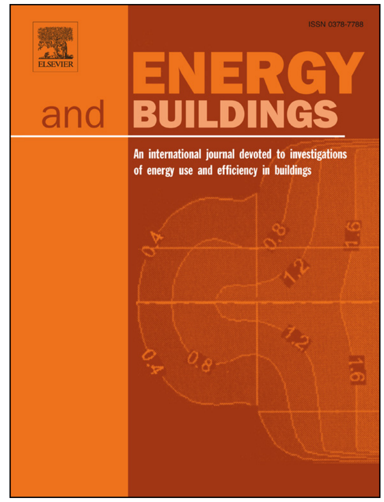


Calibration method for an open source model to simulate building energy at territorial scale

Rit Martin, Girard Robin, Villot Jonathan, Thorel Mathieu and Abdelouadoud Yassine

PII: S0378-7788(23)00435-8
DOI: https://doi.org/10.1016/j.enbuild.2023.113205
Reference: ENB 113205

To appear in: *Energy & Buildings*

Received date: 24 February 2023
Revised date: 5 May 2023
Accepted date: 25 May 2023

Calibration method for an open source model to simulate building energy at territorial scale

Rit Martin[1], Girard Robin[2], Villot Jonathan[3], Thorel Mathieu[1], Abdelouadoud Yassine[2]

Abstract

In a context of massive renovation of residential housing, stakeholders need decision-support tools based on knowledge of the current building stock and an accurate simulation of energy demand. For this purpose, we developed a validation/calibration method on a territorial/national scale in order to represent the real consumption of housing. This methodological approach provides (1) more reliable identification of energy-saving measures (changes in technology or behaviour) and (2) improved knowledge of the energy simulation tool and its post-calibration performance for optimisation issues. The main contribution of the calibration method described in this paper is the geographical scale concerned: all French residential housing has been modelled, simulated and calibrated with national data (geometries and attributes) on buildings. Furthermore, some occupants' socio-professional characteristics have been taken into account to reflect their actual energy behaviours. This is different from traditional approaches that focus only on a few buildings or archetypes. This paper also describes the application of this methodology on an Open Source simulation software in order to be easily verifiable and usable. This linear model will also be used to optimise renovation solutions at territorial scale in future

[1]Scientific and Technical Building Research Centre, 84 avenue Jean Jaurès, Champs sur Marne, 77447 Marne-la-Vallée Cedex 2, France
[2]Mines Paris, PSL University, Centre for processes, renewable energy and energy systems (PERSEE), 06904 Sophia Antipolis, France
[3]Mines Saint-Etienne, Univ Lyon, Univ Jean Moulin, Univ Lumière, Univ Jean Monnet, ENTPE, INSA Lyon, ENS Lyon, CNRS, UMR 5600 EVS, Institut Henri Fayol, Departement GEO, F - 42023 Saint-Etienne France

work. All data used in this paper are Open Data and thus available to the scientific community. This method enabled more than 18 million buildings to be calibrated while reducing the Normalized Root Mean Square Error, between simulated and real energy annual consumption, from 52% to 24% for gas and from 24% to 15% for electricity. In addition, the user of the method is free to prioritise either the maximum error reduction or the number of calibration coefficients if a simpler model is desired. This paper also discusses the results obtained from this method for future improvement.

Keywords: `energy consumption`, `calibration and validation`, `urban building energy modelling`, `linear model`, `territorial scale`

Nomenclature

ACR Air Change Rate

$AUSI$ Aggregated Units for Statistical Information

BEM Building Energy Modelling

EPC Energy Performance Certificate

GIS Geographic Information System

HDH Heating Degree Hour

$MILP$ Mixed-Integer Linear Programming

$NRMSE$ Normal Root Mean Square Error

$RMSE$ Root Mean Square Error

1. Introduction

1.1. Context

Final energy consumption comes mainly from three sectors: industry, transport and buildings. In 2019, the building sector accounted for about 21% of global energy consumption, split between residential (12.6%) and commercial and utility uses (8.1%) [1]. In the European Union, the building sector accounts for almost 28% of energy consumption [2]. Energy demand forecasting is a vital part of national planning studies [3]. Energy demand analysis and its dynamic variation are related to social, economic, technical and environmental aspects. Furthermore, energy saving and conservation potential for climate change mitigation are generally determined by analysing energy demand in different scenarios [4]. In France, the national housing stock represents 30% of final energy consumption [5]. As a result, the importance of improving the energy efficiency of the French housing stock has been recognised by the government in its National Low Carbon Strategy (2020), which outlines a strategy for reducing house-hold energy costs and carbon emissions from the residential sector, in order to meet the government's carbon budget targets in 2050 [6]. Designing prospective scenarios requires knowledge of the existing stock. This knowledge is now accessible via open databases.

With the opening of many data sets in France, it is now possible to characterise the building stock in a fairly detailed way. Indeed, since 2019, the French government has allowed the opening of data on the geometry of buildings with BD TOPO v3 [7] and on thermal and physical characterisation with the French Energy Performance Certificate (EPC) database [8].

When dealing with energy building consumption assessment/prediction, analyses can be estimated using physical models [9, 10]. In the specific context of energy building renovation tools, a physical model can assess different combination of renovation action at building scale and distinguish the energy use required for individual end-uses [11, 12, 13].

The aim of this research is to simulate the energy consumption of building stocks and leverage new open data sets that include information on real consumption, exhaustive building descriptions and EPCs. The majority of these data sets are used to characterize the buildings (surface, heating system, ...) in order to launch the different energy simulations. Only the data sets containing the real consumption are used to calibrate the energy model.

1.2. Necessity to use a fast linear model

Our building simulation model is a computer-based tool that allow the energy performance of buildings to be modeled and simulated and is based on mathematical equations and typically use linear programming techniques [14] to optimize building energy performance.

It represents the building and its energy systems as a set of equations that describe the energy flows, heat transfer, and other physical processes that occur within the building. In our case, we use a fast linear model presented in 3.1 for two main reasons :

- obtain quickly (a few hours) an energy simulation of all french residential buildings.
- optimize in a next step the retrofit solutions for the building's energy systems (such as heating and dhw systems, and ventilation) at territorial scale [15, 16, 17].

1.3. Contribution of the paper

The contribution of the paper are as follows:

- A physical bottom-up energy simulation model during the heating season (from 1 October to 20 May) that can cover the scale of several buildings to the whole of France and an associated evaluation of the 47450 French residential AUSI (Aggregated Units for Statistical Information which represent approximately 2,000 French dwellings) with open source observations.

- A statistical parameterisation in the model and an estimation procedure to exploit the open source observations.
- Different segmentations of the building stock and a model selection procedure in order to choose the best one to take into account the diversity of territories and buildings.
- An evaluation of the interest of the statistical parametrisation and the model selection procedure to reduce errors.

2. Literature review

Knowledge of the national building stock is essential and its energy simulation can be carried out taking two types of approach : bottom-up models [18, 19] and top-down models [20, 21]. Top-down models exploit the statistical relationship between energy consumption and available aggregated data, such as macro-economic data and population-weighted meteorological data [22]. They cannot therefore distinguish energy use for individual end-uses, and hence cannot be used to model the effect of technology changes. In contrast, the bottom-up approach relies on the combination of a description of each building and an energy simulation model.

Most bottom-up simulation models are deployed for hundreds of buildings, at the scale of a district (see e.g. [23, 24, 25]) or even thousands of building at the scale of a city like Melbourn [26], Zaragoza [27] or Chinese city [28]. In [29], the entire region of Seville was simulated with the aim of providing decision support for policy makers via energy demand and PV production. At the larger scale of a country, model usually rely on archetypes: hundreds of buildings chosen to represent the whole of Switzerland [30] or all of Europe [31].

Statistical methods are used in connection with bottom-up physical models to estimate additional features such as EPC data in Italian buildings [32, 33] or some more features to be fed into the physical model [34]. This kind of approach constitutes an hybridisation of physical models.

The comparison of bottom-up models with real data is necessary to understand the performance of the models. A litterature review in 2021 presents papers that have performed energy simulation calibrations [35].

At the scale of one building, there is a considerable literature on dynamic model calibration employing local measurements with various temporal resolutions (see e.g. the review paper [36]). In [31], the evaluation is performed on the aggregated total annual electricity consumption at country scale for 28 European countries, and in [37] the evaluation is performed based on annual energy estimation per usage at the scale of Japan.

At urban scale, some studies use real building data to perform physical model calibrations. Thus, in a Mediterranean climate, The projected consumption of 2311 buildings is adjusted using real metered data that's available, in order to fine-tune the model to match the occupancy and behavior patterns of users in Beirut [38]. One hundred and forty residential buildings in Kuwait City[39] were used as a case study to evaluate methods by measuring their yearly energy consumption. An other method was implemented in the residential building inventory of Bilbao city [40]. The findings of these studies demonstrate that it's feasible to accurately replicate the current space heating energy usage of a large-scale residential building inventory in a city, using readily available public data sources.

At territory scale, a bottom-up model [41] was performed on 42,000 buildings in the whole of Aosta Valley in Italy and compared to the EPC obtained at a supra-municipal scale and provided by Italian ministerial data. This work concludes that the EPC data tend to overestimate actual consumption, and that in addition their model overestimates EPC data by from 9% to 23%. The representation of building geometry in the study relies only on the ground surface description, and does not take into account e.g. adjacency, different building shapes, and shadowing effects. In 2015, statistical and engineering models were compared in around 1000 buildings in Rotterdam [42] to a database of gas consumption divided into postal code areas. In this study, the mean absolute error is 49% for the engineering methods, which justifies working on improving prediction models.

To date and based on this literature review [35], it seems that no study has been performed on a national scale taking a bottom-up approach, integrating the diversity of buildings and urban areas (countryside, suburbs, mountains, city centres, etc.) and making a comparison with real consumption data at a territory or city scale.

The data in [35] also revealed that urban-scale case studies were only conducted in the United States (54%), Europe (34%), and the Middle East (12%). Notably, none of the urban-scale studies were conducted in tropical regions. Considering that the urban context, including inter-building effects and the urban microclimate, is a crucial factor that must be taken into account in UBEMs [43], it would be intriguing to assess the effectiveness of UBEM calibration methodologies in tropical regions and cities beyond the United States and Europe.

To represent the reality as best as possible and to take into account the lack of information on the characterisation of the building stock, a data scientist needs to make a number of modelling assumptions and subjective judgments based on her/his prior experience and expertise. In order to minimise the difference between the simulation prediction and reality, the calibration efforts of a simulation model are essential [44, 45]. There is a very large body of literature about methods that match building energy simulation models with measured data for one building at a time [36]. These are based on manual, iterative interventions, special tests and analytical procedures or mathematical methods of calibration [46]. Most mathematical techniques employ some form of optimisation function to reduce the difference between measured and simulated data. An objective function may be used to set a target, for example, to minimise the mean square error between measured and simulated data [47, 48, 49]. However, no optimisation method studied so far allows the calibration of a very large number of buildings by comparing them with real data.

The main objective of this study is to design and develop a calibration method to be used for a large range of buildings and territories. To achieve this, we have presented variables to describe the buildings stock and developed a statistical model.

3. Physical model operation

The methodology developed in this research work is structured around 3 main steps:

- A **physical model** to obtain a first estimation of the yearly consumption of each building b: $\bar{C}_b$.
- A **statistical inference** on different databases to obtain building data to feed the physical model.
- A **statistical correction** $\hat{C}_b$ of the physical model to lever all consumption observations.

3.1. Physical model

The physical model can estimate, for any building b, and energy vector $v \in \{gas, electricity, biomass, oil, heating\ network\}$, the yearly energy consumption $\bar{C}_b^v$ as a sum over types of usages $t \in \{h, hw, su\}$ (h stands for heating, hw for domestic hot water and su for specific use in the case when v is electricity) is defined in Equation 1:

$$\bar{C}_b^v = \bar{C}_{h,b}^v + \bar{C}_{hw,b}^v + \bar{C}_{su,b} = [1,1,1]^T \bar{\mathbf{C}}_b^v, \quad \text{with } \bar{\mathbf{C}}_b^v = [\bar{C}_{h,b}^v, \bar{C}_{hw,b}^v, \bar{C}_{su,b}] \tag{1}$$

For any vector v and building b, $\bar{C}_{h,b}^v$ is obtained as the product of an efficiency η_b and the thermal need Q_b of each buildings as a combination of losses and gains:

$$Q_b = \sum_{w \in W} Q_{w,b}^t + Q_b^v - Q_b^s - \sum_{dw \in DW} Q_{occupant,dw} \tag{2}$$

where W is the set of exchange surfaces (walls, roof, floor, windows) and $Q_{w,b}^t$ associated transmission losses, Q_b^v is the ventilation losses due to ventilation and infiltration, Q_b^s is the annual solar gains, DW the set of dwellings and $Q_{occupant,dw}$ the thermal gains provided by the occupants of each dwelling. We denote by $HDH_{w,b}$ the heating degree hour at building b for exchange surface w.

Transmission (resp. ventilation) losses $Q_{w,b}^c$ (resp. Q_b^v) are assumed to be linear with the exchange surface (resp. volume) and $HDH_{w,b}$:

$$Q_{w,b}^t = HDH_{w,b} * A_{w,b} * U_{w,b} \quad Q_b^v = HDH_b * \frac{c_{air}}{3.6} * ACR_b * V_b \tag{3}$$

where ACR_b is the air change rate of building b.

We calculate the total solar gains Q_b^s as the sum of the transmitted radiation during the heating season (from 1 October to 20 May).

$$Q_b^s = \sum_{w \in Wb} A_w * (1 - \frac{AOI_w}{90})^5 * TF_w * DR_w \tag{4}$$

where A_w is the area of the window, AOI_w the angle of incidence of the sun on the window in degrees, TF_w the window transmission factor and DR_w the direct radiation if no solar mask is observed (otherwise the value is 0).

To summarize, this model allows to simulate the need for heating (primary and secondary), dhw, ventilation, cooking as well as specific needs. The cooling needs are a future development objective.

3.2. Building data inference

For each building b the physical model has to be fed with meteorological data (irradiation received by the building and outside temperature), and with building data : $DD_b = \{(U_{w,b}, A_{w,b})_{w \in W}, V_b, ACR_b, \eta_b\}$

For any building b in France, we obtain these data by an inference procedure using three databases summarized in Table 1. We can notice that these databases have different scales of information as well as often non-exhaustive data. The geometric characteristics of all buildings are obtained from BD Topo v3 [4] which contains 24 million buildings in France. For each building, the

[4]Databse BD TOPO : https://geoservices.ign.fr/bdtopo

base provides among other things the number of dwellings, the surface of the dwellings, the age of the building and the materials of the walls.

Source	Scale	Completude
BD TOPO	Building	Yes
INSEE	AUSI	Yes
EPC	Dwelling	No

Attributes	Source
AUSI code	All
Year of construction	All
Residential type	All
Heating system	INSEE & EPC
Altitude	BD TOPO
Height	BD TOPO
Wall and roof features	BD TOPO
Number of dwellings	BD TOPO
Surface area	INSEE
Number of dwellings	INSEE
Air change	EPC
Windows	EPC
Thermal features	EPC

Table 1: Database and attributes description used as input in the inference process. The aim of the inference is to have all features at building scale.

Other information characterizing the building is necessary to carry out the energy simulation and is referenced in the Table 1. We will try to estimate the information of the non-exhaustive bases from the available information and it is what we call inference. For example, the heating system will be inferred from the INSEE database [5] by relying on the common characteristics with the DPE database [6]. A geographical inference is carried out: we first look at the buildings with the same IRIS code, then in the same city before going to the department, regional or national level if no concordance is found.

[5]INSEE census :https://www.insee.fr/fr/information/2008354

[6]EPC databse : https://data.ademe.fr/datasets/dpe-france

Other information is deduced from the French database of energy performance certificates (DPE) using the same principle.

We agree the EPC are not exhaustive and real but in Open Data for the physical and thermal characterization of a building, the most complete data set remains this data set (more than 9 million EPC in France (40% of housing)). The EPC is not exhaustive, but the age, type, heating vector and location of the building allow us to infer the building's thermal characteristics and the efficiency of its heating system : $(U_{w,b})_w$, ACR_b, η_b. It is not possible to obtain certain characteristics for all buildings, and the aim of the inference procedure is to obtain a representative description of all buildings at the scale of a district (AUSI scale). Thus, for buildings without EPC information, we infered before the simulation some thermal features informations, heating system or ventilation system characteristics.

The corresponding model and data inference has an open source implementation named Building Model [50]. The complexity of this linear model is in line with the data source. The linearity allows for fast execution and for use in optimisation models such as [51].

3.3. Climate data

Building model uses weather data averaged over the past years to simulate buildings energetically. The format used is the EPW file from the One Building website [52]. A total of 122 weather files are used in France with data collected by field measurements and averaged between the years 2004 and 2018.

4. Calibration method

4.1. Statistical correction, a generic formulation

In many countries in Europe, a consumption observation $C_a = \sum_{b \in a} C_b$ is available for each AUSI $a \in \mathcal{A}$ (Aggregated Units for Statistical Information). With this data, large scale evaluation of any energy simulation model $\hat{C}^v$ is possible; the estimation error $\mathcal{E}_a^v$ for any AUSI a, and the total root mean square error (RMSE) are given by:

$$\mathcal{E}_a^v(\hat{C}^v) = C_a^v - \sum_{b \in a} \hat{C}_b^v \qquad RMSE^v(\hat{C}^v) = \sqrt{\frac{1}{|\mathcal{A}|}\sum_a \mathcal{E}^{v\,2}_a} \tag{5}$$

The Normalised RMSE (NRMSE) is the ratio of the RMSE and the total average consumption. It was performed on 40,660 AUSI for electricity and 18,608 for gas, employing available real data.

In this paper, we propose a general framework to enhance the physical model with a statistical correction of the form

$$\hat{C}_b^v = \alpha_b^T \bar{\mathbf{C}}_b^v + \beta_b^T X_b + \gamma_b \tag{6}$$

Explanatory variable $\bar{\mathbf{C}}_b^v$ is the three-elements vector from Equation 1, but it could contain for other models a more detailed breakdown of the building consumption simulation. X_b is a D-dimensional explanatory variable which can be used to add D other explanatory variables than physical ones (see in 4.3) in the calibration method.

For a given building b, the vector $\Theta_b = [\alpha_b, \beta_b, \gamma_b]$ is unknown and of size $3 + D + 1$. It has to be estimated from the AUSI aggregated consumption data. Since no observation is available at the building level it is not a simple regression problem but rather a source separation problem.

4.2. Partition of the set of buildings

To allow the model parameter to adapt to different kinds of building the vector Θ is assumed to depend on the building b through a partition $\mathcal{C}$ of the set of buildings.

There are many different way to form the building set partition e.g. using existing building archetypes from the literature [53, 54]. In this paper, we propose to use the auxiliary variables given in Table 2 to build several types of partitions and to let our algorithm identify the one that gives the best results (i.e. minimise the total root mean square error in Equation 5).

Type	Main heating energy	Use	Year of construction
Principal	Electricity	Individual	[1000,1945]
Secondary	Gas	Collective	[1946,1970]
	Biomass		[1971,1990]
	Oil		[1991,2005]
	Heating network		[2006,2100]

Table 2: Variables used to categorise the built environment

In addition, the ”Vacant” variable will remain a unique partition for the entire model due to its low number of elements. It is therefore necessary to add an element to the set of combinations made in Table 2.

A partition can be obtained according to one or several variables. For example $\mathcal{C} = \{\textbf{Type}\}$ will be the partition according to ”Type” values (with 3 elements), $\mathcal{C} = \{\textbf{Type} \otimes \textbf{Heating energy}\}$ will be the (2x5+1) elements partition with all combination of Types and Heating energy. The most complex partition that can be obtained from these variables is $\{\textbf{Type} \otimes \textbf{Heating energy} \otimes \textbf{Year} \otimes \textbf{Use}\}$ and it contains 101 (2x5x2x5+1) elements. The number of elements in $\mathcal{C}$ will be denoted $|\mathcal{C}|$. For all partitions, it is necessary to add a case for vacant buildings without any precision of the heating energy, use or year of construction. This adds a single element to all considered partitions. With all possible combinations of our 4 variables, this gives 14 possible partitions.

Some of the elements of a given partition can have very few buildings, and the associated parameter can then be difficult to estimate. For this reason, we introduce a procedure, that groups all small elements of a size lower than ρ, and a hyper-parameter to be chosen in 4.4, into a single ”garbage” element of the partition. The corresponding new partition is denoted $\mathcal{C}_\rho$. When $\rho = 0$, $\mathcal{C}_\rho$ tends to be similar to the initial partition, and when ρ is very large, $\mathcal{C}_\rho$ tends to have only one element. The set of all possible partitions, obtained by varying ρ along the 14 initial partitions, is denoted Γ.

Now that the model parameters depend on the buildings through categories, let us denote by $c(b)$ the category associated with building b, by $\bar{\mathbf{C}}_{c,a}$ (resp. $X_{c,a}$), the sum of corresponding explanatory variables along buildings from category c in AUSI a, and by $N_{c,a}$ the number of buildings from category c in AUSI a. If we use the notation $Y = (Y_c)_{c \in \mathcal{C}}$ for any variable Y indexed by c, the error in Equation 5 can be written

$$\begin{aligned}
\mathcal{E}_a^v(\hat{C}^v) &= C_a^v - \sum_{b \in a} \alpha_{c(b)}^T \bar{\mathbf{C}}_b^v + \beta_{c(b)}^T X_b + \gamma_{c(b)} && (7) \\
&= C_a^v - \sum_{c} (\alpha_c^T \bar{\mathbf{C}}_{c,a}^v + \beta_c^T X_{c,a} + \gamma_c N_{c,a}) && (8) \\
&= C_a^v - (\alpha^T \bar{\mathbf{C}}_a^v + \beta^T X_a + \gamma^T N_a) && (9)
\end{aligned}$$

In the initial problem we seek a $(3 + D + 1)$ dimensional statistical model to correct the physical model for each building. However, observations are only available at the AUSI scale. For a given partition $\mathcal{C}$, this sequence of equations allows us to transform the initial problem into a more convenient linear regression with $(3 + D + 1) * |\mathcal{C}|$ parameters with associated observations at the AUSI scale.

4.3. External explanatory variable

In addition to the result of the physical model, our statistical energy simulation model can use other explanatory variables. It is denoted X_b at the building level in Equation 6 but, as shown in Equation 7, we only need X_a at the AUSI scale. In this section, we describe how this part of the model is used to take into account occupants' behaviour.

There is a consensus [55] [56] that income, energy price, number of occupants, age of the reference person, professional status, and individual preferences play an important role in explaining the variability of energy consumption.

According to statistical studies [57] [58] carried out on the French building stock, the type of occupation of the building (owner or tenant), the professional activity (employed, unemployed or retired) and the financial income of households are the sociological characteristics that have the greatest influence on the energy consumption of residential buildings. Two data sets are available for these features as open data for the French building stock at the AUSI level and are listed in Table 3. As one of the priority of this study is to be reusable, only open database will be add in the model. For that reason, the poverty rate and the pensioner rate will therefore be used to describe the occupants of dwellings in this study.

Database	Year used	Description
FILOSOFI	2018	Poverty rate at 60% Ω_a^{60} of median disposable income per consumption unit in metropolitan France (%)
INSEE	2018	Pensioner rate Δ_a (%)

Table 3: Data description of occupants for all AUSI a

We take into account the occupant data defined previously by adding an additional heating consumption related to either the percentage of pensioners or the percentage of poverty. Thus, a heating consumption term is added as follows:

$$X_a^{INSEE} = \bar{C}_{t,h,c}^{v} * \Delta_a, \quad X_a^{FILO} = \bar{C}_{t,h,c}^{v} * \Omega_a^{60} \tag{10}$$

In the case when both pensioner rate and poverty rate are used, the X variable is a concatenation of X_a^{INSEE} and X_a^{FILO}. In this case the dimension D of X is 2.

4.4. Parameter estimation algorithm

For a given partition $\mathcal{C}$, the model parameter Θ has $(3 + D + 1) * |\mathcal{C}|$ components that we seek to recover with all AUSI consumption observations (several thousands in France). Equation 7 shows that it can be reformulated as a linear regression model; if $\hat{C}^v(\Theta)$ is the consumption model obtained with parameter $\Theta = [\alpha, \beta, \gamma]$, the RMSE can be rewritten as

$$RMSE^v(\hat{C}^v(\Theta)) = \|C - (\alpha^T \bar{\mathbf{C}} + \beta^T X + \gamma^T N)\|_2 \quad (11)$$

We have to choose among different models. First, there are many different possible partitions in Γ : 14 possible choices for the initial partition, and more depending on the choice of hyper-parameter ρ. Second, our model can integrate different types of explanatory variables (e.g. different choices for X). We want our algorithm to choose the best model automatically. Depending on the number of parameters in a given model, there is a risk of overfitting. For this reason, we propose to use an estimation procedure with a LASSO penalty [59]. :

$$\hat{\Theta}_\lambda = \text{Argmin}_{\mathcal{C} \in \Gamma, \Theta} \ \|C - (\alpha^T \bar{\mathbf{C}} + \beta^T X + \gamma^T N)\|_2 + \lambda \|\Theta\|_1 \quad (12)$$

The error presented in the results is calculated by K-fold cross-validation procedure [60] [61] (we took $K = 10$ in our numerical experiment) to take into account the risk of overfitting. Only the λ hyperparameter is optimized on the whole dataset.

This procedure selects the best model and estimate the associated parameters. This means that an optimal model and an associated optimal partition are found as by-products of our procedure.

5. Results

5.1. Case study and performances of initial model

We demonstrate our algorithm in the case of France. France is composed of around 47,450 residential AUSI (named IRIS in France). In France, for the gas [62] and electricity vectors [63], yearly observations of residential energy consumption are made available by Distribution System Operators (DSO) for thousands of AUSI.

A mean of all Open Data accessible have been considered : 2018 and 2019 for gas consumption, 2015 to 2019 for electricity consumption.

For any AUSI a and vector v, the normalised deviation between simulated and observed consumption is given by:

$$E_a^v(\hat{C}^v) = \frac{\mathcal{E}_a^v(\hat{C}^v)}{C_a^v} * 100, \tag{13}$$

In order to obtain comprehensive results, we also define a normalised RMSE.

$$NRMSE^v(\hat{C}^v) = \frac{RMSE^v(\hat{C}^v)}{\frac{1}{\mathcal{A}}\sum_{a\in\mathcal{A}} C_a^v} \tag{14}$$

In all results we have removed AUSI with extreme errors (the 5% of AUSI with the largest errors).

As discussed in the inference section 3.2, the construction of the building stock gives 2 different results for 2 different simulations. To limit these uncertainties, we carried out different simulations of the building stock. Electricity and gas consumption were averaged for each AUSI over all of these simulations for each partition $\mathcal{C}$ of the set of buildings. After many tests, we found that the average error did not change after 6 sets of simulations. Thus, all of the results presented in this article were presented with 6 sets of simulated buildings.

The performance of the initial physical model proposed in section 3.1 is shown in Table 4 for electricity and gas consumption. The error is low compared to those presented in the literature review in section 2.

	Electricity	**Gas**
Mean error	-3.35 %	-7.28 %
Initial NRMSE	23.86 %	52.35 %
Minimum deviation error	-94.87 %	-94.76 %
Maximum deviation error	99.90 %	99.92 %
Standard deviation error	0.019	0.18
AUSI simulated	40 925	19 147

Table 4: Simulation results of Building Model on all French buildings

Figure 1 shows the distribution of the corresponding errors. It confirms the overestimation of the model. This overestimation is not a surprise since this model is similar to an EPC model, which is built to estimate conventional consumption, rather than consumers' behaviour and real consumption. The error distribution is also slightly skewed. Bias, normalised error, standard deviations are larger for gas than for electricity.

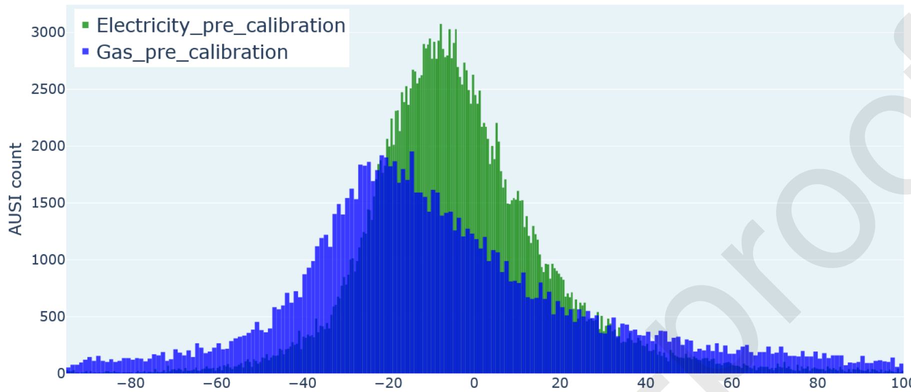


Figure 1: Normalised deviation distribution for electricity and gas (%)

The geographical results obtained for this initial physical model can be seen in Figure 7a for electricity.

There are large areas with no data, which cane be explained by two reasons:

- The lack of GIS data on certain characteristics (such as the number of dwellings in each building), which is essential for the energy simulation.
- The lack of energy consumption data. Indeed, some French departments do not depend on the state organisations ENEDIS (the main electricity provider company) or GRDF (the national gas provider company).

The differences between the different French territories are also quite significant. Indeed, for the north-west and most of the central territories of France, the Building Model overestimates energy consumption. The presence of rural territories with high biomass consumption could be the cause. These buildings would have been identified as consuming electricity instead of biomass heating. On the other hand, warm regions like the Mediterranean basin or the South Atlantic coast are underestimated. This can be explained by the fact that air conditioning is not taken into account in the simulation model and that there may be more second homes than those identified in our model.

5.2. Calibration results

We now present the results obtained from the calibration method developed in section 4.1. As a global result, the calibration method reduced the initial error from 52% to 24% for gas consumption and from 24% to 15% for electricity consumption. Figures 2 and 3 are presented for one climatic zone, the H1B climate zone (described in Annex in Figure 8). These figures show the results of the best models created for electricity and gas. Thus, the models presented are those with the lowest NRMSE for a partition containing 1 to 4 variables. The 95% confidence interval presented as vertical bars below comes from the 10-fold crossed validation. Our optimal model and our associated optimal partition are found with hyper-parameter $\rho\_gas = 325$, $\lambda\_gas = 0.2$, $\rho\_elec = 5186$, and $\lambda\_elec = 0.1$.

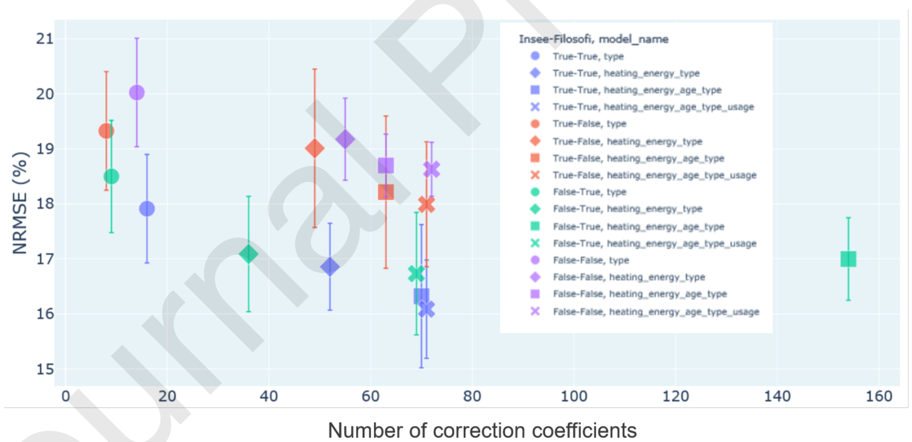


Figure 2: Number of coefficients obtained by lasso method in electricity calibration

This means that the variables chosen allow the simulated building stock to be efficiently segmented. The influence of complementary variables containing information on the occupants is also important and allows in each case to improve the post calibration error. It also appears that the age of the occupants (in

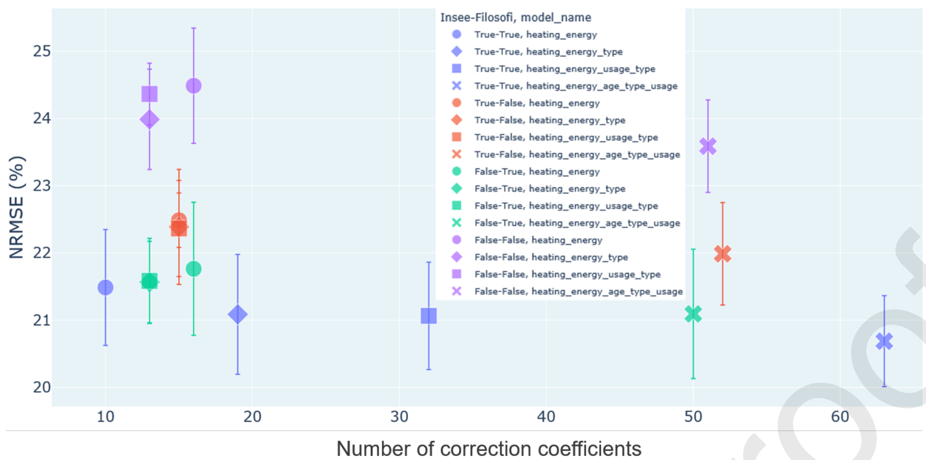


Figure 3: Number of coefficients obtained by lasso method in gas calibration

our case the presence or absence of pensioners in the INSEE database) reduces the NRMSE more than the poverty rate in the AUSI considered (FILOSOFI database). The use of the INSEE database allows a 2% improvement of NRMSE on average, contrary to the FILOSOFI database, which reduces it by 1% on average.

An important result is the number of parameters needed to describe the calibration problem. Indeed, as we can see in figure 2, models with a higher NRMSE reduction generally have more descriptive coefficients. However, we note that it is possible to optimise the number of descriptive coefficients if reducing the NRMSE is not the priority. Indeed, between the model with the least and the most coefficients, a 2% reduction of the NRMSE is observed for 10 times fewer descriptive coefficients. Thus, it is possible to calibrate the consumption model with only 7 coefficients and obtain a final NRMSE of 18.5% for electricity consumption for instance.

In our case, the objective is to obtain the lowest possible NRMSE. We have therefore chosen for each climate zone and each energy vector, the partition allowing to obtain it either $\mathcal{C} = \{\textbf{Type} \otimes \textbf{Heating energy} \otimes \textbf{Year} \otimes \textbf{Use}\}$

For each considered segmentation of the building stock, corrective coefficients will be applied to the final consumption of electricity or gas to increase or decrease it. The analysis of these coefficients presented in figure 4 shows that few categories of dwellings are subject to real changes in consumption. The major consumption changes are concentrated on the main electricity consumption (heating).

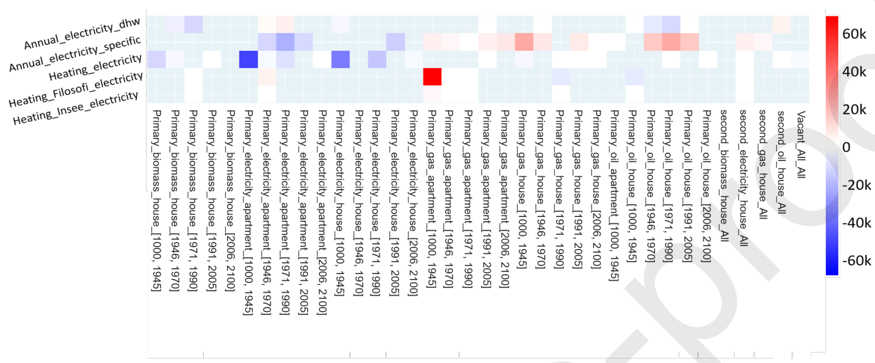


Figure 4: Change in consumption (kWh) post electricity calibration by fleet segmentation

The calibration of all French simulated residential buildings considerably reduce the global error and thus refines the energy simulation. The results of the calibration method for each climate zone (described in Annex) in figure 5 with its standard error deviation. The reduction of the NRMSE is broadly similar for all climate zones for electricity and gas consumption.

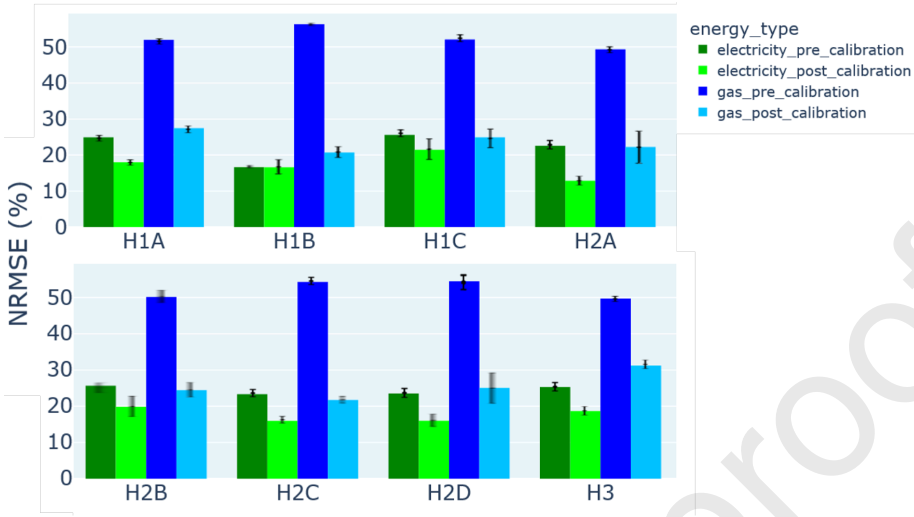


Figure 5: NRMSE results post calibration for each climate zone (Annex Figure 8)

For the whole France, Figure 6 shows the refocusing of the simulations around actual electricity and gas consumption data.

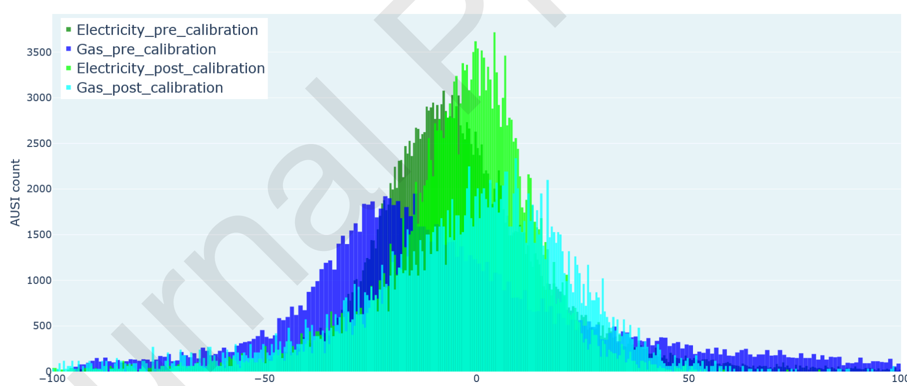
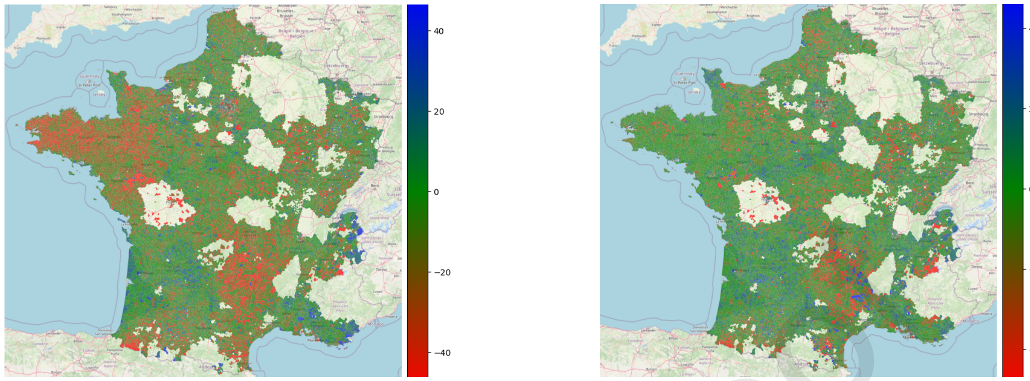


Figure 6: Comparison of consumption normalised deviation pre/post calibration

There is a clear improvement in the prediction of consumption thanks to the calibration and a strong reduction in the differences in predictions in the north-west and centre of France as we can see on the Map 7b.

However, there is some debate concerning the simulation assumptions used in the software and in the calibration method in order to explain the difference



(a) Deviation from electricity data (%)

(b) Deviation from electricity data post calibration (%)

Figure 7: Comparison map of France pre/post calibration

between simulated and predicted data.

6. Discussion

6.1. Simulation assumptions

Simulation assumptions were made to simplify the calculation time for the 18 million buildings. For example, the temperature was set at 19°C, considering the dwellings as a single thermal zone, and the heating period was from 1 October to 20 May. Other simulation assumptions would have led to different results and comparisons of local energy data would also have differed.

Some physical parameters are not taken into account for now : cooling, relative humidity and inhabitants schedules. Cooling for instance will have a strong impact on energy consumption and the calibration method should be done again.

In order to improve the simulation software and to better represent reality, future developments are planned. In addition, adapting heating periods according to the climatic region can have an impact and could be studied. In addition, state of maintenance and loss of efficiency are not directly taken into account. However, with the EPC database, a building with an old year of construction will have a higher wall or window 'U-value' and so a higher energy consumption.

Furthermore, in France, we don't have for now Open Data on previous retrofits in the buildings.

Similarly for cooling, we therefore planned to take into account the need for cooling in our model but the same problem of access to the cooling system presence data arises to simulate this consumption and it represents in 2020 3.3% of the annual energy consumption of the french residentiel sector [64] despite local difference. Moreover, with climate change and inhabitants' higher demands in terms of comfort, the use of air conditioning will increase in the coming years in France [64].

6.2. Calibration assumptions

The calibration model used could be improved by using other scientific methods. Indeed, neural networks could certainly be used. However, because the study conducted does not follow a standard statistical framework with observation of aggregates of results, this complicates the use of other methods.

For the calibration method, different assumptions have also been made which necessarily change the final results. The description of the building stock was carried out in the calibration method (in 4.2) by a partition obtained according to one or several variables. The choice of these variables was based on our expertise and the importance of these data. However, it would be interesting to test other variables such as the surface area of the dwellings, and the EPC energy class, and see whether they lead to an improvement of the calibration method.

One of these explanatory variables is the type of heating system. This corresponds to the main heating system in the dwellings but does not represent the great diversity of heating system present in the built-up areas today. In fact, many rural houses have 2 heating systems (e.g. gas + wood fuel) but can only be calibrated for gas and electricity. This necessarily distorts the calibration method.

In addition, other sociological variables than the rate of retirement or poverty could influence the results as indicated in 4.3. However, it can already be noted that only a few NRMSE points were gained using these variables considered among the most influential for energy consumption [57, 58]. Thus, it is likely that adding other variables would have a small impact on the final result.

It will be considered in future renovation works based on this calibration that the current building stock covers enough configurations to use the coefficients obtained on the new renovation solutions. In order to do without this assumption, a calibration model that allows for modifications of the physical input parameters would be a desirable future development.

7. Conclusion

To conclude, this paper contributes to presenting a physical bottom-up model and to evaluating its performance on the French territory. In a second step, a statistical parametrisation to reduce the error was performed.

The bottum-up approach makes it possible to model at the scale of buildings and not to use archetypes. The building data is realized with Open data bases in France and is easily replicable in the simulation model [50]. This simple and fast simulation model is used to simulate all french residential buildings.

The calibration methodology is aloso based solely on open data to allow the entire scientific community to use this process. The methodology described in this paper allows to describe the main steps to realize this calibration and the source code can be requested to realize it.

However, the ways in which the simulation model could be improved are numerous. Testing building archetypes, in comparison with current data inference processes, should bring knowledge on model behaviour. Thus, the model could be used in its present form for a specific geographical area to help public decision makers. In the private domain, it could prove useful to real estate asset managers such as national train company or national hotel groups with no specific geographic unit.

Furthermore, the calibration method could also be improved. As an example, it would be useful to try to estimate the physical parameters rather than having some form of statistical post-processing.

The calibration method is fully open. Indeed, all of the data used as well as the usage model are open data and open source. Thus, this method and the associated results can be used again in the Building Model or any other simulation software.

In future work, this calibrated model will be used in a decision support tool for the renovation of residential buildings on a territorial scale to continue the work presented in [15]. The interest of having calibrated it on real data is that the real consumption of buildings before renovation corresponds more closely to reality, and the proposed solutions will thus be more relevant.

References

[1] World energy balances: Overview 2021 (2021).
URL https://www.iea.org/reports/world-energy-balances-overview

[2] Statistical office of the european union (eurostat). final energy consumption in the households sector, eu-28 (2018).
URL https://www.eea.europa.eu/data-and-maps/daviz/total-final-energy-consumption-by-sector-4#tab-googlechartid_chart_41

[3] M. M. Kh A.H. Fakehi, S. Ahmadi, The impact of energy conservation policies on the projection of future energy demand, Energy Technol. Policy 2 (2015) 104–121doi:10.1080/23317000.2015.1068140.

[4] A. Hui, G. Walker, Concepts and methodologies for a new relational geography of energy demand: Social practices, doing-places and settings 36 21–29. doi:10.1016/j.erss.2017.09.032.
URL https://linkinghub.elsevier.com/retrieve/pii/S221462961730316X

[5] Ministère de la Transition Ecologique et Solidaire, Bilan énergétique de la france pour 2020 (2020).
URL https://www.statistiques.developpement-durable.gouv.fr/bilan-energetique-de-la-france-pour-2020

[6] M. for the ecological {and} solidary transition, National low carbon strategy.
URL https://www.ecologie.gouv.fr/sites/default/files/en_SNBC-2_complete.pdf

[7] Ign database.
URL https://geoservices.ign.fr/bdtopo

[8] Dpe database.
URL https://data.ademe.fr/datasets/dpe-france

[9] L. Le-Thanh, H. Nguyen-Thi-Viet, J. Lee, H. Nguyen-Xuan, Machine learning-based real-time daylight analysis in buildings 52 104374. doi:10.1016/j.jobe.2022.104374.
URL https://linkinghub.elsevier.com/retrieve/pii/S2352710222003874

[10] Y. Ding, L. Fan, X. Liu, Analysis of feature matrix in machine learning algorithms to predict energy consumption of public buildings 249 111208. doi:10.1016/j.enbuild.2021.111208.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378778821004928

[11] M. Panagiotidou, L. Aye, B. Rismanchi, Optimisation of multi-residential building retrofit, cost-optimal and net-zero emission targets 252 111385. doi:10.1016/j.enbuild.2021.111385.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378778821006691

[12] Y. He, N. Liao, J. Bi, L. Guo, Investment decision-making optimization of energy efficiency retrofit measures in multiple buildings under financing budgetary restraint 215 1078–1094. doi:10.1016/j.jclepro.2019.01.119.
URL https://linkinghub.elsevier.com/retrieve/pii/S0959652619301362

[13] M. Tavakolan, F. Mostafazadeh, S. Jalilzadeh Eirdmousa, A. Safari, K. Mirzaei, A parallel computing simulation-based multi-objective optimization framework for economic analysis of building energy retrofit: A case study in iran 45 103485. doi:10.1016/j.jobe.2021.103485.
URL https://linkinghub.elsevier.com/retrieve/pii/S2352710221013437

[14] A. Schrijver, Theory of linear and integer programming, John Wiley & Sons, 1998.

URL https://books.google.fr/books/about/Theory_of_Linear_and_Integer_Programming.html?id=zEzW5mhppB8C&redir_esc=y

[15] A. Rogeau, R. Girard, Y. Abdelouadoud, M. Thorel, G. Kariniotakis, Joint optimization of building-envelope and heating-system retrofits at territory scale to enhance decision-aiding 264 114639. doi:10.1016/j.apenergy.2020.114639.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261920301513

[16] D. Zhang, S. Evangelisti, P. Lettieri, L. G. Papageorgiou, Optimal design of chp-based microgrids: Multiobjective optimisation and life cycle assessment, Energy 85 (2015) 181–193. doi:https://doi.org/10.1016/j.energy.2015.03.036.
URL https://www.sciencedirect.com/science/article/pii/S0360544215003308

[17] C. Weber, N. Shah, Optimisation based design of a district energy system for an eco-town in the united kingdom, Energy 36 (2) (2011) 1292–1308. doi:https://doi.org/10.1016/j.energy.2010.11.014.
URL https://www.sciencedirect.com/science/article/pii/S0360544210006407

[18] 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, Building and Environment 45 (7) (2010) 1683–1697. doi:https://doi.org/10.1016/j.buildenv.2010.01.021.
URL https://www.sciencedirect.com/science/article/pii/S0360132310000338

[19] Y.J. Huang, J. Brodrick, A bottom-up engineering estimate of the aggregate heating and cooling loads of the entire u.s. building stock, 2000 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove,

CA (2000).
URL https://www.researchgate.net/publication/237253593_A_Bottom-Up_Engineering_Estimate_of_the_Aggregate_Heating_and_Cooling_Loads_of_the_Entire_US_Building_Stock

[20] K. Fabbri, M. Zuppiroli, K. Ambrogio, Heritage buildings and energy performance: Mapping with GIS tools 48 137–145. doi:10.1016/j.enbuild.2012.01.018.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378778812000333

[21] Y. Yamaguchi, Y. Shimoda, Database and simulation model development for modelling the energy use of non-residential buildings.
URL https://www.aivc.org/sites/default/files/BS09_1482_1489.pdf

[22] S. Heiple, D. J. Sailor, Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles 40 (8) 1426–1436. doi:10.1016/j.enbuild.2008.01.005.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378778808000200

[23] E. Garreau, Y. Abdelouadoud, E. Herrera, W. Keilholz, G.-E. Kyriakodis, V. Partenay, P. Riederer, District MOdeller and SIMulator (DIMOSIM) – a dynamic simulation platform based on a bottom-up approach for district and territory energetic assessment 251 111354. doi:10.1016/j.enbuild.2021.111354.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378778821006381

[24] L. Diao, Y. Sun, Z. Chen, J. Chen, Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation 147 47–66.

doi:10.1016/j.enbuild.2017.04.072.
URL https://linkinghub.elsevier.com/retrieve/pii/S037877881731469X

[25] P. Nageler, G. Zahrer, R. Heimrath, T. Mach, F. Mauthner, I. Leusbrock, H. Schranzhofer, C. Hochenauer, Novel validated method for GIS based automated dynamic urban building energy simulations 139 142–154. doi:10.1016/j.energy.2017.07.151.
URL https://linkinghub.elsevier.com/retrieve/pii/S036054421731335X

[26] S. Li, B. Rismanchi, L. Aye, A simulation-based bottom-up approach for analysing the evolution of residential buildings' material stocks and environmental impacts – a case study of inner melbourne 314 118941. doi:10.1016/j.apenergy.2022.118941.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261922003609

[27] C. J. García-Ballano, A. Ruiz-Varona, L. Casas-Villarreal, Parametric-based and automatized GIS application to calculate energy savings of the building envelope in rehabilitated nearly zero energy buildings (nZEB). case study of zaragoza, spain 215 109922. doi:10.1016/j.enbuild.2020.109922.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378778819332888

[28] T. Huo, L. Xu, W. Feng, W. Cai, B. Liu, Dynamic scenario simulations of carbon emission peak in china's city-scale urban residential building sector through 2050 159 112612. doi:10.1016/j.enpol.2021.112612.
URL https://linkinghub.elsevier.com/retrieve/pii/S030142152100478X

[29] P. E. Camporeale, P. Mercader-Moyano, A GIS-based methodology to increase energy flexibility in building cluster through

deep renovation: A neighborhood in seville 231 110573. doi:10.1016/j.enbuild.2020.110573.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378778820333594

[30] K. N. Streicher, P. Padey, D. Parra, M. C. Bürer, S. Schneider, M. K. Patel, Analysis of space heating demand in the swiss residential building stock: Element-based bottom-up model of archetype buildings 184 300–322. doi:10.1016/j.enbuild.2018.12.011.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378778818325660

[31] T. Gulotta, M. Cellura, F. Guarino, S. Longo, A bottom-up harmonized energy-environmental models for europe (BOHEEME): A case study on the thermal insulation of the EU-28 building stock 231 110584. doi:10.1016/j.enbuild.2020.110584.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378778820333703

[32] K. G. Droutsa, S. Kontoyiannidis, E. G. Dascalaki, C. A. Balaras, Mapping the energy performance of hellenic residential buildings from EPC (energy performance certificate) data 98 284–295. doi:10.1016/j.energy.2015.12.137.
URL https://linkinghub.elsevier.com/retrieve/pii/S0360544216000050

[33] F. Pagliaro, F. Hugony, F. Zanghirella, R. Basili, M. Misceo, L. Colasuonno, V. Del Fatto, Assessing building energy performance and energy policy impact through the combined analysis of EPC data – the italian case study of SIAPE 159 112609. doi:10.1016/j.enpol.2021.112609.
URL https://linkinghub.elsevier.com/retrieve/pii/S0301421521004754

[34] P. Schetelat, L. Lefort, N. Delgado, Urban data imputation using multi-output multi-class classification, in: Proceedings of uSim Conference 2020: 2nd uSim Conference of IBPSA-Scotland, Vol. 2 of uSim Conference, IBPSA-Scotland, Edinburgh, Scotland, 2020, pp. 126–133.
URL https://publications.ibpsa.org/conference/paper/?id=usim2020_B1_3_Schetelat

[35] A. Chong, Y. Gu, H. Jia, Calibrating building energy simulation models: A review of the basics to guide future work, Energy and Buildings 253 (2021) 111533. doi:https://doi.org/10.1016/j.enbuild.2021.111533.
URL https://www.sciencedirect.com/science/article/pii/S0378778821008173

[36] D. Coakley, P. Raftery, M. Keane, A review of methods to match building energy simulation models to measured data 37 123–141. doi:10.1016/j.rser.2014.05.007.
URL https://linkinghub.elsevier.com/retrieve/pii/S1364032114003232

[37] A. Taniguchi-Matsuoka, Y. Shimoda, M. Sugiyama, Y. Kurokawa, H. Matoba, T. Yamasaki, T. Morikuni, Y. Yamaguchi, Evaluating japan's national greenhouse gas reduction policy using a bottom-up residential end-use energy simulation model 279 115792. doi:10.1016/j.apenergy.2020.115792.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261920312757

[38] A. Krayem, A. Al Bitar, A. Ahmad, G. Faour, J.-P. Gastellu-Etchegorry, I. Lakkis, J. Gerard, H. Zaraket, A. Yeretzian, S. Najem, Urban energy modeling and calibration of a coastal mediterranean city: The case of beirut, Energy and Buildings 199 (2019) 223–234. doi:https://doi.org/10.1016/j.enbuild.2019.06.050.

URL https://www.sciencedirect.com/science/article/pii/S0378778819312460

[39] C. Cerezo, J. Sokol, C. Reinhart, A. Al-Mumin, THREE METHODS FOR CHARACTERIZING BUILDING ARCHETYPES IN URBAN ENERGY SIMULATION. a CASE STUDY IN KUWAIT CITY.

[40] J. Fernandez, L. del Portillo, I. Flores, A novel residential heating consumption characterisation approach at city level from available public data: Description and case study, Energy and Buildings 221 (2020) 110082. doi:https://doi.org/10.1016/j.enbuild.2020.110082.
URL https://www.sciencedirect.com/science/article/pii/S0378778819331512

[41] V. D'Alonzo, A. Novelli, R. Vaccaro, D. Vettorato, R. Albatici, C. Diamantini, P. Zambelli, A bottom-up spatially explicit methodology to estimate the space heating demand of the building stock at regional scale 206 109581. doi:10.1016/j.enbuild.2019.109581.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378778819306978

[42] R. Nouvel, A. Mastrucci, U. Leopold, O. Baume, V. Coors, U. Eicker, Combining GIS-based statistical and engineering urban heat consumption models: Towards a new framework for multi-scale policy support 107 204–212. doi:10.1016/j.enbuild.2015.08.021.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378778815302061

[43] T. Hong, Y. Chen, X. Luo, N. Luo, S. H. Lee, Ten questions on urban building energy modeling, Building and Environment 168 (2020) 106508. doi:https://doi.org/10.1016/j.buildenv.2019.106508.
URL https://www.sciencedirect.com/science/article/pii/S0360132319307206

[44] T. Reddy, Literature review on calibration of building energy simulation programs: Uses, problems, procedure, uncertainty, and tools, in: ASHRAE Transactions, Vol. 112 PART 1, 2006, pp. 226–240, 2006 Winter Meeting of the American Society of Heating, Refrigerating and Air-Conditioning Engineers, ASHRAE ; Conference date: 21-01-2006 Through 25-01-2006.
URL https://asu.pure.elsevier.com/en/publications/literature-review-on-calibration-of-building-energy-simulation-pr

[45] J. A. Clarke, P. A. Strachan, C. Pernot, AN APPROACH TO THE CALIBRATION OF BUILDING ENERGY SIMULATION MODELS 13.
URL https://www.esru.strath.ac.uk/Documents/93/calibration.pdf

[46] T. A. Reddy, I. Maor, C. Panjapornpon, Calibrating detailed building energy simulation programs with measured data—part i: General methodology (RP-1051) 13 (2) 221–241, publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/10789669.2007.10390952. doi:10.1080/10789669.2007.10390952.
URL https://www.tandfonline.com/doi/abs/10.1080/10789669.2007.10390952

[47] W. Carroll, R. Hitchcock, Tuning simulated building descriptions to match actual utility data: Methods and implementation, Vol. 99, 1993, p. 928 – 934, cited by: 46.
URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-0027812532&partnerID=40&md5=29d777b111cbf43075ac19c96326f7d8

[48] J. Sun, T. A. Reddy, Calibration of building energy simulation programs using the analytic optimization approach (rp-1051), HVAC&R Research 12 (1) (2006) 177–196. arXiv:https://www.tandfonline.com/doi/pdf/10.1080/10789669.2006.10391173, doi:10.1080/10789669.2006.10391173.

URL https://www.tandfonline.com/doi/abs/10.1080/10789669.2006.10391173

[49] T. A. Reddy, I. Maor, C. Panjapornpon, Calibrating detailed building energy simulation programs with measured data—part i: General methodology (rp-1051), HVAC&R Research 13 (2) (2007) 221–241. arXiv:https://www.tandfonline.com/doi/pdf/10.1080/10789669.2007.10390952, doi:10.1080/10789669.2007.10390952.
URL https://www.tandfonline.com/doi/abs/10.1080/10789669.2007.10390952

[50] Buildingmodel git.
URL https://gitlab.com/energytransition/buildingmodel

[51] M. J. Risbeck, C. T. Maravelias, J. B. Rawlings, R. D. Turney, A mixed-integer linear programming model for real-time cost optimization of building heating, ventilation, and air conditioning equipment 142 220–235. doi:10.1016/j.enbuild.2017.02.053.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378778817306473

[52] Climate data.
URL https://www.climate.onebuilding.org/WMO_Region_6_Europe/FRA_France/index.html

[53] P. Borges, O. Travesset-Baro, A. Pages-Ramon, Hybrid approach to representative building archetypes development for urban models – a case study in andorra 215 108958. doi:10.1016/j.buildenv.2022.108958.
URL https://linkinghub.elsevier.com/retrieve/pii/S0360132322002001

[54] J. Palmer Real, J. K. Møller, R. Li, H. Madsen, A data-driven framework for characterising building archetypes: A mixed effects modelling approach 254 124278. doi:10.1016/j.energy.2022.124278.

URL https://linkinghub.elsevier.com/retrieve/pii/S0360544222011811

[55] S. Bakaloglou, D. Charlier, The role of individual preferences in explaining the energy performance gap 104 105611. doi:10.1016/j.eneco.2021.105611.
URL https://linkinghub.elsevier.com/retrieve/pii/S0140988321004771

[56] D. Brounen, N. Kok, J. M. Quigley, Energy literacy, awareness, and conservation behavior of residential households 38 42–50. doi:10.1016/j.eneco.2013.02.008.
URL https://linkinghub.elsevier.com/retrieve/pii/S0140988313000327

[57] F. Belaïd, Understanding the spectrum of domestic energy consumption: Empirical evidence from france 92 220–233. doi:10.1016/j.enpol.2016.02.015.
URL https://linkinghub.elsevier.com/retrieve/pii/S030142151630057X

[58] F. Belaïd, T. Garcia, Understanding the spectrum of residential energy-saving behaviours: French evidence using disaggregated data 57 204–214. doi:10.1016/j.eneco.2016.05.006.
URL https://linkinghub.elsevier.com/retrieve/pii/S0140988316301244

[59] R. Tibshirani, Regression shrinkage and selection via the lasso 58 (1) 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.
URL https://onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1996.tb02080.x

[60] M. B. Lyons, D. A. Keith, S. R. Phinn, T. J. Mason, J. Elith, A comparison of resampling methods for remote sensing classification and

accuracy assessment, Remote Sensing of Environment 208 (2018) 145–153. doi:https://doi.org/10.1016/j.rse.2018.02.026.
URL https://www.sciencedirect.com/science/article/pii/S0034425718300324

[61] M. W. Browne, Cross-validation methods 44 (1) 108–132. doi:10.1006/jmps.1999.1279.
URL https://linkinghub.elsevier.com/retrieve/pii/S0022249699912798

[62] Grdf database.
URL https://opendata.grdf.fr/explore/dataset/consommation-annuelle-de-gaz-par-iris-et-code-naf/information/

[63] Enedis database.
URL https://data.enedis.fr/explore/dataset/consommation-electrique-par-secteur-dactivite-iris/information/

[64] ADEME, La climatisation de confort dans les bâtiments résidentiels et tertiaires 13.
URL https://librairie.ademe.fr/air-et-bruit/4745-la-climatisation-de-confort-dans-les-batiments-residentiels-et-tertiaires.html

8. Annex

8.1. Annex 1 : Description of French climate zones

For this study, we used the French climate zones shown in the figure below.

Figure 8: Climate Zone

Declaration of interests

☐ 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.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

RIT Martin reports was provided by Scientific and Technical Building Research Centre.