

# Hourly energy profile determination technique from monthly energy bills

Mario Lamagna<sup>1</sup>, Benedetto Nastasi<sup>2</sup> (✉), Daniele Groppi<sup>1</sup>, Meysam Majidi Nezhad<sup>1</sup>, Davide Astiaso Garcia<sup>2</sup>

1. Department of Astronautical, Electrical and Energy Engineering (DIAEE), Sapienza University of Rome, Via Eudossiana, 18 - 00184 Rome, Italy

2. Department of Planning, Design, Technology of Architecture, Sapienza University of Rome, Via Flaminia 72 – 00196 Rome, Italy

## Abstract

Hourly energy consumption profiles are of primary interest for measures to apply to the dynamics of the energy system. Indeed, during the planning phase, the required data availability and their quality is essential for a successful scenarios' projection. As a matter of fact, the resolution of available data is not the requested one, especially in the field of their hourly distribution when the objective function is the production-demand matching for effective renewables integration. To fill this gap, there are several data analysis techniques but most of them require strong statistical skills and proper size of the original database. Referring to the built environment data, the monthly energy bills are the most common and easy to find source of data. This is why the authors in this paper propose, test and validate an expeditious mathematical method to extract the building energy demand on an hourly basis. A benchmark hourly profile is considered for a specific type of building, in this case an office one. The benchmark profile is used to normalize the consumption extracted from the 3 tariffs the bill is divided into, accounting for weekdays, Saturdays and Sundays. The calibration is carried out together with a sensitivity analysis of on-site solar electricity production. The method gives a predicted result with an average 25% MAPE and a 32% cvRMSE during one year of hourly profile reconstruction when compared with the measured data given by the Distributor System Operator (DSO).

## Keywords

building load profile, building energy demand model, simplified hourly distribution, simulated hourly demand

## Article History

Received: 19 May 2020

Revised: 08 July 2020

Accepted: 24 July 2020

© Author(s) 2020

## 1 Introduction

Building sector is highly relevant for energy transition since it is responsible for 39% of CO<sub>2</sub> gas emissions and 36% of global final energy use (IEA 2018). It is noteworthy that the adoption of innovative solutions can be strongly affected by the behaviour of building users as tested by Fink (2011), and the time distribution of energy consumption is a crucial aspect to analyze in the view of smart solutions and flexibility enablers as verified by Mancini and Nastasi (2019). Indeed, understanding building energy consumption is the first step for its energy retrofitting (Tian 2013) which will lead to a subsequent performance improvement with a relative reduction of costs and emissions. The claim for just increasing the renewable energy supply cannot alone support the transition toward decarbonization scenarios especially for their non-programmability nature (Loorbach 2010) coupled

with the performance gap between design and operation of buildings (Manfren and Nastasi 2020). Smart energy approach is, then, required to link different production and consumption nodes (Rosenbloom and Meadowcroft 2014) keeping in mind that automation and communication in models and reality is fundamental for its success (Tronchin et al. 2018). Energy Management Systems (EMS) are keen in it, composed of hardware and software for optimal control and rational use of energy (Li et al. 2019) enabling strategy as demand-response. For the aforementioned systems, availability and quality of the information are essentials for their effective outcome (Erdinc and Uzunoglu 2011). Moreover, even in the design phase for any intervention to improve the building performance such as lowering energy consumption or including it in a micro-grid, the data on hourly resolution are of primary interest. There are two possibilities to obtain these data at the needed time-step:

E-mail: benedetto.nastasi@outlook.com

(i) acquiring it or (ii) simulating it if absent. It entails instrumental and mathematical approaches, respectively.

The first approach can be referred to as “instrumental”, and it is based on the use of smart meters allowing to directly measure the consumptions with different time resolutions. The smart meters are being widely installed in buildings, as mentioned by Birt et al. (2011), for time-of-use pricing, meter reading and the identification of outages among all. Yet, the installation has to have taken place before the required analysis and to last for a sufficient amount of time, generally more than one year. However, data reliability is not guaranteed due to the errors occurring in metering. To prevent this, data filtering is necessary as suggested by Ford et al. (2014), before proceeding with their analysis. As mentioned, sufficient large time-series are challenging to have unless the process slows down. By the way, privacy issues occur when data from private building are recorded as mentioned by Finster and Baumgart (2014). As a matter of fact, further method such as short term load forecasting can benefit from the installation of sensors promoting the transition to a manageable Smart Cities (Massana et al. 2017).

In the absence of installed smart meters, an already existing database of other buildings’ data could be used to carry out the analysis. Here, the mathematical approach can be used that relies on different statistical and probabilistic methods as summarized by Grandjean et al. (2012), where they propose a cross analysis of existing methods capable of building up a residential electric load curve from available dataset. Those techniques for modelling energy consumptions can be separated into two categories: top-down and bottom-up.

Top-down models start from aggregated data related to building stock or portfolio and call for inserting certain variables and constraints to identify the energy consumption of a single building. Bottom-up models allow estimating the energy consumption of a single or small group of buildings

starting from the addition of single units from components like appliances to buildings in a district. The opportunity is related to the position of data inputs, according to Swan and Ugursal (2009). Finally, data can be mined for the required resolution from lower detail or extracted for a single or typical user from a group, clustered earlier. As concisely explained by Ramos et al. (2015), data mining regards detect patterns in large data sets using the techniques introduced above, while clustering is the process of subdividing a dataset into clusters based on similarity or proximity among data. In this wide range of techniques, a strong statistical background is requested as well as a proper size of the database and amount of recorded parameters. Even neural networks to forecast energy demand as in Ford et al. (2014), Dudek (2016), Biswas et al. (2016) are used, while other studies propose to implement Fourier analysis on data as performed in Ji et al. (2015), Niu et al. (2018). Here, the Fourier series periodic pattern is used to reconstruct a variation in hourly energy use. Actually, there are also simplified models which resort to cluster dataset and external information such as temperature (Kipping and Trømborg 2016) and/or degree days (Eto 1988; De Rosa et al. 2014). In addition, benchmarks models and utility bills can be used as proposed by Fumo et al. (2010) and further examined by Smith et al. (2011). The fundamental of Fumo et al. (2010) method is the use of fixed benchmark models accessible from EnergyPlus libraries to generate normalized energy profiles, used to make hourly energy consumption from monthly ones reported in the utility bills. The models are selected to describe energy buildings performance with similar characteristics in the U.S. climate zones (DOE 2015). Table 1 summarizes those contributions.

This paper proposes a similar approach but, with a different benchmark for normalization. A piecewise function, described in three non-connected intervals based on the Italian tariff subdivision, is built to distribute and normalise

**Table 1** Literature review

Source	Input	Output	Case study	Smart meter
Kipping and Trømborg (2016)	External temperature; survey; electric hourly data	Hourly energy consumption for heating and cooling	Real case	Yes
Eto (1988)	Degree day	Monthly energy consumption for heating and cooling	Building archetype	No
De rosa et al. (2014)	Degree day; external temperature; wind intensity and direction; normal direct radiation and diffused horizontal radiation; altitude; latitude	Hourly energy consumption for heating and cooling	Building archetype	No
Fumo et al. (2010)	EnergyPlus benchmark; utility bills	Hourly energy consumption divided in base load and variable consumption	Building archetype	No
Smith et al. (2011)	EnergyPlus benchmark; utility bills	Hourly energy consumption divided in base load and variable consumption	Building archetype	No

the monthly data obtained by utility bills on an hourly load curve. The function represents the typical trend of energy consumption of a building with a specific end use similar to the case study. Therefore, in the absence of recorded hourly energy consumption data, the proposed method provides expeditious results for designing strategies and scenarios that call for this data resolution.

The three intervals of electricity tariffs refer to a time subdivision for peak, mid-level, and off-peak hours during the week. A calibration based on the peculiarity of the case study (details already available such as opening hours) is carried out. An error analysis is also performed for the first testing and for calibrating the model.

## 2 Material and methods

### 2.1 Italian billing system

From January first 2007, the metering consumption interval is regulated according to Delibera n.181/06 (GU 2006) which was anticipated by Delibera n.19/06 (ARERA 2006), dividing the electric billing measurement in three period called (i) F1 representing the peak period, where the highest consumptions are recorded during the day, (ii) F3 representing the off-peak period, mostly the base load consumption, and (iii) F2 representing the mid-level period, which is the transition interval between peak and off-peak consumption intensities. The subdivisions (Fs) are motivated by a careful trend observation in electricity exchange prices, deeply analysed on the Italian Electricity Market Operator website (GME 2020), which have a very distinguished trend between day and night, and between weekdays (WDs), holidays and weekends. F1 is identified in the time interval going from 8:00 to 19:00 during WDs. F2 is identified from 7:00 to 8:00 and from 19:00 to 23:00 during the WDs and from 7:00 to 23:00 on Saturdays (STs). Ultimately, F3 is recorded from 23:00 to 7:00 during WDs and STs and all day long during Sundays and holidays (both simplified with the acronym SN), that information is schematically summarised in Table 2. The colours used in the table, will be applied always in the same way to guide the reader in easier understanding, accordingly the F1 is exemplified by red, the F2 by yellow and the F3 by green.

As it can be noticed in Table 2, only F1 is recorded always in the same time interval while F2 and F3 change according to the day. Furthermore, F3 is recorded all day long on every holiday day. Accounting for national holidays as established by ARERA, which is the Italian authority in charge (ARERA 2007), it is possible to report precisely, and in advance, the amount of hours during F1, F2 and F3 for each month of the year 2018 as reported in Table 3.

**Table 2** Description of Fs subdivision along a week

	WD					ST	SN
	Mon	Tue	Wen	Thu	Fry	Sat	Sun/Hol
00:01	F3	F3	F3	F3	F3	F3	F3
01:01	F3	F3	F3	F3	F3	F3	F3
02:01	F3	F3	F3	F3	F3	F3	F3
03:01	F3	F3	F3	F3	F3	F3	F3
04:01	F3	F3	F3	F3	F3	F3	F3
05:01	F3	F3	F3	F3	F3	F3	F3
06:01	F3	F3	F3	F3	F3	F3	F3
07:01	F2	F2	F2	F2	F2	F2	F3
08:01	F1	F1	F1	F1	F1	F2	F3
09:01	F1	F1	F1	F1	F1	F2	F3
10:01	F1	F1	F1	F1	F1	F2	F3
11:01	F1	F1	F1	F1	F1	F2	F3
12:01	F1	F1	F1	F1	F1	F2	F3
13:01	F1	F1	F1	F1	F1	F2	F3
14:01	F1	F1	F1	F1	F1	F2	F3
15:01	F1	F1	F1	F1	F1	F2	F3
16:01	F1	F1	F1	F1	F1	F2	F3
17:01	F1	F1	F1	F1	F1	F2	F3
18:01	F1	F1	F1	F1	F1	F2	F3
19:01	F2	F2	F2	F2	F2	F2	F3
20:01	F2	F2	F2	F2	F2	F2	F3
21:01	F2	F2	F2	F2	F2	F2	F3
22:01	F2	F2	F2	F2	F2	F2	F3
23:01	F3	F3	F3	F3	F3	F3	F3

**Table 3** Summary of each Fs duration month by month for 2018

2018	$t_{F1}$ [h]	$t_{F2}$ [h]	$t_{F3}$ [h]	$t_{Tot}$ [h]
January	242	183	319	744
February	220	188	264	672
March	242	217	285	744
April	209	182	329	720
May	242	200	302	744
June	231	194	295	720
July	242	200	302	744
August	242	200	302	744
September	220	205	295	720
October	253	206	285	744
November	231	194	295	720
December	209	182	353	744
Mean	232	196	302	730

### 2.2 The building: Procida City Hall

Thanks to a strict collaboration with Procida municipality, it was possible to gather all the utility bills for the local City Hall offices that are summarised in Table 4 for 2018 according

to the Italian price time-slots previously introduced. The building is not connected to the gas grid, thus both thermal and electric consumptions are covered by electricity. The building serves as an office for the City Hall and for the Municipal police employees. Occasionally, the meeting room is conceded in lease for local activities. The building has a 160 m<sup>2</sup> PV plant installed on the roof, for a total of 100 panels, 20 kW peak power production and 9 inverters. The City Hall has a total useful surface of 1740 m<sup>2</sup> divided in basement, first floor, second floor and roof. It has 9 unused rooms in the basement, 19 offices and 2 bathrooms on the ground floor, 9 offices and 1 bathroom on the first floor. All the offices (i.e. 28) have a computer workstation. The building hosts almost 40 employees, with a variable timetable according to different activities carried out in the building. Regarding building's consumptions, the air conditioning system consists of a heat pump and various systems split conditioning. The heat pump is currently not working. Due to this situation, the cooling system relies on indoor units located in each office, for a total of 28 units, which can be controlled and regulated by the employees. Similar conditions are recorded for the heating system, whose time usage limit

in Italy is regulated by (GU 1993), relying on single electric resistance heaters, with variable power, used at will of the employees.

Table 4 reports the monthly consumptions for the studied building, it can be appreciated a consumption reduction from April due to PV production especially during F1 hours and the end of the heating season. The consumption then tends to slightly increase during warmer months due to the use of cooling systems. The values recorded in the other two Fs remain constants along the year, since they are computed almost completely during night-time, especially F3, thus not being affected by the PV production and outside of the working time schedule hence not being affected by heating and cooling loads. There is not a meter connected to the PV plant, which could gauge the real balance between self-consumption and electricity sent to the power grid. Thus, the online free software PVWATT Calculator, released by the US National Renewable Energy Laboratory (NREL), has been used to evaluate the energy production. As regards the excess electricity production, delivered to the power grid, the official data has been obtained from the Distributor System Operator (DSO). In order to validate the proposed method, the hourly measures data have been gathered from the DSO. To obtain the real consumption data the energy supplied by the grid had to be summed to the PV production and the energy delivered to the grid had to be subtracted.

**Table 4** Breakdown of City Hall utility bills consumptions

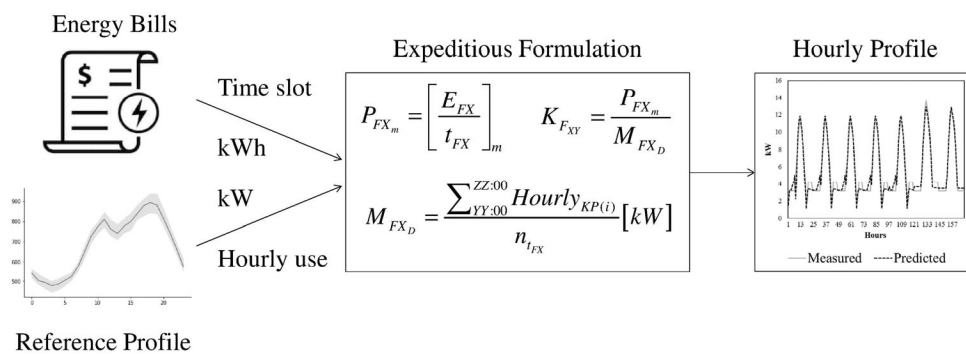
2018	$E_{F1}$ [kWh]	$E_{F2}$ [kWh]	$E_{F3}$ [kWh]
January	3,271.70	995.2	1,556.3
February	4,087.80	1188.8	1,447.4
March	2,376.2	1161.2	1,408.6
April	359.3	593	1,013.3
May	387.5	622.7	1,011.8
June	269.7	560	1,002.4
July	468.8	767.7	1,215.4
August	513.3	837.3	1,393.9
September	680.2	716.2	1,106.9
October	686.1	813.2	1,045.6
November	1,359.60	919.3	1,239.2
December	2,741.10	1,218.30	1,984.30
Total	12,902.10	7,063.90	11,022.60

### 2.3 Method description

The paper objective is to present an easily replicable procedure to obtain an hourly profile consumption from monthly bills. The entire flow is depicted in Figure 1.

Once collected the required monthly values, the first step is to evaluate the power consumption recorded during the month ( $m$ ) in each time interval  $t_F$  identified in Table 3 for 2018 as shown in Eqs. (1), (2) and (3).

$$P_{F1_m} = \left[ \frac{E_{F1}}{t_{F1}} \right]_m \tag{1}$$



**Fig. 1** Flow of the proposed method

$$P_{F2,m} = \left[ \frac{E_{F2}}{t_{F2}} \right]_m \tag{2}$$

$$P_{F3,m} = \left[ \frac{E_{F3}}{t_{F3}} \right]_m \tag{3}$$

With the computed values, a preliminary hourly load curve is obtainable shown in Figure 2 for three typical days (from Friday to Sunday), where the colours still indicate the same F time interval as in the previous tables. The values coming from Eqs. (1), (2) and (3) fix the height of the respective rectangles, and the total area under those lines represent the energy consumption, while their bases are defined by the time length according to Table 2 details.

The obtained profile shows in sequence a WD, in this case it can be assumed a Friday for convention, a ST and a SN, which can be assumed to be a Sunday. In doing so, the profile shows only three hourly consumption levels according to the 3 intervals, which remain constant in the same F interval independently from the day or the hour. Because of this, the consumptions recorded in the F2 interval are the same during Friday morning or Saturday afternoon.

Although mathematically correct, the predicted value behind this consumption could be distant from the real measure. Similarly, the same problem appears for the F3 interval, showing the same consumption during nights and days. Those situations must be studied to weight correctly the consumptions for different days but with same Fs, to avoid unrealistic consumption in specific hours. As regards F1, being recorded only during the WDs and always at the same time slots, it is not affected by this kind of misrepresentation. Starting from this latter consideration, the attention now will be focused on the WDs profile, and especially on the F1 interval. From literature, it was possible to select a typical WD load curve presented by Luo et al. (2017), which will be used as a known profile (KP) during the normalization for all the studied case (SC) monthly consumptions.

Figure 3 depicts the values obtained for the SC obtained by using Eqs. (1), (2) and (3) for a WD. Just for graphic purposes, the KP hourly consumption curve (in cyan) is shown after being scaled down to the size of the values obtained for the SC.

The mean consumption value recorded during F1 for

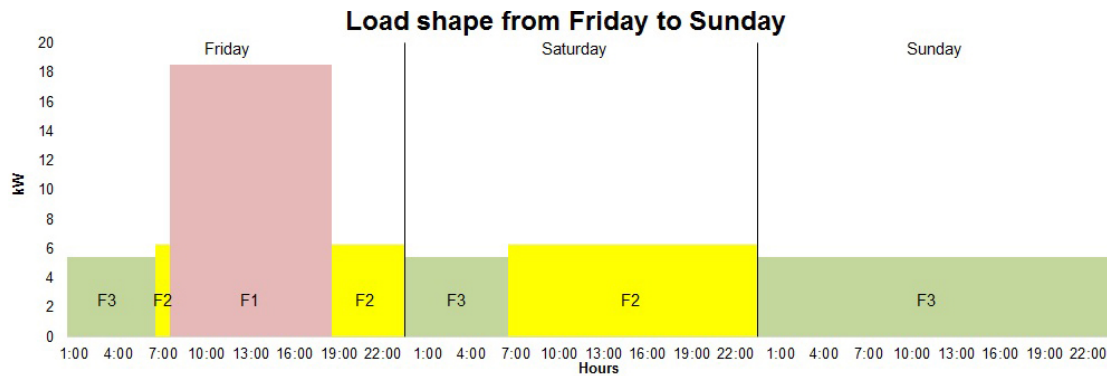


Fig. 2 Hourly profile obtained by Eqs. (1), (2) and (3) representing a weekend

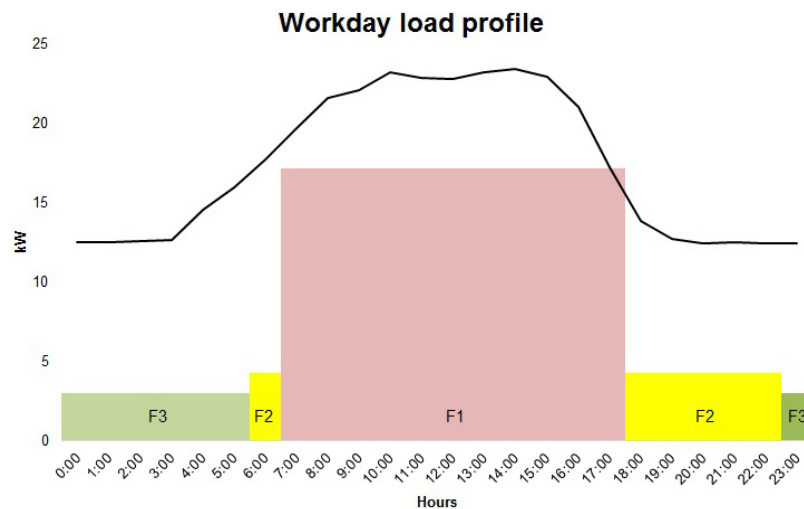


Fig. 3 Load curve comparison between known profile and study case during a WD



the KP ( $M_{F1_{WD}}$ ), and is evaluated according to Eq. (4) while its base is equal to the SC's base since both consumptions are evaluated in the same interval.

$$M_{F1_{WD}} = \frac{\sum_{8:00}^{18:00} \text{Hourly}_{KP(i)}}{n_{t_{F1}}} \quad [\text{kW}] \quad (4)$$

In Eq. (4), the  $n_{t_{F1}}$  is the number of hours elapsed in F1 during a WD (11 hours), while  $\text{Hourly}_{KP}$  is the hourly consumption recorded during that interval for the KP.

Both  $M_{F1_{WD}}$  and  $P_{F1}$  (evaluated as per Eq. (1)) define the height of rectangles with the same bases. Since there is a well-defined proportionality between hourly and mean value in the KP, the next step is to transfer the same proportionality from the KP to the SC obtaining an hourly consumption variation with the shape of the verified behaviour. The area under the two curves represents the energy request in the F1 interval for both cases, and their ratio will be used as a normalization coefficient called  $K$ , and its formula for the F1 is shown in Eq. (5).

$$K_{F1_m} = \frac{P_{F1_m}}{M_{F1_{WD}}} \quad (5)$$

As told, F1 is recorded only during WDs; hence, there is no need to provide other normalization factors because it is sure that the consumption documented in the first column in Table 4 is all measured in a time interval with the same length. By means of  $K_{F1}$ , which embodies the mathematical operation to match the two different areas, the normalization is performed as indicated in Eq. (6):

$$\text{Hourly}_{SC(i)} = K_{F1_m} \times \text{Hourly}_{KP(i)} \quad [\text{kW}] \quad (6)$$

Regarding F2 and F3, it is necessary to make further assumptions before applying the normalization process. As aforementioned, the bases of the areas under the F2 and F3 curves in previous figures change according to the day of the week. In accordance with that, the  $K$  regarding F2 and F3 must be differentiated to consider different days (i.e. WD, STs and SNs). Starting from F2, it is not measured during SN thus no  $K_{SN}$  is needed. Given that the F2 is an interval where medium intensity loads are reported, it is possible to assume that the consumption values will not change much between a WD and a ST. Indeed, ST is not a canonical working day but nonetheless some activities occasionally happen to be done on STs either way. Anyway, it is not expected a peak-load intensity load even during the central hours of the day. Thus, a homogeneous consumption level can be assumed with no need to split the F2 consumptions according to the day of the week, and graphically we can confirm the rectangle regarding F2 consumptions have the same heights, represented by Eq. (2), or the same  $K$  numerator

value, but different base values as considered in Eqs. (7) and (9) with their associated  $K$ :

$$M_{F2_{WD}} = \frac{\text{Hourly}_{KP_{7:00}} + \sum_{19:00}^{22:00} \text{Hourly}_{KP(i)}}{n_{t_{F2}}} \quad [\text{kW}] \quad (7)$$

$$K_{F2_{WD_m}} = \frac{P_{F2_m}}{M_{F2_{WD}}} \quad (8)$$

$$M_{F2_{ST}} = \frac{\sum_{7:00}^{22:00} \text{Hourly}_{KP(i)}}{n_{t_{F2}}} \quad [\text{kW}] \quad (9)$$

$$K_{F2_{ST_m}} = \frac{P_{F2_m}}{M_{F2_{ST}}} \quad (10)$$

Similarly, it is possible to make assumptions on F3 consumption distributions starting from practical considerations. During WDs and STs, F3 interval is the same as per Eqs. (13), (14), while it changes for SNs. In this latter situation, the consumptions are measured during the day, but pondering the fact that during SNs the office is vacant, similarly as during night-time, it is possible to assume the consumption is homogeneously distributed during the entire F3 interval. This statement implies that the value evaluated with Eq. (3) can be assumed, in the three F3 interval, as the height of the respective rectangles with same basis during WDs and STs and a larger one in SNs as interpreted by Eqs. (11), (13) and (15):

$$M_{F3_{WD}} = \frac{\sum_{23:00}^{7:00} \text{Hourly}_{KP(i)}}{n_{t_{F3}}} \quad [\text{kW}] \quad (11)$$

$$K_{F3_{WD_m}} = \frac{P_{F3_m}}{M_{F3_{WD}}} \quad (12)$$

$$M_{F3_{ST}} = M_{F3_{WD}} \quad [\text{kW}] \quad (13)$$

$$K_{F3_{ST_m}} = K_{F3_{WD_m}} \quad (14)$$

$$M_{F3_{SN}} = \frac{\sum_{00:00}^{23:00} \text{Hourly}_{KP(i)}}{n_{t_{F3}}} \quad [\text{kW}] \quad (15)$$

$$K_{F3_{SN_m}} = \frac{P_{F3_m}}{M_{F3_{SN}}} \quad (16)$$

Those assumptions are strongly related to the specific end uses of the investigated building. Thus, the assumptions that have been made for this case study could be extended to schools, universities or more in general to buildings with a strong utilization rate during the week. Nevertheless, those interpretations will be counterproductive in touristic buildings such as restaurants and hotels which have a consumption in countertendency with respect to the offices, with a marked utilization rate during weekends.

### 2.4 The implemented software

Thanks to the set of equations from Eqs. (1) to (16) it is now possible to construct a piecewise function, described in three non-connected intervals, and obtain a realistic hourly profile with the only input information given by utility bills, as shown in Table 4, and a KP daily hourly profile. A hierarchical code was implemented in MATLAB to gage the piecewise function described in the different intervals for different days in the previous section. All the information reported in Table 2 and Table 3 are already executed as matrices. Consequently, the upload of the needed data, the software will evaluate the related  $K$  depending on the  $F$  and type of day by using Eqs. (5, 8, 10, 12, 14, 16). Then, thanks to recursive matrix constructions, in each defined position the KP hourly value will be multiplied by the corresponding  $K$  as described in Eq. (6) compiling the entire year. At the end of this operation, as a result, the hourly profile for each month is obtained.

Due to the PV presence, a subsequent post treatment sum of delivered and produced energy to the obtained vector is required. Since the utility bills just consider the electricity supplied by the grid, in order to evaluate the real consumption, it is needed to consider the building self-consumption. Thus, the PV production vector will be summed to the obtained profile (i.e. energy extracted from the grid). Likewise, the delivered production to the power grid vector will be subtracted from consumptions. This balance is necessary in order to consider the self-consumption that is hidden from data in the utility bills and which is not directly detectable since a dedicated smart meter is not installed. This step can be avoided if renewable production is not installed in the building.

### 2.5 Error analysis

The following statistical error indicators are used to evaluate the goodness of the hourly consumption curve evaluated with the aforementioned methods against the measured data gathered from the DSO. The comparison has been made between the hourly measured value indicated with  $M_{(i)}$  and the predicted values indicated with  $P_{(i)}$ , obtained with the above explained software. The first indicator assessed is the mean error (ME), namely the arithmetic average of the sum of the committed error (Eq. (17)) in  $n$  intervals; in this case,  $n$  is the number of hours reconstructed:

$$E_{(i)} = M_{(i)} - P_{(i)} \quad [\text{kWh}] \quad (17)$$

$$ME = \frac{\sum_i^n E_{(i)}}{n} \quad [\text{kW}] \quad (18)$$

ME measures the average error deviation from the actual

value; by its nature, therefore, it is subject to a phenomenon of compensation: if a value close to zero is obtained it does not mean that there have been no errors in the forecast, but that those of overestimation and those of underestimation tend to compensate each other. For this reason, the mean absolute deviation (MAD) is analysed as well, which represents the arithmetic mean of the absolute values of the forecast errors committed in  $n$  intervals.

$$MAD = \frac{\sum_i^n |E_{(i)}|}{n} \quad [\text{kW}] \quad (19)$$

If the MAD is greater than zero, the normalized mean bias error (NMBE) is calculated since an overestimation of energy consumption determines a positive NMBE value while an underestimation determines a negative one. The NMBE is evaluated as the negative total sum of the error in the time intervals, divided by the sum of the measured energy consumption as reported in Eq. (20) and it is expressed as a percentage.

$$NMBE = -\frac{\sum_i^n E_{(i)}}{\sum_i^n M_{(i)}} \times 100 \quad (20)$$

Then, the mean absolute percentage error (MAPE) is evaluated, which is the arithmetic average of the ratio between the forecasted absolute error and the measured demand occurred in  $n$  intervals (Eq. (21)). It relates the error (in absolute value) with the measurement, in this way it is possible to ponder the error with its actual relevance.

$$MAPE = \frac{\sum_i^n \frac{E_{(i)}}{M_{(i)}}}{n} \times 100 \quad (21)$$

The coefficient of variation of root mean squared error cvRMSE (Eq. (24)) is the normalized measure of the error. It is based on Root Mean Squared Error RMSE (Eq. (23)), the sample deviation of the differences among values measured and predicted divided by  $A$  (Eq. (22)), which represents the measured average energy consumption. A lower cvRMSE value means a better calibrated model (ASHRAE 2014).

$$A = \frac{\sum_i^n M_{(i)}}{n} \quad [\text{kWh}] \quad (22)$$

$$RMSE = \sqrt{\frac{\sum_i^n E_{(i)}^2}{n}} \quad [\text{kW}] \quad (23)$$

$$cvRMSE = \frac{RMSE}{A} \times 100 \quad (24)$$

The values obtained by Eqs. (20) and (24) must be within a certain threshold to validate the method accuracy as

indicated by Fabrizio and Monetti (2015) and reported in Table 5 for hourly simulations.

Thus, values indicated in Table 5 will be used as reference to evaluate the developed method and its validation.

**Table 5** Different guideline thresholds for calibration

Metric	ASHRAE Guidelines 14	IPMVP	FEMP
NMBE (%)	$\pm 10$	$\pm 5$	$\pm 10$
cvRMSE (%)	30	20	30

### 3 Results and discussion

In this section, the hourly predicted vectors obtained applying the explained method are shown and compared to the data obtained by the DSO by means of a detailed error analysis. After considerations on the results obtained on this first test, additional constraints will be implemented during model calibration.

#### 3.1 Test results

In sequence, the hourly value obtained for February, April, July and October are shown. Those months have been selected since they embody the seasonal consumption variations.

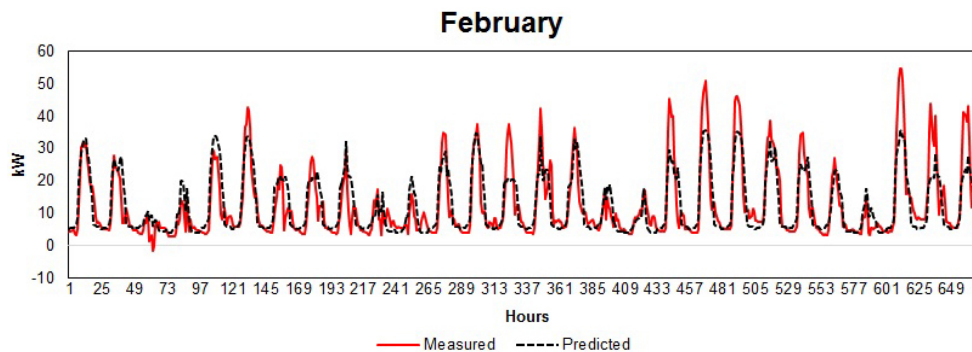
From Figures 4 to 7 and Table 6 that summarize the error analysis, it is possible to appreciate how the method is able to evaluate a monthly overall value which is identical to the measured one (i.e. the ME is almost null), but reading the

MAD and the NMBE it can be understood that there is a compensation error within the month. Similarly, a high MAPE states that predicted values occasionally suffer important divergences from the measured ones. Nevertheless, it must be considered that the measured values oscillate between 1 kWh during nights. Values lower than 1 dramatically increase the MAPE, leading to a conspicuous deviation.

Additionally, it can be noticed as during weekends, there is a countertendency with respect the other days. Indeed, the method stops underestimating the measured consumption and begins overestimating them instead.

The above revealed problem, related with measured value below 1 kWh, results even stronger during the month of April. Indeed, during April, the baseload often oscillates around 0 kWh due to the balance between energy consumption, production and delivered to the power grid. When the delivery value is higher than the supposed production the resultant consumption can also be negative. This is due to the fact that the PV production does not precisely represent the PV production in 2018 but it is built on an average year. Thus, sunny and cloudy days do not coincide with reference year, and thus the PV production does not precisely reflect the actual production. For this reason, despite lower values of MAD and RMSE, a MAPE increase occurs.

In July, a similar behaviour is shown, but this time the MAPE values are lower since the baseload is higher than the one in April. Additionally, also in this month it is possible to perceive the overproduction predicted during the weekends and in the second part of the day.



**Fig. 4** Hourly load curve comparison for February after the first model test simulations

**Table 6** Model test error analysis results

	February	April	July	October	Annual average
ME	-0.00011	-0.09	-0.03	0.00	-0.04
MAD	3.70488	1.89	1.93	0.65	2.04
MAPE	0.435848	0.83	0.40	0.17	66%
RMSE	5.379487	2.52	2.47	0.87	2.87
A	12.26777	5.00	6.93	5.54	7.32
cvRMSE	0.438506	0.50	0.36	0.16	39%
NMBE	9.1E-06	0.01	0.00	0.00	0.34%



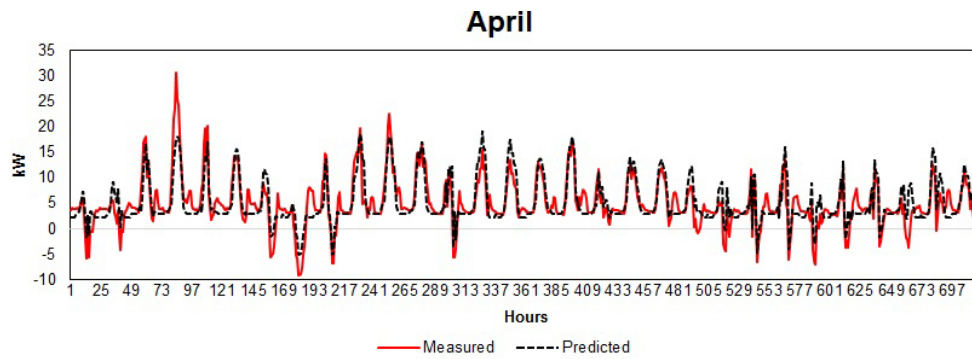


Fig. 5 Hourly load curve comparison for April after the first model test simulations

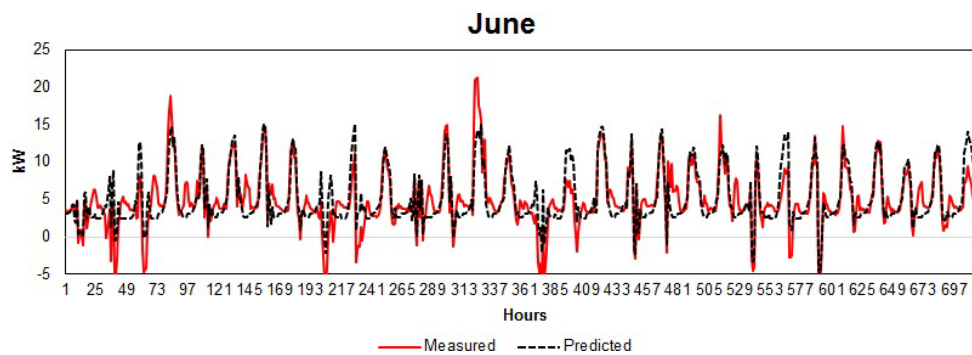


Fig. 6 Hourly load curve comparison for July after the first model test simulations

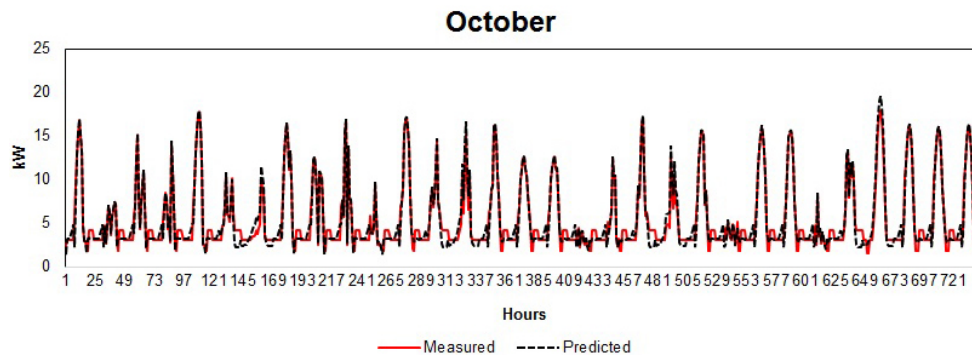


Fig. 7 Hourly load curve comparison for October after the first model test simulations

October is a cold month as February; thus, the base load consumption is higher than warm months due to heating systems and thus the MAPE decreases. Furthermore October, exemplifies the problem related to the data collection. The smart meter installed by the DSO, due to a malfunctioning, from the end of September stopped to record the hourly consumptions and started to provide as output just an hourly mean vector analogous to the results obtainable using Eqs. (1), (2) and (3). It is possible to appreciate as October offers the best scores from an error analysis point of view. This condition in part is due to the high baseload which lowers the divergences but also to the measured data given by the DSO.

At the end of this first test analysis about the direct

method use, different adroitness regarding the realistic building usage will be implemented to calibrate the model.

### 3.2 Calibration

There are some improvements that can be implemented to get closer to measured data based on the specific case study features. During the test, a profile with an average scheduled working timetable from 8:00 to 18:00 was used for the normalization. First, the information regarding the City Hall offices opening time, which represents the majority consumption invoice, will be adopted. The new profile is adapted to an office working from 8:00 to 14:00, the changes are represented in Figure 8.

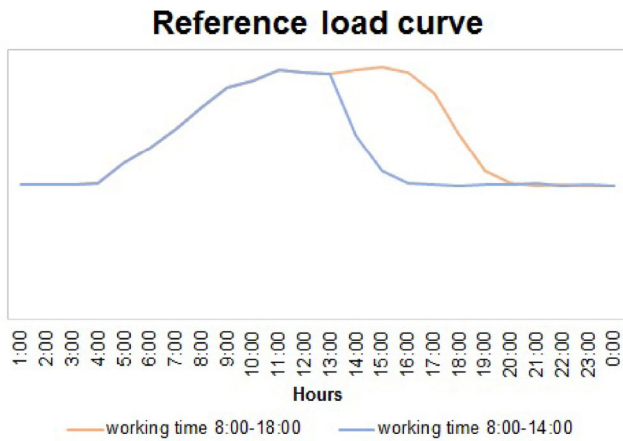


Fig. 8 Different working time load curve for normalization

Since during weekends the building is empty most of the time, it is interesting to use a flat profile instead of an hourly changing load curve. Thus, once computed the first SN and ST hourly values, rather than use their varying profile with hour peak loads, the twenty-four hours will be substituted by the daily mean values.

Third, the energy delivered to the grid, the measured consumption (from the DSO) and the PV production are not consistent with each other (i.e. the energy consumption and delivery are provided by the DSO while the PV production is evaluated by means of PVWatt) thus greatly affecting the

error analysis. Thus, the following two adjustments were made:  
 1) a daily mean profile PV production per month was used so as to avoid the random sequence of sunny and cloudy days and to smooth the PV production profile;  
 2) a sensitive analysis on the PV size is conducted, simulating an installed PV plant peak power variation of  $\pm 25\%$ , thus 16 kW<sub>p</sub> and 24 kW<sub>p</sub> production since the simulated PV production might not reflect the real PV plant.

Fourth, however, discrepancies between the predicted and the measured values are expected due to the piecewise function nature and to the hourly PV production vector non-directly matching the measured data. Finally, it is noteworthy that the expeditious nature of this method could be further strengthened by the addition of more sophisticated analyses as the short-term load forecasting method (Massana et al. 2015). This latter is demonstrated to be very effective for residential buildings due to the expected high variance of the loads (Gerwig 2015). Here, since the case study is a non-residential building, the calibration is made by introducing only the aforementioned changes.

Figures 9 to 12 show the calibrated profiles.

The new graphs return a predicted hourly profile (blue line) which follows with more precision the measured one (orange line). The error analysis is shown in Tables 7, 8 and 9. Except the NMBE values, all the remaining error indicators decrease.

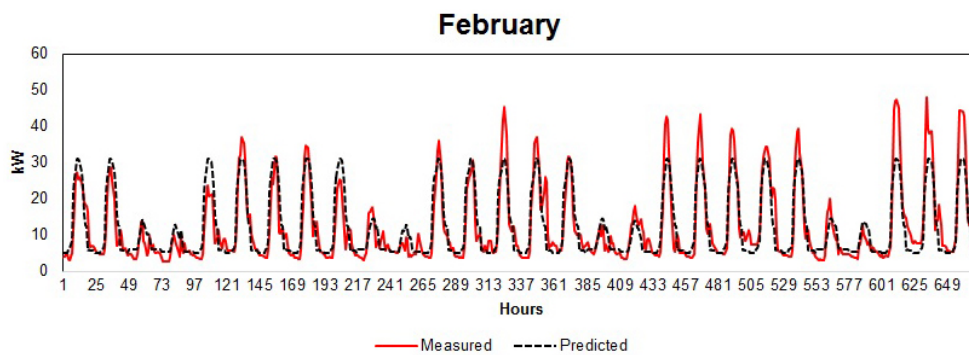


Fig. 9 Hourly load curve comparison for February after the calibration

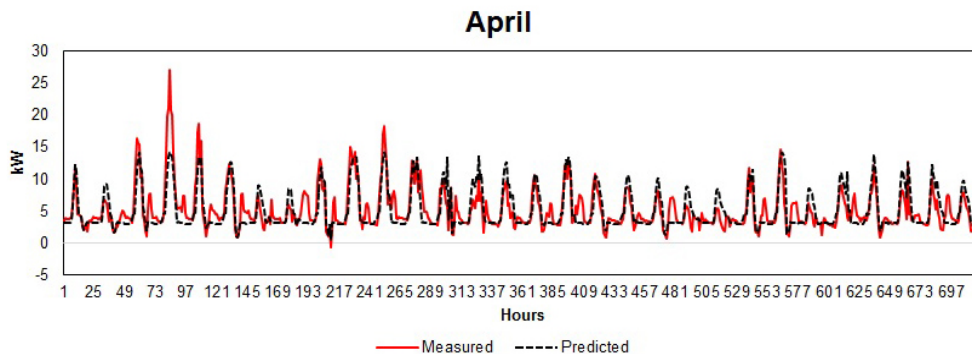


Fig. 10 Hourly load curve comparison for April after the calibration

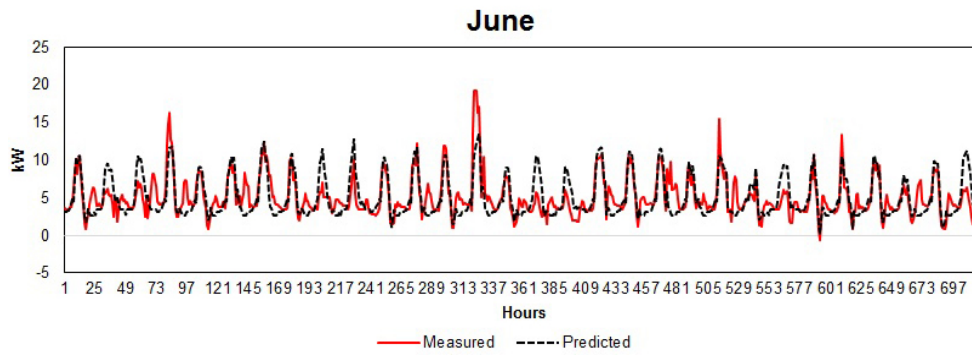


Fig. 11 Hourly load curve comparison for July after the calibration

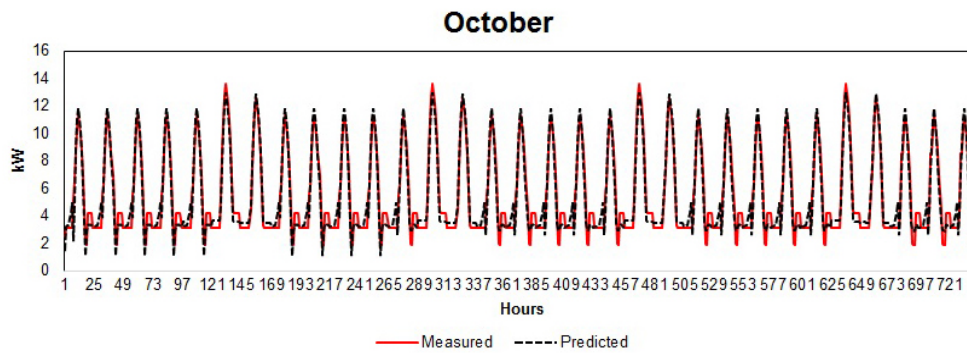


Fig. 12 Hourly load curve comparison for October after the calibration

The values shown in Tables 7, 8 and 9 are filtered, excluding hours with a MAPE value over 200%, maintaining more than 95% of the sample. Indeed, as later shown in

Table 7 Error analysis results from model calibration with 16 kW<sub>p</sub>

Indicators	February	April	July	October	Annual average
ME	0.000	0.147	-0.028	-0.005	0.097
MAD	3.122	1.367	1.666	0.654	1.709
MAPE	32%	36%	31%	17%	28%
RMSE	4.469	2.041	2.141	0.775	2.441
A	11.640	4.056	5.997	4.950	6.549
cvRMSE	38%	50%	36%	16%	37%
NMBE	0%	-2%	0%	0%	-1%

Table 8 Error analysis results from model calibration with 20 kW<sub>p</sub>

Indicators	February	April	July	October	Annual average
ME	0.000	0.119	-0.013	-0.001	0.074
MAD	3.122	1.420	1.651	0.649	1.736
MAPE	30%	29%	27%	16%	26%
RMSE	4.469	2.044	2.128	0.769	2.450
A	12.268	5.204	6.950	5.541	7.454
cvRMSE	36%	39%	31%	14%	32%
NMBE	0%	-1%	0%	0%	-1%

Table 10, few values were excluded. The MAPE reported in those tables varies between the 25% and 28%, showing a consistent reduction while increasing the PV.

Similarly, the results obtained with PV sensitive analysis are shown for April only, since it is the month mostly affected by this error as depicted in Figures 13 to 15.

Ultimately, the hours when divergences occur and their percentage on the total simulated time (8016 hours) are resumed in Table 10.

Table 9 Error analysis results from model calibration with 24 kW<sub>p</sub>

Indicators	February	April	July	October	Annual average
ME	0.46	0.03	-0.03	-0.01	0.052
MAD	3.06	1.42	1.67	0.65	1.76
MAPE	28%	27%	23%	16%	25%
RMSE	4.42	2.09	2.14	0.78	2.48
A	12.48	5.25	7.71	5.94	7.66
cvRMSE	35%	40%	28%	13%	32%
NMBE	-4%	0%	0%	0%	-4%

Table 10 Hours when a divergence occurs

Case	Divergences in hours	Divergences in percentage
24 kW <sub>p</sub>	53	0.7%
20 kW <sub>p</sub>	159	2.0%
16 kW <sub>p</sub>	190	2.4%

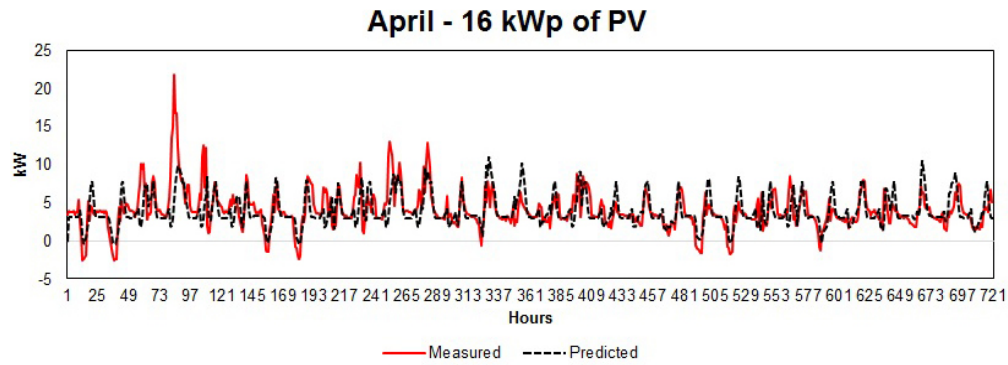


Fig. 13 Hourly load curve comparison for April after the calibration simulating a 16 kW<sub>p</sub> PV

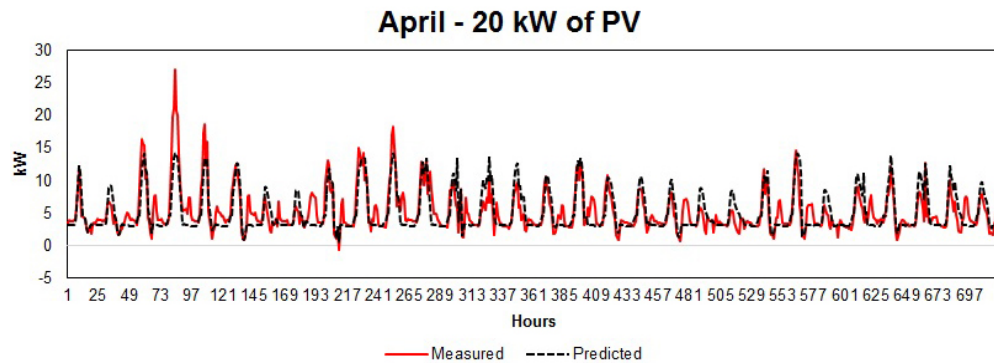


Fig. 14 Hourly load curve comparison for April after the calibration

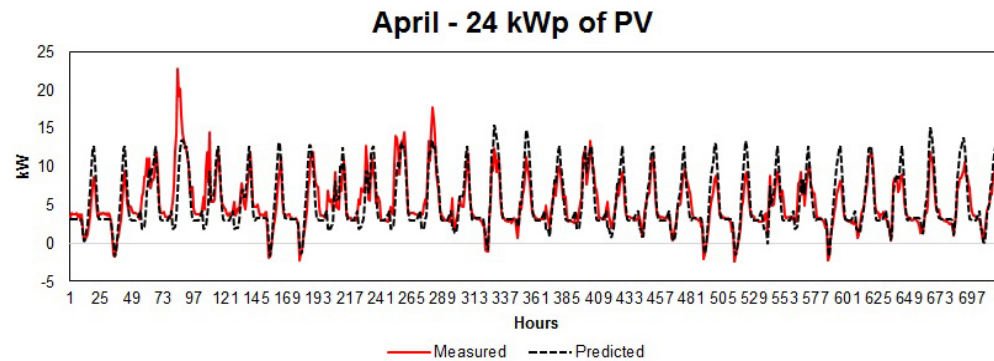


Fig. 15 Hourly load curve comparison for April after the calibration simulating a 24 kW<sub>p</sub> PV

It is possible to notice how an increase of the PV plant size lowers the encountered deviations. A possible explanation is that the PV simulator gives as a result a probabilistic hourly profile in that region, with a certain efficiency and production based on meteorological databases. Therefore, the real PV installed performances for the reference year 2018 could be better approximated by a 24 kW<sub>p</sub> instead than 20 kW<sub>p</sub>, i.e. the nominal installed power.

#### 4 Conclusion

An expeditious methodology to assess the hourly energy profile from the energy bills was presented, tested, calibrated

and finally validated.

The method starts from a preliminary assessment of reference values represented by Eqs. (1), (2), and (3) obtained from the utility bills analysis. From those latter, and a known hourly profile, it was possible to evaluate different tailored normalization factors. To obtain a proper consumption distribution along the days of the week several assumptions on building consumption were made. Nevertheless, the assumptions describe in the methodology were chosen considering the building end use as a generic office, assuring a greater replicability for similar cases. The so obtained normalization factors describe a three piecewise function capable to reconstruct the hourly consumption starting from



the utility bill information

For the considered case study, a further issue was the unknown value of self-consumption fraction of the local PV production which is usually not reported in the utility bills. The results are encouraging even though the further uncertainty introduced by the PV plant. The calibrated model, as reported in Table 8, has an almost null ME and MAD values, while the average MAPE is 26%. Likewise, the remaining error indicators testify the good model outcomes with satisfactory values. In comparison with the ASHRAE guidelines, the cvRMSE should be lower, nonetheless due to the described uncertainties it is acceptable that in this specific case, the value is slightly over the recommended threshold confirming its possible validation.

### Acknowledgements

GIFT (Geographical Islands FlexibiliTy) project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 824410.

**Funding note:** Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

**Open Access:** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

### References

- ARERA (2006). Proposte in Materia di Definizione delle Fasce Orarie per l'Anno 2007 e Successivi, atto n.19/06, documento per la consultazione. Autorità di Regolazione per Energia Reti e Ambiente, Rome, Italy. Available at [https://www.arera.it/allegati/docs/dc/06/dc\\_060703.pdf](https://www.arera.it/allegati/docs/dc/06/dc_060703.pdf). Accessed 16 Apr 2020. (in Italian)
- ARERA (2007). Aggiornamento delle Fasce orarie con decorezza 1 gennaio 2007 annex A. Delibera n.181/06. Autorità di Regolazione per Energia Reti e Ambiente, Rome, Italy. Available at <https://www.arera.it/allegati/docs/06/181-06tab.pdf>. Accessed 18 Apr 2020. (in Italian)
- ASHRAE (2014). Measurement of Energy, Demand, and Water Savings. Atlanta, GA, USA: American Society of Heating Refrigerating and Air-Conditioning Engineers.
- Birt BJ, Newsham GR, Beausoleil-Morrison I, Armstrong MM, Saldanha N, et al. (2012). Disaggregating categories of electrical energy end-use from whole-house hourly data. *Energy and Buildings*, 50: 93–102.
- Biswas MAR, Robinson MD, Fumo N (2016). Prediction of residential building energy consumption: a neural network approach. *Energy*, 117: 84–92.
- De Rosa M, Bianco V, Scarpa F, Tagliafico LA (2014). Heating and cooling building energy demand evaluation; a simplified model and a modified degree days approach. *Applied Energy*, 128: 217–229.
- DOE (2015). Guide to Determining Climate Regions by County, volume 7.3. U.S. Department of Energy, Pacific Northwest National Laboratory.
- Dudek G (2016). Pattern-based local linear regression models for short-term load forecasting. *Electric Power Systems Research*, 130: 139–147.
- Erdinc O, Uzunoglu M (2011). The importance of detailed data utilization on the performance evaluation of a grid-independent hybrid renewable energy system. *International Journal of Hydrogen Energy*, 36: 12664–12677.
- Eto JH (1988). On using degree-days to account for the effects of weather on annual energy use in office buildings. *Energy and Buildings*, 12: 113–127.
- Fabrizio E, Monetti V (2015). Methodologies and advancements in the calibration of building energy models. *Energies*, 8: 2548–2574.
- Fink HS (2011). Promoting behavioral change towards lower energy consumption in the building sector. *Innovation: the European Journal of Social Science Research*, 24: 7–26.
- Finster S, Baumgart I (2014). Privacy-aware smart metering: A survey. *IEEE Communications Surveys & Tutorials*, 16: 1732–1745.
- Ford V, Siraj A, Eberle W (2014). Smart grid energy fraud detection using artificial neural networks. In: Proceedings of IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG), Orlando, FL, USA.
- Fumo N, Mago P, Luck R (2010). Methodology to estimate building energy consumption using EnergyPlus Benchmark Models. *Energy and Buildings*, 42: 2331–2337.
- Gerwig C (2015). Short term load forecasting for residential buildings—An extensive literature review. In: Neves-Silva R, Jain L, Howlett R (eds), *Intelligent Decision Technologies*. Cham, Switzerland: Springer.
- GME (2020). Results of the Electricity Market. Gestore dei Mercati Energetici. Available at <https://www.mercatoelettrico.org/en/Default.aspx>. Accessed 18 Apr 2020.
- Grandjean A, Adnot J, Binet G (2012). A review and an analysis of the residential electric load curve models. *Renewable and Sustainable Energy Reviews*, 16: 6539–6565.



- GU (1993). Regolamento recante norme per la progettazione, l'installazione, l'esercizio e la manutenzione degli impianti termici degli edifici ai fini del contenimento dei consumi di energia, in attuazione dell'art. 4, comma 4, della legge 9 gennaio 1991, n. 10. Gazzetta Ufficiale della Repubblica Italiana, Rome, Italy. Available at <https://www.gazzettaufficiale.it/eli/gu/1993/10/14/242/so/96/sg/pdf>. Accessed 18 Apr 2020. (in Italian)
- GU (2006). Aggiornamento delle Fasce orarie con decorezza 1 gennaio 2007, Delibera n.181/06. Gazzetta Ufficiale della Repubblica Italiana, Rome, Italy. Available at <https://www.gazzettaufficiale.it/eli/gu/2006/09/11/211/sg/pdf>. Accessed 15 Apr 2020. (in Italian)
- IEA (2018). 2018 Global Status Report. Paris: International Energy Agency.
- Ji Y, Xu P, Ye Y (2015). HVAC terminal hourly end-use disaggregation in commercial buildings with Fourier series model. *Energy and Buildings*, 97: 33–46.
- Kipping A, Trømborg E (2016). Modeling and disaggregating hourly electricity consumption in Norwegian dwellings based on smart meter data. *Energy and Buildings*, 118: 350–369.
- Li Z, Xu Y, Fang S, Mazzoni S (2019). Optimal placement of heterogeneous distributed generators in a grid-connected multi-energy microgrid under uncertainties. *IET Renewable Power Generation*, 13: 2623–2633.
- Loorbach D (2010). Transition management for sustainable development: A prescriptive, complexity-based governance framework. *Governance*, 23: 161–183.
- Luo X, Hong T, Chen Y, Piette MA (2017). Electric load shape benchmarking for small- and medium-sized commercial buildings. *Applied Energy*, 204: 715–725.
- Mancini F, Nastasi B (2019). Energy retrofitting effects on the energy flexibility of dwellings. *Energies*, 12: 2788.
- Manfren M, Nastasi B (2020). Parametric performance analysis and energy model calibration workflow integration—A scalable approach for buildings. *Energies*, 13: 621.
- Massana J, Pous C, Burgas L, Melendez J, Colomer J (2015). Short-term load forecasting in a non-residential building contrasting models and attributes. *Energy and Buildings*, 92: 322–330.
- Massana J, Pous C, Burgas L, Melendez J, Colomer J (2017). Identifying services for short-term load forecasting using data driven models in a Smart City platform. *Sustainable Cities and Society*, 28: 108–117.
- Niu F, O'Neill Z, O'Neill C (2018). Data-driven based estimation of HVAC energy consumption using an improved Fourier series decomposition in buildings. *Building Simulation*, 11: 633–645.
- Ramos S, Duarte JM, Duarte FJ, Vale Z (2015). A data-mining-based methodology to support MV electricity customers' characterization. *Energy and Buildings*, 91: 16–25.
- Rosenbloom D, Meadowcroft J (2014). The journey towards decarbonization: Exploring socio-technical transitions in the electricity sector in the province of Ontario (1885–2013) and potential low-carbon pathways. *Energy Policy*, 65: 670–679.
- Smith A, Fumo N, Luck R, Mago PJ (2011). Robustness of a methodology for estimating hourly energy consumption of buildings using monthly utility bills. *Energy and Buildings*, 43: 779–786.
- Swan LG, Ugursal VI (2009). Modeling of end-use energy consumption in the residential sector: a review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13: 1819–1835.
- Tian W (2013). A review of sensitivity analysis methods in building energy analysis. *Renewable and Sustainable Energy Reviews*, 20: 411–419.
- Tronchin L, Manfren M, James PA (2018). Linking design and operation performance analysis through model calibration: Parametric assessment on a Passive House building. *Energy*, 165: 26–40.