Sustainable Cities and Society

Volume 51, November 2019, 101695

Predictive models for airtightness in social housing in a Mediterranean region

Jesica Fernández-Agüera[1], Samuel Domínguez-Amarillo[1]*, Juan José Sendra[1], Rafael Suarez[1]

[1]Instituto Universitario de Arquitectura y Ciencias de la Construcción, Escuela Técnica Superior de Arquitectura, Universidad de Sevilla, Spain. jfernandezaguera@us.es; jsendra@us.es; rsuarez@us.es;

*sdomin@us.es, Escuela Tecnica Superior de Arquitectura, Avenida de Reina Mercedes, nº2, 41014 España.

Abstract

This article describes two models developed to predict airtightness in multifamily buildings in a Mediterranean region. They are designed to enable city planners, architects and engineers to estimate airtightness in homes built from 1980 to date (predictive model 1) or prior to 1979 (predictive model 2), when the first domestic energy conservation regulations entered into effect. They are based on a series of readily accessible parameters such as winter severity, envelope exposure, presence of a bathroom window and façade type. The estimated n_{50} data can be used with energy certification software, which presently envisages the same, non-experimentally quantified mean value for all types of housing. They can also be entered into energy and comfort simulation programs to predict energy consumption and expected indoor temperatures.

Keywords: Airtightness, residential buildings, blower door test, air infiltration, southern Europe, predictive model, clustering

1. Introduction

Spain's 25 million homes account for 17 % (=14 865 kTep) of the country's yearly energy consumption and 25 % (=6 025 kTep) of its electric power consumption. Further to European guidelines, national legislation presently in place in Spain and other Mediterranean countries establishes a legally binding target of 80 % lower carbon emissions in 2050 than in 1990 (European Commission, 2011). That will call for implementing effective building rehabilitation able to both reduce energy consumption by the existing housing stock and raise indoor comfort levels.

The uncontrolled exchange of air across the elements in building envelopes, known as infiltration or leakage, affects both indoor air quality and the temperature and relative humidity conditions prevailing in built environments (N. M. M. Ramos et al., 2018). It consequently has a direct impact on comfort, health and energy use in buildings and the associated CO_2 emissions(J. Fernández-Agüera, Sendra, J., Suárez, Domínguez-Amarillo, & Oteiza, 2015; Nabinger & Persily, 2011; Salehi, Torres, & Ramos, 2017b). Envelope air permeability determines such exchanges of indoor and outdoor air.

Assessments of how air leakage across building envelopes affects energy savings (Alalouch, Al-Saadi, AlWaer, & Al-Khaled, 2019; Papadopoulos, Whiffen, Tilford, & Willson, 2018) are applied in housing rehabilitation projects to enhance airtightness as a passive measure for reducing energy consumption (Suárez & Fernández-Agüera, 2015).

In mild Mediterranean climates, the effect of infiltration on residential energy demand has generally been found to range from 5 kw/hm^2 to 10 kW/hm^2 (Domínguez-Amarillo et al., 2019; Feijó-Muñoz, Pardal, et al., 2019). Ventilation in such climes has traditionally depended on air leakage and uncontrolled opening of windows. In Spain, for instance, compulsory mechanical ventilation was not instituted until the twenty-first century (Fomento, 2013; Ministerio de Vivienda, 2006b).

Traditional ventilation practice in Mediterranean housing has a substantial impact on indoor air quality, which is unsuitable in many homes, particularly in winter when windows are normally opened for no more than 30 minutes per day (Domínguez-Amarillo, Fernández-Agüera, Sendra, & Roaf, 2018). As indoor pollutant concentration often exceeds the outdoor values in cities (Scibor, 2019), appropriate ventilation is essential for maintaining healthy conditions inside homes (Hesaraki, Myhren, & Holmberg, 2015; Salehi, Torres, & Ramos, 2017a).

An understanding of airtightness in the homes comprising the new-build and existing housing stock is consequently a key factor in planning energy consumption and environmental quality. The method most commonly accepted by the scientific community to assess airtightness is the blower door test (Hynek, 2011; M. Sherman, 1995). Airtightness in single family homes has been widely studied by researchers in northern Europe (Caillou & Van Orshoven, 2010; Gillott et al., 2016; Johnston, Wingfield, Miles-Shenton, & Bell, 2004; Maaleudstyr, 1984; Paap, Mikola, Teet-Andrus, & Kalamees, 2012; Vinha et al., 2015) and the United States (Chan, Joh, & Sherman, 2012, 2013; Walker, Sherman, Joh, & Chan, 2013) over the last three decades. In Mediterranean areas, however, where such testing is not mandatory, its cost and the population's general unawareness of its existence have limited its routine application in buildings. Nonetheless, scientific research in the area in recent years has contributed to enlarging the database in a number of countries (Alfano, Dell'Isola, Ficco, & Tassini, 2012; Alves, Fernández-Agüera, & Sendra, 2014; Pereira, Almeida, Ramos, & Sousa, 2014; N. Ramos et al., 2015; Sfakianaki et al., 2008), including Spain. All the airtightness tests conducted in Spain are run by university research teams (Feijó-Muñoz, González-Lezcano, Poza-Casado, Padilla-Marcos, & Meiss, 2019; Feijó-Muñoz, Pardal, et al., 2019; J. Fernández-Agüera et al., 2015; J. Fernández-Agüera, Sendra, & Domínguez, 2011; Jesica Fernández-Agüera, Domínguez-Amarillo, Sendra, & Suárez, 2016; Jesica Fernández-Agüera, Domínguez-Amarillo, Sendra, Suárez, & Oteiza, 2019; Jesús, Feijó-Muñoz; Irene, Poza-Casado; Roberto Alonso, González-Lezcano; Cristina, Pardal; Víctor, Echarri; Rafael, Assiego L.; Jesica, Fernández-Agüera; María Jesús, Dios-Viéitez; Víctor José, del C.-D.; Manuel, Montesdeoca C.; Miguel Ángel, Padilla-Mar, 2018; María I. Montoya, Pastor, & Planas, 2011). They are not undertaken by public or private construction companies, which have yet to be sensitised to the problems (impact on energy demand, occupant comfort and indoor air quality) posed when housing airtightness rates go uncontrolled. Such problems are the more severe in social housing, occupied by the most vulnerable segments of society that can afford neither to install HVAC systems nor to pay the high electric power bills associated with individual room heating/cooling. Despite its purportedly mild climate, Spain is the European country with the highest rate of cold weather-induced death.

Building envelope airtightness depends on many often interconnected factors. In addition to morphology, typology and construction, workmanship quality (a random component), also plays an important role. Given that infiltration is multi-parametric and at least partially stochastic, predicting and modelling this property of envelopes is particularly complex (Pan, 2010; Prignon & Van Moeseke, 2017; M. Sherman & Mcwilliams, 2007; M.H. Sherman & Chan, 2004).

The development of statistical models to predict building, in this case housing, airtightness contributes to progress in estimating energy demand and indoor air quality (Jones et al., 2015; M I Montoya, Pastor, Carrie, Guyot, & Planas, 2010; Pan, 2010; Persily, Musser, & Emmerich, 2010; Prignon & Van Moeseke, 2017). Although such models do not aspire to highly accurate predictions for individual cases, they may deliver reasonably good estimates of leakage distribution in housing stocks. The most prominent model in place in Spain was developed to estimate leakage in single-family homes in Catalonia. Based on similarities in climate and construction types between Catalonia and south-eastern France and following a procedure inspired by the LBNL (Chan et al., 2012), the model was developed by applying regression analysis to the Centre d'Études Techniques de l'Équipement de Lyon's airtightness database of single-family homes in France. It focused on determining initial leakage routes and exploring airtightness patterns by construction type, insulation, building age and occupancy (M I Montoya et al., 2010).

Models developed using artificial neural networks with human learning and adaptation capacities are also in place. Such systems are based on a few simple processing units and many connections that prompt adaptive changes in the units as new data are acquired (Cesar & da Fontoura Costa, 1997). Krstic et al. (Krstić, Koški, Otković, & Španić, 2014) recently tested a neural network to predict airtightness in a series of residential buildings in Croatia. One year later, the methodology was validated in a second study in which it was applied to a suite of residential buildings in the Republic of Serbia (Krstic, Otkovic, & Todorovic, 2015). A powerful estimation tool, it exhibits sound capacities although its full validation will call for considerable further effort.

Cluster analysis, introduced in architecture in 2007, has been used by researchers to fit predictive models to assess heating in schools or, more recently, identify energy consumption patterns (N Gaitani, Lehmann, Santamouris, Mihalakakou, & Patargias, 2010; Santamouris et al., 2007), draw heat load profiles in residential buildings (An, Yan, & Hong, 2018; do Carmo & Christensen, 2016), define decarbonisation strategies (Sousa, Jones, Mirzaei, & Robinson, 2018) or (in Spain) even aggregate the building stock based on archetypes (Mata, Sasic Kalagasidis, & Johnsson, 2014).

This study analyses the vulnerability to air leakage of social housing envelopes built prior to the entry in effect of CT79, Spain's first legislation on the subject (Gobierno, 1979). It introduces empirical data-based predictive models that can be used by public and private organisations to estimate airtightness. That issue has become a key consideration in cost optimisation analysis as envisaged in the (recast) EPBD (Ferrara, Monetti, & Fabrizio, 2018) for both residential new-builds (predictive model 1) and the existing housing stock (predictive model 2). The originality of the approach lies in the use of cluster analysis to significantly simplify predictive models based on climate zone and a small number of geometric, typological and construction parameters.

While designed for a broad sampling of multi-family residential buildings in southern Andalusia, these models can also serve as a basis for similar building typologies throughout the Mediterranean area.

The measurement protocols and procedures deployed here to locate air leakage are set out in an earlier paper describing a preliminary approach to predictive models for airtightness in recently built multi-family housing in gallery type buildings in southern Europe, based on a sample of 45 dwellings (Jesica Fernández-Agüera et al., 2016).

2. Methods

2.1. Sampling

The buildings on which blower door tests were conducted were selected by stratified random sampling. The attributes defined for stratification were the two deemed to have the greatest potential to induce differences (Alfano et al., 2012; Chan, Nazaroff, Price, Sohn, & Gadgil, 2005) in the construction solutions adopted for building envelopes: area-dependent climate and date of construction (before or after enactment of legislation on building envelope airtightness requirements) (Domínguez-Amarillo, Sendra, & Oteiza San José, 2016).

The region boasts a total of 568 455 (N) multi-family housing units ('dwellings' or 'homes'). Given the regional scale of this research, the size of sample n (a subset of population N) initially estimated as necessary to ensure a normal distribution was on the order of 150 homes. The sample ultimately comprised 159 low-income dwellings located in multi-family buildings identified as particularly representative of the construction characteristics observed at the housing stock and sub-group levels. Variance and consequently the errors committed in selecting the 159 homes in the sample were found by entering the results in Equation 1 below:

$$e = \sqrt{\frac{z^2\sigma^2}{n}} \qquad \text{(E 1)}$$

The buildings chosen were located in five climate zones, with winters ranging from very mild (zone A) to cold (zone C) and summers from warm (zone 3) to very warm (zone 4) (de la Flor, Domínguez, Félix, & Falcón, 2008), classified as per the climate categories set out in Spain's Technical Building Code (Ministerio de Vivienda, 2006b).

2.2. Blower door test

Dwelling envelope airtightness was measured with the standard blower door test and the specific methodology developed in (J. Fernández-Agüera et al., 2011). The 'Minneapolis Blower Door Model 4' kit used was connected to an automated performance testing system (flow range at 50 Pa, 25–7800 $m^3\ h^{-1}$; accuracy, ±3%). All openings in contact with the outdoors were closed and ventilation ducts were sealed. Measurements were taken at pressures ranging from 20 Pa to 70 Pa at 5 Pa intervals further to the procedure described in Spanish and European standard UNE EN 13829:2002 (ISO, 2015). Method B, which measures the performance of the building envelope overall, was deemed most suitable for categorising the sample. In other words, as the objective was to study air leakage due to the constructional parameters of the envelope, all intentional openings such as vents and shunts (which are unrelated to the materials or construction processes used) were sealed off to ensure they would not affect the measurements. The findings are summarised in Table 1.

ID.	**No. prop.**	**Year**	**Standard**	**Climate zone**	**n_{50} Med**	***std***
1	4	1954	preCT79	A3	5.13	*0.39*
2	3	1968	preCT79	A3	7.93	*0.34*
3	3	1971	preCT79	A3	5.73	*0.57*
4	4	1972	preCT79	A3	6.58	*1.01*
5	4	1974	preCT79	A3	6.89	*1.03*
6	1	1976	preCT79	A3	3.89	*0.00*
7	1	1978	preCT79	A3	13.14	*0.00*
8	3	1966	preCT79	A4	7.68	*2.82*
9	2	1969	preCT79	A4	7.16	*0.16*
10	2	1970	preCT79	A4	3.01	*0.13*
11	1	1961	preCT79	A4	11.62	*0.00*
12	1	1951	preCT79	B4	10.12	*0.00*
13	3	1963	preCT79	B4	6.24	*0.54*
14	4	1964	preCT79	B4	7.32	*1.92*
15	3	1965	preCT79	B4	9.48	*1.58*
16	2	1970	preCT79	B4	12.30	*1.09*
17	1	1973	preCT79	B4	11.80	*0.00*
18	1	1978	preCT79	B4	14.68	*0.00*
19	2	1959	preCT79	C3	5.11	*0.31*
20	4	1964	preCT79	C4	6.46	*2.59*
21	4	1967	preCT79	C4	6.80	*1.65*
22	4	2010	CT79	A3	4.36	*0.53*
23	8	2011	CT79	A3	8.41	*1.13*
24	8	2012	CT79	A3	6.46	*0.38*
25	8	2007	CT79	A4	3.93	*0.34*
26	1	1993	CT79	B4	15.57	*0.00*
27	7	1998	CT79	B4	8.45	*0.86*
28	10	2004	CT79	B4	9.06	*1.38*
29	8	2010	CT79	B4	5.30	*0.51*
30	7	2011	CT79	B4	4.17	*0.87*
31	5	2011	CT79	B4	8.37	*0.09*
32	8	2010	CT79	B4	6.90	*0.69*
33	8	2011	CT79	C3	4.70	*0.52*
34	5	2011	CT79	C4	7.38	*0.48*
35	8	2010	CT06	B4	4.95	*0.28*
36	7	2011	CT06	B4	9.95	*1.64*
37	4	2011	CT06	C3	2.74	*0.48*
Tot	159				6.52	*2.59*

Table 1. Characterization of n_{50}.

2.3. Cluster analysis

The primary aim of cluster analysis in this context was to find groups of individual dwellings exhibiting similar behaviours, i.e., common patterns or trends, not necessarily numerical proximity. Cluster analysis is a procedure designed to identify populational groups or subsets that share attributes.

It consists essentially in grouping *n* distinguishable objects or items into subsets such that the objects in any given subset or cluster are similar to one another and different from the items in all the other clusters. The most common attribute, Nn = {1, 2, …, n}, in each cluster is used to label each distinguishable object in that cluster. When a clustering algorithm is applied, the dataset is partitioned into a disorderly collection of non-empty subsets. The main problem is to identify items in terms of their similarity and differentiate among clusters (Hand, McLachlan, & Basford, 1989; Steinley & Brusco, 2011).

In complex datasets where the results depend on many factors (complex multi-dimensional systems) that may in turn be co-dependent, this approach is useful for categorising data items and parameters in the overall set to better process the information by identifying underlying patterns. Clustering consequently helps to roughly outline the data structure, which in turn serves as a support for analysis and to establish working hypotheses. The procedure makes it possible to partition the data (force them into a structure) into groups expected to exhibit similar behaviours in certain respects (Hennig et al., 2015).

In the K-means procedure chosen for the analysis, data are distributed across a set of K groups that contain the centroids representing the mean of the members of each subset. The centroid is the point that minimises the sum of the distances of all the members of the group to that point (Kanungo et al., 2002). The Howard-Harris method, based on Lloyd's algorithm (Lloyd, 1982a), was the optimisation procedure applied.

The initial assumption was that the factors determining airtightness should be related essentially to the morphological and constructional characteristics of housing envelopes. Whilst that relationship has been shown earlier to be neither linear no univocal, the performance of groups of dwellings may be expected to conform to a series of patterns. In other words, the aim was to identify clusters comprising homes very similar to one another and distinctly different from the rest of the sample.

Principal components analysis (PCA) was conducted prior to establishing the clusters to determine which descriptive variables were of greatest significance in sample morphology. Only the variables characterising the clustered items in terms relevant to the intended analysis were selected. PCA, an analytical procedure for exploring datasets to build predictive models, entails breaking the covariance matrix down into eigenvalues after normalising each element or variable. It is primarily an approach designed to extract factors. PCA was run on all the morphological variables of the dwellings in the sample to define a small number of linear combinations of the 21 variables identified that would explain the largest possible proportion of variation in the data.

As the scale for each parameter differed from that of all the others, to ensure suitable data processing with the K-means method all the values were standardised using the min-max system: [x - min(x)]/[max(x) - min(x)].

In contrast to hierarchical clustering, in the partitioning clustering used here clusters are not merged. Rather, items are assigned to clusters in keeping with an objective criterion.

The K-means method of partitioning minimises the within-cluster sum of squares (sum of squares of the distance of each item to the centroid):

$$minimise\ D = \sum_{i=1}^{nc} \sum_{\in \sum f} (x - x_i)^2 \qquad (E\ 2)$$

The standard minimisation procedures developed by Forgy (E. Forgy, 1965) and Lloyd (Lloyd, 1982b) are the ones most extensively used of all the methods reviewed by Xu and Wunsch (Taber, 2009), given their simplicity and effectiveness.

In non-hierarchical partitioning clustering, a given item can be assigned to only one cluster. The Calinski-Harabasz index, which gives a measure of the distance between clusters, was used to determine the number of clusters of greatest significance. It entails calculating the sum of squares of the between-cluster variance (SS_B) and the sum of squares of the within-cluster variance (SS_W), while also attempting to minimise the error associated with over-partitioning. The aim is to define the optimal number of clusters (Calinski & Harabasz, 1974).

The qualitative information furnished by clustering was also assessed. The model ultimately adopted was the one able to furnish the most powerful information on airtightness structure and performance trends.

2.4. Predictive models

Multiple linear regression was used to determine whether a mathematical model could be fitted to the relationship between airtightness and the parameters listed in Table 1. That method establishes the relationship between a dependent variable Y (airtightness) and a set of independent variables (X_1, X_2, ... X_K) such as year, typology, climate zone or floor area. A closer fit to actual situations can be obtained with multiple than single linear regression, for construction-related factors are complex and must consequently be explained, as far as possible, by the many variables directly or indirectly involved.

The mathematical notation for multiple linear regression is:

$$Y = a + b_1X_1 + b_2X_2 + ... + b_nX_n \qquad (E\ 3)$$

where:

Y: variable to be predicted

X_1, X_2... X_n: independent variables a, b_1, b_2, …., b_n: unknown constants to be estimated.

Some of the independent variables introduced were quantitative and others qualitative. They were included in the regression model by constructing what are known as dummy variables that ordinarily but not necessarily adopt the arbitrary values 1 and 0, although other values and more than two variables may be used.

Dummy variables were entered in the model to determine the result that afforded the best fit in two ways: with additive dummy variables and with variable categorisation.

The additive dummy variable procedure consisted in entering 'X' new dummy variables in the model, where 'X' is the 'number of existing categories, less 1'. For instance, if the variable to be entered had five categories, four dummy variables were entered and attributed a value of 1 if they pertained to the category assigned and 0 otherwise. The dummy variable coefficient measures the difference in the effect of the two y-intercepts, i.e., the difference in the expected values of the dependent variable depending on whether it features or fails to feature a given characteristic of the qualitative factor.

Categorising consisted in ranking the categories from least to greatest effect on the airtightness findings. Deploying categorical regression, SPSS software assigned the categories a coefficient in keeping with their respective impact.

The procedure followed to generate the multiple regression model was as follows:

i. cluster identification
ii. choice of parameters affecting airtightness
iii. testing for multiple collinearities in the parameters studied
iv. dummy variable classification and entry in keeping with the parameters classified
v. implementation of the stepwise method to identify the model with the smallest number of variables that best explained the dependent variable or criteriondetermination of the goodness of fit of the data to the multiple regression model
vii. estimation of equation or predictive model parameters.The predictive models were subject to a series of limitations imposed by the sample: i.e., to be included, homes had to constitute social, multi-family housing, lie in one of the climate zones defined in southern Spain and have a net floor area <105 m^2 and a window area <17 m^2.

2.5. Characteristic parameters

The independent variables with the greatest impact on building airtightness identified in a recent review of the literature informed the present selection (Prignon & Van Moeseke, 2017). The parameters deemed suitable for the study and listed and classified in Table 2 were defined on the grounds of the type of buildings sampled: reinforced concrete slab and column, multi-family housing.

Geometry	Technology and materials (Liddament, 1986; Sfakianaki et al., 2008)	Other
-A Area(Chan et al., 2012) -V Volume -A_F Façade area -A_R Roof area -A_{CA} area adjacent to communal areas -A_D area adjacent to other dwellings (Jesica Fernández-Agüera et al., 2016) -A_W Window area (Sfakianaki et al., 2008) -P_W Window perimeter(Almeida, Ramos, & Pereira, 2017),(Sfakianaki et al., 2008)	-T_F Façade type(Kalamees, 2007) -T_W Window type(Krstic et al., 2015) -T_B Blind type -B No blinds -M General condition -HVAC (yes or no)	-WS Winter severity (Chan et al., 2013) -SS Summer severity (Chan et al., 2013) -Y Year (Chan et al., 2013; Eskola et al., 2015; M I Montoya et al., 2010; Sinnott & Dyer, 2012) -E Exposure type - Window in bathroom -Separate kitchen -Regulation

Table 2. Classification of parameters studied

The categorical variables established in this sample were as follows:

- Facade type (F1: 1 or 1 and ½ foot brick fabric; F2: ½ foot brick fabric (or one-brick thick) outer wall + air cavity + hollow brick inner wall; F3: ½ foot brick fabric (or one-brick thick) outer wall + insulation layer + hollow brick inner wall; F4: ½ foot brick fabric (or one-brick thick) outer wall + air cavity + plasterboard inner wall; F5: fired clay panelling + air cavity + insulation + fired clay block).
- Window type (W1: hinge opening windows; W2: sliding windows; W3: hinged and sliding windows).
- Blind type (B1: no blinds; B2: external blinds; B3: roller shutter in splayed openings; B4: roller shutter in compact blinds).
- Exposure (E1: semi-detached, linearly aligned buildings with four homes per storey; E2: semi-detached, linearly aligned buildings with two homes per storey; E3: open gallery buildings; E4: stand-alone high rises; E5: semi-detached, linearly aligned buildings with two homes per storey and building, located at the corner of the compound or in stand-alone buildings with H, T- or X-shaped ground plans).

3. Results

3.1. Cluster analysis

The six principal components found with PCA to have eigenvalues greater than 1 were chosen for the analysis, for they explained 95.66 % of the variation in the raw data (Figure 1).

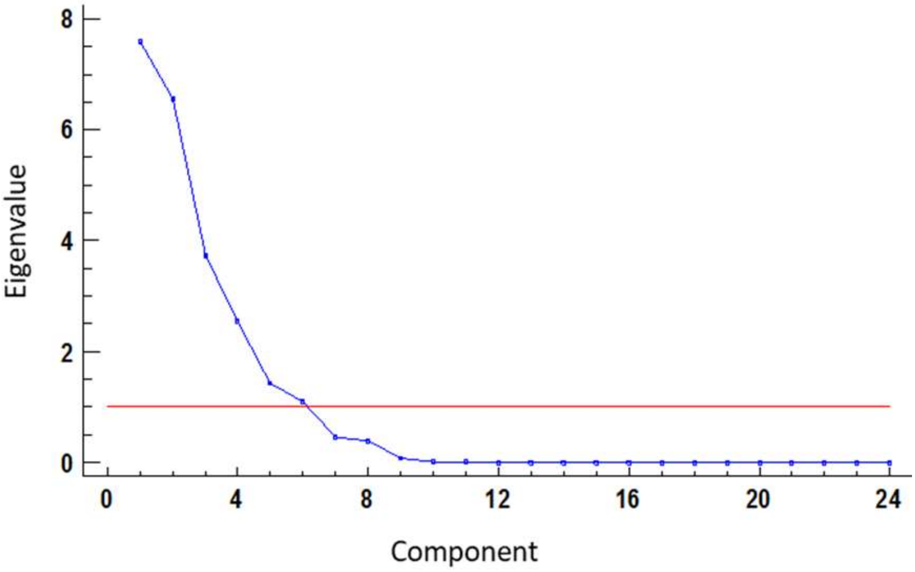

Figure 1. Components chosen

Based on that criterion, the variables chosen were those with the capacity to significantly affect clustering. The variables selected were quantitative (associated with dimensional ratios: P_W/V, P_W/A_F, A_D/V, A_F/V, A_{CA}/V, A_W/A_F, A_W/V), supplemented with a series of categorical parameters (façade (T_F), window (T_W) and blind (T_B) types). The former was normalised to indoor volume of the home or façade area for better inter-comparison.

On the grounds of the foregoing, two clusters were defined, one with 98 dwellings and the other with 53, along with eight outliers. The characteristic cluster values, i.e., the centroids for each variable, are given in Table 3.

		Centroid							Most common value		
Cluster	**No.**	**P_w/V** (m^{-1})	**P_w/A_F** (m^{-1})	**A_{CA}/V** (m^{-1})	**A_F/V** (m^{-1})	**A_{CA}/V** (m^{-1})	**A_w/A_F**	**A_w/V** (m^{-1})	**T_F**	**T_B**	**T_W**
1	*98*	0.179	0.718	0.263	0.254	0.059	0.231	0.056	3	3	1
2	*53*	0.213	0.628	0.190	0.345	0.101	0.196	0.067	2	2	2
Outliers	*8*										

Table 3. Morphological and construction characteristics defining clusters

A series of statistical tests was conducted on the clustering as a whole and on the variables selected to determine their relative weights with a view to determining the significance and representativeness of the items clustered (Table 4).

Number of clusters:	2
Number of datapoints:	149
Between-cluster sum of squares:	17.75
Total sum of squares:	78.97

The partitioning of greatest significance was delivered by the aforementioned seven variables and two clusters (with eight unassigned outliers). Of all the approaches analysed, this was the one that maximised between-cluster variance (SS_B), ensuring suitable separation among them. This metric quantifies between-cluster separation as the sum of squares of the distance between the centre of each cluster (measured as the mean value of its data points) and the centre of the entire dataset.

The resulting clustering was deemed to afford a sufficiently robust description, for the total sum of squares explained nearly 80 % of the variance.

Analysis of variance

			Model		Error	
Variable	**F-statistic**	**P-value**	**Sum of squares**	**DF**	**Sum of squares**	**DF**
A_F/V	48.92	8.714 e^{-11}	2.289	1	6.87	147
A_D/V	24.84	1.73 e^{-06}	1.166	1	6.89	147
A_{CA}/V	17.97	3.942 e^{-05}	0.553	1	4.53	147
P_W/V	16.53	7.786 e^{-05}	0.536	1	4.76	147
A_W/V	12.76	0.000479	0.330	1	3.80	147
P_W/A_F	12.48	0.000551	0.669	1	7.88	147
A_W/A_F	10.04	0.001862	0.526	1	7.71	147

(*) Categorical variables not included in the analysis of variance table

Analysis of variance (ANOVA) is a collection of statistical models and associated procedures that identify the variance both within and between the observations associated with each cluster. Here ANOVA was calculated for each variable to determine which most effectively defined the clusters.

The F-statistic value indicates the proportion of variance explained by the variable (ratio of the between-cluster variance to the total variance). The higher the value of the F-statistic, the greater the inter-cluster difference in the variable. The F-statistic values are listed in descending order in the table. The associated P-value is an indication of the statistical significance of the F distribution. The lower its value, the greater is the expected between-cluster difference of the values of the respective variable.

The mean squares model is the ratio of the between-cluster sum of squares (SS_B)to the degrees of freedom. The between-cluster sum of squares is a measure of the variance between the mean values of the clustered items. The closer the mean values, the smaller the SS_B (for a model with k-1 degrees of freedom, where k is the number of clusters).

Table 4: Analysis of variance between clusters

As the table shows, the factor with the heaviest impact on variation and consequently on clustering was the ratio between façade area and volume. The area adjacent to other dwellings was another differentiating factor, although it carried perceptibly less weight than the A_F/V variable. The remaining parameters, all significant for clustering purposes, exhibited lower values.

Window-related variables (area and perimeter), particularly the ratio to the façade area, varied the least across the sample. In other words, these parameters were similar in all the homes in the overall sample. Although the parameters defining these openings were important in establishing the principal components, due to the relative stability of the aforementioned ratios, they had a smaller effect on partitioning than the other parameters selected.

Cluster composition is shown in Figure 2, which identifies the number of components in each cluster by development, listed by date of construction (from oldest to most recent). Clustering was observed to be closely correlated to age, for cluster 1 comprised the more recent and cluster 2 the older developments.

A transition period was detected in the late nineteen seventies and early eighties, with homes alternating between the two clusters. The dwellings in the developments sampled were distributed between the two clusters on the grounds of their individual morphologies.

Development	Cluster 1	Cluster 2	Unassigned
1		4	
2			3
3		3	
4		4	
5		4	
6		1	
7		1	
8		3	
9		2	
10	2		
11		1	
12		1	
13		3	
14		4	
15		3	
16		2	
17		1	
18			1
19		2	
20	4		
21		4	
22	3	1	
23	6	2	
24	8		
25	8		
26		1	
27	3	4	
28	8	2	
29	8		
30	7		
31	5		
32	8		
33	8		
34	5		
35	8		
36	7		
37			4

Figure 2. Distribution of developments (in chronological order) by morphological cluster

The homes not assigned to any cluster were in developments 2, 18 and 37. The situations involved varied. On the one hand, developments 2 and 18, with remodelled homes (enclosed balconies, kitchens enlarged to include laundry rooms and similar), exhibited morphologies very different from the rest of the sample, especially as regards openings. On the other, the dwellings in development 37 were the least permeable of the entire sample, exhibiting a highly compact building type which, together with their windows with no blinds, distinguished them from all the others.

Clustering suggested a close correlation between airtightness and the suite of morphological and construction factors, although these results were subject to a certain degree of stochasticity stemming essentially from the variation inherent in construction processes. That consideration was particularly significant in components such as façade walls and windows in which the substantial manual labour required could well have occasioned considerable differences between apparently equivalent homes within a given development. Nonetheless, with this division into clusters or families of dwellings, similar airtightness values could be determined from construction, typological and climate parameters.

That underlying structure was used to build the multiple regression-based prediction models as discussed in the sections below. A specific performance model was generated for each cluster to predict permeability in keeping with parameters characteristic of each dwelling.

Clustering therefore established two subsamples, taking the distinguishing factor (to simplify the process) to be the period when the homes were built. Significance was greatest when the distinction was drawn between those built in the first, pre-regulation period (1950 to 1979) and those erected in the second (post-1979), which covered developments governed both by Code CT-79 (Gobierno, 1979) and by the Technical Building Code (Ministerio de Vivienda, 2006a) presently in effect.

Verification consisted initially in contrasting the two sub-samples to establish their independence and the consistency of their distributions. The parameter chosen was air infiltration rate at 50 Pa (n_{50}), the criterion routinely used to characterise envelope airtightness in dwellings. Clustering suitability would be associated with the ability to furnish information on the specific airtightness of each group, i.e., the capacity to generate prediction models for that parameter. The key statistical descriptors for the two clusters are given in Table 5, which is followed by a description of the tests run to compare the two distributions.

Cluster descriptor		
	Cluster 1	***Cluster 2***
No. elements	98	53
Mean	6.96 h^{-1}	7.51 h^{-1}
Median	6.41 h^{-1}	7.18 h^{-1}
Standard deviation	2.31 h^{-1}	2.74 h^{-1}
Coefficient of variation	36.84%	31.79%
Minimum	3.23 h^{-1}	3.88 h^{-1}
Maximum	14.14 h^{-1}	13.39 h^{-1}
Range	10.91 h^{-1}	9.50 h^{-1}

Table 5. Statistical descriptors for n_{50}, clusters 1 and 2

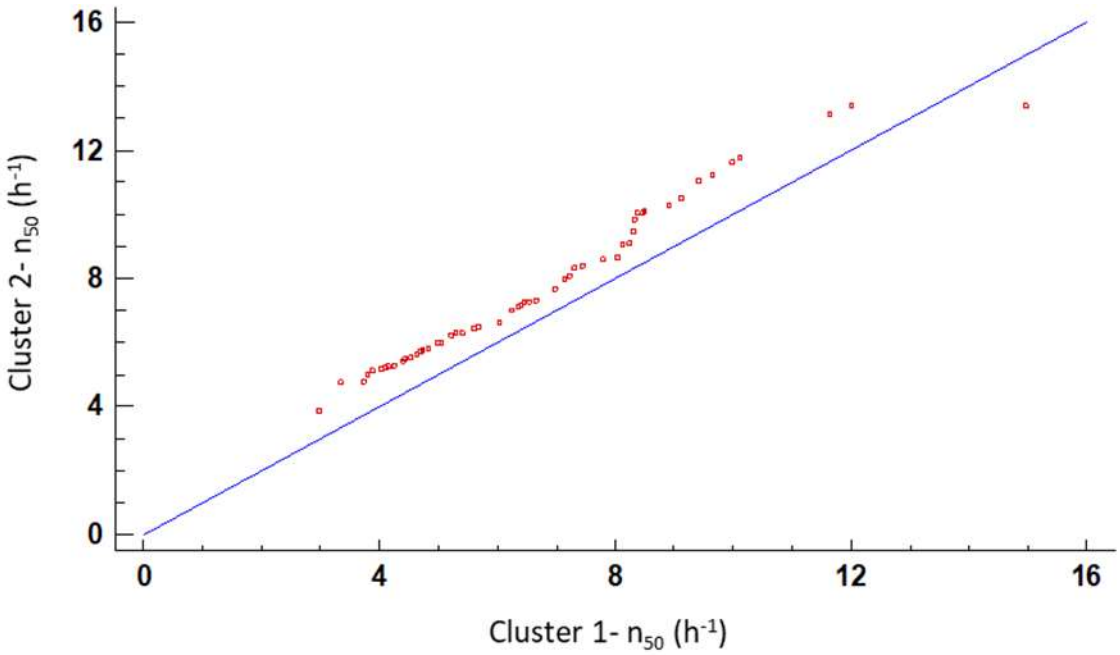

Figure 3. Q-Q (quantile-quantile) graph for cluster 1 vs cluster 2 distributions

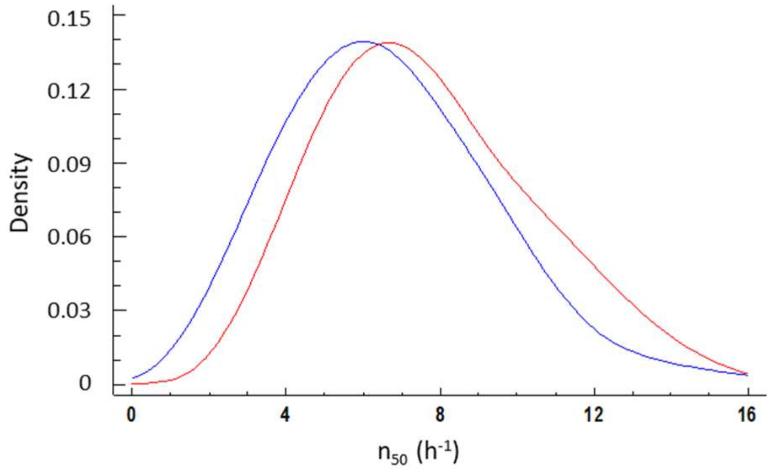

Figure 4. Smooth density curves for cluster 1 (blue) and cluster 2 (red) probability distributions

Graphic analysis showed that the distributions associated with the two clusters differed significantly. A very prominent general shift was observed on the upper end of the linear regression line (where the values should cluster if the distributions concur) in the Q-Q graph (Figure 3). The differences in the probability density curves and the misalignment of their central values (Figure 4) further supported the independence of the two groups.

The cluster 1 values had a mean n_{50} of 6.96 h^{-1} although variation was generally wide, with a high (σ: 2.31 h^{-1}) standard deviation relative to the mean. Sample values ranged over a wide interval, from a minimum of 3.23 h^{-1} to a maximum of 14.14 h^{-1} (Table 3). Cluster 2, with a mean n_{50} of 7.51 h^{-1}, likewise varied widely, with a standard deviation (σ: 2.74 h^{-1}) even higher than observed for cluster 1. The minimum (3.88 h^{-1}) and maximum (13.39 h^{-1}) values also covered a broad spectrum.

Further to the similarity tests, the two clusters exhibited distinctly different behaviours and distributions, confirming that they represented different populations.

3.2. Predictive models

A mathematical model able to fit airtightness-versus-selected-parameter curves was developed using discriminate analysis techniques and SPSS software. That exercise delivered two predictive airtightness models based on the specific characteristics of each home and building: one adapted to the characteristics of dwellings built prior to 1979 when the first general legislation to limit building energy demand was enacted in Spain, and a second to the homes built after that date.

3.2.1. Cluster 1

The best fit was obtained with backward stepwise multiple regression including a constant. For cluster 1, which covered homes built from 1979 to date, the characteristics addressed were location, morphology, construction and geometry. In the model, parameter Y_{AW} represented the coefficient for the quantitative variable window area and Y_{PW} window perimeter. The model also accommodated constants for the categorical variables: separate kitchen (β_K), blinds (β_B), bathroom window (β_{WS}), winter severity (β_{ws}), window type (β_{TW}), exposure (β_{TE}) and façade type (β_{TF}), listed with their respective category in Table 6.

Coefficient	Category	Value	Std error
Constant (α)		6.613	1.624
Window area (Υ_{AW})		1.265	0.218
Window perimeter (Υ_{PW})		-0.387	0.070
Separate kitchen (β_K)	Yes	-1.698	0.922
	No	0.000	0.000
Blind (β_B)	Yes	1.957	1.051
	No	0.000	0.000
Bathroom window (β_{WB})	Yes	1.498	0.315
	No	0.000	0.000
Winter severity (β_{WS})	WS A	*	*
	WS B	4.143	0.368
	WS C	0.618	0.618
Window type (β_{TW})	W1	-5.154	0.600
	W2	-2.445	0.493
	W3	*	*
Exposure (β_{TE})	E1	-0.075	0.473
	E2	-0.620	0.665
	E3	-1.558	0.588
	E4	-2.570	0.664
	E5	*	*
Façade type (β_{TF})	F3	1.179	0.619
	F4	-0.838	0.668
	F5	*	*

Table 6. Predictive model coefficients for cluster 1(*variable category included in constant 'α')

The predictive model was defined by the probability function shown in Equation 4 and the coefficients listed in Table 4.

$$n_{50} = \alpha + \gamma_{AW} \cdot AW + \gamma_{PW} \cdot PW + \beta_K + \beta_{PB} + \beta_{WB} + \beta_{WS} + \beta_{TW} + \beta_{TE} + \beta_{TF} \qquad \textit{(E 4)}$$

The nine independent variables included in the model explained 88.70 % of dependent variable variation. At 0.849 h^{-1} (Table 7), the standard error for the prediction was very narrow relative to cluster variation and just 35.00 % of the standard deviation for cluster 1 (Table 5). To rule out inter-variable dependence which, even if present, would not denote causality, a series of ANOVAs was run to determine whether infiltration in the homes studied could be predicted with a model based on their classificatory characteristics.

R	R^2	Adjusted R^2	Standard error for the estimate
0.943	0.887	0.865	0.849

Table 7. Predictive model for n_{50} in cluster 1: statistical descriptors

The ANOVA table delivered an F-statistic with which to test the null hypothesis that R^2 and the slope of the curve were equal to 0; in other words, that the two variables were not correlated. Since the p-value for the F-statistic was lower than the significance level, the null hypothesis was ruled out and the results obtained for the sample were deemed to be applicable to the population from which it was drawn (Table 8).

	Sum of squares	df	Root mean square	F	Sig.
Regression	443.177	12	36.931	54.455	.000
Residuals	54.934	81	0.678		
Total	498.111	93			

Table 8. ANOVA for the cluster 1 n_{50} predictive model

Model predictions proved to closely mimic actual performance. The plot of the estimated vs the empirical values for n_{50} (Figure 5) showed that all but seven of the values lay in the area between the line of symmetry and the distance defined by the standard error for the estimate (± 0.87). No bias was observed and the points on the graph were clustered around the regression line.

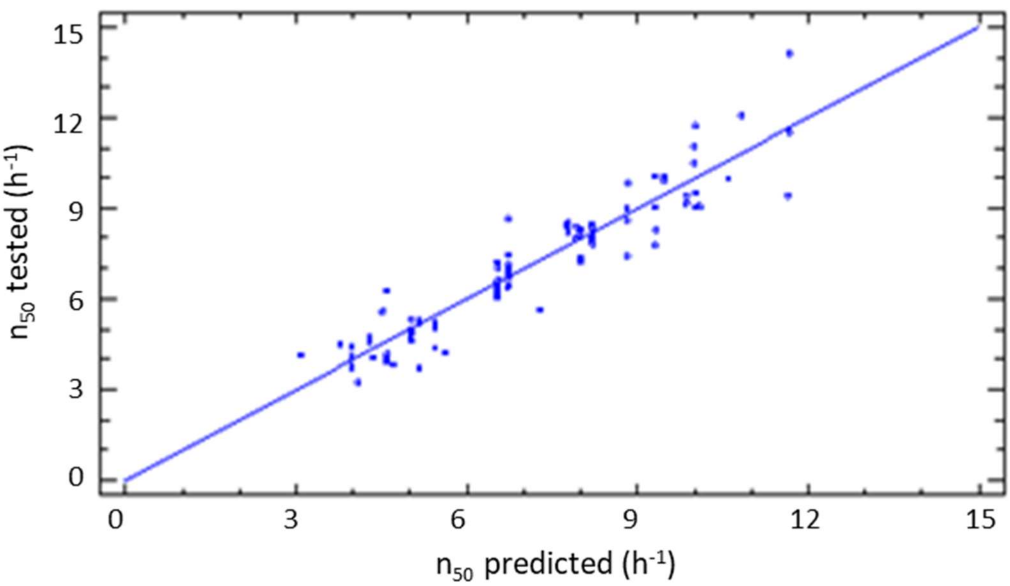

Figure 5. Tested vs estimated n_{50} values for cluster 1: scatter plot

The cumulative frequency curves for the measured and estimated n_{50} values were practically identical, deviating by 5.30 % (estimated<measured) at n_{50}=7.00 h^{-1} and by 3.00 % (estimated> measured) at n_{50}=8.00 h^{-1}. However, the actual values for the parameter showed a higher maximum infiltration rate (n_{50}=14.40 h^{-1}) than the predictive model, which delivered n_{50} values no higher than 12.00 h^{-1} (Figure 6). As the estimated probability distribution exhibited a satisfactory fit to the empirical values, the model can be deemed to be a particularly useful tool for predicting airtightness in the cluster analysed.

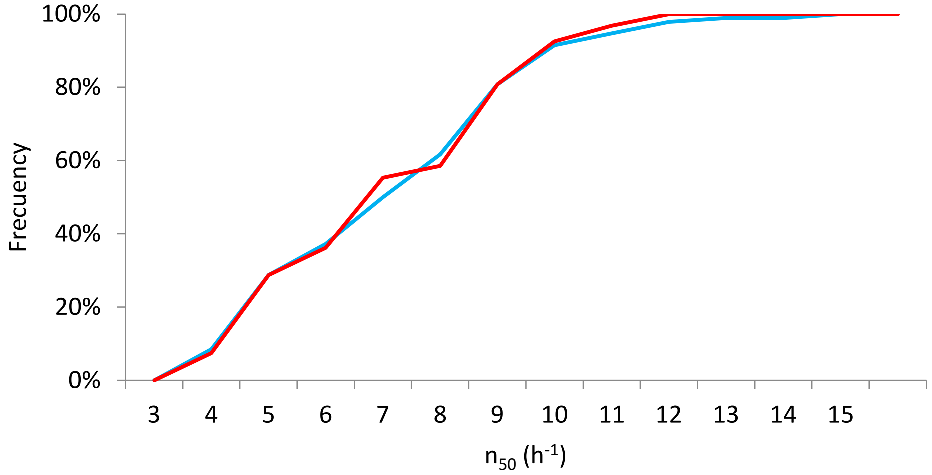

Figure 6. Cumulative frequency (%) of the measured (blue) and estimated (red) cluster 1 n_{50} values

3.2.2. Cluster 2

In the model proposed for cluster 2, which included morphological, construction and geometric characteristics, parameter Y_{SU} was the coefficient for the quantitative variable net floor area. The constants defined for the categorical variables [bathroom window (β_{WB}), general condition (β_{M}), façade type (β_{F}), exposure (β_{TE}) and winter severity (β_{WS})] and their respective categories are listed in Table 9.

Coefficient	Category	Value	Std error
Constant (α)		3.607	1.448
Floor area (γ_A)		0.048	0.021
Bathroom window (β_{WB})	Yes	2.220	0.546
	No	0	0
General condition (β_M)	Poor	5.971	0.829
	Good condition	*	*
	Energy retrofitting	-4.611	1.309
Façade type (β_F))	F 1	2.711	0.554
	F 2	*	*
Exposure (β_{TE})	E 1	-0.399	0.976
	E 2	-1.805	0.711
	E 3	**	**
	E 4	-0.968	0.919
	E 5	*	*
Winter severity (β_{WS})	WS A	*	*
	WS B	2.578	0.486
	WS C	2.699	0.730

*Table 9. Predictive model coefficients for cluster 2 n_{50} (*variable category included in constant ‘α’; ** typology not observed in the sample)*

The predictive model was defined by the probability function shown in Equation 5 and the coefficients listed in Table 9.

$$n_{50} = \alpha + \gamma_A \cdot A + \beta_{WB} + \beta_M + \beta_F + \beta_{TE} + \beta_{WS} \qquad (E\ 5)$$

The optimised model included six independent variables, together accounting for 62.60 % of the variation with a highly significant correlation coefficient. The standard error for the prediction was 1.37 h^{-1} (Table 10), which while significant afforded a more precise fit than the standard deviation of the values measured for cluster 2, for it was approximately half as wide as sigma. In light of possible inter-variable dependence, which even if present would not prove causality, ANOVAs and linear regressions were conducted to determine whether air permeability in the cluster of homes studied could be predicted with a model based on their classificatory characteristics, as premised in the initial hypothesis.

R	R^2	Adjusted R^2	Standard error for the estimate	No. of variables
0.882	0.698	0.626	1.370	6

Table 10. Predictive model for n_{50} in cluster 2: statistical descriptors

The ANOVA delivered an F-statistic with which to test the null hypothesis that R^2 and the slope of the curve were equal to 0; in other words, that the two variables were not correlated. Since the p-value for the F-statistic was lower than the significance level, the null hypothesis was ruled out and the results obtained for the sample were deemed to be applicable to the population from which it was drawn (Table 11).

	Sum of squares	df	Root mean square	F	Sig.
Regression	263.875	10	26.387	14.371	.000
Residuals	75.285	41	1.836		
Total	339.160	51			

Table 11. ANOVA for the cluster 2 n_{50} predictive model

The model results mimicked actual behaviour very closely. The plot of the estimated vs the empirical n_{50} values (Figure 7) showed that nearly all lay in the area between the line of symmetry and the distance defined by the standard error for the estimate (± 1.35), with only six outliers.

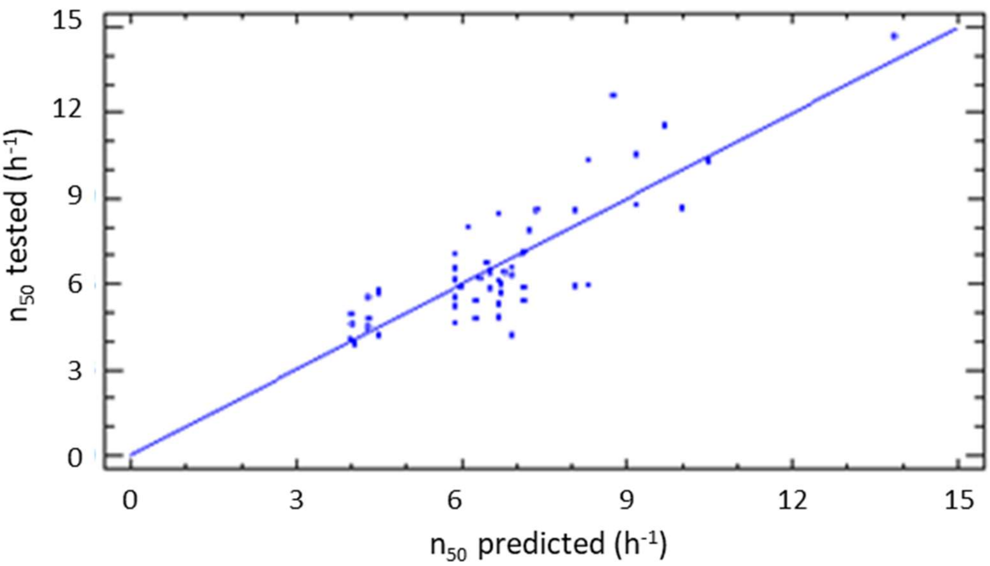

Figure 7. Tested vs estimated n_{50} values for cluster 2: scatter plot

The cumulative frequency curves for the measured and estimated n_{50} values deviated across the entire spectrum by approximately 4 % to 17 %, except at n_{50}=7 h^{-1}, where the estimated value was 30 % higher than the measured value. The most visible changes in cumulative frequency were observed between n_{50} values of 5 h^{-1} and 8 h^{-1} (Figure 8). The estimated probability distribution exhibited an approximate fit to the distribution for the measured data. The divergence found in the central values was somewhat wider than observed for cluster 1, however, a finding that may be attributed to the greater scatter and differential evolution expected in the older than in the newer housing stock.

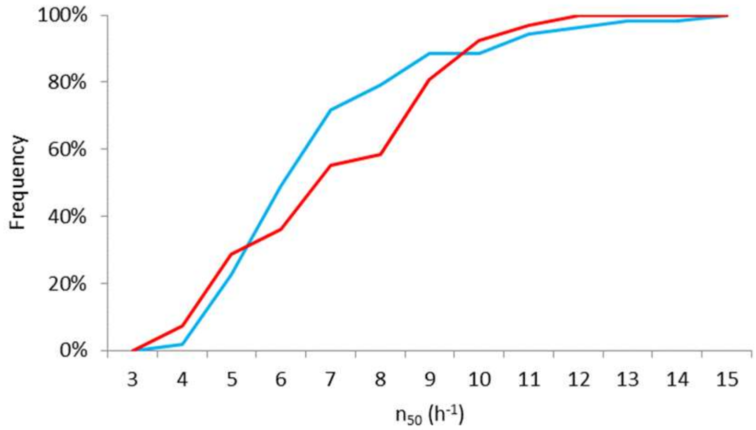

Figure 8. Cumulative frequency of the measured (blue) and estimated (red) cluster 2 n_{50} values.

4. Discussion

4.1. Model assessment and weighting

Predicting envelope airtightness is a complex endeavour due to the large number of variables and factors involved. The predictive model applied here was built with the variables best able to reduce the variation in the results, even though they were not the only factors affecting permeability. The others, associated with the building envelope or particulars such as year of construction or climate zone, were found to be inapt for model construction, either because they had a scant impact on the results or because they introduced too much variation. That notwithstanding, their effects must not be disregarded, in particular when referred to specific items in the population, as discussed in the analysis of the elements comprising building envelopes.

The models developed revealed different behaviours attributable to cluster particulars. The homes built after CT79 tended to exhibit greater uniformity in connection with their basic characteristics such as morphological ratios and construction systems. That may be related to greater stability in design regulations and ordinances for social housing which, outside of a few exceptions analysed in the section on characterisation, had a standardising effect on the housing stock. In contrast, in the cluster grouping housing built prior to 1979, the values of their basic parameters were more scattered, which may be attributed to a certain diversity of legislative provisions and construction programmes that evolved during that period. That diversity gave rise to substantial differences in housing formats, compounded by a wider variety of construction processes (Domínguez-Amarillo et al., 2016).

That situation is mirrored in the performance models proposed, for variation was narrower in the cluster 1 than in the cluster 2 model. The limitations inherent in models designed to predict building airtightness of individual homes must be assumed in their formulation, for performance, as discussed earlier, is not attributable only to deterministic factors. Rather, faulty envelope workmanship, the absence or presence of conservation or specific retrofits may translate into very different values for homes in one and the same building. Consequently, the basic factors that most differentiated the two clusters were individual retrofitting and dwelling deterioration, which had a lesser impact on the more modern than on the older cluster.

Two sets of factors can be defined in the assessment of the two predictive models: those common to both and consequently essential to determining dwelling envelope airtightness and those specific to each cluster.

The factors common to both included:

- winter severity
- degree of envelope exposure
- existence of a bathroom window
- façade type.

All the factors affecting both models were categorical (as opposed to continuous) variables, associated either with geographic location or basic envelope morphology.

The common factor with the heaviest impact was winter severity. Location in areas with scantly severe winters was factored into the model, for it predicted a lower value for parameter n_{50}. Location in areas with severe or moderately severe winters lowered the airtightness values predicted.

The better performance of mild winter (winter severity 1) than severe winter homes was associated primarily with the coastal location of the former, where the effects of wind action are normally greater (García de Pedraza & García Vega, 1990; Sánchez Gallardo, 2002). In other words, construction strategies in such areas apparently focused more on airtightness than on thermal issues. In contrast, winter severity zones 2 and 3 were located inland, where wind action is less intense and less frequent. A distinction may be drawn in these two zones between colder (severity 3) and more temperate (severity 2) winters, with the least airtight homes found in the latter. That effect was visible in the regression coefficients associated with the two models, (although with a narrower difference in the older than in the more modern homes), possibly an indication of greater concern in more modern construction about airtightness control in colder areas.

The second most significant factor was the degree of envelope exposure, which was more significant than façade area. The former, while categorical, accommodated the possibility of singular points and inter-surface abutments, as observed in the model, where type T_{E4} had a significant effect in the more modern and types T_{E2} and T_{E4} in the older homes.

The presence of windows in bathrooms, the third factor of interest, was almost as significant in the models as the preceding variable, given the problem posed by the abutments between joinery and the tile finishes normally found in bathrooms (Jesica Fernández-Agüera et al., 2016).

The fourth common factor, façade type, was somewhat less significant than the preceding variables, particularly in more contemporary buildings. This factor was of greater significance in cluster 2 buildings with single-wythe (T_{F1}) façades than in the others. Here, however, that might be attributed to co-linearity with other particulars not addressed in the analysis, such as quality or type of abutments, rather than to any specific façade system. Although given their constructional characteristics the likelihood of infiltration across such enclosures is low, such façades were associated with the oldest homes in the sample, possibly denoting the presence of secondary factors that might contribute to such poorer performance.

4.2. Factors specific to cluster 1

The predictive model for cluster 1 developments showed them to be affected by four additional parameters:

a. window area and perimeter
b. window type
c. absence of partition between kitchen and living room
d. presence of blinds on windows.

The total window area and perimeter were the factors mainly affecting predicted airtightness, for their dimensional scales were perceptibly larger than those used for the categorical variables. The two dimensions, area and perimeter, were observed to have a similar impact, for normally the ratio between them was on the order of 1 to 3, while the coefficients for these independent variables exhibited a ratio of the same order of magnitude, but inverted.

In this cluster, unlike cluster 2, the floor area of the homes taken by itself carried less weight in the airtightness model than other factors. Nonetheless, dwelling size had to be included in the model, for the analysis conducted to characterise envelopes revealed a clear relationship between home size and the size and area of its openings, primarily in response to the minimum dimensions established in the legislation (standardised in the modern period). Consequently, whilst dwelling size affected relative (more than absolute) permeability, the number and geometry of openings furnished more information on the airtightness of the homes in this cluster.

The importance of openings in airtightness was reinforced by the fact that the second most relevant factor in the model was window type. Joinery type had a heavy impact on permeability, for all other factors being equal, hinged windows were observed to be more airtight than sliding windows, with a higher n_{50} value, a finding consistent with specific studies and earlier reports (Max H. Sherman & Chan, 2006). Although window type has been widely analysed in qualitative terms, the identification here of its weight in the factor matrix is of particular utility and its significance in the more modern homes is of special interest.

The absence of partitions between kitchen and living room was the third and the presence of blinds the fourth weightiest factor in cluster 1. Both represented fairly singular situations extant in a short number of homes in the sample, although where present they modified airtightness performance in the model significantly. Kitchens opening onto the living room generated greater porosity in the envelope. The possible explanation is that in enclosed kitchens, the service piping is better confined and normally closer to the building exterior, whereas in the open arrangement it normally runs across longer distances, weakening the envelope. The homes built after 1979 that had no blinds were observed to be more airtight.

The conclusion drawn from this model was that in more modern homes air permeability tended to be concentrated around windows (or semi-transparent envelopes). In contrast to older flats, airtightness appeared to be governed more evenly by the many components of these dwellings and be less dependent upon window / façade abutments.

4.3. Factors specific to cluster 2

In addition to the four common characteristics, two specific parameters were observed in the cluster 2 developments: one continuous and quantitative, and the other categorical:

a. floor area
b. general condition.

As noted in the preceding section, air flows might be related to more of the envelope components in the older than in the newer buildings. Consequently, the most determinant of the dimensional parameters was home size, here defined as floor area as a predictor. Nonetheless, in contrast to cluster 1, here this factor carried less relative weight than the other model predictors.

The most influential factor in the cluster 2 model for predicting overall airtightness, considering both the common and specific variables, was general condition or degree of conservation. That finding is of particular interest, inasmuch as it infers potential for improvement through sealing and other measures to improve envelope airtightness. The model revealed air flow differences of up to 10 h^{-1} between poorly conserved and rehabilitated dwellings.

4.4. Adaptation of the models to other samples

The predictive models developed in this study will be cross-validated with the homes measured in the national Infiles project 'Energy impact of air permeability in residential buildings in Spain. Study and characterisation of infiltration' (Feijó-Muñoz, Jesús and Meiss, Alberto and Poza-Casado, Irene and Padilla-Marcos, Miguel and Rabanillo-Herrero, Mario and Royuela del Val, Andrés and Gonzalez-Lezcano, Roberto and Pardal, Cristina and Echarri Iribarren, Victor and Assiego de Larriva, Rafae, 2019), funded by the Spanish Ministry of the Economy and Competitiveness, to verify their applicability and accuracy (Figure 9).

The plot of the estimated vs the empirical values for n_{50} in cluster 1 showed that half of the values lay in the area between the line of symmetry and the distance defined by the standard error for the estimate (± 0.87); and in cluster 2 only two values lay in that area (standard error=± 1.35). In contrast to the findings for the sample studied here, cluster 2 had a higher R^2 than cluster 1 when a different sample was used. With R^2 values of 0.77 (cluster 1) and 0.82 (cluster 2), the sample studied was deemed to fit the model developed.

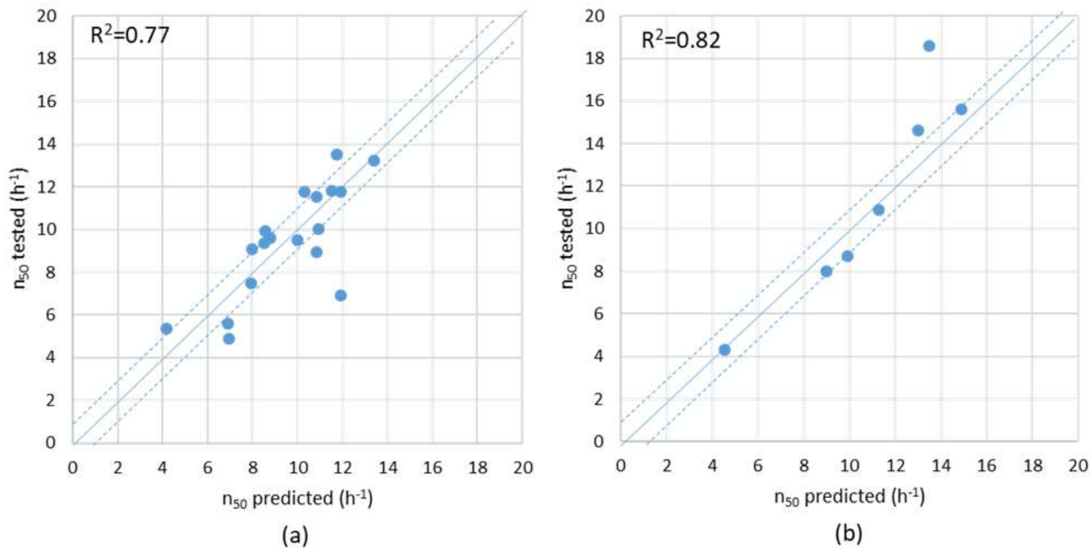

Figure 9. Tested vs estimated n_{50} values for other samples built after 1979 (a) and before 1979 (b): scatter plot

5. *Conclusions*

This paper establishes a series of predictive airtightness models based on measured data with which public and private organisations can estimate airtightness in new-builds (predictive model 1) and the existing stock (predictive model 2). Designed for the Mediterranean region, they are based on a series of readily recognisable parameters that would preclude the need for airtightness testing. The resulting n_{50} data can be used with energy certification software, which presently envisages a mean value for all types of housing, not quantified by testing. They can also be entered into energy and comfort simulation programs to predict housing consumption and expected indoor temperatures.

The air infiltration rate at 50 Pa for the stock as a whole is 7.00 h^{-1} (with a median of 6.52 h^{-1}). The values for the stock in southern Spain are widely scattered (with values fluctuating from 2.50 h^{-1} to 15.57 h^{-1}), particularly as compared to other areas, a finding associated with the breadth of the sample studied, which covered a number of time periods and construction typologies.

The housing stock can be divided into two performance-based sets based on the results attributable to their specific characteristics, from which two models can be derived. For social housing in southern Spain built from 1950 to date, one of the models developed fits pre-1979 developments and the other the homes constructed after that year.

The wider diversity of solutions and morphologies in the pre-1979 stock explains the greater variation and stochasticity observed in this (cluster 2) than in the later group (cluster 1) of dwellings.

Inter-model comparison shows that the most prominent differentiating factors are related to the probability of time-driven alteration, the appearance of individual change and the effects of deterioration. These factors carry greater weight in the cluster comprising older than in the one consisting in more modern dwellings (although individual cases exhibiting a heavier impact may be found in the latter).

Two sets of factors can be defined: those common to both clusters and consequently essential to determining dwelling envelope airtightness and those that are specific to each cluster. Both constitute sets of associated factors that while not individually able to prompt significant differences, taken together induce specific performance patterns.

The factors common to the two models include winter severity, envelope exposure, bathroom window and façade type. The parameters observed to affect the predictive model for cluster 1 (post-1979) developments only include window area, perimeter and type, separate kitchen and blinds on windows. One of the two parameters specific to the (pre-1979) developments, floor area, is quantitative, while the other, a general condition, is qualitative.

Despite the uncertainty associated with largely manual envelope construction methods, the models proposed can predict the airtightness in such dwellings with acceptable accuracy, particularly on the housing stock or development scale.

Acknowledgement

This study was funded by the Spanish Ministry of Economy and Competitiveness, as part of its National R&D+I Plan, under projects BIA2012-39020-C02-01, BES-2013-063097 and REVEN: "Rehabilitación energética de viviendas sociales, aplicando productos innovadores de ventana con marcado CE".

Symbols

- α Constant
- A Floor Area
- A_F Façade area
- A_{CA} Area adjacent to communal areas
- A_D Area adjacent to other dwellings
- A_W Window area
- B Blind
- HVAC Heating, ventilation, and air conditioning
- K Kitchen
- M General condition
- n_{50} Air infiltration rate at 50 Pa
- P_W Window perimeter
- T_B Blind type
- B1 No blinds
- B2 External blinds
- B3 Blinds in splayed openings
- B4 Compact windows blinds)
- TE Exposure type
- E1 Semi-detached, linearly aligned buildings with four homes per storey
- E2 Semi-detached, linearly aligned buildings with two homes per storey
- E3 Open gallery buildings

- E4 Stand-alone high rises
- E5 Semi-detached, linearly aligned buildings with two sper storey and building, located at the corner of the compound or in stand-alone buildings with H, T- or X-shaped ground plans
- TF: façade type
- F1: 1 or 1 and ½ foot brick fabric
- F2: ½ foot brick fabric(or one-brick thick) outer wall + air cavity + hollow brick inner wall
- F3: ½ foot brick fabric (or one-brick thick) outer wall + insulation layer + hollow brick inner wall
- F4: ½ foot brick fabric (or one-brick thick) outer wall + air cavity + plasterboard inner wall
- F5: fired clay panelling + air cavity + insulation + fired clay block
- TW Window type
- W1: hinge opening windows
- W2: sliding windows
- W3: hinged and sliding windows
- SS Summer severity
- V Volume
- WS Winter severity
- WB bathroom window
- Y Year

References

Alalouch, C., Al-Saadi, S., AlWaer, H., & Al-Khaled, K. (2019). Energy saving potential for residential buildings in hot climates: The case of Oman. *Sustainable Cities and Society*, *46*, 101442. https://doi.org/10.1016/j.scs.2019.101442

Alfano, D. F. R., Dell'Isola, M., Ficco, G., & Tassini, F. (2012). Experimental analysis of air tightness in Mediterranean buildings using the fan pressurization method. *Building and Environment*, *53*, 16–25. https://doi.org/10.1016/J.BUILDENV.2011.12.017

Almeida, R. M. S. F., Ramos, N. M. M., & Pereira, P. F. (2017). A contribution for the quantification of the influence of windows on the airtightness of Southern European buildings. *Energy and Buildings*, *139*, 174–185. https://doi.org/10.1016/j.enbuild.2017.01.012

Alves, S., Fernández-Agüera, J., & Sendra, J. J. (2014). Infiltration rate performance of buildings in the historic centre of Oporto. *Informes de La Construcción*, *66*(535), e033. https://doi.org/10.3989/ic.13.009

An, J., Yan, D., & Hong, T. (2018). Clustering and statistical analyses of air-conditioning intensity and use patterns in residential buildings. *Energy and Buildings*, *174*, 214–227. https://doi.org/10.1016/J.ENBUILD.2018.06.035

Caillou, S., & Van Orshoven, D. (2010). *Report on the building airtightness measurement method in European countries. ASIEPI: Intelligent Energy Europe*.

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, *3*(1), 1–27. https://doi.org/10.1080/03610927408827101

Cesar, R. M., & da Fontoura Costa, L. (1997). An introduction to neural networks. *Neurocomputing*, *14*(1), 101–104. https://doi.org/10.1016/S0925-2312(96)00046-X

Chan, W. R., Joh, J., & Sherman, M. H. (2012). Air leakage of US homes: Regression analysis and improvements from retrofit. In *Joint Conference 33rd AIVC Conference and 2nd TightVent Conference Optimising Ventilative Cooling and Airtightness for [Nearly] Zero-Energy Buildings, IAQ and Comfort.* (pp. 35–39).

Chan, W. R., Joh, J., & Sherman, M. H. (2013). Analysis of air leakage measurements of US houses. *Energy and Buildings*, *66*, 616–625. https://doi.org/10.1016/J.ENBUILD.2013.07.047

Chan, W. R., Nazaroff, W. W., Price, P. N., Sohn, M. D., & Gadgil, A. J. (2005). Analyzing a database of residential air leakage in the United States, *39*, 3445–3455. https://doi.org/10.1016/j.atmosenv.2005.01.062

de la Flor, F. J. S., Domínguez, S. Á., Félix, J. L. M., & Falcón, R. G. (2008). Climatic zoning and its application to Spanish building energy performance regulations. *Energy and Buildings*, *40*(10), 1984–1990. https://doi.org/10.1016/j.enbuild.2008.05.006

do Carmo, C. M. R., & Christensen, T. H. (2016). Cluster analysis of residential heat load profiles and the role of technical and household characteristics. *Energy and Buildings*, *125*, 171–180. https://doi.org/10.1016/J.ENBUILD.2016.04.079

Domínguez-Amarillo, S., Fernández-Agüera, J., Campano, M. Á., Acosta, I., Domínguez-Amarillo, S., Fernández-Agüera, J., … Acosta, I. (2019). Effect of Airtightness on Thermal Loads in Legacy Low-Income Housing. *Energies*, *12*(9), 1677. https://doi.org/10.3390/en12091677

Domínguez-Amarillo, S., Fernández-Agüera, J., Sendra, J. J., & Roaf, S. (2018). Rethinking User Behaviour Comfort Patterns in the South of Spain—What Users Really Do. *Sustainability*, *10*(12), 4448. https://doi.org/10.3390/su10124448

Domínguez-Amarillo, S., Sendra, J. J., & Oteiza San José, I. (2016). *La envolvente térmica de la vivienda social: el caso de Sevilla, 1939 a 1979 Title* (1st ed.). Madrid (Spain): Editorial Consejo Superior de Investigaciones Científicas.

E. Forgy. (1965). Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classification | BibSonomy. *Biometrics*, *21*(3), 768–769. Retrieved from https://www.bibsonomy.org/bibtex/c86383cba8cfe00d5e6ef200016aca3f

Eskola, L., Alev, Û., Arumägi, E., Jokisalo, J., Donarelli, A., Sirén, K., & Kalamees, T. (2015). Airtightness, air exchange and energy performance in historic residential buildings with different structures. *International Journal of Ventilation*, *14*(1), 11–26. https://doi.org/10.1080/14733315.2015.11684066

European Commission. (2011). *The roadmap for transforming the EU into a competitive, low-carbon economy by 2050*.

Feijó-Muñoz, Jesús and Meiss, Alberto and Poza-Casado, Irene and Padilla-Marcos, Miguel and Rabanillo-Herrero, Mario and Royuela del Val, Andrés and Gonzalez-Lezcano, Roberto and Pardal, Cristina and Echarri Iribarren, Victor and Assiego de Larriva, Rafae, M. (2019). *Permeabilidad al aire de los edificios residenciales en España. Estudio y caracterización de sus infiltraciones* (Ediciones). Spain.

Feijó-Muñoz, J., González-Lezcano, R. A., Poza-Casado, I., Padilla-Marcos, M. Á., & Meiss, A. (2019). Airtightness of residential buildings in the Continental area of Spain. *Building and Environment*, *148*, 299–308. https://doi.org/10.1016/J.BUILDENV.2018.11.010

Feijó-Muñoz, J., Pardal, C., Echarri, V., Fernández-Agüera, J., Assiego de Larriva, R., Montesdeoca Calderín, M., … Meiss, A. (2019). Energy impact of the air infiltration in residential buildings in the Mediterranean area of Spain and the Canary islands. *Energy and Buildings*, *188–189*, 226–238. https://doi.org/10.1016/J.ENBUILD.2019.02.023

Fernández-Agüera, J., Sendra, J., Suárez, R., Domínguez-Amarillo, S., & Oteiza, I. (2015). Airtightness and indoor air quality in subsidised housing in Spain. *AIVC*.

Fernández-Agüera, J., Sendra, J. J., & Domínguez, S. (2011). Protocols for measuring the airtightness of multi-dwelling units in Southern Europe. In *Procedia Engineering* (Vol. 21). https://doi.org/10.1016/j.proeng.2011.11.1992

Fernández-Agüera, Jesica, Domínguez-Amarillo, S., Sendra, J. J., & Suárez, R. (2016). An approach to modelling envelope airtightness in multi-family social housing in Mediterranean Europe based on the situation in Spain. *Energy and Buildings*, *128*, 236–253. https://doi.org/10.1016/j.enbuild.2016.06.074

Fernández-Agüera, Jesica, Domínguez-Amarillo, S., Sendra, J. J., Suárez, R., & Oteiza, I. (2019). Social housing airtightness in Southern Europe. *Energy and Buildings*, *183*, 377–391. https://doi.org/10.1016/J.ENBUILD.2018.10.041

Ferrara, M., Monetti, V., & Fabrizio, E. (2018). Cost-Optimal Analysis for Nearly Zero Energy Buildings Design and Optimization: A Critical Review. *Energies*, *11*. https://doi.org/10.3390/en11061478

Fomento, M. Orden FOM/1635/2013, de 10 de septiembre, por la que se actualiza el Documento Básico DB-HE “Ahorro de Energía”, del Código Técnico de la Edificación, Gobierno de España § (2013). Madrid, Spain: B.O.E. (Boletín Oficial del Estado).

García de Pedraza, L., & García Vega, C. (1990). Contrastes y afinidades climáticas entre el noreste y suroeste de la Península Ibérica: Cataluña - Andalucía atlántica. *Revista de Metereología A.M.E.*, *13*, 59–73.

Gillott, M. C., Loveday, D. L., White, J., Wood, C. J., Chmutina, K., & Vadodaria, K. (2016). Improving the airtightness in an existing UK dwelling: The challenges, the measures and their effectiveness. *Building and Environment*, *95*, 227–239. https://doi.org/10.1016/j.buildenv.2015.08.017

Gobierno, P. del. Real Decreto 2429/1979, de 6 de julio, por el que se aprueba la norma básica de edificación NBE-CT-79, sobre condiciones térmicas en los edificios. Boletín Oficial del Estado. (1979).

Hesaraki, A., Myhren, J. A., & Holmberg, S. (2015). Influence of different ventilation levels on indoor air quality and energy savings: A case study of a single-family house. *Sustainable Cities and Society*, *19*, 165–172. https://doi.org/10.1016/J.SCS.2015.08.004

Hynek, D. (2011). BLOWER DOOR TESTING IN MULTIFAMILY BUILDINGS. *Home Energy*, *28*(5), 32–41. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=8gh&AN=70457950&site=ehost-live

ISO. (2015). *ISO 9972: 2015 Thermal performance of buildings -- Determination of air permeability of buildings -- Fan pressurization method*.

Jesús, Feijó-Muñoz; Irene, Poza-Casado; Roberto Alonso, González-Lezcano; Cristina, Pardal; Víctor, Echarri; Rafael, Assiego L.; Jesica, Fernández-Agüera; María Jesús, Dios-Viéitez; Víctor José, del C.-D.; Manuel, Montesdeoca C.; Miguel Ángel, Padilla-Mar, M. (2018). Methodology for the Study of the Envelope Airtightness of Residential Buildings in Spain: A Case Study. *Energies*, *4*(704).

Johnston, D., Wingfield, J., Miles-Shenton, D., & Bell, M. (2004). Airtightness of UK Dwellings : Some Recent Measurements. *The RICS Foundation Construction and Building Research Conference*, (September), 7–8. Retrieved from http://www.leedsbeckett.ac.uk/as/cebe/projects/cobra04-3.pdf

Jones, B., Das, P., Chalabi, Z., Davies, M., Hamilton, I., Lowe, R., … Taylor, J. (2015). Assessing uncertainty in housing stock infiltration rates and associated heat loss: English and UK case studies. *Building and Environment*, *92*, 644–656. https://doi.org/10.1016/J.BUILDENV.2015.05.033

Kalamees, T. (2007). Air tightness and air leakages of new lightweight single-family detached houses in Estonia, *42*, 2369–2377. https://doi.org/10.1016/j.buildenv.2006.06.001

Kanungo, T., Member, S., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., … Member,

S. (2002). An Efficient k -Means Clustering Algorithm : Analysis and Implementation. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, *24*(7), 881–892.

Krstić, H., Koški, Ž., Otković, I. I., & Španić, M. (2014). Application of neural networks in predicting airtightness of residential units. *Energy and Buildings*, *84*, 160–168. https://doi.org/10.1016/j.enbuild.2014.08.007

Krstic, H., Otkovic, I. I., & Todorovic, G. (2015). Validation of a model for predicting airtightness of residential units. *Energy Procedia*, *78*, 1525–1530. https://doi.org/10.1016/j.egypro.2015.11.182

Liddament, M. (1986). A Review of European Research into Airtightness and Air Infiltration Meassurement Techniques. *Measured Air Leakage of Buildings: A Symposium, ASTM Committee E-6 on Performance of Building Construction*, (904), 407–415.

Lloyd, S. (1982a). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

Lloyd, S. (1982b). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

Maaleudstyr, S. og persienners lufttaethed. E. (1984). Air infiltration through shutters and roller blinds. A laboratory rig measuring. Denmark.

Mata, É., Sasic Kalagasidis, A., & Johnsson, F. (2014). Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK. *Building and Environment*, *81*, 270–282. https://doi.org/10.1016/J.BUILDENV.2014.06.013

Ministerio de Vivienda. (2006a). Código Técnico de la Edificación (CTE). *Real Decreto 314/2006 de 17 de Marzo*. https://doi.org/CTE-DB-SE

Ministerio de Vivienda. Real Decreto 314/2006, de 17 de marzo, por el que se aprueba el Código Técnico de la Edificación. (2006). Ministerio de Vivienda. Gobierno de España: BOE-A-2006-5515.

Montoya, M I, Pastor, E., Carrie, F. R., Guyot, G., & Planas, E. (2010). Air leakage in Catalan dwellings: Developing an airtightness model and leakage airflow predictions. *Building and Environment*, *45*(6), 1458–1469. https://doi.org/10.1016/j.buildenv.2009.12.009 ER

Montoya, María I., Pastor, E., & Planas, E. (2011). Air infiltration in Catalan dwellings and sealed rooms: An experimental study. *Building and Environment*, *46*(10), 2003–2011. https://doi.org/10.1016/j.buildenv.2011.04.009

N Gaitani, Lehmann, C., Santamouris, M., Mihalakakou, G., & Patargias, P. (2010). Using principal component and cluster analysis in the heating evaluation of the school building sector. *Applied Energy*, *87*(6), 2079–2086.

Nabinger, S., & Persily, A. (2011). Impacts of airtightening retrofits on ventilation rates and energy consumption in a manufactured home. *Energy and Buildings*, *43*(11), 3059–3067. https://doi.org/10.1016/j.enbuild.2011.07.027

Paap, L., Mikola, A., Teet-Andrus, K., & Kalamees, T. (2012). Airtightness and Ventilation of new Estonian Apartments Constructed 2001-2010. In *33nd AIVC Conference. Optimising*

Ventilative Cooling and Airtightness for [Nearly] Zero-Energy Buildings, IAQ and Comfort. Copenhagen, Denmark.

Pan, W. (2010). Relationships between air-tightness and its influencing factors of post-2006 new-build dwellings in the UK. *Building and Environment*, *45*(11), 2387–2399. https://doi.org/10.1016/j.buildenv.2010.04.011

Papadopoulos, F., Whiffen, T. R., Tilford, A., & Willson, C. (2018). Actual energy and environmental savings on energy retrofit works at the Lakes Estate, Milton Keynes. *Sustainable Cities and Society*, *41*, 611–624. https://doi.org/10.1016/j.scs.2018.01.046

Pereira, P. F., Almeida, R. M. S. F., Ramos, N. M. M., & Sousa, R. (2014). Testing for building components contribution to airtightness assessment. In *35th AIVC Conference " Ventilation and airtightness in transforming the building stock to high performance"* (pp. 322–330).

Persily, A., Musser, A., & Emmerich, S. J. (2010). Modeled infiltration rate distributions for U.S. housing. *Indoor Air*, *20*(6), 473–485. https://doi.org/10.1111/j.1600-0668.2010.00669.x

Prignon, M., & Van Moeseke, G. (2017). Factors influencing airtightness and airtightness predictive models: A literature review. *Energy and Buildings*. https://doi.org/10.1016/j.enbuild.2017.04.062

Ramos, N., Almeida, R., Curado, A., Pereira, P., Manuel, S., & Maia, J. (2015). Airtightness and ventilation in a mild climate country rehabilitated social housing buildings - What users want and what they get. *Building and Environment*, *92*, 97–110. https://doi.org/10.1016/j.buildenv.2015.04.016

Ramos, N. M. M., Almeida, R. M. S. F., Simões, M. L., Delgado, J. M. P. Q., Pereira, P. F., Curado, A., … Fraga, S. (2018). Indoor hygrothermal conditions and quality of life in social housing: A comparison between two neighbourhoods. *Sustainable Cities and Society*, *38*, 80–90. https://doi.org/10.1016/j.scs.2017.12.016

Salehi, A., Torres, I., & Ramos, A. (2017a). Assessment of ventilation effectiveness in exiting residential building in mediterranean countries: Case study, existing residential building in Portugal. *Sustainable Cities and Society*, *32*, 496–507. https://doi.org/10.1016/j.scs.2017.04.018

Salehi, A., Torres, I., & Ramos, A. (2017b). Experimental analysis of building airtightness in traditional residential Portuguese buildings. *Energy and Buildings*, *151*, 198–205. https://doi.org/10.1016/j.enbuild.2017.06.037

Sánchez Gallardo, F. (2002). REDUCCIÓN DE LA VULNERABILIDAD A LOS FENÓMENOS METEOROLÓGICOS Y CLIMÁTICOS EXTREMOS. In C. de Documentación & del I. N. de M. (INM) (Eds.), *Día Meteorológico Mundial de 2002* (pp. 1–24). Madrid (Spain): Centro de Publicaciones. Ministerio de Medio Ambiente.

Santamouris, M., Mihalakakou, G., Patargias, P., Gaitani, N., Sfakianaki, K., Papaglastra, M., … Zerefos, S. (2007). Using intelligent clustering techniques to classify the energy performance of school buildings. *Energy and Buildings*, *39*(1), 45–51. https://doi.org/10.1016/J.ENBUILD.2006.04.018

Scibor, M. (2019). Are we safe inside ? Indoor air quality in relation to outdoor concentration of PM 10 and PM 2 . 5 and to characteristics of homes, *48*(April 2018).

Sfakianaki, A., Pavlou, K., Santamouris, M., Livada, I., Assimakopoulos, M.-N., Mantas, P., & Christakopoulos, A. (2008). Air tightness measurements of residential houses in Athens, Greece. *Building and Environment*, *43*(4), 398–405. https://doi.org/10.1016/J.BUILDENV.2007.01.006

Sherman, M. (1995). The Use of Blower-Door Data. *Indoor Air*, *5*(3), 215–224. https://doi.org/10.1111/j.1600-0668.1995.t01-1-00008.x

Sherman, M., & Mcwilliams, J. (2007). Air Leakage of U . S . Homes : Model Prediction. *Thermal Performance of Exterior Envelopes of Whole Buildings X International Conference*, (January), 18.

Sherman, M.H., & Chan, R. (2004). Building Airtightness: Research and Practice. *Lawrence Berkeley National Laboratory*, (February), 1–46. https://doi.org/10.4324/9781849770620

Sherman, Max H., & Chan, W. R. (2006). Building Air Tightness: Research and Practice. In M. Santamouris & P. Wouters (Eds.), *Building Ventilation: The State of the Art* (1 edition, pp. 137–162). Routledge.

Sinnott, D., & Dyer, M. (2012). Air-tightness field data for dwellings in Ireland. *Building and Environment*, *51*, 269–275. https://doi.org/10.1016/j.buildenv.2011.11.016

Sousa, G., Jones, B. M., Mirzaei, P. A., & Robinson, D. (2018). An open-source simulation platform to support the formulation of housing stock decarbonisation strategies. *Energy and Buildings*, *172*, 459–477. https://doi.org/10.1016/J.ENBUILD.2018.05.015

Suárez, R., & Fernández-Agüera, J. (2015). Passive energy strategies in the retrofitting of the residential sector: A practical case study in dry hot climate. *Building Simulation*. https://doi.org/10.1007/s12273-015-0234-7

Taber, R. (2009). Clustering (Xu, R. and Wunsch II, D.C.; 2009) [Book review]. *IEEE Computational Intelligence Magazine*, *4*(3), 92–95. https://doi.org/10.1109/MCI.2009.933101

Vinha, J., Manelius, E., Korpi, M., Salminen, K., Kurnitski, J., Kiviste, M., & Laukkarinen, A. (2015). Airtightness of residential buildings in Finland. *Building and Environment*, *93*, 128–140. https://doi.org/10.1016/j.buildenv.2015.06.011

Walker, I. S., Sherman, M. H., Joh, J., & Chan, W. R. (2013). Applying large datasets to developing a better understanding of air leakage measurement in homes. *International Journal of Ventilation*, *11*(4), 323–338. https://doi.org/10.1080/14733315.2013.11683991