Comprehensive Bottom-Up Methodology for Generating High-Resolution Yearly Building Load Profiles: A Case Study in Temperate Oceanic Climate

Mathieu PATIN[1a], Sylvie BEGOT[1], Frédéric GUSTIN[1], Valérie LEPILLER[1]

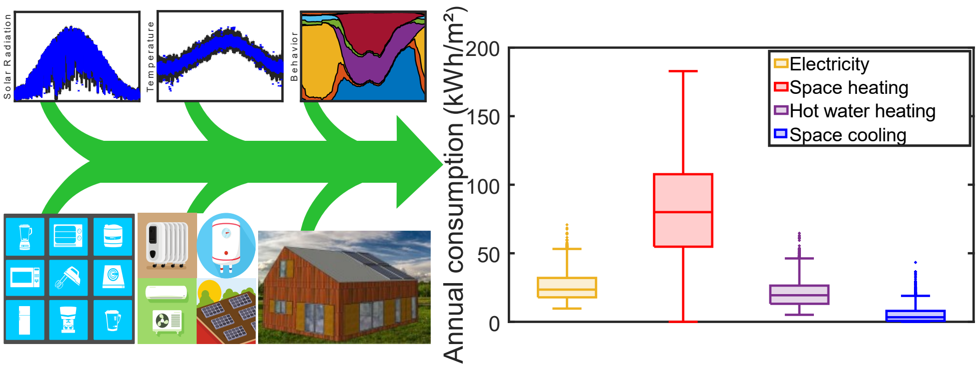
[1]Université de Franche-Comté, FEMTO-ST, FCLAB, UTBM, CNRS, Belfort, France

[a]mathieu.patin@femto-st.fr; 13 rue Ernest Thierry-Mieg, 90010 Cedex Belfort, France

Abstract

To facilitate the transition of residential buildings towards decarbonized energy sources, various energy systems are currently being investigated within the scientific community. The accurate sizing and performance evaluation of these systems heavily rely on the quality of input profiles. Addressing this necessity, a method for generating diverse, high-resolution, continuous, consistent demand and production profiles for a whole year is proposed. This method is structured in a modular fashion and draws upon a widely recognized demand model from literature. Each module of the method is systematically presented, and the parametrization process is detailed through a case study focusing on single-family houses in temperate climate. This comprehensive description facilitates the replication of the method in different geographical regions. Subsequently, a Monte Carlo simulation is employed, incorporating variations in weather conditions, building properties, and occupant behaviors. This simulation generates an openly accessible dataset comprising thermal, electrical and photovoltaic profiles for 3500 configurations. The generated weather and electricity demand profiles exhibit trends and variations that closely match the measured data. Photovoltaic production profiles were validated against PVGIS data, showing similar monthly variations and diversity. The generated dataset includes houses with energy consumption profiles that correspond to Energy Performance Certificates ranging from A to E.

Graphical abstract



Highlights

*Modeling of domestic energy load and PV production throughout the year

*Generation of diverse, high-resolution, continuous and consistent energy profiles

*Presentation of the model parametrization procedure through detailed case study

*Sharing of a dataset comprising 3500 configurations derived from French data

*Trend and dispersion analysis of the dataset in comparison to measured data

Keywords: Residential, thermal demand, electricity demand, Stochastic, photovoltaics, open data, bottom-up, Passive house, Central Europe

Nomenclature

Symbols

D	Day
T	Temperature (°C)
G	Solar irradiance (W.m^{-2})
t	time (minute)
η	Efficiency
Pr	Probability
C	Capacitance (J.K^{-1})
H	Heat loss coefficient (W.K^{-1})
$\dot{T}$	Temperature variation (°C.s^{-1})
$\dot{m}$	Hot water mass flow rate (kg.s^{-1})
P	Power (W)
A	Global irradiance multiplier (m²)
Obj	Objectif function

Indices

ref	Reference
mp	Maximum power point
amb	Ambient
irrad	Irradiation

hw	Hot water
heat	Heating
i	Interior of building
w	Water
cw	Cold water
set	Setpoint
db	Dead band
b	Building envelope node
rad	Radiator
cool	Cooling
s	Solar
o	Outside of Building
v	Ventilation

Abbreviations

PV	Photovoltaic
CREST	Centre for Renewable Energy Systems Technology
NOCT	Normal Operating Cell Temperature
TUS	Time Use Surveys
INSEE	Institut National de la Statistique et des Études Économiques (France)
ADEME	Agence De l'Environnement et de la Maîtrise de l'Énergie, (France)
DHW	Domestic Hot Water
RMSE	Root Mean Square Error
GHI	Global Horizontal Irradiance

1. Introduction

1.1. Research Background

Climate change represents a global phenomenon requiring urgent international action. The building sector stands out as one of the most energy-intensive industries worldwide, accounting for 30% of global final energy consumption and 26% of total greenhouse gas emissions from the energy sector in 2022 [1]. Addressing climate change, mandates a reduction in energy consumption within this sector and a transition away from high-carbon energy sources like natural gas.

Various energy solutions are being explored globally to meet building demands in a decarbonized manner, including photovoltaics, thermal solar panels, battery storage, hydrogen storage, heat pumps, biogas boilers, new generation heat networks or waste heat recovery systems. However, accurately assessing and sizing these energy systems require high-resolution electric and thermal load profiles.

Monitoring building consumption in situ is often costly and time-consuming. Jin et al. examined available open datasets and their utility in the literature [2], identifying only 33 open datasets, among which only 25 provide yearly data, 6 offer monthly data, and just 2 offer hourly or sub-hourly data. Additionally, only 6 datasets pertain to the European Union, none of which provide sub-yearly resolution. To address this data gap, building energy models are commonly utilized to generate time series suitable for multi-energy simulations.

Building energy models employ two main approaches: top-down and bottom-up [3]. Top-down models utilize macroscopic data on national, regional, or local scales, which are then disaggregated based on various economic, social (employment rate, income, etc.), technical (set temperature, equipment, etc.), or physical parameters (vintage, insulation type, etc.) to the building level.

However, these models may lack the precision required for energy system simulations at the individual building level, which is primordial for the assessment of individual energy systems [4]. Furthermore, top-down approaches may struggle to maintain consistency between different profiles (e.g., thermal demand, electric demand, local electric production), which is also crucial for evaluating energy systems, particularly in urban photovoltaic modeling. For instance, a decrease in solar radiation can lead to simultaneous increases in energy demand (for lighting and heating) and decreases in local PV production, resulting in a sudden gap between energy production and demand. Capturing these gaps is essential for evaluating the potential of energy storage technologies. For these reasons, top-down models are less suitable for generating electrical and thermal load profiles needed for evaluating the performance of residential multi-energy systems.

On the contrary, bottom-up models begin with the specific attributes of a building to derive its thermal or electrical load profiles. Consequently, extensive data on household equipment, occupants' behavior, and weather conditions are necessary for electricity load simulations. Similarly, thermal data necessitates information on inhabitants' behavior and weather, along with detailed data on the physical properties of the building. These methods, through higher parameterization and increased modeling granularity, generate coherent and precise profiles suitable for multi-energy systems assessment.

1.2. Literature review

Numerous bottom-up models currently exist in the literature. They can be further categorized into black, white, and gray box methodologies. Black box models, also known as data-driven models, utilize historical data to project load profiles based on a new set of inputs such as building characteristics, occupancy, and weather conditions. Techniques include linear regression, support vector machines, and neural networks [5]. Wang et al. conducted a comparative study on the performance of 12 black box algorithms, identifying linear, ridge, and lasso regression as underperforming methods, while Extreme Gradient Boost and Long Short-Term Memory were highlighted as superior long-term and short-term prediction methods, respectively [6]. However, black box approaches are limited to extrapolating existing data and cannot generate datasets from scratch.

Conversely, white box models, also referred to as physical models, utilize the physical properties of buildings to compute thermal exchanges between different zones within the building and with the external environment. Commonly employed white box software includes EnergyPlus, TRNSYS, Pleiades, and Dymola.

Hong et al. proposed a method to generate synthetic smart meter data using EnergyPlus via OpenStudio (2020) [7]. This model incorporates variations in four sectors (weather, building envelope, building operation, and inhabitant behavior) across sixteen types of commercial buildings (e.g., offices, restaurants, schools, hotels) under sixteen U.S. climates, for five vintages (2004, 2007, 2010, 2013, and 2016) and three building operation scenarios (good, average, and poor). An agent-based model simulates building occupancy, while appliance usage is represented by a linear function dependent on occupancy count. The resulting profiles can serve as baselines for testing energy algorithms.

Chaudhary et al. introduced in 2023 a bottom-up approach for generating profile datasets suitable for deep neural network training, known as synconn_build [8]. This method employs Python scripts to automate the setup of EnergyPlus software, which then simulates temperature variations and the corresponding heating and cooling loads. Variations are introduced through solicitation profiles (weather, occupancy, lighting, appliances), three perturbation signals (heating/cooling setpoint, control and windows' opening), and noise on the heating/cooling temperature setpoint signal.

Ferando et al. presented in 2020 eight commonly used bottom-up physics-based urban energy models, three of which employ a white box approach, all based on EnergyPlus (umi, CityBES, and URBANopt) [9]. However, for accuracy, the design of multi-energy systems often requires extensive dataset inputs to conduct uncertainty and off-design analyses. While classical white-box models can be computationally expensive for generating larger datasets, there is a growing need for faster methods such as gray-box approaches. The most prevalent simulation technique in this domain is reduced-order resistor capacitance (RC). Ferando et al. identified five commonly used tools employing this approach: CitySim, SimStadt, OpenIDEAS, CEA, and TEASER [9].

The complexity of reduced-order RC models varies depending on their order. Shamsi et al. proposed a procedure for determining the most suitable order of RC models for commercial building studies [10]. This method relies on various building characteristics, with significant impacts identified including total interior floor area, glazed area, number of floors, number of zones, presence of solar facades, heat demand profile, installed heating/cooling systems, and renovation history. Validation of this approach was conducted using a forward selection method, demonstrating consistency between the order identified by the proposed method and the forward selection procedure.

Roth et al. (2020) created SynCity which adopts a hybrid approach, combining elements of both top-down and bottom-up methodologies to generate hourly electric and thermal load profiles on a city scale [11]. Initially, a machine learning algorithm was developed using annual consumption and physical properties from 15,000 buildings in New York City, which then estimated the annual consumption of 1 million buildings based on their physical properties. Subsequently, estimated annual consumptions were used to allocate each building to three reference physical models among nineteen. The respective weightings of these three reference models for each building were determined using convex optimization, comparing the aggregated results to citywide consumption. This method allows the authors to generate over 1 million hourly energy profiles in New York City utilizing only open datasets.

Guo et al. (2023) also utilized a mixed approach [4]. Initially, bottom-up models were generated based on building characteristics such as building types, roof types, vintage, building layout, and footprints. Subsequently, a top-down approach was employed to reduce uncertainties in the model inputs. Evaluation on a district in Leeste, Germany, indicated a mean absolute error percentage of 2%, contrasting with 15% when solely employing bottom-up approaches.

The majority of building dataset-generating tools use, for computing time reasons, time steps above the minute. Yearly values are used for (SimStadt, umi), hourly time steps for (CitySim, Roth et al., CEA, Urbanopt, Teaser, SynCity), and 15-minute time steps for (Smart-E, synconn_build). However, to accurately capture the performance of energy systems, high resolution is required. For example, some energy systems exhibit slow start-ups and ramp-ups, and assessing the impact of these slow dynamics on performance requires fully capturing the abrupt changes that can appear in real domestic loads.

Some models, like CityBES and OpenIDEAS, can produce data with a resolution of 1 minute. However, OpenIDEAS does not produce PV production profiles, and CityBES does not allow for stochastic variation of occupancy, heating setpoints, appliances ownership and usage, etc. Moreover, these two models, as well as most aforementioned tools, require expertise specific to building modeling, which can limit the number of users able to use them.

Conversely, premade open datasets offer easier utilization for researchers specializing in energy systems (electric, thermal, or multi-energy) who may lack expertise in building modeling. The drawback is that those datasets cannot be adapted to new parameters (new building, new regions, etc.). For the European region in a similar way as for measures datasets, available synthetic datasets are limited. Ali et al. developed a synthetic building dataset for 1 million buildings, including heating, electricity, and hot water consumption, as well as PV production in Dublin but only for comprising annual values [12]. The Joint Activity Scenarios and Modeling share open dataset for Swiss buildings, notably in the work of Murray et al. that generated hourly profiles for 14 residential and commercial buildings under 7 retrofit scenarios [13]. In the same way, Iturralde et al. shared hourly profile for one multifamily house and one single-family house in Central Europe [14].

For the purpose of energy system design and sizing, datasets need to be high-resolution to assess dynamic system performances, consistent to capture gaps between local production and consumption, continuous throughout the year to evaluate storage potential, pre-made to accommodate a larger number of researchers, and varied enough to allow for uncertainty and off-design analysis. In the literature there is a lack of corresponding openly available dataset tailored to Europe.

The Centre for Renewable Energy Systems Technology (CREST) demand model, widely recognized in the literature, has the capability to generate such profiles without necessitating specific expertise in building energy modeling. Leveraging statistical techniques, this model stochastically generates consumption profiles for single-family houses. Initially conceived as a domestic occupancy model, it has since evolved into an electricity demand model. Subsequent enhancements incorporated a thermal component, encompassing heating, and more recently, cooling [15, 16, 17, 18, 19].

The electrical and thermal load profiles, as well as PV production profiles generated by the model, are consistent and high resolution (1 minute). Only a few inputs are necessary to generate profiles, making this method usable by a large range of researchers. Moreover, the model is able to generate profile datasets with high diversity, relying on variations in weather conditions, building envelopes, building sizes, inhabitant occupancy, lighting/appliances (variation in ownership and usage), and inhabitant heating/cooling habits. The CREST model can be compared to the other presented bottom-up methods in Table 1.

However, the current iteration of the model still has limitations. It simulates one day at a time, resulting in discontinuities between consecutive days when constructing yearly profiles. Additionally, compared to conventional scientific programming languages, the VBA code environment used, can exhibit slower performance, resulting in prolonged execution times for yearly simulations. Furthermore, the downside of not requiring complex user inputs is that the model is rigid and is limited to the specific regional setting it originates.

Therefore, a new method is created to enhance the model's capabilities to an annual simulation by enabling annual simulations while maintaining consistency across days. A reproducible and comprehensive procedure is proposed for the reparameterization of the CREST model for different weather conditions, building types, inhabitant behaviors… The procedure is illustrated through a case study using French statistical data and weather data from a temperate oceanic climate (as defined by the Köppen-Geiger climate classification [20]). Finally, utilizing the enhanced model, Monte Carlo simulation is conducted to generate an open dataset suitable for assessing energy system performance and sizing.

The intended novelty of this paper includes:

- Development of a method for the generation of yearly, high-resolution, consistent residential electricity, heating and cooling load profiles as well as PV production profiles
- Presentation of the model parametrization process, using the case study of single-family houses in temperate oceanic climate, such that it can be replicated for other regions
- Analysis and discussion of profiles generated by Monte Carlo simulation for 3500 unique houses' configurations
- Provision of the generated dataset for public access.

To achieve these objectives, the following tasks must be undertaken:

- Modeling the varying weather conditions, including solar radiation and temperature fluctuations
- Modeling the appliance and lighting loads typical of Central European households
- Modeling the space heating, hot water, and cooling demands specific to typical single-family houses in a temperate oceanic climate.
- Integration of these models into a Monte Carlo simulation framework to generate a comprehensive load profile dataset incorporating variations in weather, building characteristics, occupant behavior, and equipment usage

Table 1: Summarize of existing residential profiles generation methodologies

Model	**Modeling approach**	**Resolution**	**Profiles generated**	**PV**	**Variation**
Umi (2013) [21]	White box (Energy plus)	Yearly	Electricity, hot water, heating and cooling	Yes	Weather, building properties
CitySim (2015) [22]	Gray-box (reduced-order RC)	Hourly	Electricity, hot water, heating and cooling	No	Weather, building properties
Simstadt (2015) [23]	Gray-box (reduced-order RC)	Yearly	Electricity, hot water, heating and cooling	Yes	Weather, building properties
OpenIDEAS (2015) [24]	Gray-box (reduced-order RC)	1 minute	Electricity, hot water, heating and cooling	No	Weather, Building properties, occupancy
CityBes (2016) [25]	White box (Energy plus and OpenStudio)	1 minute	Electricity, hot water, heating and cooling	Yes	Weather, building properties
CEA (2016) [26]	Gray-box (reduced-order RC)	Hourly	Electricity, hot water, heating and cooling	Yes	Weather, building properties
TEASER (2018) [27]	Gray-box (reduced-order RC)	Hourly	Electricity, hot water, heating and cooling	No	Weather, building properties
URBANopt (2020) [28]	White box (Energy plus)	Hourly	Electricity, hot water, heating and cooling	Yes	Weather, building properties
Hong et al. (2020) [7]	White box (Energy plus and OpenStudio)	15 minutes	Electricity, heating and cooling	No	Weather, building properties, occupancy and building operation
CREST (2020) [15]	Gray-box (reduced-order RC)	1 minute	Electricity, hot water, heating and cooling	Yes	Weather, building properties, occupancy, building operation, lighting and appliances (ownership and usage)
SynCity (2020) [11]	Hybrid (top-down and bottom-up)	Hourly	Electricity, hot water, heating and cooling	No	Building properties
Guo et al. (2023) [4]	Hybrid (top-down and bottom-up)	Hourly	Electricity, hot water and heating	No	Building properties
synconn_build (2023) [8]	White box (Energy plus)	15 minutes	Heating and cooling	No	Building properties, occupancy, lighting, appliances, setpoint, and building operation

2. Method

As previously stated, the decision was made to use the model established by CREST as foundation, the original model is available as an open-source VBA code [15]. The proposed new model allows for:

- Continuous simulations for a given time period (weekly, monthly, yearly),
- Faster execution time, through programming environment change (Matlab®) and structural changes (vectorization),
- Finer clearness modeling to capture monthly weather patterns,
- Smother temperature variation through the year,
- More complex PV modeling to capture temperature dependency,
- Flexible lighting calibration function allowing to dynamically align to a specific consumption,
- Handling of conditional probability for appliances ownership,
- More temporal variation of hot water demand through variation of cold-water inlet,
- Inclusion of solar radiative gain on the building envelope,
- Variation in window shutters use through variation in internal solar radiative gain,
- Coherence between the occupancy and heating schedules,
- More stable cooling loads profiles even for high power cooling systems.

The model parametrization procedure is presented in detail using the case study of temperate oceanic climate as well as the addition of a low consumption building. Moreover, the model is updated with more recent data, notably for the introduction of IT appliances, and corrected when needed. The complete methodology regarding the final framework and the parameterization process are provided in the subsequent sections.

2.1. Weather model

2.1.1. Solar radiation

The CREST model employs stochastic processes to generate daily profiles of solar irradiation and outdoor temperature based on historical data. A comprehensive description of the solar irradiation modeling is provided by Richardson and Thomson [29]. Initially, the program computes an irradiance profile for a clear sky scenario using solar angle and optical depth. The impact of cloud cover on this value is estimated through a clearness coefficient ranging from 0 to 1. To account for its diurnal variation, a Markov chain is employed. The final irradiation is obtained by multiplying the clear sky value by the clearness index.

In transitioning to year-round modeling, the daily clear sky irradiation model is executed for each day of the year and then aggregated to generate yearly irradiation profiles. For the clearness index, the Markov chain is computed for every minute of the year.

To address the complexity of cloud coverage variations throughout the year, the model is enhanced by employing 12 transition matrices, each specific to a month. This refinement acknowledges the reality that certain months are more prone to cloudy days than others. Utilizing monthly matrices helps mitigate the issue of overestimating solar irradiation in winter months and underestimating it in summer months. This correction is particularly evident in PV production profiles, which tend to be significantly overestimated in January and December in the CREST model.

In terms of parameterization, the clear sky irradiation module already incorporates variations in longitude, latitude, and meridians and thus requires no further adaptation. However, in the original CREST model, the clearness model was initially parametrized using weather data from the UK and remained unchanged even after incorporating Indian climate zones. It's important to note that the clearness index is closely linked to specific weather patterns, necessitating adaptation to the region under study in order to accurately capture these patterns.

The Markov chain utilized for modeling relies on a 101 X 101 transition matrix, which necessitates reparameterization. The parametrization method utilizes necessitate weather data with a 1-minute resolution. In the present case weather data from the FEMTO ST/FCLAB laboratory weather station in Belfort (France) are used. Irradiation profile data from 2015 to 2021 are separated by month in 12 sets. These sets are compared with clear-sky irradiance at the same time of the year to estimate the cloud cover index sets. These indices are then converted into 101 intervals with a resolution of 1%. Finally, the matrix is generated by calculating for each state the frequency of transition to the 101 states.

2.1.2. Outside Temperature

The model for outdoor temperature is detailed in the work of McKenna and Thomson [19]. It employs an autoregressive moving-average model based on monthly reference values to predict the average temperature for a given day. The cumulative solar irradiance throughout the day is utilized to estimate both the daily maximum and minimum temperatures. The minimum temperature corresponds to sunrise, while the maximum temperature is determined by the moment when the ratio of the sum of ground irradiance to the sum of extraterrestrial irradiance is maximized. Between these two extremes, temperature variation is influenced by solar irradiance during the day and cloud cover at night.

Transitioning from a daily to a yearly model necessitates a change in approach. In the original model, the temperature at sunrise (minimum temperature) on day D dictates the temperature decrease during the preceding evening and following morning of day D. This setup introduces inconsistencies between consecutive days, which are addressed by utilizing the sunrise temperature of day D+1 to determine the nighttime temperature evolution between days D and D+1.

Additionally, the original model employs monthly references, leading to abrupt switches in daily temperature between months. To address this issue, the proposed model now utilizes sinusoidal functions to create smoother day-to-day transitions.

The parametrization methodology uses openly available data from the Open-Meteo historical weather API [30, 31, 32, 33]. The daily mean temperatures for each day from 2000 to 2023 are extracted and used to fit the sinusoidal functions. The results and presented in (Figure 1).

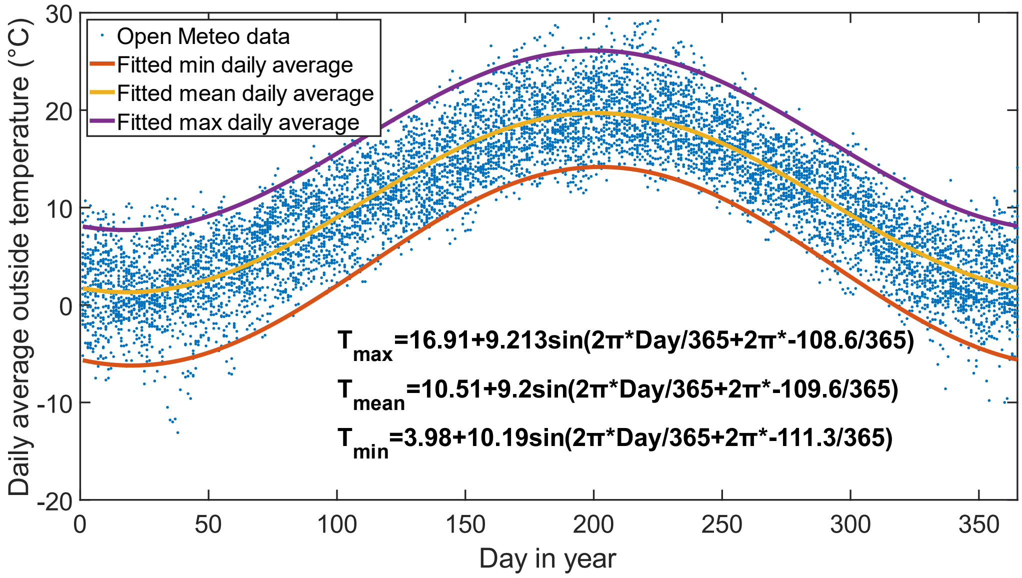


Figure 1: Adjustment of the average reference temperature function

2.2. Photovoltaics model

The CREST model incorporates a module for the self-generation of electricity by photovoltaic (PV) panels. The operation of this module is comprehensively detailed in the work of Richardson and Thomson [29]. The PV module utilizes irradiance data from the climate module to calculate the incident irradiance on the solar panel, considering its tilt and orientation. Subsequently, the power output of the panel is estimated based on the given irradiation, the panel's efficiency, and its surface area.

Similar to the adjustments made to other modules, modifications are applied to the PV module to enable its operation over variable periods instead of just a single day. Additionally, the estimation of power output for a given irradiation is refined to incorporate the effect of panel temperature on efficiency. To address this, the model proposed by Arsalis et al. is utilized [34]. Panel temperature is estimated using its Normal Operating Cell Temperature (NOCT) condition and reference condition performances (Equation (1)). This temperature value is then utilized to modulate the panel's efficiency (Equation (2)) with its maximum power point efficiency temperature coefficient (μ_{mp}). In the profile generation, all PV panels face south and are inclined by 40°. For performances an efficiency of 16.9% and a maximum power point efficiency temperature coefficient of -0.38%.C^{-1} are used [35].

$$T_{Panel} = \frac{G_{Panel}(t)}{G_{ref}}\left(1 - \frac{\eta_{PV,ref}}{0.9}\right)\left(T_{NOCT} - T_{amb,NOCT}\right) + T_{amb} \tag{1}$$

$$\eta_{PV} = \eta_{PV,ref}\left(1 + \mu_{mp}\left(T_{Panel} - T_{amb,NOCT}\right)\right) \tag{2}$$

2.3. Behavior model

Model for inhabitant behavior is comprehensively described in the work of McKenna et al. [36]. Similar to the clearness index modeling, behavioral modeling relies on a Markov chain that governs both the number of active inhabitants and the number of active inhabitants. Given the significant variation in the probability of a sleeping resident waking up between 2 a.m. and 7 a.m., distinct matrices are employed for each 10-minute interval throughout the day, totaling 144 matrices. Different sets of matrices are utilized based on the number of inhabitants (up to 5) and the day of the week (weekday or weekend), resulting in a total of 1440 matrices ranging from 16 (4*4) states for 1 inhabitant to 1296 (36*36) states for 5 inhabitants.

To run the model over a year while maintaining an alternation between 5 typical weekdays and 2 typical weekend days, the matrices are concatenated into a weekly matrix. The program iterates over these matrices every 7 days.

Transition matrices are parameterized using data from Time Use Surveys (TUS). Ideally, the model would be parameterized with the studied country data to align fully with the test case study. However, in our test case for France, the French surveys have a different structure compared to those in the UK, as they only select one inhabitant per dwelling [37]. This limitation prevents the data from capturing the interactions between inhabitants, which are fundamental components of the behavior model. Therefore, parameterization is conducted using UK data [38], representing an update of the CREST, moving from 2003 data to 2015.

The 2015 dataset comprises 16533 notebooks from 4230 dwellings. Each notebook records the activities and positions of residents over the course of a day at 10-minute intervals. Dwellings are first classified by inhabitant numbers and the survey day. Then, the state corresponding to each dwelling at each time step is determined based on the number of present inhabitants present and the number of active inhabitants. Finally, the frequency of each state transition is calculated and used to populate the transition matrices.

2.4. Lighting model

The operational details of the lighting consumption module are outlined in the work of Richardson et al [17]. This module utilizes a list of bulb configurations to assign lighting fixtures to a building. Upon initialization of a building, a draw is conducted to assign it to one of the hundred configurations. At each time step, a probability, comprising an irradiation threshold and factors related to occupancy, calibration, and usage, is compared against a random draw to determine whether a bulb is switched on (Equation (3)). Once a bulb is activated, it remains illuminated for a stochastically determined duration (ranging from 1 to 259 minutes) or until there are no active inhabitants present.

$$Pr_{act} = flag_{irrad} \times Factor_{occupancy} \times factor_{relative,usage} \times factor_{calibration} \qquad (3)$$

The lighting model utilizes only one temporal loop, which is simply extended to cover a year's duration. As described earlier, the model employs a calibration factor to align lighting consumption with a specified annual value. However, in the original model, this factor remains fixed, despite expectations that buildings of different sizes will exhibit varying annual consumption levels. To address this, it is proposed to implement a function dependent on the targeted consumption to adjust this factor accordingly.

Parametrizing the model for a new region involves several steps. Firstly, it is essential to create 100 new bulb configurations from which to select. This process involves drawing from a normal distribution to assign a number of bulbs, with the mean based on the French average (31.1 bulbs) [39] and a standard deviation of 4.4. Subsequently, each bulb is paired with a technology and a power according to the distribution of technologies obtained from the same study (see Table 2).

For the calibration factor function, lighting simulations are conducted for 100 buildings with calibration factors varying from 0 to 0.025. A curve is then fitted to the results (see Figure 2). This approach enables the selection of the calibration factor during initialization based on the desired annual consumption. In the present case, the calibration factor was adjusted depending on the building's area, with a reference value for France set at 1.7 $kWh.m^{-2}.year^{-1}$ [39].

Table 2: Distribution of bulb technologies (from [39])

Technology	**Proportion in the home**	**Average power per bulb (W)**
LED	48%	6.9
Halogen	28%	45.6
Incandescent	9%	52.9
Fluorescent	15%	18.5
Average per dwelling	31.1 bulbs	733 W

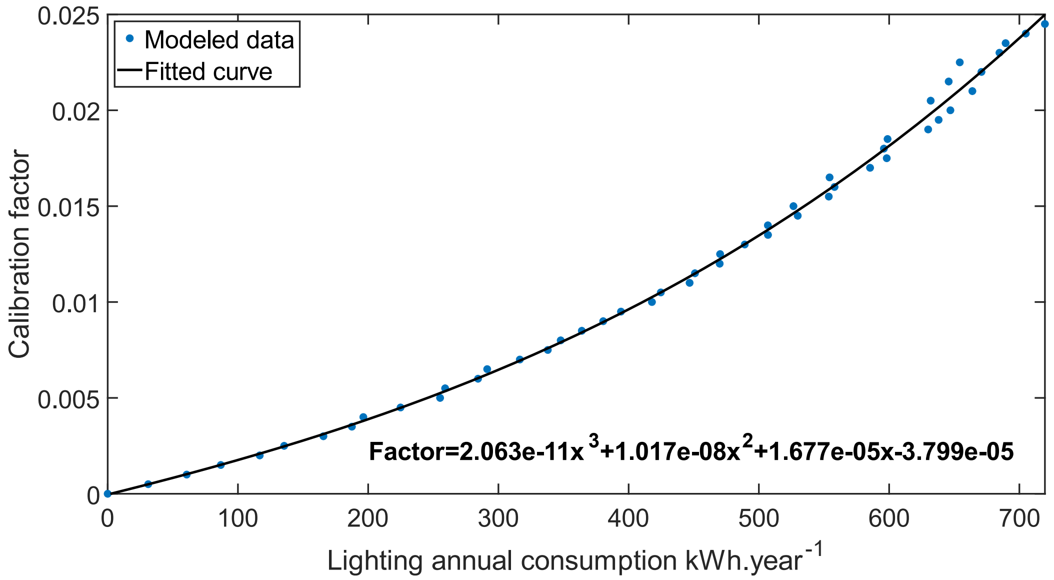


Figure 2: Example of adjustment of the calibration factor, French test case

2.5. Appliances model

The electrical appliance load modeling constitutes a fundamental aspect of the electrical module within the CREST model, extensively elaborated in the work of Richardson et al. [18]. Its operation can be summarized in three steps. Firstly, a set of electrical appliances is stochastically associated with a building, utilizing random draws and ownership probabilities. Secondly, for each appliance, the switch-on times are estimated by amalgamating corresponding daily activity probability profiles throughout the day with specific annual consumption data. Once switched on, an appliance operates for a predetermined reference period or until the building becomes unoccupied. Finally, these operational periods are coupled with a constant electrical load derived from a normal draw around a reference mean, with a standard deviation set at one-tenth of this value.

Similar to the behavior model, to simulate the model over a year while maintaining an alternation between 5 typical weekdays and 2 typical weekend days, the activity matrices are concatenated into a weekly matrix. The program iterates over these matrices every 7 days.

The parametrization of the model to a new region requires the modification of the appliance list, ownership, mean cycle power, mean cycle length, and annual consumption. For this, in our test case for France, three studies with varying levels of precision are utilized. These include studies from the Institut National de la Statistique et des Études Économiques (INSEE) (16,000 households, ownership data only) [40], one from Gifam, a consortium of domestic appliance brands, (6,500 households, ownership data only) [41] ,and one from the Agence De l'Environnement et de la Maîtrise de l'Énergie (ADEME) (101 households, ownership, usage, and consumption data) [39]. Leveraging findings from these studies alongside values from the reference model, the appliance model is reparameterized, with summarized values provided in Table 3. Notably, one significant change from the reference model is the utilization of conditional probability for appliances beyond the first for televisions, laptops, and desktops to models the difference in consumption between a main television and a secondary one.

The activity profiles are updated using the same time use survey as the behavior model [38]. Within each of the 4230 households, activities conducted within the buildings are extracted at each time step, and their frequencies are utilized to reconstruct the activity profiles. The considered activities remain consistent with the reference model (cooking, laundry, house cleaning, ironing), except that watching TV is altered to multimedia usage to accommodate the increased prevalence of digital appliances in households (laptops, desktops, etc.). Appliances not linked to these activities can either be automatically switched on (at any time step) or associated with occupancy (at any time step with active occupancy).

Table 3: Summary of electrical appliance specifications

Appliance	**Associated activity**	**Probability of ownership**	**Mean cycle length (minute)**	**Mean cycle power (W)**	**Average annual consumption (kWh.year^{-1})**
Refrigerator	Automatic	1.00	18	200	344
Freezer	Automatic	0.56	22	300	288
Wine cellar	Automatic	0.09	18	200	193
Washing machine	Laundry	0.96	138	222	101
Dishwasher	Cooking	0.61	82	714	162
Tumble dryer	Laundry	0.34	97	1017	301
TV 1	Multimedia	0.95	73	70	189
TV 2	Multimedia	0.44	73	70	58
TV receiver	Automatic	0.90	13	13	87
Games console	Multimedia	0.66	162	52	103
Hi-Fi and speakers	Multimedia	1.00	60	100	25

Internet box	Automatic	0.86	1315	12	97
Laptop 1	Multimedia	0.82	193	19	22
Laptop 2	Multimedia	0.28	193	19	22
Laptop 3	Multimedia	0.35	193	19	22
Desktop 1	Multimedia	0.33	234	76	123
Desktop 2	Multimedia	0.18	234	76	123
Cell phone or tablet	Occupation	0.96	60	5	3
Printer	Multimedia	0.83	4	355	12
Electric stove	Cooking	0.34	3	1000	164
Cooker hob	Cooking	0.67	16	1265	138
Built-in oven	Cooking	0.56	51	919	146
Tabletop oven	Cooking	0.27	51	371	59
Microwave	Cooking	0.89	30	1250	39
Multi-cooker	Cooking	0.38	51	101	16
Kettle	Cooking	0.62	3	2000	49
Coffee machine	Cooking	0.97	3	1000	28
Vacuum cleaner	Housework	0.82	20	2000	9
Iron	Ironing	0.59	30	1000	27
Various constants	Automatic	1.00	Constant	5	48
Ventilation	Automatic	1.00	1253	32	241

2.6. Hot water model

The Domestic Hot Water (DHW) tank is modeled based on its heat capacity (C_{tank}) and heat loss coefficient (H_{loss}), with its operation detailed in the work of McKenna and Thomson [19]. This model governs the variation in tank temperature ($\dot{T}_{tank}$) in response to hot water demand flow rate ($\dot{m}_{water}$), heat exchanges with the air, and gains from the heating system ($P_{heat,hw}$) (Equation (4)). Hot water demand profiles are modeled similarly to electrical appliance use, considering only four sources (basin, sink, bath, and shower) and two activities (cooking and self-care).

The heating power supplied to the hot water tank is regulated by a thermostat signal with a dead band of 5°C. Hot water setpoints are assigned, at initialization, via a draw from a discrete probability distribution. When heat is required, the corresponding power is calculated using Equation (5). To ensure a realistic profile, the power is then limited by the maximum power of the heating system to obtain the thermal demand associated with hot water usage.

In the CREST model, the dead band is not factored in the power calculation. Consequently, the domestic hot water (DHW) tank only warms up to the setpoint and never exceeds it. As a result, the DHW tank remains continuously ON for the entire simulation after being switched on once. To address this issue, the dead band is incorporated into the power calculation (Equation (5)) resulting in less constant heat intermittent spikes in demand. Furthermore, the model initially utilized a constant cold-water inlet temperature, leading to a failure in capturing the seasonality of DHW demand observed in measurements [39]. To rectify this, the model now incorporates varying cold-water inlet temperatures based on monthly reference values. Additionally, the model has been modified to exclude heating systems' efficiencies from the calculation. This adjustment aims to provide generic heat demand profiles that can be more readily adapted to various types of energy systems under study.

$$C_{tank}\dot{T}_{tank} = P_{heat,hw} - H_{loss}(T_{tank} - T_i) - c_w\dot{m}_{hw}(T_{tank} - T_{cw}) \tag{4}$$

$$P_{heat,hw,ideal} = C_{tank}(T_{set} + T_{db} - T_{tank}) + H_{loss}(T_{tank} - T_i) + c_w\dot{m}_{hw}(T_{tank} - T_{cw}) \tag{5}$$

The parametrization procedure involves determining both the volume of hot water drawn and the corresponding power load. In our French test case, the average hot water consumption is estimated at 56 liters at 40°C per inhabitant [42], with an average of 2.15 inhabitants per household [43], it results in 120.4 liters per household. This aligns with the 120L value utilized in the reference model, hence no changes are made to the consumption volumes. However, the power loss associated with the drawn hot water is now calculated using a value of 40°C to correspond to the reference value (Equation (6)). Furthermore, the daily constant cold-water temperature is determined by a normal draw, with the average and standard deviation specified in Table 4 (sourced from [43]).

$$C_{tank}\dot{T}_{tank} = P_{heat,hw} - H_{loss}(T_{tank} - T_i) - c_w\dot{m}_{hw}(40 - T_{cw}) \tag{6}$$

Table 4: Average monthly cold-water temperature (from [43])

Month	**Mean (°C)**	**Standard deviation (°C)**
January	11	2
February	11	2
March	12	2
April	15	2
May	17	3
June	19	3
July	21	3
August	21	3
September	20	3
October	17	3
November	15	2
December	12	2

2.7. Building thermal model

The reference thermal model was initially described for the heating part in the work by McKenna and Thomson [19], and later for the cooling part in the work by Barton et al. [15]. The change to annual modeling is made through an increase of the time loop. To enhance execution efficiency within the Matlab® environment, the thermal model is implemented using vectorization. Meaning that instead of looping through every building one by one the stats' values of the thermal are concatenated in vectors which are then introduced in the equations. Leveraging the architecture of the Matlab® environment in this manner significantly improves execution speed, particularly for simulations involving a large number of buildings. As a result, larger datasets can be generated rapidly. Moreover, this vectorization allows for the coupling of the thermal part of the model with complex energy system modeling. This integration enables the assessment of how a given energy system influences the thermal behavior of buildings throughout the year and vice versa.

2.7.1. Inside temperature

The model adopts a gray box approach as illustrated in Figure 3. The equations governing temperature variations are Equation (7) for building envelope temperature ($\dot{T}_b$), Equation (8) for indoor temperature ($\dot{T}_i$), and Equations (10) and (11) for heating ($\dot{T}_{rad}$) and cooling ($\dot{T}_{cool}$) emitter temperatures respectively.

Those equations are similar to those in the reference model. However two changes were made. First, for the introduction of solar radiation into the energy balance at two distinct points: within the building envelope node (through $A_{s,exterior}$) and through the windows in the house interior node (through $A_{s,interior}$), adopting a methodology similar to that of Berthou et al [44]. Furthermore, different values of $A_{s,interior}$ are employed for the cooling and heating periods . Indeed, when interior temperature increases the first reaction is often to decrease the admission of solar radiation before resorting to cooling systems. Adding this second value of $A_{s,interior}$ aims at capturing the change in windows occultation resulting from this behavior, which is not possible in the CREST model.

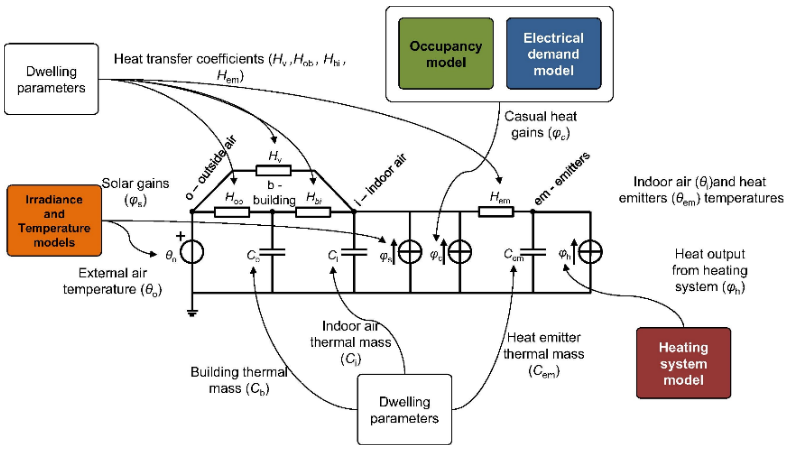


Figure 3: CREST thermal model, from [19]

$$C_b\dot{T}_b = H_{ob}(T_o - T_b) + H_{bi}(T_i - T_b) + A_{s,exterior}G \tag{7}$$

$$C_i\dot{T}_i = H_{bi}(T_b - T_i) + H_v(T_o - T_i) + H_{rad}(T_{rad} - T_i) + H_{cool}(T_{cool} - T_i) + H_{loss}(T_i - T_{tank}) + P_{intern} + A_{s,interior}G \tag{8}$$

With :

$$P_{intern} = P_{occupacy} + P_{lighting} + P_{appliances} \tag{9}$$

$$C_{rad}\dot{T}_{rad} = H_{rad}(T_i - T_{rad}) + P_{heat,air} \quad (10)$$

$$C_{cool}\dot{T}_{cool} = H_{cool}(T_i - T_{cool}) + P_{cool} \quad (11)$$

To parameterize the thermal model or to add a new building archetype, identification of characteristic values of the building model such as transmission coefficients, heat capacities, and solar radiation multiplication are necessary. The procedure is detailed using as a case study the addition of a very low consumption single-family house.

The procedure consists of using a complex white-box simulation of the building as a reference to identify the parameters of the reduced order RC model. For the present study the software Pléiades® is used to generate the reference indoor temperature profiles and heating loads [45]. The simulation is conducted on a detached house located near Dijon, France.

Detailed information on the architecture and composition of the house is provided by Topoïein Studio, an architecture and urban planning firm. This house has been awarded Passif house certification by the Passivhaus Institute [46], aligning with the desired low-carbon building type. The house is insulated with glass wool, cellulose, and polyurethane, with thermal conductivities of 0.04, 0.04 and 0.02 $W.m^{-1}.K^{-1}$, respectively. The complete composition of the exterior and interior walls, as well as the windows inventory and thermal bridges, can be found in Tables A1 to A8.

The dwelling comprises two floors, totaling approximately 114 m^2 of living space. The first floor includes the living room, kitchen, bathroom, toilet, storeroom, and bedroom. Half of the first floor consists of a hollow space over the living room, while the other half contains two habitable rooms. The asymmetrical roof features 56.3 m^2 facing south and 38.5 m^2 facing north. A visual representation of the dwelling is depicted in Figure 4.

Figure 4: Representation of the modeled house

Heating and cooling setpoints are maintained at 20 °C and 28 °C respectively. Internal gains and heating/cooling schedules adhere to the normative French energy calculation method [47]. Ventilation is provided at 0.3 volumes per hour by a heat recovery ventilator with an efficiency of 89%. Weather profiles correspond to those of Belfort (France) and are based on French thermal regulations for buildings [47]. During the cooling period (May 6th to September 23rd), strategies such as window shading with shutters during the day and increased ventilation during the night are employed to manage interior temperatures.

The building is simulated over the course of a full year, with a time resolution of one minute. The estimated annual heating consumption is 1561 kWh (equivalent to 13.7 kWh per square meter), while air-conditioning consumption is estimated at 58 kWh (0.51 kWh per square meter).

The results from the simulation are utilized for the identification of parameters in the gray model. Firstly, the global irradiance multiplier is determined using the annual irradiation gain through windows calculated from Pleiades®. The parameter is identified by minimizing the Root Mean Square Error (RMSE). Two values for the global irradiance parameters are estimated: one for the windows occultation period (summer) and one for the period outside it (winter) by comparing the irradiation for the corresponding periods.

Other model parameters are identified using the Matlab® Multistart optimization tool, which enables the fmincon nonlinear solver to be used in parallel for several starting points. The Pleiades® profiles for solar irradiation, outdoor temperature, internal gains (lighting, occupancy, power dissipation), and heating/cooling loads are utilized as inputs to the gray box model. The internal temperature profile for the first 16 weeks of the year is compared through the RMSE.

A comparison between the indoor temperature profiles obtained with the Pleiades® model and the reduced model is presented in Figure 5 (for 51 weeks). This new building type as well as the 6 buildings other types taken from [15] can be used for profile generation (Table 5). It is to be noted that the 6 standard building models are taken from the CREST model and thus do not take into account different windows shutter management in summers.

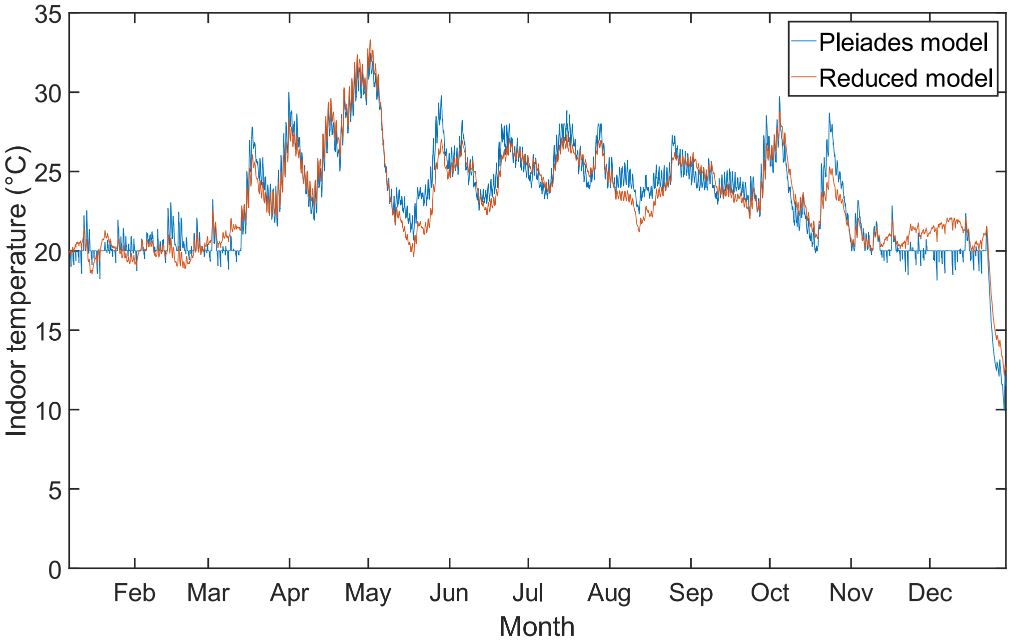


Figure 5: Indoor temperature comparison between the reduce and complete models

Table 5: Building types

Building index	Building type	insulation type	Floor area (m^2)
1	Detached	Standard insulation	136
2	Detached	Improved insulation	136
3	Semi-detached	Standard insulation	87
4	Semi-detached	Improved insulation	87
5	Terraced	Standard insulation	58
6	Terraced	Improved insulation	58
7	Detached	Passive house	114

1: taken from [19]

2.7.2. Space heating and cooling loads

In the CREST Model, the heating and air-conditioning control module determine the power required to maintain comfort based on the building's indoor temperature. Space heating is regulated by two signals: a thermostat signal and a clock signal. The thermostat signal combines two thermostats—one controlling the indoor air temperature (with a 2°C dead band) and the other controlling the emitter temperatures (with a 5°C dead band). Heating setpoint temperatures for indoor air are stochastically drawn from a discrete distribution. The setpoint temperatures for emitters are consistent across all buildings: 50°C for heating emitters and 0°C for cooling emitters. Indoor air thermostats are drawn from a discrete distribution for heating and placed 5 °C above heating for cooling.

When the systems are activated, Equation (12) calculates the heating power required to attain the setpoint, while Equation (13) is utilized for cooling. These calculated ideal power requirements are then adjusted based on the actual capabilities of the heating and cooling systems. Given that the heating system also caters to the demand for hot water, the power supplied will be constrained by the power already allocated for water.

In the present model, a notable change concerns the cooling control strategy. Given that the cooling emitter possesses a smaller thermal capacity compared to the heating counterpart, utilizing the same control approach as in the CREST model leads to operational issues. Specifically, when the cooling power available is sufficiently high, the system cools the emitter within a time frame shorter than the resolution of the simulation, resulting in constantly switch ON and OFF. To prevent this, it is chosen to only use the indoor air thermostat. Another change is that instead of directly using a cooling setpoint 5 °C above the heating setpoint, it is chosen from a rounded normal distribution with an average of 28°C and a standard deviation of 2.5°C. This value is compared with the heating temperature to ensure it is never lower than or equal to the heating setpoint.

One limitation of the CREST thermal model, as highlighted by its authors, is that the profiles generated for the clock signal are not synchronized with the occupancy model, leading to occasional inconsistencies. For example, a house may be heated between 10 a.m. and noon, while the occupants are only present from 8 a.m. to 10 a.m. To mitigate this, in the heating part, the authors generate the clock signal through pattern assignation. The heating clock signal is correlated with the occupancy model in three distinct patterns: heating always on (pattern 1), heating off when no occupants are present (pattern 2), and heating setpoint lowered when no occupant is present (pattern 3). As an example, these patterns represent 17.6%, 27.50%, and 54.90% of French households respectively [48]. During system control's initialization, a random draw determines which pattern the building corresponds to. If temperature reduction during absences is implemented, the setpoint temperature is decreased by 5°C during absences. For cooling, this signal is established using a Markov chain, mirroring the approach of the occupancy model. As an example, the resulting probability of the signal to be activated can be observed in Figure 6 (during both the heating and cooling seasons). Consistent with the white box model, heating is switched off from May 6 to September 23, while air conditioning is not utilized from September 23 to May 6.

As already mention above, the generated thermal demand profile remains generic. To derive specific demands (electricity, gas, etc.), the efficiency of the heating and cooling systems pertinent to the studied systems must be applied to modulate the profiles.

$$P_{heat,air,ideal} = C_{rad}\left(T_{set,rad} + T_{db} - T_{rad}\right) + H_{rad}(T_{rad} - T_i) \tag{12}$$

$$P_{cool,ideal} = C_{cool}\left(T_{set,cool} - T_{cool}\right) + H_{cool}(T_{cool} - T_i) \tag{13}$$

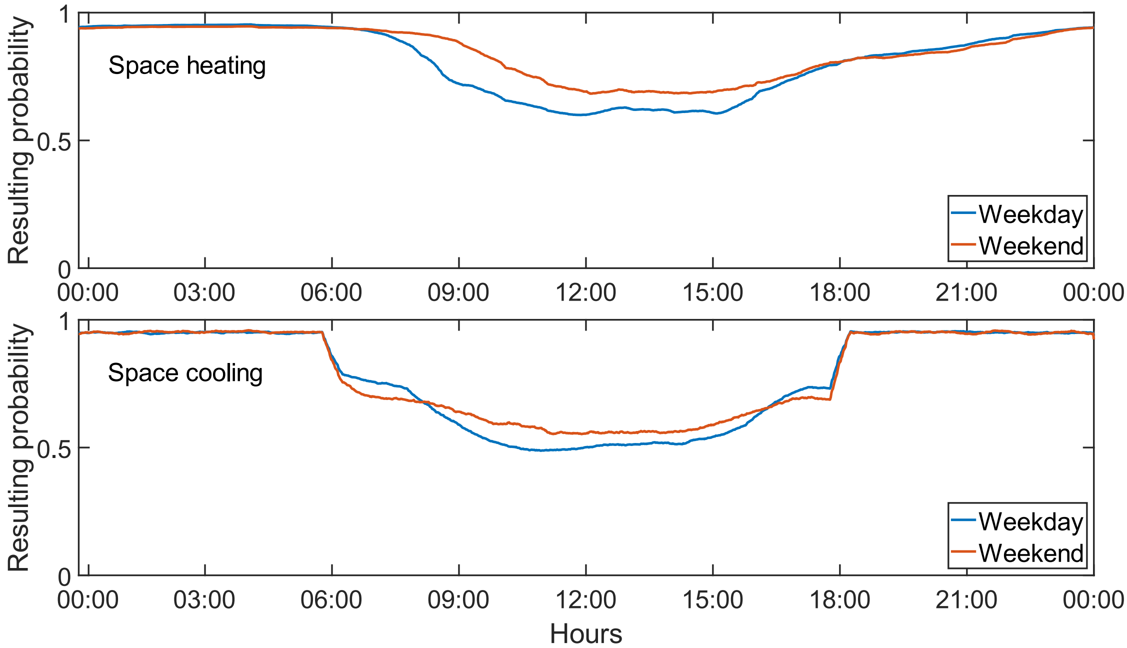
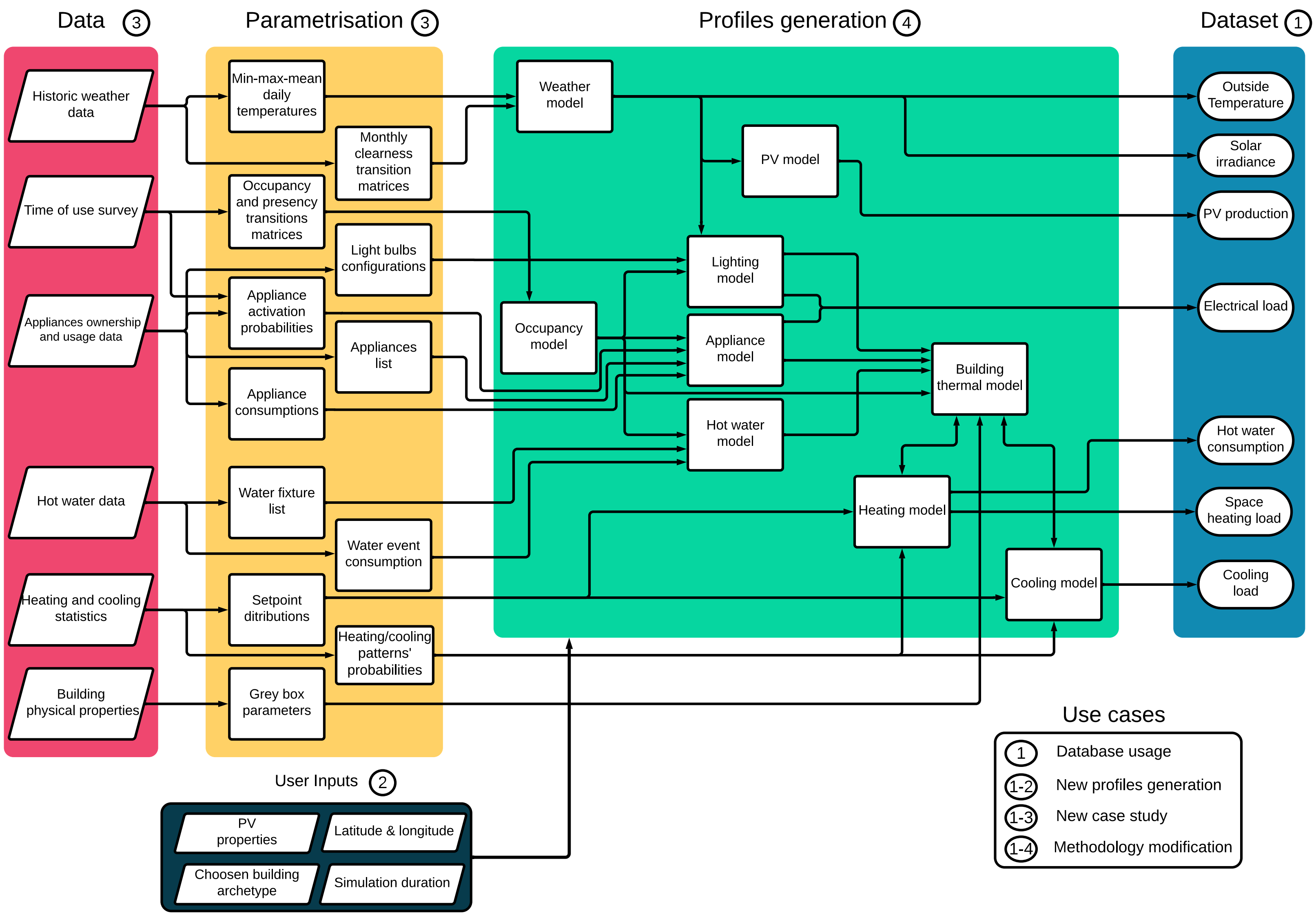


Figure 6: Examples of probabilities of the clock signal being one through the day (during heating/cooling season)

The complete behavior of the model is summarized in the Figure 7. The parametrization procedure as well as the necessary data to undertake this procedure are also described in the figure. Depending on the use case, different parts of the methodology are required. A researcher that only wants to assess the potential of an energy system can directly use the open dataset. For the generation of new profiles, the user also needs to modify the inputs of the openly shared Matlab® algorithm. If the case study needs to be changed, the whole parameterization process needs to be redone. Finally, to improve the methodology every part of the procedure needs to be studied.



626

627 Figure 7: Flowchart of the proposed methodology

3. Test case, results and discussion

Using the model parametrized for Belfort (France), 3500 annual profiles, with a 1-minute time step, are generated (500 for each type of house) using a Monte Carlo simulation. The data generated includes, weather profile: temperature and solar radiation, load profile: electricity, space heating, hot water and cooling as well as PV production profiles. Important house properties are also saved for profile identification (inhabitant number, type of house, appliance ownership…).

3.1. Overall trend

For each module of the model profiles annuals average results are assessed. In the weather model, with the new clearness matrix for Belfort (France), the solar radiation model yields an average of 1211 kWh.m^{-2}.year^{-1}, closely resembling the average of 1199 kWh.m^{-2}.year^{-1} for the weather station used as input (from 2018 to 2020). This average can also be compared with Open-Meteo data, which reports an average of 1244 kWh.m^{-2}.year^{-1} for Belfort from 2000 to 2022 [30, 31, 32, 33]. For temperature the synthetic profiles exhibit an annual average of 10.48°C against 10.51 C for Open-Meteo data from 2000 to 2022 [30, 31, 32, 33]. For both radiation and outside temperature, the models is able to reproduce the measured weather data.

The PV production profiles present an annual PV production of 223 kWh.m^{-2}. This value can be compared to the value of 211 kWh.m^{-2} obtain for a similar panel position with PVGIS [49]. In both cases only losses in the power invertor are taken into account to produce generic profiles, which can then be adjusted to specific profiles by adding specific losses as shading, dirt, snow, mismatch, wiring, etc. Having these PV productions profiles coupled with the electricity and heating demand allow to accurately assess the temporal matching and mismatching between production and demand, which is not possible in methods without integrated PV modeling (OpenIDEAS, synconn_build, TEASER etc.).

Intermediary variables such as inhabitant behavior and disaggregated electric profiles are tested for only 700 profiles, 100 for each house type. The occupancy data are compared using the number of inhabitant present and active, which is crucial for determining appliance usage and internal thermal gain. The number of active and present inhabitants in all houses is calculated for every 10-minute interval throughout the day and then averaged across all houses. The profiles obtain are compared with the TUS data using the root-mean square error. Results are summarized in Table 6. A RMSE of 0.5 mean on average a 0.5 inhabitant active and present difference between TUS and synthetic data. Additionally, Figure 8 provides a direct comparison of the frequency of behavior states for an individual living alone, comparing TUS data with synthetic profiles (for the 266 houses among the 700 with 1 inhabitant). In both state and active presence comparison the model closely follows the behavior of the TUS data.

Table 6: Root mean square error in average active presence daily profile between the synthetic and TUS data (i.e., RMSE 0.5 mean on average 0.5 inhabitant active and present difference between)

Inhabitant number	RMSE Weekdays	RMSE Weekend
1	0.0190	0.0454
2	0.0211	0.1556
3	0.0220	0.3096
4	0.1242	0.5447
5	0.3107	0.6691

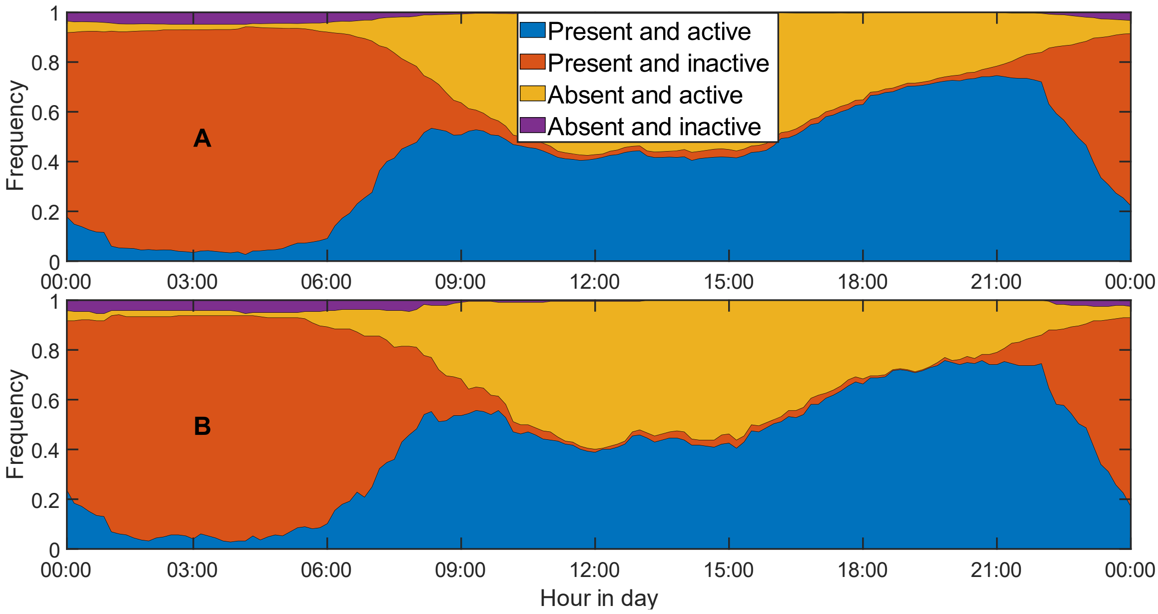


Figure 8: Probability of the state of a single resident during the week, a) input data b) synthetic data

In the ADEME measurements, the average domestic electricity consumption, excluding heating and cooling loads, is reported at 2183 kWh.year^{-1} [39]. The modeled profiles present a similar value, totaling 2241 kWh.year^{-1}. Within this consumption in the ADEME data, lighting accounts for an annual consumption of 148.6 kWh.year^{-1} [39], whereas the modeled result for the 700 dwellings yields 182.6 kWh.year^{-1}. However, this value is influenced by the floor area of the house. Down to the square meters the profiles exhibit a value of 1.9 kWh.m^{-2}.year^{-1} that is close to the ADEME data measure of 1.7 kWh.m^{-2}.year^{-1} [39].

The appliances consumption for the 700 configurations presents an average annual load of 2074 kWh.year^{-1}. Comparatively, the ADEME average in a sample of 101 homes is 2163 kWh. A disaggregated comparison by end use is provided in Table 7. The overestimation in cooking consumption is attributed to the decision to replace gas appliances with their electric counterparts, aligning with the prevailing trend towards reduced natural gas usage. Similarly, the overestimation in hygiene/self-care consumption stems from differences in ownership rates between the larger Gifam study [41] and the ADEME study [39]. To conform to French building regulations [47], ventilation ownership is set at 100%, contributing to potential overestimation in this category. Conversely, the outdoor and electric mobility categories are underestimated due to the exclusion of appliances such as in-ground swimming pools and electric cars, which have high annual consumption but low ownership rates. Overall, certain sectors with lower consumption are not modeled, as they encompass very broad range of electrical appliances.

Table 7: Annual appliances consumption by end use

End-use	**ADEME study of 101 homes [39] (kWh.year^{-1})**	**Average over 700 profiles (kWh.year^{-1})**
Cold production	535	540
Audiovisual	328	380
Washing and drying	308	315
Cooking	299	360
Informatics	191	166
Not monitored	189	0
Ventilation	114	241
Outdoor	89	0
Various constant	49	48
Electric mobility	28	0
Hobbies	16	0
Safety	6	0
Hygiene/self-care	6	24
Health	2	0
Other	3	0
Total	**2163**	**2074**

For hot water profiles, an average annual thermal demand of 1795 kWh.year^{-1} is calculated. This annual demand closely aligns with the measured annual demand of 1676 kWh.year^{-1} reported by ADEME for 57 Joule heating tanks [39]. Seasonal variation induced by variation in cold water temperature is tested using seasonality coefficient (daily consumption divided by average annual consumption). In Figure 9, obtained seasonal coefficients are compared with the ADEME data as well as the data obtain using constant temperature (with 100 runs). The seasonal variation is better assessed in the model that with the CREST method. However, the decrease in hot water consumption during summer is still underestimated. Likely because the change in behavior is not considered (colder shower, less interest in hot water).

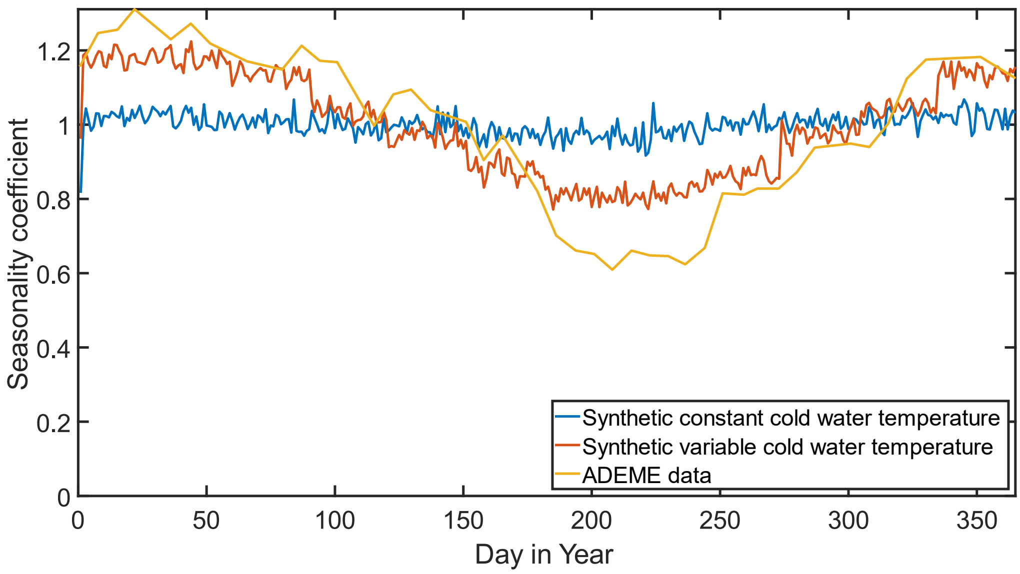


Figure 9: Seasonal variation of hot water consumption

On the space heating and cooling side, the annual load significantly depends on the type of insulation use and the house size. The annual values for each house type are summarized in Table 8. With these average space heating and cooling annual consumption, each building type can be associated with Energy Performance Certificates using French labeling rules [47]. However, the energy consumption associated with those thermal needs depend on the type of energy systems used to answer them. Table 9 summarizes the different houses label for 5 combinations of space heating and DHW systems, cooling is considered always provided by heat pumps. Label varies from A to E and no house type corresponds to label F or G. Only the passive house type manages to get an A label when using efficient heating systems. Because of the conversion factor of 2.3 between final and primary energy for electricity [47], the worst labels are when using inefficient electric systems (radiator and joule effect DHW tank). Even if gas boilers are penalized through the carbon emission criteria this penalty does not compensate for the increase in primary energy consumption.

Only the passive house type (index 7) originates from the present study, and thus, detailed white box modeling is available only for this case. As previously stated, the white box Pleiades® model give a heating load of 13.7 $kWh.m^{-2}.year^{-1}$ and a cooling load of 0.51 $kWh.m^{-2}.year^{-1}$, whereas the average annual consumption over 500 profiles is 20.3 $kWh.m^{-2}.year^{-1}$ of heating and 0.18 $kWh.m^{-2}.year^{-1}$ for cooling. It can be observed that those consumptions are quite different and do not respect the passive house heating norm limit (15 $kWh.m^{-2}.year^{-1}$ [46]). However, when subjected to the same conditions as the Pleiades® model (including weather, internal gain, temperature setpoints), the model predicts an annual heating consumption of 13.4 $kWh.m^{-2}.year^{-1}$ and a cooling consumption of 1.3 $kWh.m^{-2}.year^{-1}$ (averaged over 100 homes with similar climates). This comparison demonstrates that under standard conditions, the houses comply with the passive house norm and replicate the heating consumption of the white box models. The increase in load is therefore attributed to the coupling with the other models presented above, which provide a more complex estimation of the demands instead of constant scheduled values.

Table 8: Average annual heating and cooling consumption

Building index	Heating load ($kWh.m^{-2}.year^{-1}$)	Cooling load ($kWh.m^{-2}.year^{-1}$)
1	119	0.55
2	60	3.34
3	114	3.38
4	72	10.39
5	98	3.63
6	70	7.81
7	20	0.18

Table 9: Energy Performance Certificates for each building type

Building index	Heat pump for space heating and DHW	Heat pump for space heating and Joule heating for DHW	Heat pump for DHW and electric radiators for space heating	Electric radiator for space heating and Joule heating for DHW	Gas boiler for space heating and DHW
1	C	C	E	E	D
2	B	B	C	D	C
3	C	C	E	E	D
4	B	C	D	D	C
5	C	D	E	E	D
6	C	C	D	E	D
7	A	A	B	B	C

3.2. Dispersion and Diversity

In each module of the domestic load model, various factors are incorporated to generate diversity in the profiles. The sources of variation outlined in the previous sections are summarized in Table 10. These models introduce diversity between houses (inhabitant number, building properties, setpoints, heating schedules, ownership…) and temporal diversity (sky clearness, exterior temperature, switch on, cycle length…). The diversity between houses is used as the foundation of the Monte Carlo simulation, distinguishing configurations from one another. This variation between profiles allows for the generation of a diverse dataset, usable for off-design analysis, which is not possible with tools that can only introduce variation through inputs and building properties (CityBES, TEASER, CitySim, etc.).

Table 10: Profiles variating parameters

Parameter	**Model**	**Comments**
Sky clearness	Markov chain	
Mean exterior temperature	Autoregressive moving average	
Max variation around mean	Normal distribution	Based on cumulative irradiation
Inhabitant number	Discrete probability distribution	1 to 5 (based on [50])
Inhabitant behavior	Markov chain	Occupancy and activity
Light bulbs ownership	Random configuration selection	From 100 configurations
Bulbs switch ON	Single probability check per bulbs	Time in dependent
Bulbs ON cycle length	Double probability draw	9 equally likely ranges
Appliances’ ownership	Single probability check per appliance	
Appliances’ switch on	Single probability check per appliance	Time in day dependent
Appliances’ cycle length	Normal distribution	
Water fixture ownership	Single probability check per fixture	
Fixtures switch ON	Single probability check per fixture	Time in day dependent
Fixtures consumption	Discrete probability distribution	Poisson law
Hot water setpoint	Discrete probability distribution	Between 42 °C and 62 °C
Cold water temperature	Normal distribution	Month dependent
Space heating setpoint	Discrete probability distribution	Between 13 °C and 27 °C
Space heating schedule	Pattern assignation	
Space cooling setpoint	Normal distribution	Superior to the heating one
Space cooling schedule	Markov chain	
Building properties	Archetype assignation	Among 7 models

3.2.1. Weather model dispersion

In the weather model, variation in the daily mean Global Horizontal Radiation (GHI) is illustrated in Figure 10 through 365 box diagrams. It can be observed when compared with the Open Meteo GHI data for Belfort from 2000 to 2022 [30, 31, 32, 33], that the synthetic variation in the profile is closely matching the measured one. Similarly, the variation in mean daily temperatures can be observed in Figure 11. In this case also, the mean daily values reproduce the typical temperature profile as well as the extreme values. The model is thus well able to reproduce the variation in climate conditions for a specific region. Moreover, no abrupt steps between month can be observed in the temperature profiles contrary to the CREST method.

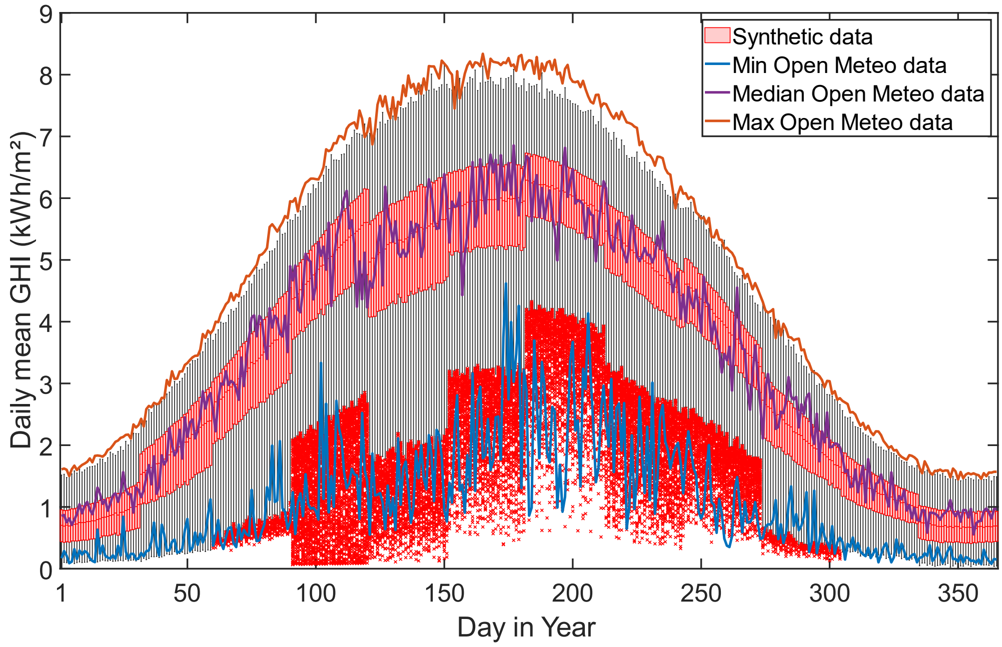


Figure 10: Dispersion of the daily mean GHI (synthetic data represented as box graph with median quartiles, and outliner)

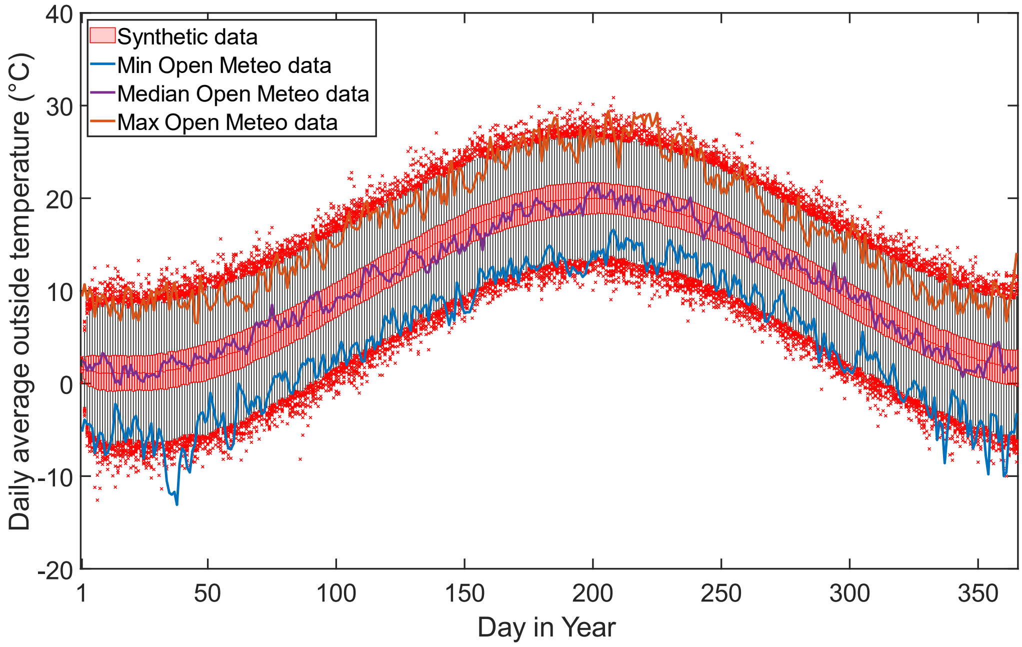


Figure 11: Dispersion of the daily mean temperature (synthetic data represented as box graph with median quartiles, and outliner)

3.2.2. Dispersion of photovoltaics production

The dispersion of PV monthly PV production is compared, in Figure 12, with PVGIS as well as profiles obtain with CREST method (over 3500 runs). Overall, the dispersion due-to-year to year variation estimated by the PVGIS tool is well reproduced in the profiles. A clear advantage of using 12 monthly matrices for clearness modeling is apparent, especially for the winter months where the production is heavily overestimated with the CREST method.

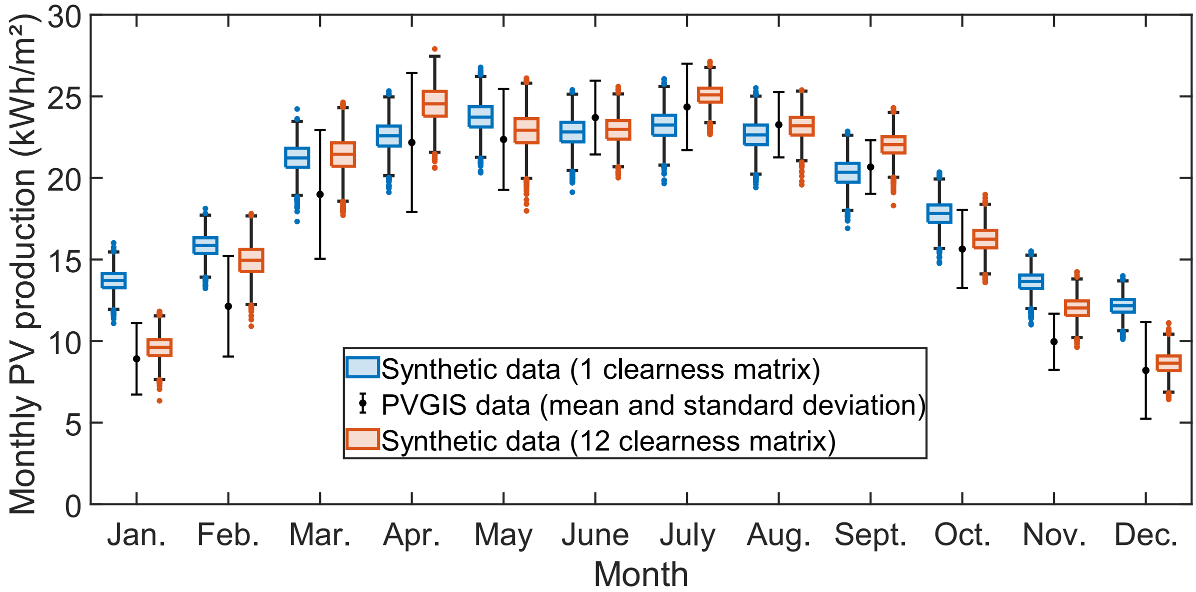


Figure 12: Dispersion of PV production profiles compared with PVGIS

3.2.3. Dispersion of electricity consumption

The electricity consumption profiles can be distinguished by inhabitant number, appliance/bulbs ownership and floor area. Ownership generates variation through the annual reference values listed in Table 3 and floor area through lighting reference consumption 1.8 $kWh.m^{-2}$. Inhabitant number introduces variation through the frequency of appliance and lighting usage. However, this variation is more challenging to directly infer from the input data. The dispersion of the electricity consumption depending on the inhabitant number is shown in Figure 13. It can be observed that the consumption stabilizes after 3 inhabitants because of appliances sharing. The dispersion is compared with the ADEME study used as input (101 dwellings) and corresponding dispersion is observed [39]. This comparison helps validate the model's ability to replicate real-world variations in electricity consumption based on the number of inhabitants.

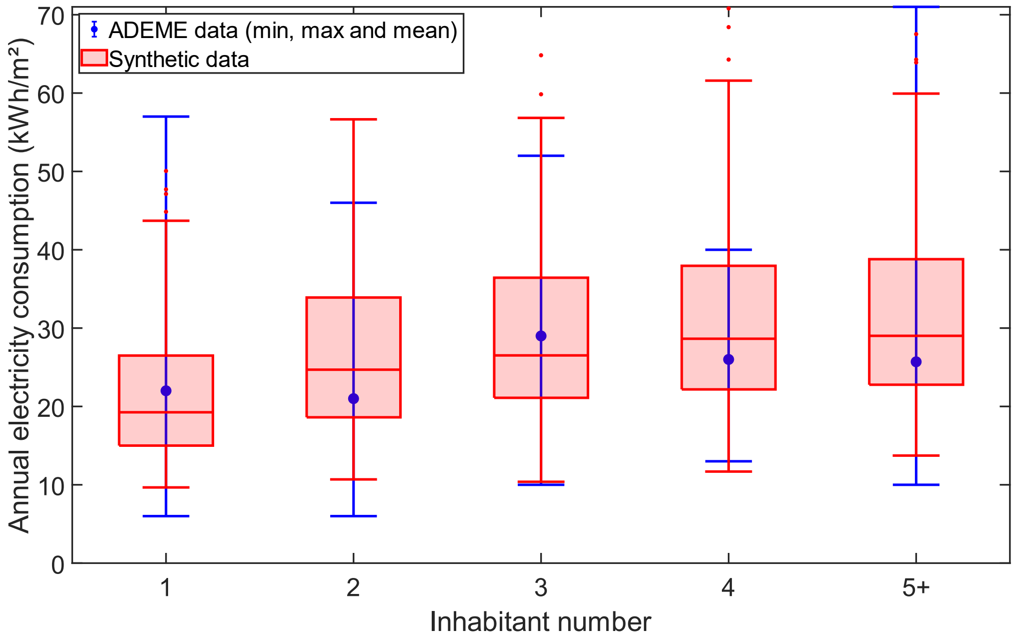


Figure 13: Dispersion of the electricity consumption depending on the inhabitant number (ADEME data use mean and synthetic use median)

3.2.4. Dispersion of thermal consumption

The Figure 14 compare the generated hot water thermal consumption profiles with the measurements from ADEME (for Joule effect water tank) [39]. The dispersion around the median is well represented, but extreme values are not fully captured. One reason for this discrepancy is that in the ADEME measurements, some dwellings were not consuming hot water, and the thermal demand only represented static losses, which is not reflected in the synthetic data, where every house uses its DHW tank. Additionally, in the present study, the volume of the DHW tank is fixed at 125 L, whereas in the measurements, volumes vary from 15 L to 300 L.

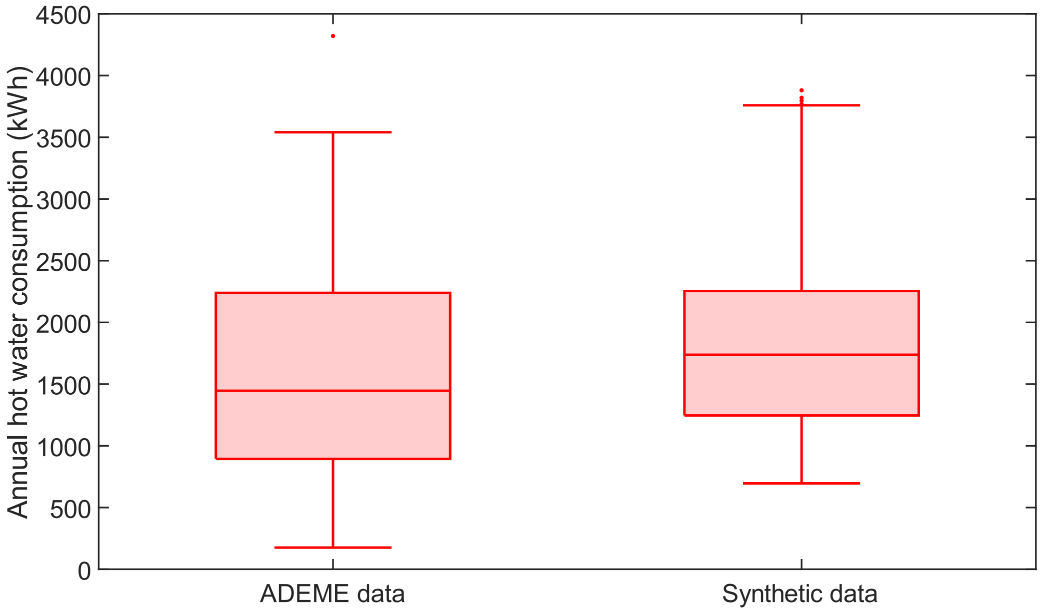


Figure 14: Dispersion of the hot water consumption profiles

The space heating and cooling consumption profiles are mainly differentiated by temperature setpoints, insulation types and heating patterns. Regarding heating patterns, the mean annual space heating consumption is 84.5 kWh.m^{-2}.year^{-1} for houses heated continuously, 77.1 kWh.m^{-2}.year^{-1} for those with heating lowered during absences and 79.6 kWh.m^{-2}.year^{-1}, for those with heating switch off during absences. Temperature setpoints also significantly impact consumption with the mean annual heating consumption varying from 34.2 kWh.m^{-2}.year^{-1} for houses with setpoint below 15 C to 112.2 kWh.m^{-2}.year^{-1} for setpoints above 25°C. Conversely, annual cooling consumption varies from 20.3 kWh.m^{-2}.year^{-1} for houses with setpoint bellow 23°C to 0.53 kWh.m^{-2}.year^{-1} for setpoints above 33 °C.

The impact of insulation type is illustrated in Figure 15. Single family houses with annual heating consumption up to 183 kWh.m^{-2}.year^{-1} are represented in the set. Introducing a building archetype with lower thermal insulation could complement the dataset to represent the worst-performing houses in the housing stock. The passive house archetype can vary from not needing heating at all to having needs similar to less insulated archetypes, depending on heating management (setpoints, schedule, etc.), underscoring the importance of considering variations in heating management for energy system sizing. Given the relatively cool climate of Belfort, most houses have low or non-existent cooling consumption. Moreover, houses archetype where cooling demand is the most important (improved insulation), are archetype from the CREST model and thus do not use windows shutter usage scenarios to mitigate interior temperature in summer. These cooling differences, highlights again the importance of accessing variations in inhabitant interior temperature management. In most cases, the cooling needs are low enough that inhabitants would likely not invest in a cooling system.

To create an extensive dataset, the authors produced an equal number of each house archetypes in the profiles (500 per archetype). However, in reality, some house thermal archetypes are more prevalent in the housing stock than others, this complicates direct comparisons between the dataset and national space heating and cooling consumption. However, specific targeted annual consumption profiles can be selected from the dataset to conform to desired criteria. For example, a benchmark for energy system assessment can be created by eighter directly using the model for a specific set of building archetypes or by selecting the right profiles from the openly available pre-made dataset to represent the studied population. For example, in France the repartition of Energy Performance Certificates for houses is 8% of A, 4% of B, 19% of C, 28% of D and 41% of E or worst [51], using the Table 9 the correct number of each building index can be picked to reproduce this repartition. Similarly, a district load can be generated using either the model or the dataset. However, in the dataset, the weather for each house is distinct, whereas in the case of a district, each house is subjected to similar conditions. This type of configuration, can only be achieved with the model by running the weather module once and applying the same climate to every house in the district.

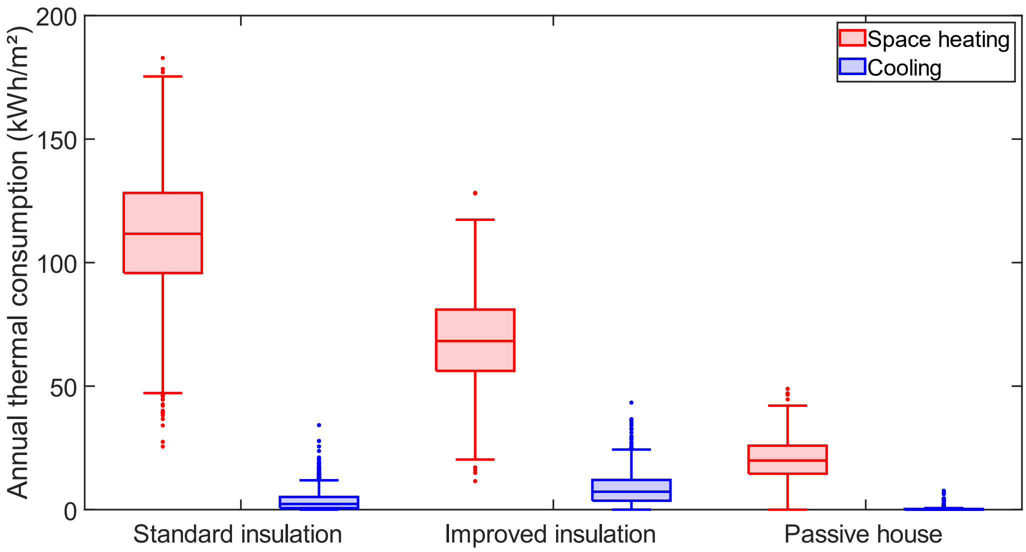


Figure 15: Dispersion of the thermal consumption depending on the insulation type

4. Conclusion

The objective of this study is to develop a comprehensive method to generate yearly energy domestic load profiles usable in energy systems sizing and performance assessment. For this type of use, profiles need to be varied for off design analysis, consistent to capture load/production mismatch, high-resolution for system dynamics analysis and continuous for storage potential assessment. Open dataset profile also allows for researcher without specific knowledge of building energy modeling to conduct accurate domestic energy systems study, enhancing the usefulness of the profiles.

To achieve this goal, starting from a commonly used open-source model each module was presented, as well as all the modification and amelioration necessaries to the generation of profiles. Among other modifications, the weather module was improved to append monthly cloud cover patterns, temperature effects on PV production were added, seasonal variation of water inlet was introduced, coherence between occupancy and heating schedule was established, and seasonal variation in shutter usage was made possible.

Additionally, in each module, the parametrization procedure was extensively presented, using the use case of a region with a temperate oceanic climate, to facilitate the reproduction of the approach in specific geographic regions. Using the model in a Monte Carlo simulation, a dataset was generated comprising electricity, hot water, space heating, and cooling load profiles, as well as local PV production profiles for 3500 single-family houses in a temperate oceanic climate. The profiles in the generated dataset, vary in weather (solar radiation, temperature), inhabitant behavior (presence, appliances usage/ownership, interior temperature management…) and building properties (floor area, thermal properties…).

The generated weather profiles show a relative difference of approximately 0.3% and 2.7% for mean annual temperature and annual solar irradiation, respectively, compared to data from Open-meteo. The models also generate sufficient weather variation to replicate the diversity observed in the last 23 years of Open-meteo data. Regarding PV production profiles, the mean annual production exhibits a relative difference of 5.7% compared to PVGIS data. The dataset also reproduces the monthly variation and diversity estimated by PVGIS. In terms of electricity profiles, the mean annual demand shows a relative difference of 2.7% compared to French ADEME agency measurements. Due to the absence of houses without hot water consumption, the associated thermal consumption is overestimated by a relative error of 7.1% compared to ADEME measurements. The presented archetypes allow for the generation of houses with annual heating consumption ranging from 0 to 25 $MWh.year^{-1}$, covering Energy Performance Certificates from A to E. Analyzing the results shows that the generated profiles reproduce the diversity present in measurement studies while retaining the underlying trends, highlighting the potential of both the generation method and the openly available dataset for accurate domestic energy system assessment.

Suggestions for further study on the subject include the following improvement of the methodology:

- Enhancing the PV model to incorporate specific losses such as shading and dirt
- Improving the precision of the thermal model by increasing the RC order.
- Delving deeper into inhabitant interior temperature management (window opening, shutter usage, etc.).

Other suggestions for further study on the subject include the following expansion of the methodology:

- Applying the methodology in other climates, particularly hotter climates, to enable a more precise study of house cooling behaviors
- Broadening the scope of considered appliances to include electric vehicles, small digital appliances, etc
- Introducing a new archetype for a poorly insulated house to expand the range of heating consumption that can be modeled
- Expand the variety of building types (multi-family houses, apartment buildings, office buildings, etc.)

Declaration of competing interest

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

Acknowledgments

This work has been supported by the EIPHI Graduate School (contract ANR-17-EURE-0002) and the Region Bourgogne Franche-Comté.

Computations have been performed on the supercomputer facilities of the Mésocentre de calcul de Franche-Comté

5. Appendix A. Passive house characteristics

Table A1: Exterior wall composition

Element	Thickness (cm)	Conductivity ($W.m^{-1}.K^{-1}$)	Resistance ($m^2.K.W^{-1}$)
Wood fiber panel	10	0.04	2.5
OSB panel	1.6	0.09	0,18
Adjusted cellulose	14	0.04	3.41
Vapor barrier	0.1	0	0
Glass wool	8	0.04	2.29
Drywall	1.3	0.25	0.05
Total	35		8.43

Table A2: Ground floor composition

Element	Thickness (cm)	Conductivity ($W.m^{-1}.K^{-1}$)	Resistance ($m^2.K.W^{-1}$)
Reinforced concrete	13	2.5	0.05
Vapour barrier	0.1	0	0
Polyurethane	20	0.02	8.7
Reinforced concrete	5	2.5	0.02
Floor covering (wood)	2	0.15	0.13
Total	40.1		8.9

Table A3: Roof composition

Element	Thickness (cm)	Conductivity ($W.m^{-1}.K^{-1}$)	Resistance ($m^2.K.W^{-1}$)
Wood fiber panel	10	0.05	2.13
Adjusted cellulose	36	0.04	9
Vapor barrier	0.1	0	0
OSB panel	1.6	0.09	0.18
Drywall	1.3	0.25	0.05
Total	49		11.36

Table A4: Bearing inner wall composition

Element	**Thickness (cm)**	**Conductivity ($W.m^{-1}.K^{-1}$)**	**Resistance ($m^2.K.W^{-1}$)**
Drywall	1.3	0.25	0.05
Air gap	4	0.19	0.21
Glass wool	6	0.04	1.5
OSB panel	1.6	0.09	0.18
Drywall	1.3	0.25	0.05
Total	14.2		1.99

Table A5: inner wall composition

Element	**Thickness (cm)**	**Conductivity ($W.m^{-1}.K^{-1}$)**	**Resistance ($m^2.K.W^{-1}$)**
Drywall	1.3	0.33	0.04
Air gap	1.5	0.19	0.16
Drywall	1.3	0.33	0.04
Total	4.1		0.24

Table A6: Intermediate floor composition

Element	**Thickness (cm)**	**Conductivity ($W.m^{-1}.K^{-1}$)**	**Resistance ($m^2.K.W^{-1}$)**
OSB panel	1.6	0.13	0.12
Glass wool	10	0.04	2.5
Drywall	1.3	0.25	0.05
Total	12.9		2.67

Table A7: Window inventory

Surface (m^2)	Orientation	Solar factor	U value ($W.m^{-2}K^{-1}$)	Quantity
0.795	North	0.508	0.864	2
0.795	East	0.508	0.864	1
1.575	East	0.635	0.804	1
1.89	South	0.636	0.805	2
6.3	South	0.808	0.719	1
4.24	West	0.774	0.736	1
1.545	West	0.635	0.804	1
1.26	West	0.593	0.824	1

Table A8: Thermal bridge inventory

Name	**Linear thermal transmittance ($W.m^{-1}.K^{-1}$)**	**Length concerned (m)**
Exterior wall - roof (parallel)	0.024	19.6
Exterior wall - roof (perpendicular)	0.017	16
Ground floor - exterior wall	0.024	35.6
Ground floor - interior wall	0.05	9.9
Interior floor - exterior wall	0.04	18.6
Bearing interior wall - exterior wall	0.033	5
Outer angle	0.032	15.5
Roof angle	0.018	9.8

6. Appendix B. Supplementary data

The following are supplementary data to this article. The method presented as a MATLAB® algorithm is available at hal.science/hal-04574032. A dataset of consumption profiles related to electricity, heating, hot water and air conditioning, as well as photovoltaic production profiles for 3500 single family generated with this method can be found at dx.doi.org/doi:10.25666/DATAUBFC-2024-05-03

7. References

[1] International Energy Agency (IEA), "Tracking Clean Energy Progress 2023 - Buildings," Paris, 2023.

[2] X. Jin, C. Zhang, F. Xiao, A. Li and C. Miller, "A review and reflection on open datasets of city-level building energy use and their applications," *Energy and Buildings,* vol. 285, p. 112911, 2023.

[3] A. Grandjean, J. Adnot and G. Binet, "A review and an analysis of the residential electric load curve models," *Renewable and Sustainable Energy Reviews,* vol. 16, p. 6539–6565, 2012.

[4] T. Guo, M. Bachmann, M. Kersten and M. Kriegel, "A combined workflow to generate citywide building energy demand profiles from low-level datasets," *Sustainable Cities and Society,* vol. 96, p. 104694, 2023.

[5] Y. Chen, M. Guo, Z. Chen, Z. Chen and Y. Ji, "Physical energy and data-driven models in building energy prediction: A review," *Energy Reports,* vol. 8, pp. 2656-2671, 2022.

[6] Z. Wang, T. Hong and M. A. Piette, "Building thermal load prediction through shallow machine learning and deep learning," *Applied Energy,* vol. 263, p. 114683, 2020.

[7] T. Hong, D. Macumber, H. Li, K. Fleming and Z. Wang, "Generation and representation of synthetic smart meter data," *Building Simulation,* vol. 13, p. 1205–1220, 2020.

[8] G. Chaudhary, H. Johra, L. Georges and B. Austbø, "Synconn_build: A python based synthetic dataset generator for testing and validating control-oriented neural networks for building dynamics prediction," *MethodsX,* vol. 11, p. 102464, 2023.

[9] M. Ferrando, F. Causone, T. Hong and Y. Chen, "Urban building energy modeling (UBEM) tools: A state-of-the-art review of bottom-up physics-based approaches," *Sustainable Cities and Society,* vol. 62, p. 102408, 2020.

[10] M. H. Shamsi, U. Ali and J. O'Donnell, "A generalization approach for reduced order modelling of commercial buildings," *Journal of Building Performance Simulation,* vol. 12, p. 729–744, November 2019.

[11] J. Roth, A. Martin, C. Miller and R. K. Jain, "SynCity: Using open data to create a synthetic city of hourly building energy estimates by integrating data-driven and physics-based methods," *Applied Energy,* vol. 280, p. 115981, 2020.

[12] U. Ali, S. Bano, M. H. Shamsi, D. Sood, C. Hoare, W. Zuo, N. Hewitt and J. O'Donnell, "Urban residential building stock synthetic datasets for building energy performance analysis," *Data in Brief,* vol. 53, p. 110241, 2024.

[13] P. Murray, J. Marquant, M. Niffeler, G. Mavromatidis and K. Orehounig, "Optimal transformation strategies for buildings, neighbourhoods and districts to reach CO2 emission reduction targets," *Energy and Buildings,* vol. 207, p. 109569, 2020.

[14] J. Iturralde, L. Alonso, A. Carrera, J. Salom, M. Battaglia and D. Carbonell, "Energy demands for multi-family buildings in different climatic zones D1.1," 2019.

[15] J. Barton, M. Thomson, P. Sandwell and A. Mellor, "A Domestic Demand Model for India," in *Advances in Energy Research, Vol. 1*, Singapore, 2020.

[16] I. Richardson, M. Thomson and D. Infield, "A high-resolution domestic building occupancy model for energy demand simulations," *Energy and Buildings,* vol. 40, pp. 1560-1566, 2008.

[17] I. Richardson, M. Thomson, D. Infield and A. Delahunty, "Domestic lighting: A high-resolution energy demand model," *Energy and Buildings,* vol. 41, pp. 781-789, 2009.

[18] I. Richardson, M. Thomson, D. Infield and C. Clifford, "Domestic electricity use: A high-resolution energy demand model," *Energy and Buildings,* vol. 42, p. 1878–1887, 2010.

[19] E. McKenna and M. Thomson, "High-resolution stochastic integrated thermal–electrical domestic demand model," *Applied Energy,* vol. 165, pp. 445-461, 2016.

[20] M. C. Peel, B. L. Finlayson and T. A. McMahon, "Updated world map of the Köppen-Geiger climate classification," *Hydrology and Earth System Sciences,* vol. 11, p. 1633–1644, 2007.

[21] C. Reinhart, T. Dogan, J. Jakubiec, T. Rakha and A. Sang, "Umi – An Urban Simulation Environment For Building Energy Use, Daylighting And Walkability," 2013.

[22] M. Baratieri, V. Corrado, A. Gasparella, F. Patuzzi, E. Walter and J. Kämpf, "A verification of CitySim results using the BESTEST and monitored consumption values," *Proceedings of the 2nd Building Simulation Applications Conference,* January 2015.

[23] J.-L. Scartezzini, R. Nouvel, K.-H. BRASSEL, M. BRUSE, E. Duminil, V. Coors, U. Eicker and D. Robinson, "SimStadt, a new workflow-driven urban energy simulation platform for CityGML city models," January 2015.

[24] R. Baetens, R. De Coninck, F. Jorissen, D. Picard, L. Helsen and D. Saelens, "OpenIDEAS – An Open Framework for integrated District Energy Simulations," 2015.

[25] T. Hong, Y. Chen, S. H. Lee and M. Piette, "CityBES: A Web-based Platform to Support City-Scale Building Energy Efficiency," 2016.

[26] J. A. Fonseca, T.-A. Nguyen, A. Schlueter and F. Marechal, "City Energy Analyst (CEA): Integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts," *Energy and Buildings,* vol. 113, pp. 202-226, 2016.

[27] M. M. M. F. T. O. Peter Remmen and D. Müller, "TEASER: an open tool for urban energy modelling of building stocks," *Journal of Building Performance Simulation,* vol. 11, p. 84–98, 2018.

[28] R. E. Kontar, B. Polly, T. Charan, K. Fleming, N. Moore, N. Long and D. Goldwasser, "URBANopt: An Open-Source Software Development Kit for Community and Urban District Energy Modeling," in *2020 Building Performance Analysis Conference and SimBuild co-organized by ASHRAE and IBPSA-USA*, Virtual, 2020.

[29] I. Richardson and M. Thomson, "Integrated simulation of photovoltaic micro-generation and domestic electricity demand: a one-minute resolution open-source model," *proceedings of the institution of Mechanical Engineers, Part A: Journal of Power and Energy,* vol. 227, p. 73–81, 2013.

[30] P. Zippenfenig, *Open-Meteo.com Weather API,* 2023.

[31] H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee and J.-N. Thépaut, *ERA5 hourly data on single levels from 1940 to present,* ECMWF, 2023.

[32] J. Muñoz Sabater, *ERA5-Land hourly data from 2001 to present,* ECMWF, 2019.

[33] S. Schimanke, M. Ridal, P. Le Moigne, L. Berggren, P. Undén, R. Randriamampianina, U. Andrea, E. Bazile, A. Bertelsen, P. Brousseau, P. Dahlgren, L. Edvinsson, A. El Said, M. Glinton, S. Hopsch, L. Isaksson, R. Mladek, E. Olsson, A. Verrelle and Z. Q. Wang, *CERRA sub-daily regional reanalysis data for Europe on single levels from 1984 to present,* ECMWF, 2021.

[34] A. Arsalis, A. N. Alexandrou and G. E. Georghiou, "Thermoeconomic modeling of a completely autonomous, zero-emission photovoltaic system with hydrogen storage for residential applications," *Renewable Energy,* vol. 126, p. 354–369, 2018.

[35] M. Hosseini, I. Dincer and M. A. Rosen, "Hybrid solar-fuel cell combined heat and power systems for residential applications: Energy and exergy analyses," *Journal of Power Sources,* vol. 221, p. 372–380, 2013.

[36] E. McKenna, M. Krawczynski and M. Thomson, "Four-state domestic building occupancy model for energy demand simulations," *Energy and Buildings,* vol. 96, pp. 30-39, 2015.

[37] Institut national de la statistique et des études économiques, *Enquête Emploi du temps 2009-2010,* 2010.

[38] Office for National Statistics, *United Kingdom Time Use Survey,* 2019.

[39] Agence de l'environnement et de la maîtrise de l'énergie (Ademe), *PANEL USAGES ELECTRODOMESTIQUES —Consommations électrodomestiques françaises basées sur des mesures collectées en continu dans 100 logements,* 2021.

[40] Institut national de la statistique et des études économiques (INSEE), *L'équipement des ménages de 1996 à 2019, Enquêtes EPCV et SRCV,* 2021.

[41] Gifam, le groupement des marques d'appareils pour la maison, *Conférence de presse, 12 février 2021,* 2021.

[42] Agence de l'environnement et de la maîtrise de l'énergie (Ademe), *Guide technique: Les besoin d’eau chaude sanitaire en habitat individuelle et collectifs,* 2016.

[43] Institut national de la statistique et des études économiques, *Taille des ménages, Données annuelles de 1968 à 2020,* 2023.

[44] T. Berthou, P. Stabat, R. Salvazet and D. Marchio, "Development and validation of a gray box model to predict thermal behavior of occupied office buildings," *Energy and Buildings,* vol. 74, p. 91–100, 2014.

[45] IZUBA énergies, *Pleiades version 5.23.4.4,* 2023.

[46] W. Feist, Z. Bastian, W. Ebel, E. Gollwitzer, J. Grove-Smith, O. Kah, B. Kaufmann, B. Krick, R. Pfluger, J. Schnieders and others, "Passive House Planning Package Version 7," *The Energy Balance and Design Tool for Efficent Buildings and Retrofits,,* 2012.

[47] Ministère de la Transition écologique et de la Cohésion des territoires, *RT-RE Bâtiment,* 2022.

[48] F. Belaïd, D. Roubaud and E. Galariotis, "Features of residential energy consumption: Evidence from France using an innovative multilevel modelling approach," *Energy Policy,* vol. 125, pp. 277-285, 2019.

[49] T. Huld, I. Pinedo Pascua and A. Gracia Amillo, "PVGIS 5: Internet tools for the assessment of solar resource and photovoltaic solar systems," 2017.

[50] Institut national de la statistique et des études économiques (INSEE), *Taille des ménages : répartition des résidences principales selon le nombre d'occupants,* 2020.

[51] Agence de l'environnement et de la maîtrise de l'énergie, *L'observatoire DPE-Audit, Diagnostic de Performance Énergétique Audit Énergétique,* 2024.