considered as insufficient to evaluate interactions with the control as they represent more sporadic use of the systems.

To make sure that a realistic value of the interactions with the controllers is obtained, the absolute number of interactions was normalized with the number of hours of operation. This implies that the parameter being evaluated was the number of Set-point changes per hour.

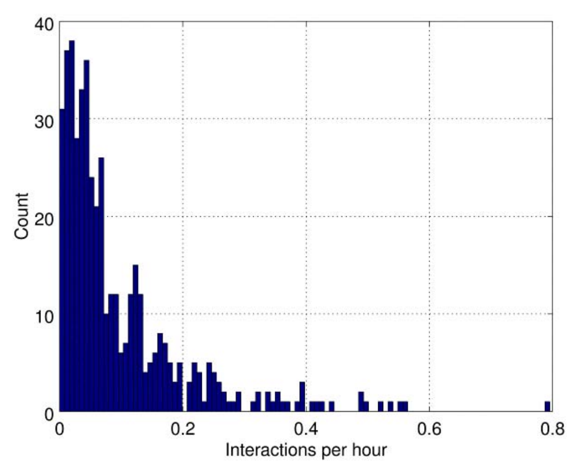

Fig. 1. Histogram of the Interactions per hour of the 460 users analyzed.

Fig. 1 shows the histogram of interactions of the users with their controllers, understanding as interactions the act of changing the thermostat temperature. This graph showed how the majority of the users interact little with their controller. However, due to the nature of the phenomenon being studied we showed that the study of this normalized variable as such was not providing enough information. This was mainly because most of the information gets masked as the variable gets closer to its natural constrain zero.

To overcome this difficulty, we used a manifold that transforms the number of interactions into an unconstrained variable. This manifold was the natural logarithm function. The function transformed a variable defined within the interval [0, inf) into a variable that can takes values within the interval (-inf, inf).

B. Values chosen by the users for the set-point

On the next part of the study, we did a comprehensive analysis of the actual values that the users choose when setting up their conditioning units. This part is a necessary first step for the creation of a centralized control system that could provide comfort to all users of a centrally controlled building such an office building.

The first test to do in this respect was to represent the time series representing the set-point temperatures of the 460 devices and the mean of all of them. The results are shown in Fig. 2.

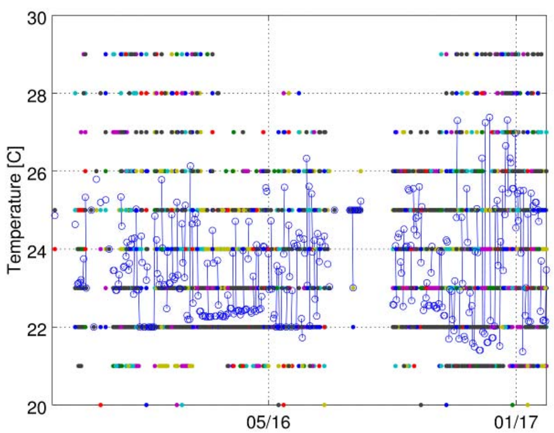

Fig. 2. Set-point temperatures and mean (blue circles).

From this representation we saw that the raw data was going to be difficult to interpreted and to relate to year patterns or tendencies. For this reason, we designed a further analysis.

For this, and considering the large amount of data available we did some preprocessing of the data. We wanted to investigate the fluctuations of data that could be found in thermostat values in the year. This was because Nicol and Humphreys have already established that preferences have been seen to be different along the year. Principle that has also been used on the ASHRAE Standard 55 for thermal environmental conditions. It was crucial to compare the thermostat values on their constant term as on their yearly swing for two reasons: (1) The data used in other studies consider comfort temperatures, but not set-point temperatures although the latter is directly linked with consumption, and (2) the data studied in this paper is unprecedented for the Spanish territory and may outline differences with previous studies due to different thermal preferences.

In our study, a sinusoidal function has been chosen to extract information to the rather chaotic set-point time series (see Fig. 5). The time series of each one of the users were fit a sinusoidal function with fixed wavelength of 365 days. With this we wanted to fit the simplest curve that would smooth the time series yet imposing the yearly periodicity that is expected on a variable of this kind. A constant factor was added to the sinusoidal, this constant factor with the year swing and the phase were the parameters searched to fit the curve. The equation of the curve used for this fitting was:

$$T_{set}(\text{time}) = a + b \sin(2\pi\,\text{time}/365 + \gamma) \tag{1}$$

Where $T_{set}(\text{time})$ is the fitted sinusoidal representing the set-point temperature series, a is the constant term, b is the yearly swing, and γ is the phase or lag.

C. Comparison of values with the zones of comfort

Part of the rational of developing a centralized thermostatic controller is to maximize the comfort of the users and at the same time minimize the energy consumption. The results of this study can be used to complement previous studies that regard energy consumption [14, 15]. This has two consequences: (1) for the environment; as the Carbon Emissions are reduced, and (2) for the running costs; savings of 15% can be seen on the heating and cooling bills with the right behaviors of the occupants, also productivity can increase up to 2% in thermally adequate environments (this is not to be disregarded as salaries normally account for an order of magnitude more than energy bills in office buildings).

To study if the set-points are used adequately, the comfort limits of the ASHRAE Standard 55 were used. These boundaries are defined considering temperature and relative humidity, as in our case no information about RH was available we have considered that this was fixed at 50%. The bounds taken from ASHRAE of comfort for summer and winter are defined in Table I.

TABLE I. COMFORT LIMITS USED ON THE STUDY.

	Bounds of the Comfort Temperatures		
	Min	***Max***	***Clothing***
Summer	25	27	0.5
Winter	21	24	1.0

To evaluate the number of hours that the systems were pushed outside the comfort ranges, and the intensity with which this was done, the integral of the area defined by the curve representing the set-point temperature and the upper or lower bound was calculated. This provided with an indicator of overheating or overcooling on Kelvin-hour, Kh a measurement well recognized on the Building Physics community.

These parameters were only examined when the system was in operation to ensure that an inadequate set-point left on the machine during non-used period was not modifying the data.

D. Who are the big consumers?

Taking advantage of the completeness of the data and its relevance on energy consumption of the building we investigated relationships between variables that could lead to further knowledge creation about thermostat use. To do this, we compared the user's groups identified in section A and compared them with the overheating and overcooling indicators. This and the rest of results are shown in the following section.

IV. RESULTS

A. Interactions with the thermostat

To test the hypothesis of the existence of two different groups of users we plotted the kernel density functions of the set-point changes per hour. The representation showed a clear bimodal nature of this parameter. This proves the hypothesis that considered that two groups of users can be defined; one that interacts with the controller often and another that tends to leave the controller to a given temperature and do not interact with the set-point. This is shown in Fig. 3.

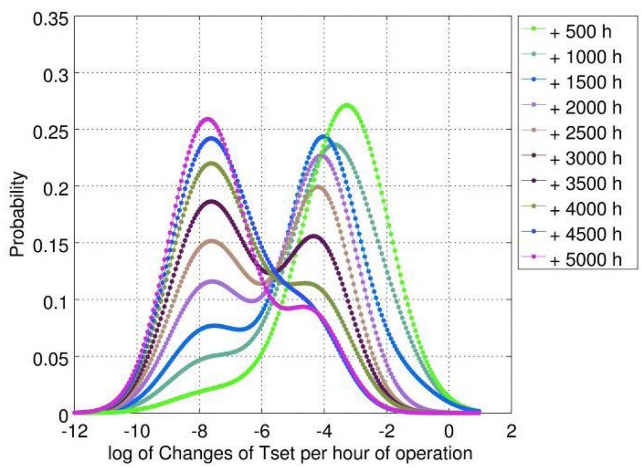

Fig. 3. Kernel density distributions of the natural logarithms of the changes normalized per hour. The different curves represent the subgroups that had a number of hours of operation larger than that on the legend.

It should be noted that different subgroups were used for the representation in Fig. 3. The groups consist on clusters that used the machines for more than the hours given in the legend. This shows that as the subgroup of users is reduced (towards higher consumers) they tend to interact less with the machines. It should be reminded at this point that this is particularly interesting considering that the measurement of interactions in our case is a specific value per hour.

With respect to this, the peak shown on the left of Fig. 3 with a value of approximately -8 [log(1/h)] corresponds to approximately 3e-4 [1/h] what considering 4000 hours of operation represents around 1 or 2 changes of the set-point in the whole period studied (1 year and 4 months). In opposition, a value of -4 [log(1/h)] is translated for a period of 2000 hours into 36 changes, or approximately a change every week. We see then clearly that this bi-modality of the phenomenon is representing two different attitudes toward the use.

To illustrate these two groups, we have included four examples from the data that are representative of these two groups. The first two are represented in Fig. 4 and show the examples of users that did almost not change the set-point for the whole period. The internal temperature in this graph has been represented with a fitted sinusoidal curve of wavelength 365 days to remove any noise and extract only the yearly tendency.

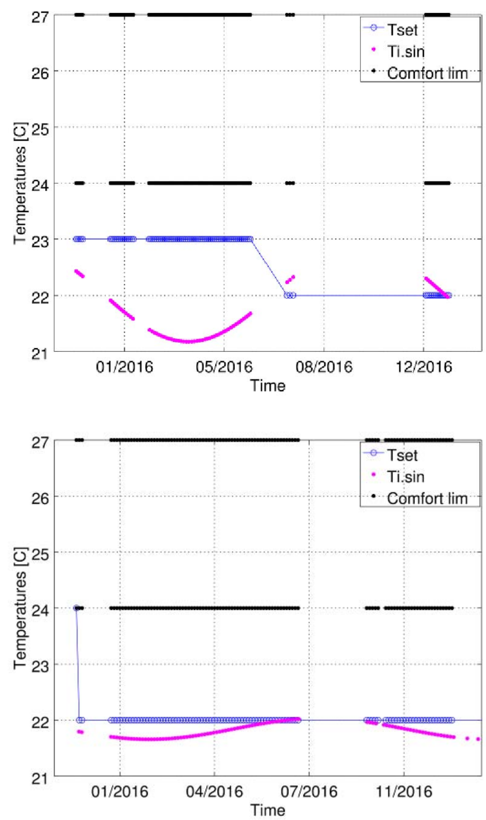

Fig. 4. Examples of users that have little interaction with the thermostats. (The data has been resampled to allow plotting, as the number of points was making fail the rendering engine of the software).

In Fig. 5, however we show two users that interact often with the controllers of the conditioning machine. In this two graphs the internal temperature tendency has also been represented and it is notorious that its value is higher in winter than it is in summer for the second case. Fact that is justified when one sees that the set-points in summer are particularly high.The comfort limits according to ASHRAE 55 have also been included in Fig 4 and Fig 5 with black dotted lines. As the Relative Humidity of the spaces was not recorded, a relative humidity of 50% has been chosen to define the comfort limits. However, this assumption should be taken into account when judging the set-points chosen by the users. Also, depending on the operation of the A/C machine (heating or cooling) the comfort band has been considered as those for winter or summer.

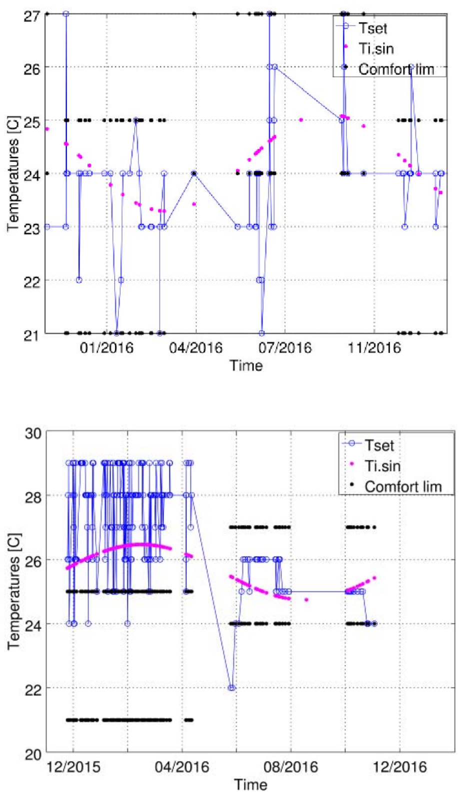

Fig. 5. Examples of users that have large interaction with the thermostats. (The data has been resampled to allow plotting, as the number of points was making fail the rendering engine of the software).

B. *Values chosen by users*

As mentioned before, the values that are chosen by users are key to develop the knowledge that would take us to the creation of a centralized controlled thermostat. As mentioned before, to facilitate the study of the values of the thermostats, sinusoidal curves were fitted to each one of the time series. These curves representing the preferences of the users under study were defined using three parameters, namely the constant value, the yearly swings and the phase or lag.

To perform an analysis of the curves representing the preferences in set-point temperatures, we created a box-plot that is shown in Fig. 6. This shows how the constant term of the curves is rather adequate. Considering that the ASHRAE limits go from 21 to 27 degree, one expects to see this value centered on 24, what is exactly what was seen in the boxplot of the constant term of the curves. Moreover, we see that the whiskers of the box-plots (representing the 97.5% of the data) are almost coincidental with the 21 to 27 margin established by ASHRAE. Although the yearly swing would open up this range in cases, one can assume that in average, the users are not biased when setting up the controllers.

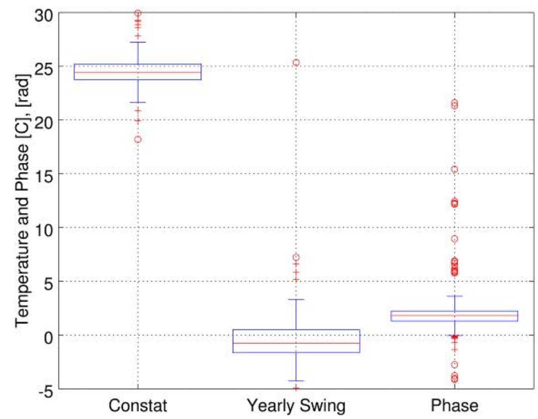

Fig. 6. Box plots of the parameters fitted to the thermostat value time series.

When studying the parameters of the curves, it was seen that the data is rather consistent, and that, excluding a few outliers most of the curves maintain the same characteristics. This can be seen in Table II and in Fig. 6. However, it should be mentioned that we observed a large variability of the yearly swing and phase. This implies that not clear tendency has been seen of having higher set-point temperatures in summer than they are set in winter. This is an interesting finding, as ASHRAE and the body of knowledge on adaptive thermal comfort has shown that the temperature preferences should change along the year.

TABLE II. PARAMETERS OF THE FITTED CURVES OF TSET.

	Parameters		
	Constant term [C]	***Year swing [C]***	***Phase [rad]***
Mean	24.48	-0.66	1.54
std	1.186	1.48	3.76
Medians	25.19	0.47	2.23

To add some light to this fact, we have also plotted all the fitted sinusoidals representing the set-point temperatures of each unit. We have seen after studying this curves, that although the means and medians were not able to outline this, we see a clear tendency of higher temperatures in winter than in summer. This is shown in Fig 7. This would contradict the work of Nicol and Humphreys that suggested in [16] that desired temperatures would increase when outdoor temperature rise (as soon as outdoor temperature does not go below zero). We have seen the opposite in Fig 7.

C. *Comparison of values with the zones of comfort*

The evaluation of the degree-hours outside the area of comfort was very illustrative. It was seen that although the means are very near the recommended temperatures, some machines are being used with a substantial overheating but also overcooling. This section shows the magnitude of this overheating and overcooling. It was seen that users tend to tolerate high temperatures much more than cold temperatures and use values of the thermostat for cooling that are close to the upper bound of the ASHRAE comfort range.

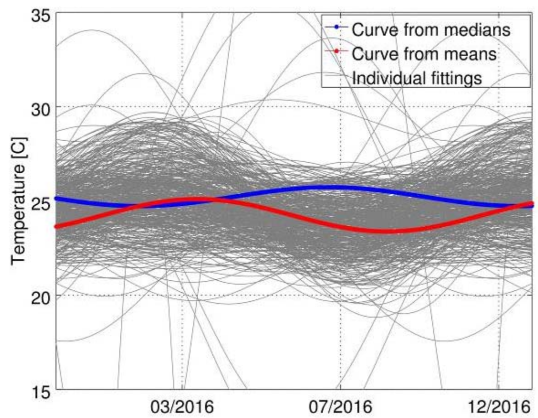

Fig 7. Representation of the independent fitting curves and those based on median and mean values.

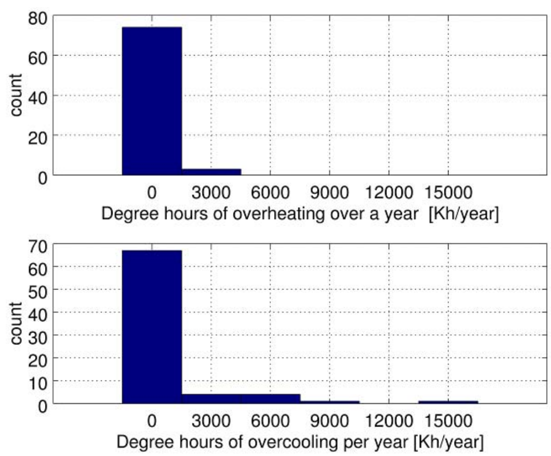

Fig. 8. Histograms representing degree hours of overheating(top) and overcooling (bottom).

The histograms showed in Fig. 8 show how the overheating is rather limited in the set of users under study. The mean of the distribution shown in the histogram has a value of 292.7 Kh, what is substantial. However, in the case of overcooling the figures were more prominent. The mean of the indicators of overheating show a value of 866.4 Kh what is much more significant than the mean value for overheating.

In both cases, the energy used to overheat or overcool does not provide anything to the occupant, so these behaviors are an opportunity to save energy. It was detected that overcooling happens substantially on a large amount of devices. This should be considered in future interventions, as it is very likely that this will happen due to a bad understanding of the controls of the machine.

D. Identification of high and low-consumers

The previous section A showed how users of the conditioning machines can be classified in two different groups those that interact often with the controllers, and those that interact little with the controller, which interestingly are those that use the equipment for more hours in the year.

The discovery of these two groups have been a relevant finding for the development of an automated build-level personalized thermostatic control. It is for this reason that we wanted to exploited further and look for patterns that will identify a particular feature of those two groups.

As the overcooling was detected as the most inadequate behavior (for resulting on the largest waste of energy) we tried to look for a relationship and check if the users falling into an inefficient behavior are somehow related to the two groups previously defined. Fig. 9 shows the degree-hours of overheating against the log of the interactions per hour. It was interesting to see how the group of users that were considered to interact often with the machines, from exp (-5) to exp(0) do not show in general any overcooling. In opposition, are those users that interact with the machines less often those that have registered the larger overcooling. This result, although rather counter intuitive may serve well when defining a cloud based app that has to communicate with users regarding controllers and thermal preferences.

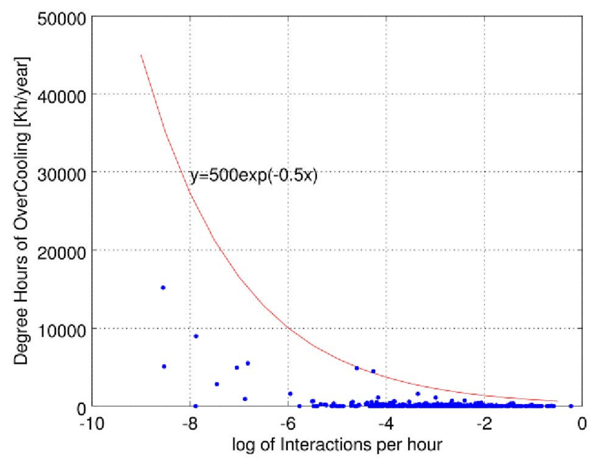

Fig. 9. Relationship between the log of interactions and the overcooling. The red line has been found as upper boundary of the data-points.

No particular relationship between number of interactions with the controls and overheating was found on this study. The users that tend to overheat their rooms (few and with little intensity) were spread between the group that have large interactions and the group that had few.

E. Comparison with external temperature

Adaptive Thermal Comfort demonstrates that people have different comfort temperatures depending on the temperature outside. Particularly, daily means have been used to create comfort temperatures that vary with the weather.

This approach of considering the comfort temperatures a moving parameter depending on the outside temperature may or may not be related with what happen with the set-point temperatures chosen by the users. It is for this reason that we have performed an evaluation of the thermostat temperatures with respect to the outside temperature.

Due to the large volume of data, the graphical representations did not give light to the phenomenon at hand. However, a variety of statistical tests have been used to evaluate on a quantitative manner if there was a relationship between the outside temperature and the set-point temperature chosen by buildings' users. Due to the large amount of data we have performed a linear regression with a subsample of 10% the size of the sample (11e6). The results are shown in Figure 10.

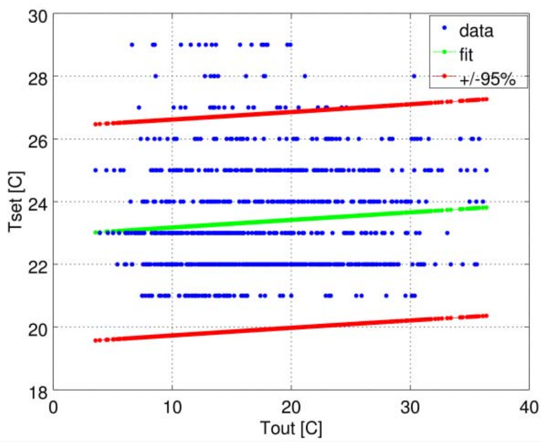

Figure 10. Regression between Outside Temperature and Set Point temperature.

The results of this test show that no clear linear relationship can be found between the set-point temperature and the outside temperature. This is considering the dataset as a whole, if each single individual would have been evaluated and checked for relationship between the dependency of their set-point values and the outside temperature the result might have been different (this is recommended for further work.

TABLE III. STATISTICAL PARAMETERS OF THE REGRESSION BETWEEN TOUT AND TEST.

R2	1.73e-2
F	1.84e4
p-value	4.6e-13
Estimated error variance	3.13

The statistical indicators of the model can be seen in Table III. Also, the parameters obtained in his regression can be seen in Table IV. Other tests have been done trying to fit this data with polynomial of order two and three to try to undercover higher other effects. The results have shown that there is no relationship that can be identify in a higher order. This lack of relationship of the thermostat temperature with respect to the outside temperature shows that in average, large building in which many controllers are available should not take into account the variety of the weather. However, it is possible that independent users show certain correlation between outside temperature and set-point temperature.

TABLE IV. PARAMETERS OF THE REGRESSION.

	Value	*Lower bound*	*Upper bound*
m [C/C]	0.0356	0.0351	0.0362
n [C]	22.71	22.701	22.720

To evaluate to what extent there could be some correlation of outside temperature and set-point temperature on some of the building user's, independent regression analyses were done of the independent users, and their regression coefficient R^2 measured. These R^2 values have been represented on the box plot shown in Figure 11.

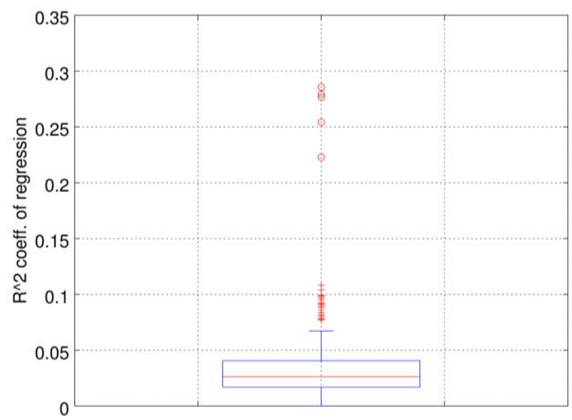

Figure 11. Regression between Outside Temperature and Set Point temperature.

Figure 11 shows how in the majority of cases no correlation was seen between the outside temperature and the set-point temperature. However, it was interesting to see that five out of the 420 units in which the regression was performed (the datasets of the rest of the units did not have enough quality data to perform the test) shows a larger correlation. Although the value is still small in this group (around 27%) it should be noted that it was also on those cases were de largest set-points were seen, and the largest slopes on the fitting line.

V. CONCLUSION

This study shows a seminal work investigating the way in which people interact with their conditioning set-points. There is starting to be growing interest on this aspect as more smart thermostats are starting to appear at the same time that the larger number of IoT devices are allowing the installation of cloud-based algorithms capable of controlling air condition machines, and with the help of crowd-sensing, maximize comfort in all users minimizing energy costs.

This study has shown that users can be classified into two groups, one that tends to interact often with the set-point and keeps changing its value at least once a week and the other that tends to ignore the set-point and do not interact with the machine. This is interesting as if one wanted to develop a solution that should "communicate" with the user, one should consider that some of the users will need continuous feedback informing them that "everything is under control" and that "the system is aware of the current situation (today) and has fitted the system accordingly" fail to take this into consideration could render the solution *unliked*.

In addition to that, we have seen that users when considered in average use thermostat values that are those expected from the comfort studies. However, we have also seen that the individual variations can be rather large. This means that unsatisfied users are likely to appear when a large building is controlled with a single realization of the expected set-point. Nevertheless, we have seen that these individual preferences are rather easy to capture and only few times implying that the characterization of individual user's preferences can be rather easy using crowd-sensing or any other method that allows to capture preferences with a larger granularity.

It was seen that, contrary to what was expected, users that interact with the controllers often, changing the values of the thermostats frequently are less likely to overcool the spaces, problem that has been seen to appear more in the data studied. This implies that feedback reminding the users to set up the thermostat appropriately might be a way of stopping overcooling.

Additionally, it was evaluated if the set-point temperature changed with the specific temperature of the moment. This served to reinforce the previous finding of the no modification of the set point temperature in different times of the year. In this test, we saw no statistical significance that allowed us to state that the set-point temperature changes with respect to the external temperature. As this was preliminary study, simple statistics were used to do this verification, and although one can find them inconclusive, we believe we have at least given the first step in this regard, more evaluation should be done to be conclusive about such important aspect of human behavior on thermostats.

Overall, we have seen that set-point preferences is a phenomenon difficult to capture completely. There exists a great deal of variability, particularly from user to user, and that should be taken into account when implementing cloud based solutions. However, the analysis done in this work has shed light to the phenomenon clarifying important aspects of it. We truly believe that the information obtained in this paper can place developers one step closer to accepted cloud-based thermostat controllers for large buildings, that reduce energy use and maximize users' satisfaction.

ACKNOWLEDGMENT

This paper has been made possible thanks to the support of the European Commission through the H2020-ENTROPY-649849, the Spanish National Project CI-CYT EDISON (TIN2014-52099-R) and MINECO TIN2014-52099-R project (grant BES-2015-071956) granted by the Ministry of Economy and Competitiveness of Spain (including ERDF support). Ramallo-González would like to thank the program Saavedra Fajardo (grant number220035/SF/16) funded by Consejería de Educación y Universidades of CARM , via Fundación Séneca-Agencia de Ciencia y Tecnología de la Región de Murcia.

REFERENCES

[1] IPCC, 2013: Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK and New York, NY, USA.

[2] International Energy Agency. Online Resource: www.iea.org. Accessed 14/03/2018.

[3] Anna Carolina Menezes, Andrew Cripps, Dino Bouchlaghem, Richard Buswell, Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap, Applied Energy, 97 (2012) 355-364.

[4] Ian Richardson, Murray Thomson, David Infield, A high-resolution domestic building occupancy model for energy demand simulations, Energy and Buildings, 40, 8, 2008, 1560-1566.

[5] Rouleau, J., Ramallo-González, A. P., Gosselin, L., Towards A Comprehensive Tool To Model Occupant Behaviour For Dwellings That Combines Domestic Hot Water Use With Active Occupancy, Proceedings of the 15th IBPSA Conference. San Francisco, CA, USA, Aug. 7-9, 2017.

[6] Nataliya Mogles, Ian Walker, Alfonso P. Ramallo-González, JeeHang Lee, Sukumar Natarajan, Julian Padget, Elizabeth Gabe-Thomas, Tom Lovett, Gang Ren, Sylwia Hyniewska, Eamonn O'Neill, Rachid Hourizi, David Coley, How smart do smart meters need to be? Building and Environment 125 (2017) 439-450,

[7] Darby, S. Smart metering: what potential for householder engagement? Building Research & Information 38 (2010) 442-457

[8] Hargreaves, T.; Nye, M. and Burgess, J. Making energy visible: A qualitative field study of how householders interact with feedback from smart energy monitors, Energy Policy 38 (2010) 6111 - 6119

[9] Shipworth, M.; Firth, S. K.; Gentry, M. I.; Wright, A. J.; Shipworth, D. T. & Lomas, K. J. (2010), 'Central heating thermostat settings and timing: building demographics', Building Research & Information 38(1), 50-69.

[10] Andersen, R. V., Olesen, B. W., & Toftum, J. (2011). Modelling occupants' heating set-point preferences. In Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association (pp. 1451-1456)

[11] Stancovic, J. A., Research directions for the internet of things, IEEE Internet of Things Journal 1 (2014) 3-9. Research directions for the internet of things IEEE Internet of Things J., 1 (1) (2014), pp. 3-9.

[12] EC, Benchmarking smart metering deployment in the EU-27 with a focus on electricity. Report from the Commission (2014).

[13] Cheng Fan, Fu Xiao, Yang Zhao. A short-term building cooling load prediction method using deep learning algorithms, Applied Energy 195 (2017) 222-233.

[14] González-Vidal, A., Ramallo-González, A. P., Terroso-Sáenz, F., & Skarmeta, A. (2017, December). Data driven modeling for energy consumption prediction in smart buildings. In Big Data (Big Data), 2017 IEEE International Conference on (pp. 4562-4569). IEEE.

[15] González-Vidal, A., Moreno-Cano, V., Terroso-Sáenz, F., & Skarmeta, A. F. (2016). Towards energy efficiency smart buildings models based on intelligent data analytics. Procedia Computer Science, 83, 994-999

[16] J.F. Nicol, M.A. Humphreys, Adaptive thermal comfort and sustainable thermal standards for buildings, Energy and Buildings, Volume 34, Issue 6, 2002, 563-572.