

Article

Finding Patterns of Construction Systems in Low-Income Housing for Thermal and Energy Performance Evaluation through Cluster Analysis

Aline Schaefer, Taylana Piccinini Scolaro  and EneDir Ghisi *

Research Group on Management of Sustainable Environments, Laboratory of Energy Efficiency in Buildings, Department of Civil Engineering, Federal University of Santa Catarina, Florianópolis 88040-900, SC, Brazil; alineschaefer.au@gmail.com (A.S.); taylanaps@gmail.com (T.P.S.)

* Correspondence: enedir.ghisi@ufsc.br

Abstract: Energy consumption in buildings depends on many factors, including envelope materials. This paper aims to propose a method based on cluster analysis for finding reference models based on actual construction systems of low-income housing. Such reference models can be used in future thermal and energy performance studies. Data on the envelope composition of a sample of 106 dwellings were obtained through a field survey in Florianópolis, southern Brazil. Cluster analyses were performed to group similar materials and construction systems together, and therefore, a reference model was obtained for each cluster. Computer simulations and hypothesis tests were performed to verify whether the reference models represented the sample. Three reference models were obtained from the cluster analysis. Cluster 1 comprised houses with ceramic-brick walls, concrete floor, and concrete slabs. Cluster 2 comprised houses with ceramic-brick walls, concrete floor, ceramic tiles, and wooden ceilings. Cluster 3 comprised houses with wooden walls, wooden floor, cement tiles, and wooden ceilings. Cluster 1 performed better than the other clusters in the cold season (mean degree-hour was 1299 for cooling and 1361 for heating in the reference model). Cluster 2 performed better the other clusters over the hot season (mean degree-hour was 1014 for cooling and 1451 for heating). Cluster 3 showed the worst performance (mean degree-hour was 3793 for cooling and 2988 for heating). Thus, the hypothesis tests have shown that the three reference models differ from each other and can represent their clusters properly. Cluster analysis was a practical and objective method for obtaining reference models.

Keywords: low-income housing; cluster analysis; thermal behaviour assessment; energy performance; hypothesis tests; reference model; field survey; computer simulation



Citation: Schaefer, A.; Scolaro, T.P.; Ghisi, E. Finding Patterns of Construction Systems in Low-Income Housing for Thermal and Energy Performance Evaluation through Cluster Analysis. *Sustainability* **2023**, *15*, 12793. <https://doi.org/10.3390/su151712793>

Academic Editor: Antonio Caggiano

Received: 6 July 2023

Revised: 14 August 2023

Accepted: 22 August 2023

Published: 24 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In Brazil, the development of public policies aimed at low-income housing has contributed to reducing the housing deficit (which currently is approximately 5.9 million units, according to [1]), of which more than 87% correspond to families with monthly income lower than three minimum wages (i.e., lower than BRL 3960.00 per month). BRL stands for Brazilian Real; as of 6 July 2023, USD 1.00 equals BRL 4.93. In the long term, this process will certainly impact the growth of the Brazilian construction industry, promoting more jobs and income, but also an increase in energy demand. Buildings represent a large share of energy consumption around the world. Global energy consumption in the building sector is estimated at around 30% [2]. In Brazil, according to the National Energy Balance [3], the residential sector was responsible for 28.3% of Brazil's electricity consumption in 2021, second only to the industrial sector, which consumed 40.1%. Due to this context, energy efficiency in low-income housing in Brazil represents an important research area.

Energy consumption in buildings depends on many factors. The combination of the materials that compose the structural systems and the envelope sealing is one of them.

The materials impact energy consumption as they regulate the thermal changes between the building and the external environment, requiring, for example, users to make use of artificial air-conditioners to obtain thermally comfortable environments. The roof and floor allow solar radiation to enter the interior. Due to properties such as thermal storage, solar reflectance, and thermal conductivity, these components have a significant impact on the energy consumption of the building. In turn, the thermal properties of building walls and windows impact indoor temperature and occupant comfort [4]. In this sense, simulation techniques can be used to optimise building components to achieve higher thermal comfort and lower energy consumption. The impact of construction systems on thermal comfort and energy consumption in buildings has been studied by many researchers [5–10].

Thermal and energy performance of buildings can be analysed through computer simulation, but it is necessary to invest substantial time and work, even though the study may be unfeasible. A study on thermal and energy performance for a whole building stock, as for the creation of construction guidelines, requires simplifications due to the impossibility of studying every case. Thus, a possibility is to define a reference model based on the characteristics of a group of existing buildings obtained using statistical analysis. This allows different buildings to be studied in order to reduce their energy consumption by running fewer simulations [11,12]. From the variation of the characteristics of the model, information regarding the improvement of thermal performance, comfort conditions, and reduction in energy consumption, among others, can be obtained.

Governments have made efforts to obtain reference models that can be used in thermal and energy performance studies. The US Department of Energy (DOE) has several reference models of commercial buildings. Also, such models are used to support the development of new versions of ASHRAE Standard 90.1 [13]. In Europe, efforts to determine reference models of each nation's building stock have also been made. In particular, the Typology Approach for Building Stock Energy Assessment (TABULA) project is cited as a reference for determining residential building models and has the collaboration of twenty countries [14]. Brazil has no standardised reference models for analysing the thermal and energy performance of buildings. Thus, studies adopt different building models [15,16], sometimes not representative of the building stock, making it challenging to compare their results. Furthermore, no standard currently determines the best way to obtain a reference model.

As seen, the use of thermal and energy simulations assists in the process of designing increasingly efficient and comfortable buildings. Using reference models in computer simulations allows a better representation of the building as to its geometry, materials, systems, and operation [12], and it is a promising tool in the thermal and energy studies in buildings. Therefore, reference models must be determined for specific localities and countries, considering the different cultures and building typologies. Finally, the reference model can be developed by finding the most representative features of a building stock through cluster analysis. Within this context, this paper proposes a method based on cluster analysis for finding reference models based on actual construction systems of low-income housing. A case study was carried out in Florianópolis, southern Brazil. Thus, it is intended to identify which materials and construction systems are used in such buildings' envelopes and to define reference models that represent the buildings' samples and can be used in future thermal and energy performance studies.

2. Background

2.1. Reference Models for Thermal and Energy Simulations

Reference models are indispensable tools for large-scale studies involving thermal and energy performance of buildings, such as those developed by [17–20]. These studies were developed to determine the building stock and its energy consumption [21,22], classify the existing buildings regarding some specific characteristics [18,20], or even study the impact of new energy efficiency guidelines and materials when applied to existing or new buildings [12,23]. These models can be used to assess the thermal performance of a

building with improvement measures by comparing them with a known, standardised envelope. The reference model should reflect the characteristics typically found in a given building typology and locality. In this way, the building with improvement measures (e.g., different materials or construction strategies) would present better thermal and energy performance than the reference model if it obtains better results in the considered performance indicators [24].

Many studies have been developed worldwide to determine reference models and evaluate their thermal and energy performance. Fumo et al. [25] applied reference models to simplify the process of estimating the hourly energy consumption of a building. They compared the hourly consumption of electricity and fossil fuel obtained from simulation with actual energy consumption from energy bills. They found an error of up to 10% on the hourly energy consumption estimation, proving it unnecessary to perform dynamic simulations of the building.

Theodoridou et al. [26] developed an analysis of the built stock in Greece, structured through the classification of existing data in national terms, categorising the inventory according to the age of the buildings. That allowed the authors to summarise a set of information related to the construction period, such as building typology, construction typology, and equipment related to energy performance. A sample of actual buildings was selected for an energy audit, thermal comfort, and air quality monitoring. Simulations considering user behaviour were conducted to evaluate energy consumption. The simulation results were compared with actual data on average annual energy consumption for heating. The authors affirm that the scenario mapped for the built stock of Greece represents a tool to identify the potential points of improvement of the energy performance of the buildings.

Attia et al. [27] developed reference models for the energy consumption of the residential sector in Egypt. A database on the profiles and patterns of energy use, equipment, construction, and size of a building sample was built. Two representative typologies of flats were developed based on the data on the intensities of internal loads and the energy end-use patterns. Results obtained from the simulation of the reference models were compared with the estimated monthly consumption averages. The comparison showed a good correlation between simulation data and survey data. The simulation of the reference models found a difference of about 2% for the total annual consumption.

Geraldi et al. [28] developed reference buildings based on information about school building stock shapes in Brazil, aiming to evaluate the impact of the shape on buildings' energy benchmarking. The reference buildings were simulated for different scenarios, and the simulation results were used to develop regressive benchmarking models to reproduce the building stock performance. Results showed that the different shapes of reference buildings significantly impact benchmarking buildings using regressions.

Palladino [29] defined reference models representative of the Italian residential building stock considering thermal properties and geometry. Parametric energy simulations of such reference models were performed to evaluate the theoretical deviation of the energy performance gap (i.e., the difference between the calculated and actual energy consumption). This approach allowed the quantification of the energy performance gap and could help develop national energy policies.

The studies presented above show that using reference models can be an efficient way to obtain indicators of thermal and energy performance on a large scale, in the sense that they simplify the studies in this area. However, it is important to note that although the composition of the model is based on an integrated set of characteristics that describe it, the methods for obtaining such models are still largely based on the application of univariate statistical techniques. Therefore, searching for methods based on multivariate statistics is necessary to determine the models. An appropriate response to this problem relies on cluster analysis.

2.2. Cluster Analysis to Determine Reference Models

Cluster analysis is an exploratory, non-theoretical, and non-inferential statistical technique encompassing various algorithms whose purpose is to divide a sample of entities into a smaller number of mutually exclusive groups based on their similarities [30]. There are three considerations to be taken to perform a cluster analysis, i.e., (1) the treatment of input data (variables), (2) the selection of the similarity measure, and (3) the selection of the partitioning technique.

Firstly, the variables selected should characterise the aggregated objects and relate specifically to clustering objectives (i.e., in an energy efficiency study, variables that impact the building performance). Furthermore, it is advisable to standardise the variables in order to obtain more homogeneous variances [30]. Finally, before starting the clustering process, it is also important to check that there are no outliers in the sample. The similarity measure (or dissimilarity) represents the differences or similarities between two entities through a mathematical value [31]. Finally, partition techniques are procedures used as a criterion for separating clusters from a distance between objects (similarity measure). They are classified as hierarchical or non-hierarchical. In hierarchical techniques, only two objects are attached to each other in each step, thus characterising tree formation, called a dendrogram (Figure 1), which represents the similarity level obtained with each union of two clusters. In non-hierarchical partitioning techniques, each cluster has no tree formation, and this occurs interactively. Objects are distributed in clusters simultaneously from seed points. Some objects can be relocated to only resemble the cluster they were allocated to [30–34].

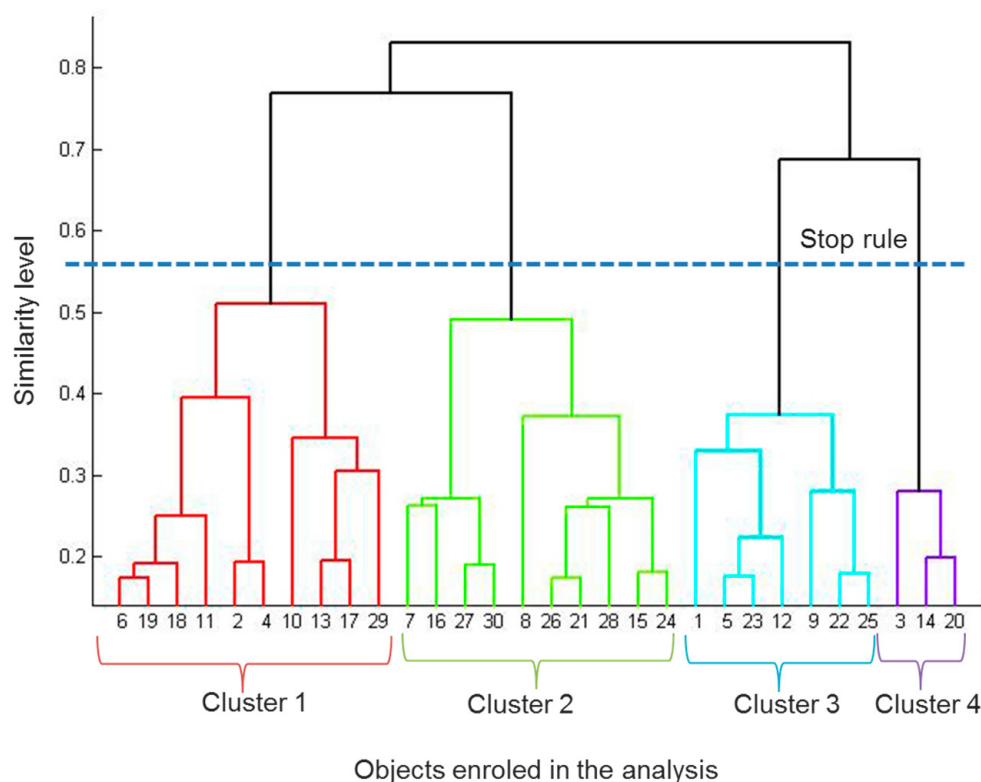


Figure 1. Example of a dendrogram. The horizontal axis presents the objects involved in the analysis; in the case of this work, each number would represent a house. The vertical axis presents the level of similarity obtained in each new union. The cut line, also called a stop rule, indicates when the partitioning process will be interrupted, determining the amount and formation of the clusters.

But how do you obtain a reference model at the end of the whole process? According to [35], one can determine theoretical or real models from the results. The theoretical models are hypothetical models constructed from the average characteristics of each grouping

(average of each variable). As a real model, one solution is to identify the case closest to the centre of the cluster (the nearest distance from the object to the centre of the cluster) and adopt it as a model. This latter process is more practical since there is no need to define characteristics beyond those used in the analysis.

Cluster analysis has been used to investigate the potential for energy savings in buildings and propose classifying buildings as their energy efficiency, among others. The most common objectives of using these studies are to identify different groups in a population according to some specified criteria or to find a way to represent a large group through data simplification. Some applications of cluster analysis to determine reference models can be found in the literature.

Schaefer and Ghisi [19] determined two reference models for low-income housing in southern Brazil to carry out energy efficiency studies for this group of buildings. Data were collected on 100 dwellings over a year, referring to their geometry, such as internal and external dimensions, spatial distribution of environments, solar orientation, and information about openings such as operation mode, sizes, and shading. The collected data were submitted to cluster analysis, which aimed to find subgroups within a sample with high internal homogeneity and heterogeneity between the groups. Hypothesis tests were also conducted to verify the statistical independence between the groups found. Two reference models were determined from the buildings closest to the centre of each group. Finally, each building from the sample and the models were submitted to computational simulation to prove that the model represented the group's thermal performance.

Wang et al. [36] used cluster analysis to determine the shape of a typical office building in Shanghai, China. Based on a field survey, data about the morphology of 181 buildings were collected. Three building characteristics were considered: single-floor area, storeys, and floor height. After the analysis, four clusters were determined. The geometric information of the most representative cluster (61 office buildings) was used to assess the effects of urban morphology factors on the energy performance of neighbourhood-scale building stocks.

A cluster analysis was carried out by Mitra et al. [37] to identify typical schedules followed by the population of the United States. American Time Use Survey data were obtained and split into groups based on age and weekday or weekend. Three main patterns were obtained from cluster analysis: day-work, stay-home, and night-work. Furthermore, different patterns based on age and weekday or weekend were observed. Thus, the results of this analysis can be used to estimate the users' occupancy schedule based on their age and characteristics, which significantly impact the building energy performance.

Liu et al. [38] assessed the correlations between occupancy patterns and socio-demographic characteristics in Chinese households since occupancy patterns significantly impact building energy consumption. For that, cluster analysis was used to determine representative occupancy patterns based on household occupancy data collected through a field survey. Several representative occupancy patterns were observed; each cluster formed showed the probability of users being at home in specific environments and periods.

3. Method

The method applied in this research follows the method proposed by Schaefer and Ghisi [19]. It aims to answer these four main questions: (1) Which are the materials and construction systems used in low-income houses of Florianópolis? (2) Can we split the sample into clusters with similar features and determine a construction system reference model to represent each cluster? (3) Do the clusters differ based on the thermal performance of their houses? (4) Are the reference models found representative of their cluster?

To answer such questions, this study was developed in three steps: (1) obtaining data, (2) clustering data, and (3) checking the representativeness of clusters and models.

In the first step, data regarding the materials and construction systems which compose the dwellings envelope of the sample were obtained. These data were summarised and submitted to cluster analysis in the second step. This analysis allowed us to split the

sample of houses into clusters with similar features. A reference model of features of each cluster was determined. In the third and last step, all the samples, including the reference models, were submitted to computer simulation, from which the degree-hour, a thermal performance indicator, was obtained for each house. That allowed us to verify if the clusters differ in their thermal performance and if the reference models represent their cluster.

3.1. Obtaining Data

Measurements and interviews were conducted through semi-structured questionnaires in low-income single-family houses in Florianópolis, Brazil. The houses considered for the study should have family income equal to or less than three minimum monthly wages or be located in urban areas intended for low-income housing, as established by Brazil's guidelines of national housing programmes. Information about the envelope of 106 low-income houses in Florianópolis was obtained. Although there is no generally accepted rule regarding minimum sample size to perform cluster analysis [39], some authors recommend a sample size of at least ten times the number of variables [40]. This study considered ten variables (representing the composition of the wall, roof, floor, and frames) in cluster analysis, as explained in Section 4.1.

Data collected refer to the materials that compose the walls, roof, floor, and opening frames of each house. Although the most convenient and correct form to acquire this type of data would be from the executive design project of the building, such a design was not available for most of the houses. Thus, this information was obtained by interviewing the residents, who, in many cases, have participated in the construction process and could indicate the materials used. When the residents could not provide such information, the researcher visually inspected its composition.

This inspection identified the combination of materials that make up the construction system. As this is social housing—built with low resources—in some cases, there was no painting or plastering on all internal surfaces, so it was possible to identify the different layers of materials through one or more walls of the house. Even in buildings with plaster, the identification could be made considering the thickness of the walls, resistance to touching, or even observing some leftover material in the backyard of the building. There is usually a standard in the thickness of different wall enclosure materials (e.g., ceramic-brick or concrete block) and slight variation in the available options. All other systems, such as the floor and roof, could also, in most cases, be identified using the same process. A concrete slab or wooden floor could be observed from the outside of the dwelling, which was normally not in contact with the ground. The researcher could also easily identify the type of interior finish (wood, ceramic, etc.). In the case of the roof, the existence of the ceiling and the material used could be identified by the researcher inside the building. The tiles, in turn, could be identified from the outside. The existence of a slab on the roof can be identified through a trapdoor or by observing roof beams. The frame materials were identified in the same way, either by the thickness of the frame or through visual observation.

This procedure was adopted to collect data on the layers of materials of the walls, floor, roof, and frames in each room, as described below.

- Walls: thickness and materials that compose the wall, plaster, mortar, ceramic-brick or concrete block, wood, etc.;
- Floor: floor covering, structure and contact or not with the ground;
- Roof: type of tile, existence or not of a concrete slab and ceiling material, when applicable;
- Frames: frame material.

Despite the data collected having a low level of detail, the reliability regarding the compositions of the construction systems was considered acceptable due to the little variability of the existing compositions for this object of study.

3.2. Cluster Analysis

In this study, the method proposed by Schaefer and Ghisi [19] was used to split the sample into clusters and to determine the reference models for each group. Firstly, data about the houses' envelope obtained in the field were normalised in Z-scores (Equation (1)) prior to cluster analysis, so the results would not be impacted by the differences in the dispersion of the variables involved in the analysis. Also, after the normalisation process, the existence of outliers was assessed using the D^2 of Mahalanobis measure (Equation (2)). The D^2 limit value adopted in this study was 0.001, as suggested by Schaefer and Ghisi [19]. No atypical values were found for the Mahalanobis D^2 test, so the 106 dwellings were submitted to cluster analysis.

$$Z_{x_i} = \frac{(x_i - \bar{x})}{s} \quad (1)$$

where Z_{x_i} is the standardised value of x_i , $\bar{x}$ is the average of variable values, and s is the standard deviation of the variable values.

$$D^2_{nm} = \sqrt{(x_n - x_m)C^{-1}(x_n - x_m)'} \quad (2)$$

where D^2_{nm} is the Mahalanobis measure, C^{-1} is the covariance matrix, x_n is the value of "n" for each variable, x_m is the value of "m" for each variable, "n" represents each object, and "m" represents the multivariate average.

Cluster analysis started with the hierarchical procedure, which was applied in order to identify the ideal number of clusters (grouping of houses) in the sample. The first step to initiate the hierarchical procedure is defining a similarity measure. The Squared Euclidean distance was adopted for all pairs of objects. With this measure, it is possible to accumulate empirical evidence on similarity levels, differentiating itself from Euclidean distance by highlighting the differences between more distant objects [32]. Equation (3) was used to obtain the distances between the objects using the Squared Euclidean distance.

$$d_{AB} = \sum_{i=1}^p (x_{i_A} - x_{i_B})^2 \quad (3)$$

where d_{AB} is the Squared Euclidean distance from A to B, x_{i_A} is the value of A for each variable, and x_{i_B} is the value of B for each variable. In this study, points A and B can be considered houses of the sample, as shown in Figures 2 and 3.

The Ward Method was selected as the partitioning algorithm, a set of rules defining which pairs of clusters are combined in each step. This method considers the best solution the combination that minimises the residual increase in the squares across all the variables in all the clusters [31]. Once the partitioning algorithm and the similarity matrix were defined, the hierarchical procedure was performed. In this procedure, the pairs of objects (dwellings) were combined according to the distance value obtained with the specified similarity measure and the partition rules of the selected algorithm. At each stage, two distinct clusters were combined to form a new cluster. The value of the distance in which each cluster was formed indicates their similarity level.

The formation of the clusters from the hierarchical method can be represented in a dendrogram, a stepped graph in the form of a tree where the levels of similarity obtained to each union are observed. In the early stages, this measure is small and grows as different clusters are combined. Figure 2 exemplifies the construction of a dendrogram.

The dendrogram helped to determine the preliminary clustering solutions, corresponding to the stage in which there was a high increase in the level of similarity in relation to the previous stages. In addition to the dendrogram, the percentage variations of the heterogeneity obtained in each union were also analysed to obtain a suitable number of clusters. A large increase in heterogeneity means that two considerably distinct clusters were joined. The number of clusters to be formed was then determined by the previous solution to a step where the combination of clusters generated a percentage increase in the

agglomeration coefficient (a measure of heterogeneity for each new cluster provided by the statistical programme) more considerable than the previous steps.

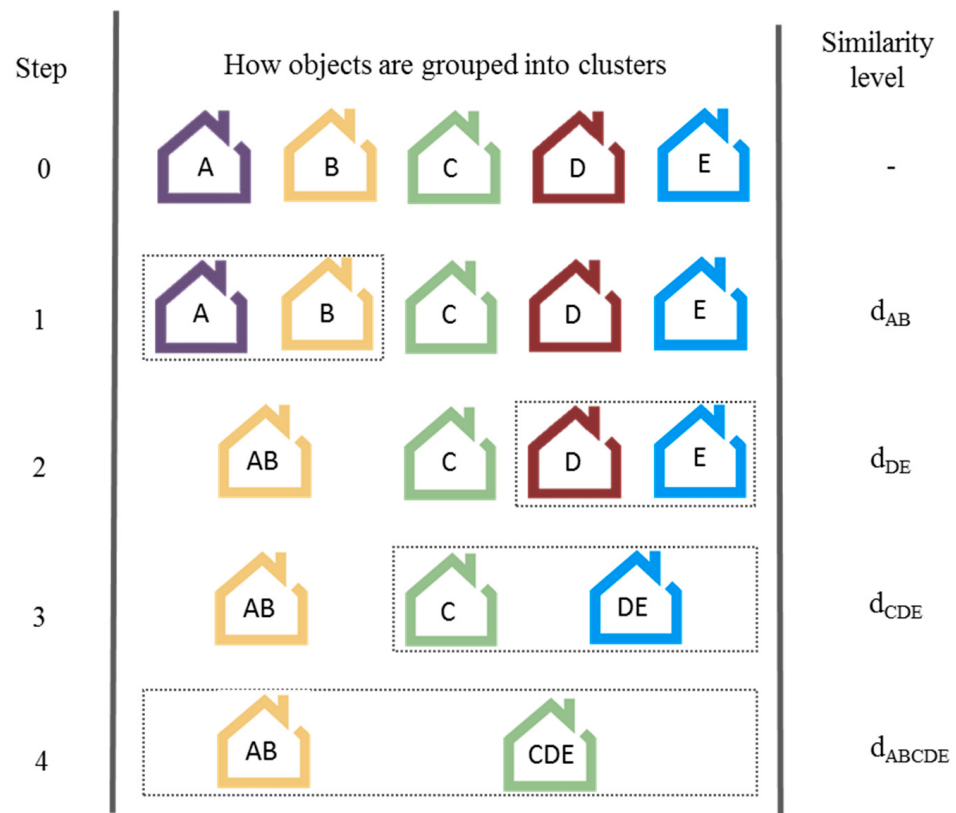


Figure 2. Dendrogram of the construction process.



Figure 3. Floor plan drawing of the model used in computational simulations [19].

After determining the optimal number of clusters, the non-hierarchical procedure was applied to obtain the final solution using the k-means algorithm. In this procedure, the objects (dwellings) are distributed simultaneously in clusters according to their proximity

to the seed points (reference points where the clustering process starts). These points represent the centre of the cluster (centroid), which changes as new objects are grouped. As centroids keep changing, some objects initially assigned to one cluster may become closer to the centroid of another cluster during the process and therefore are re-assigned. The same procedure is repeated the number of times necessary until the convergence is reached (when no object is assigned to a new cluster due to the change of its centre).

Based on this process, the clusters were formed, grouping dwellings with similar envelope materials. Because cluster analysis is an exploratory rather than inferential analysis, there is no “ p_{value} ” reference for comparing the final result to verify if the clustering formation was suitable, since the statistical variable is defined by the data involved in the analysis itself [30]. Therefore, hypothesis tests (Chi-square test) were performed with all variables (representing the composition of the wall, roof, floor, and frames) to verify significant differences between clusters. A significance level of 0.05 was adopted, assuming that the variables are statistically independent when the value obtained using the test is less than 0.05. The reference models were considered as the objects (dwellings) with the smallest distance to the centroid of each cluster, adopted as actual reference models (defined from characteristics of actual dwellings) and not from theoretical reference models (resulting from the combination of the means obtained for each variable).

At last, the construction systems that comprise the walls, floor, roof, and frame of each model were presented.

3.3. Thermal Performance through Computer Simulation

In the last step, to verify the suitability of the clusters and their reference models in terms of thermal performance, the thermal performance of all houses was assessed using EnergyPlus (version 8.1). It was considered the climate of Florianópolis (latitude 27° south, longitude 48° west). For that, virtual models were configured for each composition of the envelope materials found in the field surveys. The other parameters, such as geometry, operation, and internal gains, were fixed for all files so that the performance of the models reflected only the effects caused by the difference in the composition of the envelope materials. These parameters were adopted as proposed by Schaefer and Ghisi [19] and are presented below.

The geometry was configured using the Open Studio plugin within the Sketch Up software (13.0.3689). It is a three-bedroom house with an independent living room and kitchen, a floor plan area corresponding to 76.00 m², and a front façade facing east (Figure 3). Figure 4 shows the room’s occupancy rate concerning the maximum occupancy of each room. It was considered a maximum occupancy of two persons per bedroom, power density equal to 119 W in the living room, 1465 W in the kitchen, 33 W in the main bedroom, and 17 W in the other bedrooms. Figure 5 shows the electrical equipment use of the reference model considering the rate of use for installed power density (119 W for the living room, 1465 W for the kitchen, 33 W for the main bedroom, and 17 W for other bedrooms).

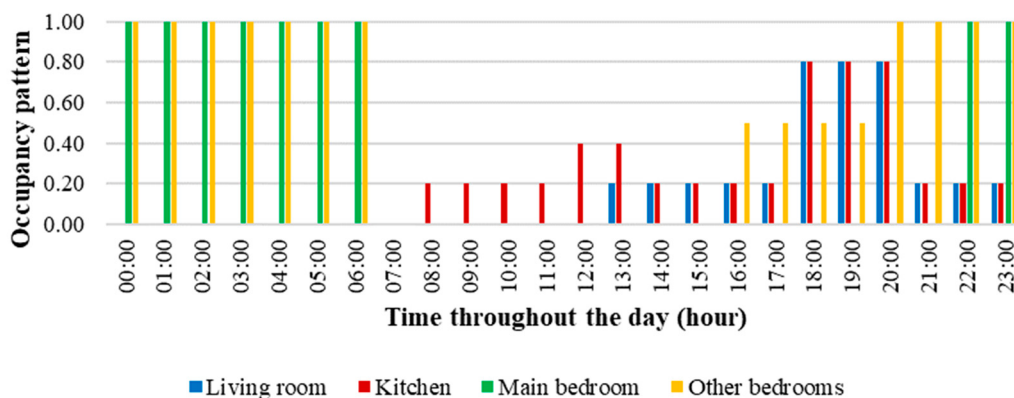


Figure 4. Hourly occupancy of long-term rooms [19].

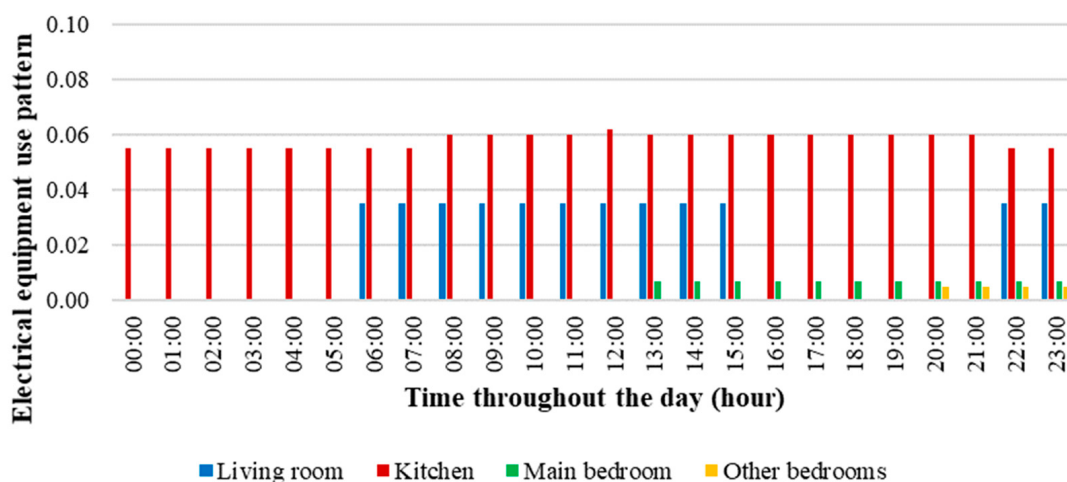


Figure 5. Hourly electrical equipment use in long-term rooms [19].

Table 1 shows the artificial lighting use and the opening operation patterns. It was considered lighting power equal to 30 W, 40 W, 60 W, and 20 W in the living room, kitchen, main bedroom, and other bedrooms, respectively. This table shows the periods when the artificial lighting system is on, and the openings are open.

Table 1. Artificial lighting use and the opening operation patterns.

Use and Operation Patterns				
Artificial Lighting	Environment			
Lightning turned on	Living room 18:00–23:59	Kitchen 18:00–23:59	Main bedroom 21:00–23:59	Other bedrooms 20:00–21:59
Openings operation	Openings			
Openings opened	External door 08:00–19:59	Internal doors 08:00–22:59	Living room and kitchen windows 08:00–22:59	Bedrooms windows 08:00–17:59

Source: based on [19].

From the operative temperatures obtained from the simulations, the heating and cooling degree-hours were calculated for all houses, including the reference model for the living room, bedrooms, and the whole house (Equations (4) and (5), respectively). Additionally, a third indicator, representing the total degree-hours in the year, was obtained by summing the degree-hours of heating and cooling. In order to obtain the total degree-hours for the whole dwelling, a weighting average (Equation (6)) was calculated according to the area of each room. In the end, nine variables were obtained for each dwelling (cooling, heating, and total degree-hours for the living room, bedroom, and whole house).

$$GH_c = \sum (T_o - 26^\circ\text{C}) \quad (4)$$

$$GH_h = \sum (18^\circ\text{C} - T_o) \quad (5)$$

$$GH_{house} = \sum \frac{(GH \times AU)}{AU_{total}} \quad (6)$$

where GH_c is the cooling degree-hour ($^\circ\text{Ch}$) for each long-term room, GH_h is the heating degree-hour ($^\circ\text{Ch}$) for each long-term room, T_o is the operative temperature for each long-term room for each hour of the year ($^\circ\text{C}$), GH_{house} is the weighted average of degree-hours

per floor plan area of each room ($^{\circ}\text{Ch}$), GH is the sum of cooling and heating degree-hours of each room ($^{\circ}\text{Ch}$), AU is the floor plan area of each room (m^2), AU_{total} is the sum of the floor plan areas of all rooms (m^2).

Through the degree-hour obtained, two analyses were used to verify if the clusters and their reference models were adequate. First, a hypothesis test was performed to verify if each cluster mean was the same or differed between clusters, meaning that the clusters had similar or different performances. The significance value was set at 0.05. Thus, variables whose p_{value} was lower than 0.05 were considered not equal.

The second analysis regards the sample distribution for each cluster variable through a boxplot. A boxplot displays the sample data divided into four equal parts (quartiles), each containing 25% of sample cases. The horizontal lines indicate the separation of each quartile. The area between the first and third quartiles makes up 50% of the sample data. Above the third quartile line are 25% of cases, and below the first quartile line, the remaining 25%. The centre of the distribution is indicated by the median, which is the line dividing the rectangle into two equal parts, each containing 25% of the data. According to its position in the rectangle, it is possible to understand the asymmetry of the sample. The height of the rectangle can determine the dispersion of the sample; the higher it is, the more dispersed the sample is. Horizontal lines outside the rectangle represent the upper and lower limits. The cases that fall outside the range delimited by the upper and lower limits may represent outliers and should be analysed by the researcher to determine if they represent a particular case or should be withdrawn from the sample. Thus, the performance of each model was compared to the performance of all houses of its cluster using a boxplot to check its suitability. For that, the reference model value should be located between the first and third quartiles of the distribution. If that criterion is satisfied, the model presents a central tendency and can represent its cluster. This step aims to verify the ability of the reference building to represent the sample and not to evaluate its thermal performance. Boxplots were analysed to verify if the degree-hours of the reference model were close to the median.

4. Results and Discussion

This section shows the results of the three steps of the method proposed.

4.1. Data on Materials and Construction Systems

The compositions of the construction systems found in the field were considerably heterogeneous, requiring simplification. Therefore, the characteristics of the dwellings were summarised in ten variables, representing the composition of the wall, roof, floor, and frames. The wall composition was split into two variables, which described the composition of the walls in the dry areas (such as living room and bedrooms) and the composition of walls in wet areas (such as kitchen and bathroom). The walls were classified into three different types, based on the structural elements of the partitions: wooden walls, ceramic-brick walls, and concrete block walls (Figure 6). The floor was also divided into two variables, considering dry areas and wet areas. The compositions considered were wooden floors, concrete floors without ceramic coating, or concrete floors with ceramic coating. As for the floor, it was also considered whether or not this surface was in contact with the ground (Figure 7). The roofs were classified into three variables: tile type (fibre cement, ceramic, or none), the existence of concrete slab and ceiling (wood, PVC, gypsum, or none) (Figure 8). Finally, the frame material was divided into two variables: door frames (wood, aluminium, or PVC) and window frames (wood, aluminium, PVC, or mixed) (Figure 9).

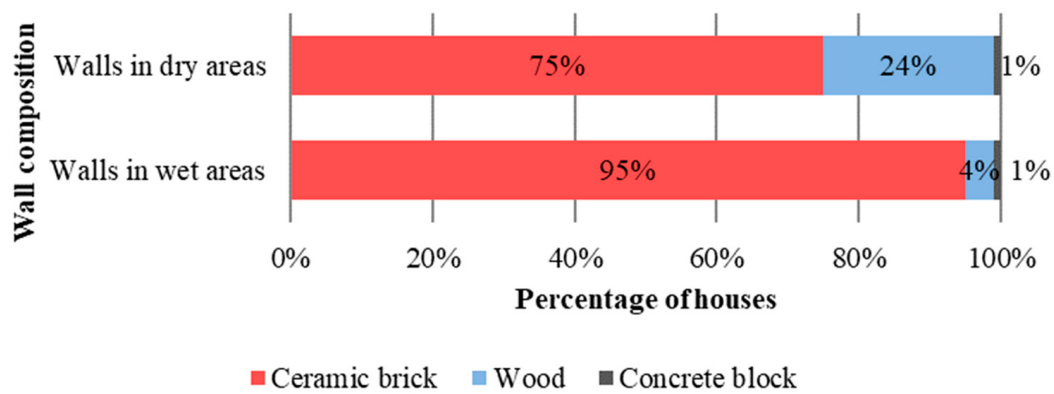


Figure 6. Materials used in the walls.

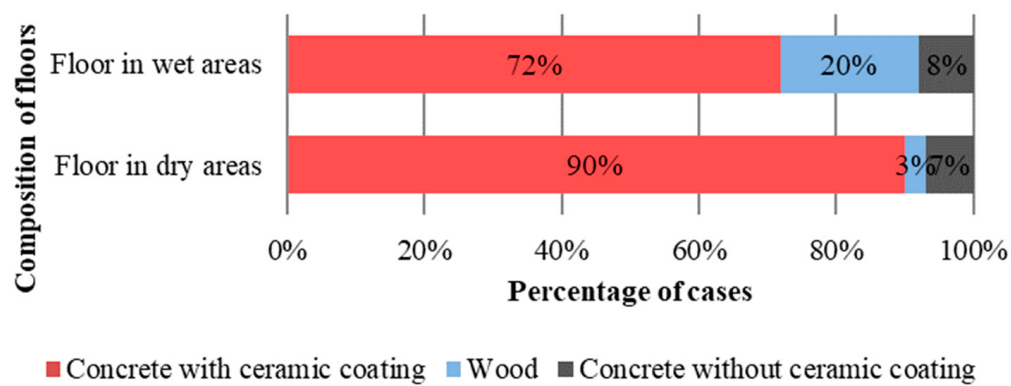


Figure 7. Materials used in the floors.

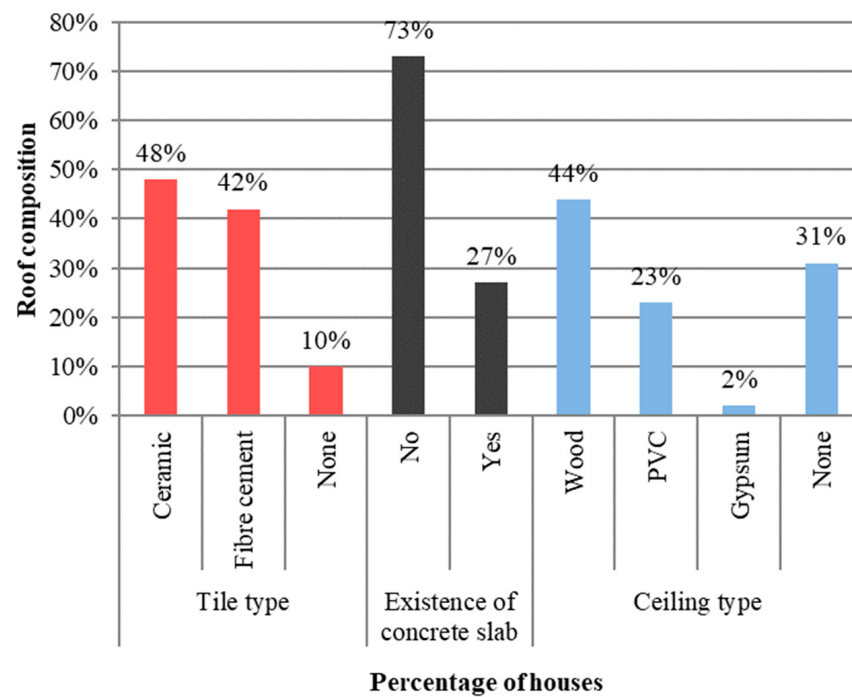


Figure 8. Materials used in the roofs.

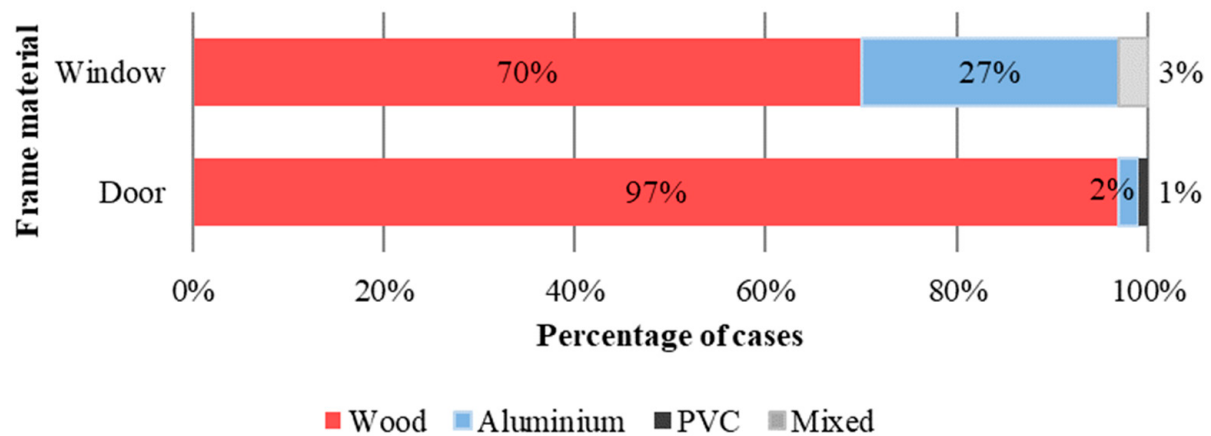


Figure 9. Materials used in the frames.

As can be seen in Figures 6–9, most walls are composed of ceramic brick (75% for dry and 95% for wet areas). Ceramic and fibre cement tiles, the absence of concrete slabs (73% of cases, most of them with ceiling), and wooden ceilings were also predominant. As for floors, the composition of concrete slabs with ceramic coating predominates (72%). Door and window wooden frames are the most common, accounting for 97% of cases.

Figure 10 shows the percentage of cases found for each roof composition. The most frequent was fibre cement tile, with no slab and wooden ceiling (25%), followed by ceramic tile, without slab and with wooden ceiling (19%).

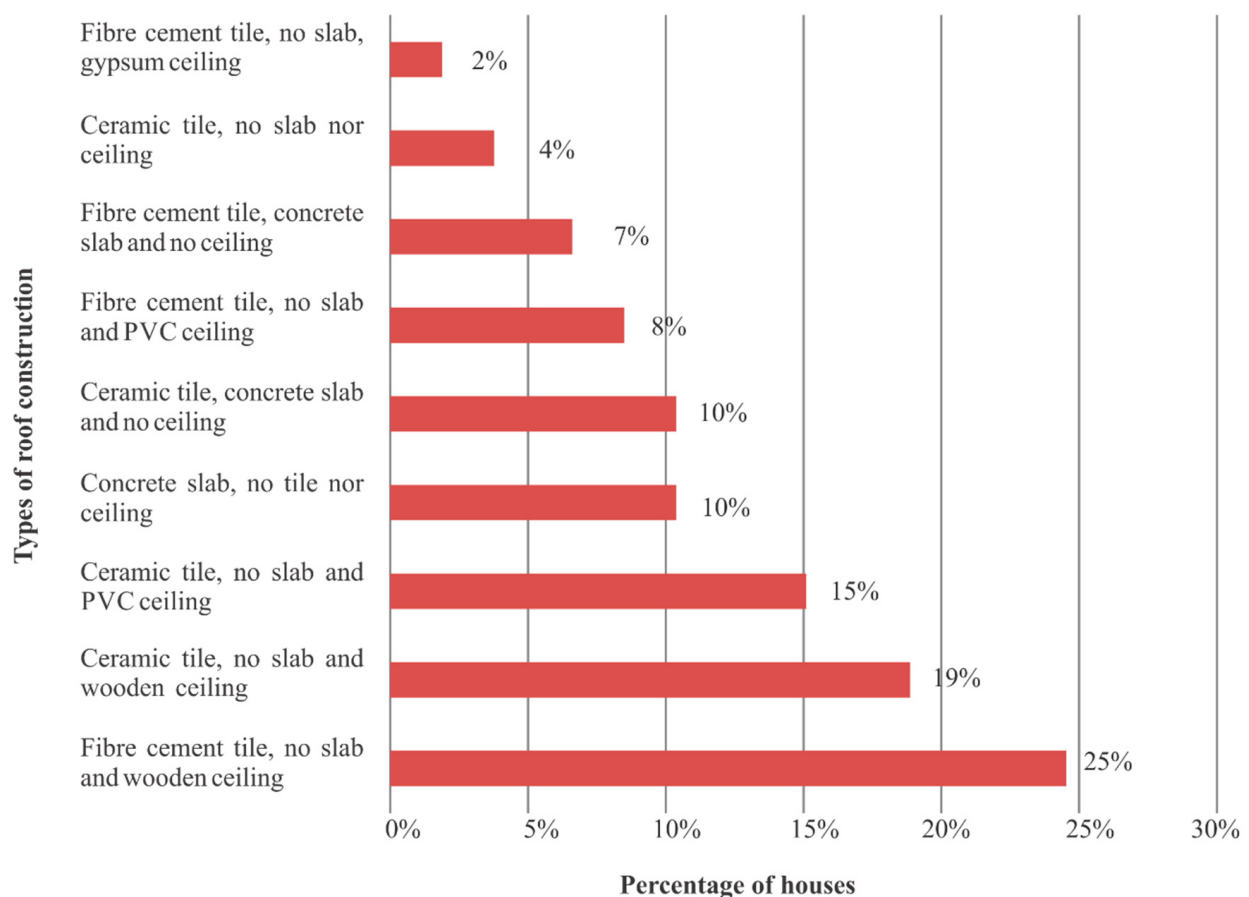


Figure 10. Percentage of occurrence of roof compositions in the sample.

4.2. Clusters

The hierarchical procedure yielded the dendrogram shown in Figure 11. It was possible to verify a substantial increase in the level of similarity after joining two clusters in two moments: in the joining of the two final clusters (similarity level of approximately 25), resulting in two clusters (clusters 1 and 2, with 76 and 30 dwellings, respectively), and the previous step (similarity level of approximately 18), resulting in three clusters (clusters 1, 2 and 3, with 26, 50 and 30 dwellings, respectively).

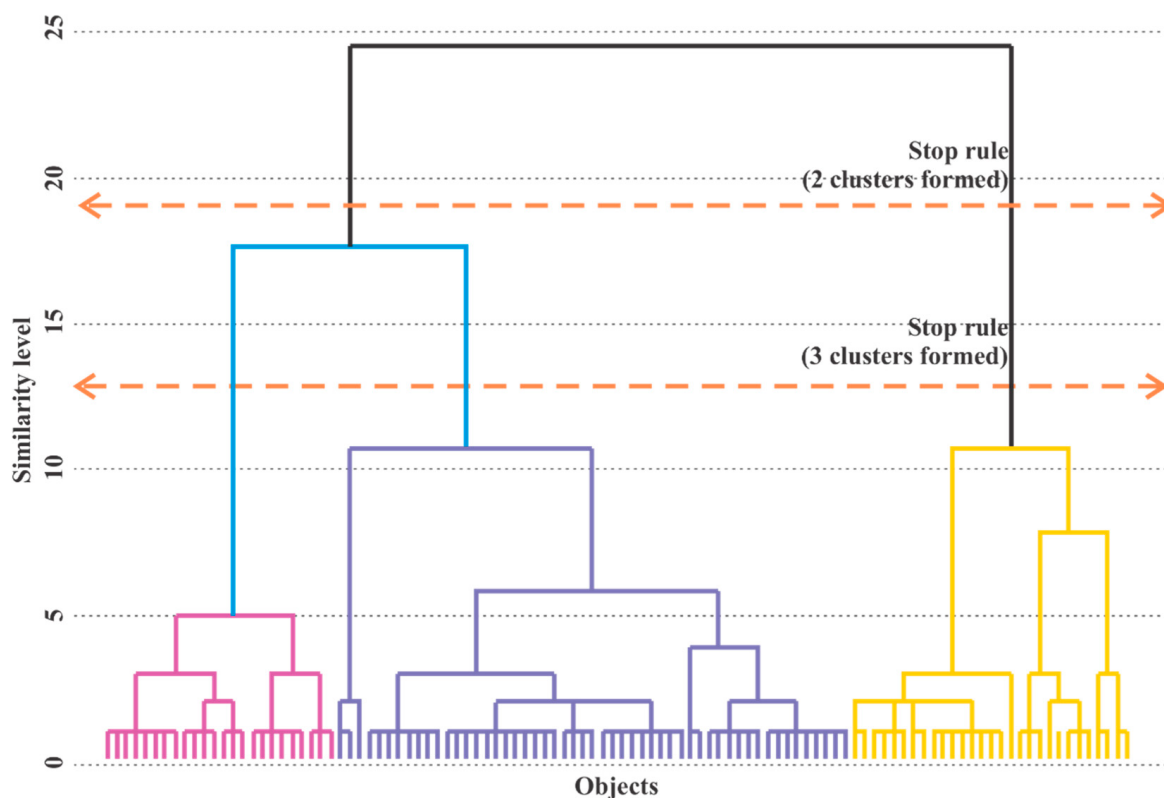


Figure 11. Dendrogram obtained in the hierarchical process.

Table 2 shows the heterogeneity percentages' variation at each stage of the hierarchical process. It can be seen that there are two suitable solutions for this sample, as also shown in the dendrogram. The first one was to stop the clustering in step 103, obtaining three clusters due to the increase of 20.2% in the agglomeration coefficient compared to the previous union. The second solution would be to stop in step 104 and obtain two clusters, with an increase of 24.2% in the agglomeration coefficient.

As in the hierarchical procedure, two solutions were obtained regarding the number of clusters to be formed. Therefore, the non-hierarchical procedure (k-means) was performed for the two alternatives. The final decision to adopt the two- or three-cluster solution was based on two criteria: statistical significance and practical significance.

Table 3 shows the significance values obtained using the Chi-square test for each variable. The best solution for this comparison was identified when having the lowest significance values, i.e., closer to zero. It was verified that for the two cases, the dry area wall composition, dry and wet floor composition, ground contact, and concrete slab on the roof were significant for determining the clusters. The tile type and ceiling material were significant only for the solution of three clusters, while the window frame material was significant only for the solution of two clusters. The wet areas' wall composition and the door frame material were insignificant in any case.

Table 2. Heterogeneity percentages' variation at each stage of the hierarchical process.

Step	Stop Rule		
	Number of Clusters in Each Step	Agglomeration Coefficient	Increase in the Agglomeration Coefficient Compared to the Previous Step (%)
(Previous steps omitted)			
98	8	424.6	8.6
99	7	466.9	9.9
100	6	517.0	10.7
101	5	587.9	13.7
102	4	679.3	15.5
103	3	772.9	13.7
104	2	929.2	20.2
105	1	1155.0	24.2

Table 3. Comparison between both solutions based on the statistical significance (Chi-square test).

Variables	Two-Cluster Solution	Three-Cluster Solution
	(<i>p</i> -Value < 0.05)	(<i>p</i> -Value < 0.05)
Dry areas wall composition	<0.00	<0.00
Wet areas wall composition	0.25	0.48
Dry areas floor composition	<0.00	<0.00
Wet areas floor composition	<0.00	<0.00
Ground contact	<0.00	<0.00
Tile type	0.25	0.00
Existence of concrete slab on the roof	0.04	<0.00
Ceiling material	0.48	<0.00
Door frame material	0.30	0.22
Windows frame material	0.02	0.07

The three-cluster solution is statistically more significant than the two-cluster solution. In addition to having more significantly impacting variables, the variables in which the two-cluster solution is more significant are not as important as those of the three-cluster solution. Also, even though some variables are significant for both solutions, their significance is higher for the three-cluster solution (existence of concrete slab). Thus, it is concluded through the statistical significance that the three-cluster solution is more appropriate.

Regarding the practical significance, the two-cluster profiles are presented in Table 4. The two-cluster solutions differed in drywall composition, wet floor composition, ground contact, and tile type. Thus, cluster 1 would be described as a house with ceramic-brick walls and ceramic-coated concrete floor, both in dry and wet areas, a roof composed of ceramic tile and wooden ceiling, without concrete slab and wooden frames. Cluster 2 would be a house with wooden walls and floor in dry areas and ceramic-brick walls and ceramic-coated concrete floor in wet areas without ground contact. The roof would have fibre cement tiles and a wooden ceiling with no concrete slab. The opening frame would also be made of wood.

Table 4. Comparison between both solutions based on practical significance.

Variables	Two Clusters Solution		Three Clusters Solution		
	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 3
Dry areas wall composition	Ceramic brick	Wood	Ceramic brick	Ceramic brick	Wood
Wet areas wall composition	Ceramic brick	Ceramic brick	Ceramic brick	Ceramic brick	Ceramic brick
Dry areas floor composition	Ceramic-coated concrete floor	Wood	Ceramic-coated concrete floor	Ceramic-coated concrete floor	Wood
Wet areas floor composition	Ceramic-coated concrete floor	Ceramic-coated concrete floor	Ceramic-coated concrete floor	Ceramic-coated concrete floor	Ceramic-coated concrete floor
Ground contact	Yes	No	Yes	Yes	No
Tile type	Ceramic tile	Fibre cement tile	None	Ceramic tile	Fibre cement tile
Existence of concrete slab on the roof	No	No	Yes	No	No
Ceiling material	Wood	Wood	None	Wood	Wood
Door frame material	Wood	Wood	Wood	Wood	Wood
Windows frame material	Wood	Wood	Wood	Wood	Wood

For the three-cluster solution, clusters 1 and 2 were more similar than cluster 3. As for the composition of the walls, clusters 1 and 2 presented ceramic-brick walls in the dry and wet areas, while cluster 3 had wooden walls in dry areas and ceramic-brick in wet areas. Ground contact was observed in clusters 1 and 2, but not in cluster 3. The roof composition was the only difference between clusters 1 and 2, the first consisting of a flat slab, without tile and ceiling, while the second was composed of ceramic tile and wooden ceiling, without concrete slab. In cluster 3, the roof would comprise fibre cement tile and a wooden ceiling without a concrete slab. The material of the frames is wood in all clusters. Cluster 3 was considerably similar to cluster 2 of the two-cluster solution, while clusters 1 and 2 appeared to merge, forming a single cluster. The differences observed for these two clusters (1 and 2) were only due to the roof, of which their thermal properties differ significantly (for cluster 1: $U = 3.73 \text{ W/m}^2\text{K}$ and $CT = 220 \text{ kJ/m}^2\text{K}$; for cluster 2: $U = 2.02 \text{ W/m}^2\text{K}$ and $CT = 26 \text{ kJ/m}^2\text{K}$). Thus, as the assessment of statistical significance also pointed to the solution of three clusters as more adequate, three clusters were adopted in this work.

Three dwellings were selected as the reference models of clusters 1, 2, and 3 due to their higher proximity to the centroid of their cluster (0.620, 0.908, and 1.125, respectively). Figures 12–14 show the compositions of the materials of the reference models in each cluster.

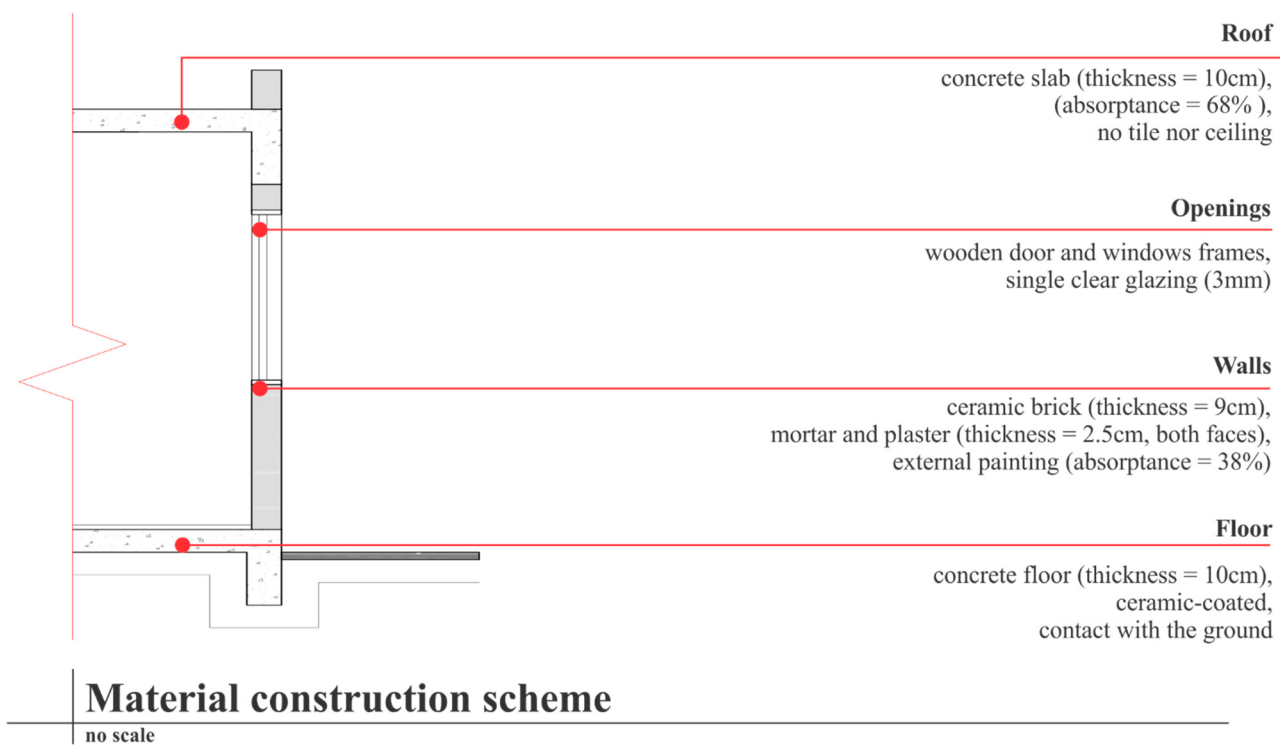


Figure 12. Scheme of the composition of the walls, floor, roof, and frames of the reference model of cluster 1.

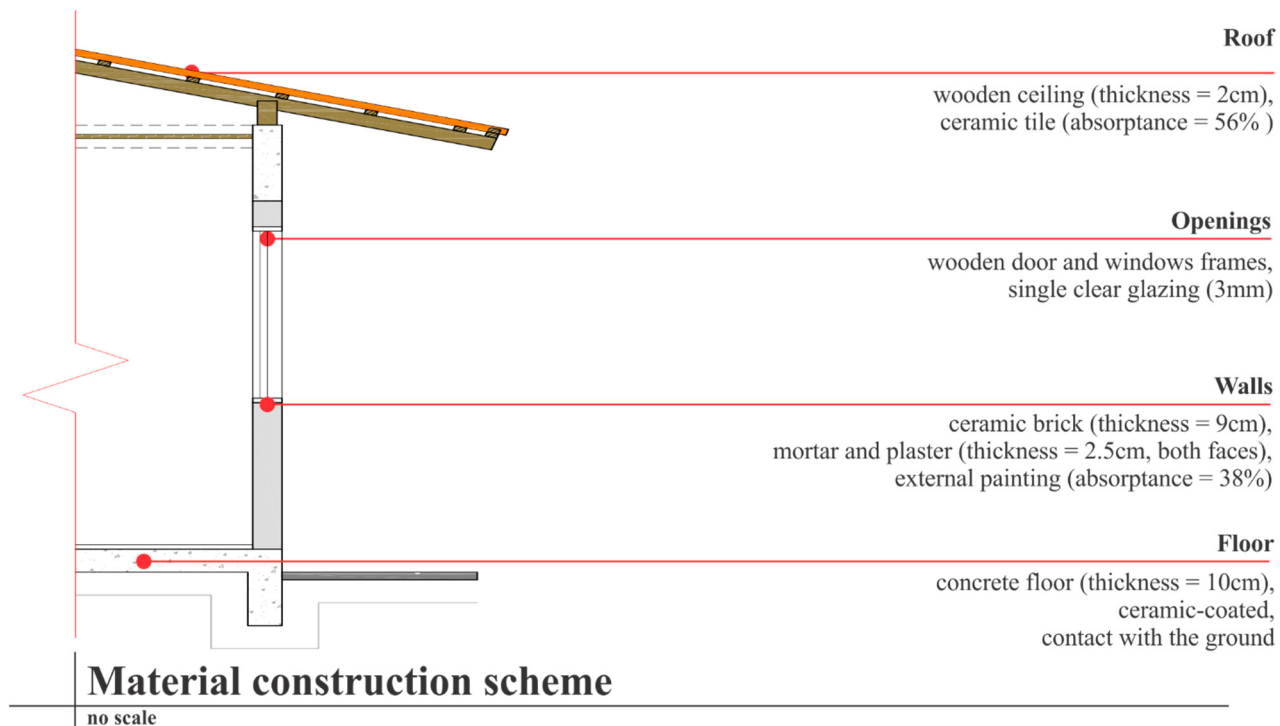


Figure 13. Scheme of the composition of the walls, floor, roof, and frames of the reference model of cluster 2.

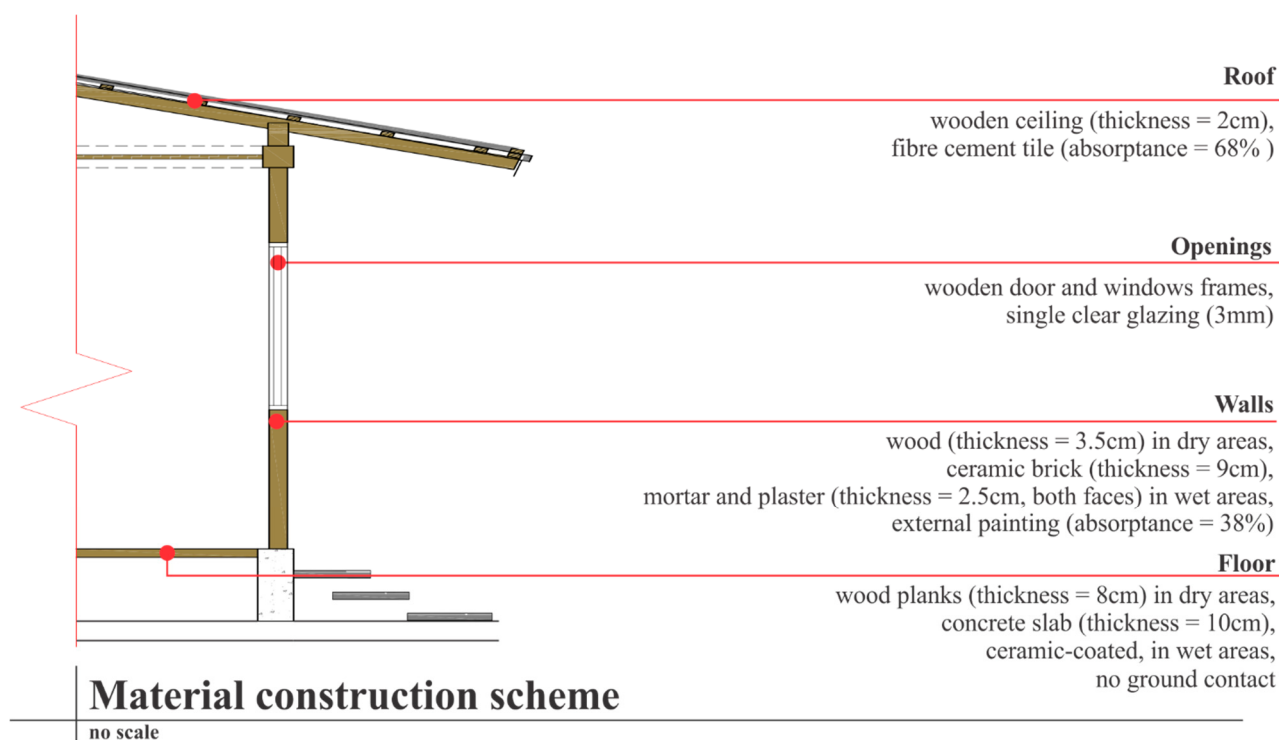


Figure 14. Scheme of the composition of the walls, floor, roof, and frames of the reference model of cluster 3.

The reference model of cluster 1 is a house with ceramic-brick walls, both in dry and wet areas. It was determined that the walls would have 2.5 cm of mortar laying on both sides, and external painting. The average wall absorptance found in the sample was adopted. It was not possible to obtain the roof tile absorptance, so it was adopted based on the literature. The wall thickness was adopted considering the average obtained for the cluster. The floor was composed of a concrete slab and ceramic coated throughout the house, including dry and wet areas. The thickness adopted was 10 cm, based on what is usual, since this characteristic was not obtained in the surveys. The floor is in contact with the ground. As for the roof, it is composed of a concrete slab without tile and ceiling. Doors and windows frames are made of wood. Details on the glass were not used in the analysis due to the impossibility of collecting this information in the field, and 3 mm thick single-pane clear glass was adopted in all windows.

The reference model of cluster 2 has the same characteristics as reference model 1 for walls, floor, and frames, differing only by the composition of the roof. It is composed of ceramic tile and wooden ceilings.

Cluster 3 is the one that differs most among the three clusters. The floor of dry areas is composed of planks of wood, while the wet areas consist of ceramic-coated concrete slabs. Likewise, the walls of the dry areas are different from the walls of the wet areas, as they are made of wood, while those of the wet areas are made of ceramic brick with mortar. The wall thickness was 3.5 cm, corresponding to the average for the cluster. The absorptance was the same as the sample mean. The roof composition was a wooden ceiling and cement tile, whose absorptance was adopted as found in the literature. The frames have the same characteristics as the other clusters.

The definition of the reference models based on their characteristics seems to fit well with what was observed in the surveys, highlighting the existing typologies. Based on the concepts of cluster analysis, it was concluded that the formation of clusters was suitable.

4.3. Suitability of the Clusters and Their Reference Models Based on Thermal Performance

The results of the comparison of means for the three clusters are presented in Table 5. The degree-hour cluster means and the standard errors for each cluster are presented in columns, and the variables are presented in each row. As there are three clusters and it was not possible to prove the normality of data, the Kruskal–Wallis test was applied to verify if the means of the clusters differ. It was observed that the cluster analysis formed clusters with significantly different means for all variables since the p_{value} was less than 0.05 for all cases.

Table 5. Comparison between both solutions based on practical significance.

Rooms	Annual Degree-Hour	Cluster 1 (26 Houses)		Cluster 2 (50 Houses)		Cluster 3 (30 Houses)		<i>p</i> -Value
		Mean (°Ch)	Sd. Error (°Ch)	Mean (°Ch)	Sd. Error (°Ch)	Mean (°Ch)	Sd. Error (°Ch)	
Living room	Cooling	1358	231	1081	143	3861	380	<0.00
	Heating	1553	99	1713	91	3473	325	<0.00
	Total	2925	332	2807	235	7334	708	<0.00
Bedroom	Cooling	1206	225	910	144	3688	364	<0.00
	Heating	1066	61	1047	56	2242	209	<0.00
	Total	2283	288	1967	201	5929	577	<0.00
House	Cooling	1299	228	1014	143	3793	374	<0.00
	Heating	1361	83	1451	77	2988	279	<0.00
	Total	2660	305	2464	210	6781	627	<0.00

It is observed that cluster 1 and cluster 2 have higher similarity than cluster 3. That is because their walls and floor construction systems are similar, mainly differing in their roof construction system. Yet, a slight difference in their performance due to the roof composition can be observed. The cooling degree-hour means for cluster 2 are lower than the ones obtained for cluster 1, while the heating degree-hour means are higher for cluster 2. That proves that cluster 1, composed mainly of houses with concrete slabs and no tile nor ceiling, performs better in the cold season.

In contrast, cluster 2, composed of houses with only ceramic tile and wood ceilings, performed better over the hot season. These results are related to their thermal properties, i.e., thermal transmittance and thermal capacity, which are higher for cluster 1 roof construction systems. The differences in their performance would be even higher if the absorptances were the same. The absorptance of cluster 1 is higher (68%) than that of cluster 2 (56%), contributing to heat absorption. Cluster 3 showed the worst performance, both for heating and cooling, as expected. It is made of wood (walls and floor) and has only fibre cement tiles on the roof with no concrete slab, which is known for promoting poor thermal performance.

The results refer to a virtual building model with specific geometry and occupation patterns and might not be valid for another model. These findings aimed to obtain different building reference models based on the envelope materials.

Figure 15 shows the boxplot for all variables (in rows) and rooms (in columns) of each cluster and allows one to check the central tendency of a value in a sample. Each cluster is represented by a boxplot dot colour (blue, orange, and green for clusters 1, 2, and 3, respectively). The big yellow dot in each boxplot represents the cluster reference model, and its position in the distribution is related to its sample representativeness. As expected, due to what was seen in the hypothesis tests, it is observed that for all variables of all rooms, the reference model degree-hour is close to the sample median. That indicates that it takes on a central position in the sample, proving a good representativeness of the reference model. Additionally, it can be observed that clusters differ in the dispersion of

their data, which highlights the different behaviour of houses with different construction systems.

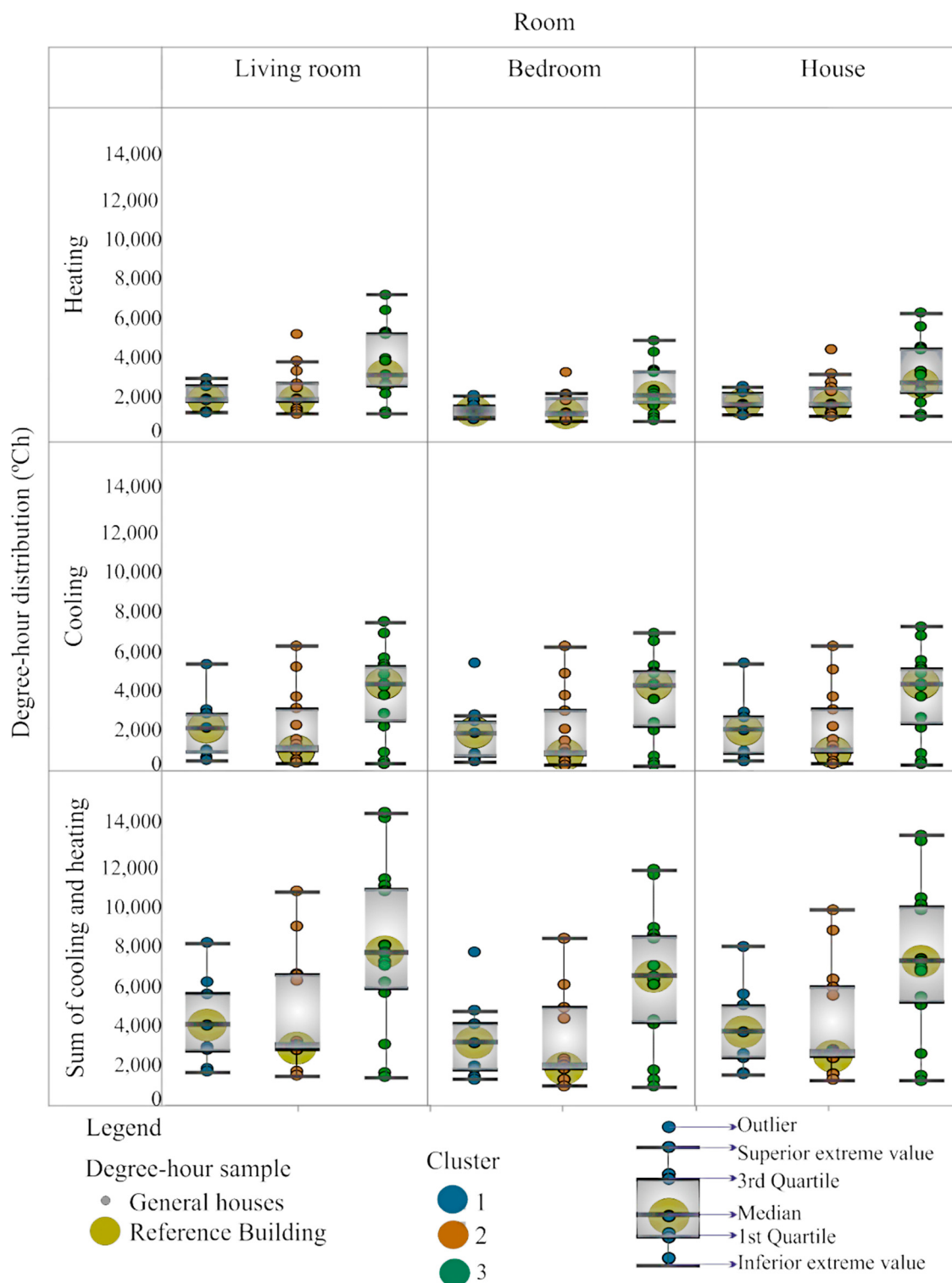


Figure 15. Boxplot referring to the degree-hour value of all dwellings of each cluster for each variable.

Also, one can observe that cluster 3 is the cluster with the most dispersed sample. It means that the differences in their houses' construction systems impact their thermal performance higher than in the other clusters, and it is even higher for the bedrooms in the hot season. Few outliers were found, two belonging to cluster 2 and one to cluster 1.

They differ from their sample mainly due to their tile type (both have fibre cement tile; in cluster 2, ceramic tile was the usual, and there was no tile in most cases in cluster 1).

Based on the results, it is considered that the reference models found in the cluster analysis are adequate to represent the sample. Thus, the envelope materials from these models can be used in future studies to evaluate the thermal performance of low-income housing in southern Brazil.

5. Conclusions

This paper aimed to find actual construction system reference models of low-income housing in Florianópolis, southern Brazil, through cluster analysis, to be used in future thermal and energy performance studies. For that, three steps were conducted in the proposed method.

First, a data acquisition process was developed, and information about the envelope of 106 low-income houses in Florianópolis was obtained. A great difficulty in obtaining data from this type of building was observed. Many of them were built without project or approval in the city hall. The imprecision and the high heterogeneity in the information found hinder the creation of a pattern, being necessary to simplify and even withdraw much of the collected data to proceed with this study. However, despite simplifications, it was possible to identify patterns and improve the knowledge about the characteristics of these buildings.

Three clusters were found, each represented by their reference models. The main difference is their roof construction systems, which differ in all three clusters. The definition of the reference models based on their characteristics seems to fit well with the reality found, highlighting the existing typologies. Beyond that, the thermal properties of the reference models' construction systems from the clusters obtained are quite different, as confirmed by the hypothesis tests, which can provide a good panorama of the thermal performance of that type of building in future studies. From the computer simulation step, the internal operative temperatures of all long-term rooms were obtained for all houses of the sample. The degree-hour temperatures, a thermal performance indicator, were calculated from these output data, and hypothesis tests were performed to check if the clusters' means differed significantly between clusters. The statistical independence between the clusters for all variables was confirmed. In addition, it was verified in the boxplot diagram that the degree-hours of the reference models were within the range of the first to the third quartile.

From these results, it is concluded that the clusters are well-represented by the reference models found. In general, cluster 1 performed better over winter, while cluster 2 performed better over summer than the other clusters (see degree-hour means in Table 5). Cluster 3 showed the worst performance, both for cooling and heating degree-hour. This was already expected since it was composed of wood and fibre cement tile houses, materials known for their poor thermal performance in buildings. It is highlighted that these findings are based only on the envelope materials since other parameters such as geometry, operation, and internal gains were fixed in all virtual models in the computer simulations.

This study contributes to the literature by proposing and validating a method to determine building reference models. In addition, representative reference models of low-income housing in a city in southern Brazil were developed, helping to form a database that can be used in future thermal and energy performance studies. Knowing the possible building systems of a given location helps in the development of new housing designs. The design project should reduce the adverse effects of climate, and its performance should be analysed by means of comparison to usual building systems. In addition, it assists in establishing public policies to better meet performance criteria. Therefore, reference models are helpful in analyses for improving the thermal performance of a building stock and not only of a specific building. This is particularly important for studies developed as a basis for modifying or developing standards.

At last, it can be concluded that cluster analysis was a practical and objective method for obtaining reference models. On the one hand, because it is an exploratory rather

than inferential data analysis, the researcher must choose wisely the variables to compose the database and be aware of its relation with the objectives of the study. On the other hand, even though cluster analysis is an exploratory analysis, it is essential to apply other statistical techniques to ensure adequate results, such as the hypothesis tests applied.

Author Contributions: Conceptualization, A.S. and E.G.; methodology, A.S.; validation, A.S. and T.P.S.; formal analysis, A.S.; investigation, A.S.; data curation, A.S.; writing—original draft preparation, A.S. and T.P.S.; writing—review and editing, T.P.S. and E.G.; visualization, A.S. and T.P.S.; supervision, E.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fundação João Pinheiro. *Déficit Habitacional No Brasil Por Cor Ou Raça (2016–2019)*; Fundação João Pinheiro: Brasília, Brazil, 2022. (In Portuguese)
2. International Energy Agency. *World Energy Outlook 2022*; International Energy Agency: Paris, France, 2022.
3. Empresa de Pesquisa Energética. *Balanço Energético Nacional: Ano Base 2021*; Empresa de Pesquisa Energética: Brasília, Brazil, 2022. (In Portuguese)
4. Mousavi, S.N.; Gheibi, M.; Waclawek, S.; Smith, N.R.; Hajiaghahi-Keshteli, M.; Behzadian, K. Low-energy residential building optimisation for energy and comfort enhancement in semi-arid climate conditions. *Energy Convers. Manag.* **2023**, *291*, 117264. [\[CrossRef\]](#)
5. Alaidroos, A.; Krarti, M. Optimal design of residential building envelope systems in the Kingdom of Saudi Arabia. *Energy Build.* **2015**, *86*, 104–117. [\[CrossRef\]](#)
6. Yong, S.G.; Kim, J.H.; Gim, Y.; Kim, J.; Cho, J.; Hong, H.; Baik, Y.J.; Koo, J. Impacts of building envelope design factors upon energy loads and their optimization in US standard climate zones using experimental design. *Energy Build.* **2017**, *141*, 1–15. [\[CrossRef\]](#)
7. Yousefi, F.; Gholipour, Y.; Yan, W. A study of the impact of occupant behaviors on energy performance of building envelopes using occupants' data. *Energy Build.* **2017**, *148*, 182–198. [\[CrossRef\]](#)
8. El-Darwish, I.; Gomaa, M. Retrofitting strategy for building envelopes to achieve energy efficiency. *Alex. Eng. J.* **2017**, *56*, 579–589. [\[CrossRef\]](#)
9. Charisi, S. The Role of the Building Envelope in Achieving Nearly-zero Energy Buildings (nZEBs). *Procedia Environ. Sci.* **2017**, *38*, 115–120. [\[CrossRef\]](#)
10. Loukaidou, K.; Michopoulos, A.; Zachariadis, T. Nearly-zero Energy Buildings: Cost-optimal Analysis of Building Envelope Characteristics. *Procedia Environ. Sci.* **2017**, *38*, 20–27. [\[CrossRef\]](#)
11. Famuyibo, A.A.; Duffy, A.; Strachan, P. Developing archetypes for domestic dwellings—An Irish case study. *Energy Build.* **2012**, *50*, 150–157. [\[CrossRef\]](#)
12. Corgnati, S.P.; Fabrizio, E.; Filippi, M.; Monetti, V. Reference buildings for cost optimal analysis: Method of definition and application. *Appl. Energy* **2013**, *102*, 983–993. [\[CrossRef\]](#)
13. Torcellini, P.; Deru, M.; Griffith, B.; Benne, K. DOE commercial building benchmark models. In Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, USA, 17–22 August 2008; ACEEE: Washington, DC, USA, 2008.
14. Intelligent Energy Europe Programme. Typology Approach for Building Stock Energy Assessment. Energy Performance Indicator Tracking Schemes for the Continuous Optimisation of Refurbishment Processes in European Housing Stocks. Available online: www.building-typology.eu (accessed on 5 August 2023).
15. Vettorazzi, E.; Figueiredo, A.; Rebelo, F.; Vicente, R.; Feiertag, G.A. Beyond passive House: Use of evolutionary algorithms in architectural design. *J. Build. Eng.* **2023**, *76*, 107058. [\[CrossRef\]](#)
16. Balvedi, N.; Giglio, T. Influence of green roof systems on the energy performance of buildings and their surroundings. *J. Build. Eng.* **2023**, *70*, 106430. [\[CrossRef\]](#)
17. Benejam, G.M. benjam. In Proceedings of the Retrofit 2012 Conference, Manchester, UK, 24–26 January 2012.
18. Filogamo, L.; Peri, G.; Rizzo, G.; Giaccione, A. On the classification of large residential buildings stocks by sample typologies for energy planning purposes. *Appl. Energy* **2014**, *135*, 825–835. [\[CrossRef\]](#)
19. Schaefer, A.; Ghisi, E. Method for obtaining reference buildings. *Energy Build.* **2016**, *128*, 660–672. [\[CrossRef\]](#)
20. Dascalaki, E.G.; Kontoyiannidis, S.; Balaras, C.A.; Droutsas, K.G. Energy certification of Hellenic buildings: First findings. *Energy Build.* **2013**, *65*, 429–437. [\[CrossRef\]](#)
21. Fracastoro, G.V.; Serraino, M. A methodology for assessing the energy performance of large scale building stocks and possible applications. *Energy Build.* **2011**, *43*, 844–852. [\[CrossRef\]](#)
22. Dascalaki, E.G.; Droutsas, K.; Gaglia, A.G.; Kontoyiannidis, S.; Balaras, C.A. Data collection and analysis of the building stock and its energy performance—An example for Hellenic buildings. *Energy Build.* **2010**, *42*, 1231–1237. [\[CrossRef\]](#)

23. Petcharat, S.; Chungpaibulpatana, S.; Rakkwamsuk, P. Assessment of potential energy saving using cluster analysis: A case study of lighting systems in buildings. *Energy Build.* **2012**, *52*, 145–152. [\[CrossRef\]](#)
24. Krelling, A.F.; Eli, L.G.; Olinger, M.S.; Machado, R.M.E.S.; Melo, A.P.; Lamberts, R. A thermal performance standard for residential buildings in warm climates: Lessons learned in Brazil. *Energy Build.* **2023**, *281*, 112770. [\[CrossRef\]](#)
25. Fumo, N.; Mago, P.; Luck, R. Methodology to estimate building energy consumption using EnergyPlus Benchmark Models. *Energy Build.* **2010**, *42*, 2331–2337. [\[CrossRef\]](#)
26. Theodoridou, I.; Papadopoulos, A.M.; Hegger, M. A typological classification of the Greek residential building stock. *Energy Build.* **2011**, *43*, 2779–2787. [\[CrossRef\]](#)
27. Attia, S.; Evrard, A.; Gratia, E. Development of benchmark models for the Egyptian residential buildings sector. *Appl. Energy* **2012**, *94*, 270–284. [\[CrossRef\]](#)
28. Geraldi, M.S.; Gnecco, V.M.; Barzan Neto, A.; de Mafra Martins, B.A.; Ghisi, E.; Fossati, M.; Melo, A.P.; Lamberts, R. Evaluating the impact of the shape of school reference buildings on bottom-up energy benchmarking. *J. Build. Eng.* **2021**, *43*, 103142. [\[CrossRef\]](#)
29. Palladino, D. Energy performance gap of the Italian residential building stock: Parametric energy simulations for theoretical deviation assessment from standard conditions. *Appl. Energy* **2023**, *345*, 121365. [\[CrossRef\]](#)
30. Hair, J.F.; Black, W.C.; Babin, B.J.; Anderson, R.E.; Tatham, R.L. *Análise Multivariada De Dados [Multivariate Data Analysis]*, 6th ed.; Bookman: Porto Alegre, Brazil, 2009.
31. Mingoti, S.A. *Análise De Dados Através De Métodos De Estatística Multivariada: Uma Abordagem Aplicada [Data Analysis Using Multivariate Statistical Methods: Na Applied Approach]*; Ed. da UFMG: Belo Horizonte, Brazil, 2005.
32. Bussab, W.O.; Miazaki, E.S.; Andrade, D.F. Introdução à Análise de agrupamentos. In Proceedings of the IX Simpósio Nacional de Probabilidade e Estatística; Caderno de Resumos: São Paulo, Brazil, 1990.
33. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*; Library of Congress: Camden, NJ, USA, 1998.
34. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; John Wiley: Hoboken, NJ, USA, 2005.
35. European Parliament. *EU Directive 2010/31/EC, European Parliament and of the Council on the Energy Performance of Buildings*; European Parliament: Brussels, Belgium, 2010; Available online: [https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:153:0013:0035:en:PDF#:~:text=This%20Directive%20promotes%20the%20improvement,climate%20requirements%20and%20cost%20effectiveness.&text=\(g\)%20independent%20control%20systems%20for,performance%20certificates%20and%20inspection%20reports](https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:153:0013:0035:en:PDF#:~:text=This%20Directive%20promotes%20the%20improvement,climate%20requirements%20and%20cost%20effectiveness.&text=(g)%20independent%20control%20systems%20for,performance%20certificates%20and%20inspection%20reports) (accessed on 5 August 2023).
36. Wang, M.; Yu, H.; Yang, Y.; Jing, R.; Tang, Y.; Li, C. Assessing the impacts of urban morphology factors on the energy performance for building stocks based on a novel automatic generation framework. *Sustain. Cities Soc.* **2022**, *87*, 104267. [\[CrossRef\]](#)
37. Mitra, D.; Chu, Y.; Cetin, K. Cluster analysis of occupancy schedules in residential buildings in the United States. *Energy Build.* **2021**, *236*, 110791. [\[CrossRef\]](#)
38. Liu, X.; Hu, S.; Yan, D. A statistical quantitative analysis of the correlations between socio-demographic characteristics and household occupancy patterns in residential buildings in China. *Energy Build.* **2023**, *284*, 112842. [\[CrossRef\]](#)
39. Sarstedt, M.; Mooi, E. *Cluster Analysis*. In *A Concise Guide to Market Research*; Springer: Berlin/Heidelberg, Germany, 2018.
40. Qiu, W.; Joe, H. Random Cluster Generation (with Specified Degree of Separation), R Package Version 1.3.7. 2022. Available online: <https://cran.r-project.org/web/packages/clusterGeneration/clusterGeneration.pdf> (accessed on 5 August 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.