



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Novel domestic building energy consumption dataset: 1D timeseries and 2D Gramian Angular Fields representation

Abdullah Alsalemi^{a,*}, Abbas Amira^{a,b}, Hossein Malekmohamadi^a, Kegong Diao^c

^a Institute of Artificial Intelligence, De Montfort University, Leicester, UK

^b Department of Computer Science, University of Sharjah, Sharjah, UAE

^c Institute of Energy and Sustainable Development, De Montfort University, Leicester, UK

ARTICLE INFO

Article history:

Received 16 December 2022

Revised 23 January 2023

Accepted 8 February 2023

Available online xxx

Dataset link: [Novel domestic building energy consumption dataset: 1D timeseries and 2D Gramian Angular Fields representation \(Original data\)](#)

Keywords:

Energy efficiency
Internet of things
Environmental sensing
Occupancy
Smart plug
Image processing
Visualization

ABSTRACT

This data article describes a dataset collected in 2022 in a domestic household in the UK. The data provides appliance-level power consumption data and ambient environmental conditions as a timeseries and as a collection of 2D images created using Gramian Angular Fields (GAF). The importance of the dataset lies in (a) providing the research community with a dataset that combines appliance-level data coupled with important contextual information for the surrounding environment; (b) presents energy data summaries as 2D images to help obtain novel insights from the data using data visualization and Machine Learning (ML). The methodology involves installing smart plugs to a number of domestic appliances, environmental and occupancy sensors, and connecting the plus and the sensors to a High-Performance Edge Computing (HPEC) system to privately store, pre-process, and post-process data. The heterogenous data include several parameters, including power consumption (W), voltage (V), current (A), ambient indoor temperature (°C), relative indoor humidity (RH%), and occupancy (binary). The dataset also includes outdoor weather conditions based on data from the Norwegian Meteorological Institute (MET Norway) including temperature (°C), outdoor humidity (RH%), barometric pres-

* Corresponding author.

E-mail address: abdullah.alsalemi@my365.dmu.ac.uk (A. Alsalemi).

<https://doi.org/10.1016/j.dib.2023.108985>

2352-3409/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Please cite this article as: A. Alsalemi, A. Amira and H. Malekmohamadi et al., Novel domestic building energy consumption dataset: 1D timeseries and 2D Gramian Angular Fields representation, Data in Brief, <https://doi.org/10.1016/j.dib.2023.108985>

sure (hPa), wind bearing (deg), and windspeed (m/s). This dataset is valuable for energy efficiency researchers, electrical engineers, and computer scientists to develop, validate, and deploy and computer vision and data-driven energy efficiency systems.

© 2023 The Authors. Published by Elsevier Inc.

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

1 Specifications Table

Subject	Renewable Energy, Sustainability and the Environment
Specific subject area	Appliance-level electric power consumption with ambient environmental conditions in domestic building
Type of data	Timeseries Generated Images from Timeseries Figures Charts
How the data were acquired	The dataset was acquired using a smart plugs and ambient environmental sensors. The sensors were installed in a kitchen, living area, and a room of a household. The data was collected through a combination of local WiFi (smart plugs) and ZigBee protocol (environmental sensors). The data was collected on an edge computing hub, the ODR0ID-XU4 equipped with WiFi and CC2531 wireless dongles. It uses specialized energy management, Home Assistant, to store, pre- and post-process data. The following instruments were used in acquiring the data: 7 LocalBytes Power Monitoring Smart Plug for power consumption, 2 SONOFF SNZB-03 for occupancy detection, 3 SONOFF SNZB-02 for temperature and humidity measurement. Outdoor weather conditions were aggregated based on data from MET Norway.
Data format	Raw Analyzed
Description of data collection	The data was collected in a domestic household between April 2022 and November 2022 in a household in the United Kingdom from eleven sensors. Initially, the heterogenous data has been collected in time-series format. Additionally, the data has been later transformed into a two-dimensional (2D), heat map-like format using Gramian Angular Fields (GAF) for to aid in classification and data visualization. In the case of the 2D dataset, the readings were normalized using min-max normalization.
Data source location	· City/Town/Region: Leicester · Country: United Kingdom
Data accessibility	Repository name: Novel domestic building energy consumption dataset: 1D timeseries and 2D Gramian Angular Fields collection Data identification number: 10.17632/v2wr7grbbg.1 Direct URL to data: http://dx.doi.org/10.17632/v2wr7grbbg.1

2 Value of the Data

- Data driven research: The data is valuable to the research community in which researchers can use to train Machine Learning (ML) energy data classification models on 1D timeseries and 2D GAF appliance-level dataset;
- Comprehensive parameters: The heterogenous dataset provides a large set of useful parameters that combines appliance-level data and contextual information for the surrounding environment such as temperature, humidity, and occupancy;
- Visualization: The data also presents energy data summaries as 2D images to help obtain novel insights from the data using data visualization; and

1. Objective

In the advancing field of energy efficiency, developing robust computational methods for analyzing energy efficiency behavior models necessitate creating correspondingly robust and rich

14 dataset [1]. In domestic households, collecting appliance-level and ambient environmental data
15 can lead to producing more effective energy efficiency models using Artificial Intelligence (AI)
16 methods [2]. This is especially when working on energy-saving research for domestic house-
17 holds where economic intrinsic incentives exist to save for the cost and the environment [3]. Ac-
18 cordingly, this dataset aims to help researchers train ML energy classification and recommender
19 systems using 1D and 2D data formats, whereby, classical time-series models can be utilized
20 as well as well-developed Deep Learning (DL) systems that can perceive the intricate details of
21 GAF-generated snapshot of data.

22 2. Data description

23 The dataset described is contained in three data containers (1) raw data for power consump-
24 tion time-series, (2) raw data for ambient environmental conditions, and (3) raw 2D GAF data.
25 The data format includes tables as Comma-Separated Values (CSV) files and images as Portable
26 Network Graphics (PNG) files inside either folders or compressed zip files.

27 2.1. Raw Data: Power Consumption Time Series

28 A collection of tables (CSV files) that depict pre-processed power appliance-level consump-
29 tion data as follows:

- 30 - Sheet 1: Power consumption of plug 1 (Television).
- 31 - Sheet 2: Power consumption of plug 2 (Kettle).
- 32 - Sheet 3: Power consumption of plug 3 (Computer Setup 1).
- 33 - Sheet 4: Power consumption of plug 4 (Toaster).
- 34 - Sheet 5: Power consumption of plug 5 (Washing Machine).
- 35 - Sheet 6: Power consumption of plug 6 (Computer Setup 2).
- 36 - Sheet 7: Power consumption of plug 7 (Fridge).

37 Each sheet includes the following columns: timestamp in %year-%month-%day%hour:%minute:
38 %second.% (%Y-%m-%d%H:%M:%S.%f) format, parameter value (W), and UNIX timestamp.

39 2.2. Raw Data: Ambient Environment Time Series

40 A collection of tables (CSV files) that depict pre-processed ambient indoor (sheets 1–6) and
41 outdoor (7) environment conditions as follows:

- 42 - Sheet 1: Temperature in kitchen/living room (°C).
- 43 - Sheet 2: Temperature in office room (°C).
- 44 - Sheet 3: Humidity in kitchen/living room (RH%).
- 45 - Sheet 4: Humidity in office room (RH%).
- 46 - Sheet 5: Occupancy in kitchen (binary).
- 47 - Sheet 6: Occupancy in living room (binary).
- 48 - Sheet 7: Outdoor weather data based on data from MET Norway temperature (°C), out-
49 door humidity (RH%), barometric pressure (hPA), wind bearing (deg), and windspeed
50 (m/s).

51 Each sheet includes the following columns: timestamp in %Y-%m-%d%H:%M:%S.%f format, pa-
52 rameter value, and UNIX timestamp, with the exception of the outdoor weather data which is
53 organized as follows: datetime%Y-%m-%d%H:%M:%S format UNIX timestamp, weather state, tem-
54 perature, humidity, barometric pressure, wind bearing (direction), and wind speed.

55 2.3. Analyzed Data: 2D GAF Energy Data

56 Zip files representing data for the following:

- 57 - Power consumption of plug 1 (Television).
- 58 - Power consumption of plug 2 (Kettle).
- 59 - Power consumption of plug 3 (Computer Setup 1).
- 60 - Power consumption of plug 4 (Toaster).
- 61 - Power consumption of plug 5 (Washing Machine).
- 62 - Power consumption of plug 6 (Computer Setup 2).
- 63 - Temperature in kitchen/living room.
- 64 - Temperature in office room.
- 65 - Humidity in kitchen/living room.
- 66 - Humidity in office room.
- 67 - Occupancy in kitchen.
- 68 - Occupancy in living room.

69 In the above, each file includes:

- 70 - Files list (CSV): includes a description of all the training GAF files
- 71 - GAF raw data folder: includes raw GAF data, each image min-max normalized to 1-hour
- 72 snapshots

73 Examples for time-series data are depicted in Fig. 1. Also, sample raw GAF data is shown in
74 Fig. 2.

75 3. Experimental Design, Materials and Methods

76 To start, the data is collected at a small-size residential household in the UK. Prior to config-
77 uring sensors and smart plugs, a central data management edge computing hub is needed. Fig. 3
78 shows the main data collection setup components. For this dataset, we have used the ODROID-
79 XU4 edge platform, which has been chosen based on its cost and performance effectiveness in
80 data-driven workflows [4]. Connected to a local network using an Ethernet cable, the ODROID-
81 XU4 runs Home Assistant, which is an open-source smart home management system. Initially
82 the raw data is collected in the Structured Query Language (SQLite) format before further post-
83 processing.

84 Afterwards, smart plugs and sensors are installed in specified locations at the household.
85 First, smart plugs are connected to appliances including a kettle, TV, toaster, computer setup,
86 fridge, and washing machine. Every plug is calibrated separately with a power meter to mini-
87 mize reading errors. Also, environmental condition sensors, namely temperature, humidity, and
88 occupancy sensors are placed in strategic locations (e.g., occupancy sensors are placed at the
89 corner of a given room in order to maximize occupancy detection angle) to capture contextual
90 information that support power consumption data. The occupancy sensors use Passive InfraRed
91 (PIR) technology to presence detection. Similarly, the temperature and humidity sensor are cali-
92 brated against reference meters to ensure accuracy. The smart plugs and sensors are installed in
93 the living room, kitchen, and study room. Table 1 describes the dataset's appliances.

94 Once all smart plugs and sensors are installed and configured, they are tested and validated
95 wireless connectivity tests, occupancy detection accuracy tests, and data integrity tests on the
96 ODROID-XU4. Table 2 shows parameter validation of the used sensors.

97 With data acquisition frequency averaging between 1 and 5 s, the dataset grows in size quite
98 quickly. In the span of seven months, the data has reached more than 8.5 million datapoints. As
99 continuous readings are accumulated into Home Assistant's SQLite database file, the data grows
100 quite large, exceeding more than 5 GB. Accordingly, the raw database is restructured to only
101 have the required data columns for further post-processing (i.e., timestamp, appliance name,

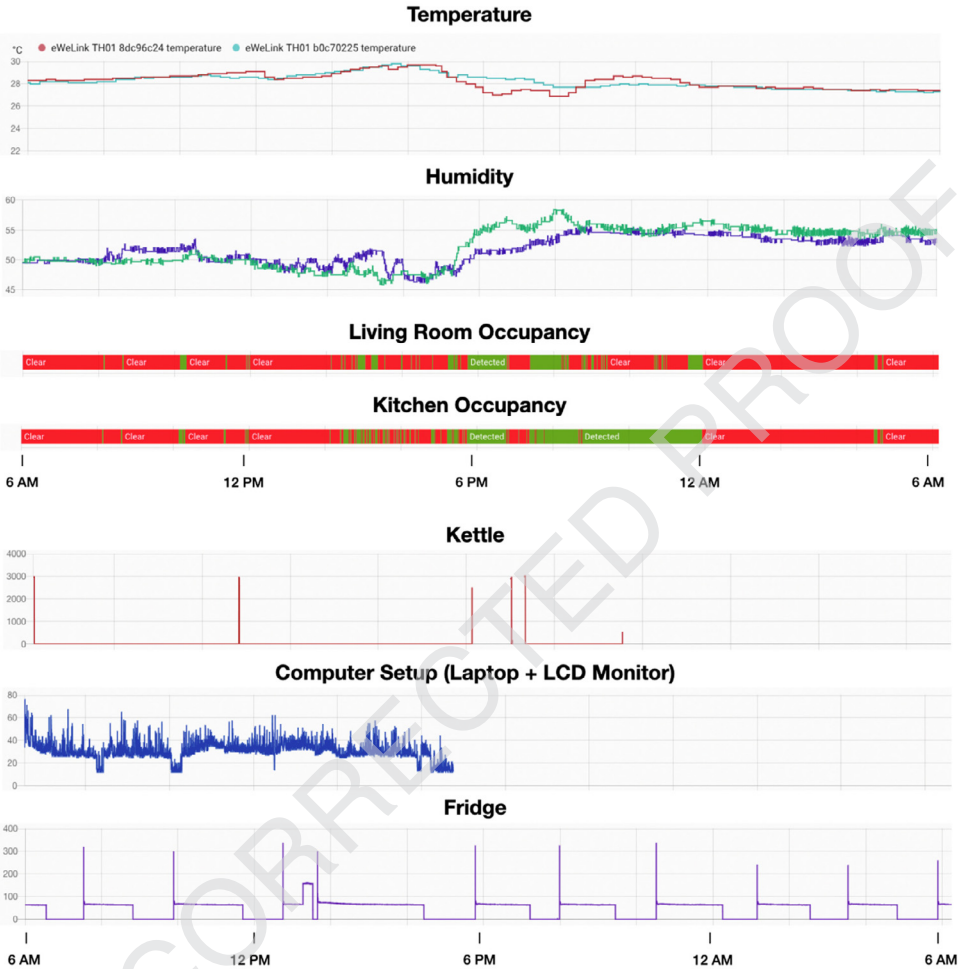


Fig. 1. Examples of collected time-series data (from top to bottom) including temperature, humidity, occupancy, kettle, computer setup, and fridge.

102 power consumption, temperature, humidity, and occupancy). Following, the data is exported to
 103 a CSV file, after which the file size of the database is significantly compressed. It is noteworthy
 104 to mention that the outdoor condition's data acquisition frequency varies depending on the
 105 received data from MET Norway live data sources.

106 As described earlier, 1D time series are also be transformed into GAF representation by (a)
 107 converting cartesian points to the polar coordinates and using the Gramian Angular Summation
 108 Field (GASF) [5,6] in Eq. (1) [7]:

$$GASF = \begin{pmatrix} \cos(\vartheta_1 + \vartheta_1) & \cdots & \cos(\vartheta_1 + \vartheta_N) \\ \vdots & \ddots & \vdots \\ \cos(\vartheta_N + \vartheta_1) & \cdots & \cos(\vartheta_N + \vartheta_N) \end{pmatrix} \quad (1)$$

109 where ϑ the is derived in Eq. (2):

$$\{\vartheta = \arccos(\tilde{x}_i), 0 < \tilde{x}_i < 1 \in x\} \quad (2)$$

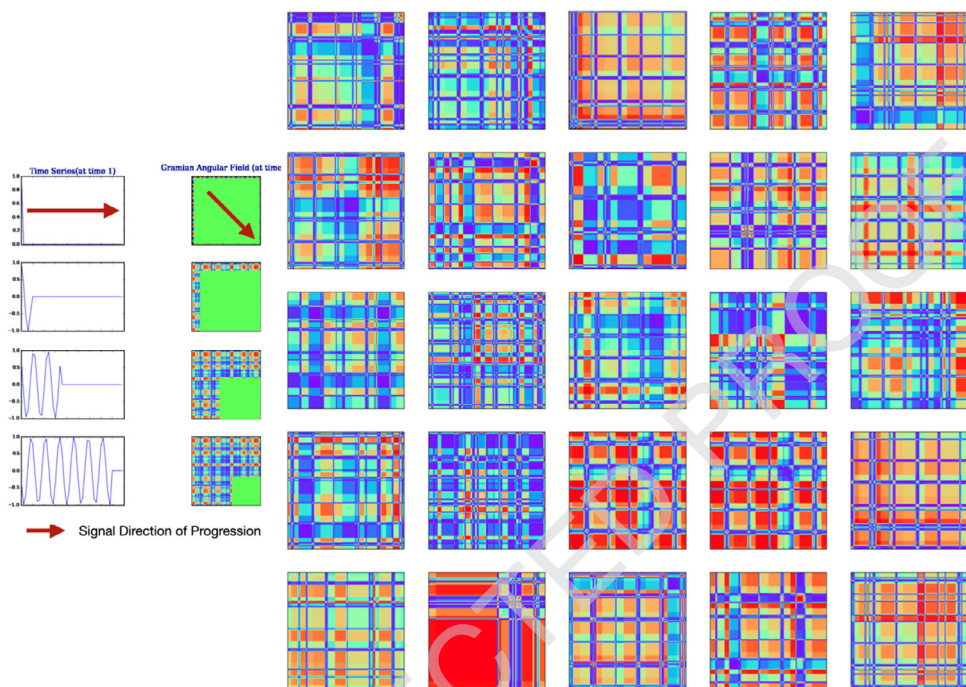


Fig. 2. Sample 2D GAF images of 1-hour snapshots of energy consumption data of a TV.

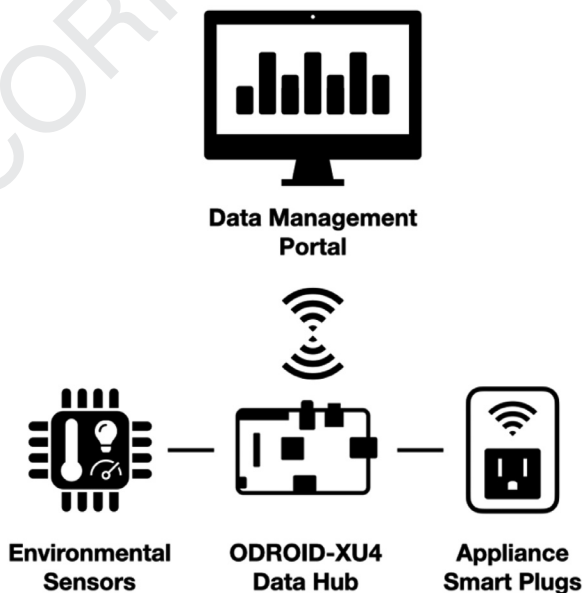


Fig. 3. Overview of the Data Collection Setup. Edge computing board icon obtained from fredly from the Noun Project.

Table 1

Specifications of appliances used in the dataset.

#	Appliance Name	Manufacturer	Power Rating (W)	Location
1	Television	Hisense	74	Living Room
2	Kettle	Tesco	2550–3000	Kitchen
3	Computer Setup 1 (MacBook Pro + Dell Monitor)	Apple/Dell	61/18	Office Room
4	Toaster	Tesco	750	Kitchen
5	Washing Machine	Indesit	2200	Kitchen
6	Computer Setup 2 (laptop + charging hub)	Hewlett Packard	119.9/38	Office Room
7	Fridge	Iceking	70–100	Kitchen

Table 2

Overview of sensor validation.

Parameter	Sensor	Average Accuracy
Temperature	SONOFF SNZB-02	±1 °C
Humidity	SONOFF SNZB-02	±5%
Occupancy	SONOFF SNZB-03	Up to 6m (100°)
Power	LocalBytes Power Monitoring Smart Plug	2–3.5%

Henceforth, the timeseries data is fed into the GAF processing program developed in [8] to produce 2D GAF image files, each file representing a 1-hour fragment of the data.

Ethics statements

This data was collected in accordance the Declaration of Helsinki and have obtained ethical approval from Faculty of Computing, Engineering and Media at De Montfort University (CEM ID No G414200).

CRediT author statement

Conceptualization, A. Alsalemi, A. Amira, H. Malekmohamadi, and K. Diao; methodology, A. Alsalemi, A. Amira, H. Malekmohamadi, and K. Diao; writing—original draft preparation, A. Alsalemi.; writing—review and editing, A. Alsalemi, A. Amira, H. Malekmohamadi, and K. Diao; validation, A. Alsalemi, A. Amira, H. Malekmohamadi, and K. Diao; supervision, A. Amira, H. Malekmohamadi, and K. Diao; All authors have read and agreed to the published version of the manuscript.

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

Novel domestic building energy consumption dataset: 1D timeseries and 2D Gramian Angular Fields representation (Original data) (Mendeley Data)

129 Acknowledgments

05 This work was supported by the Doctoral College Support and Development Fund and Insti-
130 tute of Artificial Intelligence, De Montfort University, Leicester, United Kingdom.
131

132 References

- 133 [1] C. Benavente-Peces, N. Ibadah, Buildings energy efficiency analysis and classification using various machine learning
134 technique classifiers, *Energies* 13 (2020) 3497, doi:[10.3390/en13133497](https://doi.org/10.3390/en13133497).
- 135 [2] S. Attia, Data on residential nearly Zero Energy Buildings (nZEB) design in Eastern Europe, *Data Brief* 43 (2022)
136 108419, doi:[10.1016/j.dib.2022.108419](https://doi.org/10.1016/j.dib.2022.108419).
- 137 [3] M.A. Andor, K.M. Fels, Behavioral economics and energy conservation – a systematic review of non-price interven-
138 tions and their causal effects, *Ecol. Econ.* 148 (2018) 178–210, doi:[10.1016/j.ecolecon.2018.01.018](https://doi.org/10.1016/j.ecolecon.2018.01.018).
- 139 [4] A. Alsalemi, A. Amira, H. Malekmohamadi, K. Diao, Facilitating deep learning for edge computing: a case study on
140 data classification, in: 2022 IEEE Conf. Dependable Secure Comput. DSC, 2022, pp. 1–4, doi:[10.1109/DSC54232.2022.](https://doi.org/10.1109/DSC54232.2022.9888939)
141 [9888939](https://doi.org/10.1109/DSC54232.2022.9888939).
- 142 [5] Y.-Y. Hong, J.J.F. Martinez, A.C. Fajardo, Day-ahead solar irradiation forecasting utilizing gramian angular field and
143 convolutional long short-term memory, *IEEE Access* 8 (2020) 18741–18753, doi:[10.1109/ACCESS.2020.2967900](https://doi.org/10.1109/ACCESS.2020.2967900).
- 144 [6] K.P. Thanaraj, B. Parvathavarthini, U.J. Tanik, V. Rajinikanth, S. Kadry, K. Kamalanand, Implementation of Deep Neural
145 Networks to Classify EEG Signals using Gramian Angular Summation Field for Epilepsy Diagnosis, *ArXiv200304534*
146 *Cs Eess.* (2020). <http://arxiv.org/abs/2003.04534> (Accessed 22 September 2021).
- 147 [7] Z. Wang, T. Oates, Imaging time-series to improve classification and imputation, in: Twenty-Fourth Int. Jt. Conf. Artif.
148 *Intell.*, 2015.
- 149 [8] A. Alsalemi, A. Amira, H. Malekmohamadi, K. Diao, F. Bensaali, Elevating energy data analysis with M2GAF: micro-
150 moment driven Gramian angular field visualizations, in: *International Conference on Applied Energy*, 2021.

151