Energy & Buildings 335 (2025) 115549

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

Clustering of archetypal building-inhabitant pairs to improve energy efficiency: The case of the Walloon region in Belgium

Guirec Ruellan [a,*], Shady Attia [a], Gentiane Haesbroeck [b]

[a] Sustainable Building Design Lab, Dept. UEE, Faculty of Applied Sciences, Université de Liège, Belgium
[b] Department of Mathematics, Faculty of Sciences, Université de Liège, Belgium

ARTICLE INFO

Keywords:
Renovation strategies
Socio-economic typologies
Statistical analysis
Bottom-up model
Energy policies
Clustering
Walloon region

ABSTRACT

Considering the pressing need to renovate existing buildings across Europe, there exists a shared consensus on the importance of developing targeted strategies. Despite research, regulation and action in all the countries of Western Europe, the rate of energy renovation is stagnating. Turning this consensus into effective action requires models that accurately represent the diverse building stock and its inhabitants. This article focuses on Wallonia, aiming to identify typical combinations of buildings and inhabitants representative of the region's situation. The methodology employs ordinal logistic regression and beta regression algorithms to analyze correlations, leveraging extensive databases encompassing technical and socio-economic data. Subsequently, K-means clustering is utilized to distill the building stock into several characteristic typologies, offering insights into the diversity of the region. Notably, our findings highlight certain underestimated building types, constituting a significant portion of the Walloon building stock. Firstly, a typology of low energy performance houses, mainly attached and semi-detached, inhabited by low-income households makes up more than 17% of Walloon housing. More unexpectedly, a type of low energy performance house, mostly 4-fronted, inhabited by high-income households, still makes up more than 11% of the built stock. These results underscore the efficacy of our methodology in harmonizing disparate datasets and provide novel insights into the building stock and its occupants. Furthermore, the identified typologies empower researchers and policymakers to address the renovation challenge by directing targeted actions at appropriate scales, whether regional or local. Overall, this article contributes to a deeper understanding of the complexities surrounding building renovation and offers practical implications for policy formulation and implementation.

1. Introduction

1.1. Background

The war in Ukraine and the subsequent rise in energy prices (Fig. 1) have brought a fresh look at the issue of energy savings, particularly building renovation. According to Ari et al. [5], "By the end of the first quarter of 2022, crude oil prices had doubled, coal prices tripled, and natural gas prices increased more than five-fold relative to early 2021″. This inflation has also affected the cost of materials and loan rates [17]. This is happening in a context that has been largely disrupted by COVID-19 and the increase in teleworking, which have already had a profound effect on housing market prices [31,80]. In the space of a few months, Europe has found itself faced with the dual urgency of having to continue to combat the climate crisis while at the same time redirecting a large part of its energy supply [32,61].

In the short term, the energy crisis has become priority number one. In the medium term, however, it has put the spotlight back on

Abbreviations: CEHD, Centre d'Etudes en Habitat Durable (Centre for Sustainable Housing Studies); CREG, Commission de Régulation de l'Electricité et du Gaz (Belgian Electricity and Gas Regulatory Commission); CWaPE, Commission Wallonne Pour l'Energie (Walloon Energy Commission); DG TLPE, Direction Général Territoire, Logement, Patrimoine et Environnement (Walloon Administration for Land, Housing, Heritage and the Environment); EPBD, Energy Performance of Buildings Directive; EPC, Energy Performance Certification; GDPR, General Data Protection Regulation; GHGE, Green House Gas Emissions; ILR, Isometric Log Ratio; IWEPS, Institut Wallon de l'Evaluation, de la Prospective et de la Statistique (Walloon Institute of Evaluation, Foresight and Statistics); LCA, Life Cycle Assessment; LCC, Life Cycle Cost; PEB, Performance Energétique des Bâtiments (Walloon EPC); PCA, Principal Component Analysis; SPF, Service Publique Fédéral (Belgian Federal Public Service); STATBEL, Belgian Statistical Office.

* Corresponding author.
E-mail address: guirec.ruellan@uliege.be (G. Ruellan).

https://doi.org/10.1016/j.enbuild.2025.115549
Received 30 July 2024; Received in revised form 14 January 2025; Accepted 2 March 2025
Available online 3 March 2025

commitments already made by the European countries and regions. The 55 % reduction in GHGEs (Green House Gas Emissions) by 2030 goes hand in hand with a reduction in energy demand [18,32]. Among the various options considered, the efficient energy renovation of the built stock would make it possible to reduce both the amount of energy consumed and dependence on Russian gas [80]. In this respect, the particularly old Belgian buildings are both a difficulty and a major challenge [18,70].

Energy renovation was sometimes considered primarily through the prism of the environment [26,65] or even solely in terms of the climate. However, given the urgency of the situation, governments have remembered that the issues of political and diplomatic independence and maintaining social peace must also be considered [14,31,82,83].

The rise in energy prices has had a direct impact on the moderation of energy consumption and the increase in energy efficiency work [75]. But this increase has also had a direct impact on public finances through the subsidies introduced to protect a growing proportion of the population. In Belgium, the CREG (Belgian Electricity and Gas Regulatory Commission) estimates the cost of the social tariff at €1.7 billion in 2021, 2022 and 2023, including €1 billion in 2022 alone [19]. Similar policies have been implemented throughout Europe. This makes it more important to transform this lost expenditure into a longer-term investment in building stock, regardless of the geopolitical situation in Eastern Europe. Regulations would benefit from being more geared towards renovation [102]. While renovation remains a costly policy, both on a household and national scale, it is even more important to compare it with the expected savings [102].

One of the main challenges is to gain a better understanding of the building stock and its behavior. For the moment, we can see a correlation between interest in renovation work and energy prices (Fig. 2) while observing a stagnation in applications for renovation permits [29]. This is proof of the reluctance of private individuals to commit themselves to cumbersome administrative, technical, and financial procedures. For similar reasons, the Walloon Region has decided to return to a simplified system of grants for energy-saving work [90]. This is a long-standing debate between better control of the work carried out to improve quality, and simplification of the procedures for renovation work to improve quantity. For the moment, renovation work that generates energy savings of over 60 % only concerns 0.2 % of projects [31].

A previous article has already assessed the statistical correlation between the technical characteristics of the home and the socio-economic characteristics of its occupants[84]. This article complemented other studies which had looked at these aspects but did not attempt to correlate them[4]. But this better understanding of the link between technical factors in the building and socio-economic factors in the inhabitant still needs to be refined to propose policies in favor of renovation that activate the right levers.

1.2. Determinants of energy consumption and renovation

There has been a significant increase in the number of articles devoted to sustainable renovation over the last few years [42]. At the same time, political and legislative tools are being put in place to accelerate the rate of major renovation. Article 23 of the Belgian Constitution defines the right to decent housing. However, Belgian efforts are mainly focused on the technical aspect of building, with the regional integration of the EPBD (Energy Performance of Buildings Directive) standard. Scientific publications show that passive renovation of buildings is possible [33]. Questions are being raised about the optimum level to be reached [102] and the right scale of intervention [64], about heritage issues [78] and workforce skills [52]. All these studies approach the building as a technical object whose performance must be optimized. At the same time, there is a lack of research [73] into renovation strategies and their social sustainability [49,94].

Yet there are important links between the energy and social characteristics of the residential sector [79,85], even if “connections between energy, housing affordability and well-being are still under-researched” [17]. In the Walloon region, for example, the weighted average annual bill has risen sharply (58.9 % for electricity and 63.8 % for gas). [21]. The various types of energy insecurity (measured, hidden, perceived) [68] will now affect 20.6 % of Belgian households in 2021 and 28.8 % of Walloon households [67]. There is considerable heterogeneity between and within municipalities, while housing inequalities are still relatively low in Belgium. [25]. Fuel poverty exacerbates poverty levels and disparities in well-being [96]. In the long term, housing inequalities may be associated with 18.5 % of deaths. [74]. Acting solely on energy prices can also be counter-productive: a reduction in VAT on electricity was accompanied by an increase in consumption. [45]. Occupant behavior plays an important role in the risk of the rebound effect [11,40,43], which has been much discussed recently [76]. It is important to note, however, that this Energy Performance Gap is due not only to the rise in temperature but also to the improvement in air quality [100]. Whatever the method used to define and calculate a rebound effect [38], it is undeniable that it tends to

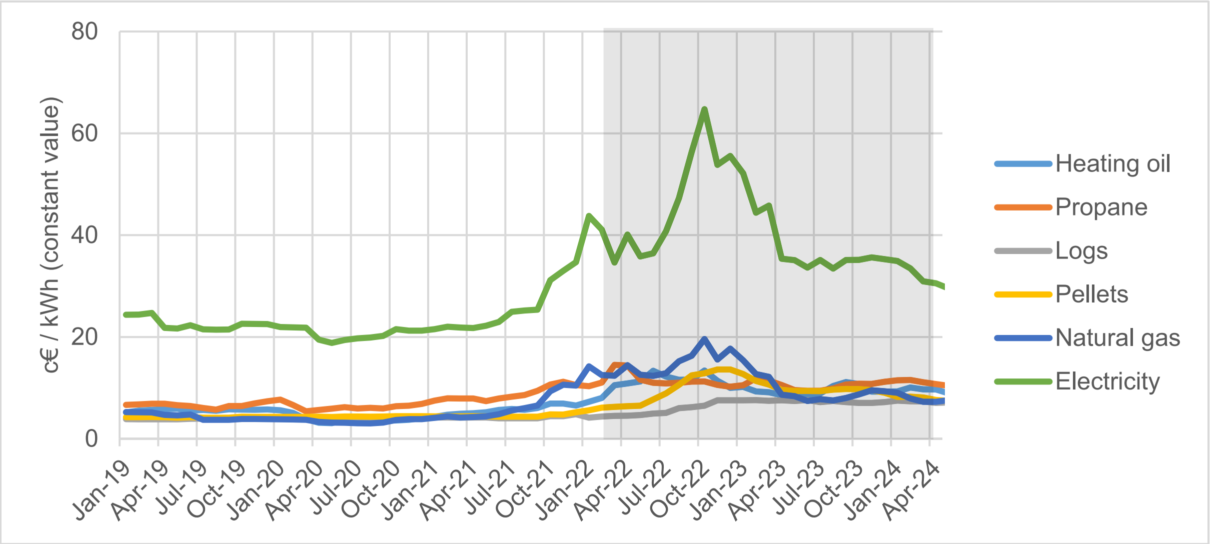

Fig. 1. Evolution of the price of energy purchased by households in recent years in Belgium. Value at constant currency. Source CREG, SPF économie, ValBiom; Statbel, 2024.

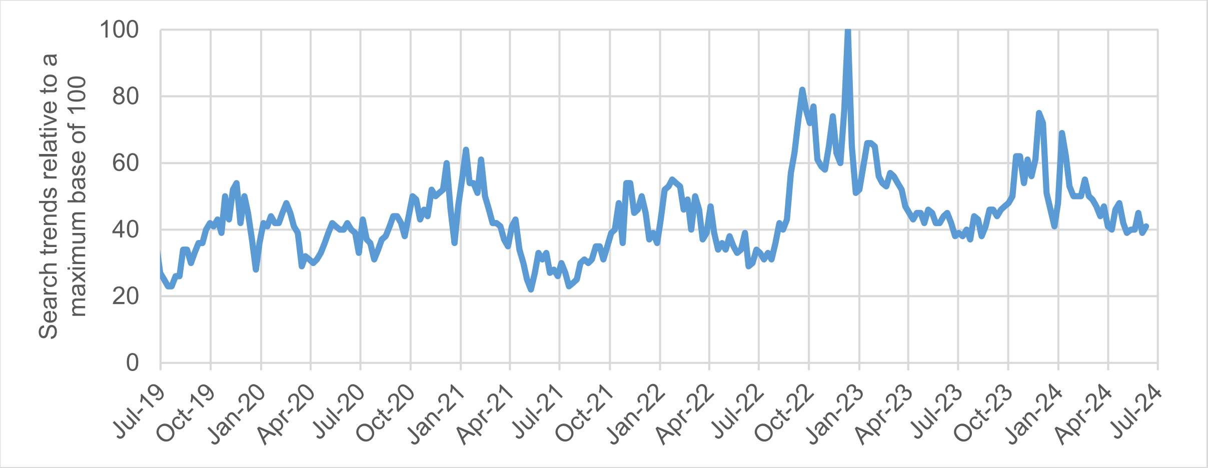

Fig. 2. “Building Thermal Insulation” research trends on Google in Belgium. . Source: *trends.google.com*, 2024

reduce the results of energy efficiency work [7]. All the more so, the inhabitants of low energy performance buildings are already tending to adopt energy-saving behaviors [92].

The occupant is also a determining factor in the decision to renovate [82]. For example, reducing carbon emissions is rarely a factor in the decision to renovate. Would-be renovators are motivated by aesthetics, comfort or energy savings[52]. Despite this, the role of residents remains largely unintegrated in policy initiatives[9]. Although financial tools [12]) have been put in place to reach the various profiles, they are not very effective for certain profiles. [57]. Private rental housing is a poor relation when it comes to financial aid.[13,59,62], because of the gap between the objectives of owners and tenants[6]. Rising house prices also contribute to widening existing gaps[80]. The economic issue remains the main obstacle to renovation work[2,52], even though such work would be cost-optimal from a macro-economic point of view[18]. More progressive work could help to broaden the target audience[69]. More generally, far-reaching renovation must no longer be the sole objective; we need to multiply strategies and approaches [50]. It is, therefore, essential to integrate the social aspects of the inhabitant into the article of the evolution of the built stock through its renovation.

1.3. *Built stock modeling*

Among studies of energy in buildings, modeling the built stock is one of the proven techniques [18,42]. There are two main families of residential energy consumption models [93]. Top-down models are based on macro-economic variables, which estimate the breakdown between different sectors. These models do not consider possible improvements at the end-user level. Bottom-up models construct and extrapolate end-user consumption. They are better suited to estimating political measures and energy renovation strategies [72].

Table 1
Belgian and Walloon built stock models.

Sources	Dwelling typologies	Geographic area	Approach
[44]	960	Belgium	Representative
[99]	5	Belgium	Typical
[54]	8	Walloon region	Mixed
[70]	11	Belgium	Typical
[3]	16	Belgium	Representative
[22]	25	Belgium	Mixed
[41]	8	Walloon region	Typical
[71]	4	Walloon region	Typical
[39]	992	Walloon region	Representative
[60]	42 (22 existing)	Walloon region	Mixed
[18]	20	Walloon region	Representative

Table 1 summarizes the main bottom-up models developed for Belgium and the Walloon region. In 2001, Hens et al. [44] proposed 960 typologies representative of the built stock, classified according to several criteria (age, type, surface area, primary energy, central heating or not) to article the energy-saving potential of the Belgian built stock. The same team of researchers considered 5 individual references separated according to architectural criteria (classic villa, typical Flemish rural house, modern commissioned house, small working-class terraced dwelling, large upper-class terraced dwelling) houses to articulate the financial viability of renovation [99]. Kints and De Herde [54] used a matrix of 28 typologies classified by age and typology based on the results of the 2001 Census [97]. They combine them to distinguish 8 priority architectural typologies in terms of renovation for the Walloon region. The Low Energy Housing Retrofit (LEHR) article [70] proposes 11 architectural typologies representative of the stock of buildings in need of renovation in Belgium. Allacker et al. [3] used 16 typologies representative of the existing Belgian building stock, according to age and type, for the SuFiQuaD project to evaluate LCAs and LCCs. Based on the previous work, the TABULA article [22] proposes a mixed classification of 25 typologies based on 5 construction periods and 5 types of housing units. Monfils and Hauglustaine use the same typologies as Kints for their article on the Walloon region, with 8 representative typologies [41]. The first cost-optimum article COZEB 1 [71] studies a large number of building archetypes, but focuses on 4 typologies of existing dwellings among those proposed by Kints to set the optimal energy performance of these dwellings. Gendebien et al. [39] rely heavily on the above work. The tree-based approach creates 992 building typologies to analyze the impact of different renovation scenarios on a national scale. The second cost-optimum article, COZEB 2 [60], studies 40 building typologies, including 22 typologies of existing housing, to refine and update the results of COZEB 1[71]. The TIMES model [18] has been developed for the Walloon region, including 20 categories of building typologies depending on the period of construction and the number of facades. TIMES models are general bottom-up energy system models based on a set of criteria. These typologies are mainly based on the COZEB 2 dataset. In addition, it should be noted that some studies focus on a single typology [8] to specify its characteristics.

Two approaches can be distinguished in the creation of typologies, according to Cyx et al. [22]. The representative approach creates simulated buildings based on technical characteristics (age of the building, type of dwelling, surface area, etc.), which are aggregated into matrices. The typical approach uses existing buildings that are representative of the built stock. This approach is often based on architectural archetypes: mansions, workers’ houses, and four-fronted villas.

However, the two approaches tend to merge, with typical buildings being selected based on representative criteria and vice versa.

The quality and availability of the data studied is a key issue in data analysis [25,53,77]. While more and more data are being collected in our societies, the question of how to aggregate it is an unavoidable problem. It is necessary to standardize and improve methodologies and datasets to obtain more robust findings [58].

The transposition of the EPBD [24]/91/CE sur le performance énergétique des bâtiments, 2002) provides a large amount of information on the composition of the Walloon building stock [15,47]. These data can be used to describe energy use and renovation potential [35,56], by statistical analysis [91] or linear regression [84,101]. Clustering improves the use and cross-referencing of these data [86] while maintaining a reasonable number of representative typologies.

The models by Hens [44] and Gendebien [39] are designed for algorithmic use, which explains the large number of typologies (~900). The other models are less precise but allow us to work on a few typologies (from 4 to 25) that are easier to identify.

The 2001 Census [97] and the 2011 Census [66] provide the other major databases on which most of these studies base their models. They also offer the advantage of providing information on both the composition of the built stock and the socio-economic aspects of the population.

A review of the literature shows that the integration of socio-economic data into building stock modeling is still in its early stages [42,53]. The social factors that are essential for understanding the decision to renovate are not considered [10]. “Building stock modeling has so far focused on modeling technological aspects of the development of energy use in the stock and neglect complex interactions between technology, economics and policy when modeling the development” [72]. Sartori et al.[87]go even further in Norway by estimating the renovation rate based on population statistics alone.

A few studies nevertheless include technical and socio-economic factors based on surveys[63]. In the Walloon region, in particular, the CEHD (Centre for Sustainable Housing Studies) produced a report on housing quality in 2014 [16]. A new report is expected for 2024. This article combines socio-economic data on the occupant (age, tenure status, socio-economic status, nationality, household composition, household size, length of tenure, rent or credit, occupation of the dwelling) and technical data (type of dwelling, period of construction, surface area, heating method, energy consumption, wall insulation, condition of the building, building environment). The main drawback of this extremely comprehensive article is that it does not provide a representative typology of the built stock. However, it does offer the most detailed view of the link between technical and socio-economic characteristics in the Walloon region. In the models of the built stock studied in Table 1, only Verbeeck and Hens [99]make slight mention of this criterion by distinguishing two models of terraced dwelling: lower class and upper class.

As far as geographical delimitation is concerned, there are as many Belgian models as there are regional models. However, it seems simpler and more coherent to create a regional model. From a practical point of view, building and population databases are mainly regionalized. From a political point of view, building, energy and climate policies are largely regionalized in Belgium [18]. This probably explains the predominance of Walloon models in recent years. In any case, while there are significant differences between regions, spatial disparities in the built stock go beyond the north–south divide [98]. The European Commission also recognizes the importance of regions and cities in defining appropriate energy-saving solutions [32].

1.4. Objectives

This article aims to help increase the renovation rate by enabling policies to promote the energy-efficient renovation of housing in the Walloon region to be better targeted. Prior to this article, several factors were identified as playing a major role in the decision to renovate or not to renovate a dwelling.

The first objective of the article is to cross-reference technical data on the building with socio-economic data on the residents. The inclusion of socio-economic data enhances the understanding of occupant behavior, financial capacity, and decision-making processes related to building renovations. This information is critical for identifying barriers to energy efficiency, such as financial constraints, knowledge gaps, or behavioral tendencies. By clustering building-inhabitant pairs based on both technical and socio-economic attributes, the study identifies typologies that allow for tailored interventions. For example, low-income households in low energy performance buildings may require targeted subsidies or simplified renovation processes, while high-income households might benefit from tax incentives or stricter regulations. This alignment ensures that renovation strategies are not only technically sound but also socio-economically feasible, thereby increasing the likelihood of adoption and achieving significant energy savings. The second objective is to use the results of this cross-referencing to create clusters of building-individual pairs. These clusters will enable a new analysis of the composition of the built stock.

The third objective is to examine the distribution of these clusters. Their distribution across the Walloon region and the different municipalities will already provide keys to understanding the actions that can be envisaged at the different levels of government.

This research holds broader significance beyond the Walloon Region, offering a replicable methodology for integrating socio-economic and technical building data to identify actionable renovation strategies. European countries share similar challenges with aging building stocks and diverse occupant profiles, complicating the implementation of effective energy renovation policies. By demonstrating a scalable approach that aligns building typologies with resident characteristics, this study provides insights that can be adapted to other regions facing the dual pressures of climate commitments and energy crises. The findings emphasize the importance of localized interventions that balance technical efficiency with social equity, reinforcing the utility of this framework for advancing renovation strategies across Europe.

This article is innovative in two ways. First, it presents a solution for crossing datasets by using regression models to predict missing socio-economic information based on technical building data and vice versa, followed by clustering to group building-occupant pairs into representative typologies. This approach harmonizes distinct datasets, allowing for the extraction of actionable insights that connect occupant characteristics with building energy performance. The same methodology can be applied to other regions with similar databases, such as all European countries with similar EPC [1] and census data [95]. This methodology is also not limited to the proposed characteristics; it is easily possible to subsequently integrate other information that could have been collected. This methodology also has the advantage of avoiding problems related to the privacy of individuals because the criteria studied are all anonymous. This methodology also has the advantage of avoiding issues related to the privacy of individuals, as the criteria studied are all anonymous.

Regarding the results, this article offers a new perspective on the constitution of the built stock. It focuses on the correlations that exist between the building and its occupant. This building-occupant couple is essential to understanding the response of the built stock to the challenges of energy renovation. While the numerical results are specific to the Walloon case study, the methodology provides a framework that could potentially inform renovation strategies in other regions with comparable challenges, such as aging building stocks and the need for policies integrating socio-economic and technical data [27,34,36,37,89].

This article relies on a large amount of data already available. On the one hand the data from the PEB certification, result of the implementation of the EPBD policy. On the other hand, the statistical data collected by the official institutes. It should be noted that the everincreasing amount of data generated is leading to an interest in how to exploit them to understand the dynamics of many issues better. Once this data has been selected, cleaned, and sorted, it is possible to cross-reference it statistically to deduce potential correlations and extrapolate general profiles. The result is not only the classification of the Walloon building stock into some major archetypes. But also, the possibility to article the different Walloon municipalities according to the distribution of these archetypes to better understand their composition and to better target energy efficiency policies.

After this introduction, the methodology section allows us to go into more detail on the different steps of data processing. This part will explain how the databases were identified, collected, and cleaned. Then how different regression algorithms crossed the different databases. Finally, the clustering methods were used to simplify the expression. The result section will present the main results of the article. We will first focus on the clustering result. But we will also detail the implications of these results on the assumed energy efficiency of the built stock and the distribution of the clusters in the different Walloon municipalities. Finally, some validation steps will confirm the quality of the obtained results from a statistical point of view. These results will then be exploited through a discussion part that will highlight them with respect to what is already known about the Walloon building stock. We will also take the opportunity to broaden the conclusions and propose new avenues for the article.

The issue of energy renovation is of particular concern to the countries of Western Europe. The Netherlands, the United Kingdom, France, Denmark, Belgium and others are all characterized to some degree by a building stock that largely pre-dates the first energy regulations and by slow growth. Despite a great deal of research and regulation at the national and European level, as well as numerous awareness-raising, funding and training initiatives, the rate of energy renovation is growing slowly. Of all the obstacles that have been identified, this article focuses on improving knowledge of the composition of the housing stock and its occupancy applied to the Walloon region. This should enable more effective renovation strategies to be designed.

This article proposes a new model of the Walloon housing stock, in 6 or 10 clusters in this case. This model is based not only on the geometry and assumed energy efficiency of the dwellings. Above all, it incorporates certain occupant characteristics that are essential for understanding the decision to renovate, such as income and whether they are homeowners. This study, therefore, fills a research gap between a highly technical approach to the composition of the built environment and housing policies geared towards residents or homeowners. This article proposes a new methodology. By drawing on existing databases in all the countries of Western Europe that are subject to the same difficulties, it promotes a result that is both robust and reproducible. It is also easy to improve the model by integrating new databases using tools like those used or by updating the databases already in use. Finally, the results provided by this article make it possible to approach the question of renovation strategies in the Walloon region differently. By way of example, the cluster representing very energy-intensive housing occupied by low-income households represents more than 17 % of the built stock. This represents almost 1 household in 5 at high risk of fuel poverty. In contrast, the cluster representing very energy-intensive housing occupied by high-income households represents more than 11 % of the built stock. Renovation strategies for this cluster are totally different from those for the first group.

2. Methodology

2.1. Tools

The entire statistical processing work was done with the R programming language version 4.2.1. The programming interface is RStudio. Table 2 presents the packages that were used during this article in addition to the libraries natively present in R.

Table 2
Extra packages use.

Library	Use	Version
“readxl “	read excel table	1.3.1
“compositions”	isometric log ratio (ILR) transform	2.0–4
“VGAM”	logistic ordinal regression	1.1–5
“lmtest”	likelihood-ratio test	0.9–38
“betareg”	beta regression	3.1–4
“ggplot2“	plotting	3.3.5
“Hmisc”	weighted statistics	4.7–0

The developed code is available on Zenodo (https://doi.org/10.5281/zenodo.7708755) under the GPL-3 license.

2.2. Methodology flow chart

The article’s methodology is based on several successive statistical operations, the general structure of which is shown in Fig. 3. This flow chart highlights the link between databases relating to cities to the left and databases relating to energy certificates to the right. The methodology is divided into three main stages, each requiring several sub-stages: data collection and treatment (2.3), data crossing (2.4) and clustering (2.5). The first part of the work involves identifying ’individuals,’ which, in the statistical sense, refer to data points representing specific entities within the datasets. In this study, ’individuals’ correspond to building units associated with socio-economic data at the household level. These data points are aggregated and expressed at the level of ’individual-cities,’ meaning that all data within a given city are pooled to form representative values for that city. This allows the methodology to connect the characteristics of the building stock with socio-economic variables at the municipal scale rather than focusing on individual buildings or households. These correlations are used to deduce the probable missing information on the second type of “individuals”: for each energy certificate, probable socio-economic data are predicted. If these “individual-certificates” do not have value independently of each other, they can, on the other hand, draw trends by grouping them by cluster. The distribution of these clusters can then be studied at different territorial scales, either directly to highlight specific clusters or through a PCA to highlight specific cities. The various stages and the results obtained are described in greater detail below.

Data collection and treatment.

Data Collection

The article relies on existing databases. Table 3 summarizes references of the different databases used. Most of this data is published by public institutions. The data is then freely available. The energy data is directly taken from the PEB (Performance Energétique des Bâtiments) database. The PEB is the application for the Walloon region of the European EPC (Energy Performance Certificates) rating scheme. In the rest of the article, this database will be referred to as EPC. To respect the GDPR (General Data Protection Regulation), the EPC data collected are anonymized by the absence of any precise information (names, exact addresses).

All the data collected are compiled in two tables. The CITY table (Table 4) compiles all known data (typology, income, ownership, households) on the constitution of Walloon cities (268 cities). For each city is indicated the name of the city, the name of the province, the number of inhabitants and the proportion relative to the Walloon population, the proportion of owners, the proportion of each housing typology (apartments, 2, 3 and 4 facades) and the proportion of the different income categories.

The EPC table (Table 5) compiles data from the EPC database (632,833 certificates) that will be used in this article. For each certificate is indicated the name of the city, the specific primary energy consumption, the energy label, the destination (public, apartment, single-family house) and the number of free facades. The EPC data are corrected so that the city names are written the same in both databases. The

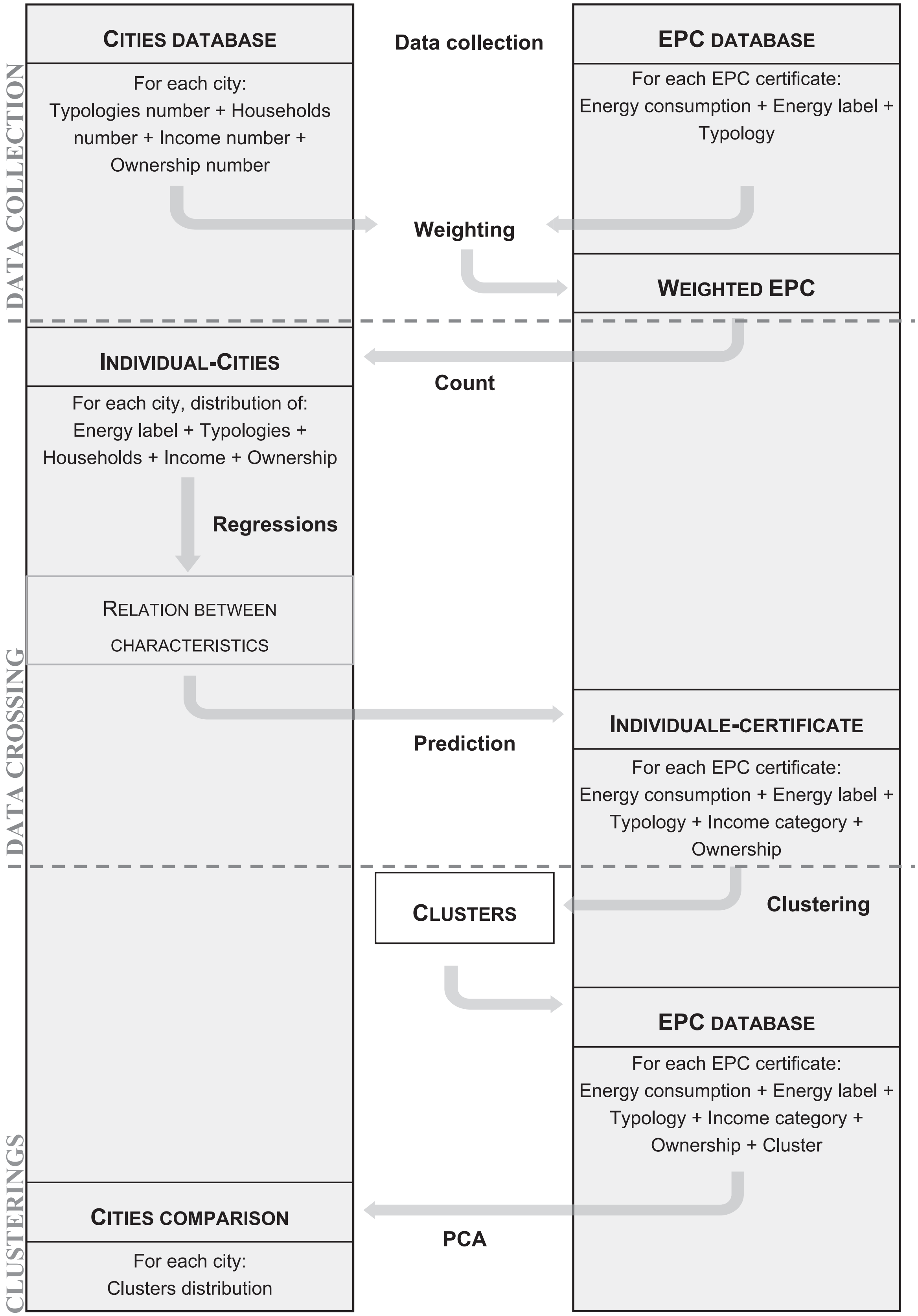
DATA COLLECTION
DATA CROSSING
CLUSTERINGS
CITIES DATABASE
For each city:
Typologies number + Households number + Income number + Ownership number
Data collection
EPC DATABASE
For each EPC certificate:
Energy consumption + Energy label + Typology
Weighting
WEIGHTED EPC
INDIVIDUAL-CITIES
Count
For each city, distribution of:
Energy label + Typologies + Households + Income + Ownership
Regressions
RELATION BETWEEN CHARACTERISTICS
Prediction
INDIVIDUALE-CERTIFICATE
For each EPC certificate:
Energy consumption + Energy label + Typology + Income category + Ownership
CLUSTERS
Clustering
EPC DATABASE
For each EPC certificate:
Energy consumption + Energy label + Typology + Income category + Ownership + Cluster
CITIES COMPARISON
PCA
For each city:
Clusters distribution

Fig. 3. Methodology framework for the cross-exploitation of technical data from energy certificates and socio-economic data from official databases.

Table 3
Summary table of the dataset.

Category name	Data name	Source	Data year
Energy	EPC label and consumption	DG TLPE	2020
Geometry	Type of building in housing	STATBEL	2018
Income	Categories of income	IWEPS	2016
Ownership	Property title	Census	2011
Households	Number of households	STATBEL	2018

A, A + and A++ labels are combined.

The EPC data are also cleaned of some unnecessary or unusable data:

- 1,522 certificates for municipal buildings.
- 1,012 certificates without geometry details.
- 3 certificates with a specific consumption lower than −100 kWh/m^2.year. Buildings with renewable energy production (such as photovoltaic panels) can achieve negative energy consumption. However, a production value that is too high is difficult to explain in terms of energy logic.
- 8 certificates with a specific consumption higher than 10,000 kWh/m^2.year. If there are very energy-intensive buildings, the analysis of certificates presenting such extreme values shows inconsistencies between the quantities of heated surfaces and loss-producing surfaces.

The EPC table finally consists of 630,278 certificates representing a built stock of 1,749,879 dwellings in the Walloon Region.

Weighting.

The weighting process ensures the representativeness of the EPC data relative to the Cities Dataset. This was checked by comparing the distributions of building typologies (e.g., apartments, attached houses) in both datasets using chi-square tests (“chisq.test” function in library “stats”).

Discrepancies were corrected by applying proportional weights to the EPC data.

2.3. *Data crossing*

Count.

The matching of datasets was conducted by linking municipal-level data from the Cities Dataset with the geographic identifiers in the EPC data.

This weighting step allows us to obtain information on the corrected distribution of the certificates and to calculate a weighted energy efficiency (“wtd.mean” function in library “Hmisc”) of the Walloon building stock.

At this stage, all the data collected was therefore pooled at the same scale, that of cities.

Regressions.

The new combined database contains the following data for each city: income, geometry, energy label, ownership, province, and city size. The interest of this first pooling is to be able to calculate the correlations that exist between the different characteristics studied.

Income, geometry, and energy labels are expressed as percentages of representation of different categories. The sum of these categories is equal to 1 (100 %). It is not possible to use a traditional regression model on this type of redundant data. Applying an Isometric Log Ratio (ILR) transformation [28] on these distributions (“ilr” function in library “compositions”) allows for use in a regression model while reversibly keeping all the information [46].

The first regression model looks for explanatory variables for the distribution of income categories. The specificity of this regression is the existence of a hierarchy between the different income categories (from lowest to highest). Ordinal logistic regression [88] (“vglm” function in library “VGAM”) takes this hierarchy into account during the statistical

Table 4
CITY table structure.

City	Typology				Income						Owner	Households	City size	Province
262 Walloon cities (name)	attached (%)	semidetached (%)	detached (%)	apartments (%)	Cat. 1. (%) inc. < 10.000€	Cat. 2 (%) 10.001€ < inc. < 20.000€	Cat. 3 (%) 20.001€ < inc. < 30.000€	Cat. 4 (%) 30.001€ < inc. < 40.000€	Cat. 5 (%) 40.001€ < inc. < 50.000€	Cat. 6 (%) 50.001€ < inc.	(%)	(number)	relative city size (%)	(name)

Table 5
EPC table structure.

Certificate	City	Destination	Free Facade	Specific primary energy	Energy label
630,278 certificates	(name)	(public / apartment / single family)	(detached / one free / two free / three free)	(kWh/m^2.an)	(A / B / C / D / E / F / G)

treatment. Different explanatory variables are tested to find the best formula using several likelihood ratio tests [30] (“lrtest_vglm” function in library “VGAM”). The full model with non-proportional odds performs better than other models. The most efficient model (modINCOME) explains the ordered income variable with typology, energy label, province, and city size.

The second regression model looks for explanatory variables for the distribution of ownership rates in different cities. The ownership rate must be between 0 and 100 %. The use of a beta regression algorithm [20] (“betareg” function in library “betareg”) allows to maintain this requirement of a response variable between 0 and 1. Different explanatory variables are tested to find the best formula using several likelihood ratio tests (“lrtest” function in library “lmtest”). The most efficient model (modOWNER) explains the ownership variable with typology, energy label, income, and province.

Prediction

These two regression models (modINCOME and modOWNER) allow income and property to be deduced based on typology, energy label, province and city size. They can be used to complete the characteristics of the EPC certificates. For each certificate, we can deduce a probability of belonging to the different income (“predictvglm” function in library “VGAM”) and ownership categories (“predict” function in library “stats”).

2.4. Clusterings

Certificates clustering.

For certificate clustering, the specific primary energy consumption of each certificate will be considered rather than the energy label. This makes it easier to differentiate between certificates. All certificates are considered, regardless of their location (province or city). These choices allow us to consider only quantitative variables in the clustering.

The main objective of this article is a better understanding of the composition of the built stock to renovate it. The A, A + and A++ energy certificates can be considered outliers. They are few, and might require one or more clusters to represent them. Consequently, it is decided to remove them during the clustering.

At this point, each certificate has a probability of being attached to each income category. In practice, each certificate is occupied by a single income category. A draw determines which income category each certificate belongs to, considering the probability of belonging to each category. The income category is drawn at random 10 times. This allows us to check that this draw has no effect on their distribution within the built stock. A slight variance appears after clustering on the distribution of the different clusters: a change in cluster size of up to 5 % of the total population may be observed. However, the other characteristics of the clusters show variations in values of less than 1 %, making their analysis identical.

To ensure that each criterion has the same weight when clustering, the energy, ownership, and income categories are standardized: distribution of mean 0 and standard deviation 1.

The first method for estimating the number of clusters [55] to be created is to look for an inflection in the within groups sum of squares. This inflection is between 2 and 11 but needs to be refined. The “NbClust” function in the library “NbClust” compares different ways to obtain the optimal number of clusters but can only be used on small databases for hardware reasons. The function is used on samples drawn from our database, but the results vary from one test to another between 3 et 10. Based on these initial results, it was decided to limit the number of clusters tested to 10. The minimum number of clusters tested is set in accordance with the state of the art (Table 1), in which no study proposes fewer than 4 clusters. Clustering will, therefore be tested from 4 to 10 centers.

A k-means clustering method is used [48] (“kmeans” function in library “stats”). This is a clustering method that allows to manage of large numerical databases in a reasonable amount of time. Multiplying the iterations (200 maximum iterations) and testing different numbers of centers (from 4 to 10) reduces the known drawbacks.

The cluster results are corrected by de-standardizing the variables and calculating the representativeness (see Weighting) of each certificate in the cluster. A quick analysis of the distribution of cluster centers enables the clusters to be sorted according to the most explicit distributions. In the results section, we will look at the two clusters that seem most interesting to analyze from a technical and socio-economic point of view.

Cities PCA.

Once each certificate has been assigned to a cluster, it is possible to calculate the distribution of these different clusters in each city in proportion to their representativeness. We can see that certain specific clusters are more represented in certain cities, while other clusters correspond to a different type of cities.

In this case, we will try to identify general trends rather than focusing on specific clusters. To do this, we run a Principal Component Analysis (PCA) [51] of these distributions (“prcomp” function in library “stats”). PCA is a way of reducing the number of dimensions in a database while limiting the loss of information. In our case, the number of dimensions is 6 or 10, corresponding to the number of clusters chosen during clustering. PCA is used to obtain new dimensions chosen to represent the variability of the old dimensions best. The representation of the first two dimensions of these PCAs allows us to highlight in a 2D-plot the cities with the most distinctive k-means cluster distributions, considering a maximum of information present in the 6 or 10 initial dimensions.

This type of representation should allow highlighting groups of cities with a similar composition or, on the contrary, to highlight cities with a very different composition from the majority group.

2.5. Validation

The entire data collection and statistical processing process was subject to various validation stages. We have already explained in the previous sections how the databases were processed and corrected, as well as the various stages that led to the calibration of the different algorithms used to obtain the most accurate results.

However, to ensure the validity of the regression algorithms, it is also necessary to ensure that the results are statistically exploitable. For the ordinal logistic regression model, the Pearson distribution of residuals highlights potential outliers that could call into question the validity of the model. For the beta regression model, the calculation of the generalized leverage [81] allows us to identify the individuals having the most influence on the regression.

Robustness and Limitations.

On one hand, the article relies on well-known and validated databases of large size to maximize their representativeness. A cleaning step allows to removal of outliers, and a calibration step enables to correction of some of the known biases of the EPC database. Moreover, the open access code allows everyone to verify it and reproduce the experiment with other databases.

On the other hand, the article is based on databases collected at different times. Not only is there a gap of several years between the different databases, but also within the same database (EPC certificate).

This limitation must be put into perspective because the characteristics of the buildings – particularly the Walloon buildings – change slowly. The differences generated affect the accuracy of the results but do not compromise the conclusions drawn. The proposed methodology also leads to several simplification steps (sampling, drawing) whose impact has been verified but which reduce accuracy.

Furthermore, energy certificates are widely considered to be imprecise tools for measuring energy performance [23] for two main reasons: the lack of precise knowledge of existing buildings, which implies many default values, and the lack of consideration of user behavior. However, these very real weaknesses are moderate for the use made of them here. Even if the performance of a certificate can be largely underestimated or overestimated, analyses on large samples will smooth out the most significant errors. Furthermore, what interests us is not so much the actual consumption as the hierarchy of these consumptions between high energy performance, moderate energy performance and low energy performance buildings. Finally, if user behavior can indeed have a large impact on energy performance, the very idea of the article to jointly study socio-economic data makes it possible to integrate, in a very simplified way, a first behavioral issue.

3. Results

3.1. Certificates clustering and built stock characterization

The main result of the article is the characterization of the built stock by clustering the aggregated data. Two sets of clusters are kept. These two sets represent different characteristics of the built stock. Two digits number each cluster obtained: the first corresponds to the importance of this cluster in the clustering and the second corresponds to the number of clusters in the clustering. For example, cluster 1/6 is the cluster with the most individuals out of 6 clusters. The first set of 6 clusters (Table 6) is characterized primarily by its energy performance, ownership rate, and income. One of the most interesting clusters is Cluster 5/6, which contains 12 % of the housing stock. It gathers houses with low energy performance occupied by high-income owners. We also notice that one of the clusters with the highest rate of home ownership (Cluster 3/6) is also characterized by low energy performance for low-income households. This typically represents households in potential fuel poverty.

On the other hand, clusters 4/6 and 6/6, which contain the highest energy performance housing (excluding label A), are mainly occupied by households with higher incomes. From this first clustering, we can deduce the importance of accentuating renovation aids for low-income households in clusters 1/6, 2/6 and 3/6, which group ∼ 63 % of the dwellings, mainly with average and low energy performances. But it is also necessary to put in place more restrictive regulations to push cluster 5/6 to invest in energy efficiency. Following this analysis, each cluster can be designated not only by a number but also by its main characteristics.

The representation of clusters (Fig. 4) by province highlights the similarities between some provinces. On the one hand, the former industrial provinces of Hainaut and Liège, which are home to the two largest agglomerations, are more likely to be made up of clusters 1/6 “mid performance – compact – renter” and 3/6 “low performance – low income – owner”. In contrast, the provinces of Walloon Brabant, Namur and Luxembourg are composed of more 4-front and semi-detached houses with moderate (2/6 “mid performance – low income”) or high (4/6 “high performance – house – owner”) performances.

The distribution in Belgian municipalities of clusters 3/6 “low performance – low income – owner” and 5/6 “low performance – high income – owner”, which deserve particular attention, will be highlighted in Section 3.3.

The second set of 10 clusters (Table 7) completes the observations by highlighting characteristics specific to housing typologies. We observe 6 clusters specific to specific typologies (3 for apartments due to their over-representation in the certificates, 1 for each other typology) and 4 clusters with mixed typologies (2 high performance and 2 low performance).

Among the clusters representing specific typologies, we see that detached (5/10) and semi-detached (4/10) houses are more owner-occupied than attached houses (2/10) and apartments (8/10, 9/10 and 10/10). We also note that the most energy-intensive clusters are overwhelmingly owner-occupied. However, it constitutes only 2 % of the built stock, cluster 10/10 concentrates on the weaknesses with poor energy efficiency, low ownership rates and low-income categories. With cluster 1/10 much more owner-occupied, they constitute the two clusters at risk of fuel poverty. Following this analysis, each cluster can be designated not only by a number but also by its main characteristics.

This multiplication of clusters allows us to see more detail in the distribution by province (Fig. 5). For example, Walloon Brabant is characterized by its high rate of relatively high-performance dwellings (clusters 3/10 “high performance – high income – house – owner”, 7/10 “high performance – high income – renter”, 9/10 “high performance – low income – apartment – renter”), which is explained by its recent urbanization. Conversely, the province of Liège is characterized by its relatively high rate of low-performance apartments (cluster 10/10 “low performance – low income – apartment – renter”). The cluster 1/10 “low performance – low income – apartment –– renter” is particularly present in the provinces of Hainaut and Liège.

In the same way, as for clusters 3/6 “low performance – low income – owner” and 5/6 “low performance – high income – owner”, a more detailed analysis of the distribution of clusters 1/10 “low performance – low income – house – owner” and 6/10 “low performance – high income – owner” with similar characteristics will be carried out in Section 3.3.

3.2. Certificate distribution analysis

The weighting of the certificates before statistical analysis provides information on their distribution and the constitution of the built stock. The weighting correction of the certificates of the 4 typologies in each city (Fig. 6) highlights the lower weight of apartment certificates in all cities. Apartments are overrepresented in the certificates. All other typologies are given greater weight, which means they are under-represented. The attached houses have the greatest variability in their representativeness.

This is explained by Walloon legislation which requires the creation of certificates for new housing and rented housing. However, these are two categories of housing in which apartments are largely over-represented, while terraced houses are generally older and more occupied by owners.

Table 6
Centers characteristics of 6 clusters.

Cluster	Proportion	Specific primary energy	Apartment	Attached	Semidetached	Detached	Owner	Income
1/6 “mid performance – compact – renter”	∼23 %	331	70 %	30 %	0 %	0 %	37 %	2.0
2/6 “mid performance – low income”	∼23 %	361	0 %	7 %	50 %	43 %	67 %	2.2
3/6 “low performance – low income – owner”	∼17 %	709	16 %	23 %	34 %	27 %	79 %	2.1
4/6 “high performance – house – owner”	∼15 %	277	0 %	14 %	23 %	63 %	75 %	5.5
5/6 “low performance – high income – owner”	∼12 %	702	4 %	21 %	29 %	46 %	84 %	5.3
6/6 “high performance – compact – renter”	∼10 %	276	75 %	25 %	0 %	0 %	42 %	5.1

Income: average of income categories (from 1 to 6) of cluster members (see Table 4).

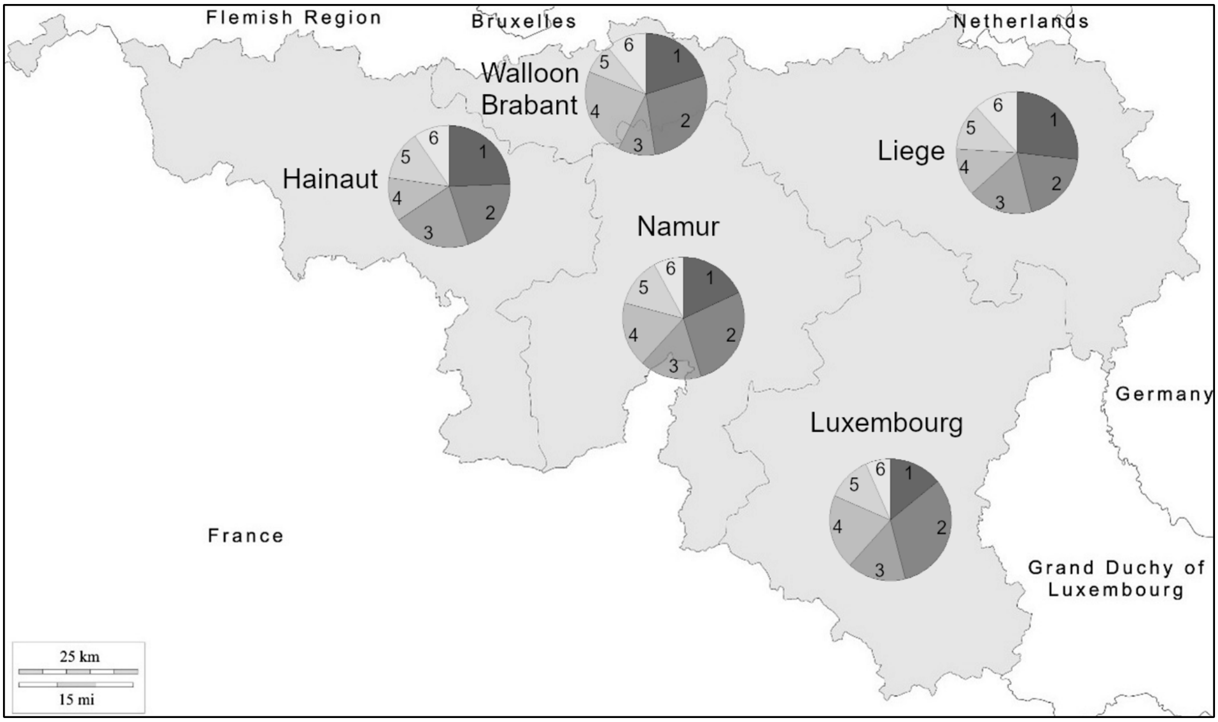

Fig. 4. Distribution of the 6 clusters by province.

Table 7
Centers characteristics of 10 clusters.

Cluster	Proportion	Specific primary energy	Apartment	Attached	Semidetached	Detached	Owner	Income
1/10 “low performance – low income − house – owner”	~16 %	705	0 %	29 %	43 %	28 %	85 %	2.1
2/10 “mid performance – low income – attached – renter”	~15 %	373	0 %	100 %	0 %	0 %	44 %	2.2
3/10 “high performance – high income − house – owner”	~14 %	265	0 %	15 %	25 %	60 %	75 %	5.6
4/10 “mid performance – low income – semidetached – owner”	~12 %	376	0 %	0 %	100 %	0 %	63 %	2.2
5/10 “mid performance – low income – detached – owner”	~11 %	397	0 %	0 %	0 %	100 %	73 %	2.2
6/10 “low performance – high income – owner”	~11 %	700	3 %	24 %	21 %	52 %	84 %	5.4
7/10 “high performance – high income – renter”	~7%	267	73 %	26 %	1 %	0 %	42 %	5.7
8/10 “mid performance – low income − apartment – renter”	~6%	372	100 %	0 %	0 %	0 %	22 %	2.2
9/10 “high performance – low income − apartment – renter”	~6%	173	100 %	0 %	0 %	0 %	48 %	2.5
10/10 “low performance – low income − apartment – renter”	~2%	704	100 %	0 %	0 %	0 %	55 %	2.5

Income: average of income categories (from 1 to 6) of cluster members (see Table 4).

An analysis of the specific energy statistics of the unweighted and weighted certificates according to the distribution of the typologies (Table 8) also highlights the overrepresentation of the best-performing certificates. The average consumption of the certificates is 417 kWh/m^2.an. After weighting, this average consumption becomes 437 kWh/m^2.an. In the same way, the different quartiles confirm that a certification of the whole building stock would probably lead to a decrease in the estimated performance of the building stock.

3.3. Cities distribution

During the clustering analysis (3.1), two typologies of clusters emerged: clusters of low performance buildings occupied by low-income households (3/6 “low performance – low income – owner” and 1/10 “low performance – low income − house – owner”) and clusters of low performance buildings occupied by wealthy households (5/6 “low performance − high income – owner” and 6/10 “low performance – high income – owner”). The distribution of these clusters in the different communes of the Walloon region is shown in Fig. 7. The same representation can be used for the other clusters obtained in the cluster analysis. However, the choice has been made to concentrate here on those whose characteristics appear to be the most interesting in terms of the objective of improving energy efficiency. In addition, the representation of clusters with redundant characteristics makes it possible to check the homogeneity of the results between 6-clustering and 10-clustering.

For these two cluster typologies, 6-groups (top maps) and 10-group clustering (bottom maps) give practically identical distributions. It can also be seen that some municipalities concentrate high rates of these two clusters, reflecting an overall inefficiency of their built stock and, more generally, their older urbanization. Finally, while we might expect large towns (>50,000 inhabitants) to be poorly represented by clusters representing wealthy households, it is more surprising to find that this is also the case for clusters representing low-income households. This can be explained by the greater compactness of urban buildings (flats and terraced housing), which implies lower energy consumption, and which

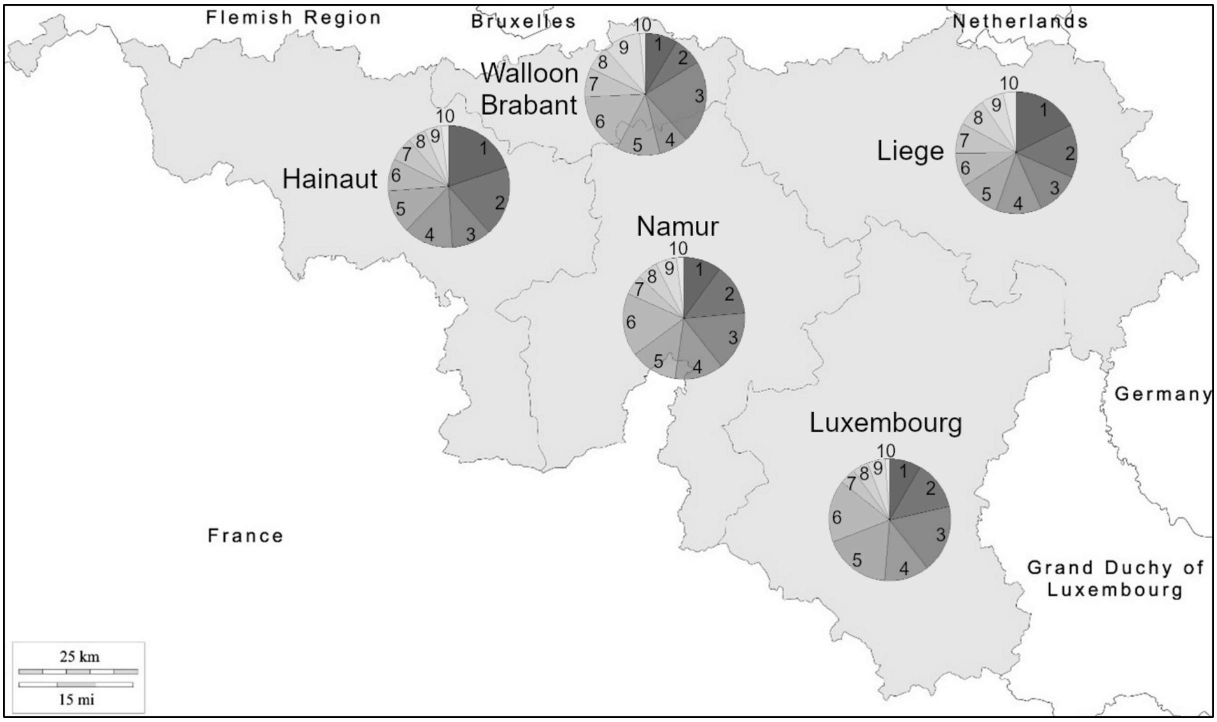

Fig. 5. Distribution of the 10 clusters by province.

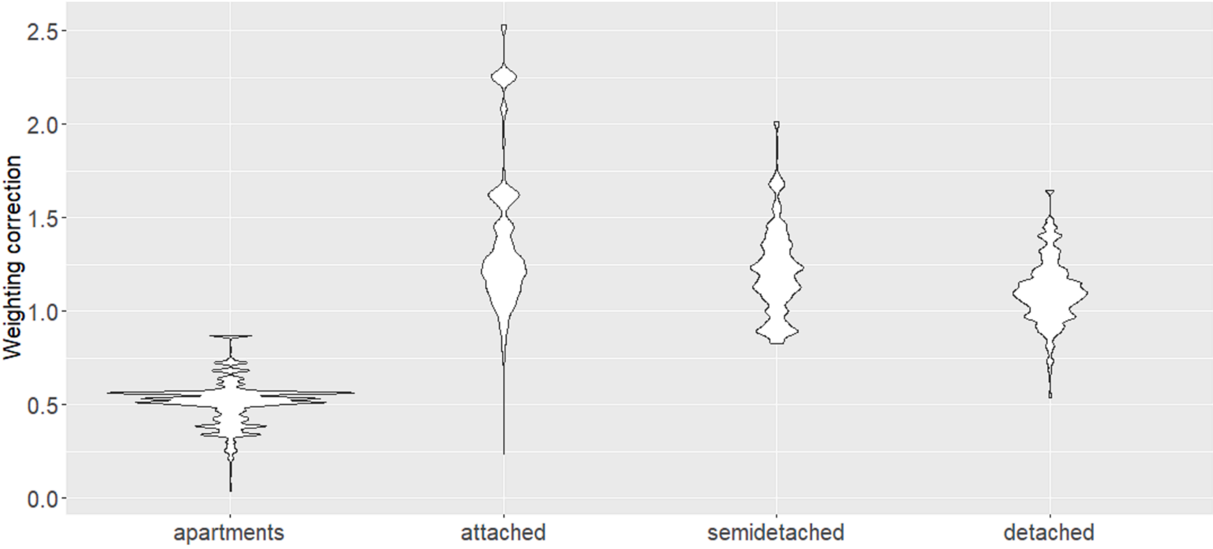

Fig. 6. Distribution of the certificate’s weights depending on the typologies.

Table 8
Certificate and weighted certificate statistics.

	Q 25 %	Median	Mean	Q 75 %
Certificate	253	386	417	534
Weighted certificate	282	409	437	550

would be better represented by clusters 2/6 “mid performance – low income”, 2/10 “mid performance – low income – attached –– renter”, 4/10 “mid performance – low income – semidetached – owner” and 8/10 “mid performance – low income – apartment –– renter”.

More comprehensively, the PCA of the distribution of clusters in Walloon cities presented in Fig. 8 should be interpreted as the difference in the distribution of clusters in a city compared to an ’average’ distribution in the Walloon region. The more peripheral the towns, the more the distribution of the clusters differs from this average distribution. First, it confirms the similarities between the two clusterings. Certain towns with compositions that merit closer articles are enhanced by their exterior positioning. We find the same cities as extreme cases, which means that the characteristics of the two clusters are homogeneous.

Dimensions 1 and 2 cannot be directly interpreted because they are the result of an integration of all the initial dimensions. The percentage on the graph represents the quality of these dimensions to represent the information. Thus, in the graph on the left concerning the cluster with 6 groups, dimension 1 represents 68.9 % of the initial data and dimension 2 represents 26.4 % of the initial data, for a total of 95.3 % of the initial data represented. On the graph on the right concerning the cluster with

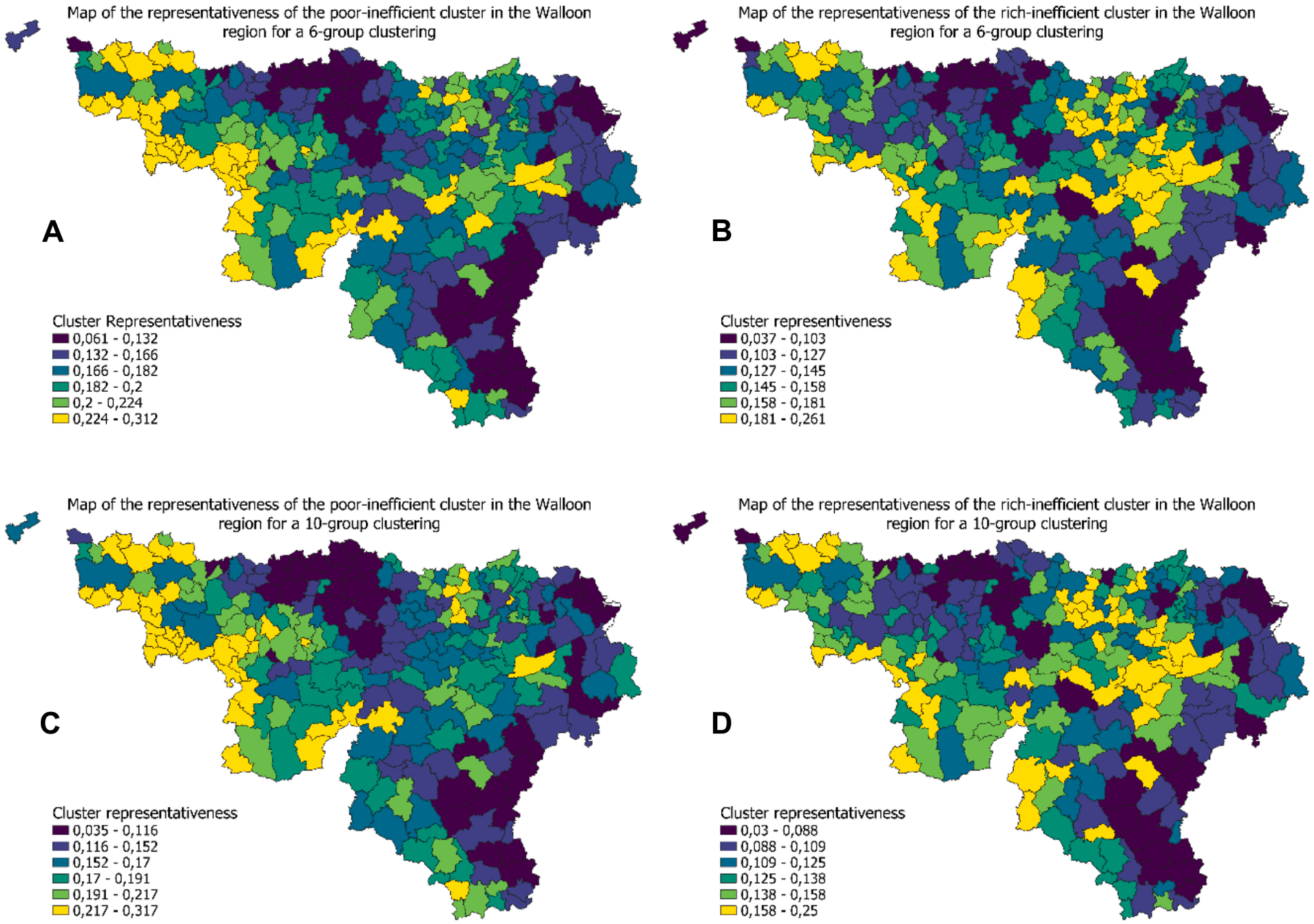
Map of the representativeness of the poor-inefficient cluster in the Walloon region for a 6-group clustering
A
Cluster Representativeness
0,061 - 0,132
0,132 - 0,166
0,166 - 0,182
0,182 - 0,2
0,2 - 0,224
0,224 - 0,312
Map of the representativeness of the rich-inefficient cluster in the Walloon region for a 6-group clustering
B
Cluster representiveness
0,037 - 0,103
0,103 - 0,127
0,127 - 0,145
0,145 - 0,158
0,158 - 0,181
0,181 - 0,261
Map of the representativeness of the poor-inefficient cluster in the Walloon region for a 10-group clustering
C
Cluster representativeness
0,035 - 0,116
0,116 - 0,152
0,152 - 0,17
0,17 - 0,191
0,191 - 0,217
0,217 - 0,317
Map of the representativeness of the rich-inefficient cluster in the Walloon region for a 10-group clustering
D
Cluster representativeness
0,03 - 0,088
0,088 - 0,109
0,109 - 0,125
0,125 - 0,138
0,138 - 0,158
0,158 - 0,25

Fig. 7. Distribution in Walloon municipalities of clusters 3/6 "low performance – low income – owner" (A), 5/6 "low performance – high income – owner" (B), 1/10 "low performance – low income – house – owner"(C) and6/10 "low performance – high income – owner" (D).

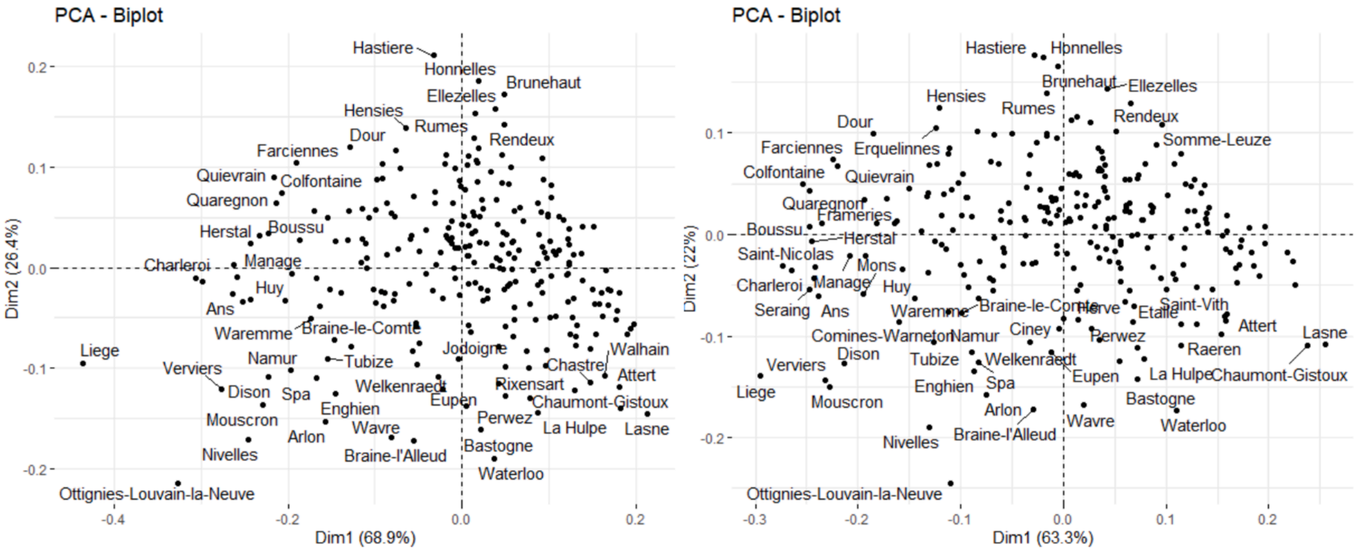

Fig. 8. PCA distribution of the 6 (left) and the 10 clusters (right) in the Walloon region cities.

10 groups, dimension 1 represents 63.6 % of the initial data and dimension 2 represents 22 % of the initial data, for a total of 85.3 %. We can, therefore, consider that these graphs highlight very well the individuals with the most remarkable distribution, even if this quality logically decreases with the increase in the number of initial dimensions. Despite this, we can hypothesize that the horizontal axes put forward information related to the distribution of geometries (more apartments and semi-detached houses on the left, more 4-front houses on the right). In contrast, the vertical axes seem to put forward information related to energy performance (better performances at the bottom). On the

diagonal, we find information related to income, with a higher proportion of high incomes on the bottom right.

What is most interesting is probably to highlight the cities that should rely on a differentiated approach to their building stock. Deux towns clearly have a very specific built stock, which is consistent with their size (Liège) and their history (Ottignies-Louvain-la-Neuve). Some towns, such as several former industrial towns in the Sambre et Meuse region (Charleroi, Ans, Manage, Herstal, Huy, Boussu, etc.), can be included in the same group. Other towns (Arlon, Braine-L'Alleud, Wavre, Waterlo, Bastogne, Enghien, etc.) are characterized by a more recent stock of buildings due to their location on the outskirts of Brussels or Luxembourg. This classification could be used to identify the most effective housing and building policies in similar towns. Policies that have proved successful in a specific city or region could be replicated in cities with a similar distribution of clusters. On the other hand, we must be careful not to generalize too quickly results between two entities with very different distributions.

3.4. Validation

Fig. 9 allows us to check the consistency of the logistic ordinal model's results used for the income regression. It can be noted that the majority of "individual-cities" have a similar weight in linear regression. However, the residuals of this model for the two metropolises of Liège and Charleroi are particularly high, but this is explained by the size of these two cities, which bring together 11 % of the region's households. From a statistical point of view, this check does not invalidate the results of the ordinal logistic model.

The representation of the fitted values according to the generalized leverage (Fig. 10) confirms a relatively homogeneous distribution of the individuals for the beta regression model. Even though some individuals logically have more influence than others, none of them stand out specifically. From a statistical point of view, this check does not invalidate the results of the beta regression model.

The various statistical validation tests carried out as the processing progressed ensured that the statistical regression models could indeed be used.

4. Discussion and conclusions

Previous research has already identified the potential of exploiting existing databases to build a mixed model of the built stock. This article succeeded in combining several databases from different sources and representing varied information (residents and housing). With the help of advanced statistical analysis involving regression and clustering techniques, the article managed to provide useful and insightful knowledge on the correlation between occupancy profiles and housing energy efficiency. The article could thus provide a methodological base on which future studies could rely to compile the multiple databases that exist on the composition of the built stock and the socio-economic characteristics of the population. The article's methodology results in the grouping of homes into several representative groups. Each group has its characteristics and its occupant profile. Thus, renovation policies can be adapted to match the characteristics of each group, particularly socio-economic characteristics. As the state of the art has shown, it is necessary to distinguish better and group the real estate stock to achieve high renovation rates. The analysis confirms what had already been envisaged by Cassilde [15], that we have more EPC certificates for high-performance housing and fewer certificates for low-performance housing. Consequently, the EPC database must be used with caution to characterize and improve the performance of low energy performance buildings, which constitute most of the Belgian and European building stock.

The article also revealed an interesting phenomenon among two groups of housing users. The first group, low-income people, with an ownership rate of 79 %, occupies 17 % of residential buildings in Wallonia. Unfortunately, they occupy the lowest energy performance housing with EPC F and G. This high level of ownership, which can be explained by Belgian housing policy, implies great difficulty in carrying out heavy renovation work. This group must become a priority in all renovation policies. The second group is high-income occupants, with a high property rate of 84 %, living in houses whose specific energy is, on average, 704 kWh/m^2.year. In other words, being rich and owner is not enough to ensure energy-saving work either. This result confirms one of the hypotheses of our study: the inclusion of socio-economic data enhances the understanding of occupant behavior, financial capacity, and decision-making processes related to building renovations.

Furthermore, any renovation strategies should be planned and

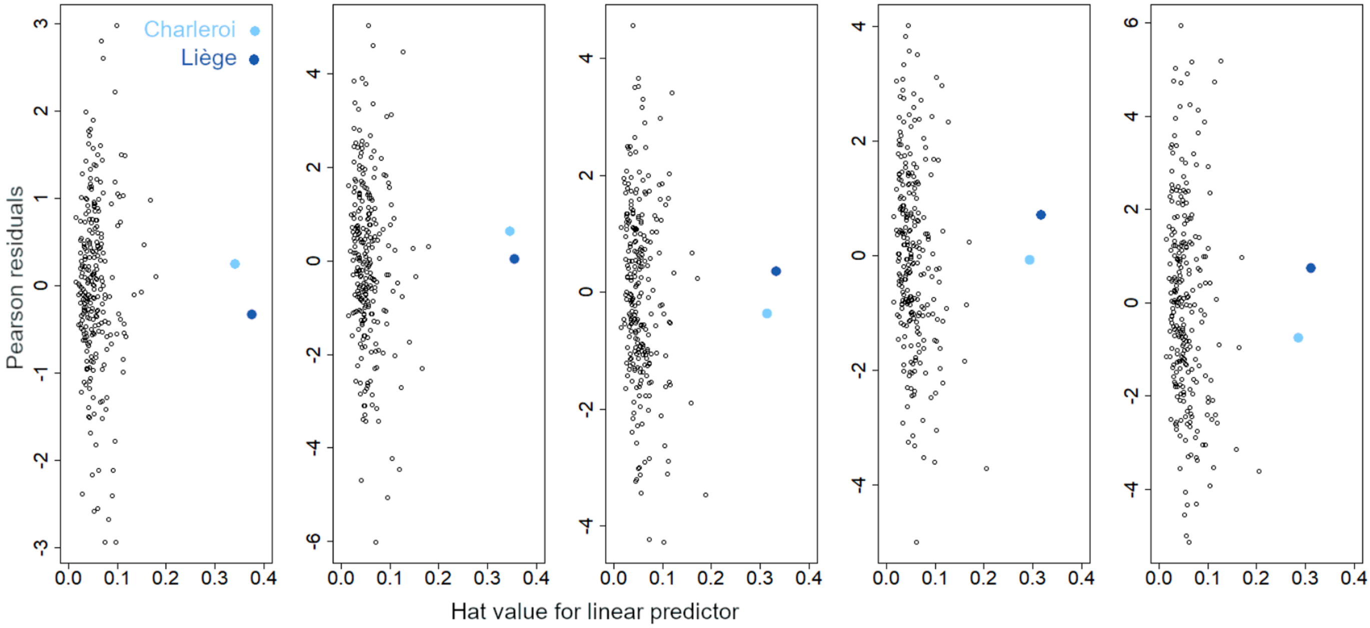

Fig. 9. Pearson residuals depending on Hat values of the income logistic ordinal.

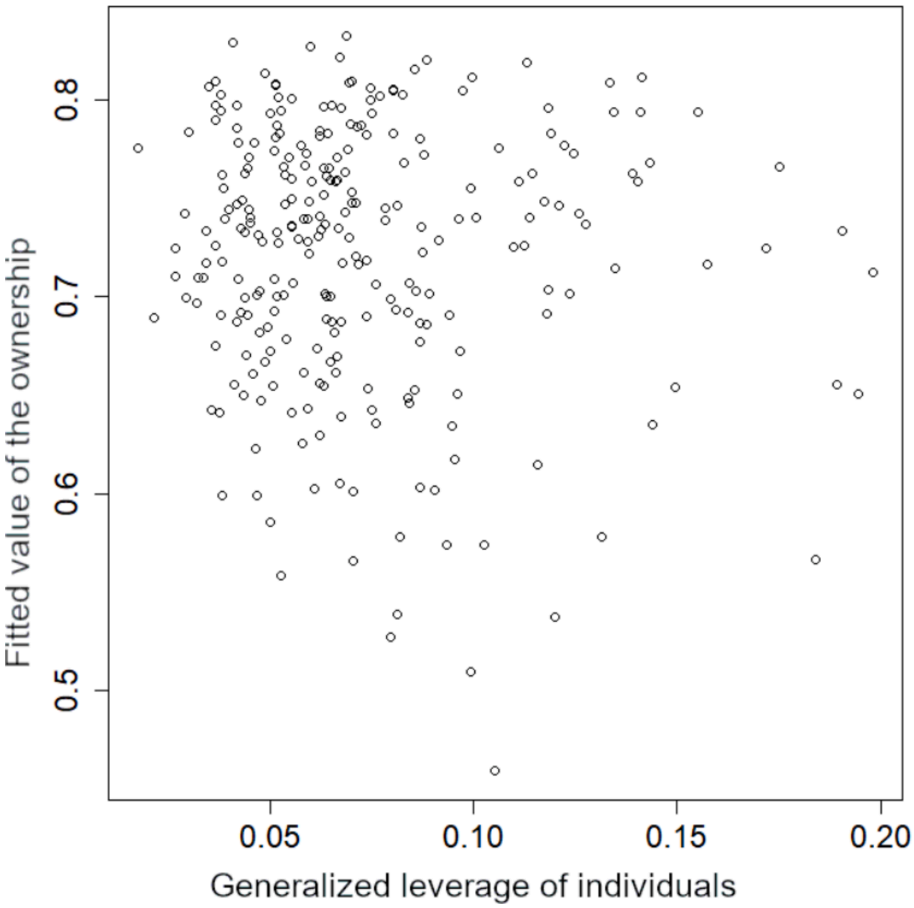

Fig. 10. Fitted values of the ownership beta regression model depending on the generalized leverage.

implemented on the city and municipal level and not only at the regional or national level, to achieve a breakthrough by increasing the renovation rate in the Walloon region. Our article confirms that only customized and tailor-made renovation strategies and plans can become effective. These targeted strategies will help to address the diversity of socio-economic profiles highlighted by the clustering, which is a determining factor in the decision to renovate [57,69,82]. These targeted strategies could be adapted to the heterogeneous distribution of these clusters across the territory highlighted by the PCA [25]. There is also an urgent need to create energy performance certificates for all existing buildings without certification. Without this step, any city or municipality will not have a complete picture of the energy efficiency and clusters of its residential stock. In the current socioeconomical context of Wallonia a fact check confirms the low progress of renovation ineffective subsidies program. The article revealed that the existing incentive program failed to support low-income households with significant renovation measures nor high-income dwellers. Therefore, we suggest revising the policy of support for property ownership for social classes and low incomes. Empowerment via ownership is not enough and must be coupled with proactive energy efficiency measures and incentives, at the risk of putting these households in a vicious circle of energy expenditure. For high-income house owners who live in low energy performance dwellings, a high taxation of greenhouse gas emissions could be an example of future reforms or regulations that should address this category.

The article succeeds in addressing the representativity problem by coupling different databases. This is unique and novel. Until now, similar models (Table 1) were built with homogeneous technical data whose distribution in the built stock is well documented. This method is transferable and reproducible and can be supplemented with additional data sets. This coupling of a data set on the energy efficiency of buildings with a socio-economic data set via regression and clustering techniques makes it possible to create completely new typologies. If we already knew the essential link between these two aspects [16], we now have a tool to put them together to propose appropriate policies. The influence of the variance of the dataset temporality might be significant but does not jeopardize the clustering results. The known limitations of the EPC database are specifically considered. Its representativeness is corrected, possible measurement errors and default values are smoothed by a large-scale analysis, and the introduction of socio-economic characteristics makes it possible to anticipate the real energy behavior of households better. The article requires several steps of simplification to perform statistical modeling and assure hypothetical values for each different calculation method. We consulted statistical experts to ensure that all our assumptions were within acceptable ranges and best practices. However, to increase confidence in this model, it would be interesting to validate the results with an in-situ survey that would confirm the results obtained.

The article's results implicate revising existing strategies in the European Union and elsewhere to make sure that they can be executed operationally on the city or municipal level. Without detailed classification and clustering of different archetypes, generic renovation programs will not be able to target priority households. There is a need to use real case studies in the form of neighborhoods or municipalities to test renovation strategies and implement our suggested clustering approach. Missing information on the EPC of existing buildings in those neighborhoods or municipalities will require targeted audits to complete the full picture of the building stock efficiency states. The impact of the new European carbon tax that will be implemented in 2027 needs to be well-studied. Also, the influence of the Ukrainian-Russian war and associated energy price increases require a detailed article. Between 2022 and 2023, the Belgian government will have paid out about 5 billion euros to protect households and businesses against price increases. It would be interesting to article other alternatives to use the money from this social protection in targeted renovation plans.

CRediT authorship contribution statement

Guirec Ruellan: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Shady Attia:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Gentiane Haesbroeck:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by Liege University, which is gratefully acknowledged. We would like to acknowledge the DG TLPE to give access to EPC data. We express our thanks to Dr Gilles Tihon, prof. Griet Verbeeck and prof. Sigrid Reiter and appreciate their valuable comments and feedback.

Data availability

EPC data is confidential. Other data is public and can be sent by the author on request.

References

[1] C. Ahern, B. Norton, A generalisable bottom-up methodology for deriving a residential stock model from large empirical databases, *Energ. Buildings* 215 (2020) 109886.

[2] J. Albrecht, S. Hamels, The financial barrier for renovation investments towards a carbon neutral building stock – An assessment for the Flemish region in Belgium, *Energ. Buildings* 248 (2021) 111177.

[3] Allacker, K., De Troyer, F., Trigaux, D., Geerken, T., Spirinckx, C., Debacker, W., Van Dessel, J., Janssen, A., Delem, L., & Putzeys, K. (2011). *Sustainability, Financial and Quality evaluation of Dwelling Types-SuFiQuaD-FINAL REPORT*.

[4] Anfrie, M.-N., Coban, E., Hubert, J., Kryvobokov, M., & Pradella, S. (2021). *Chiffres clés du logement en Wallonie—Cinquième édition* (5ème édition; p. 225). Centre d'Études en Habitat Durable de Wallonie.

[5] Ari, M. A., Arregui, M. N., Black, M. S., Celasun, O., Iakova, M. D. M., Mineshima, M. A., Mylonas, V., Parry, I. W. H., Teodoru, I., & Zhunussova, K. (2022). *Surging Energy Prices in Europe in the Aftermath of the War : How to Support the Vulnerable and Speed Up the Transition Away from Fossil Fuels*. International Monetary Fund.

[6] B. Ástmarsson, P.A. Jensen, E. Maslesa, Sustainable renovation of residential buildings and the landlord/tenant dilemma, *Energy Policy* 63 (2013) 355–362, https://doi.org/10.1016/j.enpol.2013.08.046.

[7] S. Attia, Spatial and Behavioral Thermal Adaptation in Net Zero Energy Buildings : An Exploratory Investigation, *Sustainability* 12 (19) (2020).

[8] S. Attia, T. Canonge, M. Popineau, M. Cuchet, Developing a benchmark model for renovated, nearly zero-energy, terraced dwellings, *Appl. Energy* 306 (2022) 118128.

[9] S. Azizi, G. Nair, T. Olofsson, Analysing the house-owners' perceptions on benefits and barriers of energy renovation in Swedish single-family houses, *Energ. Buildings* 198 (2019) 187–196.

[10] F. Bartiaux, C. Vandeschrick, M. Moezzi, N. Frogneux, Energy justice, unequal access to affordable warmth, and capability deprivation : A quantitative analysis for Belgium, *Appl. Energy* 225 (2018) 1219–1233.

[11] Berkhout, P. H. G., Muskens, J. C., & W. Velthuijsen, J. (2000). Defining the rebound effect. *Energy Policy*, *28*(6), 425-432.

[12] P. Bertoldi, M. Economidou, V. Palermo, B. Boza-Kiss, V. Todeschi, How to finance energy renovation of residential buildings : Review of current and emerging financing instruments in the EU, *Wires Energy Environ.* 10 (1) (2021) e384.

[13] S. Bird, D. Hernández, Policy options for the split incentive : Increasing energy efficiency for low-income renters, *Energy Policy* 48 (2012) 506–514.

[14] N. Campbell, L. Ryan, V. Rozite, E. Lees, G. Heffner, *Capturing the Multiple Benefits of Energy Efficiency*, International Energy Agency, 2014.

[15] Cassilde, S. (2017). *Analyse de la base de données des certificats PEB en Wallonie* [Report]. Centre d'Etudes en Habitat Durable.

[16] Cehd, *Enquête sur la Qualité de l'Habitat en Wallonie 2012-2013—Résultats clés*, CEHD. (2014).

[17] K. Čermáková, E. Hromada, Change in the Affordability of Owner-Occupied Housing in the Context of Rising Energy Prices, *Energies* 15 (4) (2022).

[18] L. Coppens, M. Gargiulo, M. Orsini, N. Arnould, Achieving −55% GHG emissions in 2030 in Wallonia, Belgium : Insights from the TIMES-Wal energy system model, *Energy Policy* 164 (2022) 112871.

[19] Creg, *Neuvième rapport de monitoring concernant l'extension de l'application des tarifs sociaux électricité et gaz naturel aux bénéficiaires de l'intervention majorée* ((RA)2556, Commission de Régulation de l'Électricité et du Gaz, CREG, 2023, p. 16.

[20] F. Cribari-Neto, A. Zeileis, Beta Regression in R, *J. Stat. Softw.* 34 (2010) 1–24.

[21] CWAPE. (2023). Analyse des prix de l'électricité et du gaz naturel en Wallonie (clients résidentiels) sur la période de janvier 2007 à décembre 2022. (CD-23b23-CWaPE-0113). CWAPE.

[22] Cyx, W., Renders, N., Van Holm, M., & Verbeke, S. (2011). *IEE TABULA - Typology Approach for Building Stock Energy Assessment* (p. 81). Flemish Institute for Technological Research.

[23] M. Delghust, W. Roelens, T. Tanghe, Y. De Weerdt, A. Janssens, Regulatory energy calculations versus real energy use in high-performance houses, *Build. Res. Inf.* 43 (6) (2015) 675–690.

[24] Directive 2002/91/CE sur le performance énergétique des bâtiments (2002).

[25] Dom'enech-Arum, G., Gobbi, P. E., & Magerman, G. (2022). *Housing inequality and how fiscal policy shapes it : Evidence from Belgian real estate*. National Bank of Belgium.

[26] M. Dowson, A. Poole, D. Harrison, G. Susman, Domestic UK retrofit challenge : Barriers, incentives and current performance leading into the Green Deal, *Energy Policy* 50 (2012) 294–305.

[27] D'Oca, S., Ferrante, A., Ferrer, C., Pernetti, R., Gralka, A., Sebastian, R., & Op 't Veld, P. (2018). Technical, Financial, and Social Barriers and Challenges in Deep Building Renovation : Integration of Lessons Learned from the H2020 Cluster Projects. *Buildings*, *8*(12), Article 12.

[28] J.J. Egozcue, V. Pawlowsky-Glahn, G. Mateu-Figueras, C. Barceló-Vidal, Isometric Logratio Transformations for Compositional Data Analysis, *Math. Geol.* 35 (3) (2003) 279–300.

[29] Embuild. (2023, décembre 18). *Conférence de presse*.

[30] A. Erkan, Z. Yildiz, Parallel lines assumption in ordinal logistic regression and analysis approaches, *International Interdisciplinary Journal of Scientific Research* 1 (3) (2014) 8–23.

[31] European Commission. (2020, octobre 14). A Renovation Wave for Europe—Greening our buildings, creating jobs, improving lives.

[32] European Commission. (2022, mai 18). *REPowerEU Plan*.

[33] Feist, W. (2012). EnerPHit and EnerPHit+ i. Certification criteria for energy retrofits with passive house components. *Darmstadt: Passivhaus Institute. Retrieved November*, *1*, 2012.

[34] F. Filippidou, N. Nieboer, H. Visscher, Are we moving fast enough? The energy renovation rate of the Dutch non-profit housing using the national energy labelling database, *Energy Policy* 109 (2017) 488–498.

[35] P. Florio, O. Teissier, Estimation of the Energy Performance Certificate of a housing stock characterised via qualitative variables through a typology-based approach model : A fuel poverty evaluation tool, *Energ. Buildings* 89 (2015) 39–48.

[36] N. Galiotto, P. Heiselberg, M.-A. Knudstrup, Integrated Renovation Process : Overcoming Barriers to Sustainable Renovation, *J. Archit. Eng.* 22 (1) (2016).

[37] R. Galvin, Why German homeowners are reluctant to retrofit, *Build. Res. Inf.* 42 (4) (2014) 398–408.

[38] R. Galvin, Making the 'rebound effect'more useful for performance evaluation of thermal retrofits of existing homes : Defining the 'energy savings deficit'and the 'energy performance gap', *Energ. Buildings* 69 (2014) 515–524.

[39] S. Gendebien, E. Georges, S. Bertagnolio, V. Lemort, Methodology to characterize a residential building stock using a bottom-up approach : A case study applied to Belgium, *International Journal of Sustainable Energy Planning and Management* 4 (2015) 71–88.

[40] O. Guerra Santin, L. Itard, H. Visscher, The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock, *Energ. Buildings* 41 (11) (2009) 1223–1232.

[41] Hauglustaine, J.-M., & Monfils, S. (2013). Réno2020 : Etude énergétique et typologique du parc résidentiel wallon en vue d'en dégager des pistes de rénovation prioritaires.

[42] C. He, Y. Hou, L. Ding, P. Li, Visualized literature review on sustainable building renovation, *Journal of Building Engineering* 44 (2021) 102622.

[43] H. Hens, W. Parijs, M. Deurinck, Energy consumption for heating and rebound effects, *Energ. Buildings* 42 (1) (2010) 105–110.

[44] H. Hens, G. Verbeeck, B. Verdonck, Impact of energy efficiency measures on the CO 2 emissions in the residential sector, a large scale analysis, *Energ. Buildings* 33 (3) (2001) 275–281.

[45] J. Hindriks, V. Serse, The incidence of VAT reforms in electricity markets : Evidence from Belgium, *Int. J. Ind Organiz* 80 (2022) 102809.

[46] K. Hron, P. Filzmoser, K. Thompson, Linear regression with compositional explanatory variables, *J. Appl. Stat.* 39 (5) (2012) 1115–1128, https://doi.org/10.1080/02664763.2011.644268.

[47] Hubert, J., Anfrie, M.-N., Kryvobokov, M., & Pradella, S. (2021). *Performance énergétique du parc de bâtiments résidentiels en Wallonie—Édition 2021* (Rapport du Centre d'Etudes en Habitat Durable de Wallonie, p. 150). Centre d'Études en Habitat Durable de Wallonie.

[48] A.M. Ikotun, A.E. Ezugwu, L. Abualigah, B. Abuhaija, J. Heming, K-means clustering algorithms : A comprehensive review, variants analysis, and advances in the era of big data, *Inf. Sci.* 622 (2023) 178–210.

[49] P.A. Jensen, E. Maslesa, J. Brinkø Berg, Sustainable Building Renovation : Proposals for a Research Agenda, *Sustainability* 10 (12) (2018).

[50] P.A. Jensen, L. Thuvander, P. Femenias, H. Visscher, Sustainable building renovation – strategies and processes, *Constr. Manag. Econ.* 40 (3) (2022) 157–160.

[51] I.T. Jolliffe, J. Cadima, Principal component analysis : A review and recent developments, *Philos. Trans. r. Soc. A Math. Phys. Eng. Sci.* 374 (2065) (2016).

[52] B. Kaveh, M.U. Mazhar, B. Simmonite, M. Sarshar, B. Sertyesilisik, An investigation into retrofitting the pre-1919 owner-occupied UK housing stock to reduce carbon emissions, *Energ. Buildings* 176 (2018) 33–44.

[53] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, M. Djurovic-Petrovic, A review of bottom-up building stock models for energy consumption in the residential sector, *Build. Environ.* 45 (7) (2010) 1683–1697.

[54] C. Kints, A. De Herde, La rénovation énergétique et durable des logements wallons : Analyse du bâti existant et mise en évidence de typologies de logements prioritaires, *Architecture & Climat.* (2008).

[55] T.M. Kodinariya, P.R. Makwana, Review on determining number of Cluster in K-Means Clustering, *Int. J.* 1 (6) (2013) 90–95.

[56] J. Kragh, K.B. Wittchen, Development of two Danish building typologies for residential buildings, *Energ. Buildings* 68 (2014) 79–86.

[57] W. Kuckshinrichs, T. Kronenberg, P. Hansen, The social return on investment in the energy efficiency of buildings in Germany, *Energy Policy* 38 (8) (2010) 4317–4329.

[58] M. Lanau, G. Liu, U. Kral, D. Wiedenhofer, E. Keijzer, C. Yu, C. Ehlert, Taking Stock of Built Environment Stock Studies : Progress and Prospects, *Environ. Sci. Tech.* 53 (15) (2019) 8499–8515.

[59] Z. Lejeune, G. Xhignesse, M. Kryvobokov, J. Teller, Housing quality as environmental inequality : The case of Wallonia, Belgium, *J. Hous. Built Environ.* 31 (3) (2016) 495–512.

[60] Leveau, C., Cherdon, M., Vismara, M., & Huberlant, B. (2018). *Détermination du niveau de performance énergétique optimal des bâtiments en fonction des coûts : Etude Cost optimum PER-PEN 2017*.

[61] Lonergan, K., Gabrielli, P., & Sansavini, G. (2022). *Energy justice analysis of the European Commission REPowerEU plan* [Working Paper]. ETH Zurich.

[62] M. Mangold, M. Österbring, C. Overland, T. Johansson, H. Wallbaum, Building Ownership, Renovation Investments, and Energy Performance—A Study of Multi-Family Dwellings in Gothenburg, *Sustainability* 10 (5) (2018).

[63] M. Mangold, M. Österbring, H. Wallbaum, L. Thuvander, P. Femenias, Socio-economic impact of renovation and energy retrofitting of the Gothenburg building stock, *Energ. Buildings* 123 (2016) 41–49.

[64] A.-F. Marique, S. Reiter, Retrofitting the suburbs : Insulation, density, urban form and location, *Environmental Management and Sustainable Development* 3 (2) (2014).

[65] A.-F. Marique, B. Rossi, Cradle-to-grave life-cycle assessment within the built environment : Comparison between the refurbishment and the complete reconstruction of an office building in Belgium, *J. Environ. Manage.* 224 (2018) 396–405.

[66] Meyer, S., & Anastasia, J. (2019). Census 2011 Logement Synthèse des principaux résultats pour les 4 grands centres urbains wallons. *Rapport de recherche du projet Energ-Ethic*.

[67] Meyer, S., & Coene, J. (2023). *Baromètre de la précarité énergétique (2023)* (9). Fondation Roi Baudoin.

[68] S. Meyer, H. Laurence, D. Bart, L. Middlemiss, K. Maréchal, Capturing the multifaceted nature of energy poverty : Lessons from Belgium, *Energy Res. Soc. Sci.* 40 (2018) 273–283.

[69] K. Mjörnell, P. Femenías, K. Annadotter, Renovation Strategies for Multi-Residential Buildings from the Record Years in Sweden—Profit-Driven or Socio-economically Responsible? *Sustainability* 11 (24) (2019).

[70] E. Mlecnik, W. Hilderson, J. Cre, I. Desmidt, V.D. Uyttebroeck, S. Abeele, A. Van Quathem, L. Vandaele, L. Delem, F. Dobbels, O. Lesage, S. Prieus, P. Van Den Bossche, J. Vrijders, A. De Herde, A. Branders, J. Desmedt, T. De Meester, C. Kints, O. Henz, *Low energy housing retrofit (LEHR), final report*. Belgian Science, Policy (2010).

[71] Mouton, C., De Meyer, A., & Feldheim, V. (2013). *COZEB : Rapport final du projet*.

[72] C. Nägeli, M. Jakob, G. Catenazzi, Y. Ostermeyer, Towards agent-based building stock modeling : Bottom-up modeling of long-term stock dynamics affecting the energy and climate impact of building stocks, *Energ. Buildings* 211 (2020) 109763.

[73] A.N. Nielsen, R.L. Jensen, T.S. Larsen, S.B. Nissen, Early stage decision support for sustainable building renovation – A review, *Build. Environ.* 103 (2016) 165–181.

[74] M. Otavova, C. Faes, C. Bouland, E. De Clercq, B. Vandeninden, T. Eggerickx, J.-P. Sanderson, B. Devleesschauwer, B. Masquelier, Inequalities in mortality associated with housing conditions in Belgium between 1991 and 2020, *BMC Public Health* 22 (1) (2022) 2397.

[75] A.M. Papadopoulos, T.G. Theodosiou, K.D. Karatzas, Feasibility of energy saving renovation measures in urban buildings : The impact of energy prices and the acceptable pay back time criterion, *Energ. Buildings* 34 (5) (2002) 455–466.

[76] C. Peñasco, L.D. Anadón, Assessing the effectiveness of energy efficiency measures in the residential sector gas consumption through dynamic treatment effects : Evidence from England and Wales, *Energy Econ.* 117 (2023) 106435.

[77] C. Protopapadaki, G. Reynders, D. Saelens, Bottom-up modelling of the Belgian residential building stock : Impact of building stock descriptions. *Proceedings of the 9th International Conference on System Simulation in Buildings-SSB2014*, 2014.

[78] K. Qu, X. Chen, Y. Wang, J. Calautit, S. Riffat, X. Cui, Comprehensive energy, economic and thermal comfort assessments for the passive energy retrofit of historical buildings—A case study of a late nineteenth-century Victorian house renovation in the UK, *Energy* 220 (2021) 119646.

[79] T.G. Reames, Targeting energy justice : Exploring spatial, racial/ethnic and socio-economic disparities in urban residential heating energy efficiency, *Energy Policy* 97 (2016) 549–558.

[80] P. Reusens, F. Vastmans, S. Damen, The impact of changes in dwelling characteristics and housing preferences on Belgian house prices, *Econ. Rev.* (2022) 1–40.

[81] A.V. Rocha, A.B. Simas, Influence diagnostics in a general class of beta regression models, *TEST* 20 (1) (2011) 95–119.

[82] G. Ruellan, *Interviews sur la rénovation du stock bâti en Belgique*, Uliège. (2016).

[83] G. Ruellan, *La rénovation du bâti résidentiel en Belgique*, Uliège. (2016).

[84] G. Ruellan, M. Cools, S. Attia, Analysis of the Determining Factors for the Renovation of the Walloon Residential Building Stock, *Sustainability* 13 (4) (2021).

[85] M. Santamouris, K. Kapsis, D. Korres, I. Livada, C. Pavlou, M.N. Assimakopoulos, On the relation between the energy and social characteristics of the residential sector, *Energ. Buildings* 39 (8) (2007) 893–905.

[86] M. Santamouris, G. Mihalakakou, P. Patargias, N. Gaitani, K. Sfakianaki, M. Papaglastra, C. Pavlou, P. Doukas, E. Primikiri, V. Geros, M. N. Assimakopoulos, R. Mitoula, S. Zerefos, Using intelligent clustering techniques to classify the energy performance of school buildings, *Energ. Buildings* 39 (1) (2007) 45–51.

[87] I. Sartori, N.H. Sandberg, H. Brattebø, Dynamic building stock modelling : General algorithm and exemplification for Norway, *Energ. Buildings* 132 (2016) 13–25.

[88] S.C. Scott, M.S. Goldberg, N.E. Mayo, Statistical assessment of ordinal outcomes in comparative studies, *J. Clin. Epidemiol.* 50 (1) (1997) 45–55.

[89] A. Serrano-Jiménez, P. Femenías, L. Thuvander, Á. Barrios-Padura, A multi-criteria decision support method towards selecting feasible and sustainable housing renovation strategies, *J. Clean. Prod.* 278 (2021) 123588.

[90] SPW. (2022). Arrêté du Gouvernement wallon instaurant un régime d'aides accordées pour la réalisation d'investissements économiseurs d'énergie et de rénovation d'un logement.

[91] K.N. Streicher, P. Padey, D. Parra, M.C. Bürer, M.K. Patel, Assessment of the current thermal performance level of the Swiss residential building stock : Statistical analysis of energy performance certificates, *Energ. Buildings* 178 (2018) 360–378.

[92] M. Sunikka-Blank, R. Galvin, Introducing the prebound effect : The gap between performance and actual energy consumption, *Build. Res. Inf.* 40 (3) (2012) 260–273.

[93] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector : A review of modeling techniques, *Renew. Sustain. Energy Rev.* 13 (8) (2009) 1819–1835.

[94] L. Thuvander, P. Femenías, K. Mjörnell, P. Meiling, Unveiling the Process of Sustainable Renovation, *Sustainability* 4 (6) (2012).

[95] Valente, P. (2014). Innovative Approaches to Census-Taking : Overview of the 2011 Census Round in Europe. In F. Crescenzi & S. Mignani (Éds.), *Statistical Methods and Applications from a Historical Perspective : Selected Issues* (p. 187-200). Springer International Publishing.

[96] R. Van Dam, V. Geurts, I. Pannecoucke, Housing tenure, housing costs and poverty in Flanders (Belgium), *J. Hous. Built Environ.* 18 (1) (2003) 1–23.

[97] Vanneste, D., Thomas, I., Goossens, L., & others. (2007). *Le logement en Belgique* (p. 211). Direction générale Statistique et Information économique.

[98] D. Vanneste, I. Thomas, L. Vanderstraeten, The spatial structure(s) of the Belgian housing stock, *J. Hous. Built Environ.* 23 (3) (2008) 173–198.

[99] G. Verbeeck, H. Hens, Energy savings in retrofitted dwellings : Economically viable? *Energ. Buildings* 37 (7) (2005) 747–754.

[100] J. Vivian, L. Carnieletto, M. Cover, M. De Carli, At the roots of the energy performance gap : Analysis of monitored indoor air before and after building retrofits, *Build. Environ.* 110914 (2023).

[101] T. Walter, M.D. Sohn, A regression-based approach to estimating retrofit savings using the Building Performance Database, *Appl. Energy* 179 (2016) 996–1005.

[102] S. Welch, E. Obonyo, A.M. Memari, A review of the previous and current challenges of passive house retrofits, *Build. Environ.* 245 (2023) 110938.