UBEM's archetypes improvement via data-driven occupant-related schedules randomly distributed and their impact assessment

Martina Ferrando[a]*, Sibilla Ferroni[a], Martina Pelle[a,b], Anita Tatti[a], Silvia Erba[c,d], Xing Shi[e,f], Francesco Causone[a]

[a] Department of Energy, Politecnico di Milano, Milano, Italy
[b] Institute for Renewable Energy, EURAC Research, Bolzano, Italy
[c] Department of Architecture and Urban Studies, Politecnico di Milano, Italy
[d] Department of Ocean Operations and Civil Engineering, NTNU, Ålesund, Norway
[e] Key Laboratory of Ecology and Energy-Saving Study of Dense Habitat, Ministry of Education, Shanghai, China
[f] College of Architecture and Urban Planning, Tongji University, Shanghai, China
*Corresponding author, martina.ferrando@polimi.it

Highlights:

1. In UBEM buildings are usually modelled via archetypes with fixed occupants-schedules
2. Data-driven occupants-related schedules randomly distributed are proposed
3. The impact of these occupants-related schedules on UBEM results is assessed
4. Data-driven schedules are significant if the analysis targets hourly or daily values
5. Data-driven schedules are significant if the analysis targets 5 or fewer buildings

Abstract:

In Urban Building Energy Models (UBEMs), buildings are usually modelled via archetypes describing occupants' behaviour via fixed schedules. This research (i) creates data-driven schedules for electric use and occupancy from smart meter readings randomly distributed in the model to improve residential archetypes, (ii) assesses the impact of these schedules on UBEMs' energy results at different temporal resolutions and spatial scales. The novel assessment procedure exploits integrated heat maps based on coefficients of variation of the root means square error (CVRMSE). The outcomes show that differences in energy needs, with randomized schedules, range based on temporal and spatial aggregation. Yearly, for the entire neighbourhood, heating and cooling energy needs, and electric uses are estimated -2%, +1%, and +18% compared to the base case. The outputs show that, when simulations are focused on the entire district, fixed schedules can be enough to describe energy patterns. However, if the simulation is focused on small groups of buildings (e.g., 5 or fewer), randomising the schedules can create variability in the model in terms of electric use and occupancy among buildings characterized by the same archetype. The followed methodology can be exploited also with larger databases and eventually verified with also other types of data.

Keywords: Urban Building Energy Model, UBEM, Building Archetype, Smart Meter, Clustering, Occupant Behaviour, OB, urban modelling interface (umi)

1. Introduction

Building simulation is increasingly used to quantify and optimize energy use in buildings and cities [1,2]. Especially, for urban applications, novel bottom-up physics-based urban building energy modelling tools (UBEMs) are emerging [3]. This typology of UBEMs allows the modelling of numerous buildings together with a physics-based approach that usually can quantify the building energy use to hourly and even sub-hourly values, being based on multizone dynamic thermal simulation models or reduced-order resistor-capacitor models [4]. At the single-building scale, it is widely recognized that occupants play a fundamental role in the simulated energy outputs [5,6] and several approaches are present in the literature to model different aspects related to occupants, also considering their intrinsic variability and dynamic [7]. Conversely, in UBEMs, buildings are usually modelled via archetypes (standard fully-characterized buildings models) [8] which define the building by setting geometries, systems, constructions, and Occupant Behaviour (OB) mainly via daily profiles (usually called schedules). This characterization approach generates no distinctions among buildings described by the same archetype and often provides unrealistic energy outcomes [9]. As an extension of the single-building energy simulation, the OB is supposed to be relevant also in UBEMs, thus, new approaches to model OB in UBEMs are emerging [9], although this remains still an open topic because of the lack of data regarding occupants, their movements, or actions at the urban scale. In UBEMs, the focus is not only on the energy result of the single building or the hourly values but the output is examined at different spatial scales (e.g., considering more than one building together) and temporal scales (i.e., mainly annual or monthly values), which can bring to different effects of OB on specific energy results. The quantification of the impact of occupants-related inputs in UBEMs energy outputs considering different spatial and temporal scales is, indeed, still missing in the literature, and this study aims to reduce this knowledge gap.

The exploitation of smart meter readings to improve the description of large-scale building models is a topic that has been growing in the last years [10–12]. The installation of smart meters is making available a huge amount of data that, with the use of advanced data analytics, offers new opportunities in the UBEM field including, but not limited to, energy benchmarking [13], customer classification [14], schedules creation [11], calibration [15]. Occupants behave differently, especially in their houses, bringing different building load shape profiles [16] that could change the design and optimization of systems. For example, Spoladore et al. [17] exploited the daily and hourly registrations of the energy demand of buildings, combined with the power demand peak and the geographical location to design district heating networks. In particular, they used the hourly natural gas consumption of the entire town of Genoa (Italy) for a decade and their correlation with the outside air temperature to develop a simulator for district heating networks. El Kontar et al. [18] used hourly energy registrations and a tuned clustering approach to profile energy use for different day types (i.e., workdays and holidays) and various occupants-related aspects (i.e., heating, cooling, lighting, and appliances usage), to generate input schedules for UBEMs. The framework is validated in a residential neighbourhood, demonstrating that the energy load patterns outputs can be more accurate than via simple archetypes. Gianniou et al. [19] combined smart meter data for 14000 households with weather data and building data from a national database to assess the temperature setpoints to improve the housing stock model inputs and achieve a more reliable result. Park et al. [20] exploited the smart meter readings of almost 4000 buildings to set three fundamental load shapes and classify the buildings according to their main load profiles. This study proved that smart meters can be used to discover typical load shapes that are relevant for the characterization of buildings. The classification, using the developed fundamental load shapes, was able to group the buildings better than considering standard information such as the floor area, location, usage type, etc. Rakha et al. [21] used data clustering of measured data to calibrate a UBEM focusing on OB aspects in buildings. Usage daily profiles are generated through clustering to calibrate a case study of a community in Texas reducing the error derived by the comparison between simulated and registered data. Razavi et al. [22] analysed the electric consumption behaviour of more than 5000 households for more than one year to predict the occupancy status of the households in the present and future exploiting a wide array of machine learning algorithms. In particular, the algorithms were trained on a ground truth registered through an occupancy survey on different predictive characteristics of the demand (e.g., mean, standard deviation, etc.). Time use surveys (TUS) data are often used as an alternative to smart meter data to develop OB models [23], especially in single-building energy models. With TUS the time people spend doing various activities is collected [24] (e.g., work, household, social life, travel, etc.). The surveys usually include a household interview, a personal interview, and a weekly diary. Clustering of TUS is often used as a methodology to produce occupancy schedules (like in the works of Buttitta et al. [25] and Mitra et al. [26]). Among the whole literature, several methodology examples are available for United States [23,26,27], United Kingdom [25,28,29], Denmark [30,31] and France [32]. However, applications in the Italian Building Simulation context are not available.

Clustering analysis to derive schedules for both electricity [33–35] and occupancy [36,37] is a widely used methodology in single-building energy modelling. However, their direct application in UBEM is not yet standardized, especially regarding the random distribution of the schedules. For example, in the work of Fonseca et al. [38], in which the focus was the assessment of buildings' consumption by looking at the temporal and spatial scale, the schedules characterizing occupancy and electricity loads are fixed. Moreover, it is well known that occupant behaviour has a higher impact on hourly and daily energy results than on monthly and yearly ones [39], but the study on the different spatial scales has been addressed only in a few publications without direct applications on a case study [40], or on different building types from the residential one [41], or focusing mainly on yearly energy results [42]. This study exploits typical sub-hourly smart meter readings to generate multiple data-driven occupant-related schedules randomly distributed among the same archetype and assesses their impact on the energy outputs of a UBEM simulation using the urban modelling interface (umi) [43], one of the main bottom-up physics-based UBEM tools currently available [3]. The work is based on the use of machine learning techniques to cluster the daily smart meter readings and to exploit the clustering centroid (the multi-dimensional average of a cluster), to create more realistic schedules for UBEMs regarding the occupancy and electric use. The procedure is applied to the case study of the Chiaravalle neighbourhood, located in the South-East area of Milan (Italy), including 49 multi-family residential buildings. The smart meter readings are registered from multi-family buildings in the same area of Milan. This paper aims firstly at defining a procedure to create schedules for electric use and occupancy for UBEMs exploiting a limited number of smart meter readings and, secondly, at understanding which is the impact of occupant-related schedules on the UBEM energy results at different temporal scales (i.e., annual, monthly, weekly, daily, and hourly) and spatial scales (i.e., looking at the single building, 5 buildings, etc.). For now, results are related to residential buildings in a specific area of Milan and submitted to the assumption that the smart meter dataset is representative of the simulated buildings. However, the followed methodology can be exploited also with larger databases and, in the future, verified with also other types of data (i.e., an Italian TUS).

The novelty of this research relies on the direct application of an easy-to-follow procedure to set data-driven schedules in UBEM and the direct assessment of their impact on the energy results to understand at which temporal and spatial scale the random distribution of occupants-related schedules can actually be beneficial and worth of investigation. In particular, this assessment is done in an integrated way, using a heat map based on the coefficient of variation of the root mean square error (CVRMSE), as defined in the ASHRAE Guideline 14 [44].

Section 2 presents the methodology used in this study and subsection 2.1 reports the schedules creation procedure. The case study is presented in Section 3, including the description of the neighbourhood (Section 3.1.), the archetypes, the weather file used (Section 3.2 and Section 3.3.), and the smart meter dataset (Section 3.4.). In Section 4 the creation procedure of schedules and scenarios is shown applied to a specific case study. Section 5 collects the main results and in Section 6 the discussions are outlined. Finally, Section 7 reports the conclusions, the limitations of the study, and future outlooks.

2. Methodology

2.1. Workflow

The research aims to assess the impact of OB modelling in a residential neighbourhood. The exploited workflow consists of three main phases: (i) the development of the urban building energy model based on standard archetypes, (ii) the creation of schedules based on smart meter readings, and (iii) the comparison between the UBEM scenarios. Figure 1 schematizes the process.

(i) The starting point is the creation of the urban building energy model and different tools are available [3]. Generally, the geometry of the district can be deduced using geographic information system (GIS) data with footprints and buildings' heights. This information is translated into a geometric model built in a virtual 3D environment. Building archetypes, collecting both the envelope and the system specifications, are then utilized to characterize the geometry and to create the energy model. The software can be a UBEM with a built-in computer-aided design (CAD) tool [45,46] or a simulation plug-in [47]. In this case study, the geometry provided by the Municipality of Milan as GIS data [48] has been coupled with the archetypes developed for the North of Italy [8], and version 2.0 of umi [43] has been used. In Section 3 the model is described in detail.

(ii) To create schedules more representative of the actual energy use and occupancy in the buildings, the first step is to collect from energy operators a series of smart meter readings, consisting of a yearly dataset of electric energy use. Data are usually provided completely anonymized, having a registration time step of 15-minutes, without errors or gaps. If errors are present, data processing should be taken into consideration to improve the quality of results [11]. In Section 2.1 the procedure followed to create the schedules is detailed and in Section 3.4. the smart meter dataset used in the study is described.

(iii) The hourly schedules for electric load and occupancy, generated in step (ii), are combined to create different scenarios that are compared with the standard case defined by the archetypes. Here, to perform this comparison, the umi energy results in terms of Total Operational Energy (TOE) (in kWh or kWh/m^2), cooling energy needs (in kWh or kWh/m^2), heating energy needs (in kWh or kWh/m^2), electric energy uses (in kWh or kWh/m^2) are considered. In which, TOE is the sum of energy needs for cooling, heating, domestic hot water, and electric use (including lighting and appliances). Finally, the CVRMSE, as defined in the ASHRAE Guideline 14 [44] is applied to further analyse the results. In this paper, the CVRMSE is used to compare the energy outputs of Case 0 (reference scenario) to the cases characterized by different schedules' settings. The analyses are performed at different temporal scales (i.e., hourly, daily, weekly, monthly, yearly) and different spatial scales (i.e., single buildings, 5 buildings together, 10 buildings, 20 buildings, and the whole neighbourhood). The scenarios are compared in terms of energy needs as defined by ISO 52000-1:2017 [49].

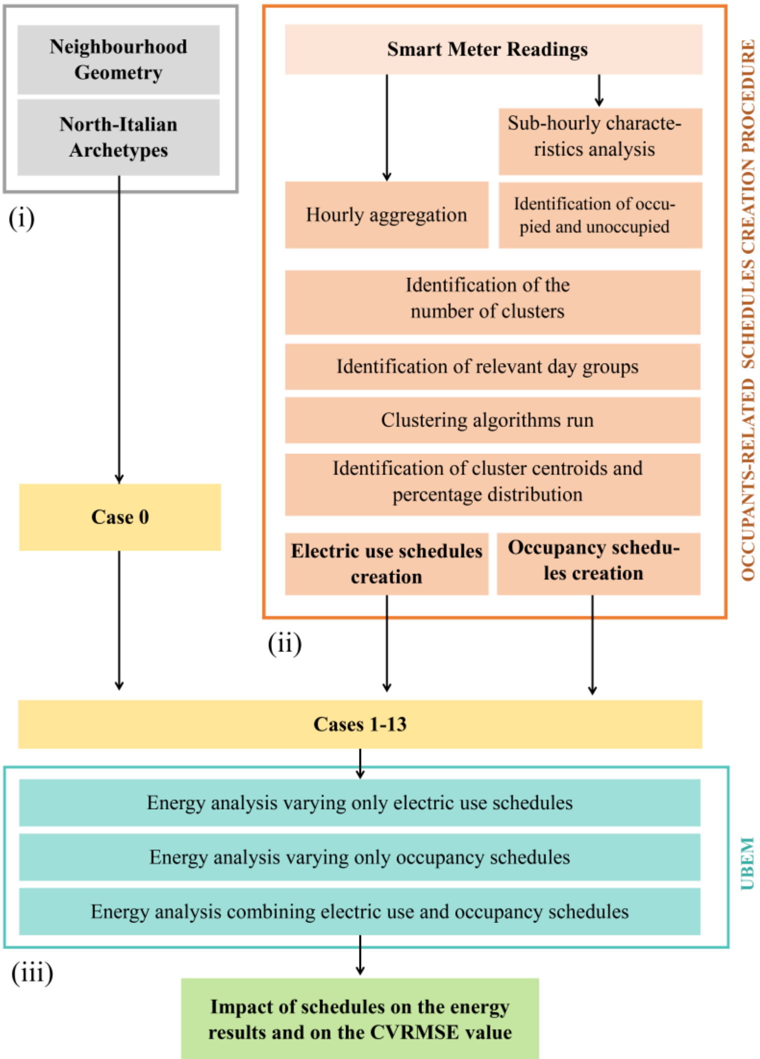


Figure 1: Scheme of the workflow

2.2. Schedules creation procedure

The proposed procedure aims to determine occupants-related schedules for UBEM based on smart meter readings, in particular, the schedules related to electric energy use (i.e., lighting and appliances) and those referred to occupancy. Smart meter readings can be collected from energy operators, DSOs, owners, etc., providing a yearly dataset of the electric energy uses. In the beginning, data processing might help if errors or gaps are present in the dataset.

In the case of electric-load schedules, if the smart meter readings are provided with 15-minute time steps, it is advisable to aggregate them in hourly values. Then, the clustering phase can start. Clustering involves categorizing a dataset into an N number of clusters C_i, where i = {1, 2, …, N}, and it is performed via the following steps: (i) set the clusters' centroids (i.e., a centroid can be imagined as the multi-dimensional average of the samples in a cluster); (ii) group the samples; (iii) revise the cluster centroids; (iv) if the centroids are unchanged, terminate, if not, go back to step (ii). This clustering aims to obtain daily schedules, formed by values ranging from 0 to 1, that are multiplied by density levels of electric usage and occupancy. To do so, it is fundamental to normalize the database to submit to the algorithm a set of 24 values (the daily pattern) ranging from 0 to 1. Thus, each group for every building is divided by its daily maximum value so that each day may result in a profile of 24 values included between 0 and 1. To solve the clustering problem, k-means is exploited [50], which is a typical and widely used unsupervised machine learning algorithm used for these applications [34,51]. However, this algorithm needs a predefined number of clusters. To set the number of clusters, the Davies-Bouldin index (DBI) is evaluated [52]. The DBI evaluates the similarities between clusters, comparing the distance between clusters with the size of the clusters themselves. The lower this metric, the better the clustering result; zero is the lowest possible value. Afterwards, the clustering algorithm is run and the cluster, together with the distribution percentages of the clusters in the database, are adopted to create hourly electric-load schedules.

To create occupancy schedules, the sub-hourly values are directly used. Generating data-driven schedules related to the occupancy of the buildings, starting from the smart meter readings, is one of the non-intrusive ways to create occupants-related schedules [53]. As presented in different studies (e.g. Carlucci et al. [7]), the idea is to consider some numerical features of the electric use within an hour as indicative of the presence of occupants. For example, a high standard deviation corresponds to high variability in the electric use within the hour that can be associated with turning on/off the devices and related to the presence of people in the flat [54,55] since no building automation system is installed. However, in most cases, a ground truth (even small) is necessary to understand when the space is occupied. For this reason, Allik et al. [56] suggest that without ground truth a simple yet reasonable way to understand occupancy is analysing the standard deviation within the hour during nighttime (from 11 pm to 5 am) when people are usually inactive and sleeping, and to assume that the third quartile can be the threshold to detect the presence of people during daytime. This threshold is thus used as a detector in the database, resulting in an unoccupied (0) and occupied (1) value for each hour. Later, the new daily 0 or 1 occupancy hours are clustered into groups with the same logic as the electric-load data. Thus, occupancy probability schedules ranging from 0 to 1 are generated.

3. The case study

This section describes the different aspects of the case study: the neighbourhood, the archetypes, the weather dataset, and the smart meter dataset.

3.1. The neighbourhood

In UBEMs the modelling of buildings involves two steps: the setting of the geometry and its characterization to create the actual energy model. The first step is usually simplified, buildings are modelled as simple boxes created by the extrusion of the footprint along the heights. In this case study, this phase was relatively easy to complete because the Municipality of Milan provided a GIS of the entire area that includes footprints and heights. GIS data allows to correctly define the ground floor areas, the mutual position, and the height of the buildings. Shape data are adapted to be used in the visual programming software Grasshopper 3D [57] and to build the volumetric model in Rhinoceros® [58].

The model consists of 49 buildings with an overall gross floor area of 56787 m^2. The height of the buildings varies between 3.5 to 16 m, with a minimum of 1 and a maximum of 5 floors. An average window-to-wall ratio of 10% is set for vertical surfaces, and the floor-to-floor height is fixed to 3 m, based on the building descriptions and onsite visits. The model is also populated with surrounding buildings, which are constructions no longer or not yet used (unoccupied, abandoned, or under a complete renovation), accounted only for shading.

Buildings are all residential and built between 1960 and 2010, according to the land registries of Milan [48].

3.2. The archetypes

The second step of modelling (i.e., characterization of the geometry) is traditionally the most difficult to solve in UBEMs [4]. Usually, archetypes are assigned to the geometry grouping the buildings based on some data (e.g., intended use, building shape, construction year, etc.). These archetypes include all the minimum characteristics to treat the geometry as a proper energy model. In this case study, the archetypes developed by Carnieletto et al. [8] for the North area of Italy are exploited. They include the characterization of systems and construction while a few attributes (e.g., window-to-wall ratio, floor-ceiling height) are left as parametric values customized to the footprint and height of the specific building. The 16 archetypes are differentiated based on the construction year and the construction type (i.e., traditional or prefabricated), which are the parameters used to assign the archetype to the geometry [59]. In the case study, 6 out of 16 residential archetypes are used and their description is reported in Table A.1 in the Appendix. In particular, 30 buildings (around 62% of the case study) are characterized by the “multi-family traditional built before 1930” archetype, 7 buildings (around 14% of the whole) by the “multi-family prefabricated built between 2000 and 2005” archetype, 5 buildings (around 10% of the whole) by the “multi-family traditional built between 1961 and 1970” archetype, 4 buildings (around 8% of the whole) by the “multi-family prefabricated built between 1981 and 1990” archetype, 2 buildings (around 4% of the whole) by the “multi-family traditional built after 2010” archetype, and lastly, 1 building (around 2% of the whole) by the “multi-family traditional built between 2005 and 2010” archetype. The prefabricated or traditional characteristics have been assumed based on visual inspections and/or based on available documentation (e.g., project specifications) provided by the Municipality of Milan. The archetypes are already completed with schedules for electric appliances, lighting, and occupancy mainly derived from the standards EN 16798-1 [60] and ISO 18523 – 1 [61]. Moreover, average density values per square meter for occupancy, lighting and appliances are set referencing specific standards (EN 16798-1 [60], EN 15193-2 [62], and EN 16798-1 [60] respectively). The schedules and the average density of the loads are fixed for all buildings and all days. In particular, an appliances density level of 3 W/m^2 and an occupancy density of 0.0353 people/m^2 is set for all archetypes. For the lighting density level, a range from 6 to 4 W/m^2 is used to take into account that old buildings can still have installed inefficient lightbulbs. To analyse the energy needs of buildings, the efficiency of the systems is set to 100% and their capacities to infinite [49]. In this way, the simulated results correspond to the energy that must be delivered (heating, or electricity) or extracted (cooling) from the internal spaces to maintain the setpoint conditions. However, the heating system is natural gas-based with a setpoint of 20 °C. The heating activation is set from mid-October to mid-April, according to the national regulation. The cooling activation is set from mid-April to mid-October with a setpoint of 26 °C.The weather dataset

The Milano-Linate weather file is used (Latitude 45°26', longitude 9°17', height 103 m) [63]. It is based on a 20 years (1951-1970) period of recording. The related weather station is less than 4 km far from the neighbourhood. The average annual temperature is 11.6 °C with a maximum monthly average of 23 °C in July and a minimum of 0 °C in January. The maximum hourly global horizontal irradiation ranges from 194 Wh/m^2 in December to 965 Wh/m^2 in June.

3.3. The smart meter dataset

The dataset provides the electric energy metering of 21 multi-family buildings for the year 2019. These buildings are partially included in the modelled area and partially situated at a maximum distance of 10 km from the modelled buildings. All the buildings are located in Milan and are used as multi-family residential buildings. Data is completely anonymous, no other survey or characterization of the tenants involved is available. The dataset has a registration time step of 15-minutes and no errors or gaps are present due to a filling algorithm based on historical electric data series, used by the utility company. The electric data is the accumulated electric absorption of all the electric uses in each apartment; thus, the registration involves lighting and electric appliances including possible small space cooling or heating devices and plug loads. This exemplifies a classic dataset accessible to energy modellers. The mean registered value of electric energy use in the database is around 4.2 Wh/m^2. The average building gross area is around 3500 m^2, with a maximum of 9322 m^2 and a minimum of 702 m^2. The buildings included in the dataset vary a lot in dimensions, even if they are all multi-family buildings divided into flats. A few buildings show a relatively high electric load per square meter, even if they are characterized by a small building area. In particular, high electric consumption is registered during summertime, probably due to the use of fans and small cooling devices.

4. Schedules creation

4.1. Electric load and occupancy schedules

Profiles are generated by clustering the dataset into groups of days with similar daily patterns following the steps described in Section 2.1. The original dataset includes 365 days for 21 buildings, thus a total of 7665 hourly daily patterns. Firstly, the dataset is divided into two groups corresponding to workdays (i.e., from Monday to Friday) and holidays (i.e., Saturday, Sunday, and national holidays). A minimum number of three clusters is intended to be interesting for this case study, to give variability to the schedules. In Figure 2, the DBI is plotted for a number of clusters that goes from 3 to 29, already after 12 clusters, the variation of the index is small or null. As explained in Section 2.1, 4 is chosen as the final number of clusters since it corresponds to the lowest values of DBI (Figure 2). This value provides a good outcome based on data characteristics, without creating not-representative clusters.

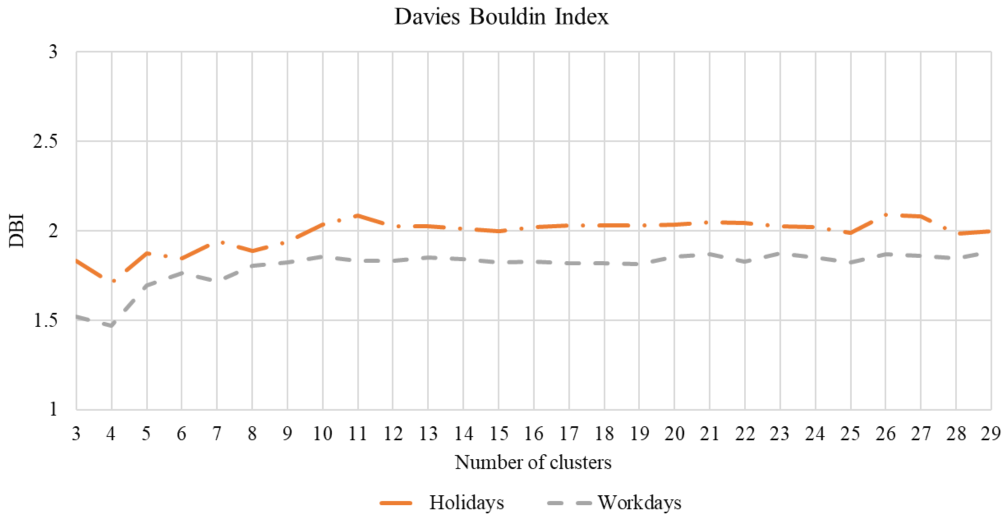


Figure 2: DBI varying the number of clusters for the two Holidays and Workdays databases

The k-means is run, and the clusters plotted in Figures 3a and 3b are generated. The percentage of distribution of each cluster in the original database is shown in Figures 3c and 3d. The same number of clusters as in the electric use case is exploited to simplify the generation of the occupancy schedules. The result, up to this point, is the setting of probability schedules ranging from 0 to 1 that represent the presence of occupants. To be used in the energy model as schedules, eventually, this probability is multiplied by the nominal number of people that are supposed to live in a building, which is estimated based on the area (i.e., 0.0353 people/m^2). Figure 4 (a and b) shows the occupancy probability for holidays and workdays, while, in Figure 4 (c and d), the percentage distribution of each cluster is plotted.

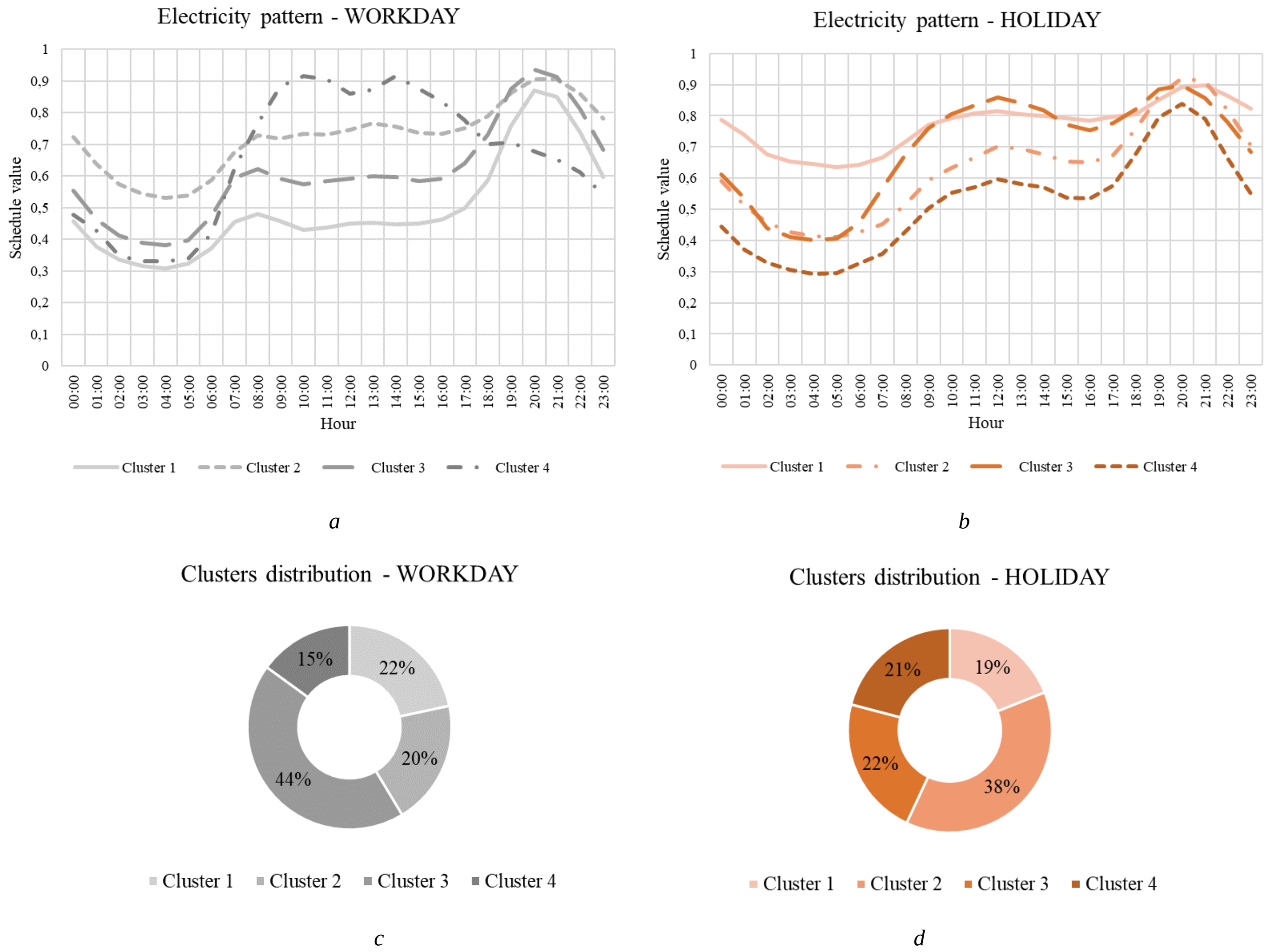


Figure 3: Electric load clusters for workdays (a) and Holidays (b) databases and electric load clusters distribution in Workdays (c) and Holidays (d) databases

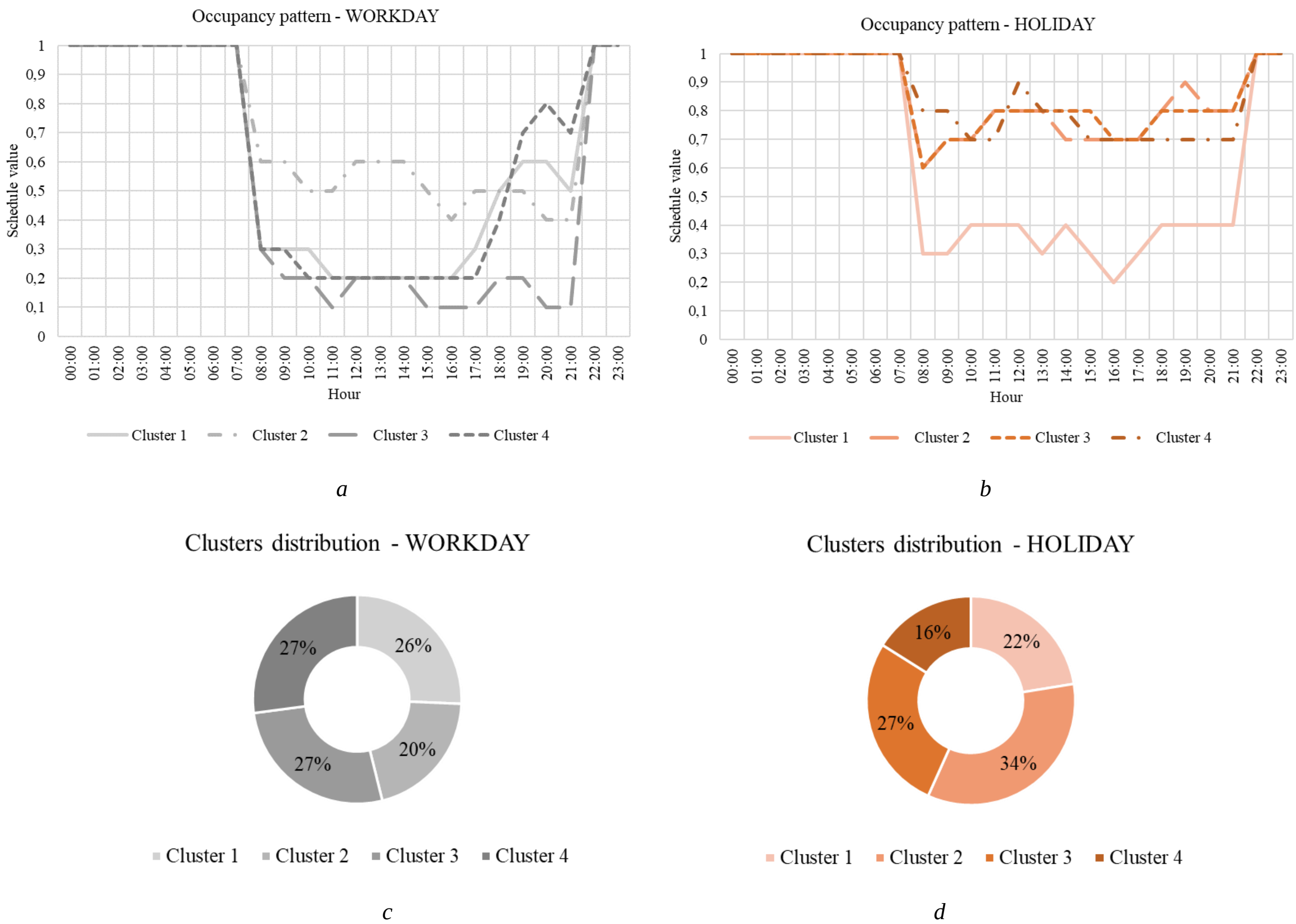


Figure 4: Occupancy clusters for Workdays (a) and Holidays (b) databases and Occupancy clusters distribution in Workdays (c) and Holidays (d) databases

4.2. Cases' generation

In umi, schedules are set as hourly patterns in a day that are then organized in weekly schedules. These weekly schedules are possibly different for each month, creating a yearly schedule. In this case study, to simplify the creation of the schedules, the holiday and workday schedules for both electricity usage and occupancy, are combined in weekly schedules, considering the different distributions of the clusters in the database (Figure 3c and 3d, and Figure 4c and 4d). In particular, the weekly schedule for Cases 1-2-3-4 in Table 1 is obtained by first comparing the distribution percentage of each electric load cluster for workdays with the distribution percentage of the holidays clusters and then selecting the two clusters which show the most similar percentage (e.g. cluster 1, in figure 3c, which shows a percentage of distribution equal to 22% has been matched with cluster 3, in figure 3d, which shows the same value). The same approach has been used for creating the weekly schedules of occupancy, presented in Cases 5-6-7-8. The weekly schedules, generated through the clustering and the distribution percentages, are kept fixed for the whole year.

With respect to Case 0, in which the schedules are set based on standards, the schedules developed by the clustering of real electric smart meter readings and the derived occupancy are used to generate new scenarios of internal gains. Firstly, to better understand the effect of the single schedules on the simulation results, 4 cases have been generated varying, with respect to Case 0, only the electric load schedules, and assigning the same schedule to all buildings (Cases 1 - 4). Secondly, other 4 cases are created in which only the occupancy schedules are varied compared to Case 0 (Cases 5 – 8). In the third step, the electric and occupancy schedules have been combined to create other 4 cases with the same schedules combination assigned to all buildings (Cases 9 - 12). Cases 1-12 are based on data-driven fixed schedules derived by clustering. The difference with respect to Case 0, in which the schedules are set based on standards, is that the schedules are developed by clustering electric smart meter readings. The resulting schedules from the clustering are used to propose different options for the same archetype to be exploited in UBEM models. Finally, the last case (Case 13) is created by exploiting the distribution percentages of the clusters in the database. In particular, to perform a more realistic scenario and to provide differentiation in the model, different schedules are assigned to a random subset of buildings in the model based on the distribution percentages of the electric workday clusters (Figure 3c). Table 1 summarizes the clusters' combinations that bring to the creation of the cases while Figure 5 shows the final distribution of the modified archetypes in Case 13. The generation of the schedules developed with this method of clustering from an existing database does not allow distinguishing between electricity used for lighting or electric appliances. For this reason, a single density level for both uses must be considered in defining the schedules. The analysis of the database shows that the average density level of electric usage is around 5 W/m^2 (resulting from the overall electric consumption of lighting and electric appliances). Moreover, to be consistent with the database results and the assumption that old buildings have a larger consumption due to more outdated appliances and lighting systems, for the buildings constructed before 1970 a total value of 6 W/m^2 is used, for the ones built between 1971 and 1999 a value of 5 W/m^2 is chosen and for building constructed after 2000, a value of 4 W/m^2 is applied. These are the values used to multiply the schedules and to create internal gains from lighting, appliances, and occupancy in the energy model.

Table 1: Clusters combinations that generate the different Case studies to be compared to Case 0

Case	Electric usage Workdays schedule(s)	Electric usage Holidays schedule(s)	Occupancy Workdays schedule(s)	Occupancy Holidays schedule(s)	Percentage of applications of buildings in the model
Case 0	From archetypes	From archetypes	From archetypes	From archetypes	100%
Case 1	Cluster 1	Cluster 3	From archetypes	From archetypes	100%
Case 2	Cluster 2	Cluster 4	From archetypes	From archetypes	100%
Case 3	Cluster 3	Cluster 2	From archetypes	From archetypes	100%
Case 4	Cluster 4	Cluster 1	From archetypes	From archetypes	100%
Case 5	From archetypes	From archetypes	Cluster 1	Cluster 3	100%
Case 6	From archetypes	From archetypes	Cluster 2	Cluster 4	100%
Case 7	From archetypes	From archetypes	Cluster 3	Cluster 2	100%
Case 8	From archetypes	From archetypes	Cluster 4	Cluster 1	100%
Case 9	Cluster 1	Cluster 3	Cluster 3	Cluster 2	100%
Case 10	Cluster 2	Cluster 4	Cluster 1	Cluster 3	100%
Case 11	Cluster 3	Cluster 2	Cluster 2	Cluster 4	100%
Case 12	Cluster 4	Cluster 1	Cluster 4	Cluster 1	100%
Case 13 (Figure 5)	Cluster 1 Cluster 2 Cluster 3 Cluster 4	Cluster 3 Cluster 4 Cluster 2 Cluster 1	Cluster 3 Cluster 1 Cluster 2 Cluster 4	Cluster 2 Cluster 3 Cluster 4 Cluster 1	around 22% around 20% around 44% around 15%

Chiaravalle neighbourhood		
Gross Floor Area (m²)	56787	
Site Ground Area (m²)	582111	
Floor Area Ratio	0.10	
Max Building Height (m)	16	
Archetypes		
Name	Color	Gross Floor Area
R_P_B1990_1		3435.47
R_P_B1990_2		3570.87
R_P_B1990_3		5461.89
R_P_B2005_1		4459.36
R_P_B2005_2		1836.82
R_P_B2005_3		11362.72
R_T_A2010_4		2070.57
R_T_B1930_1		3644.19
R_T_B1930_2		10240.02
R_T_B1930_3		3723.60
R_T_B1930_4		4661.26
R_T_B1970_1		1670.76
R_T_B1970_2		377.45
R_T_B1970_4		68.02
R_T_B2010_3		204.09

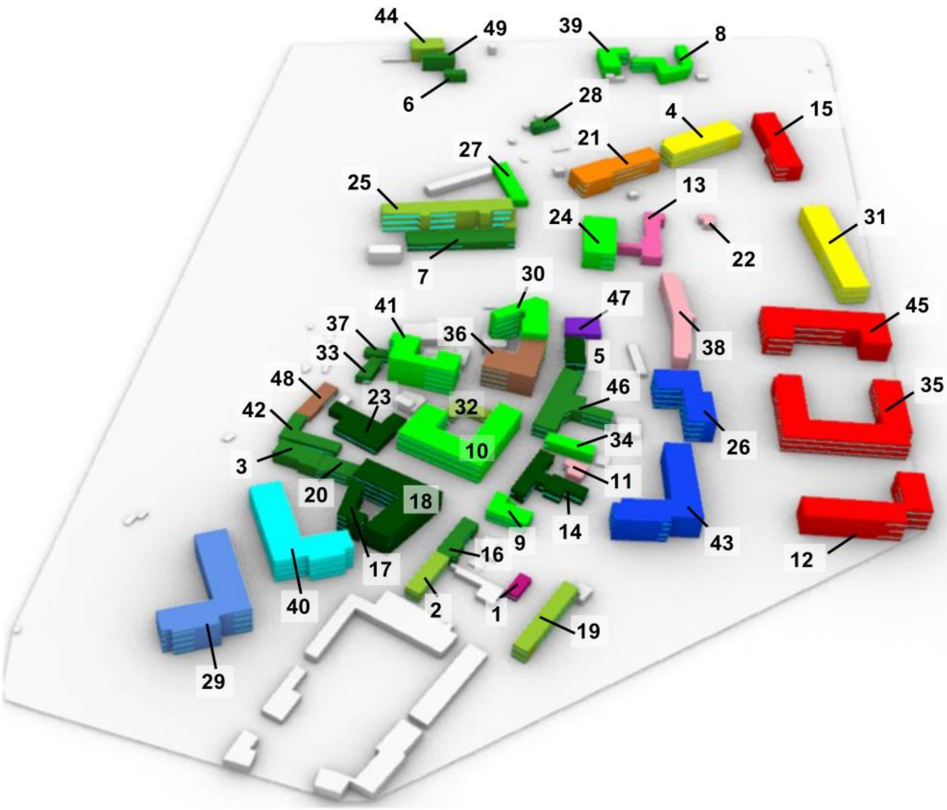


Figure 5: Distribution of the modified archetypes in the model. In the Archetypes name code, the first letter R stands for "Residential", the second can be P if "Prefabricated structure" or T if "Traditional", the third letter-number identifies the period of construction, where B stands for "Before" and A for "After", and the last number corresponds to the assigned workday electricity usage cluster (Table 1). In the neighbourhood map, the numbers correspond to the Building ID, an identification number used in the results plots.

5. Results

5.1. Case 0

The model characterized by the standard archetypes (Case 0) is adopted as a reference for the other scenarios. Figure 6 shows the annual normalized (on the gross area) results of the TOE for the entire site (Figure 6a), the energy needs for cooling (Figure 6b), the energy needs for heating (Figure 6c), and the electric energy use (Figure 6d). The colour scale represents the distribution of the energy results in kWh/m^2 in the neighbourhood. The average TOE is around 134 kWh/m^2, with a few buildings (in red) consuming around double. In general, the energy needs increase inversely to the construction years. The most efficient buildings (in blue) are all constructed after 1990. These buildings use lower amounts of energy for heating and cooling, being well insulated, and with high-efficiency windows. The heating is set from mid-October to mid-May, following the regulated heating season in Milan, with a setpoint of 19 °C. While the cooling is set from mid-May to mid-October, and it is activated when the internal temperature is above the setpoint, which has been defined to be 26 °C. The energy need for domestic hot water is fixed, with negligible differences among the buildings, and counts for an average of 29 kWh/m^2. Domestic hot water is not deeply investigated in this case study because it is kept fixed among the Cases. The deterministic setting of electric use for lighting and appliances in three different groups brings net distinctions in the electric uses (Figure 6d).

A complete validation of this model is relatively useless for this scenario since the focus is on the contrast among results of different scenarios, more than on the value of the energy results. However, the values have been compared with North Italian benchmarks present in the literature, which show to be consistent [64,65].

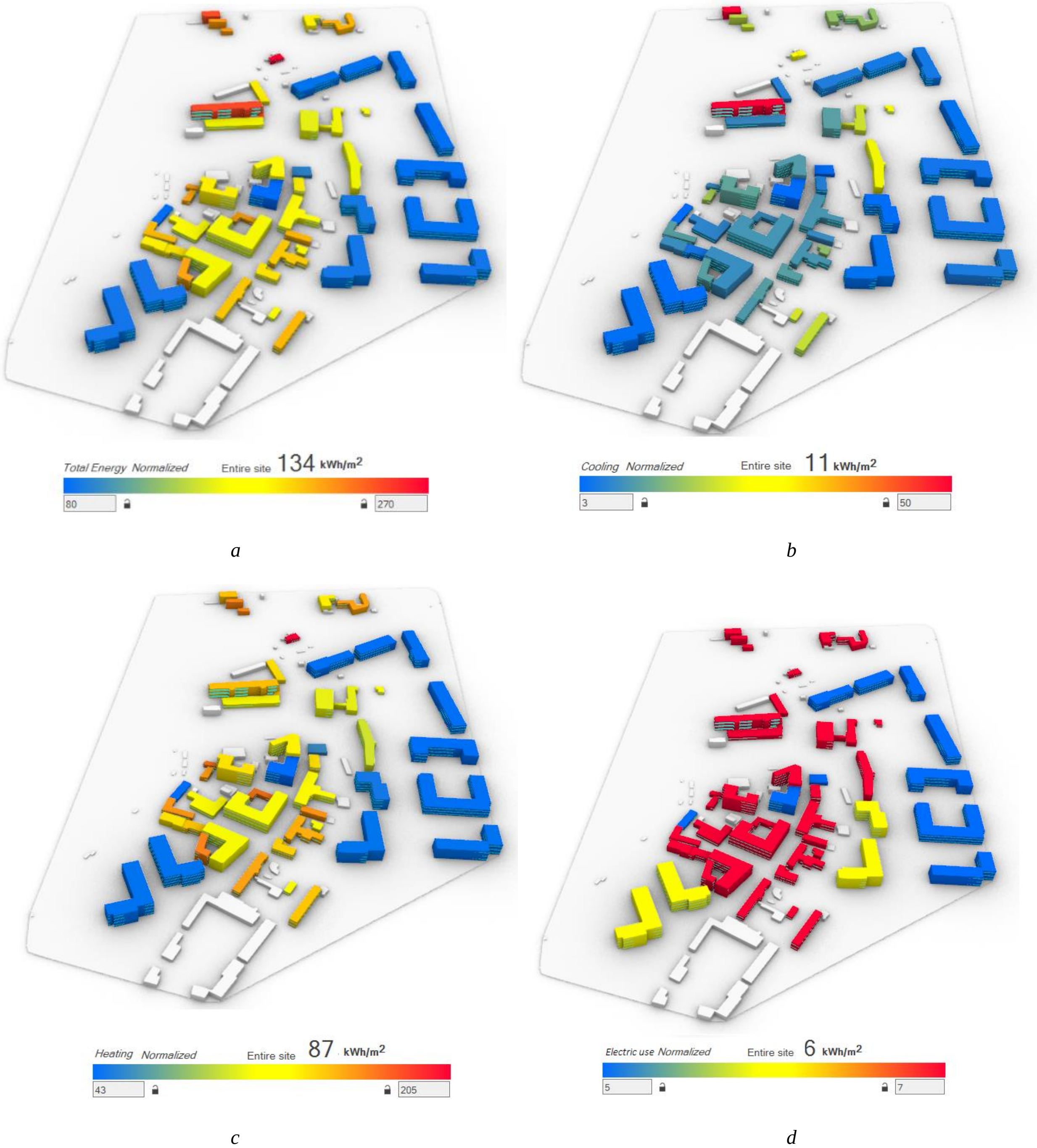


Figure 6: Case 0: Yearly TOE (a). Average and distribution of energy needs for cooling (b), heating (c), and electric use (d).

Figure 7 shows the daily results in terms of daily energy needs (kWh) for the neighbourhood. In particular, Figure 7 shows the daily sum of TOE, cooling energy needs, heating energy needs, and electric energy use. The cooling peak is reached in the hottest months

(July and August), with large variability between days due to the different external weather conditions. The highest consumption in terms of the TOE is reached in January with a few peaks related to the coldest days. Finally, the electric usage is constant over the days.

Figure 8 highlights the daily and weekly electric energy use due to lighting and appliances in the buildings of the model. It is visible how all buildings show the same repetitive pattern with the only change in the absolute values due to the different sizes of the buildings. This study aims to insert a realistic variability among buildings characterized by the same archetype.

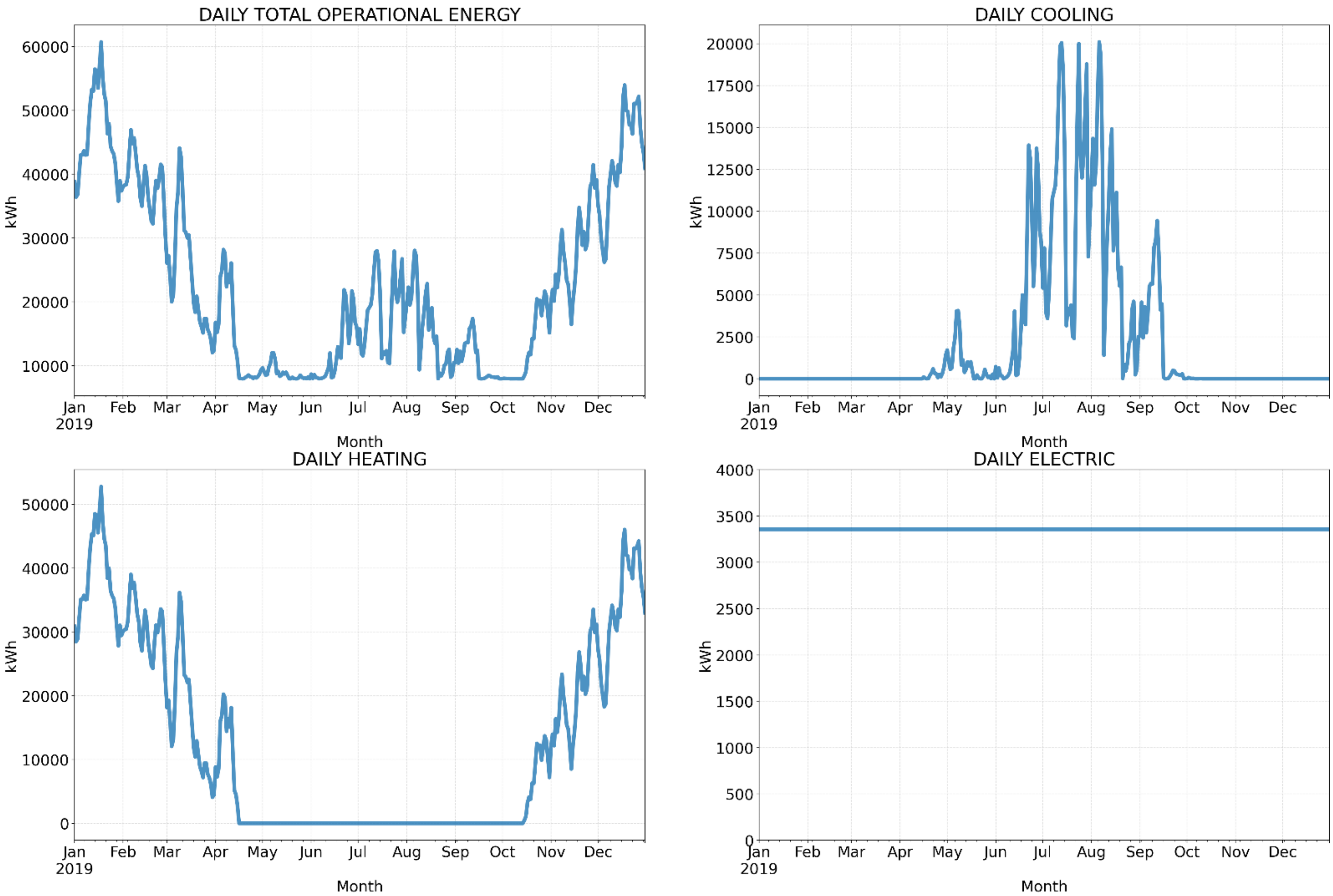
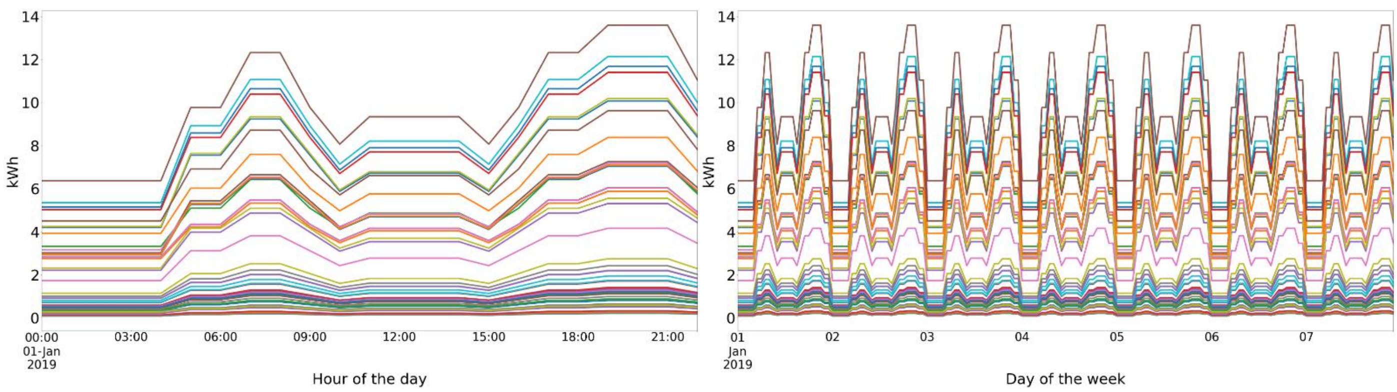


Figure 7: Neighborhood daily energy need for Case 0.



a

b

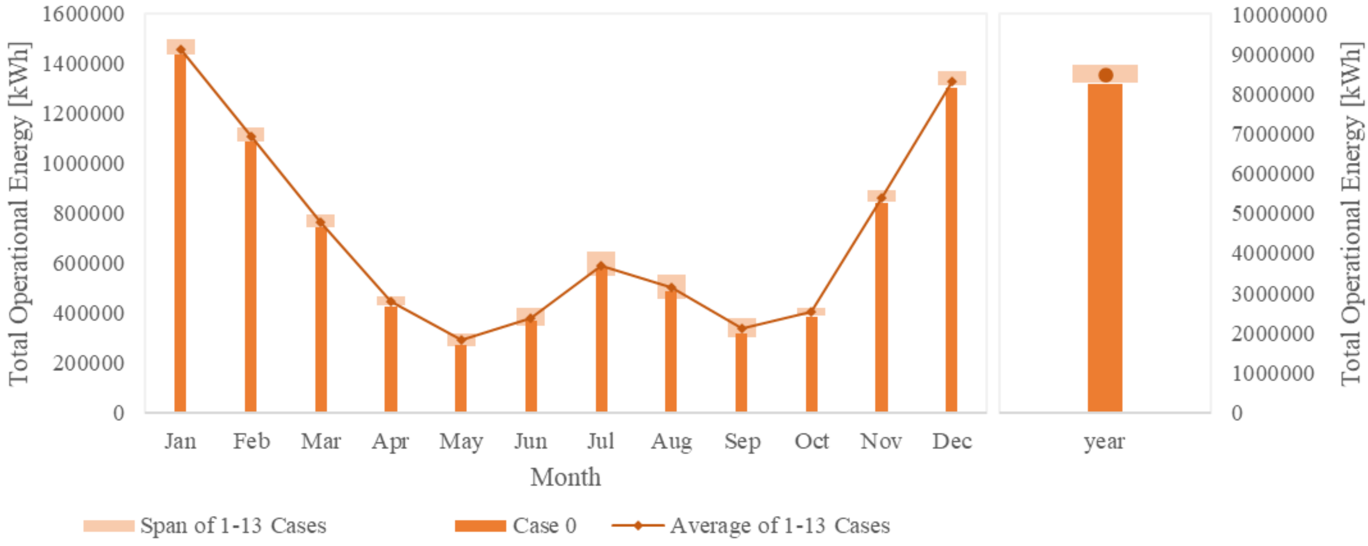
Figure 8: Hourly daily pattern (a) and weekly (b) electric use for the single buildings in the model for Case 0. The day of the week is set as a number from Sunday (01) to Saturday (07).

5.2. Comparison with schedules scenarios

To understand the effect that the schedules have on the energy results, a comparison among the scenarios is performed in terms of TOE, cooling energy needs, heating energy needs, and electric energy uses (Figure 9-12). In Figures 9-12, the Case 0 results are shown together with the results span of the other cases (1-13) and their average, to point out the variability of outputs due to the implementation of the different schedules.

Comparing the TOE of Case 0 with respect to Cases 1-13 average, a yearly average increase of +2.7% is registered (Figure 9b). The maximum monthly increase between the Case 0 and the average of Cases 1-13 is +6.1% in May, the minimum is 1.6% in January with an average of +3.4% (Figure 9a). The yearly maximum result is for Case 13, which, compared to Case 0, has a TOE of +6% and +3.2% compared to the Cases 1-13 average. The yearly minimum result is associated with Case 8 with an increase of +0.2% with respect to Case 0 and of -2.4% with respect to Cases 1-13 average. In Cases 1-4, the schedules related to electric use are changed with respect to Case 0, while in Cases 5-8 the change is related only to the occupancy schedules. Cases 9-13 are characterized by a change in both electric and occupancy schedules. These schedules variations, from an energy-balance point of view, change directly the electric use of the buildings and indirectly the internal gains due to appliances and occupancy. The electric schedules will affect both the electric use and the internal gains, while the occupancy schedules affect only the internal gains. Thus, Case 13 has a combination of schedules that brings to an overall increase of TOE with respect to Case 0 and this is expected to be one of the cases in which both electric and occupancy schedules vary. Conversely, Case 8 is one of the cases in which only the occupancy schedule changes with respect to Case 0, bringing to an overall lower difference.

Regarding the single buildings TOE results (Figure 9c), the average increase of Cases 1-13 average compared to Case 0 is +3% with a maximum of +7.7% and a minimum of +1.0%. The maximum among Cases 1-13 with respect to Case 0 is +11.8% on average, with a maximum of 56.9% for building 13 and a minimum of 3.2% for building 45. The minimum of Cases 1-13, in this case, is lower than Case 0 with an average of -0.5%, a maximum of +0.9% for building 32, and a minimum of -6.9% for building 22.

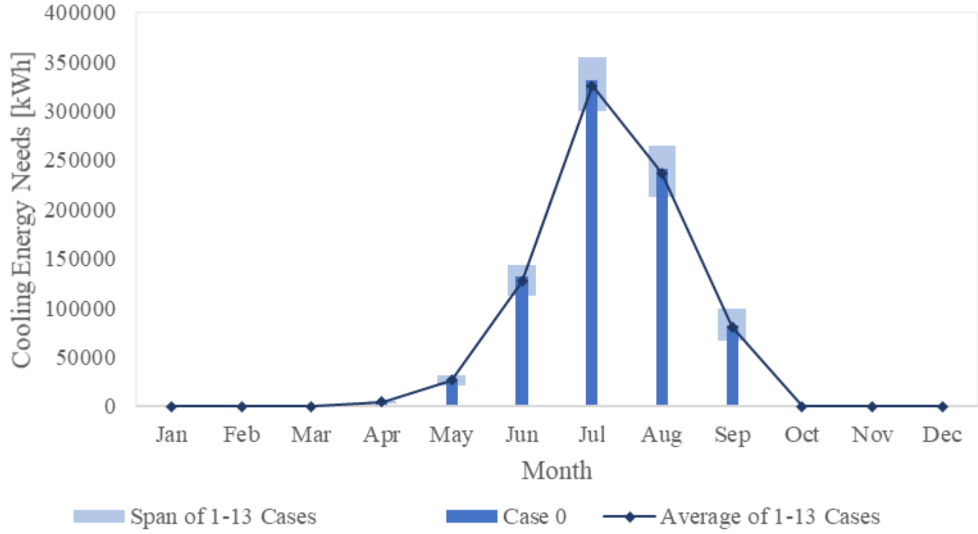


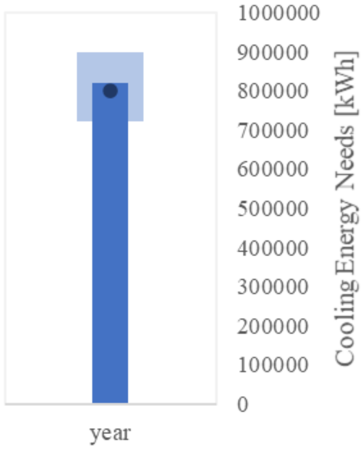
a *b*

c

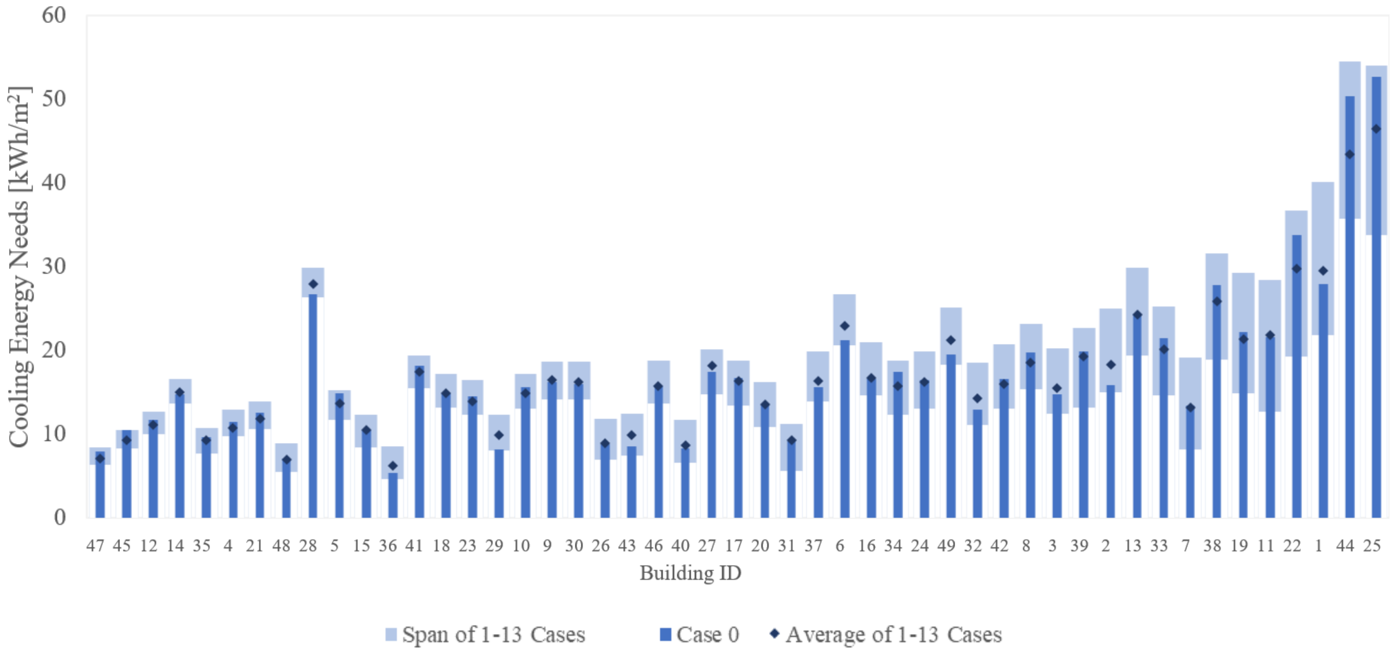
Figure 9: Monthly (a) and Annual (b) Total Operational Energy ranges within all cases for the whole neighbourhood. Annual Total Operational Energy ranges for each building (c).

Comparing the cooling energy needs of Case 0 with respect to Cases 1-13 average, a yearly average decrease of -2.2% is registered (Figure 10b). The minimum monthly increase between the Case 0 and the average of Cases 1-13 is -28.3% in October, and the maximum is -1.6% in July with an average of -7.2% (Figure 10a). The yearly maximum result is simulated for Case 12, that compared to Case 0 has a cooling energy need of +9.6% and +12.1% compared to the Cases 1-13 average. The yearly minimum result is associated with Case 5 with a decrease of -12.4% respect to Case 0 and -10.4% respect to Cases 1-13 average. The increase in cooling needs is directly correlated to an increase in internal gains and vice versa. Thus, due to the specific combination of schedules, Case 12 is the one with the maximum amount of total internal gains (i.e., electric appliances and occupancy), while Case 5 has the lowest internal gains. Regarding the single buildings' cooling needs (Figure 10c), the average decrease of Cases 1-13 average compared to Case 0 is -0.2% with a maximum of +21.7% for building 29 and a minimum of -13.8% for building 44. The maximum among Cases 1-13 with respect to Case 0 is +21.9% on average, with a maximum of 58.5% for building 36 and a minimum of 0.2% for building 45. The minimum of Cases 1-13, in this case, is lower than Case 0 with an average of -19.5%, a maximum of -0.3% for building 29, and a minimum of -42.9% for building 22.





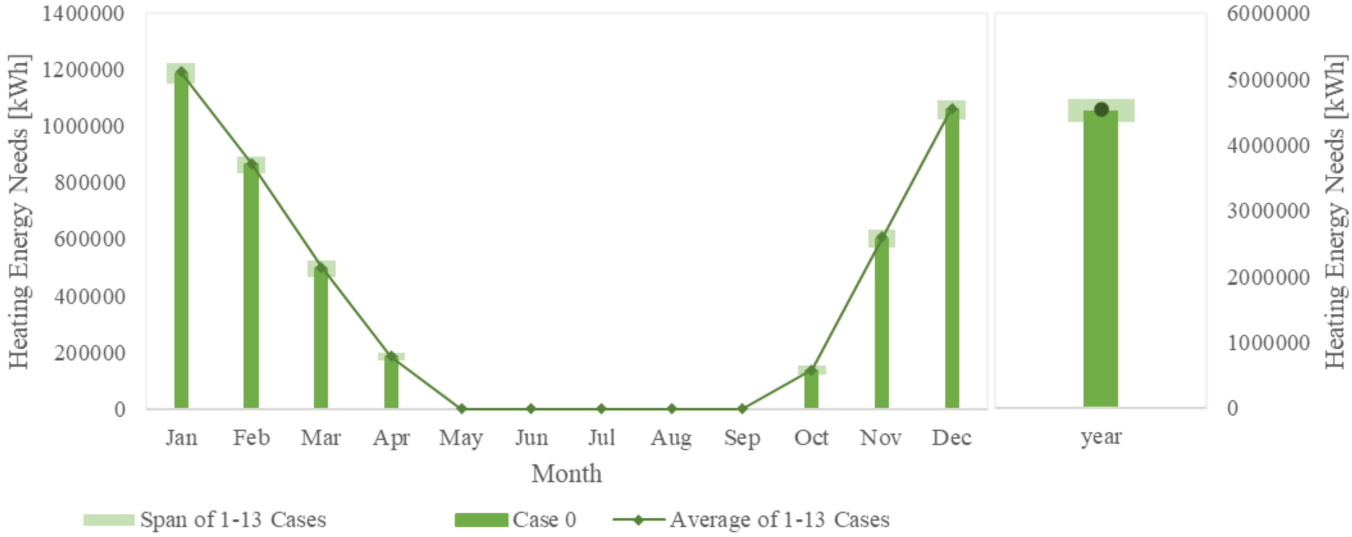
a *b*



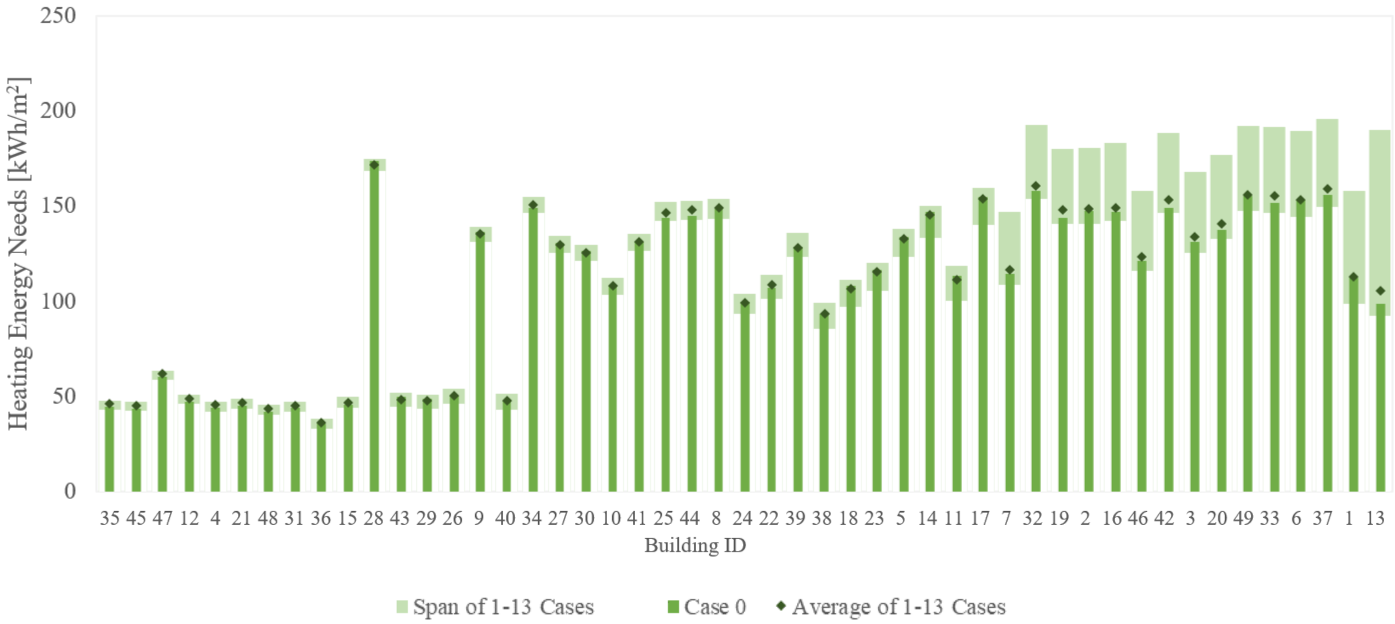
c

Figure 10: Monthly (a) and Annual (b) cooling energy ranges within all cases for the whole neighbourhood. Annual cooling energy ranges for each building (c).

Comparing the heating energy needs of Case 0 with respect to Cases 1-13 average, a yearly average increase of +0.6% is registered (Figure 11b). The minimum monthly increase between the Case 0 and the average of Cases 1-13 is +0.4% in January and December, the maximum is +1.5% in October with an average of +0.8% (Figure 11a). The yearly maximum result is simulated for Case 5, that compared to Case 0 has an energy need for heating of +4.0% and +3.4% compared to the Cases 1-13 average. The yearly minimum result is associated with Case 12 with a decrease of -4.4% respect to Case 0 and -5.0% respect to Cases 1-13 average. . Conversely to the cooling needs, the increase of heating needs is correlated to a decrease of internal gains and vice versa. Thus, as resulting also from the discussion about the cooling needs, Case 12 is the one with the maximum amount of total internal gains (i.e., electric appliances and occupancy), while Case 5 has the lowest internal gains. Regarding the single buildings' heating needs (Figure 11c), the average increase of Cases 1-13 average compared to Case 0 is +1.0% with a maximum of +6.9% for building 13 and a minimum of -2.5% for building 43. The maximum among Cases 1-13 with respect to Case 0 is +13.0% on average, with a maximum of +92.2% for building 13 and a minimum of 0.9% for building 28. The minimum of Cases 1-13, in this case, is lower than Case 0 with an average of -5.3%, a maximum of -1.2% for building 25, and a minimum of -12.3% for building 1.



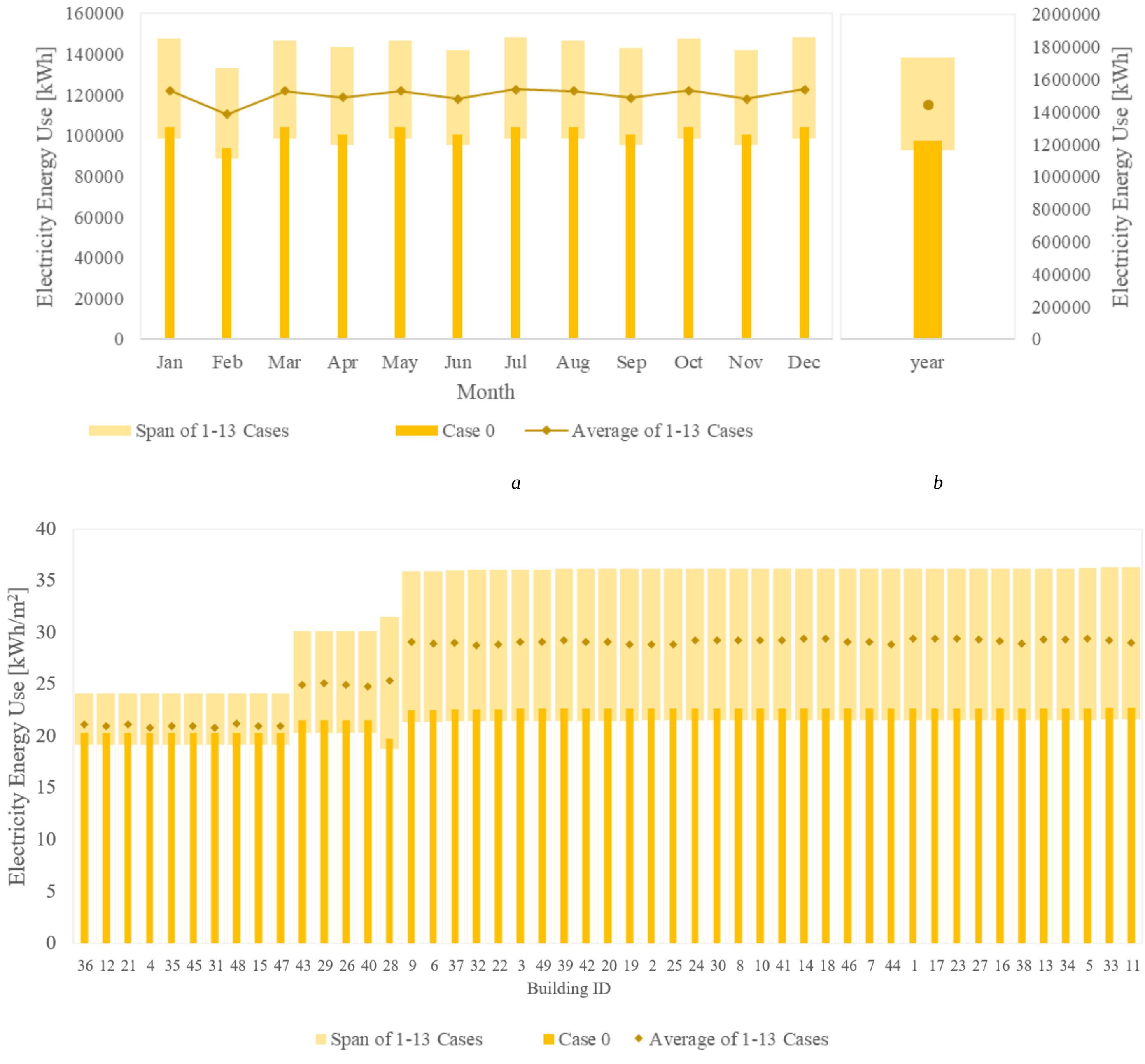
a *b*



c

Figure 11: Monthly (a) and Annual (b) heating energy ranges within all cases for the whole neighbourhood. Annual heating energy ranges for each building (c).

Lastly, the same analysis is made for electric use (Figure 12). This is purely the sum of the electric use in the buildings of the model. The variability throughout the year is small, due to the different number of hours of each month. Compared to Case 0, the scenarios increase the electric use by +17.5% on average. The maximum result among the scenarios is +41.7% with respect to Case 0 and the minimum is -5.4% lower than the Case 0 result (Figure 12a and Figure 12b). Regarding the single buildings' electric use (Figure 11c), the average increase of Cases 1-13 average compared to Case 0 is +22.4% with a maximum of +29.7% for building 5 and a minimum of +2.4% for building 31. The maximum among Cases 1-13 with respect to Case 0 is +49.5% on average, with a maximum of +59.5% and a minimum of +18.5%. The minimum of Cases 1-13, in this case, is lower than Case 0 with an average of -5.3%, a maximum of -5.2%, and a minimum of -5.7%. Looking at the single building results, by the average of Case 1-13, three groups emerged, connected to the characterization via archetypes.



a *b*

c

Figure 12: Monthly (a) and Annual (b) electric use ranges within all cases for the whole neighbourhood. Annual electric use ranges for each building (c).

6. Discussions

6.1. Cases from 1 to 12

To compare scenarios 1-13 with the reference Case 0, the coefficient of variation of the root mean square error (CVRMSE) defined in the ASHRAE Guideline 14 [44] is exploited. According to the ASHRAE Guideline, a model to be considered calibrated should have a maximum CVRMSE of 15% relative to monthly data and 30% for hourly data. In this case study, this quantity is not used to calibrate the model, but as an index to quantify the variation or randomness between Case 0 and the other cases in which the schedules are changed. The CVRMSE values are calculated at various temporal scales (i.e., hourly, daily, weekly, monthly, yearly) and spatial scales (i.e., single building, 5 buildings together, 10 buildings, 20 buildings, and the whole neighbourhood). The results are shown in a heat map that enables the visualization and quantification of the resulting maximum CVRMSE.

Figure 13 reports the results of the CVRMSE related to TOE between Case 0 and the cases from 1 to 12. Figures A.1-A.3 in the Appendix show the results in terms of energy needs for heating, energy needs for cooling, and electricity energy uses. From Case 1 to Case 4 the changes regard the electric use schedule. From Case 5 to Case 8, the occupancy schedules vary, while from Case 9 to Case 12 both electric use and occupancy schedules are modified with respect to Case 0.

Analysing the maximum CVRMSE for the whole neighbourhood, it remains low basically at all temporal scales in terms of the TOE. However, it increases progressively going towards the single building and the hourly time step. In particular, the change in the electric use schedules (Cases 1-4) has a higher influence than the only change in occupancy (Cases 5-8), since both of them indirectly impact the internal loads, and consequently the cooling and heating energy needs, but the electricity schedules directly impact on the electric use of the buildings. The combination of the two (Cases 9-12) shows a higher influence with a peak of 20% for Case 10 on the single building hourly scale. This result is lower but not so far from the limit imposed by the ASHRAE Guideline 14 [44] of 30% which indicates that a model is uncalibrated for its reference data. The same plots regarding heating energy needs (Figure A.1), cooling energy needs (Figure A.2), and electric uses (Figure A3) can be found in the appendix. Particularly, the temporal scale results are in line with the literature [39] and now are confirmed for the urban scale too. However, the spatial scale results are innovative and bring new discussions to UBEM and its OB modelling.

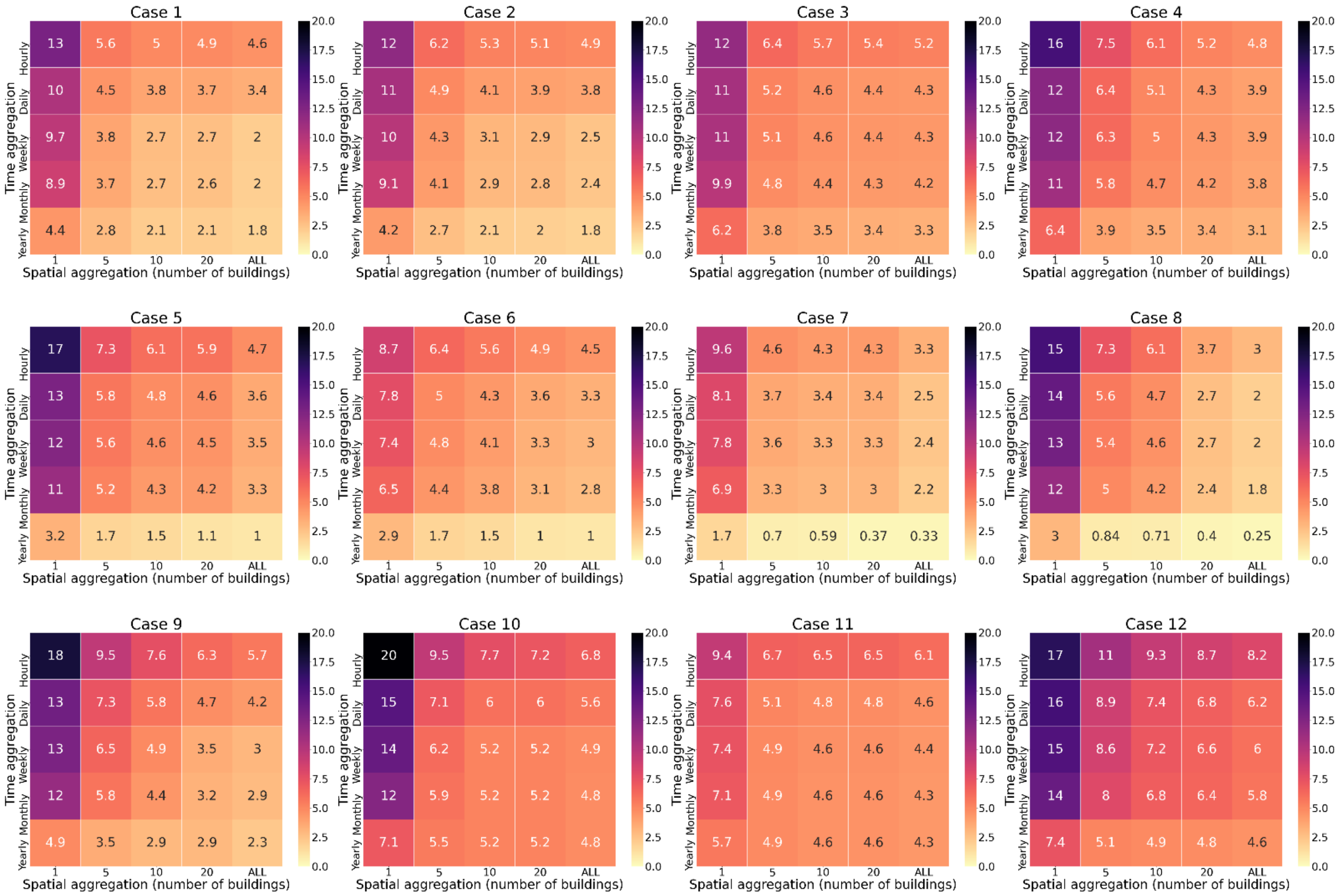
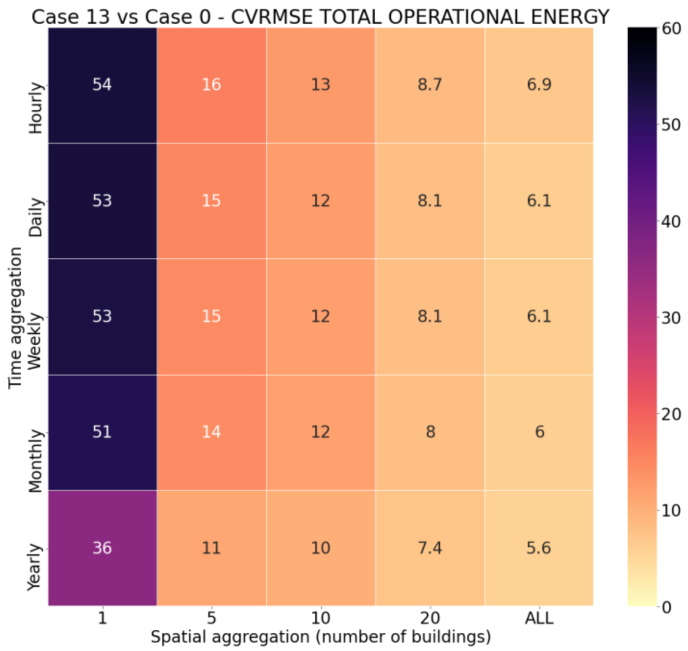


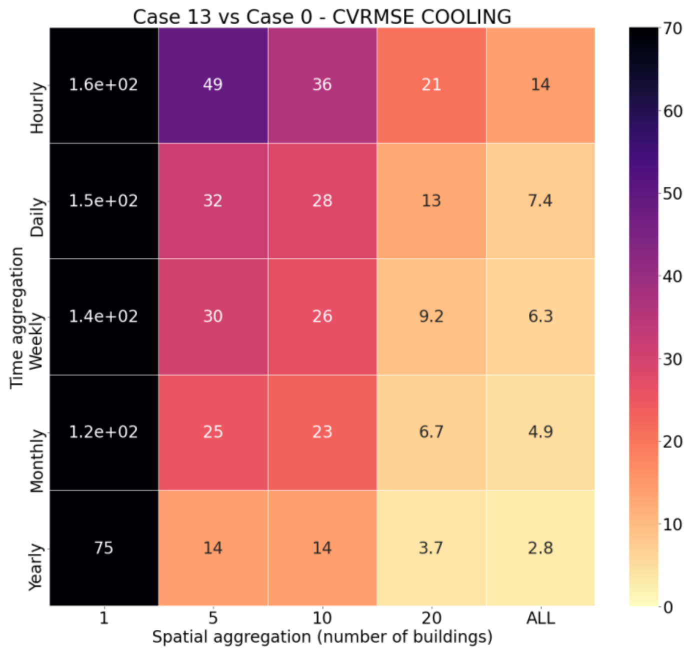
Figure 13: CVRMSE values related to the TOE difference between Case 0 for Cases from 1 to 12, with the different spatial (horizontally) and temporal (vertically) aggregations.

6.2. Case 13

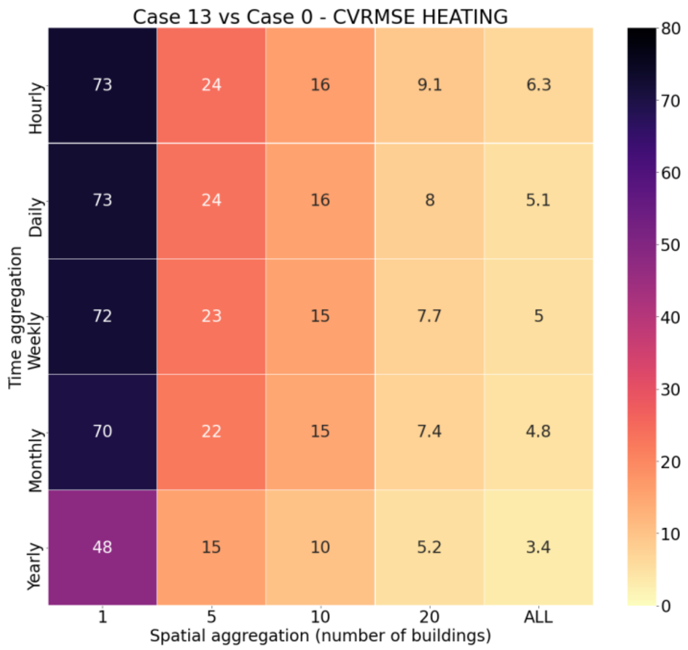
Case 13 involves the combination of the different schedules' scenarios spread within the neighbourhood model (Figure 5). The aim is to create a more realistic data-driven representation of the schedules' variability in the model. Figure 14 shows the CVRMSE values for the TOE (Figure 14a), cooling and heating energy needs (Figure 14b and 14c respectively), and electricity energy uses (Figure 14d). In terms of TOE (Figure 14a), the CVRMSE values show that there is a large difference in the single building scale that decreases strongly already aggregating 5 buildings together. This means that there are a few buildings in the model showing a large difference between Case 13 and Case 0 but on average the values are quite similar. However, investigating the cooling, heating, and electricity energy separately, the CVRMSE increases in its maximum value compared to the TOE. In particular, the electric use difference for time and spatial aggregation has a higher neighbourhood yearly value, but it shows a lower deviation. The high value of CVRMSE for the yearly neighbourhood value, compared to one of the other energy indicators, is justified by the fact that the schedules for electricity are directly changing the electric use of the buildings, while the other indicators are only indirectly affected. Also in this case, the largest values are registered for the cooling, as outlined in Cases 1-12. The high results of the cooling at 1 building spatial aggregation can be due to the combination of schedules that brings to high internal gains and consequently to increased activation of the cooling systems.



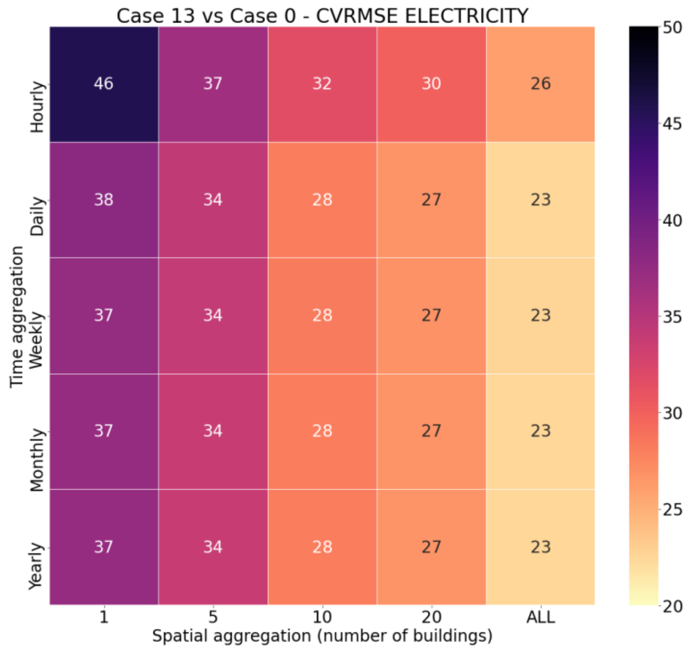
a



b



c



d

Figure 14: CVRMSE values related to the results difference between Case 0 and Case 13, with the different spatial (horizontally) and temporal (vertically) aggregations, related to Total Operational Energy (a), Cooling (b), Heating (c), Electricity (d)

7. Conclusions

In this research, smart meter readings are exploited to derive occupancy schedules and electric load profiles for UBEMs. Smart meters are widely installed in cities, but, still, in Italy, data covering entire districts or neighbourhoods are not easily accessible. With the developed procedure, readings from a few buildings are exploited to develop schedules for an entire neighbourhood inserting variability among buildings characterized by the same archetypes. This research has two main goals. Firstly, it proposes a general procedure to create data-driven schedules for electric use and occupancy derived from smart meter readings to be used in UBEMs for residential buildings with a random distribution. Secondly, it assesses the impact of the schedules on UBEMs' energy results at different temporal resolutions (i.e., annual, monthly, weekly, daily, hourly values) and spatial scales (i.e., looking at the single building, five buildings, ten buildings, etc).

The outcomes show that the differences in energy needs, resulting from the application of various schedules, widely range and change based on temporal and spatial aggregation. In particular, Case 0, with fixed and predefined schedules, tends to underestimate the energy results compared to scenarios where schedules based on measured data are used. The average yearly TOE of Case 1-13, with data-driven schedules, is +2.7% with respect to Case 0. Particularly, for the entire neighbourhood on the yearly results, the heating energy needs, the cooling energy needs, and the electric uses are estimated respectively to be -2%, +1%, and +18% compared to Case 0. While looking at the results for the single buildings, the cooling energy needs difference between Case 0 and the average of the scenarios ranges from -14% to +22%. The heating energy needs difference ranges from -2% to +7%, and for the electrical energy use from +2% to +30%. The use of heat maps showing the CVRMSE at different temporal and spatial aggregations allows a visual and numerical overview of the discordance between the scenarios and Case 0, considering that the threshold for a single building model to be intended as "calibrated" is 30% and 15% for the hourly and monthly values. Especially, for Case 13, the one in which the schedules are randomly mixed in the neighbourhood, the hourly results are higher than the threshold set by the standard for single buildings. While, in general, yearly values are always relatively low. This means that if the focus of the analysis is hourly energy values, special attention must be given to the schedules. In general, we can say that the importance of schedules is relatively low looking at the maximum spatial and temporal aggregation (i.e., whole model at yearly value). However, when the focus of the analysis is done on hourly values or a few buildings, the schedules take more and more importance. This, for instance, is the case of renewable energy communities, where UBEM simulations are used to inform the designer both to dimension district storage system, and to evaluate economic incentives, typically provided on hourly values of shared energy. For this kind of analysis, the schedule-creation procedure proposed in this paper may prove extremely useful when only smart meter readings for a few buildings in the urban model are provided. For the direct user of the results (e.g., energy community developers, transmission and distribution system operators, etc.), often the simulation is needed at the urban scale with high temporal resolution (i.e., hourly). Thus, a workflow like the presented one can solve the problem, setting the framework for further development. In particular, for the matching between energy production and use, focused on the study of the hourly UBEMs results, even more realistic schedules must be set differentiating them also among months.

The results of the study are limited to residential buildings and one specific area of Milan. Moreover, the assumption that the dataset is representative of the simulated buildings is done. This assumption can be made because the buildings from which the smart meter readings are available, are multi-family residential buildings as in the modelled area, in the same area of the same city. However, the development of an Italian Time Use Survey may enable a generalization of the process. In the future, different building types will be included (e.g., offices, commercial buildings, etc.) and the analysis will be extended to a larger number of buildings, potentially reaching the entire city. TUS can be adopted to validate or expand the occupants' model including also other activities rather than the presence or absence in the building and the related electric use. Moreover, the weekly schedules created by the coupling of a workday and holiday could be created not only based on the percentage distribution of the clusters but by randomly choosing among the options. Also, the schedules could be differentiated between months.

Acknowledgement

The study was developed within the framework of the projects SHAR-LLM (Sharing Cities) and NRG2peers, which have received funding from the European Union's Horizon 2020 research and innovation program respectively under grant agreement No 691895, and No 890345.

References

[1] F. Causone, A. Sangalli, L. Pagliano, S. Carlucci, Assessing energy performance of smart cities, Build. Serv. Eng. Res. Technol. 39 (2017) 99–116. doi:10.1177/0143624417725220.

[2] F. Causone, A. Sangalli, L. Pagliano, S. Carlucci, An Exergy Analysis for Milano Smart City, Energy Procedia. 111 (2017) 867–876. doi:10.1016/j.egypro.2017.03.249.

[3] M. Ferrando, F. Causone, T. Hong, Y. Chen, Urban building energy modeling (UBEM) tools: A state-of-the-art review of bottom-up physics-based approaches, Sustain. Cities Soc. 62 (2020) 102408. doi:10.1016/j.scs.2020.102408.

[4] T. Hong, Y. Chen, X. Luo, N. Luo, S.H. Lee, Ten questions on urban building energy modeling, Build. Environ. 168 (2020) 106508. doi:10.1016/j.buildenv.2019.106508.

[5] S. Carlucci, F. Causone, S. Biandrate, M. Ferrando, A. Moazami, S. Erba, On the impact of stochastic modeling of occupant behavior on the energy use of office buildings, Energy Build. (2021) 111049. doi:https://doi.org/10.1016/j.enbuild.2021.111049.

[6] M. Ferrando, T. Hong, F. Causone, A simulation-based assessment of technologies to reduce heat emissions from buildings, Build. Environ. 195 (2021) 107772. doi:10.1016/j.buildenv.2021.107772.

[7] S. Carlucci, M. De Simone, S.K. Firth, M.B. Kjærgaard, R. Markovic, M.S. Rahaman, M.K. Annaqeeb, S. Biandrate, A. Das, J.W. Dziedzic, G. Fajilla, M. Favero, M. Ferrando, J. Hahn, M. Han, Y. Peng, F. Salim, A. Schlüter, C. van Treeck, Modeling occupant behavior in buildings, Build. Environ. 174 (2020) 106768. doi:https://doi.org/10.1016/j.buildenv.2020.106768.

[8] L. Carnieletto, M. Ferrando, L. Teso, K. Sun, W. Zhang, F. Causone, P. Romagnoni, A. Zarrella, T. Hong, Italian prototype building models for urban scale building performance simulation, Build. Environ. (2021) 107590. doi:https://doi.org/10.1016/j.buildenv.2021.107590.

[9] G. Happle, J.A. Fonseca, A. Schlueter, A review on occupant behavior in urban building energy models, Energy Build. 174

(2018) 276–292. doi:10.1016/j.enbuild.2018.06.030.

[10] E. Himpe, A. Janssens, Identification of energy use time patterns of occupied dwellings using smart meter data, E3S Web Conf. 111 (2019). doi:10.1051/e3sconf/201911105011.

[11] F. Causone, S. Carlucci, M. Ferrando, A. Marchenko, S. Erba, A data-driven procedure to model occupancy and occupant-related electric load profiles in residential buildings for energy simulation, Energy Build. 202 (2019) 109342. doi:10.1016/J.ENBUILD.2019.109342.

[12] B. Dong, Y. Liu, H. Fontenot, M. Ouf, M. Osman, A. Chong, S. Qin, F. Salim, H. Xue, D. Yan, Y. Jin, M. Han, X. Zhang, E. Azar, S. Carlucci, Occupant behavior modeling methods for resilient building design, operation and policy at urban scale: A review, Appl. Energy. 293 (2021) 116856. doi:10.1016/j.apenergy.2021.116856.

[13] X. Luo, T. Hong, Y. Chen, M.A. Piette, Electric load shape benchmarking for small- and medium-sized commercial buildings, Appl. Energy. 204 (2017) 715–725. doi:10.1016/j.apenergy.2017.07.108.

[14] C. Beckel, L. Sadamori, S. Santini, T. Staake, Automated customer segmentation based on smart meter data with temperature and daylight sensitivity, 2015 IEEE Int. Conf. Smart Grid Commun. SmartGridComm 2015. (2016) 653–658. doi:10.1109/SmartGridComm.2015.7436375.

[15] M.H. Kristensen, R.E. Hedegaard, S. Petersen, Hierarchical calibration of archetypes for urban building energy modeling, Energy Build. 175 (2018) 219–234. doi:https://doi.org/10.1016/j.enbuild.2018.07.030.

[16] G. Ren, M. Sunikka-blank, X. Zhang, Young urban households in Shanghai , China : Characteristics of energy use and attitudes, Sustain. Cities Soc. 60 (2020) 102174. doi:10.1016/j.scs.2020.102174.

[17] A. Spoladore, D. Borelli, F. Devia, F. Mora, C. Schenone, Model for forecasting residential heat demand based on natural gas consumption and energy performance indicators, Appl. Energy. 182 (2016) 488–499. doi:10.1016/j.apenergy.2016.08.122.

[18] R. El Kontar, T. Rakha, Profiling Occupancy Patterns to Calibrate Urban Building Energy Models (UBEMs) Using Measured Data Clustering, Technol. + Des. 2 (2018) 206–217. doi:10.1080/24751448.2018.1497369.

[19] P. Gianniou, C. Reinhart, D. Hsu, A. Heller, C. Rode, Estimation of temperature setpoints and heat transfer coefficients among residential buildings in Denmark based on smart meter data, Build. Environ. 139 (2018) 125–133. doi:10.1016/j.buildenv.2018.05.016.

[20] J.Y. Park, X. Yang, C. Miller, P. Arjunan, Z. Nagy, Apples or oranges? Identification of fundamental load shape profiles for benchmarking buildings using a large and diverse dataset, Appl. Energy. 236 (2019) 1280–1295. doi:https://doi.org/10.1016/j.apenergy.2018.12.025.

[21] T. Rakha, R. El Kontar, Community energy by design: A simulation-based design workflow using measured data clustering to calibrate Urban Building Energy Models (UBEMs), Environ. Plan. B Urban Anal. City Sci. 46 (2019) 1517–1533. doi:10.1177/2399808319841909.

[22] R. Razavi, A. Gharipour, M. Fleury, I.J. Akpan, Occupancy detection of residential buildings using smart meter data: A large-scale study, Energy Build. 183 (2019) 195–208. doi:10.1016/j.enbuild.2018.11.025.

[23] Y.S. Chiou, K.M. Carley, C.I. Davidson, M.P. Johnson, A high spatial resolution residential energy model based on American Time Use Survey data and the bootstrap sampling method, Energy Build. 43 (2011) 3528–3538. doi:10.1016/j.enbuild.2011.09.020.

[24] Eurostat, Time use survey (tus), (2020). https://ec.europa.eu/eurostat/cache/metadata/en/tus_esms.htm (accessed July 11, 2022).

[25] G. Buttitta, W. Turner, D. Finn, Clustering of Household Occupancy Profiles for Archetype Building Models, Energy Procedia. 111 (2017) 161–170. doi:10.1016/j.egypro.2017.03.018.

[26] D. Mitra, Y. Chu, K. Cetin, Cluster analysis of occupancy schedules in residential buildings in the United States, Energy Build. 236 (2021) 110791. doi:10.1016/j.enbuild.2021.110791.

[27] D.M. Koupaei, F. Hashemi, V. Tabard-Fortecoëf, U. Passe, A technique for developing high-resolution residential occupancy schedules for urban energy models, Simul. Ser. 51 (2019) 95–102.

[28] D. Sood, I. Alhindawi, U. Ali, J. Mcgrath, M. Byrne, J.O. Donnell, Development of occupancy-based multi-scale building archetypes, in: REHVA 14th HVAC World Congr., Rotterdam, The Netherlands, 2022: pp. 1–8.

[29] G. Buttitta, D.P. Finn, A high-temporal resolution residential building occupancy model to generate high-temporal resolution heating load profiles of occupancy-integrated archetypes, Energy Build. 206 (2020) 109577. doi:10.1016/j.enbuild.2019.109577.

[30] K. Foteinaki, R. Li, C. Rode, R.K. Andersen, Modelling household electricity load profiles based on Danish time-use survey data, Energy Build. 202 (2019). doi:10.1016/j.enbuild.2019.109355.

[31] V.M. Barthelmes, R. Li, R.K. Andersen, W. Bahnfleth, S.P. Corgnati, C. Rode, Profiling occupant behaviour in Danish dwellings using time use survey data, Energy Build. 177 (2018) 329–340. doi:10.1016/j.enbuild.2018.07.044.

[32] U. Wilke, F. Haldi, J.L. Scartezzini, D. Robinson, A bottom-up stochastic model to predict building occupants' time-dependent activities, Build. Environ. 60 (2013) 254–264. doi:10.1016/j.buildenv.2012.10.021.

[33] Z. Yu, B.C.M. Fung, F. Haghighat, H. Yoshino, E. Morofsky, A systematic procedure to study the influence of occupant behavior on building energy consumption, Energy Build. 43 (2011) 1409–1417. doi:10.1016/j.enbuild.2011.02.002.

[34] M. Ferrando, A. Marchenko, S. Erba, F. Causone, S. Carlucci, Pattern Recognition and Classification for Electrical Energy Use in Residential Buildings, in: Proc. Build. Simul. 2019 16th Conf. IBPSA, Rome, Italy, 2019: pp. 2246–2253. doi:10.26868/25222708.2019.210750.

[35] V.M. Barthelmes, C. Becchio, S.P. Corgnati, Occupant behavior lifestyles in a residential nearly zero energy building: Effect on energy use and thermal comfort, Sci. Technol. Built Environ. 22 (2016) 960–975. doi:10.1080/23744731.2016.1197758.

[36] T.S. Blight, D.A. Coley, Sensitivity analysis of the effect of occupant behaviour on the energy consumption of passive house

dwellings, Energy Build. 66 (2013) 183–192. doi:10.1016/j.enbuild.2013.06.030.

[37] A. Muroni, I. Gaetani, P.J. Hoes, J.L.M. Hensen, Occupant behavior in identical residential buildings: A case study for occupancy profiles extraction and application to building performance simulation, Build. Simul. 12 (2019) 1047–1061. doi:10.1007/s12273-019-0573-x.

[38] J.A. Fonseca, A. Schlueter, Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts, Appl. Energy. 142 (2015) 247–265. doi:10.1016/j.apenergy.2014.12.068.

[39] H. Polinder, M. Schweiker, A. Van de Aa, K. Schakib-Ekbatan, V. Fabi, R. Andersen, N. Morishita, C. Wang, S. Corgnati, P. Heiselberg, D. Yan, B. Olesen, T. Bednar, A. Wagner, Final Report Annex 53 - Occupant behavior and modeling (Separate Document Volume II), (2013) 153. http://www.iea-ebc.org/fileadmin/user_upload/images/Pictures/EBC_Annex_53_Appendix_Volume_2.pdf.

[40] I. Gaetani, P.J. Hoes, J.L.M. Hensen, Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy, Energy Build. 121 (2016) 188–204. doi:10.1016/j.enbuild.2016.03.038.

[41] S. Gilani, W. O'Brien, H.B. Gunay, Simulating occupants' impact on building energy performance at different spatial scales, Build. Environ. 132 (2018) 327–337. doi:10.1016/j.buildenv.2018.01.040.

[42] R. Evins, K. Orehounig, V. Dorer, Variability between domestic buildings: the impact on energy use, J. Build. Perform. Simul. 9 (2016) 162–175. doi:10.1080/19401493.2015.1006526.

[43] MIT Sustainable Design Lab, umi - Urban Modeling Interface, (2013). http://web.mit.edu/sustainabledesignlab/projects/umi/index.html (accessed December 20, 2018).

[44] ASHRAE, ASHRAE Guideline 14-2014: Measurement of Energy, Demand, and Water Savings, ASHRAE Guidel. 14-2014. 4 (2014) 1–150.

[45] T. Hong, Y. Chen, M.A. Piette, X. Luo, CityBES, (2017). https://citybes.lbl.gov/ (accessed December 20, 2018).

[46] J.A. Fonseca, T.-A.A. Nguyen, A. Schlueter, F.F. Marechal, City Energy Analyst (CEA): Integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts, Energy Build. 113 (2016) 202–226. doi:10.1016/j.enbuild.2015.11.055.

[47] C.F. Reinhart, T. Dogan, A.J. Jakubiec, T. Rakha, A. Sang, Umi - an Urban Simulation Environment for Building Energy Use , Daylighting and Walkability, Proc. BS2013. (2013) 476–483. doi:10.1016/bs.pbr.2015.01.004.

[48] C. di Milano, Portale Open Data | Comune di Milano, (2021). https://dati.comune.milano.it/ (accessed August 18, 2021).

[49] European commitee for standardization, ISO 52000-1: 2017 - Energy performance of buildings – Overarching EPB assessment – Part 1: General framework and procedures, (2017).

[50] D.W. Aha, D. Kibler, M.K. Albert, Instance-based learning algorithms, Mach. Learn. 6 (1991) 37–66. doi:10.1007/BF00153759.

[51] S. Yilmaz, J. Chambers, S. Cozza, M.K. Patel, Exploratory study on clustering methods to identify electricity use patterns

in building sector, J. Phys. Conf. Ser. 1343 (2019) 0–6. doi:10.1088/1742-6596/1343/1/012044.

[52] D.L. Davies, D.W. Bouldin, A Cluster Separation Measure, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1 (1979) 224–227. doi:10.1109/TPAMI.1979.4766909.

[53] C. Feng, S. Member, A. Mehmani, J. Zhang, S. Member, Deep Learning-based Real-time Building Occupancy Detection Using AMI Data, 11 (2020) 4490–4501.

[54] W. Kleiminger, C. Beckel, S. Santini, Household occupancy monitoring using electricity meters, Proc. 2015 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput. - UbiComp '15. (2015) 975–986. doi:10.1145/2750858.2807538.

[55] W. Kleiminger, C. Beckel, T. Staake, S. Santini, Occupancy Detection from Electricity Consumption Data, Proc. 5th ACM Work. Embed. Syst. Energy-Efficient Build. - BuildSys'13. (2013) 1–8. doi:10.1145/2528282.2528295.

[56] A. Allik, S. Muiste, H. Pihlap, Smart Meter Data Analytics for Occupancy Detection of Buildings with Renewable Energy Generation, (2020) 248–251.

[57] McNeel, Grasshopper - algorithmic modeling for Rhino, (n.d.). https://www.grasshopper3d.com/ (accessed December 31, 2018).

[58] McNeel, Rhino 6 per Windows, (n.d.). https://www.rhino3d.com/ (accessed December 31, 2018).

[59] L. Carnieletto, M. Ferrando, L. Teso, K. Sun, W. Zhang, F. Causone, P. Romagnoni, A. Zarrella, T. Hong, L. Carnialetto, M. Ferrando, L. Teso, K. Sun, W. Zhang, F. Causone, P. Romagnoni, A. Zarrella, T. Hong, L. Carnieletto, M. Ferrando, L. Teso, K. Sun, W. Zhang, F. Causone, P. Romagnoni, A. Zarrella, T. Hong, Italian prototype building models for urban scale building performance simulation, Build. Environ. 192 (2021) 107590. doi:https://doi.org/10.1016/j.buildenv.2021.107590.

[60] European Committee for Standardization-CEN, BS EN 16798 - Energy performance of buildings - Ventilation for buildings - Part 1: Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acousti, Brussels, Belgium, 2019.

[61] International Standard Organisation - ISO, ISO 18523-1:2016 Energy performance of buildings — Schedule and condition of building, zone and space usage for energy calculation, 2016.

[62] European Committee for Standardization-CEN, BS EN 15193: Energy performance of buildings — Energy requirements for lighting - Part 2: Explanation and justification of EN 15193-1, Module M9, Brussels, Belgium, 2017.

[63] U.S. Department of Energy's (DOE) Building Technologies Office (BTO), EnergyPlus - weather, (2021). https://energyplus.net/weather.

[64] P. Zangheri, R. Armani, M. Pietrobon, L. Pagliano, M. Fernandez Boneta, A. Müller, Heating and cooling energy demand and loads for building types in different countries of the EU, (2014) 86. http://www.entranze.eu/files/downloads/D2_3/Heating_and_cooling_energy_demand_and_loads_for_building_types_in_different_countries_of_the_EU.pdf.

[65] F. Causone, R. Scoccia, M. Pelle, P. Colombo, M. Motta, S. Ferroni, Neighborhood energy modeling and monitoring: A

case study, Energies. 14 (2021). doi:10.3390/en14123716.

Appendix

Table A.1: Archetypes for multi-family buildings adopted in this study

Archetypes for multi-family residential buildings	Construction Element	Envelope Description	Thermal Transmittance [W/(m2 K)]	Window Description	Window Thermal Transmittance [W/(m2 K)]	System (SH = space heating)	Cooling	Centralized/ Independent
Traditional built before 1930	Roof	Wooden roof/No insulation	2.50	Single glass, metal frame without thermal break	5.70	Traditional gas boiler for SH and DHW	No	Independent for each unit
	External wall	Stone masonry	2.58					
	External floor	Concrete ground slab	1.88					
	Internal floor/Ceiling	Beams-wooden slab	1.22					
Traditional built between 1961 and 1970	Roof	Reinforced brick-concrete slab, traditional screed	1.45	Single glass, wood frame	4.90	Traditional gas boiler for SH and DHW	No	Independent for each unit
	External wall	Hollow wall brick masonry	0.98					
	External floor	Floor with reinforced concrete slab, traditional screed	1.88					
	Internal floor/Ceiling	Brick-concrete slab - lightweight screed	1.51					
Prefabricated built between 1981 and 1990	Roof	Reinforced brick-concrete slab	0.84	Double glass, air-filled, metal frame without thermal break	3.70	Traditional gas boiler for SH and DHW	No	Centralized for the whole apartment block
	External wall	Precast Reinforced concrete wall, low insulation	0.70					
	External floor	Reinforced concrete slab, lightweight screed	1.88					
	Internal floor/Ceiling	Brick-concrete slab - lightweight screed	1.54					
Prefabricated built between 2000 and 2005	Roof	Reinforced brick-concrete slab, insulated	0.52	Double glass, air-filled, metal frame with thermal break	3.40	Traditional gas boiler for SH/DHW with an electric boiler	Yes	Centralized SH/ Independent DHW
	External wall	Precast reinforced-concrete wall, low insulation	0.47					
	External floor	Reinforced concrete slab, lightweight screed, insulated	0.85					
	Internal floor/Ceiling	Brick-concrete slab, lightweight screed, insulated	0.85					
Traditional built between 2005 and 2010	Roof	Reinforced brick-concrete slab, insulated	0.33	Low-e double glass, air or other gas-	2.20	Condensing gas boiler for	Yes	Independent for each unit

	External wall	Perforated bricks and medium insulated	0.30	filled, wood frame		SH and DHW		
	External floor	Reinforced concrete slab, lightweight screed, insulated	0.33					
	Internal floor/Ceiling	Brick-concrete slab, lightweight screed, insulated	0.54					
Traditional built after 2010	Roof	Reinforced brick-concrete slab, insulated	0.33	Low-e double glass, air or other gas-filled, wood frame	1.80	Condensing gas boiler for SH and DHW	Yes	Independent for each unit
	External wall	Brick blocks and external high insulated	0.23					
	External floor	Reinforced concrete slab, lightweight screed, insulated	0.33					
	Internal floor/Ceiling	Brick-concrete slab, lightweight screed, insulated	0.34					

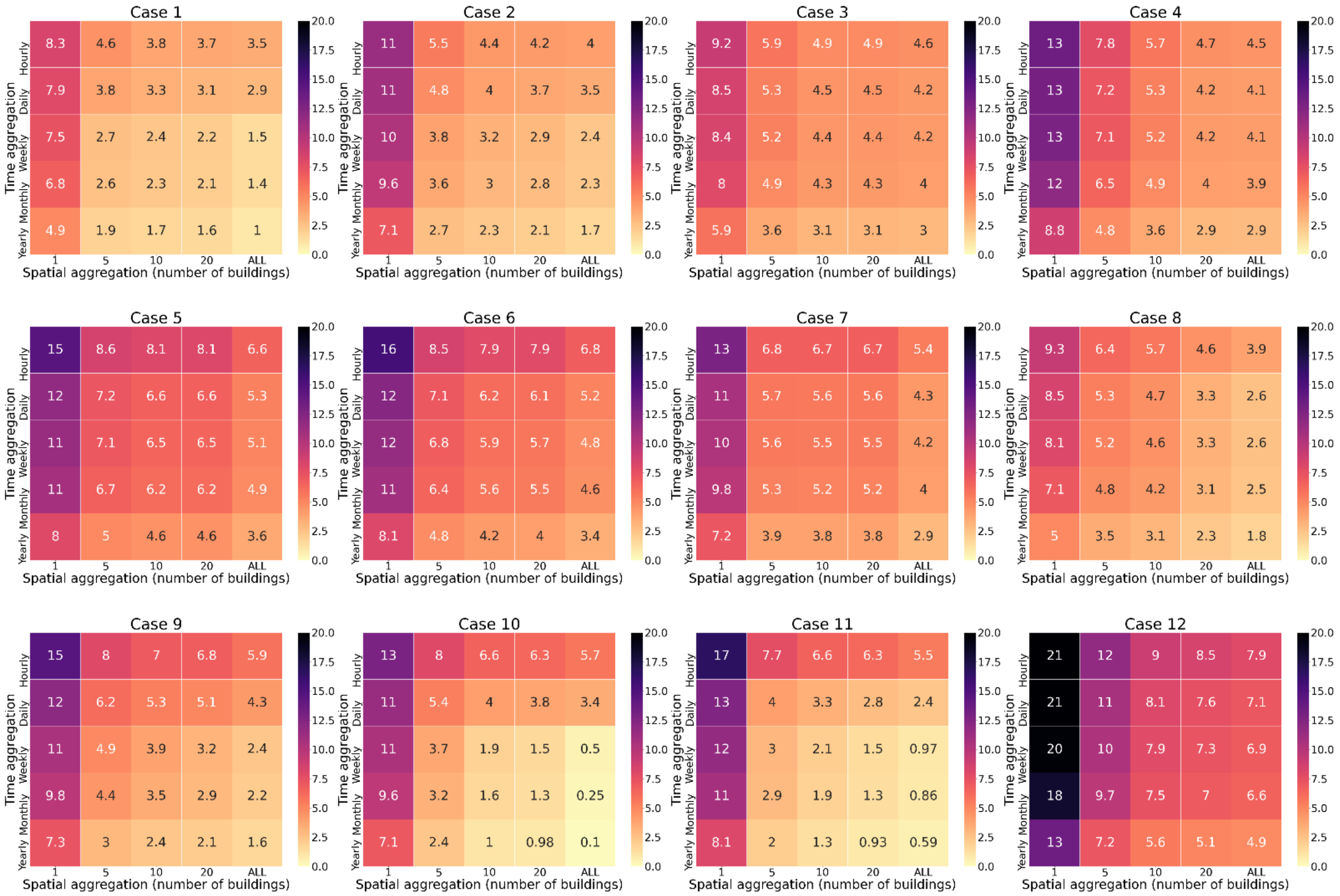


Figure A.1: CVRMSE values related to the heating energy needs difference between Case 0 for Cases from 1 to 12, with the different spatial (horizontally) and temporal (vertically) aggregations.

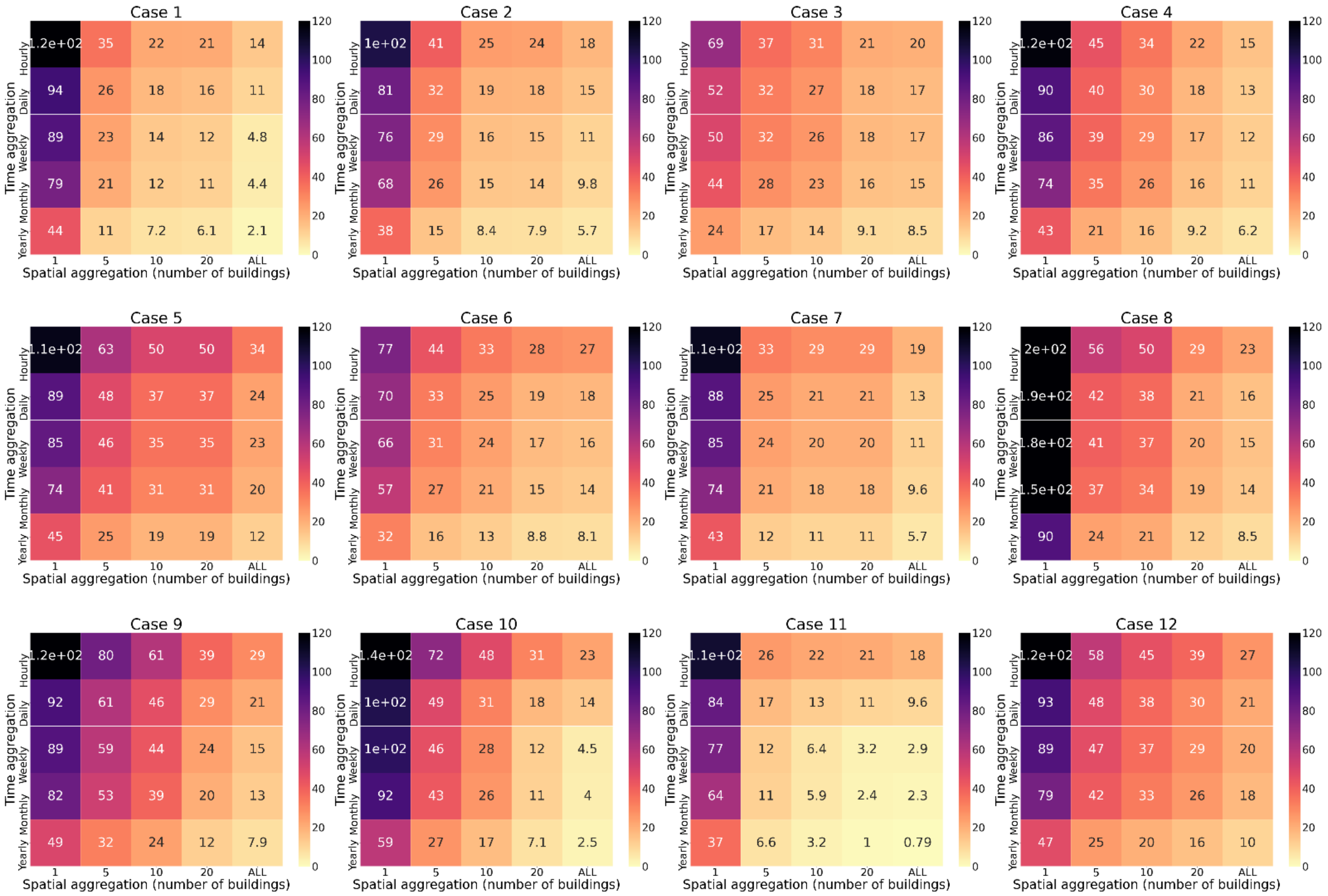


Figure A.2: CVRMSE values related to the cooling energy needs difference between Case 0 for Cases from 1 to 12, with the different spatial (horizontally) and temporal (vertically) aggregations.

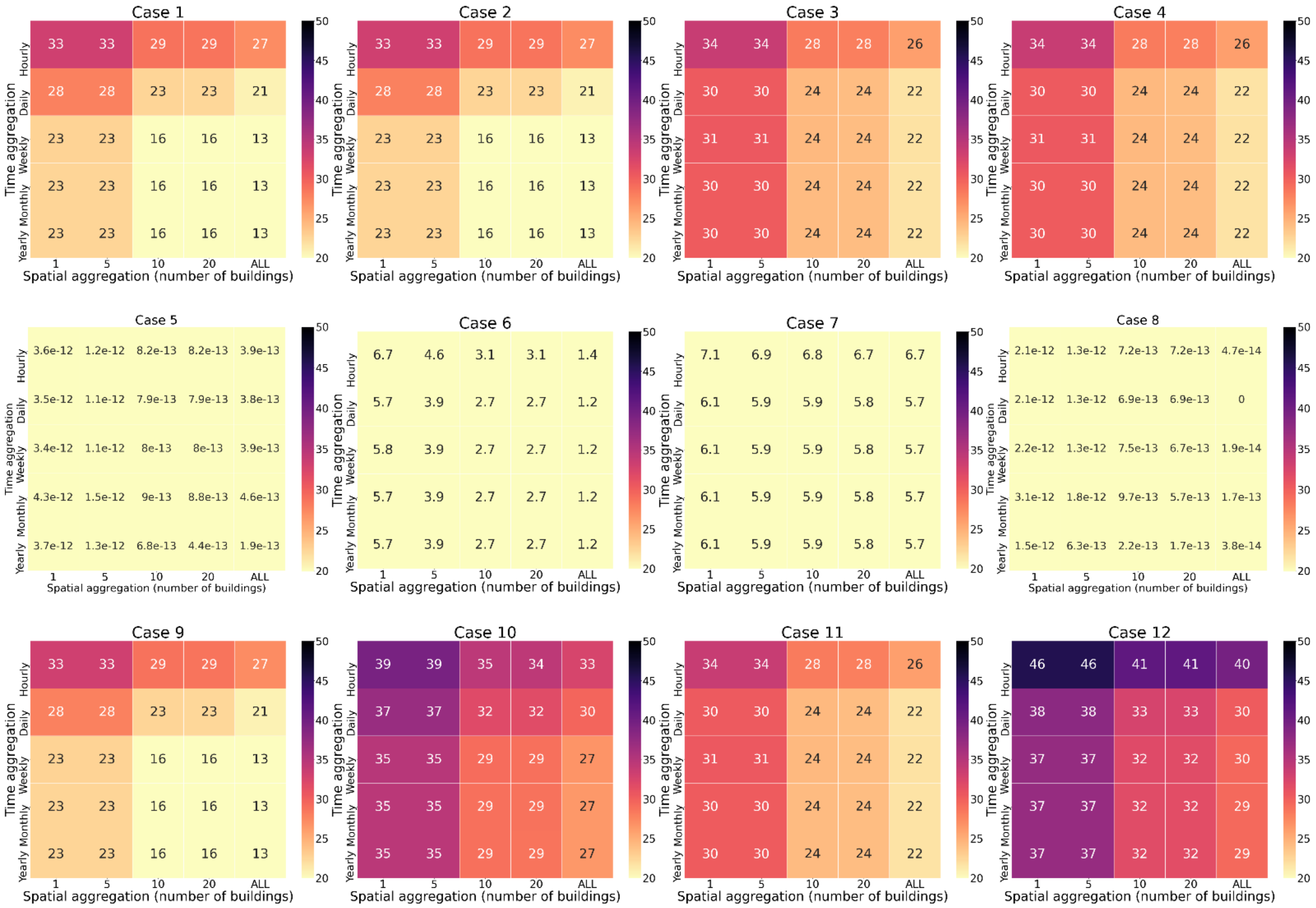


Figure A.3: CVRMSE values related to the electricity energy uses difference between Case 0 for Cases from 1 to 12, with the different spatial (horizontally) and temporal (vertically) aggregations.