



HAL
open science

Towards a classification of the building stock in terms of potential energy savings through advanced control

Alexis Wagner, Marie Ruellan, Matthias Heinrich, Romain Bourdais

► To cite this version:

Alexis Wagner, Marie Ruellan, Matthias Heinrich, Romain Bourdais. Towards a classification of the building stock in terms of potential energy savings through advanced control. *Energy and Buildings*, 2024, 310, pp.114050. 10.1016/j.enbuild.2024.114050 . hal-04500720

HAL Id: hal-04500720

<https://hal.science/hal-04500720v1>

Submitted on 12 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journal Pre-proofs

Towards a classification of the building stock in terms of potential energy savings through advanced control

Alexis Wagner, Marie Ruellan, Matthias Heinrich, Romain Bourdais

PII: S0378-7788(24)00166-X
DOI: <https://doi.org/10.1016/j.enbuild.2024.114050>
Reference: ENB 114050

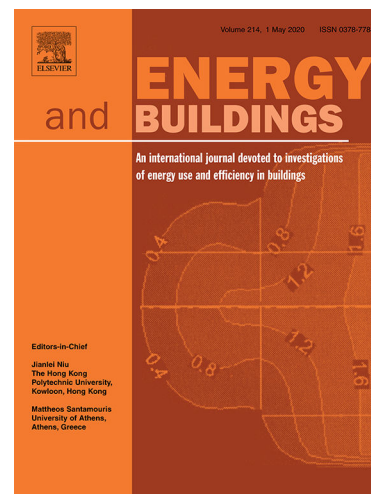
To appear in: *Energy & Buildings*

Received Date: 10 October 2023
Revised Date: 30 January 2024
Accepted Date: 3 March 2024

Please cite this article as: A. Wagner, M. Ruellan, M. Heinrich, R. Bourdais, Towards a classification of the building stock in terms of potential energy savings through advanced control, *Energy & Buildings* (2024), doi: <https://doi.org/10.1016/j.enbuild.2024.114050>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier B.V.



Towards a classification of the building stock in terms of potential energy savings through advanced control

Alexis Wagner ^{a*}, Marie Ruellan ^b, Matthias Heinrich ^c, Romain Bourdais ^a

^a Institut d'Electronique et des Technologies du numérique, Centrale Supélec, Avenue De Boulaie, Cesson-Sévigné, 35700, France

^b SATIE Laboratory, CY Cergy Paris University, Cergy, 95000, France

^c Laboratoire Techniques, Territoires et Sociétés, Ecole des Ponts ParisTech, Avenue Blaise Pascal, Marne-la-Vallée, 77455, France

* corresponding author: alexis.wagner@centralesupelec.fr

ABSTRACT

Developing advanced control strategies is a relevant pathway to achieve energy savings in heating, ventilation, and air conditioning systems. Nevertheless, the studies presenting these controls usually quantify the potential energy savings by analysing a single housing configuration: a thermal envelope, a behavioural scenario and a type of external conditions. By doing this, the study omits that these savings highly depend on the material and behavioural environments. Sensitivity studies are helpful but insufficient to characterize the vast variability of housing situations. This research proposes a methodology that describes the relevance of automation in the living spaces in which they are installed and thus improves the social and technical relevance of the abundant literature on this topic. The methodology is divided into three parts. First, a database of representative housing situations is constructed. Second, the theoretical energy savings through advanced control are estimated using thermal simulation. Third, the database on energy savings and housing situation is segmented through a factorial analysis. This clustering is then used to identify the housing situations where the implementation of controls is most favourable. The article presents first the methodology and then an example of implementation. The computation is based on the TABULA database and studies the use of predictive control of heating systems in French housing situations. The results underline that the housing type, the insulation level, and the surface are significant determinants of the automation interest. The large segment of estimated average gains (between 1.8% and 17% of the reference energy consumption) underlines the high need to improve the targeting of the field.

KEYWORDS: advanced control, HVAC systems, methodology, principal component analysis, clustering,

1. Introduction

1.1. Context

In accordance with the national strategy for greenhouse gas emissions reduction, France is committed to becoming carbon neutral by 2050. From this perspective, the building sector offers particularly promising prospects for energy saving and emission reduction, since they account for around 20% of national GHG emissions and 40% of national primary energy consumption (MTES,2015). The residential housing stock in France is still predominantly made of buildings that have not yet been renovated and the annual renovation rate for the building stock is still well below the target recommended by the Energy Performance of Buildings Directive (EPBD) in order to achieve the 2050 neutrality objectives (1.7% of the stock compared with the recommended 3%). A supplementary solution may be the addition of suitable energy management systems that provide energy savings and, more specifically, the management of heating, ventilation, and air-conditioning (HVAC) systems, which account for almost 40% of a home's average energy consumption.

1.2. Advanced control of HVAC systems in buildings

The literature provides a wide range of strategies in order to manage HVAC systems [1]. They all aim to meet the comfort needs of users (indoor air quality, thermal comfort) while offering the best performance from an energy point of view. These two objectives are contradictory, and improving one criterion without degrading the other is a complex task [2]. The most common algorithms are hysteresis (ON/OFF) and PID control [3]. Although effective in maintaining the desired thermal comfort, these controls were designed without explicit energy criteria, unlike more recent advanced control strategies. Among them, the most popular strategy is predictive control. It utilizes models that, over time, incorporate the evolution of disturbances (such as external temperature, solar gains, and occupancy) and controlled variables (such as indoor air temperature) [4]. At each time step, an optimization problem is solved

in order to determine the control sequence that minimizes a cost function defined over this future horizon, while considering a set of process constraints. The first element of this sequence is then sent to the controlled system, and the process is iterated.

Following this principle, numerous energy management strategies have been developed in the field of building management, encompassing a diverse array of advanced control approaches. These approaches vary in several aspects, including the methods used to predict the evolution of indoor air temperature and system disturbances. Among the most recent observed works, there are control approaches based on the use of white-box, pseudo-physical models [5], as well as models relying on machine learning methods [6], [7]. Other studies have explored direct HVAC system control based on collected data from sensors [8]. Hybrid control approaches have also emerged, including fuzzy cooperative predictive control [9] or adaptive predictive control [10], which aim to address specific limitations of predictive control. Furthermore, the selection of criteria for management strategies is a distinguishing characteristic of advanced control methods. The authors of [11] evaluate the influence of environmental factors and occupant thermal comfort based on the minimized criterion, such as emitted CO₂, final energy consumption, or energy tariffs. Other studies consider room-by-room occupancy of buildings occupant, energy storage, and energy prices for multi-zone air temperature [12]. Furthermore, other research work has focused on control approaches aimed at minimizing energy expenses by considering energy prices or photovoltaic production [13].

The previous works assessed the performance of their control approaches using two primary metrics: the achievement of the desired thermal comfort for occupants and the reduction of energy consumption associated with the implemented energy strategy. These metrics can be evaluated through implementation on a real building or through simulation.

To our knowledge, one of the first field implementation of predictive control for energy management in a building was carried out during the winter period of 2009 in a building at the Czech Technical University [14]. This experiment demonstrated promising energy savings estimated at 40% compared to the previous HVAC

management control. Since then, more recent works have implemented predictive controls on real buildings, showing observed savings ranging from 17% to 58% [6], [15], [16], [17].

While in-situ experiments provide valuable practical insights, most studies emphasize the benefits of control strategies through thermal simulation [5], [10], [11], [13], [18], [19], [20], [21]. These works demonstrate energy savings ranging from 0.2% to 40%.

What emerges from this literature review, aside from the wide diversity of approaches, is that the associated energy savings are estimated for a single housing situation. However, these gains depend on the characteristics of the building (geometry, surface area, insulation), the climate, and the behaviours. Sensitivity studies on the thermal envelope [22], occupancy [12], [23], and outdoor temperature [24] have shown the dependency of gains on the observed housing situation. In the meantime, the extreme diversity of dwellings, climates, and behaviours on a great large scale questions the reliability of sensitivity studies to enable us to characterise the gains.

This leads us to raise the following question: **How can we quantify the energy-saving potential of a heating management strategy at the housing stock level?**

1.3. Assessment of the energy-saving potential of renovation measures on a housing stock scale.

In France, the housing stock is composed of millions of buildings, subject to a high variety of climates, materials, and housing types.

In the scientific literature, several studies aim to quantify the energy-saving potential of technical solutions for a large-scale assessment of environmental impact reduction in the building sector. The significant differences among these works primarily lie in the choice of the energy consumption prediction model. Three types of models stand out: data-driven models, physical models, and hybrid models [25]. The first and third categories cannot be employed to address the issue identified in the previous subsection. Data-driven models are still specialised: they provide reasonable predictions for a given set of dwellings. Furthermore, as far as we know, no data-driven models are capable of simulating a diverse range of dwellings with good accuracy.

Physics-based models, on the other hand, require simulating thermally both before and after renovation for each case study to estimate its energy potential. To address the challenge of studying the entire housing stock, several studies have incorporated methodologies based on archetypes [25], [26], [27], [28]. Indeed, archetypes represent virtual buildings, households, or climates that are representative of homogeneous segments. The assumption is made that the study of an archetype provides similar results on each building, household, or climate included in its segment. Therefore, only the archetypes are simulated. The authors of [25] integrated behavioural household archetypes to assess the impact of energy renovation techniques. The authors concluded that the cost-effectiveness of technical solutions depends on household behaviour. Geographical zone archetypes have also been integrated into large-scale thermal studies, synthesizing the entire area by examining four climatic time series (external temperature, solar gains) from examples of cities [26].

Additionally, twenty-four archetypes of the Greek housing stock have been used to generate scenarios quantifying heating energy demand at the scale of the housing stock [27]. Thus, the impact of renovating the housing stock has allowed the creation of an ambitious scenario assessing the evolution of consumption if the entire housing stock is renovated.

However, the parameters of the thermal models for archetypes are virtual and represent the average of the studied segment of situations. Prediction errors in consumption may occur. To compensate for these errors, it is necessary to use many archetypes. The authors of [28] proposed a methodology based on the housing archetypes from the TABULA project [29] to assess consumption following the renovation

of a medium-sized city. They achieved results close to an approach combining statistical analysis and building-level thermal simulation, highlighting the potential of a based archetypal methodology. The authors of [30] combined, based on the degree-hour method, representative housing archetypes of the Albanian housing stock with climate archetypes to propose a methodology optimizing the thickness of insulation materials within the scope of renovation measures. However, they do not incorporate variability in building usage.

Previous studies primarily target public decision-makers, urban planners, and industrial stakeholders by providing decision-support tools for renovation. These tools can be categorized into several categories: generating consumption scenarios on a large scale for the housing stock, evaluating the profitability of renovation measures, providing sizing tools, or quantifying the impact of renovation on a territorial scale. To our knowledge, these studies need to incorporate advanced management algorithms. Furthermore, no works combine climate archetypes and building archetypes for diverse building uses.

As mentioned, the energy gains from HVCA control depend on housing, climate and behaviour. Drawing on the archetype-based approaches presented in this section we propose to use a segmentation of housing situations which are defined as a combination of a dwelling, a household, and climate conditions [31].

With this paper review, the question posed in paragraph 1.2 is completed: **Which housing situations, are most favourable or unfavourable for the implementation of these strategies?**

1.4. Objective and structure of this work

Therefore, this work aims to propose a methodology for exploring the savings potential of an advanced energy control algorithm for a representative number of housing situations.

This article is structured as follows. In the first section, we propose a methodology based on classification works to construct a representative panel of housing situations defined by the combination of housing archetypes, climate archetypes, and diversity of household behaviour. The novelty of this classification is that it incorporates the energy-saving potentials of advanced control obtained through simulation to describe each situation. The second section presents an implementation of this methodology to single-family houses, considering model predictive control with explicit consideration of occupancy intermittency as the heating management algorithm, thus highlighting the potential of this methodology. For this implementation, results are presented and discussed in Section 4. The final section concludes and offers future perspectives.

2. Methodology

In this first part, we introduce a methodology in order to compute a robust estimation of the energy-saving potential of a heating management algorithm. The main steps are presented in the figure 1. This methodology is divided into two parts.

First, a database of housing situations is constructed by associating four types of variables. The housing situations are generated by the combination of three types of variables: a typology of residential buildings, a classification of outdoor conditions, and behavioural scenarios. These inputs are based on the results of existing classification works. The fourth type of variable is the energy-saving potential which is estimated through thermal simulation for each housing situation of the database.

The second step consists of analysing the results of the database of housing situations. The computation is done using Principal Component Analysis (PCA). The results obtained from this algorithm feed into an agglomerative hierarchical clustering algorithm. The objective of this clustering is to introduce a classification that groups together housing situations in which the implementation of a management algorithm offers similar energy-saving potential.

2.1. Construction of the database of housing situations

2.1.1. Inputs

Three key factors have been identified that significantly impact residential energy consumption: the thermal envelope, external conditions (outdoor temperature), and occupant behaviour. To streamline the housing simulation process, we choose these inputs as archetypes based on classification studies. Utilizing archetypes as inputs within our methodology significantly diminishes the number of simulations required, given that each archetype delineates a homogeneous segment. The set of representative housing situations investigated is defined by all the combinations between the three inputs. However, these archetypes must encompass all essential information required for thermal simulations in heating management. In Section 3, we will present an example of classification work that fits our methodology.

2.1.2. Generation of the thermal model of the building archetype

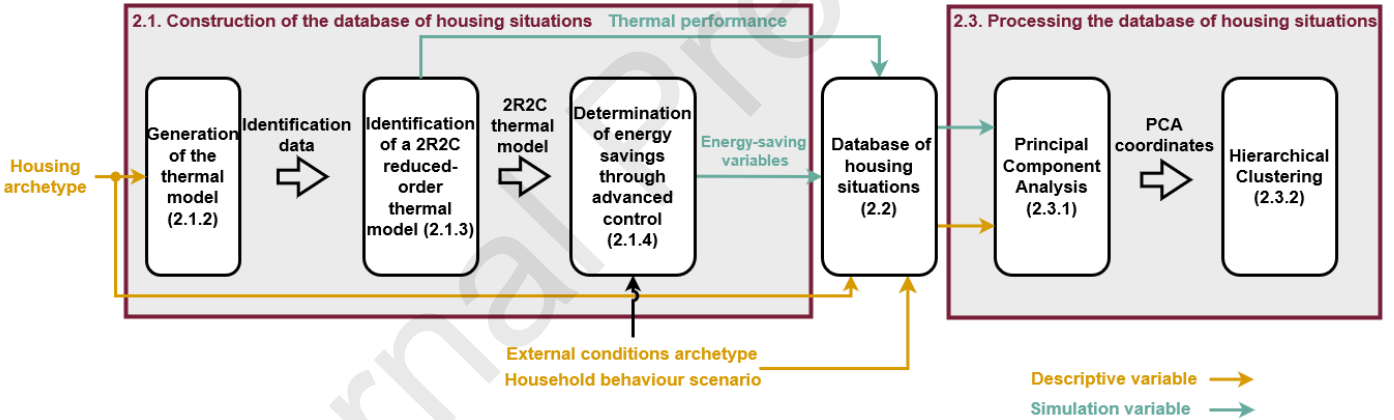


FIG. 1. Flowchart of the presented methodology

2.1.3. Identification of a reduced-order thermal model

The methodology can be directly applied to the thermal models from the previous section. However, the combination of the three inputs, defining all the housing situations to be addressed results in a significant number of thermal simulations, further compounded by the implementation of advanced control.

To address this difficulty, the use of reduced-order thermal models, whose parameters are identified from the thermal model of the housing archetypes, is proposed. The choice has been made to use second-order "grey box" models of the 2R2C type, which are sufficient to characterize the dynamic thermal behaviour of the building air indoor temperature [35]. The parameters of this model are identifiable [36]. The 2R2C models offer a time-saving advantage, facilitating the numerical processing of the entire set of housing situations. Even though these are reduced-order models, they remain physically interpretable, due to their pseudo-physical nature. The thermal capacity C_i characterises the fast dynamics of the indoor air, while the thermal capacity C_s describes the internal masses and the thermal envelope. Additionally, the thermal resistances R_i and R_s model the heat exchanges between the external air

The estimation of energy-saving potential through thermal simulation requires thermal models of the archetypes. The housing input provides information on the thermal envelope of each archetype, including the materials used, their thickness, and the type of windows and doors. Using these parameters, a thermal model of the archetypes can be constructed using dynamic thermal simulation software. The most recognised software for thermal simulation includes EnergyPlus [32], TRNSYS [33], and other methods such as using the Aixlib library [34] developed in the Modelica language or Simulink in MATLAB SimScape. These software packages model the dwelling using the thermal/electrical analogy. To summarize, these models describe the temperature evolution using the following discrete state equation:

$$\begin{aligned} X_{k+1} &= f(X_k, u_{h_k}, T_{out_k}, w_k) \\ T_{i_k} &= CX_k \end{aligned} \quad (1)$$

Considering:

- X : State vector composed of the relevant temperatures in the thermal simulation software model
- u_h : Heating power
- T_{out} : Outdoor air temperature
- T_i : Indoor air temperature
- w : Any other disturbances (solar irradiations, ...)

temperature T_{out} , the surface temperature T_s , and the indoor air temperature T_{2R2C_i} (Figure 2). The temperature evolution is described by the following discrete state model:

$$\begin{aligned} X_{2R2C_{k+1}} &= AX_{2R2C_k} + Bu_k + ET_{out_k} \\ T_{2R2C_{i_k}} &= CX_{2R2C_k} \end{aligned} \quad (2)$$

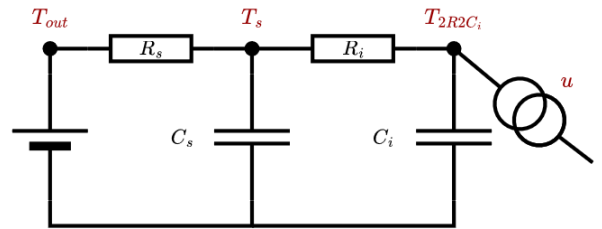


FIG. 2. 2R2C grey-box model of the housing thermal envelope.

The 2R2C parameters are obtained through an identification process using thermal models from the DTS software. Initially, heat

power and outdoor temperature disturbances are applied as inputs to the open-loop DTS model, allowing for the excitation of all the frequencies present in the building. The air temperature at each iteration, heat power, and outdoor temperature are extracted and used as training data. The identification of the 2R2C model is performed using external software. The model parameters are identified by considering the following optimization problem for a number N of training data samples.:

$$\begin{aligned} \min J(R_i, R_s, C_i, C_s) &= \sum_{k=0}^{N-1} (T_{2R2C_{i_k}} - T_{i_k})^2 \\ \text{submit to:} \quad T_{2R2C_{i_0}} &= T_{i_0} \\ 0 < R_i &\leq R_{i_{max}} \\ 0 < R_s &\leq R_{s_{max}} \\ 0 < C_i &\leq C_{i_{max}} \\ 0 < C_s &\leq C_{s_{max}} \end{aligned} \quad (3)$$

This problem can be solved using any suitable method (e.g. nonlinear solver, genetic algorithms, ...) The maximum constraint imposed on the search for each parameter is estimated based on the physical interpretation of the parameters and the materials described in the housing archetype descriptions. The reliability between the 2R2C model and the DTS software model is estimated by calculating the mean squared error (MSE) and the percentage of fit (Fit%) between the evolution of the indoor air temperature predicted by the 2R2C model and the DTS software model for the training data considered. This can be expressed as follows:

$$MSE = \frac{\sum_{k=0}^{N-1} (T_{2R2C_{i_k}} - T_{i_k})^2}{N} \quad (4a)$$

$$Fit_{\%} = 100 \cdot \left(1 - \frac{\sqrt{MSE}}{|T_{i_{max}} - T_{i_{min}}|}\right) \quad (4b)$$

Once the 2R2C parameters are obtained, the potential energy savings of the heating management algorithm can be determined using the reduced-order models, speeding up the processing of multiple housing situations.

2.1.4. Determination of energy savings through advanced control

The simulation environment is defined by the variables associated with the inputs described in section 2.1.1. It allows for the estimation of the potential of the observed control strategy for a given housing situation. Two thermal heating management simulations are computed. The first simulation estimates the energy consumption for a reference algorithm, denoted as E_{ref} . The second simulation is conducted for the investigated algorithm and estimates the energy consumption, denoted as E_{algo} . In addition, three quantitative metrics of interest complement the housing situation description, which are the **Energy-saving potential variables** represented in Figure 1.

- The « **Consumption (kWh/m²)** » (E_{ref}): this variable quantifies the consumption per square meter for the reference algorithm.
- The « **Energy saving (kWh/m²)** » (ΔE_{kWh}): this variable quantifies the energy saving per square meter in comparison with the reference algorithm.

$$\Delta E_{kWh} = E_{ref} - E_{algo}$$

- The « **Energy saving (%)** » ($\Delta E_{\%}$): this variable quantifies the energy saving in comparison with the reference algorithm in percentage.

$$\Delta E_{\%} = 100 \cdot \frac{E_{ref} - E_{algo}}{E_{ref}}$$

2.2. Database of housing situations

As specified in subsection 2.1.1., the entirety of the housing situations comprising the database is constructed from thermal simulations of all possible combinations of the various inputs. As a result, these situations are described by both the energy-saving potential variables and the descriptive variables of the different inputs. The set of variables used to analyse the database will be detailed in the implementation section (number 3).

2.3. Processing of the study case database

2.3.1. Principal Component Analysis

A first treatment of the database is performed using Principal Component Analysis (PCA) [37]. This initial representation allows for the exploration of the relationships between the variables defining the housing situations and the energy-saving potential variables determined by thermal simulation. Furthermore, representing the data along the two axes that exhibit the highest variability enables the visualization of all the housing situations in a two-dimensional space. This provides a comprehensive view of the entire set of housing situations and facilitates the identification of any patterns or groupings among them.

This algorithm synthesizes the information contained in the database into a reduced-dimensional space where the axes are constructed as linear combinations of the different variables while being pairwise orthogonal. PCA takes the quantitative data from the previous table as input.

From a methodological standpoint, the data is normalized and centred around its mean. A covariance matrix is then calculated to assess the relationship between the variables. The eigenvectors represent the directions of maximum variance, while the associated eigenvalues represent the amount of variance in the data. The eigenvectors with the highest eigenvalues are defined as the principal components. The data is then projected onto these components, creating a new reduced-dimensional space while preserving their information as best as possible by determining the projection coefficients that maximize the variance of the dataset.

2.3.2. Hierarchical clustering

Agglomerative hierarchical clustering is an unsupervised analysis method used to group similar individuals into hierarchical clusters [38]. The clustering algorithm is fed with the resulting coordinates from the PCA. In the first step, a distance matrix is calculated. For the proposed methodology, the distance is defined as the Euclidean distance between the housing situations in the PCA-derived space. The algorithm then groups similar housing situations into the same cluster. The process is repeated until a single cluster is formed. Once all iterations are completed, the optimal number of clusters is determined based on the resulting dendrogram from hierarchical clustering, which minimizes the similarity between two clusters.

Regarding the posed problem, agglomerative hierarchical clustering aims to classify the housing situations based on their overall profile and not only the energy-saving potential variables. The objective is to observe whether there exists any link between the housing situations and the energy-saving potential of the management

algorithm within the same cluster. If such links exist, it would be possible to estimate the range of energy-saving potential for an algorithm in a housing situation defined by the type of housing, the year of construction, the insulation level, the climatic zone, and the occupancy of the dwelling, without the need for a case-by-case study to assess the energy efficiency of its implementation.

3. Implementation – case application

The proposed implementation aims to illustrate the potential of the methodology. Thus, the energy-saving potentials of each housing

situation are not accurately calculated due to the choices of the inputs defining the housing situations. However, the results will lead to observed trends between the inputs defined and the energy-saving potential variables. Figure 3 details the tools used to implement the methodology and the inputs used. Each choice made is detailed in the following section. The structure of this flowchart is similar to that shown in Figure 1. The implementation choices made are highlighted in purple.

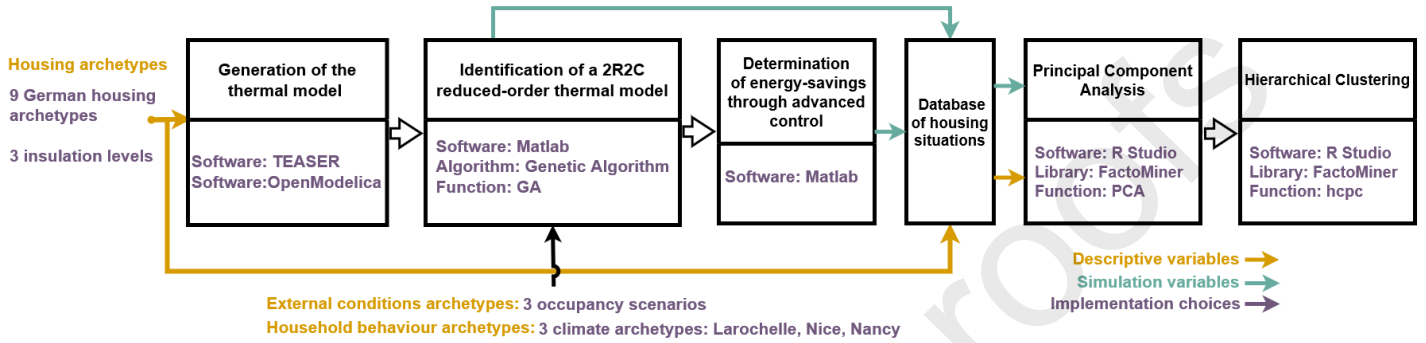




FIG. 3. Flowchart of the implementation. The implementation choices made are detailed in purple for each block of the methodology flowchart.

3.1. Definition of the inputs and generation of a database of housing situations

3.1.1. Inputs

For the input housing archetype, we use the dataset produced in the TABULA project. This European project classifies the housing stock of 20 countries into homogeneous segments [29]. These segments are the result of classifying national data on energy consumption in existing dwellings. The dwellings included in the same segment are characterized by their year of construction and housing type (single-family house, semi-detached house, multi-family dwelling, apartment building). Each segment is represented by an archetype, which represents the characteristics of the dwellings within that segment (Table 1). For each range of construction years, an archetype is proposed for single-family houses, terraced houses, multi-family dwellings, and apartment blocks. The collection of these archetypes forms a representation of the overall typology of the housing stock. Thus, each existing dwelling can be associated with its corresponding TABULA archetype.

Table 1. Examples of housing segments in the stock and associated TABULA archetypes.

Segment	Housing type	Archetype
[1919 ;1948]	Single-Family House	
[1969 ;1978]	Terraced House	

[1984 ;1994]

Multi-Family House



The methodology presented in the previous section was applied to 9 TABULA archetypes of German single-family houses, covering dwellings built between 1919 and 2015. Three insulation levels, provided by TABULA, labeled 0, 1, and 2, were considered to explore the impact of thermal envelope renovations on the energy-saving potential of advanced control strategies, except for the archetypes representing the most recent single-family houses built after 2001. For each combination of archetype and insulation level of the house, a sensitivity analysis of the house's surface area is conducted. Despite the thermal models used normalizing the consumption per square meter for the same dwelling, this sensitivity analysis enhances the robustness of the archetype identification process. In the end, the database of housing situations consists of 115 thermal envelopes. Table 2 summarizes the various combinations considered. For archetypes SFH08 and SFH09, only insulation level 1 is considered. For the rest of the archetypes, additional insulation levels 2 and 3 are considered. The total number of generated housing situations for this application is 1035.

Table 2. Summary of the composition of the database of thermal envelopes for single-family houses concerning the variables of archetypes and insulation levels.

Housing archetype [29]	Year of construction	Insulation level	Floor area (m ²)
SFH01	[1919 ;1948]	0	[115,144,176,201,230]
		1	[115,144,176,201,230]
		2	[115,144,176,201,230]
SFH02	[1949 ;1957]	0	[115,144,176,201,230]
		1	[115,144,176,201,230]

		2	[115,144,176,201,230]
SFH03	[1958 ;1968]	0	[115,144,176,201,230]
		1	[115,144,176,201,230]
		2	[115,144,176,201,230]
SFH04	[1969 ;1978]	0	[115,144,176,201,230]
		1	[115,144,176,201,230]
		2	[115,144,176,201,230]
SFH05	[1979 ;1983]	0	[115,144,176,201,230]
		1	[115,144,176,201,230]
		2	[115,144,176,201,230]
SFH06	[1984 ;1994]	0	[115,144,176,201,230]
		1	[115,144,176,201,230]
		2	[115,144,176,201,230]
SFH07	[1995 ;2001]	0	[115,144,176,201,230]
		1	[115,144,176,201,230]
		2	[115,144,176,201,230]
SFH08	[2002 ;2009]	0	[115,144,176,201,230]
SFH09	[2010 ;2015]	0	[115,144,176,201,230]

As France is usually divided into eight climatic zones, we propose to use them to model the influence of the outdoor conditions as external conditions archetypes. Weather data from three cities over one month during the winter period are used as inputs for outdoor conditions. The cities are selected considering the three main climatic zones of the French topological geography. Table 3 summarizes the information about the representative cities of the time series used, as well as the values of the average temperatures of the series.

Table 3. Climatic zones and values of the associated average outside temperature variable during the heating period.

Climatic zone	City	Average outdoor temperature (°C)
H1	Nancy	1.2
H2	La Rochelle	5.4
H3	Nice	8.02

Several studies explored the use of time use survey data to characterize household occupancy. Nevertheless, a sensitivity analysis of household behaviour is suggested to observe its influence on the energy-saving potential, defined by the input **household behaviour scenario**. Three deterministic scenarios are generated: the first one is associated with an absence from the dwelling between 8 am and 7 pm, the second one is associated with an absence from the dwelling between 10 am and 4 pm, and the last one represents an average absence of 3 hours between 1 pm and 4 pm. These simplified scenarios have been defined to represent diversity in the household behaviour of

the different housing situations. Table 4 presents the information for the three absence scenarios.

Table 4. Deterministic occupancy scenarios for the household behaviour scenario input and the value of the average occupancy variable

Time of departure	Time of return	Average occupancy (h)
13	16	21
10	16	18
8	18	14

3.1.2. Building thermal models

To our knowledge, work has yet to develop thermal models for the French housing stock archetypes. Therefore, we used thermal models from the German housing stock obtained through the TEASER software. Indeed, the TEASER software is an open-source tool that extracts white-box thermal models in the Modelica language from the German housing archetypes provided by TABULA [39]. The consumption prediction of these models is scientifically validated through comparison with simulations under TRNSYS. Furthermore, the German housing stock has developed thermal building regulations similar to the French housing stock [40].

These models are constructed using the AIXLIB library and consist of the thermal capacity of the air, two envelope resistances, and a thermal capacity representing each building element, such as external walls, ceiling, and floor. All housing models are exported to the Modelica language based on the data from Table 2. For each TEASER thermal model, the identification process described in subsection 2.1.3 is carried out under OpenModelica with a time step of 6 minutes, for a training period of 14 days. The frequencies and amplitudes of the external temperature and heating power noise are estimated to provide rich training data for identifying the 2R2C model.

3.1.3. Identification of 2R2C reduced thermal model

The identification process (described in 2.1.3) is performed using MATLAB with a genetic algorithm using the GA function. Discretization is carried out using the Euler method. Figure 4 presents the distribution of the identified parameters normalized per unit area for each housing archetype. The variability of the parameters is linked to variations in the surface and thermal renovation levels. For the same archetype, thermal renovation increases the values of resistance R_i and R_e . However, it has little impact on the values of the thermal capacities, which remain almost the same regardless of the observed insulation level. However, the thermal capacity of the air is sensitive to the identification process, as shown by significant standard deviations.

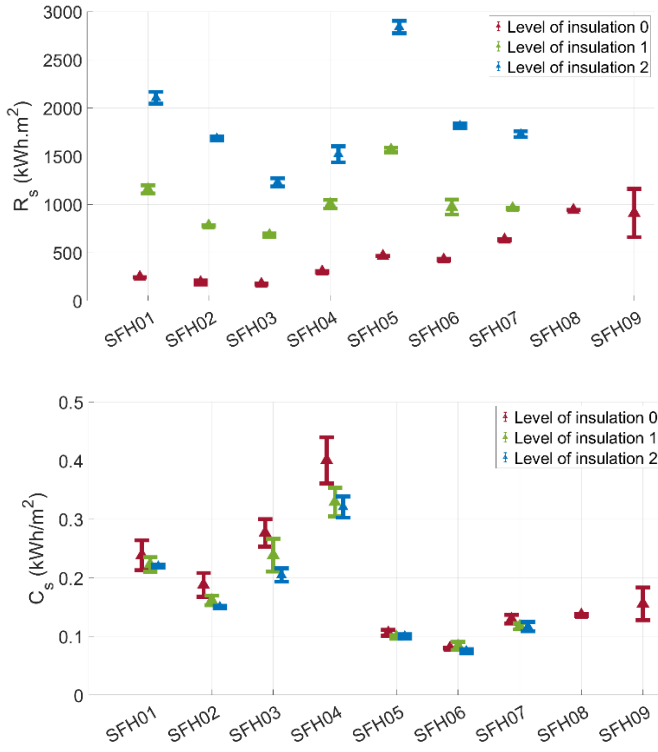


FIG. 4. Average identified parameters normalized per unit area for each single-family house archetype considering the sensitivity on the floor area. Red: Insulation level 0, green: Insulation level 1, blue: Insulation level 2

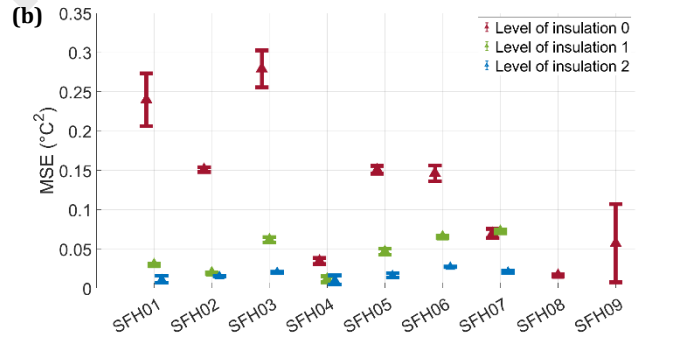
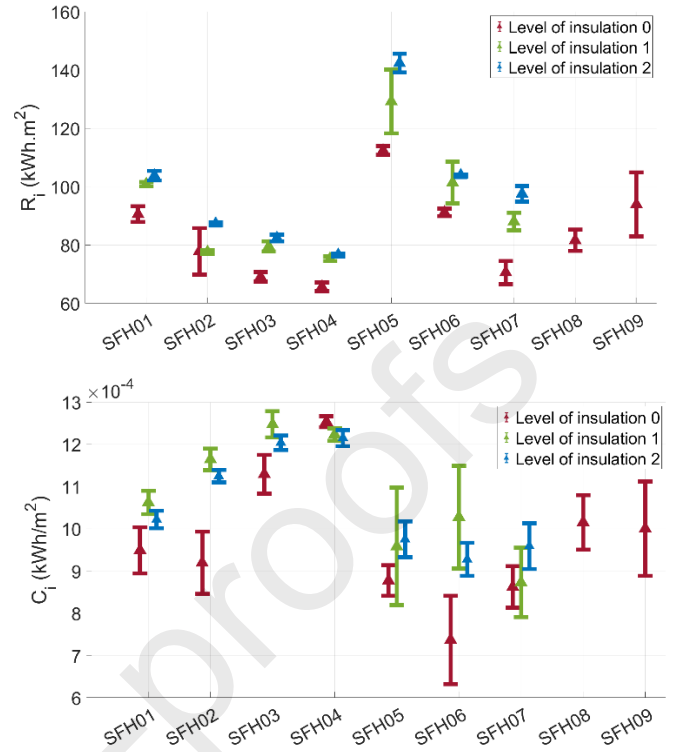
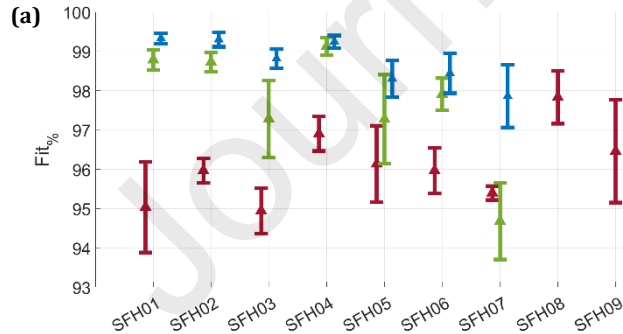


FIG. 5. (a) average fit percentage for each housing archetype and degree of isolation. (b) mean squared error per housing archetype and degree of isolation.

The performance of the identification process is evaluated by calculating the fit percentage for each building (Figure 5). The average minimum fit percentage is 95 % for the SFH07 archetype with an insulation level of 2. For all buildings, the average fit percentage is 97 % ($\sigma=1.6$ %). Furthermore, all mean squared errors are below 0.3°C for the insulation level 1 of the SFH03 archetype. Overall, the observation of fit percentages and mean squared errors are sufficiently high to validate the parameters. Each building is retained for the generation of test cases



3.1.4. Determination of energy saving through linear model predictive control

The heating control algorithm studied is a linear predictive control with intermittent management [39]. At each time instant k , the controller determines the control sequence of heating power $u_{K|k}$ over a prediction horizon of length K , minimizing an optimization criterion. Only the first element of the sequence is applied to the heating system, and a new calculation is performed at the next time step. The constraint to be respected is related to the comfort when the occupant is present in the dwelling, which is indicated by the occupancy matrix $OCC_{K|k}$. To ensure the feasibility of the optimization problem, this constraint is relaxed by a deviation variable, denoted $\varepsilon_{T_{K|k}}$, and penalized by the variable pen_T . The optimization problem takes the following form:

$$\min J(u_{K|k}, \varepsilon_{T_{K|k}}) = 1_{N_p}^T u_{K|k} + pen_T 1_{N_p}^T \varepsilon_{T_{K|k}} \quad (5)$$

$$\begin{aligned}
\text{submit to: } & 0 \leq u_{K|k} \leq u_{max} \\
& T_{2R2C_i|K|k} + \varepsilon_{T_{K+1|k}} \geq OCC_{K+1|k} W_{T_{K+1|k}} \\
& \varepsilon_{T_{K+1|k}} \geq 0
\end{aligned}$$

The propagation of the state model (2) allows expressing the predicted outputs as a function of the current state X_{2R2C_k} , which is the air temperature over the entire prediction horizon:

$$T_{2R2C_{air|K+1|k}} = F_{2R2C} X_{2R2C_k} + H_{u_c} u_{c|K|k} + H_T T_{ext|K|k} \quad (6)$$

The prediction matrices are obtained recursively, and we denote the number of prediction horizon iterations as $N_p = K - k$:

$$F_{2R2C} = \begin{pmatrix} CA \\ \vdots \\ CA^{N_p} \end{pmatrix} \quad H_{u_c} = \begin{pmatrix} CB & 0 & 0 & \dots & 0 \\ CAB & CB & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ CA^{N_p-1}B & CA^{N_p-2}B & CA^{N_p-3}B & \dots & CB \end{pmatrix}$$

$$H_T = \begin{pmatrix} CB & 0 & 0 & \dots & 0 \\ CEB & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ CE^{N_p-1}B & CE^{N_p-2}B & CE^{N_p-3}B & \dots & CB \end{pmatrix}$$

The maximum heating power u_{max} is defined in such a way that it can maintain a steady-state temperature difference between $T_{2R2C_i} = 20$ °C and $T_{out} = -5$ °C, which is given by: ΔE_{kWh}

$$u_{max} = \frac{T_{2R2C_i} - T_{out}}{R_s + R_i} \quad (7)$$

For the proposed implementation, the penalty pen_T and the heating power are adjusted to minimize energy consumption while ensuring user comfort at all times. The prediction of the household occupancy is perfect. Thus, the energy-saving potentials of different housing situations are comparable because they all exhibit the same comfort index. These settings place this implementation in estimating the **maximum potential** of the control without compromising comfort.

The variables related to the energy saving potential, such as **energy saving (kWh/m²)**, **energy saving (%)** and **consumption (kWh/m²)**, are determined through thermal simulation over 31 days using MATLAB with a frequency of 6 minutes, corresponding to an adapted trade-off between the accuracy of the air temperature prediction and the computation time.

The reference algorithm used is a heating management algorithm without occupancy and outdoor temperature prediction. It solely compensates for heat losses to maintain the indoor air temperature at the set point determined by the behaviour variables, fixed at 20°C. Figure 6 shows the evolution of the indoor air temperature (a) for both the predictive linear control (blue curve) and the reference (red curve) algorithms, the control sequences for the heating management algorithm and the reference (b), the outdoor temperature for the Nancy climatic zone (d), and the occupancy scenario with an absence from 10 am to 4 pm for 7 days (c). The housing situation corresponds to archetype SFH05 with a floor area of 140 m². All disturbances are perfectly predicted, so there is no thermal discomfort. The predictive linear control strategy involves waiting until the last moment to preheat the air and reach the desired set point of 20°C when occupants return at 4 pm.

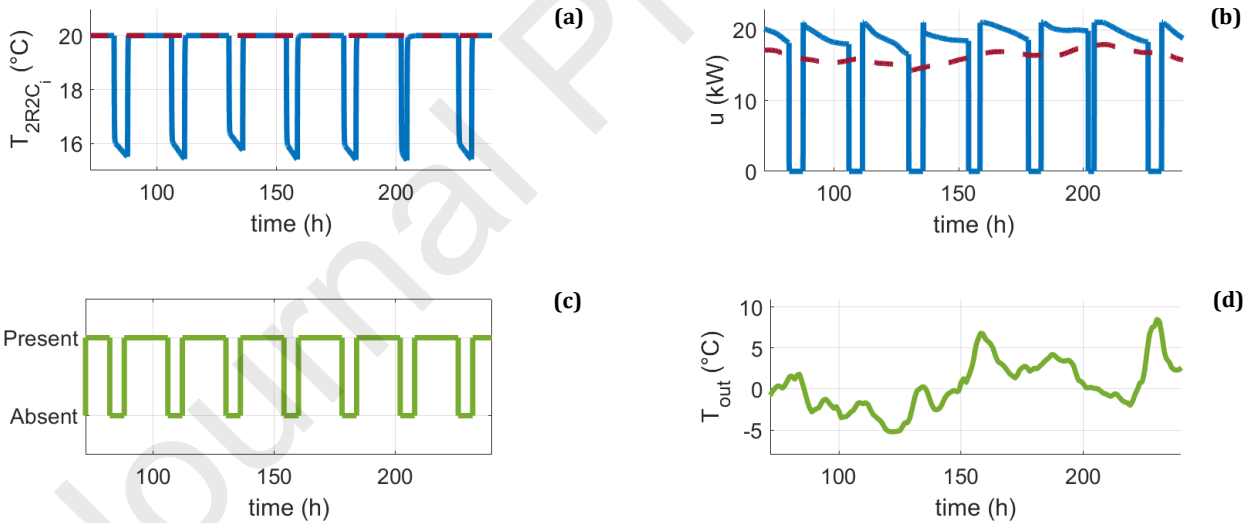


FIG. 6. Control sequences of the predictive control algorithm (in blue) and the reference algorithm (in red) for the SFH05 archetype, a 140 m² surface area, an outdoor temperature of Nancy, and an occupancy scenario with an absence between 10 a.m. and 4 p.m.

The energy consumptions of the buildings for the reference algorithm E_{ref} and the predictive control E_{algo} are obtained by integrating their respective control sequences over the entire simulation time t_{cont} , as follows:

- $E_{ref} = \int_0^{t_{cont}} u_{ref}(t) dt$
- $E_{MPC} = \int_0^{t_{cont}} u_{MPC}(t) dt$

From the energy consumption values for each housing situation, the potential energy savings variables are calculated as defined in subsection 2.1.4.

3.2. Construction of the database of housing situations

For the analysis of the results proposed in the following section, the external conditions are limited to the outside temperature. It is defined qualitatively by the climatic zone and its time series average, which constitutes useful quantitative information for the database processing, as follows:

- « **Average outdoor temperature** »: $\overline{T_{out}}$ average outdoor temperature for the considered outdoor conditions archetype.

The household behaviour is represented by its occupancy of the dwelling, characterized by its daily average:

- « **Average occupancy** »: $\overline{t_{occ}}$ Average occupancy time of the dwelling for the considered scenario.

In addition, for the database processing, a variable synthesizes information regarding the thermal envelope quality of the TABULA archetype based on the identified 2R2C parameters and the floor area of the dwelling, S_{floor} .

- « **Thermal performance** »: $P_t = (R_i + R_s) \cdot S_{floor}$

The energy-saving potential of each housing situation is quantified using the metrics described in the previous subsection. Table 5 summarizes all the variables considered.

Table 5. Summary and description of variables defining a housing situation used for the following application. The list can be expanded.

Variable	Type of variable	Description
TABULA Archetype		
Archetype	qualitative	TABULA denomination
Type of housing	qualitative	Type of housing
Insulation level	qualitative	Level of renovation of the thermal envelope for the considered archetype
Thermal parameters	/	Useful thermal parameters for the building of the STD model
Outdoor conditions archetype		
Climate zone	qualitative	Name of the archetype city
Outdoor temperature	/	Time series
Average outdoor temperature	quantitative	$\overline{T_{out}}$
Household behaviour scenario		
Occupancy scenario	qualitative	Occupancy scenario type
Average occupancy	quantitative	$\overline{t_{occ}}$
Energetic criteria		
Thermal performance	quantitative	P_t
Consumption (kWh/m ²)	quantitative	E_{ref}
Energy saving (kWh/m ²)	quantitative	ΔE_{kWh}
Energy saving (%)	quantitative	$\Delta E_{\%}$

3.3. Processing of the database of housing situations

3.3.1. Principal component analysis

As presented in subsection 2.2.1, the Principal Component Analysis (PCA) algorithm takes the quantitative values from Table 2 as input. PCA is performed in R using the PCA function from the FactoMiner package. The first two axes are retained, synthesizing 74 % of the variance in the dataset. The **thermal performance**, **energy saving (kWh/m²)**, **energy saving (%)** and **consumption (kWh/m²)** contributed to the construction of the horizontal axis, while **average occupancy**, **average outdoor temperature**, **consumption (kWh/m²)**, and **energy saving (%)** contributed to the construction of the vertical axis (Table 6). The correlation circle shows that the **thermal performance** and the

average outdoor temperature variable are negatively correlated with the energy savings potential variables. On the other hand, the **average outdoor temperature** is positively correlated with the energy savings potential variables (Figure 7).

Table 6. Contribution in per cent of variables to the construction of the two resulting axes from PCA.

Variable	Symbol	Dimension 1	Dimension 2
Thermal performance (K/kW.m ²)	P_t	19.60	8.37
Average occupancy (h)	$\overline{t_{occ}}$	4.21	38.25
Average outdoor temperature (°C)	$\overline{T_{out}}$	0.01	23.01
Energy saving (%)	$\Delta E_{\%}$	24.84	13.21
Energy saving (kWh/m ²)	ΔE_{kWh}	28.86	0.13
Consumption (kWh/m ²)	E_{ref}	22.48	17.01

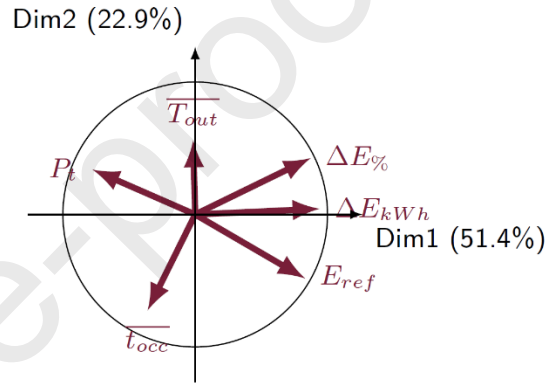


FIG. 7. Correlation circle between qualitative variables and the two principal component axes.

Figure 8 depicts the representation of data points in the principal components space. Insulation levels and TABULA archetypes are plotted with colour codes to support an analysis of the result. In Figure (8a) we observe that non-renovated buildings have the best energy-saving potential. This is coherent with the fact that the thermal performance P_t is negatively correlated with the first component. In the meantime, the dispersion of the housing situations for non-renovated dwellings indicates significant variability in the energy-saving potential. In conjunction with the observation of Figure (8b), the archetypes corresponding to buildings constructed after 1995 that have not undergone renovations are close to older buildings that have undergone a first level of renovation in the PCA space.

In a nutshell, the analysis of this work suggests that not only the housing type but also the renovation insulation and the surface are major determinants of the automation interest. The methodology proposes to go a bit further by describing segments of housing situations discriminated by their potential energy savings.

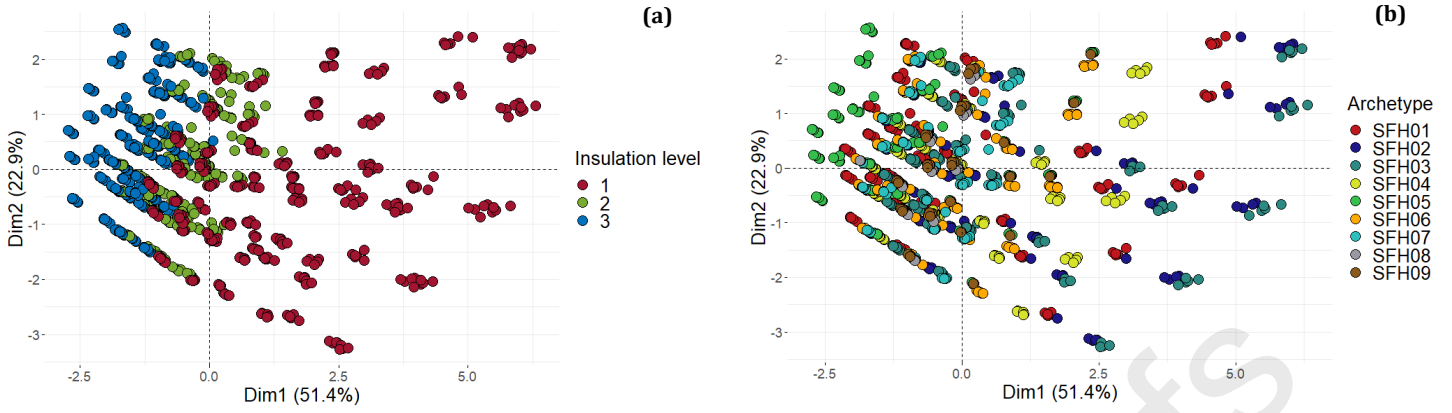


FIG. 8. (a) The insulation level of the housing situations in the resulting PCA database. (b) The archetypes of the housing situations in the resulting PCA database.

3.3.2. Ascending hierarchical clustering

The resulting coordinates from the PCA (Principal Component Analysis) feed into an ascending hierarchical clustering algorithm to group the different housing situations into clusters. The first step is to select the desired number of clusters. For this purpose, the elbow method is used to estimate the optimal number of clusters based on intra-cluster inertia [34]. Intra-cluster inertia represents the coherence of data within the same group. The more clusters there are, the lower the intra-cluster inertia becomes, indicating homogeneity among the housing situations in the groups. The elbow method involves choosing a number of clusters that show a significant drop in inertia. In this study, significant inertia drops are observed for the number of clusters equal to 3, 4, 7, 9, and 10. The choice has been made to use 9 clusters. Figure 9 represents the dendrogram of the housing situations for classification into 9 groups.

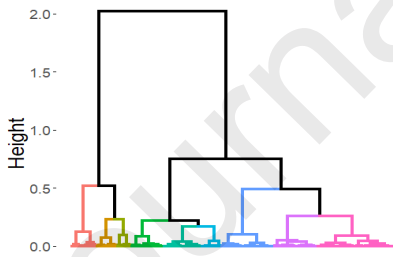


FIG. 9. Dendrogram resulting from ascending hierarchical clustering.

4. Results and discussions

We observe significant variability in the energy-saving potential variables. Indeed, the minimum gain in kWh/m² (respectively, %) among all the housing situations is 0.04 kWh/m² (respectively, 0.7%), while the maximum is 12 kWh/m² (respectively, 23 %). In terms of kWh/m², this represents a ratio of 300 between the housing situation with the highest gain and the one with the minimal gain. Figure 10 shows the distribution of the predictive control gain for the 9 generated clusters in the form of boxplots. The median distribution of the gain per kWh/m² increases from group 1 to 9. The analysis of the clusters allows us to differentiate clusters 1 to 6 as low interest for automation and clusters 7 to 9 as high-interest cases for automation. Observing the

variables defining the housing situations within the groups leads to the identification of housing profiles.

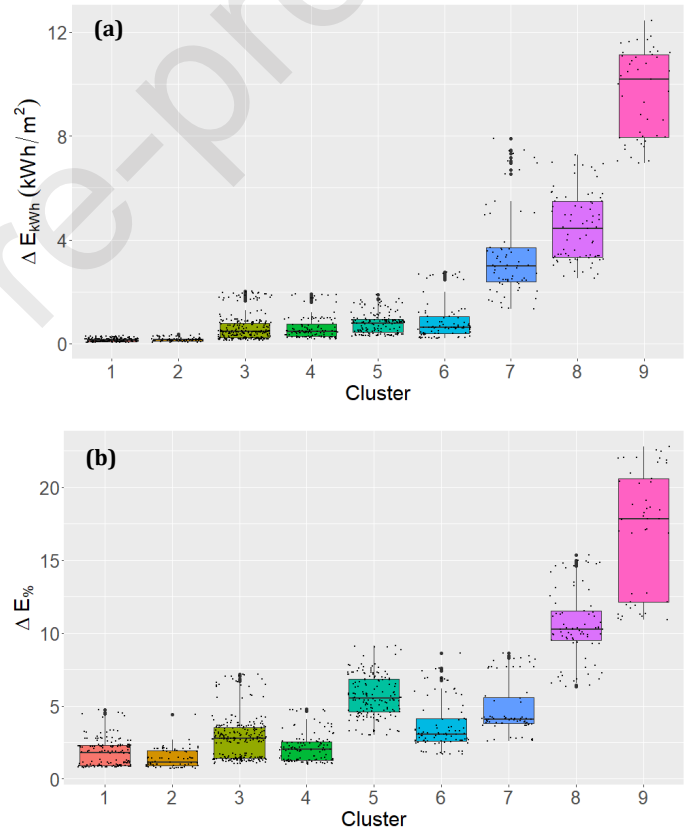


FIG. 10. (a) Box diagram of the Energy saving in kWh/m², ΔE_{kWh} , achieved by each cluster of clusters of housing situations. (b) Box diagram Energy saving in %, $\Delta E_{\%}$, achieved by each cluster of housing situations

Cluster 1: This group consists of dwellings that have undergone a third level of renovation associated with the climatic zones of La Rochelle (50%) and Nice (50%), and are predominantly of low and medium occupancy (46%). The average gain in this group, in terms of percentage, is 1.8% ($\sigma=0.9\%$), and in terms of kWh/m², it is 0.14 ($\sigma=0.07$ kWh/m²).

Cluster 2: This cluster consists of dwellings that have undergone a second level of renovation. 100% of the housing situations have a temperature scenario associated with the city of Nancy. The

households are of low or medium occupancy. The average gain in this group, in terms of percentage, is 1.40% ($\sigma=0.64\%$), and in terms of kWh/m², it is 0.16 ($\sigma=0.07$ kWh/m²).

Cluster 3: This cluster consists of homes that fall into three categories: non-renovated dwellings built after 1968, dwellings with a first level of renovation constructed between 1918 and 1948, and dwellings with a second level of renovation constructed between 1958 and 1968. The households are predominantly of low or medium occupancy. Among them, 47% have an absence of 6 hours and 53% have an absence of 3 hours. These distributions are identical for temperature scenarios between La Rochelle and Nice. The average gain in this group, in terms of percentage, is 2.80% ($\sigma=1.5\%$), and in terms of kWh/m², it is 0.61 ($\sigma=0.50$ kWh/m²).

Cluster 4: This cluster consists of two types of dwellings: non-renovated dwellings after 1978 and all dwellings that have undergone a first level of renovation. Additionally, the households are highly present in the housing (100%). The climate zone for all housing situations is Nancy (100%). The average gain in this group is 2.17% ($\sigma=0.97\%$), and in terms of kWh/m², it is 0.62 ($\sigma=0.50$ kWh/m²).

Cluster 5: This cluster consists of two types of dwellings: non-renovated dwellings built after 1994 and dwellings renovated 1 or 2 times built before 1994. All households in this group have an average daily absence of 10 hours. The average gain in this group, in terms of percentage, is 5.70% ($\sigma=1.4\%$), and in terms of kWh/m², it is 0.79 ($\sigma=0.40$ kWh/m²).

Cluster 6: The housing situations in this group consist of dwellings located in Nancy (100%) with a very significant daily absence of 10 hours per day (100%). These dwellings are energy-efficient: they include non-renovated dwellings built after 1994 and dwellings built before 1994 that have undergone thermal renovation. The average gain in this group, in terms of percentage, is 3.67% ($\sigma=1.7\%$), and in terms of kWh/m², it is 0.90 ($\sigma=0.76$ kWh/m²).

Cluster 7: This group consists of non-renovated dwellings built between 1919 and 1968. The households are predominantly very sparsely occupied. Specifically, 70% of the housing situations have a daily absence of 3 hours. Regarding the climate conditions, we found that these dwellings are located near Nancy (60%). The average gain in kWh/m² for this group is in percentage, lower than that of groups 9 and 10 but it still has an average gain of 3.5 kWh/m².

Cluster 8: This cluster consists of non-renovated dwellings built between 1919 and 1994 (100%). The geographical zone associated is either La Rochelle (46.8%) or Nice (46.8%), with an average daily absence of 6 hours (48%) or 10 hours (45%). Regarding the gain variables, we observe an average gain in percentage of 11% ($\sigma=2.4\%$), and in terms of kWh/m², it is 4.8 ($\sigma=1.3$ kWh/m²).

Cluster 9: This group consists of dwellings built between 1919 and 1968 that have not undergone any renovation (100% of the cases in the cluster). Additionally, all the households are sparsely occupied (100%). The geographical zone is not a dominant variable in the construction of this group. One-third of the housing situations have an outdoor temperature scenario in Nancy, La Rochelle, or Nice. The average gains are 9.7 kWh/m² and 17%. It is mainly the outdoor temperature variable that generates a significant standard deviation of 1.7 kWh/m² and 4.2%. The housing situations in this group represent the highest potential for energy savings through heating management.

The analysis of the previous clusters and Figure 10 allows for the characterization of the impact of different variables on the energy-saving potential of dwellings.

In the first instance, non-renovated buildings constructed between 1919 and 1968 stand out from others due to their low thermal performance. Coupled with extended periods of household absence, these dwellings represent the most significant energy-saving potential among all clusters. A closer examination of cluster 8 reveals that outdoor temperature generates significant variability in gains both in percentage and in kWh. Furthermore, even when associated with medium or short absences, these buildings still show interesting energy savings (clusters 7 and 8). In contrast, recent buildings or renovated dwellings combined with predictive control exhibit relatively low energy-saving potential, especially in terms of kWh/m² (clusters 1 to 6). This trend is amplified when occupants are highly present in the dwelling (cluster 3). Additionally, very high thermal performance indicates that managing intermittence has very little impact on consumption (clusters 1, 2, and 3), except in cases where the absences are extended.

Furthermore, clusters associated with warmer temperatures (clusters 3-5-8) stand out in terms of percentage gain. This is due to lower energy requirements to raise the temperature during the return to the dwelling, as there is a smaller temperature drop during the heating cutoff, requiring lesser heating time before the occupants return. In terms of kWh/m², the best algorithm performances for non-renovated homes built before 1979 are associated with outdoor temperatures linked to the city of La Rochelle, owing to a better balance between overall dwelling consumption and the percentage gain associated with preheating shortly before the occupants' return. This trend changes as the thermal performance of homes improves. From cluster 6 and lower (renovated buildings or those constructed after 1994), with equivalent absence scenarios and housing conditions, the best heating management gains are observed for an outdoor temperature associated with the city of Nancy.

This classification and its subsequent analysis present two main advantages. First, it shows that not all housing situations offer a significant theoretical potential for energy savings. Even more, the computation offers a first approximation of the potential energy savings. Second, it allows to identify patterns of situations where the deployment of advanced heating control is most favourable (clusters 7, 8 and 9). Each housing situation is defined by simple metrics: construction period of the dwelling, type of dwelling, degree of insulation, outdoor temperature, and average daily household absence time. This information is readily accessible. This classification can be useful for associating real housing situations with one of the 9 clusters and extracting an estimate of their energy-saving potential through predictive control heating management.

5. Conclusion

5.1. Proposed work

This work brings forth three main contributions. First, a methodology has been developed to generate a database of housing situations coupled with their energy-saving potential through heating management. This is achieved through the combination of classification works on housing and climate with a variety of occupancy scenarios. The methodology is based on the creation of thermal models for housing archetypes and the identification of reduced-order thermal models to simplify the treatment of multiple combinations of housing, external temperature and household absence. These simulations allow for the extraction of identified metrics characterizing the algorithm's energy-saving potential for each housing situation. The processing of the database is done through a combination of Principal Component Analysis (PCA) and hierarchical clustering. This processing of the database helps to identify the influence of variables on the energy-saving potential for all housing situations. Additionally, hierarchical

clustering allows for the grouping of similar situations together, which facilitates the extraction of patterns. The second main contribution is the implementation of the developed methodology to individual housing situations, based on the DTS models developed by the TEASER software in order to estimate the potential of a linear predictive control. This implementation highlights the potential of the methodology by emphasizing the variability of the predictive control's energy savings, the identification of patterns within the clusters, and the influence of renovation. The results for the variables used are coherent with the literature.

5.2. Limitations and future work

The use of German TABULA archetypes is a primary limitation. Utilizing this data may introduce errors in the identified housing patterns. To enhance this aspect, it is essential to employ thermal models using data from the French housing stock sourced from TABULA or to compare French archetypes with those of Germany. Furthermore, the use of deterministic household occupancy scenarios constitutes a major limitation of the proposed work. While they allow us to observe the influence of household absence, these scenarios are not directly linked to existing households, preventing the interpretation of the composition of households defining the housing situations. Another limitation is the perfect prediction of disturbances for linear predictive control. This assumption leads to an inaccuracy of the energy-saving potential for each housing situation in the proposed application. However, it enables only the estimation of the maximum potential for energy-saving through heating management for each situation. Additionally, the absence of any discomfort situations is observed, which means the management algorithm is not evaluated from that perspective.

Based on the previous limitations, interdisciplinary work prospects seem very interesting. The methodology can be strengthened by using household behaviour classifications. It would then be possible to identify patterns enabling us to target households whose behaviour represents a rich source of potential energy savings through advanced control. Behavioural archetypes based on French data are available. They quantify additional variables of interest for the problem addressed in this study, such as the desired thermostat temperature of households and their ability to adopt regulatory behaviours (e.g., lowering heating when absent). However, the occupancy of each archetype is not quantified at every moment of the day but as an average tendency of households to be more or less absent. Therefore, these archetypes cannot be directly integrated into our methodology [40]. Several studies have treated this challenge by classifying time-use survey data and extracting a small number of typical profiles of activity scenarios [41]. This methodology was applied by the authors of [42] to evaluate the influence of these profiles on energy consumption. However, in the context of our work and from an interdisciplinary perspective, household behaviour classifications require the descriptive variables of the households making up each group to be linked to their activity time series. To our knowledge, this link has not been made in existing studies. Finally, the analysis of energy-saving potentials of the housing situations of groups 1 and 2 raises questions about the environmental life cycle profitability of the control devices (sensor, microcontroller) necessary for implementing such heating control (0.14 kWh/m² for group 1, 0.16 kWh/m² for the second). It would be interesting, therefore, to couple this methodology with a life cycle assessment to quantify the environmental benefits of managing heating systems in this way [43], [44].

CRediT authorship contribution statement

A. Wagner: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing an original draft, Writing – review & editing, Visualization **M.Heinrich:** Investigation, Validation, Writing – review & editing **M. Ruellan:** Methodology, Validation, Investigation, Writing – review & editing, Supervision. **R.Bourdais:** Methodology, Validation, Investigation, Writing – review & editing, Supervision.

References

- [1] W. Perera, C. F. Pfeiffer, et N.-O. Skeie, « Control of temperature and energy consumption in building - a review », *Energy Env.*, vol. 5, p. 471-484, janv. 2014.
- [2] P. H. Shaikh, N. B. M. Nor, P. Nallagownden, I. Elamvazuthi, et T. Ibrahim, « A review on optimized control systems for building energy and comfort management of smart sustainable buildings », *Renew. Sustain. Energy Rev.*, vol. 34, p. 409-429, juin 2014, doi: 10.1016/j.rser.2014.03.027.
- [3] H. Mirinejad, H. Sadati, G. Maryam, et T. Hamid, « Control Techniques in Heating, Ventilating and Air Conditioning (HVAC) Systems 1 », *J. Comput. Sci.*, vol. 4, sept. 2008, doi: 10.3844/jcssp.2008.777.783.
- [4] A. Afram et F. Janabi-Sharifi, « Theory and applications of HVAC control systems – A review of model predictive control (MPC) », *Build. Environ.*, vol. 72, p. 343-355, févr. 2014, doi: 10.1016/j.buildenv.2013.11.016.
- [5] F. Bünning *et al.*, « Physics-informed linear regression is competitive with two Machine Learning methods in residential building MPC », *Appl. Energy*, vol. 310, p. 118491, mars 2022, doi: 10.1016/j.apenergy.2021.118491.
- [6] S. Yang, M. P. Wan, W. Chen, B. F. Ng, et S. Dubey, « Model predictive control with adaptive machine-learning-based model for building energy efficiency and comfort optimization », *Appl. Energy*, vol. 271, p. 115147, août 2020, doi: 10.1016/j.apenergy.2020.115147.
- [7] A. Agouzoul, E. Simeu, et M. Tabaa, « Enhancement of Building Energy Consumption Using a Digital Twin based Neural Network Model Predictive Control », in *2023 International Conference on Control, Automation and Diagnosis (ICCAD)*, mai 2023, p. 1-6. doi: 10.1109/ICCAD57653.2023.10152308.
- [8] W.-H. Chen, S. Yang, et F. You, « Thermal Comfort Control on Sustainable Building via Data-Driven Robust Model Predictive Control », in *2023 American Control Conference (ACC)*, mai 2023, p. 591-596. doi: 10.23919/ACC5779.2023.10155818.
- [9] M. Killian, B. Mayer, et M. Kozek, « Cooperative fuzzy model predictive control for heating and cooling of buildings », *Energy Build.*, vol. 112, p. 130-140, janv. 2016, doi: 10.1016/j.enbuild.2015.12.017.
- [10] T. Zeng et P. Baroah, « An Adaptive Model Predictive Control Scheme for Energy-Efficient Control of Building HVAC Systems », *ASME J. Eng. Sustain. Build. Cities*, vol. 2, n° 031001, juin 2021, doi: 10.1115/1.4051482.
- [11] P. J. C. Vogler-Finck, P. D. Pedersen, P. Popovski, et R. Wisniewski, « Comparison of strategies for model predictive control for home heating in future energy systems », in *2017 IEEE Manchester PowerTech*, juin 2017, p. 1-6. doi: 10.1109/PTC.2017.7979747.
- [12] M. Frahm, T. Dengiz, P. Zwickel, H. Maaß, J. Matthes, et V. Hagenmeyer, « Occupant-oriented demand response with multi-zone thermal building control », *Appl. Energy*, vol. 347, p. 121454, oct. 2023, doi: 10.1016/j.apenergy.2023.121454.
- [13] R. Godina, E. M. G. Rodrigues, E. Pouresmaeil, J. C. O. Matias, et J. P. S. Catalão, « Model Predictive Control Home Energy Management and Optimization Strategy with Demand Response », *Appl. Sci.*, vol. 8, n° 3, Art. n° 3, mars 2018, doi: 10.3390/app8030408.
- [14] S. Prívvara, J. Široký, L. Ferkl, et J. Cigler, « Model predictive control of a building heating system: The first experience », *Energy Build.*, vol. 43, n° 2, p. 564-572, févr. 2011, doi: 10.1016/j.enbuild.2010.10.022.
- [15] S. Freund et G. Schmitz, « Implementation of model predictive control in a large-sized, low-energy office building », *Build. Environ.*, vol. 197, p. 107830, juin 2021, doi: 10.1016/j.buildenv.2021.107830.
- [16] T. Hilliard, L. Swan, et Z. Qin, « Experimental implementation of whole building MPC with zone based thermal comfort adjustments », *Build. Environ.*, vol. 125, p. 326-338, nov. 2017, doi: 10.1016/j.buildenv.2017.09.003.

- [17] D. Sturzenegger, D. Gyalistras, M. Morari, et R. S. Smith, « Model Predictive Climate Control of a Swiss Office Building: Implementation, Results, and Cost-Benefit Analysis », *IEEE Trans. Control Syst. Technol.*, vol. 24, n° 1, p. 1-12, janv. 2016, doi: 10.1109/TCST.2015.2415411.
- [18] L. Sha, Z. Jiang, et H. Sun, « A control strategy of heating system based on adaptive model predictive control », *Energy*, vol. 273, p. 127192, juin 2023, doi: 10.1016/j.energy.2023.127192.
- [19] S. Huang, Y. Lin, V. Chinde, X. Ma, et J. Lian, « Simulation-based performance evaluation of model predictive control for building energy systems », *Appl. Energy*, vol. 281, p. 116027, janv. 2021, doi: 10.1016/j.apenergy.2020.116027.
- [20] P. Stoffel, L. Maier, A. Kämpel, T. Schreiber, et D. Müller, « Evaluation of advanced control strategies for building energy systems », *Energy Build.*, vol. 280, p. 112709, févr. 2023, doi: 10.1016/j.enbuild.2022.112709.
- [21] H. Shi et Q. Chen, « Building energy management decision-making in the real world: A comparative study of HVAC cooling strategies », *J. Build. Eng.*, vol. 33, p. 101869, janv. 2021, doi: 10.1016/j.jobee.2020.101869.
- [22] V. Taboga, A. Bellahsen, et H. Dagdougui, « An Enhanced Adaptivity of Reinforcement Learning-Based Temperature Control in Buildings Using Generalized Training », *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 6, n° 2, p. 255-266, avr. 2022, doi: 10.1109/TETCI.2021.3066999.
- [23] « Neural network-based model predictive control system for optimizing building automation and management systems of sports facilities - ScienceDirect ».
- [24] F. Oldewurtel *et al.*, « Use of model predictive control and weather forecasts for energy efficient building climate control », *Energy Build.*, vol. 45, p. 15-27, févr. 2012, doi: 10.1016/j.enbuild.2011.09.022.
- [25] H. Ben et K. Steemers, « Modelling energy retrofit using household archetypes », *Energy Build.*, vol. 224, p. 110224, oct. 2020, doi: 10.1016/j.enbuild.2020.110224.
- [26] E. Ohene, S.-C. Hsu, et A. P. C. Chan, « Feasibility and retrofit guidelines towards net-zero energy buildings in tropical climates: A case of Ghana », *Energy Build.*, vol. 269, p. 112252, août 2022, doi: 10.1016/j.enbuild.2022.112252.
- [27] E. G. Dascalaki, K. G. Droutsas, C. A. Balaras, et S. Kontoyiannidis, « Building typologies as a tool for assessing the energy performance of residential buildings – A case study for the Hellenic building stock », *Energy Build.*, vol. 43, n° 12, p. 3400-3409, déc. 2011, doi: 10.1016/j.enbuild.2011.09.002.
- [28] S. Torabi Moghadam, S. Coccolo, G. Mutani, P. Lombardi, J.-L. Scartezzini, et D. Mauree, « A new clustering and visualization method to evaluate urban heat energy planning scenarios », *Cities*, vol. 88, p. 19-36, mai 2019, doi: 10.1016/j.cities.2018.12.007.
- [29] T. Loga, B. Stein, et N. Diefenbach, « TABULA Building Typologies in 20 European countries – making energy-related features of residential building stocks comparable », *Energy Build.*, vol. 132, juill. 2016, doi: 10.1016/j.enbuild.2016.06.094.
- [30] L. Malka, A. Kuriqi, et A. Haxhimusa, « Optimum insulation thickness design of exterior walls and overhauling cost to enhance the energy efficiency of Albanian's buildings stock », *J. Clean. Prod.*, vol. 381, p. 135160, déc. 2022, doi: 10.1016/j.jclepro.2022.135160.
- [31] M. Heinrich, « Habitats, Habitants et Pratiques énergétiques : Intégrer les situations d'habitation dans la modélisation de la consommation d'énergie domestique », Ecole des Ponts, ParisTech.
- [32] D. Crawley, C. Pedersen, L. Lawrie, et F. Winkelmann, « EnergyPlus: Energy Simulation Program », *Ashrae J.*, vol. 42, p. 49-56, avr. 2000.
- [33] S. A. Klein et W. A. Beckman, « TRNSYS 16: A transient system simulation program: mathematical reference », *TRNSYS*, vol. 5, p. 389-396, janv. 2007.
- [34] D. Mueller, M. Lauster, A. Constantin, M. Fuchs, et P. Remmen, « AixLib – An Open-Source Modelica Library within the IEA-EBC Annex 60 Framework », sept. 2016.
- [35] Y. Li, Z. O'Neill, L. Zhang, J. Chen, P. Im, et J. DeGraw, « Grey-box modeling and application for building energy simulations - A critical review », *Renew. Sustain. Energy Rev.*, vol. 146, p. 111174, août 2021, doi: 10.1016/j.rser.2021.111174.
- [36] A. K. Rivera, J. Sánchez, et M. Chen Austin, « Parameter identification approach to represent building thermal dynamics reducing tuning time of control system gains: A case study in a tropical climate », *Front. Built Environ.*, vol. 8, 2022, Consulté le: 22 mars 2023.
- [37] All, « Probabilités, analyse des données et statistique - SAPORTA Gilbert », Technip.
- [38] T. Hastie, R. Tibshirani, et J. Friedman, *The Elements of Statistical Learning*. in Springer Series in Statistics. New York, NY: Springer, 2009. doi: 10.1007/978-0-387-84858-7.
- [39] D.-P. Morosan, R. Bourdais, D. Dumur, et J. Buisson, « Building temperature regulation using a distributed model predictive control », *Energy Build.*, vol. 42, n° 9, p. 1445, oct. 2010, doi: 10.1016/j.enbuild.2010.03.014.
- [40] H. Matthias, M. Ruellan, L. Oukhellou, A. Samé, et J.-P. Lévy, « From energy behaviours to lifestyles: Contribution of behavioural archetypes to the description of energy consumption patterns in the residential sector », *Energy Build.*, vol. 269, p. 112249, juin 2022, doi: 10.1016/j.enbuild.2022.112249.
- [41] D. Aerts, J. Minnen, I. Glorieux, I. Wouters, et F. Descamps, « A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison », *Build. Environ.*, vol. 75, p. 67-78, mai 2014, doi: 10.1016/j.buildenv.2014.01.021.
- [42] L. Diao, Y. Sun, Z. Chen, et J. Chen, « Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation », *Energy Build.*, vol. 147, p. 47-66, juill. 2017, doi: 10.1016/j.enbuild.2017.04.072.
- [43] N. Mortimer, A. Ashley, C. Moody, J. Rix, et S. Moss, « Carbon dioxide savings in the commercial building sector », *Energy Policy*, vol. 26, n° 8, p. 615-624, juill. 1998, doi: 10.1016/S0301-4215(98)00019-6.
- [44] J.-N. Louis, A. Calo, K. Leiviskä, et E. Pongrácz, « Environmental Impacts and Benefits of Smart Home Automation: Life Cycle Assessment of Home Energy Management System », *IFAC-Pap.*, vol. 48, n° 1, p. 880-885, janv. 2015, doi: 10.1016/j.ifacol.2015.05.158.