The use of cluster analysis to assess thermal comfort in university classrooms

Inaiele Mendes da Luz[1], *Iasmin* Lourenço Niza[1], and *Evandro Eduardo* Broday[1].

[1]IEQ *Lab*, Federal University of Technology – Paraná (UTFPR), Rua Doutor Washington Subtil Chueire, 330, Jardim Carvalho, Ponta Grossa, PR, 84017-220, Brazil

Abstract. Nowadays, providing health, well-being, productivity and energy efficiency to users inside buildings is essential. Applying these aspects aligned with sustainability becomes necessary to reduce the use of heating, ventilation, and air-conditioning (HVAC) systems. These systems are currently used to provide better thermal conditions to the occupants, who spend around 80% of their time indoors. The actual thermal conditions can be affected by several factors, such as the climatic type of the region, orientation, size, building type, and energy levels, among others. To assess thermal conditions inside buildings, several thermal comfort models have been developed over the years. However, the Predicted Mean Vote (PMV) created by Fanger is still the most common model to assess thermal comfort indoors. In this context, this research aimed to analyze thermal comfort conditions in university classrooms in Southern Brazil. By collecting the environmental and personal variables of thermal comfort and the mean thermal sensation of students through measurements and questionnaires, a total of 519 responses were obtained during the Brazilian autumn. A statistical cluster analysis was performed to classify individuals according to their sensations. Differences between genders were verified and changing indoor temperatures lower in winter would therefore save HVAC energy without impacting occupant comfort.

1 Introduction

Energy consumption in environments has become increasingly prominent, especially in the construction sector. Thus, several researchers point out that the presence of variability in indoor environmental conditions makes it necessary to reduce energy consumption and improve the perception of comfort [1]. Moreover, buildings' indoor environmental quality (IEQ) and energy distribution should be made available smartly, according to their demand, and directed to people [2].

To understand what makes an environment considered thermally comfortable, several models have been developed to verify thermal comfort under environmental and personal aspects. Thermal comfort consists of a physical-physiological process that describes the thermal sensation of people [3]. Among the most usual models is the Predicted Mean Vote (PMV) developed by Fanger [4], which performs only 34% in results [5], and adaptive models, which can show more reliable results in some cases [6].

Conducting thermal comfort studies is of paramount importance for the health of environmental users, improved productivity, and sustainable development. In various situations, heating, ventilation, and air conditioning (HVAC) systems become alternatives to improve thermal comfort. However, these systems are drivers of high energy consumption in buildings and are associated with indoor environmental quality [7]. Moreover, the presence of faults in these systems causes damage to the indoor environment and reduces energy efficiency [8].

Several thermal comfort studies have been conducted in various countries and environments, such as nursing homes in Spain [9]; parks in China [10]; offices in Australia [11]; classrooms in Ecuador [12], among others. In conjunction with these studies, numerous statistical methods are applied to investigate thermal comfort, such as factor analysis [13]; Bayesian statistics [14]; logistic regression [15]; Griffiths analysis [16], and discriminant analysis [17]. Another technique used is cluster analysis which seeks to separate objects into groups according to their maximum level of internal similarity [18].

Bennetts et al. [19] applied cluster analysis to check people's characteristics to develop thermal personalities based on beliefs, ideas, and location. Lin and Tsai [20] used clusters to recognize tree species capable of enhancing thermal comfort. Anjos et al. [21] investigated climate data to check which days had similar weather conditions.

In this context, the main objective of this research is to analyze thermal comfort conditions in university classrooms in southern Brazil by using cluster analysis, to classify individuals according to Thermal Sensation Vote (TSV), Predicted Mean Vote (PMV), Thermal Preference Votes (PREF) and percentage of dissatisfied (PPD), verifying the differences between genders.

2 Materials and Methods

2.1 Characterization of the area, building, and study population

Data were collected in classrooms at the Federal University of Technology - Paraná (UTFPR), in the city of Ponta Grossa, southern Brazil, under a humid temperate climate with moderately hot summer (Cfb) [22], categorized by the Köppen-Geiger Climate Classification.

The university has several classroom blocks, a restaurant, laboratories, and places for physical activities and sports. The data were collected in blocks L and P classrooms built of traditional masonry. Figure

 https://doi.org/10.1051/e3sconf/202339601105

1 contains the floor plan of the classrooms, which have a capacity for 42 students and an area of approximately 65 m².

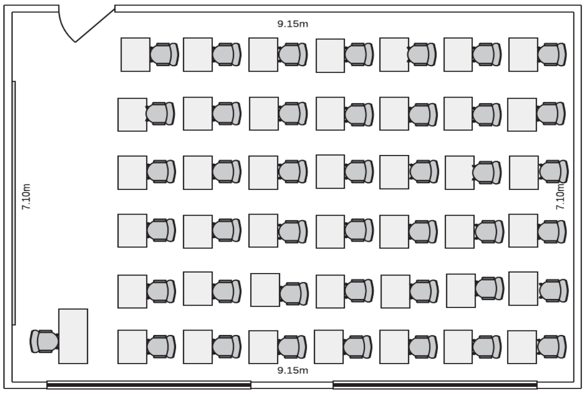

Fig. 1. Representation of the floor plan of the classrooms. [23]

The participants of this study were students regularly enrolled in some undergraduate courses at UTFPR. Regarding the environmental variables used, they had air temperature (°C), mean radiant temperature (°C), air velocity (m/s), and relative humidity (%) that were obtained through 50 measurements that corresponded to 519 valid responses raised between March 23rd to June 14th, 2022. Figure 2 shows the most relevant information about this study.

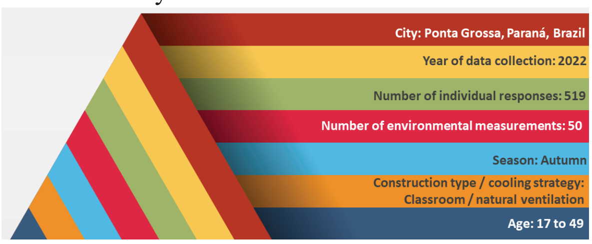

Fig. 2. Study information.

Due to the amount of data collected, it was not possible to obtain the number of people analyzed because the surveys were conducted during the morning, afternoon, and evening shifts; thus, there is the possibility of students being enrolled in courses of more than one shift that are held in the same environment. So, their preferences and thermal sensations may change throughout the day. For these factors, it was opted to count the number of participants through individual responses.

2.2 Data collection and questionnaire application

The environmental and personal variables were obtained as presented in the study by Pereira and Broday [24]. The equipment used resembles a microclimate station BABUC-A, produced by Brüel and Kjaer. The personal variables obtained were age, weight (kg), height (cm), gender (female/male), clothing worn, and Thermal Sensation Votes (TSV) according to the seven-point scale of ISO 7730 [25] (+3 hot, +2 warm, +1 slightly warm, 0 neutral, -1 slightly cool, -2 cool, -3 cold) and thermal preference (PREF) presented in ISO 10551 [26] (+3 much warmer, +2 warmer, +1 a little warm, 0 neither warmer nor cooler, -1 slightly cooler, -2 cooler, -3 much cooler).

The questionnaire was applied digitally, where the participants answered through electronic devices to speed up the information collection process; in the end, only 21 answers were excluded due to incorrectly filled-out fields. The measurements were taken during classes that lasted around 1h40min, being a satisfactory time for the application of the questionnaire and measurement of environmental variables in which the equipment was positioned in the middle of the room 0.6m above the ground as recommended by ISO 7726 [27] for sedentary activities, such as in the classroom.

For 20 minutes, the equipment was turned on before the beginning of the measurement so that it could stabilize itself with the environment in question. Soon after this period, data recording started, which lasted about 40 minutes, with readings every 3 minutes, totaling 13 variables per measurement, where the data averages were obtained. Before the measurement was finished, the questionnaire was applied to the students who agreed to participate voluntarily.

Through the questionnaire, the students indicated the clothes they used so the clothing insulation could be calculated according to ASHRAE 55 [28] and ISO 9920 [29] parameters; furthermore, the air velocity was taken as 0.1 m/s following the studies by Zhou et al. [30] and Singh, Gupta, and Sharma [31]; finally, the metabolic rate was taken as 1.2 met according to ISO 8996 [32] for sedentary activities.

2.3 Software applied in the research

The Center for the Built Environment [33] at the University of Berkeley developed a thermal comfort tool to calculate the PMV and PPD, which was adopted in this research. Then, the data obtained through the equipment and questionnaire were organized in MS Excel® so that the statistical analysis could be done in IBM SPSS Statistics, version 23. To obtain the operative temperature (T_{op}) in each of the measurements, the simple average between the air temperature (t_a) and the mean radiant temperature (t_{rm}) was calculated [34].

To analyze the individual data regarding the 519 responses, an average was performed with the values of each measurement corresponding to the TSV, PREF, PMV, and PPD for each gender. In other terms, for each of the 50 measurements taken in the classroom, it was expected to get one average value of each variable for women and another for men, that is, 100 average values. However, in six of these measurements, there were only students of a single gender, thus totaling 94 mean values, 45 for females and 49 for males.

Thus, it was possible to perform the K-means Cluster analysis to classify the individuals with the greatest thermal similarity. At first, the variables are standardized so that they contribute in a uniform way to the results. Soon after the standardization, the cluster analysis is started, generating several results, such as the variations in the centers of the clusters over the iterations; ANOVA, which allowed to find out which variable contributed more to the separation of the groups

 https://doi.org/10.1051/e3sconf/202339601105

and to classify the clusters according to their performance:

- Positive mean values for most variables: high performance/low risk.
- Negative average values for all or most variables: low performance/high risk.
- Average values close to zero: average performance/average risk.

Next, the distance matrix between the centroids of each cluster and the number of clustered cases is presented, as well as its graphical representation.

3 Results

3.1 Cluster Analysis

The variation history of cluster centers in each iteration is shown in Table 1:

Table 1. Iteration history of the clusters

Iteration	1	2	3
1	2.095	2.005	1.929
2	0.385	0.047	0.125
3	0.233	0.046	0.051
4	0.119	0.033	0.000
5	0.000	0.000	0.000

According to SPSS, the minimum distance was 4.967 between the initial centers, so the algorithm continues until there is no more significant variation in the centroids of each cluster.

Table 2 shows the ANOVA, where the variable highlighted in green is the one with the best discrimination among clusters (PPD, 74.305), and the one highlighted in red has the worst discrimination (PREF, 43.701). Thus, the F values verified how significant the variables were for forming clusters according to the level of similarity, presenting their respective contributions.

Table 2. ANOVA

	Cluster		Error	
	Mean Square	**df**	**Mean Square**	**df**
Zscore (TSV)	23.752	2	0.500	91
Zscore (PREF)	22.781	2	0.521	91
Zscore (PMV)	28.781	2	0.389	91
Zscore (PPD)	28.840	2	0.388	91
	F.		**p-Value**	
Zscore (TSV)	47.506		<0.001	
Zscore (PREF)	43.701		<0.001	
Zscore (PMV)	73.903		<0.001	
Zscore (PPD)	74.305		<0.001	

F = F-statistic; df = degree of freedom.

F-tests should only be used for descriptive purposes because the clusters were chosen to maximize the differences between cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as testing the hypothesis that the cluster averages are equal. Table 3 presents the means for each variable responsible for forming the clusters.

Table 3. Final group centers

Clusters	1	2	3
Zscore (TSV)	-1.38005	-0.01935	0.86983
Zscore (PREF)	1.22173	0.10102	-0.95125
Zscore (PMV)	-1.52049	-0.02035	0.95625
Zscore (PPD)	1.72151	-0.15940	-0.68859

Through the values obtained for the centers of the final clusters, it was possible to classify them according to their performance level, as follows:

- Cluster 1: medium performance / medium risk.
- Cluster 2: low performance / high risk.
- Cluster 3: medium performance / medium risk.

The distance matrix between the centroids of the clusters is presented in Table 4.

https://doi.org/10.1051/e3sconf/202339601105

Table 4. Matrix of distances between cluster centers

Cluster	1	2	3
1		2.983	4.661
2	2.983		1.770
3	4.661	1.770	

Table 5 contains the number of cases in each of the clusters formed.

Table 5. Number of cases in each cluster

Cluster	1	2	3
Number of cases	15	54	25
Total	94		

The values contained in Table 3 have been graphically represented in Figure 3 with the means of the variables for each cluster formed.

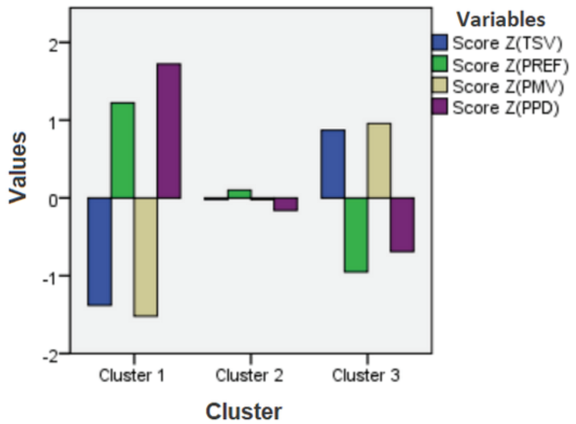

Fig. 3. Graphical representation of the cluster centers

All clusters presented some distinct cases, being:

- The variable PPD contributed the most to the formation of cluster 1, referring to the level of thermal dissatisfaction of people in this classroom.
- The PREF variable was the most prominent in cluster 2, showing that most men and women had similar thermal preferences.
- The PMV variable was in evidence in cluster 3, that around 26.6% of the cases presented values that resembled Fanger's model, close to the average predicted votes.

In Figure 3, the number of cases in each cluster was shown according to gender.

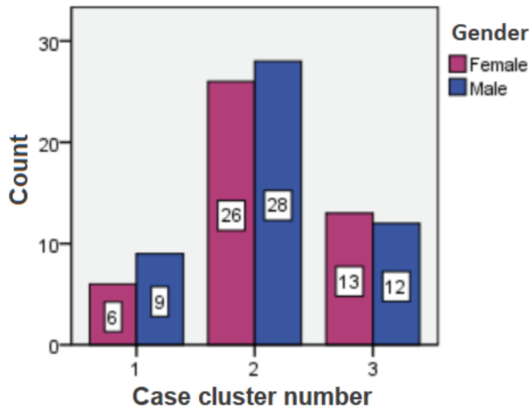

Fig. 4. Number of cases in each cluster by gender

Thus, it was seen in the clusters that the number of men and women are similar, revealing that PREF, PMV, TSV, and PPD are similar and that there is little difference between genders, most likely due to the different thermal insulation of people's clothing. As for clusters 1 and 2, men were in more significant numbers. However, this value was not significantly higher than that of the opposite gender, highlighting that the intra-groups had thermal characteristics in common.

4 Discussion

Through the cluster analysis, it was possible to classify the individuals according to Thermal Sensation Vote (TSV), Predicted Mean Vote (PMV), Thermal Preference Votes (PREF), and predicted percentage of dissatisfied (PPD) in 3 groups. In this way, each case was assigned to only one cluster, so all those with some level of similarity belonged to the same grouping, so there would be no possibility of being at the same time in two or more groups.

Over five iterations, the algorithm for forming the clusters ends when convergence of the data is achieved, containing a minimum distance between the initial centers of 4.967. This is the value that represents the centroid responsible for the separation of the clusters and their optimal solutions [35].

Another relevant detail is understanding the student's profile and preferences in classrooms, as Teli et al. [36] did in their research investigating the temperature profiles in Swedish homes. The behavior of most individuals was highlighted in the second cluster, where thermal preference (PREF) was in evidence, showing that in this environment, individuals had different preferences from the reality calculated by the PMV. Thus, it is necessary to understand the students' demands so they can have thermal comfort in this environment. In line with the second cluster, the PPD was more relevant for forming the first cluster, showing that only a portion said they were satisfied with the environment, highlighting a critical view of their preferences.

Regarding the performance of the clusters, cluster 2 had low performance and a high risk for separation of the groups; in addition, the number of men and women for all clusters were similar, revealing that PREF, PMV, TSV, and PPD have similarities for both genders and dissimilarity between the clusters formed.

Cluster analysis can be applied to indoor environments such as classrooms or outdoor areas, as Acero et al. [37] did in their studies on outdoor thermal comfort to obtain boundary conditions to represent urban microclimate models. In addition, Chang et al. [38] point out that this statistical analysis may also investigate energy consumption related to thermal comfort and occupant behavior. Chen et al. [39] verified in their cluster analysis that women are more predisposed to use adaptive strategies to reduce energy consumption, especially in issues related to lighting.

5 Conclusion

Through cluster analysis, the cases were classified based on their similarities in thermal sensation votes (TSV), thermal preference (PREF), Predicted Mean Vote (PMV), and predicted percentage of dissatisfied (PPD), identifying homogeneity in the data. For cluster 1, 15 cases were allocated; in cluster 2, there were 54 cases, and in the last cluster, there were 25 cases. The second cluster had the highest number of cases allocated, approximately 57%, characterized by intermediate values in all variables. Clusters 1 and 2 allocated the cases with extreme values, such as higher PPD and PREF for cluster 1 and higher PMV and TSV for cluster 3. The difference in cases by gender in each cluster was not statistically significant.

Among the limitations found in the research are the number of measurements taken and the collection period; therefore, if both had a more significant number and a more extended period to be taken, we could further improve the accuracy of the results and be closer to reality. It is suggested, for future work, the performance of discriminant analysis to bring even more substantiation to the results. Maroco [40] emphasizes that although cluster analysis contains a rigorous classification, in general, it is still a technique that does not have solid theoretical foundations; therefore, bringing complementary analysis to estimate the probability of errors becomes relevant to the research.

References

1. A.K. Mishra, M.G.L.C. Loomans, J.L.M. Hensen, Thermal comfort of heterogeneous and dynamic indoor conditions—An overview. Build. Environ. **109**, 82-100 (2016)
2. B. Yang, X. Li, Y. Hou, A. Meier, X. Cheng, J.-H. Choi, F. Wang, H. Wang, A. Wagner, D. Yan, A. Li, T. Olofsson, H. Li, Non-invasive (non-contact) measurements of human thermal physiology signals and thermal comfort/discomfort poses-A review. Energy Build. **224**, 110261 (2020)
3. J. Kim, S. Schiavon, G. Brager, Personal comfort models– A new paradigm in thermal comfort for occupant-centric environmental control. Build. Environ. **132**, 114-124 (2018)
4. O. P. Fanger, Thermal comfort: analysis and applications in environmental engineering. (New York: McGrawHill Book Company (1970)
5. T. Cheung, S. Schiavon, T. Parkinson, P. Li, G. Brager, Analysis of the accuracy on PMV – PPD model using the ASHRAE Global Thermal Comfort Database II. Build. Environ. **153**, 205-217 (2019)
6. Y. Jiao, H. Yu, Y. Yu, Z. Wang, Q. Wei, Adaptive thermal comfort models for homes for older people in Shanghai, China. Energy Build. **215**, 109918 (2020)
7. W.W. Che, C.Y. Tso, L. Sun, D.Y.K. Ip, H. Lee, C.Y.H. Chao, A.K.H. Lau, Energy consumption, indoor thermal comfort and air quality in a commercial office with retrofitted heat, ventilation and air conditioning (HVAC) system. Energy Build. **201**, 202-215 (2019)
8. S. Miyata, J. Lim, Y. Akashi, Y. Kuwahara, K. Tanaka, Fault detection and diagnosis for heat source system using convolutional neural network with imaged faulty behavior data. Sci. Technol. Built. Environ. **26**, 1, 52-60 (2020)
9. N. Forcada, M. Gangolells, M. Casals, B. Tejedor, M. Macarulla, K. Gaspar, Summer thermal comfort in nursing homes in the Mediterranean climate. Energy Build. 229, 110442 (2020)
10. W. Guo, L. Jiang, B. Cheng, Y. Yao, C. Wang, Y. Kou, S. Xu, D. Xian, A study of subtropical park thermal comfort and its influential factors during summer. J. Therm. Biol. **109**, 103304 (2022)
11. D.M. Rowe, Activity rates and thermal comfort of office occupants in Sydney, J. Therm.Biol. **26**, 4-5, 415-418 (2001)
12. G. Guevara, G. Soriano, I. Mino-Rodriguez, Thermal comfort in university classrooms: An experimental study in the tropics. Build. Environ. **187**, 107430 (2021)
13. Q. Xue, X. Yang, F. Wu, A two-stage system analysis of real and pseudo urban human settlements in China. J. Clean. Prod. **29**, 126272, (2021)
14. Z. Wang, T. Hong, Learning occupants' indoor comfort temperature through a Bayesian inference approach for office buildings in United States. Renew. Sust. Energ. Rev. **119**, 109593 (2020)
15. H. Du, Z. Lian, D. Lai, W. Liu, L. Duanmu, Y. Zhai, B. Cao, Y. Zhang, X. Zhou, Z. Wang, X. Zhang, Method of determining acceptable air temperature thresholds in Chinese HVAC buildings based on a data-driven model. Energy Build. **241**, 110920 (2021)
16. R.F. Rupp, J. Kim, E. Ghisi, R. de Dear, Thermal sensitivity of occupants in different building typologies: The Griffiths Constant is a Variable. Energy Build. **200**, 11-20 (2019)
17. S. Shan, E.-H. Yang, J. Zhou, V.W.-C. Chang. Human-building interaction under various indoor temperatures through neural-signal

electroencephalogram (EEG) methods. Build. Environ. **129**, 46-53 (2018).

18. A.A. Shahbazi,V. Esfahanian. Reduced-order modeling of lead-acid battery using cluster analysis and orthogonal cluster analysis method. Int. J. Energy Res. **43**, 13, 6779-6798 (2019)
19. H. Bennetts, L.A. Martins, J. Van Hoof, V. Soerbato. Thermal Personalities of Older People in South Australia: A Personas-Based Approach to Develop Thermal Comfort Guidelines. Int. J. Environ. Res. Public Health, **17**, 22, 8402 (2020)
20. Y.-H. Lin, K.-T. Tsai. Screening of Tree Species for Improving Outdoor Human Thermal Comfort in a Taiwanese City. Sustainability, **9**, 3, 340, (2017)
21. M. Anjos, A.C. Targino, P. Krecl, G.Y. Oukawa, R.F. Braga. Analysis of the urban heat island under different synoptic patterns using local climate zones. Build. Environ. **185**, 107268 (2020)
22. CLIMATEDATA.ORG, Clima Ponta Grossa (Brasil), <https://pt.climate-data.org/america-do-sul/brasil/parana/ponta-grossa-4493/> (Accessed 26 november 2022).
23. A.M. Bueno, I. M. da Luz, I. L. Niza, E.E. Broday. Hierarchical and K-means clustering to assess thermal dissatisfaction and productivity in university classrooms. Build. Environ. **233**, 110097 (2023)
24. P.F.C. Pereira, E.E. Broday. Determination of Thermal Comfort Zones through Comparative Analysis between Different Characterization Methods of Thermally Dissatisfied People, Buildings. **11**, 8, 320 (2021)
25. International Organization for Standardization, ISO 7730 Ergonomics of the thermal environment – analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria, Genève, Switzerland (2005)
26. International Organization for Standardization, ISO 10551 Ergonomics of the physical environment – subjective judgement scales for assessing physical environments, Genève, Switzerland (2019)
27. International Organization for Standardization, ISO 7726 Ergonomics of the thermal environments instruments for measuring physical quantities, Genève, Switzerland (1998)
28. American Society of Heating, Refrigerating and Air Conditioning Engineers, Ashrae standard 55-2020 Thermal environmental conditions for human occupancy, Atlanta (2021)
29. International Organization for Standardization, ISO 9920 Ergonomics of the thermal environment – estimation of thermal insulation and water vapour resistance of a clothing ensemble, Genève, Switzerland (2007)
30. M. Zhou, T.T.T. Huynh, P.W.G.G. Koerkamp, I.D.E. Van Dixhoorn, T. Amon, A.J. Aarnink. Effects of increasing air temperature on skin and respiration heat loss from dairy cows at different relative humidity and air velocity levels. J. Dairy Sci. **105**, 8, 7061–7078 (2022)
31. L.K. Singh, A.K. Gupta, A. K. Sharma. Hybrid thermal management system for a lithium-ion battery module: Effect of cell arrangement, discharge rate, phase change material thickness and air velocity. J. Energy Storage. **52**, 104907 (2022)
32. International Organization for Standardization, ISO 8996 Ergonomics of the thermal environment – determination of metabolic rate, Genève, Switzerland (2004)
33. CBE thermal comfort tool. <https://comfort.cbe.berkeley.edu/> (Accessed 16 august 2022)
34. P. Mičko, A. Kapjor, M. Holučbík, D. Hečko, Experimental Verification of CFD Simulation When Evaluating the Operative Temperature and Mean Radiation Temperature for Radiator Heating and Floor Heating. Processes. **9**, 1041(2021)
35. S. Liang, B. Li, X. Tian, Y. Cheng, C. Liao, J. Zhang, D. Liu. Determining optimal parameter ranges of warm supply air for stratum ventilation using Pareto-based MOPSO and cluster analysis. J. Build. Eng. **37**, 102145 (2021)
36. D. Teli, T. Psomas, S. Langer, A. Trüschel, J.-O. Dalenbäck. Drivers of winter indoor temperatures in Swedish dwellings: Investigating the tails of the distribution. Build. Environ. **202**, 108018 (2021)
37. J.A. Acero, E.J.K. Koh, G. Pignatta, L.K. Norford. Clustering weather types for urban outdoor thermal comfort evaluation in a tropical area. Theor. Appl. Climatol. **139**, 1-2 (2020)
38. C. Chang, N. Zhu, K. Yang, F. Yang. Data and analytics for heating energy consumption of residential buildings: The case of a severe cold climate region of China. Energy Build. **172**, 104-115 (2018)
39. C.-F. Chen, M. De Simone, S. Yilmaz, X. Xu, Z. Wang, T. Hong, Y. Pan. Intersecting heuristic adaptive strategies, building design and energy saving intentions when facing discomfort environment: A cross-country analysis. Build. Environ. **204**, 108129 (2021)
40. Maroco, J. Análise estatística com utilização do SPSS. 3 ed. Lisboa: Edições Sílabo (2007)