



Data Article

Dataset of an energy community's generation and consumption with appliance allocation



Calvin Goncalves, Ruben Barreto, Pedro Faria, Luis Gomes, Zita Vale*

Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development (GECAD), Intelligent Systems Associated Laboratory (LASI), Polytechnic of Porto (P.PORTO), P-4200-072 Porto, Portugal

ARTICLE INFO

Article history:

Received 7 July 2022

Accepted 6 September 2022

Available online 9 September 2022

Dataset link: [Dataset of an Energy Community's Consumption and Generation with Appliance Allocation for One Year \(Original data\)](#)

Keywords:

Appliance allocation
End-user energy profile
Energy community
Renewable energy

ABSTRACT

Energy data measured on-site from buildings can help describe the consumption behavior of end-users and can be used to examine and prove certain theorems and models, that require a large volume of data to be gathered. However, the direct extraction of this data can often be a lengthily and costly process. As a result, a dataset of a residential community was constructed based on real data, where sample consumption and photovoltaic generation profiles were attributed to 50 residential households and a public building (municipal library), a total of 51 buildings. In addition, the overall power consumption of these houses was desegregated into the consumption of 10 commonly used appliances using real energy profiles. First, several consumption and photovoltaic generation profiles, as well as a vast collection of appliance profiles, were gathered. These profiles were obtained from household readings in different locations, while the public building's profile was based on the consumption and photovoltaic production profiles of the research building GECAD. The profiles went through the process of normalization and new profiles were generated to complete the number of end-users needed. Moreover, these profiles were given a maximum consumption and production level at random before being accepted by one of the end-users. Therefore, fourteen of these households and the public building were

* Corresponding author.

E-mail addresses: calvi@isep.ipp.pt (C. Goncalves), rudim@isep.ipp.pt (R. Barreto), pnf@isep.ipp.pt (P. Faria), lfg@isep.ipp.pt (L. Gomes), zav@isep.ipp.pt (Z. Vale).

randomly attributed with renewable solar energy. Finally, if possible, the tool created allocated, at random in previously determined intervals, the appliances' load profiles into each end-user's available consumption areas.

© 2022 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Specifications Table

Subject	Electrical and Electronic Engineering
Specific subject area	The data describes an electrical energy community, containing photovoltaic (PV) production profiles and end-user consumption profiles, desegregated by individual appliances used.
Type of data	Table (.xlsx format)
How the data were acquired	The novel dataset presented in this publication is able to describe a full energy community. It derived from a weighted combination between other public available datasets and data acquired by the authors. From the public datasets, an end-user total consumption profile was obtained from London Households, measured by smart meters in kWh (per half hour), between November 2011 and February 2014 [1]. The other end-user total consumption profile was measured in watt-hour of active energy in a house near Sceaux, France, between December 2006 and November 2010 [2]. Additionally, the individual appliances' profiles resulted from the filtering of raw electrical consumption data of the appliances from 20 households in Watts [3]. The photovoltaic and an end-user consumption data were acquired by the authors using Saia 3-phase energy meters (ale3d5f) connected to a SCADA system, able to measure energy each ten seconds [4,5].
Data format	Filtered
Description of data collection	The data was collected from several sources, exclusively for each type of profile required, these being the renewable energy production [4], end-user [1,2,4] and appliance [3] consumption profiles.
Data source location	Renewable energy production profiles <ul style="list-style-type: none"> • Institution: GECAD - Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development • City/Town/Region: Porto • Country: Portugal End-user profiles <ul style="list-style-type: none"> • Institution: UK Power Networks • City/Town/Region: London • Country: United Kingdom • Institution: EDF R&D • City/Town/Region: Sceaux • Country: France • Institution: GECAD - Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development • City/Town/Region: Porto • Country: Portugal Appliance profiles <ul style="list-style-type: none"> • Institution: REFIT • City/Town/Region: Glasgow • Country: United Kingdom
Data accessibility	Full dataset described in this publication: Repository name: Zenodo Direct URL to data: https://zenodo.org/record/6778401 Secondary data: <ul style="list-style-type: none"> • Renewable energy production profiles Repository name: PES-ISS Direct URL to data: https://site.ieee.org/pes-iss/data-sets/

(continued on next page)

- **End-user profiles**

Repository name: UK Power Networks

Direct URL to data: <https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households>

Repository name: UCI Machine Learning Repository

Direct URL to data: <https://archive.ics.uci.edu/ml/datasets/individualhousehold+electric+power+consumption>

Repository name: PES-ISS

Direct URL to data: <https://site.ieee.org/pes-iss/data-sets/>

- **Appliance profiles**

Repository name: REFIT

Direct URL to data:

<https://pureportal.strath.ac.uk/en/datasets/refit-electrical-load-measurements>

Instructions for accessing these data:

The data does not have access controls, the readings are open access and does not compromise the anonymity of the editors and reviewers.

Related research article

R. Barreto, C. Goncalves, L. Gomes, P. Faria, Z. Vale, *Evaluation Metrics to Assess the Most Suitable Energy Community End-Users to Participate in Demand Response*. *Energies*. 15 (2022) 2380. <https://doi.org/10.3390/en15072380>

Value of the Data

- This dataset is designed for use in advanced energy management models for energy communities and smart buildings based on internet of things (IoT) devices. The area of interest might include its usage for demand response models and machine learning algorithms, targeted energy and retrofit solutions for smart buildings, survey of appliance utilization, and data on energy consumption and generation. The current dataset, publicly available in [6], was already used by the authors to demonstrate the application of new energy-related models [7,8], and [9].
- Due to the variability of renewable energy sources, like photovoltaic (PV) generation, and of the usage of individual appliances, this dataset is convenient for real-life simulating and planning of an energy community.
- This dataset presents advantages for data scientists, researchers, and decision makers in the energy field, being a good source for the application of models and algorithms. Helping these to analyze the dynamics and structure of a realistic household load profile. And to understand the consumption behavior of end-users, analyzing trends and comparing information with other sources.
- This dataset can be used to create and test distinct communities of varying sizes, with diverse characteristics.
- This type of dataset, where the total load consumption is desegregated into each individual appliance used, is hard to find in existing articles.
- The data could be used to support energy building policies at the European level, helping with a somewhat practical view on energy consumption and generation.

1. Data Description

The dataset describes a residential community comprised of 50 households and a public building, totalizing 51 end-users, where 15 of these end-users were also PV producers (i.e., prosumers).

1.1. Renewable energy generation profiles

The dataset used for energy generation was originated from the research unit GECAD, namely building N, in 2019. Only one profile was needed to represent the whole community due to

the theoretical lack of variability between production profiles waveform in the same geographic area, differing only in the power generated. For that reason, the data was normalized to be later multiplied by the level of power generated.

1.2. End-user profiles

End-user profiles were obtained from a sample of energy consumption readings of London households that took part in the United Kingdom (UK) Power Networks led Low Carbon London project between November 2011 and February 2014. The readings of the energy consumption were taken at half-hourly intervals. These consumption profiles represent a balanced sample of the Greater London area inhabitants, where the data is expressed in kWh (per half hour). However, as the focus of this dataset was on 15-minute intervals, the readings had to be duplicated for each interval.

One of the base profiles employed was acquired through an individual household's electric power consumption dataset, situated in Sceaux, France, with a one-minute sampling rate. The dataset provided 2,075,259 readings collected between December 2006 and November 2010, approximately 47 months. For this dataset, the data had to be averaged into 15-minute intervals.

As the public building offers different consumption properties, a dataset with certain characteristics was considered, in this case, the consumption profile of the research unit GECAD, namely building N, with the sampling period being every 5 minutes in 2013. This data had to be averaged into 15-minute intervals.

The four different end-user profiles were inserted into an Excel spreadsheet and in periods of 15 minutes for one year, corresponding to 35,135 periods. After that the profiles were normalized to be later multiplied by the level of power consumed.

1.3. Appliance profiles

The data was obtained from a sample of appliance load consumption readings, captured from 15 different households between October 2013 and June 2015. The readings of appliance load consumption were taken at 5-second intervals, later averaged out and converted to 15-minute intervals.

1.4. Table description

The excel spreadsheet is comprised of the total consumption and total PV production profiles, then all the separated by appliance used and their profiles. All the data were inserted into 15-minute intervals and represent a full year (366 days).

- "Total Consumers" sheet: total energy consumption profiles, in kW, of the Public building (column A) and each of the 50 residential consumers (columns B to AY) for one year (35,136 periods of 15 minutes);
- "Total Producers" sheet: total PV energy Production profiles, in kW, by the Public building (column A) and each of the 14 residential producers (columns B to O) for one year (35,136 periods of 15 minutes);
- "Weather" sheet: weather data of several months in 2019, added to the dataset for the better perception of the climate conditions when the PV energy Production profiles were obtained. Column A to J represent the timestamp (in increments of 15 minutes), atmospheric temperature and dew point in degrees Celsius ($^{\circ}\text{C}$), pressure in hectopascal (hPa), wind speed and wind speed gust in kilometer per hour (km/h), humidity in grams per cubic meter (g/m^3), hourly precipitation and daily rain in millimeters (mm), and solar radiation in watts per square meter (W/m^2);

- “PublicBuilding” sheet: public building consumption profile desegregated by appliances utilized. Column A represents each rows index, column B shows the index of number of 15-minute periods in a day, 96 total, column C retrieves the total energy consumption profile, in kW, of the public building for a year in 15-minute intervals. Columns D to M contain the profiles of each appliance used in their respective period of utilization, these appliances being air conditioning (AC) unit 1, AC unit 2, AC unit 3, AC unit 4, water heater, television (TV), microwave, kettle, lighting, and refrigerator;
- [“Consumer1” - “Consumer50”] sheets: residential household consumption profile desegregated by appliances utilized. Column A represents each rows index, column B shows the index of number of 15-minute periods in a day, 96 total, column C retrieves the total energy consumption profile, in kW, of the end-user for a year in 15-minute intervals. Columns D to M contain the profiles of each appliance used in their respective period of utilization, these appliances being AC, dish washer, washing machine, dryer, water heater, TV, microwave, kettle, lighting, and refrigerator.

2. Experimental Design, Materials and Methods

The developed method consists of three steps: construction of end-user profiles, generation of PV profiles, and identification and attribution of appliances. The community profiles were created in the first two phases using sample consumption and production profiles. These profiles were normalized, multiplied to generate new profiles, assigned a maximum consumption and production level at random, and adopted by one of the end-users. After establishing the community and gathering a large library of device profiles, the third stage assigns appliance profiles to each end-user's consumption periods, resulting in the present dataset.

2.1. Renewable energy production profiles

A PV production base profile was chosen to display the community, and, using this profile, the number of producer profiles was generated and normalized, using the function (1).

$$Value_{i,k}^{norm} = \frac{Value_{i,k}}{Max(Perfil_k)}, \forall i \in Ni, \forall k \in Nk \quad (1)$$

where $Value_{i,k}^{norm}$ corresponds to the value normalized, $Value_{i,k}$ the real value, $Max(Perfil_k)$ the maximum value of profile k , Ni and Nk the number of periods and profiles respectively.

The final producers had to be granted a particular maximum production level, a simple or double tariff, and be distributed at random based on a uniform distribution. After defining each user's maximum consumption and base profile, function (2) was utilized to build an individual production profile for each end-user.

$$Load_{i,n} = CP_n \times 0.95 \times Value_{i,k}^{norm} \times X \sim unif(0.8; 1.1), \quad \forall i \in Ni, \forall n \in Nn \quad (2)$$

where $Load_{i,n}$ represents the load for the period i of the consumer n , CP_n the contracted power of the consumer, and $X \sim unif(0.8; 1.1)$ represents a uniform distribution between 0.8 and 1.1 to assists with the variability of the profiles created. Consumers do not surpass the contracted power with the multiplication $CP_n \times 0.95$.

2.2. End-user profiles

A few end-users were selected to depict the community, then, to support the diversification of the profiles, the normalization of each profile is performed by utilizing the maximum values in each, as indicated in function (1).

2.3. Appliance profiles

The appliances are identified from end-user power usage by taking n end-user's global load profiles and k appliance profiles and provides disaggregated consumption profiles that include the appliances assigned to that end-user. The following ten appliances were chosen to be included in the dataset: AC, dishwasher, washing machine, dryer, water heater, TV, microwave, kettle, lighting, and refrigerator.

Ethics Statements

This dataset does not require ethical approval, as the data on the consumption of households, public building and PV generation was obtained through open repositories. In addition, the individual labels and description of the data was removed beforehand by the authors of the data. The data was also normalized to prevent the identification of the specific buildings, and its precise location was not disclosed. The data collection did not include direct human, animal or social media data, only energy consumption and generation values.

CRedit Author Statement

Calvin Goncalves: Conceptualization, Methodology, Software, Data curation, Writing – original draft preparation, Visualization, Investigation; **Ruben Barreto:** Conceptualization, Methodology, Software; **Pedro Faria:** Conceptualization, Methodology, Supervision, Writing – review & editing; **Luis Gomes:** Supervision, Methodology, Writing – review & editing; **Zita Vale:** Conceptualization, Methodology, Supervision, Writing – review & editing, Validation.

Declaration of Competing Interest

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

Data Availability

[Dataset of an Energy Community's Consumption and Generation with Appliance Allocation for One Year \(Original data\)](#) (Zenodo).

Acknowledgments

This article is a result of the project RETINA ([NORTE-01-0145-FEDER-000062](#)), supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the [European Regional Development Fund](#) (ERDF). We also acknowledge the work facilities and equipment provided by GECAD research center ([UIDB/00760/2020](#)) to the project team.

The authors hereby declare that they have no competing interests.

References

- [1] [Dataset] SmartMeter Energy Consumption Data in London Households - London Datastore (2021). Available at: <https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households>. Accessed June 30, 2022.

- [2] [Dataset] UCI Machine Learning Repository: Individual household electric power consumption Data Set (2021). Available at: <https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption>. Accessed June 30, 2022.
- [3] [Dataset] REFIT: Electrical Load Measurements (Cleaned) – University of Strathclyde (2021). Available at: <https://pureportal.strath.ac.uk/en/datasets/refit-electrical-load-measurements-cleaned>. Accessed June 30, 2022.
- [4] Open data sets « IEEE PES intelligent systems subcommittee (2021). Available at: <https://site.ieee.org/pes-iss/data-sets/>. Accessed June 30, 2022.
- [5] Z. Vale, P. Faria, O. Abrishambaf, L. Gomes, T. Pinto, MARTINE—a platform for real-time energy management in smart grids, *Energies* 14 (7) (2021) 1820, doi:[10.3390/en14071820](https://doi.org/10.3390/en14071820).
- [6] [Dataset] C. Gonçalves, R. Barreto, P. Faria, L. Gomes, Z. Vale, Dataset of an energy community's consumption and generation with appliance allocation for one year (1.0)", Zenodo (2022), doi:[10.5281/zenodo.6778401](https://doi.org/10.5281/zenodo.6778401).
- [7] R. Barreto, C. Gonçalves, L. Gomes, P. Faria, Z. Vale, Evaluation metrics to assess the most suitable energy community end-users to participate in demand response, *Energies* 15 (7) (2022) 2380, doi:[10.3390/en15072380](https://doi.org/10.3390/en15072380).
- [8] C. Gonçalves, R. Barreto, P. Faria, L. Gomes, Z. Vale, Energy community consumption and generation dataset with appliance allocation, in: *Proceedings of the 11th Symposium on Control of Power and Energy Systems (IFAC-CPES)*, 2022.
- [9] N. Teixeira, R. Barreto, L. Gomes, P. Faria, Z. Vale, "A trustworthy building energy management system to enable direct IoT devices' participation in demand response programs.", *Electron* 11 (6) (2022) 897, doi:[10.3390/electronics11060897](https://doi.org/10.3390/electronics11060897).