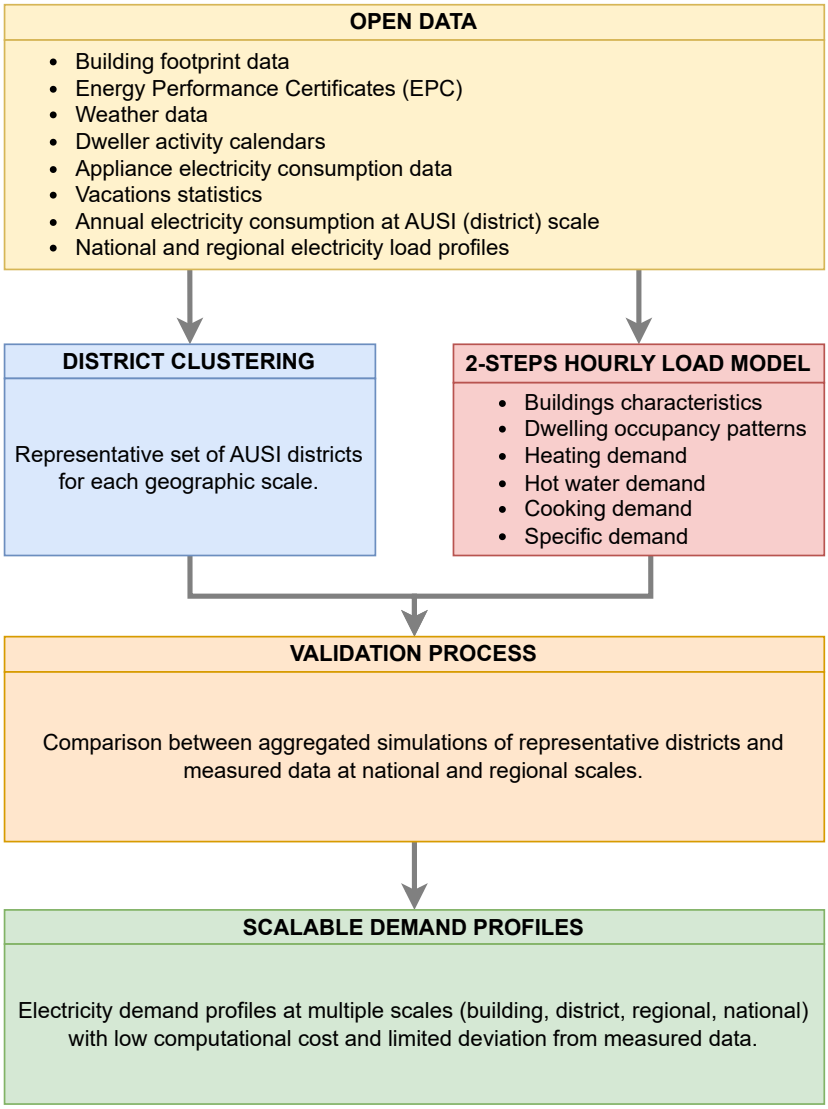
A scalable activity-based and physics-informed bottom-up model for residential electricity demand using open data

Yoann Chiche, Seddik Yassine Abdelouadoud, Anna Cocchi, Robin Girard

Graphical Abstract

A scalable activity-based and physics-informed bottom-up model for residential electricity demand using open data

Yoann Chiche, Yassine Abdelouadoud, Anna Cocchi, Robin Girard



Highlights

A scalable activity-based and physics-informed bottom-up model for residential electricity demand using open data

Yoann Chiche, Yassine Abdelouadoud, Anna Cocchi, Robin Girard

- An activity-based, physics-informed bottom-up model for residential electricity demand is developed.
- The framework relies exclusively on open data for construction, calibration and validation.
- Physics-based appliance models enable explicit representation of energy transition pathways and climate effects.
- A hierarchical clustering strategy ensures scalability and preserves spatial heterogeneity.
- The approach reconstructs aggregated load profiles accurately over large geographical areas.

A scalable activity-based and physics-informed bottom-up model for residential electricity demand using open data

Yoann Chiche[a,*], Yassine Abdelouadoud[a], Anna Cocchi[a], Robin Girard[a]

[a]PERSEE - Centre Procédés, Énergies Renouvelables, Systèmes Énergétiques, Mines Paris - PSL, CS 10207 1 rue Claude Daunesse, Sophia Antipolis Cedex, 06904, France

Abstract

In the context of the ongoing energy transition, accurately modeling the spatial and temporal heterogeneity of residential electricity load curves has become increasingly critical. Such models are essential for prospective studies ranging from emerging technology adoption to multi-scale distribution system analyses. This paper introduces an activity-based, physics-informed bottom-up model of residential electricity demand, developed exclusively from open data for its construction, calibration, and validation. The activity-based structure enables a realistic representation of occupant behavior, while the physics-based analytical expressions allow for the explicit modeling of transition pathways as well as the impact of climate change. To ensure scalability over large geographical domains, a hierarchical clustering methodology is employed to identify a minimal yet representative subset of districts. This approach enables the reconstruction of aggregated load profiles across extensive regions while preserving computational efficiency and accuracy. The resulting framework constitutes a scalable, parametrizable and robust tool for electricity consumption modeling that can be leveraged in a wide range of energy transition prospective studies.

Keywords: Residential electricity demand, Bottom-up modeling, Activity-based modeling, Physics-informed models, Open data, Load profile reconstruction, Spatial heterogeneity, Energy transition, Climate change impacts, Distribution system analysis, Hierarchical clustering, Prospective energy modeling

Acronyms

1R1C One Resistance, One Capacitance.
AUSI Aggregated Units for Statistical Information.
BEM Building energy model.
COP Coefficient Of Performance.
DHW Domestic hot water.
DN Distribution network.
DSO Distribution system operator.
EMA Exponential moving average.
EPC Energy Performance Certificate.
MSE Mean Square Error.
NRMSE Normalized Root Mean Square Error.
RMSE Root Mean Square Error.
TSO Transmission system operator.

1. Introduction

1.1. Context

Residential buildings account for 12.5% of global greenhouse gas emissions—primarily due to fossil fuel use [1]. Electrifying end-uses is therefore a key strategy for meeting international CO_2 reduction targets [2, 3].

The ongoing shift toward electrification [4], notably for heating (e.g., heat pumps and electric heaters replacing gas or oil boilers) and mobility (e.g., home charging of electric vehicles) is profoundly reshaping electricity demand patterns [5, 6]. This evolution poses new challenges for power systems, including managing peak demand and maintaining grid stability. Distribution networks (DNs) are particularly affected, as their design and sizing must adapt to evolving peak loads, increased Joule losses, and more complex voltage control requirements. To address these emerging challenges, Distribution system operators (DSOs) and Transmission system operators (TSOs) require robust, high-resolution modeling tools capable of anticipating and managing future demand profiles.

An significant difficulty in modeling DNs and performing techno-economic evaluations lies in the vast diversity of possible contexts. Energy consumption is influenced by occupant behavior, socioeconomic conditions, building typologies, construction periods, energy performance, urban density, and climatic conditions. In addition, each electrical network exhibits a unique topology, further complicating system modeling. Combined, these factors create substantial heterogeneity across both space and time, making it challenging to develop generalized models that are accurate at multiple scales.

These considerations highlight the need for flexible and scalable modeling approaches capable of capturing multi-scale variability in both electricity consumption and network characteristics. To be effective, such models must combine physics-based representations—ensuring compatibility with emerging technologies—with data-driven components that capture behavioral and contextual diversity. The increasing availability of multi-scale datasets, encompassing both physical system characteristics and detailed energy consumption records, allows for finer model granularity, improved calibration, and the filling of data gaps that previously limited the accuracy of large-scale simulations.

To address these challenges while leveraging open data sources, this paper proposes a bottom-up Building energy model (BEM) that is both computationally efficient and behaviorally realistic. A subsequent hierarchical clustering procedure identifies a minimal yet representative subset of buildings, enabling the simulation of electricity consumption over large areas without modeling the entire building stock.

*Corresponding author : yoann.chiche@minesparis.psl.eu

1.2. Literature review

BEM are widely used to analyze energy demand and inform energy policy, and are commonly classified as top-down or bottom-up approaches [7]. Top-down models rely on aggregated economic or statistical data, providing insights into macro-level trends but lacking the granularity to capture individual end-uses, building characteristics, or occupant behavior [8, 9]. In contrast, bottom-up models reconstruct energy demand from detailed data on dwellings, appliances, and user habits, enabling precise representation of specific end-uses such as space heating or domestic hot water [10]. However, high-resolution bottom-up models are often constrained by heavy computational requirements and limited data availability, which restricts their scalability [11].

To tackle these limitations, many large-scale studies adopt archetype-based methods, in which representative dwelling types are predefined to reduce model complexity [12, 13]. While this approach is effective for simulating entire regions such as the EU-28 [14] or countries like Switzerland [15] it comes at the cost of reduced spatial and temporal resolution. Moreover, archetype-based models typically rely on coarse meteorological data, limiting their ability to capture local variability and extreme weather events (e.g., heatwaves, cold spells) that critically influence peak electricity demand [16].

An alternative bottom-up strategy leverages occupant activity profiles, linking electricity consumption to detailed time-use data. By simulating energy use according to specific daily activities (e.g., cooking, showering, sleeping), these models provide a more realistic and behaviorally grounded representation of demand [17, 18, 19, 20]. Such approaches are particularly suitable for prospective analyses, including the assessment of emerging technologies or demand-side flexibility policies, but are generally limited to small spatial scales as illustrated by [21] which focuses on 5000 buildings.

Some existing studies leverage large-scale datasets to capture the heterogeneity and realism of residential electricity consumption. However, they are often limited in scope, focusing either on single buildings for energy management [22] or on district-scale simulations for urban planning [23]. A few large-scale, high-resolution models based on precise open data have been developed [24, 25, 26], but they generally provide only annual consumption estimates. At the national scale, only a few studies provide high-resolution hourly data. Moreau [27] relies on a subset of 4000 simulated buildings calibrated against a subset of 6000 archetypes, which limits its ability to capture local variability, although it achieves reasonable computational efficiency. In contrast, Thorve et al. [28] reach similar resolution using a large set of buildings, though at the cost of substantial computational demand and limited building information, which restricts scenario exploration. Moreover, their model has not been validated against large-scale measurement data.

1.3. Contribution of the paper

Addressing the limitations of existing studies, this work presents a scalable, high-resolution bottom-up BEM for the residential sector, structured as a two-level modeling architecture: a fast static model—calibrated on annual energy data—to efficiently explore a wide range of building inputs, followed by a dynamic hourly model for detailed time-series simulation.

The model combines physics-based and activity-driven components, enabling realistic representation of occupant behavior and the simulation of emerging technologies and prospective scenarios. Built entirely on open data, it provides a transparent, reproducible approach capable of capturing spatial and temporal heterogeneity in electricity consumption.

A key innovation is the use of hierarchical clustering to select a minimal yet representative subset of districts, corresponding to the Aggregated Units for Statistical Information (AUSI) [29], allowing national, regional, and sub-regional load reconstruction without simulating the full building stock.

By preserving hourly resolution and accounting for climatic variability, this methodology achieves high computational efficiency while retaining detailed system dynamics, offering a significant improvement over conventional archetype- or regression-based approaches.

The model is calibrated in terms of annual energy at the district level and validated at the hourly time step at both national and regional levels. Although it has not yet been validated at lower spatial scales due to the lack of available data (e.g., medium voltage feeders or individual buildings), the model is fully operational at these scales.

2. Modeling methodology

The load model diagram is shown in Fig. 1.

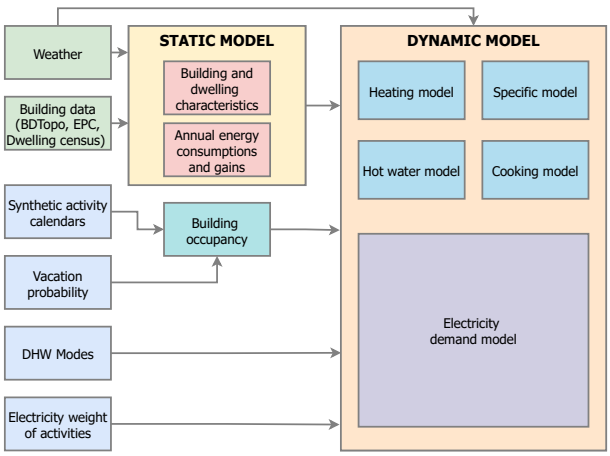


Fig. 1. Load model diagram

2.1. Data

The methodology developed in this paper relies exclusively on open-access datasets, each serving a specific purpose within the modeling framework. While the datasets listed in Tab. 1 pertain to the French building stock, the approach is fully adaptable to other geographical contexts by substituting equivalent datasets.

2.2. Static model

The first model component is the one developed in [24] and described in Section 3 of that article. It integrates localized weather data with building geometry and descriptive information to characterize the physical, thermal, and occupant-related properties needed for the simplified energy models. These models compute the annual energy consumption for each end-use and energy carrier at the building level. Detailed building characteristics, including the type and area of heat exchange surfaces, are also recorded to support subsequent simulations and analyses.

Tab. 1. Open-access datasets used in the modeling framework.

Dataset	Description
Datasets for the static model	
BDTOPO [30]	French building footprint database providing location, geometry, height and descriptive characteristics (e.g., construction year, usage).
EPC [31]	Contains detailed energy performance characteristics of dwellings for which an EPC has been established (U-values, heating system, ventilation rate).
Dwelling census database [32]	Anonymized dataset at AUSI level describing occupants, construction year class, heating energy type and occupation status.
Weather data (ERA5) [33]	Localized French weather data (temperature, irradiance, humidity, wind), also used by the dynamic model.
Datasets for the dynamic model	
Occupant activity calendars	Annual activity calendars with a 10-minute time step are generated from the processing of [34]. The raw dataset is available upon request, while the processing methodology and the resulting data are open-source.
National vacation survey [35]	Provides data on the timing and distribution of vacations across the population.
Datasets for calibration and validation	
Annual electricity consumption [36]	Used to calibrate the static model outputs.
Panel Elecdom [37]	Records specific electricity usage across 100 dwellings over one year, used for model parameter calibration.
Hourly national electricity consumption [38]	Used to validate the dynamic model at the national scale.
Hourly regional electricity consumption [39]	Used to validate the dynamic model at the regional scale.

To account for uncertainties in the relationships among the datasets, multiple random realizations of the simulation are performed, and the most representative one is selected following the procedure detailed in Section 4.1. This enables the determination of building characteristics and occupant preferences most consistent with the available data on intrinsic building performance, energy consumption, and occupant behavior.

2.3. Dynamic model

The dynamic model uses calibrated building- and dwelling-level results from the static simulation as input parameters—such as thermal envelope properties, temperature setpoints, occupancy levels, and dwelling types—alongside annual energy consumption and thermal gain values. By integrating these properties with physics-based weather modeling and activity-driven demand modeling, the model temporally disaggregates annual data into hourly time steps. The resulting hourly energy demand profiles are categorized by specific end uses:

- Heating
- Domestic hot water (DHW)
- Cooking
- Other specific electricity uses

The methodology associated with each end-use is detailed in the following sections.

2.3.1. Occupant activity calendars

The study referenced in [34] collected approximately 27,500 daily activity calendars from a representative sample of french households (Metropolitan France, La Réunion, Guadeloupe, and Martinique), with activities recorded at 10-minute intervals. Hierarchical agglomerative clustering was applied separately to weekday and weekend datasets using Ward's linkage criterion and the Jaccard distance metric. This process produced 54 weekday clusters and 52 weekend clusters, with the optimal number of clusters determined using the elbow method [40].

To generate full-year schedules, synthetic individuals were created based on key socioeconomic attributes, including age, gender, and employment status. For each individual, annual activity calendars were constructed by sampling daily calendars from the cluster whose medoid best matched their profile. These calendars provide a continuous representation of household activity patterns over the year.

The constructed annual activity calendars are then randomly assigned to the simulated housing units, without considering additional socio-demographic matching. Vacation days—during which occupants are assumed absent—are incorporated based on the national vacation distribution [35]. The number of vacation days is a model parameter that varies with dwelling type: for main residences, the probability of a vacation day follows the national distribution, whereas for second or occasional homes, which primarily serve as vacation destinations, the probabilities are reversed relative to main homes. While this random assignment may introduce minor inaccuracies at the individual dwelling level, its impact on aggregate results at regional or national scales is minimal. Appendix B illustrates the aggregation effect of the calendars and demonstrates why the random assignment does not significantly affect overall results.

Occupancy rates at the dwelling and building levels are derived from activity calendars and used as inputs to the heating model. The same calendars also drive appliance electricity consumption and domestic hot-water demand in the dynamic model, ensuring consistency across building energy simulations.

2.3.2. Heating

The heating model represents the building's thermal dynamics using a simplified One Resistance, One Capacitance (1R1C) electrical analogue. This approach models the building envelope's resistance to heat flow with a thermal resistance (R_{th} [K/W]) and the building's ability to store thermal energy with a thermal capacitance (C_{th} [Wh/K]). The model assumes a **single thermal zone per building**, linear heat transfer, and uniform internal temperature, neglecting multi-zone stratification, thermal bridges, and short-term fluctuations. This model is illustrated in Fig. 2.

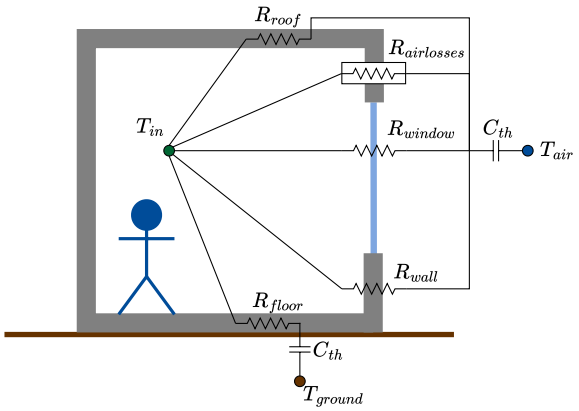


Fig. 2. Schematic of the building thermal parameters.

Thermal capacitance parameterization. Thermal capacitance values are assigned according to the building's thermal inertia category, based on the Energy Performance Certificate (EPC) classification [31]. The values are reported in Tab. 2 and correspond to the thermal storage per unit of floor area.

Tab. 2. Thermal inertia categories and corresponding capacitance per unit floor area.

Category	Light	Average	Heavy	Very heavy
Value [J K^{-1} m^{-2}]	80,000	165,000	260,000	370,000

The distribution of inertia categories across the 20 nationally representative districts, together with a sensitivity analysis of heating demand under different assumptions, is provided in Appendix A.

Heating demand calculation. Heating demand corresponds to the energy required to maintain indoor temperature at the thermostat setpoint (T_{set}) when the building's passive thermal response is insufficient. Heating power (P_{heat} [W]) is applied only when the free-floating indoor temperature (T_{free}) is below T_{set}. The required power is computed to restore the temperature to the setpoint within a one-hour timestep:

$$P_{\text{heat}} = \max\left(0, \frac{C_{\text{th}} \cdot (T_{\text{set}} - T_{\text{free}})}{\Delta t}\right), \quad \Delta t = 1\ \text{h} \qquad (1)$$

Heating season. For each district, the heating season begins when the average outdoor temperature remains below a defined threshold for a minimum number of consecutive days, and ends when it exceeds another threshold for the same duration. Exact thresholds and durations are provided in Section 2.3.6.

Setpoint temperature. The building-level setpoint temperature, T_{set}, is calculated as the surface-area-weighted average of individual dwelling setpoints. It varies dynamically throughout the day according to occupancy patterns and scheduled temperature setbacks.

Free-floating temperature. The free-floating temperature, T_{free}, captures the passive thermal response of the building due to inertia. It is computed using an Exponential moving average (EMA) of the effective external temperature ($T_{\text{effective}}$). The EMA half-life is derived from the building's thermal time constant, $\tau = R_{\text{th,total}} \cdot C_{\text{th}}$, which controls the damping and delay of indoor temperature changes. Here, $R_{\text{th,total}}$ includes contributions from walls, roof, floor, ventilation and passive air losses.

Effective external temperature. The effective external temperature $T_{\text{effective}}$ consolidates multiple thermal influences into a single equivalent temperature:

$$T_{\text{effective}} = \frac{T_{\text{air}} \cdot U_{\text{air}} + T_{\text{ground}} \cdot U_{\text{ground}}}{U_{\text{total}}} + P_{\text{gains}} \cdot R_{\text{th,total}} \qquad (2)$$

where:

- U_{air} (W/K) represents heat transfer through ventilation, infiltration, and envelope transmission (walls, windows, roof).
- U_{ground} (W/K) represents heat transfer through the floor.
- $U_{\text{total}} = 1/R_{\text{th,total}} = 1/(R_{\text{air}} + R_{\text{ground}})$ is the building's total thermal conductance.
- P_{gains} (W) represents gains from solar radiation, occupants, and appliances. Annual gains from the static simulation are distributed over the heating period, with solar gains allocated by radiation and occupant/appliance gains scaled by each hour's fraction of total appliance energy use.

Overall, this framework provides a physically grounded and reproducible method for simulating dynamic heating loads at hourly resolution across large building populations. By leveraging a fully vectorized implementation, the model also achieves high computational efficiency, allowing rapid evaluation of diverse building types, thermal inertia categories, and occupancy patterns without the need for iterative timestep loops.

2.3.3. Domestic hot water

Hourly DHW consumption per dwelling is estimated by distributing the annual demand throughout the days based on activity-driven usage patterns. Each dwelling is randomly assigned a consumption profile that specifies production hours—when water heating is prioritized—and other hours. These profiles were developed to match empirical measurements of hot water load profiles [37].

$E^{\text{DHW}}_{\text{day}}$ [Wh] is the total DHW energy demand for the day, H_{prod} [h] the number of production hours, and H_{other} [h] the number of remaining hours in the day. The maximum heating power of the water heater is denoted as P_{heater} [W] is determined based on the number of occupants, which directly defines the tank size according to the guidelines provided in [41].

For each hour t, the heating power $P^{\text{DHW}}_{\text{t}}$ [W] is determined as follows:

- **During production hours** (H_{prod}): the power is the minimum between $\frac{E^{\text{DHW}}_{\text{day}}}{H_{\text{prod}}}$, the average power needed to fulfill the daily demand evenly, and P_{heater}, the maximum heater capacity.
- **During other hours** (H_{other}):
  - If the total daily energy demand can be fully supplied within the production hours ($E^{\text{DHW}}_{\text{day}} \leq P_{\text{heater}} \times H_{\text{prod}}$), no heating power is allocated ($P^{\text{DHW}}_{t} = 0$).
  - Otherwise, the remaining energy demand after production hours is evenly distributed over the remaining hours, limited by the heater capacity.

- Any remaining energy that cannot be allocated during production or other hours is considered unmet demand.

The full equation is:

$$P_{\mathrm{DHW},t} = \begin{cases} \dfrac{E_{\mathrm{day}}^{\mathrm{DHW}}}{H_{prod}} & \text{if } t \in H_{prod} \text{ and } E_{\mathrm{day}}^{\mathrm{DHW}} \leq P_{heater} \times H_{prod} \\ P_{heater} & \text{if } t \in H_{prod} \text{ and } E_{\mathrm{day}}^{\mathrm{DHW}} > P_{heater} \times H_{prod} \\ 0 & \text{if } t \notin H_{prod} \text{ and } E_{\mathrm{day}}^{\mathrm{DHW}} \leq P_{heater} \times H_{prod} \\ \dfrac{E_{\mathrm{day}}^{\mathrm{DHW}} - P_{heater} \times H_{prod}}{H_{other}} & \text{if } t \notin H_{prod} \text{ and } E_{\mathrm{day}}^{\mathrm{DHW}} > P_{heater} \times H_{prod} \end{cases} \tag{3}$$

2.3.4. Cooking

Cooking electricity consumption is distributed according to the occupant's cooking activity. However, the energy used can vary depending on the type of meal, with dinner typically consuming more energy than breakfast. Moreover, even after the cooking activity ends, consumption can continue because some appliances remain on for a while. To model these phenomena, different weights are assigned to morning, midday, and evening cooking activities. Additionally, a post-cooking duration is included during which consumption persists, and this duration also depends on the meal type. These weights and durations have been calibrated using existing measured cooking profiles [37], ensuring that the modeled consumption reflects real-world behavior.

2.3.5. Other electricity needs

Other electricity end-uses, such as dish-washing and television, are allocated according to occupant activities. Each activity is assigned a weight representing its contribution to total electricity consumption. These weights are determined through a small optimization problem that assigns a weight to each activity group—formed based on shared characteristics—so that the aggregated load profile for a group of occupants closely matches the average appliance load profile derived from real-world measurements [37]. They are listed in Tab. 3.

Tab. 3. Electricity usage weight by activity.

Activity	Weight
Hygiene	2.9
Other at home	4.7
Meal at home	3.4
Home working	4.7
Cooking	6.5
Dish washing	5.1
Housework	5.1
Clothes washing	5.1
Ironing	5.1
Home DIY	4.7
Gardening	2.9
Television	6.4
HiFi	3.2
Video game	6.4
Computer	3.2
Electronic device	3.2
Sleep	2.6
Other outside	2.4
Meal outside	2.4
At work	2.4
At school	2.4
Shopping	2.4
Work journey	2.4
Other journey	2.4

2.3.6. Model parameters

The dynamic model relies on several parameters, each of which requires a default value. In practice, defining a default value is not straightforward, as some heaters are activated not based on a specific temperature but on a specific date. Most default values were either selected based on references or chosen through model testing. Tab. 4 lists all the parameters and their default values.

A parametric study of both heating thresholds has been conducted in Section 5.2.3 to assess their impact on the results and to verify the adequacy of the default parameter values.

3. Representative districts selection

For large-scale (national or regional) residential electricity load modeling, simulating every building is neither necessary nor computationally efficient. Instead, a representative subset of districts is selected to capture the diversity of consumption patterns across the study area .

3.1. Clustering procedure

Hierarchical agglomerative clustering was applied using Ward's linkage criterion and the Euclidean distance metric to group districts based on several key attributes: average annual electricity and gas consumption from 2020 to 2024, surface area, number of dwellings, number of buildings heated with electricity or gas, number of primary residences, proportion of houses to apartments, average year of construction, and unified degree hours. This method allows the formation of clusters at multiple levels of granularity, adaptable to different analytical scales. Within each cluster, the most representative district—defined as the medoid—is then identified.

3.2. Anomaly screening and exclusion

Before the medoid selection, districts with abnormal consumption patterns—such as extreme year-to-year variability, inconsistent long-term trends, or near-zero average consumption despite the presence of residential buildings—are excluded. Only districts with valid and consistent data are retained for the selection of representative medoids.

4. Model calibration and validation

4.1. Static model

Using the annual electricity consumption per district [36], the static model is calibrated and validated in terms of annual energy at the district level. This is achieved by running multiple simulation batches that combine different statistical inferences of building characteristics with varying heating setpoints and energy use factors for non-heating end-uses. For each iteration, the gap between the district-aggregated consumption and the measured value is computed and adjusted using a corrective factor that penalizes deviations from the national average heating setpoint of 19°C [44]. The iteration with the smallest adjusted gap is then selected as the best-fitting configuration.

Tab. 4. Dynamic model parameters, their description and default values.

Parameter	Description	Default
Heating threshold	Temperature below which heating starts (after Aug)	15,0 °C [42]
Heating end threshold	Temperature above which heating stops (after Feb)	15,0 °C [42]
Heating days	Days for rolling average of heating threshold	5 days
Heating end days	Days for rolling average of heating end threshold	5 days
Reduction temperature	Temp. reduction when occupants are absent or sleeping	−3,0 °C [43]
Absent temperature	Temp. set when dwelling is unoccupied	8,0 °C
Sleep reduction share	Fraction of dwellings reducing temp during sleep	10%
Not home reduction share	Fraction reducing temp when not home	10%
Both reduction share	Fraction reducing temp in both occupancy states	20%
Vacation parameters	Number of vacant days per residence type	main: 3, second: 300, occasional: 250
DHW modes	Fraction of dwellings and production hours per mode	A: 5% (20–23h) B: 20% (21–23h) C: 20% (22–1h) D: 10% (23–2h) E: 10% (0–3h) F: 15% (1–6h) G: 15% (4–6h / 11–16h) H: 5% (none)

The equation for the relative gap calculation is defined as:

$$GAP = GAP_{\text{raw}} \times C_{\text{factor}} \times (T_{\text{ref}} - T_{\text{setpoint}})^2 \quad (4)$$

where:

- GAP is the final relative gap value used for calibration.
- GAP_{raw} is the initial relative gap between simulated and measured consumption, before penalization.
- C_{factor} represents the corrective factor applied to adjust the weight of the penalization; in this study, it was set to 0.02.
- T_{ref} references national heating setpoint temperature (here 19 °C).
- T_{setpoint} represents the estimated average heating setpoint temperature for the district.

4.2. Dynamic model

For each geographical scale, a set of representative districts is selected using the clustering approach described in Section 3. Each representative district is assigned a scaling factor corresponding to the total annual consumption of its cluster. The aggregate load curve is then reconstructed as the weighted sum of the representative district loads and compared against measured data at the national [38] or regional [39] levels. This process is repeated for different numbers of clusters, and the configurations are evaluated to identify the one that minimizes the Mean Square Error (MSE) while maintaining reasonable computational requirements.

In addition, a parametric study (Section 5.2.3) is conducted to calibrate key model parameters, further improving the overall accuracy and robustness of the results.

5. Results

In this section, model performance is assessed using several statistical indicators, including MSE, Normalized Root Mean Square Error (NRMSE), and R^2. These metrics serve as diagnostic tools to highlight areas for improvement rather than to define strict acceptability criteria. No universal threshold is specified, as appropriate performance levels depend on the study objectives, dataset characteristics, and intended application. The interpretation of these indicators is therefore left to the user in accordance with the specific requirements of the analysis.

5.1. Static model calibration and validation results

Across 200 districts, the model achieves an R^2 score of 0.944 in reproducing annual electricity consumption, as shown in Fig. 3.

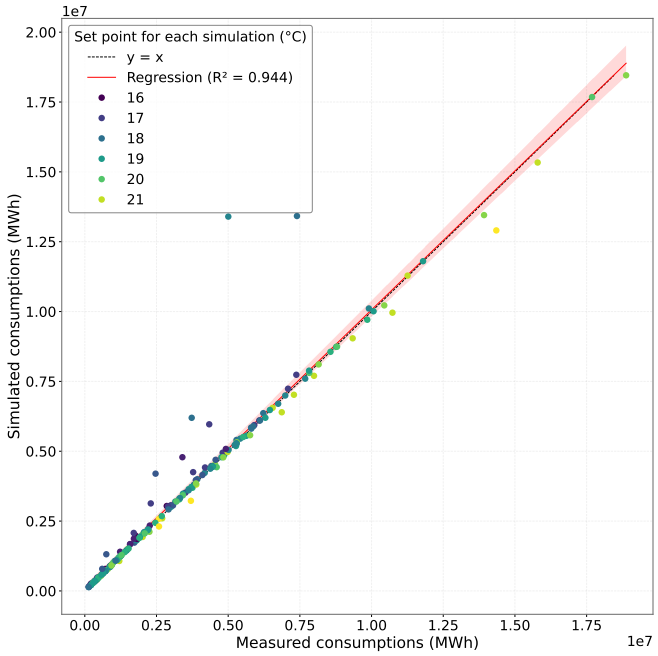


Fig. 3. Linear regression of 200 districts simulations vs measurements (2023)

As shown by Fig. 4, the distributions of the key parameters—setpoint temperature and energy use factor—are centered around 19°C and 1.0, respectively, in agreement with the expected values for France [44].

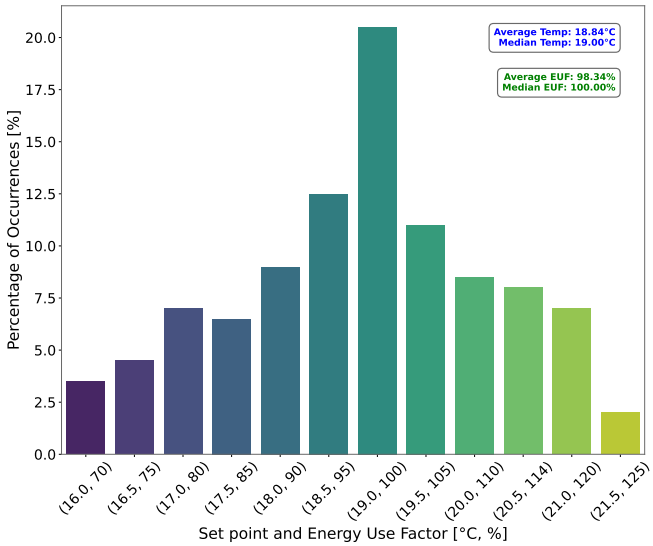


Fig. 4. Distribution of set temperature and energy use parameters, 200 districts in 2023

5.2. *National scale results*

5.2.1. *Clustering results*

Using the clustering method described in Section 3.1, the total deviation between modeled and target consumption can be quantified for each number of clusters. Fig. 5 shows the results from 50 clustering iterations, with small random perturbations applied to the clustering parameters to capture variability in the outcomes. The average computational time was 95.97 seconds per district on a single-core CPU.

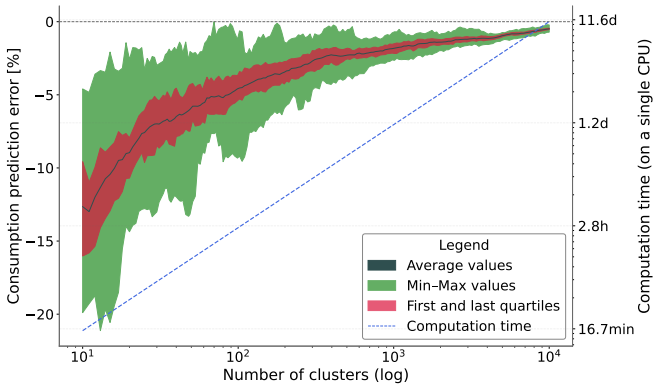


Fig. 5. Decomposition of errors and computational time as a function of the total number of clusters

For this paper, a number of clusters ranging from 10 to 200 was identified as providing an optimal balance between model accuracy and computational efficiency. The analysis therefore focused on this range, with the MSE further decomposed into its bias and variance components: the inter-day variance of the daily mean error, and the mean of the intra-day variance of the daily error. The MSE is expressed in relative terms (%²), so a value of 100% corresponds to an Root Mean Square Error (RMSE) of 10%. A cluster count of 20 was found to offer the most favorable trade-off between accuracy and computational cost.

$$\mathrm{MSE} = \mathrm{Bias}^2 + \mathrm{Var}_{\mathrm{inter}} + \mathrm{Var}_{\mathrm{intra}} \quad (5)$$

Fig. 6 shows the breakdown of the MSE for the number of clusters ranging from 1 to 500, as well as the computation time.

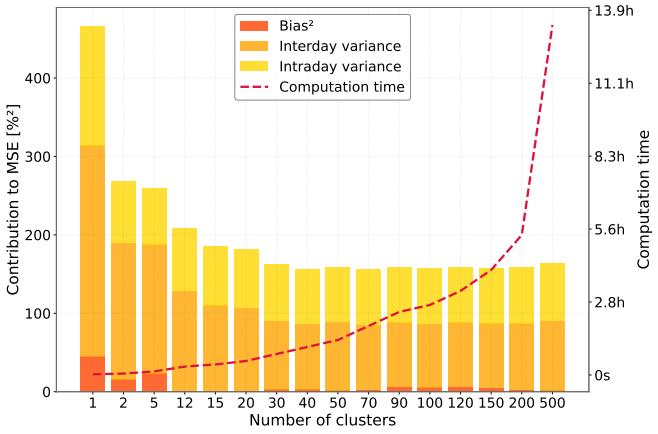


Fig. 6. Error decomposition for various total number of clusters (2023)

5.2.2. *Validation results*

Figs. 7 to 9 present a comparison between the simulated and measured hourly profiles for the years 2023, 2022, and 2021, respectively. The NRMSE, defined as the RMSE divided by the mean of the measured values, remains below 15%.

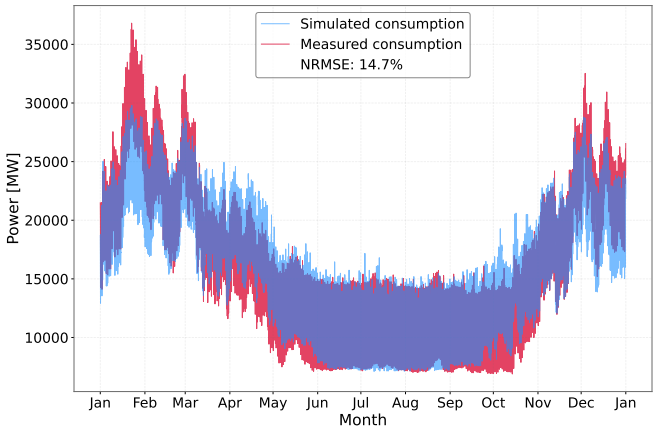


Fig. 7. Model validation for the national scale with 20 districts (2023)

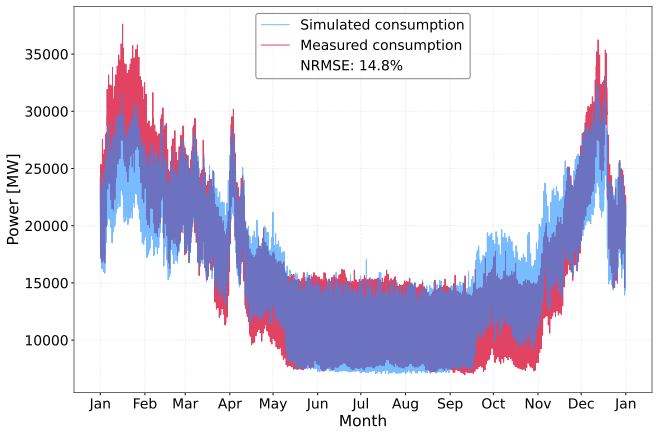


Fig. 8. Model validation for the national scale with 20 districts (2022)

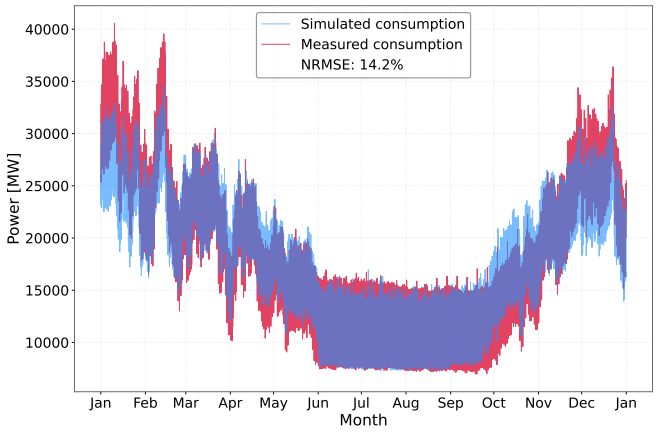


Fig. 9. Model validation for the national scale with 20 districts (2021)

5.2.3. *Parametric calibration*

The dynamic model is sensitive to several parameters, such as the outdoor temperature threshold for heating activation and deactivation, the proportion of dwellings applying temperature reductions during different occupancy states, DHW consumption patterns, and others. To further improve the results, a parametric study was conducted on the temperature threshold for heating activation and deactivation. Other parameters were either selected through optimization (Sections 2.3.3 to 2.3.5) or found to have a low impact on the overall results at the studied scale. The parametric study was performed over the years 2021 to 2023, for which sufficient data were available, and the results represent the average values across these years. Fig. 10 presents the results of the parametric study on the heating activation threshold, while Fig. 11 illustrates those corresponding to the heating deactivation threshold.

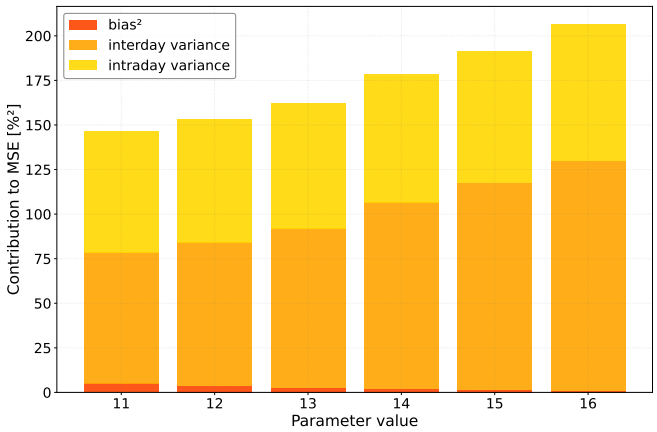


Fig. 10. National scale parametric study on heating threshold (average from 2021 to 2023)

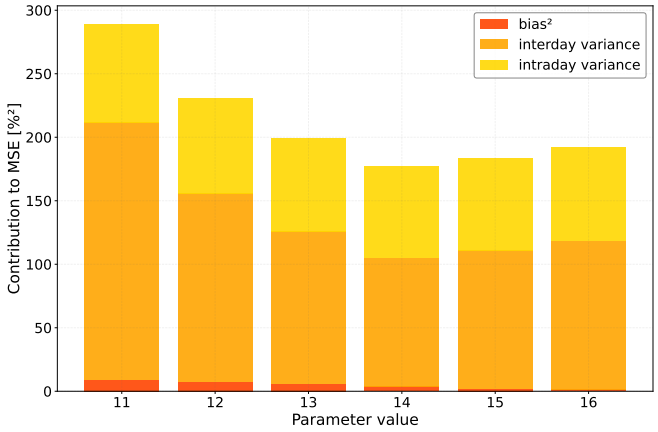


Fig. 11. National scale parametric study on heating threshold (average from 2021 to 2023)

With the best-performing parameters, the NRMSE decreases by 2.3%, as shown by Fig. 12.

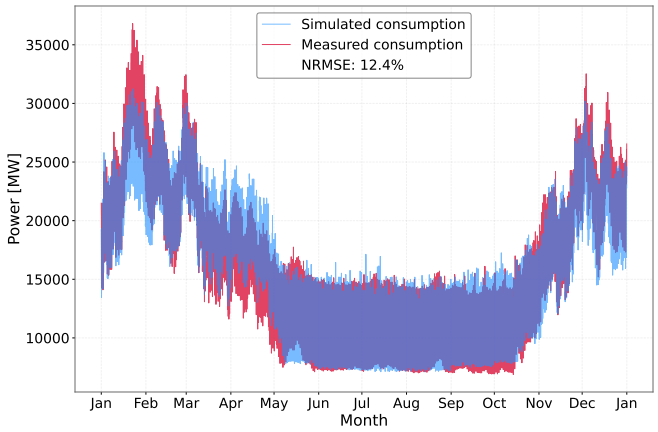


Fig. 12. Model validation for the national scale with 20 districts with heating thresholds best parameters (2023)

5.3. *Regional scale results*

As regions differ in both size and number of districts, selecting a fixed number of districts per region produces highly variable results. Therefore, at the regional scale, 1% of the total number of districts in each region was used for validation. In addition, the nationally optimized parameters identified in Section 5.2.3 were applied in this analysis.

Regional results exhibit greater variability and higher error levels than the national-scale analysis. Some discrepancies indicate that certain parameters, such as heating thresholds, are intrinsically region-specific and therefore require regional calibration. Others reveal the influence of currently unmodeled phenomena, including seasonal interregional population movements, which should be incorporated to improve model performance.

5.3.1. *Validation results*

Some regions, such as Ile-de-France in Fig. 13, exhibit unrealistically high modeled electricity consumption during summer, which should normally be lower. This discrepancy is likely due to vacation parameters that underestimate the number of days households spend away from their dwellings.

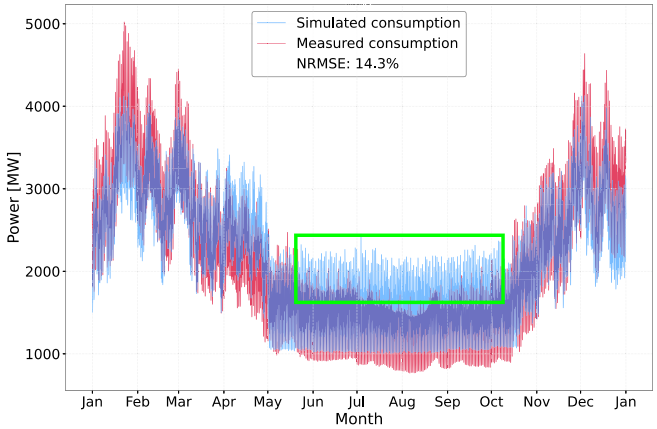


Fig. 13. Model validation for the regional scale (region Ile-de-France) with 1 % of the district (2023)

In some regions, such as Bourgogne-France-Comté in Fig. 14, the model struggles to accurately reproduce the heating season when applying parameters calibrated at the national scale

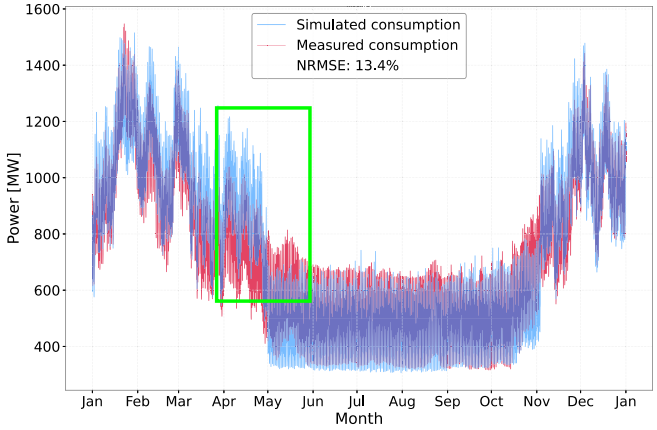


Fig. 14. Model validation for the regional scale (region Bourgogne-Franche-Comté) with 1 % of the district (2023)

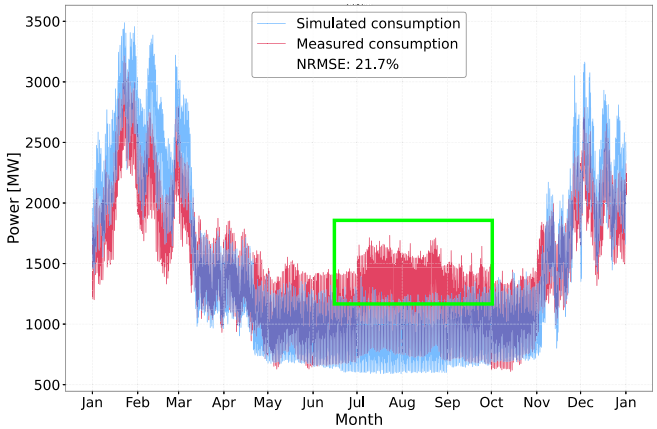


Fig. 15. Model validation for the regional scale (region Provence-Alpes-Côte-d'Azur) with 1 % of the district (2023)

Some southern regions of France, which experience higher average summer temperatures, such as PACA in Fig. 15, show a marked increase in electricity consumption during the warmer months, accompanied by a higher NRMSE. This indicates that cooling demand—currently not included in the model—plays a significant role. Additionally, seasonal population variations, such as increased occupancy during summer vacations, may contribute to these patterns, suggesting that vacation probabilities differ across regions.

The MSE and NRMSE for each region (excluding Corsica and non-continental France due to insufficient data) are shown in Fig. 16 and Tab. 5 respectively.

The MSE decomposition is presented alongside total regional consumption. Total consumption does not appear to drive increases in relative MSE (%²). Regions such as PACA, Occitanie and Auvergne-Rhône-Alpes exhibit higher relative MSE, which is likely due to vacation parameters and the absence of cooling demand in the model—a feature that should be incorporated in future work.

Tab. 5. NRMSE by region (%).

Region	NRMSE (%)
National	12,4
Île-de-France	14,3
Centre-Val de Loire	14,3
Bourgogne-Franche-Comté	13,4
Normandie	14,7
Hauts-de-France	14,1
Grand Est	12,4
Pays de la Loire	15,4
Bretagne	16,5
Nouvelle-Aquitaine	14,8
Occitanie	22,8
Auvergne-Rhône-Alpes	18,4
Provence-Alpes-Côte d'Azur (PACA)	21,7

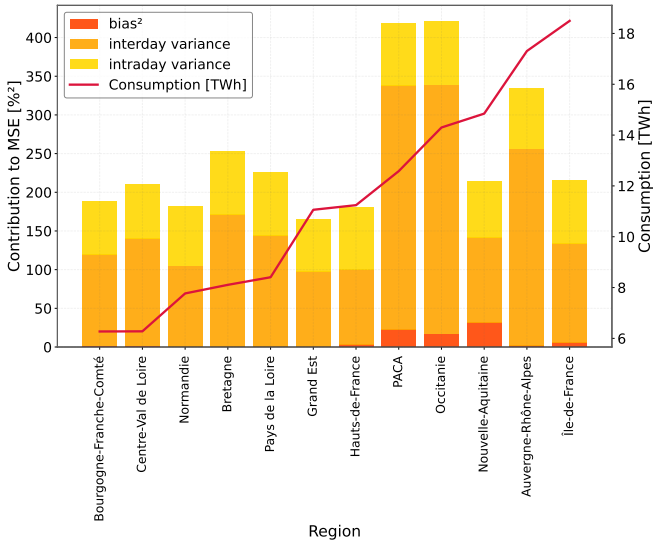


Fig. 16. MSE decomposition in %² for each region with 1% of districts (2023)

6. Discussion

This section interprets the principal findings, situating the model within the existing landscape of residential energy modeling and elaborating on the implications of its architecture and performance.

6.1. Implications of the scalable clustering methodology

A primary challenge in bottom-up energy modeling is resolving the inherent tension between computational tractability and the granular representation of heterogeneous building stocks. The results presented demonstrate that our hybrid, scalable framework provides an effective resolution to this challenge. The core finding is that the hierarchical clustering approach (Section 3.1) successfully identifies a minimal yet representative subset of districts (medoids).

The validation results—achieving a national NRMSE of 12.4% with 20 representative districts —empirically support the hypothesis that the vast heterogeneity of the national stock can be effectively compressed into data-derived district subsets.

For system operators (TSOs and DSOs), this methodology quantifies the trade-off between accuracy and computational cost (Fig. 6). This efficiency, with simulations running in under two minutes on 20 cores, is not merely a technical gain; it facilitates the high-velocity, iterative simulations required for robust prospective scenario analysis (e.g., assessing electrification pathways or climate change impacts).

6.2. Model error as a diagnostic for unmodeled dynamics

The model's performance, particularly its regional discrepancies, serves as a powerful diagnostic tool. While the national-scale aggregation is robust, the meso-scale analysis reveals a clear geographical structuration of the model error. Specifically, the significantly higher NRMSE in southern regions —such as Provence-Alpes-Côte-d'Azur (21.7%) and Occitanie (22.8%)— is highly instructive. This degradation in performance provides quantitative, empirical validation that cooling demand, a currently unmodeled end-use, is a non-negligible and regionally dominant load driver. This omission is the primary source of variance in these regions and constitutes a clear directive for future model enhancement.

Additionally, the systematic overestimation of summer consumption in regions such as Île-de-France, coupled with underestimation in southern regions, reflects population mobility effects. The current vacation modeling assumes uniform probabilities across France and does not consider socio-economic factors, which likely influence regional occupancy patterns. These limitations should be addressed in future work to improve regional accuracy, including more realistic modeling of seasonal population shifts and socio-economic heterogeneity, potentially informed by available data on non-residential tourism patterns.

6.3. The architectural implications of a hybrid framework

The architectural decision to develop a hybrid, physics-informed and activity-based framework, built exclusively on open-data sources, carries significant implications beyond immediate accuracy.

- **Prospective capabilities**: The model transcends descriptive simulation. By anchoring demand to physical (e.g., 1R1C parameters) and behavioral (e.g., DHW production schedules) drivers, it allows for ex-ante assessment of scenarios that lack historical precedent. These include technological shifts (e.g., building envelop renovation altering R_{th}) or demand-side flexibility policies (e.g., modifying activity-driven appliance use). Appendix C illustrates this capability through a case study analyzing load variations under energy system retrofit scenarios.
- **Modular extensibility**: The framework's current simplifications, such as the 1R1C thermal analogue or the static heat pump Coefficient Of Performance (COP), are not systemic weaknesses but modular components. They can be replaced by higher-fidelity modules (e.g., multi-capacitance models, temperature-dependent COPs) to incrementally enhance physical realism and reduce model variance. In this context, a cooling module could be readily integrated in future developments, leveraging cooling system data from the French EPC dataset and adopting an approach consistent with the heating model based on degree-hours and equivalent resistance–capacitance representation.
- **Behavioral fidelity**: The activity-based stochastic component provides a robust, behaviorally-grounded foundation for demand. While the current implementation randomizes schedule assignment without explicit socio-economic dependence, the framework is designed to incorporate such linkages, improving the representation of occupant-driven consumption diversity.

6.4. Validation scale and the path to local application

The validation strategy confirms the model's robustness at the macro (national) and meso (regional) scales, establishing its utility for strategic, system-level analysis. However, a key objective for DSOs is performance at the micro-scale (district or building), which has not yet been validated for hourly dynamics, due to the lack of corresponding data. Consequently, the next key step is to progress from aggregated validation toward a disaggregated assessment. This improvement will become possible once medium-voltage feeder load curves are released as open data, which is anticipated within the next few years.

7. Conclusion

This article presents a scalable, hourly-resolution residential electricity load model that is both physics- and activity-based, and built entirely from open data. Coupled with a hierarchical clustering approach, the model enables precise simulation of national and regional load profiles using a limited set of building groups, called districts, while maintaining low computational times (under 2 minutes using 20 CPU cores in parallel).

Applied to continental France, the model achieves a national NRMSE of 12.4%, and regional NRMSEs ranging from 13.4% to 22.8%. Decomposition of the MSE shows that most of the error arises from variance rather than bias. A parametric study conducted on only two key parameters reduced the NRMSE by 2.3%, suggesting that an extended study across all model parameters could further improve accuracy.

Future developments should focus on refining both thermal and behavioral representations, extending calibration to

additional parameters and regions, and integrating emerging end-uses such as electric vehicles, air conditioning, and reversible heat pumps. Strengthening the coupling between physical, behavioral, and socio-economic layers would enhance the model's representativeness and enable more robust scenario analyses under evolving climatic, technological, and behavioral conditions.

Overall, this framework provides a flexible and computationally efficient tool for extensive studies of residential electricity demand under various technology adoption and behavioral scenarios, without requiring prohibitive computational resources.

8. Code availability

The entirety of the computational tools and source code developed for this paper is made publicly available under the MIT License to ensure transparency and reproducibility of the results.

- Static Model (developed in [24]): https://gitlab.com/energytransition/buildingmodel [45]
- Dynamic Model: https://git.persee.minesparis.psl.eu/planeterr/building_eload [46]

Appendix A. Inertia class

Appendix A.1. Inertia class distribution

The distribution of building inertia classes across the 20 nationally representative districts is summarized in Tab. A.6.

Tab. A.6. Distribution of inertia classes

Inertia Class	Count	Percentage (%)
Average	3521	35.4
Heavy	2435	24.5
Light	3705	37.2
Very Heavy	291	2.9

Appendix A.2. Inertia class variation

In this study, two alternative building inertia scenarios were created and compared with the initial inertia distribution. In the high-inertia scenario, all buildings were assigned to the "very heavy" inertia class, while in the low-inertia scenario, all buildings were assigned to the "light" inertia class. Simulations were performed on the 20 nationally representative districts and then scaled to match national heating demand. Fig. A.17 and Fig. A.18 illustrate the heating demand profiles associated with the three inertia class scenarios over a winter month and a winter week, respectively.

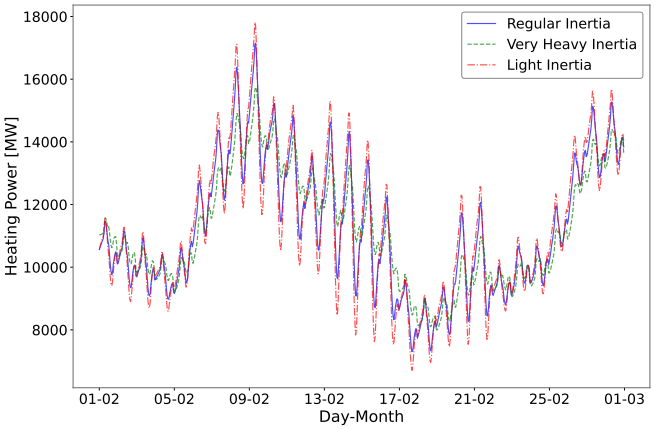


Fig. A.17. February 2023 heating demand – three inertia scenarios

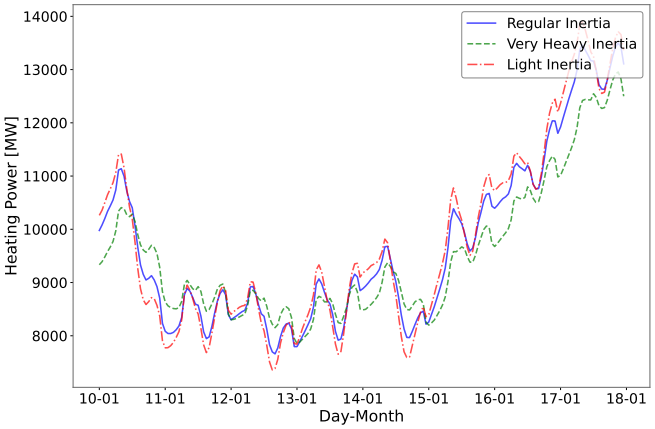
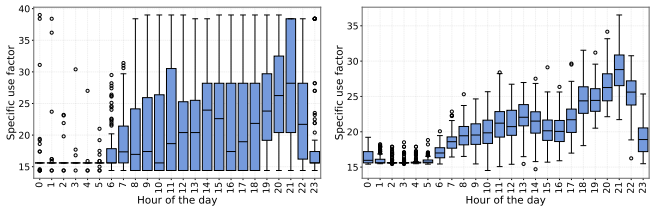


Fig. A.18. Week in february 2023 heating demand – three inertia scenarios

Compared to the baseline scenario, the high-inertia scenario exhibits reduced short-term fluctuations in heating demand, whereas the low-inertia scenario shows more pronounced variations.

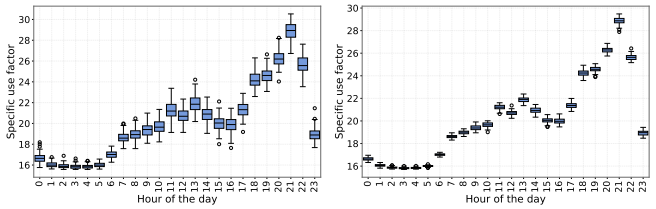
Appendix B. Empirical support for activity calendar aggregation

To validate the robustness of the random activity calendar assignment, a sensitivity analysis was conducted to measure how variability in electricity-related activities diminishes through aggregation. Using a dataset of 100,000 simulated occupants, we computed "specific use electricity weights" based on assigned activity labels, whose definitions and associated weights are described in Tab. 3. We then performed a Monte Carlo simulation, drawing 100 random combinations of occupants for different number of occupants (1, 10, 100 and 1000). The resulting Fig. B.19 illustrate the hourly distribution of these electricity weights: for a single occupant, the envelope of possible profiles is wide, representing high individual uncertainty; however, as the number of occupants increases, the mean profile stabilizes and the variance narrows significantly. This empirical evidence supports the claim that the lack of socioeconomic matching does not lead to significant error in aggregate demand modeling, as the stochastic variations of individual behaviors effectively cancel out at higher levels of aggregation.



(a) 1 occupant

(b) 10 occupants



(c) 100 occupants

(d) 1000 occupants

Fig. B.19. Variability of specific uses with respect to the number of occupant activity calendars.

Appendix C. Prospective study on energy systems

To illustrate the potential use of the model as a prospective tool for TSOs and DSOs, this section analyzes the variability of results under various energy system renovation scenarios. These scenarios are applied to the buildings of district *692020301* within the city of *Sainte-Foy-lès-Lyon (69202)*, based on the 2023 building stock. To apply the renovation scenarios, the static simulation is first run and calibrated, ensuring that building and dwelling characteristics best represent building stock. The energy systems are then modified according to the defined scenarios, and the static model's energy function is executed again without the boundary model. Finally, the dynamic model is run for each simulation scenario.

Appendix C.1. Distribution of energy systems

The distribution of energy systems across the district is summarized in Tab. C.7.

Tab. C.7. Energy systems distribution across the district

System	Quantity	%
Heating (per building)		
Biomass/coal boiler	4	2,6
Biomass/coal stove	4	2,6
Electric heater	30	19,7
Fossil gas boiler	91	59,9
Liquified petroleum gas boiler	2	1,3
Oil boiler	21	13,8
DHW (per building)		
Biomass	3	2,0
Electricity	59	38,8
Fossil gas	71	46,7
Liquified petroleum gas	1	0,7
Oil	18	11,8
Cooking (per dwelling)		
Electricity	390	41,8
Fossil gas	410	43,9
Liquified petroleum gas	132	14,1
None	1	0,1

Appendix C.2. Renovation scenarios

The retrofit scenarios studies are summarized in Tab. C.8.

Tab. C.8. Overview of energy transition scenarios

Scenario	Description
Standard	Reference case representing current building stock with no energy retrofits.
Heating HP	Replacement of all non-electric heating systems with air-to-water heat pumps of a 2.5 COP.
DHW Electric	Conversion of all fossil-fuel based DHW systems to electric storage heaters.
Heating HP + DHW	Integrated scenario combining heat pump installation for space heating and full electrification of DHW.
Electric Cooking	Complete transition of cooking appliances (gas/LPG) to electric induction or radiant systems.

Appendix C.3. Results

The results presented in Figs. C.20 and C.21 correspond to a representative winter week at the district and building scales, respectively indicate significant variability in energy consumption at both district and building scales.

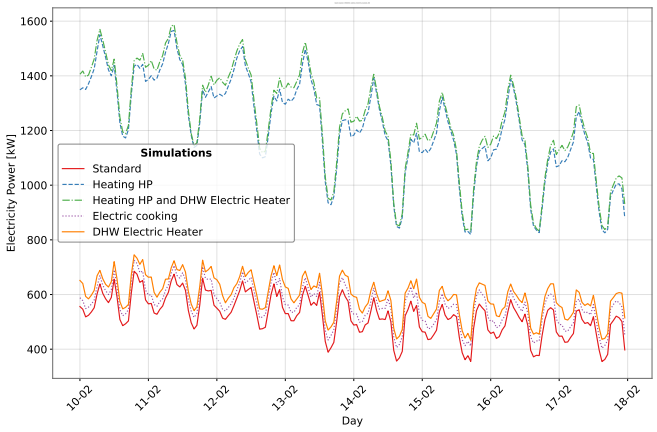


Fig. C.20. Peak-week load profiles of district *692020301* under retrofit scenarios

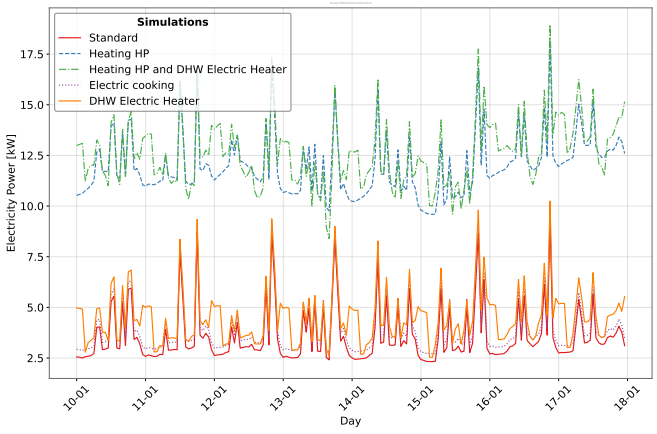


Fig. C.21. Peak-week load profiles of a building in district *692020301* under retrofit scenarios

By analyzing different deployment scenarios that combine heating system replacements and/or envelope renovation measures, network operators can assess the resulting evolution and shape of future load profiles.

References

[1] M. Ge, J. Friedrich, L. Vigna, Where Do Emissions Come From? 4 Charts Explain Greenhouse Gas Emissions by Sector, 2024. https://www.wri.org/insights/4-charts-explain-greenhouse-gas-emissions-countries-and-sectors.

[2] IEA, Breakthrough Agenda Report 2024, Technical Report, International Energy Agency, 2024. https://www.iea.org/reports/breakthrough-agenda-report-2024.

[3] M.A. d. t.T. écologique, Programmations pluriannuelles de l'énergie - 2019–2028, Technical Report, Ministères Aménagement du territoire Transition écologique, 2019. https://www.ecologie.gouv.fr/politiques-publiques/programmations-pluriannuelles-lenergie-ppe.

[4] M. Aklin, S. P. Harish, J. Urpelainen, A global analysis of progress in household electrification, Energy Policy 122 (2018) 421–428. DOI: 10.1016/j.enpol.2018.07.018

[5] L. F. Cabeza, Q. Bai, P. Bertoldi, J. M. Kihila, A. F. P. Lucena, E. Mata, S. Mirasgedis, A. Novikova, Y. Saheb, Buildings, in: P. R. Shukla, J. Skea, R. Slade, A. A. Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, J. Malley (Eds.), Climate change 2022: Mitigation of climate change. Contribution of working group III to the sixth assessment report of the intergovernmental panel on climate change, Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022, pp. 953–1048. Section: 9 Type: Book Section, DOI: 10.1017/9781009157926.011

[6] RTE, Futurs énergétiques 2050 : les chemins vers la neutralité carbone à horizon 2050, Technical Report, Réseau de transport d'électricité, 2022. https://www.rte-france.com/analyses-tendances-et-prospectives/bilan-previsionnel-2050-futurs-energetiques.

[7] L. G. Swan, V. I. Ugursal, Modeling of end-use energy consumption in the residential sector: a review of modeling techniques, Renew. Sustain. Energy Rev. 13 (8) (2009) 1819–1835. DOI: 10.1016/j.rser.2008.09.033

[8] I. C. M. Vaz, T. P. Scolaro, E. Ghisi, Understanding the environmental impact of residential electricity consumption in brazil: integrating top-down and bottom-up approaches with life cycle assessment, Sci. Total Environ. 986 (2025) 179774. DOI: 10.1016/j.scitotenv.2025.179774

[9] A. Zerroug, E. Dzelzitis, A study of modeling techniques of building energy consumption, Eng. Technol. Appl. Sci. Res. 10 (1) (2020) 5191–5194. Number: 1, DOI: 10.48084/etasr.3257

[10] S. Wang, X. Deng, H. Chen, Q. Shi, D. Xu, A bottom-up short-term residential load forecasting approach based on appliance characteristic analysis and multi-task learning, Electric Power Syst. Res. 196 (2021) 107233. DOI: 10.1016/j.epsr.2021.107233

[11] A. Kipping, E. Trømborg, Modeling aggregate hourly energy consumption in a regional building stock, Energies 11 (1) (2018) 78. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, DOI: 10.3390/en11010078

[12] C. Cerezo, J. Sokol, C. Reinhart, Three methods for characterizing building archetypes in urban energy simulation. a case study in kuwait city, MIT Web Domain (2015). Accepted: 2017-01-06T16:40:17Z Publisher: International Building Performance Simulation Association, https://dspace.mit.edu/handle/1721.1/106272.

[13] C. Ahern, B. Norton, A generalisable bottom-up methodology for deriving a residential stock model from large empirical databases, Energy Build. 215 (2020) 109886. DOI: 10.1016/j.enbuild.2020.109886

[14] T. M. 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 - ScienceDirect, 2021. DOI: 10.1016/j.enbuild.2020.110584

[15] 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, Energy Build. 184 (2019) 300–322. DOI: 10.1016/j.enbuild.2018.12.011

[16] V. Moreau, M.-H. Laurent, T. Berthou, B. Duplessis, P. Stabat, High-resolution building stock energy model for assessing the impact on the load curve of residential space heating electrification, Energy Build. 345 (2025) 116069. DOI: 10.1016/j.enbuild.2025.116069

[17] S. De Lauretis, F. Ghersi, J.-M. Cayla, Energy consumption and activity patterns: an analysis extended to total time and energy use for french households, Appl. Energy 206 (2017) 634–648. DOI: 10.1016/j.apenergy.2017.08.180

[18] M. Vellei, J. Le Dréau, S. Y. Abdelouadoud, Predicting the demand flexibility of wet appliances at national level: the case of france, Energy Build. 214 (2020) 109900. DOI: 10.1016/j.enbuild.2020.109900

[19] J. Pawlak, A. Faghih Imani, A. Sivakumar, How do household activities drive electricity demand? applying activity-based modelling in the context of the united kingdom, Energy Res. Soc. Sci. 82 (2021) 102318. DOI: 10.1016/j.erss.2021.102318

[20] P. Grünewald, M. Diakonova, The specific contributions of activities to household electricity demand, Energy Build. 204 (2019) 109498. DOI: 10.1016/j.enbuild.2019.109498

[21] T. Dogan, C. Li, H. M. Tseng, A. J. Su, P. Kastner, A bottom-up urban building energy model for evaluating thermal load electrification measures, J. Build. Perform. Simul. 0 (0) (2025) 1–28. Publisher: Taylor & Francis _eprint:, DOI: 10.1080/19401493.2025.2536261

[22] J. N. Castillo, G. G. Carrillo, L. O. Freire, B. P. Corrales, Energy modeling and simulation of a building to perform sensitivity analysis of energy consumption - ScienceDirect, 2022. DOI: 10.1016/j.egyr.2022.10.197

[23] S. Torabi Moghadam, J. Toniolo, G. Mutani, P. Lombardi, A GIS-statistical approach for assessing built environment energy use at urban scale, Sustain. Cities Soc. 37 (2018) 70–84. DOI: 10.1016/j.scs.2017.10.002

[24] M. Rit, R. Girard, J. Villot, M. Thorel, Y. Abdelouadoud, Calibration method for an open source model to simulate building energy at territorial scale, Energy Build. 293 (2023) 113205. DOI: 10.1016/j.enbuild.2023.113205

[25] C. Beltrán-Velamazán, M. Monzón-Chavarrías, B. López-Mesa, A new approach for national-scale building energy models based on energy performance certificates in european countries: the case of spain, Heliyon 10 (3) (2024). Publisher: Elsevier, DOI: 10.1016/j.heliyon.2024.e25473

[26] C. Song, Z. Deng, W. Zhao, Y. Yuan, M. Liu, S. Xu, Y. Chen, Developing urban building energy models for shanghai city with multi-source open data, Sustain. Cities Soc. 106 (2024) 105425. DOI: 10.1016/j.scs.2024.105425

[27] V. Moreau, Évaluation de scénarios d'électrification des usages thermiques sur la courbe de charge électrique du secteur résidentiel, These de doctorat, Université Paris sciences et lettres, 2023. https://theses.fr/2023UPSLM052.

[28] S. Thorve, Y. Y. Baek, S. Swarup, H. Mortveit, A. Marathe, A. Vullikanti, M. Marathe, High resolution synthetic residential energy use profiles for the united states, Sci. Data 10 (1) (2023) 1–23. Number: 1 Publisher: Nature Publishing Group, DOI: 10.1038/s41597-022-01914-1

[29] INSEE, Définition: IRIS, 2016. https://www.insee.fr/fr/metadonnees/definition/c1523.

[30] IGN, BD TOPO® | Géoservices, 2024. https://geoservices.ign.fr/documentation/donnees/vecteur/bdtopo.

[31] ADEME, Diagnostic de performance énergétique, 2022. https://observatoire-dpe-audit.ademe.fr/donnees-dpe-publiques.

[32] INSEE, Logements ordinaires en 2020, 2020. https://www.insee.fr/fr/statistiques/7705908?sommaire=7637890.

[33] B. Bell, H. Hersbach, A. Simmons, P. Berrisford, P. Dahlgren, A. Horányi, J. Muñoz-Sabater, J. Nicolas, R. Radu, D. Schepers, C. Soci, S. Villaume, J.-R. Bidlot, L. Haimberger, J. Woollen, C. Buontempo, J.-N. Thépaut, The ERA5 global reanalysis: preliminary extension to 1950, Q. J. R. Meteorologic. Soc. 147 (741) (2021) 4186–4227. DOI: 10.1002/qj.4174

[34] INSEE, Enquête emploi du temps - 2009–2010, 2010. DOI: 10.13144/LIL-0695

[35] INSEE, Congés des Français - 2010, 2010. https://www.insee.fr/fr/statistiques/1281344.

[36] ORE, Consommation annuelle d'électricité et gaz par IRIS, 2024. https://opendata.agenceore.fr/explore/dataset/consommation-annuelle-d-electricite-et-gaz-par-iris/information/?stage_theme=true&sort=-code_grand_secteur.

[37] ADEME, ENERTECH, RTE, Panel usages électrodomestiques, 2021. https://librairie.ademe.fr/batiment/4473-panel-usages-electrodomestiques.html.

[38] Enedis, Agrégats segmentés de consommation électrique au pas 1/2 h des points de soutirage <= 36kVA - Maille Nationale, 2024. https://opendata.enedis.fr/datasets/conso-inf36.

[39] Enedis, Agrégats segmentés de consommation électrique au pas 1/2 h des points de soutirage <= 36kVA - Maille Régionale, 2024. https://opendata.enedis.fr/datasets/conso-inf36-region.

[40] Wikipedia, Elbow method (clustering), 2025. Page Version ID: 1292194973, https://en.wikipedia.org/w/index.php?title=Elbow_method_(clustering)&oldid=1292194973.

[41] ADEME, Le dimensionnement des systèmes de production d'eau chaude sanitaire en habitat individuel et collectif, 2019. https://cegibat.grdf.fr/pdf/7984/3090.

[42] S. Kozarcanin, G. B. Andresen, I. Staffell, Estimating country-specific space heating threshold temperatures from national consumption data, 2019. arXiv:1904.02080 [physics], DOI: 10.48550/arXiv.1904.02080

[43] ADEME, Comment passer un hiver au chaud ?, 2025. https://librairie.ademe.fr/energies/8106-9873-comment-passer-un-hiver-au-chaud--9791029725210.html.

[44] ADEME, Pourquoi passer au thermostat programmable ?, 2022. https://librairie.ademe.fr/energies/5792-pourquoi-passer-au-thermostat-programmable-.html.

[45] S. Y. Abdelouadoud, R. Girard, A. Rogeau, buildingmodel, 2025. https://gitlab.com/energytransition/buildingmodel.

[46] Y. Chiche, building_eload, 2025. https://git.persee.minesparis.psl.eu/planeterr/building_eload.