Data-driven tool for early building energy performance diagnostic

Olaia Eguiarte[a], Antonio Garrido-Marijuan[b], Iñigo López[c], Noelia Vicente-Gómez[d] Ander Romero-Amorrortu[e]

[a] *TECNALIA, Basque Research and Technology Alliance (BRTA), 48160 Derio, Spain, olaia.eguiarte@tecnalia.com, CA,*
[b] *TECNALIA, Basque Research and Technology Alliance (BRTA), 48160 Derio, Spain, antonio.garridomarijuan@tecnalia.com,*
[c] *TECNALIA, Basque Research and Technology Alliance (BRTA), 48160 Derio, Spain, inigo.lopez@tecnalia.com,*
[d] *TECNALIA, Basque Research and Technology Alliance (BRTA), 48160 Derio, Spain, noelia.vicente@tecnalia.com,*
[e] *TECNALIA, Basque Research and Technology Alliance (BRTA), 48160 Derio, Spain, ander.romero@tecnalia.com,*

Abstract:
The building sector is responsible for a large part of the final energy consumption in Europe. One of the most relevant steps in the process of reducing energy consumption in buildings is the energy performance diagnostic. In this paper, a data-driven methodology to obtain early energy performance assessment of commercial buildings using the energy signature approach is used. As a result, a web-dashboard and API that can analyse user data input and produce streamlined outputs like suggestions for energy-saving measures is developed. In order to prove the correct functioning of the tool, a demo site of a commercial building in Dublin has been used.

Keywords:
Data-driven tool, building performance diagnostic, energy signature, commercial buildings

1. Introduction

The building sector is one of the largest consumers of resources at European level in terms of both material and energy aspects throughout all stages of a construction project [1], contributing to 40% of energy consumption and 36% of carbon emissions. An estimated 75% of buildings in the European Union are inefficient, yet only 1% undergoes renovation each year [2]. Non-residential buildings account for approximately 25% of Europe's 25 billion m^2 of useful building floor area, with 28% being wholesale or retail premises, 23% offices, 11% hotels and restaurants, and 4% sports facilities [3]. Commercial landlords often own these types of buildings and rent or lease them to one or more tenants. The split incentive problem is a significant obstacle to energy-efficient renovations in commercially rented buildings, where the benefits of a transaction do not go to the person who pays for it [4]. In this context, it is necessary to develop, test, validate and exploit new business models that lead to greater uptake of Smart Energy Services deployed via performance-based contracting in the commercial rented sector, supported by more accurate and dynamic measurement and verification of energy savings and flexible consumption in order to identify and develop business opportunities.

There are already protocols such as the guide proposed by ASHRAE [5] which set a reference frame for measurement of energy and demand savings of heating, cooling and air-conditioning. In the commercial building context, ASHRAE [6] developed a book providing standardized set of performance measurement protocols that can be applied internationally. Regarding heat load forecasting alternatives for buildings, one of the most suitable alternatives is provided by data-driven demand forecasting models. A wide variety of data-driven models exist and have been successfully implemented for early building performance diagnosis. Moreover, several authors have implemented this approach at district scale [7-9]. Data-driven models can be classified in different groups and one of them is the black-box models, which are purely based on data and statistical techniques with no physical interpretation of the building. In this sense, one of the more common types of black-box models are the energy signature models, which can provide successful results for monthly and seasonal data as demonstrated in several research [10-14]. Energy signature models predict a building's energy consumption based on external climate data. They are usually represented as a graph of overall energy use versus outdoor air temperature [15]. Furthermore, using daily or hourly intervals can provide further insights into typical energy demands in comparison to monthly or weekly patterns, allowing for a more accurate analysis [16].

https://doi.org/10.52202/069564-0297

Therefore, smart solutions are needed to identify potential flexibility opportunities and energy efficiency upgrades with high energy saving potential and communicates estimations of their expected added value to both tenants and building owners.

2. Objectives

This paper presents a new data-driven energy diagnostic approach to identify the most significant energy streams in commercial buildings using a minimal dataset. The algorithms consider general information about the building, such as location, size, usage, and HVAC characteristics, as well as overall facility energy consumption. The diagnostic provides granular data for integration with energy tariffs in real practice, dividing energy use by energy carrier and electricity use by billing schedules. The methodology for the energy use diagnostics is detailed in section 3. The present paper summarizes the work performed, including the development of a user-friendly web-dashboard and Application Programming Interface (API) that allows users to upload specific building information and datasets for early building energy diagnostics. The resulting baseline is cross-referenced against current building performance databases for benchmarking, and key performance metrics are calculated to identify energy-saving measures. Data from a real case study is used to test the algorithms. The outputs from this case study are included in section 6.

3. Methodology

This section details a data-driven methodology approach for carrying out early energy performance diagnostics of commercial buildings by considering energy building signature models. In this case, the developed algorithms are later made available for general use via a web-dashboard and an API. The calculation of the energy signature, the diagnostic requirements and main input and outputs from the developed tool are detailed in following subsections.

3.1. Energy building signature

Energy patterns in buildings are typically represented by one of the behaviours depicted in Figure 1. This model is a general approach to modelling energy loads in buildings and can be considered a good approximation in cases where heating and cooling are provided by the same energy source and where the heating and cooling loads do not overlap in temperature range.

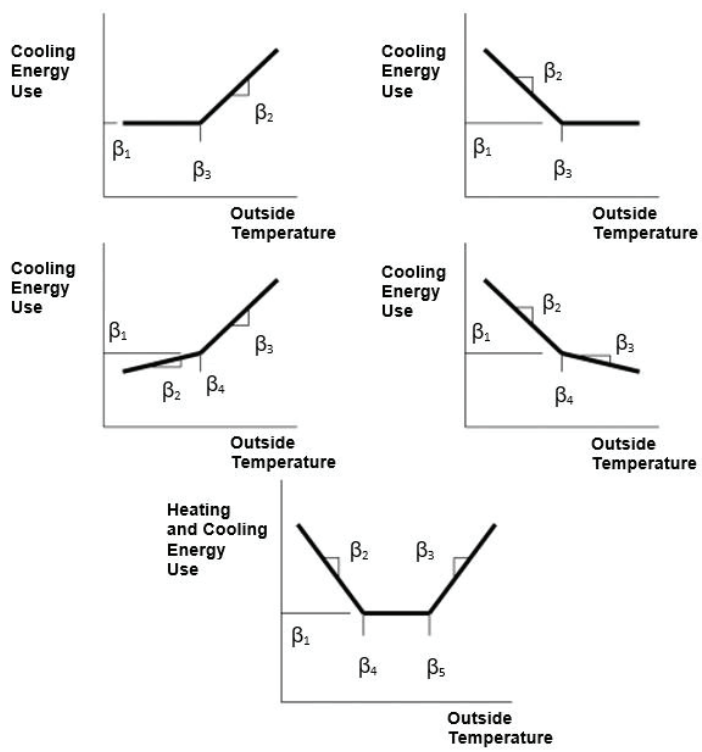

Figure 1. Overall approach to changepoint models. Top row: 3 parameter cooling and heating models. Second row from top: 4 parameter cooling and heating models. Bottom row: five parameter heating and cooling model. Adapted from [17].

Therefore, a regression model has been incorporated to depict the average performance of the commercial buildings on working days. The calculation method chosen to generate the energy signature is a segmented linear regression, in particular, the piecewise regression. Piecewise regression involves fitting a linear regression model to data that has one or more breakpoints where the slope changes. In the Python package used to perform this task, this type of regression follows the approach described by [18], where the breakpoint positions and the linear models are simultaneously fitted using an iterative method. According to [19], the general form of a one breakpoint model is implemented as in Eq. (1). As it is a non-linear relationship, it cannot be solved directly through linear regression. To take a linear approximation it is necessary to use a Taylor expansion around some initial guess for the breakpoint, ψ (0).

$$y = ax + c + \beta(x-\psi)H(x-\psi) + \zeta \tag{1}$$

This results in a linear relationship and thus a new breakpoint estimate, ψ (1), which can be used to perform ordinary linear regression using the stats models Python package [20]. After that, the process must be iterated until the breakpoint estimate converges and the algorithm stops. For multiple breakpoints, the same approach can be used with a multi-variate Taylor expansion around an initial guess for each breakpoint.

The resulted regression will just consider the outdoor temperature but no other significant factors such as solar radiation, occupancy, ventilation, etc. For that reason, 10% upper and lower boundaries are added to account those other aspects.

3.2. Energy use diagnostics

The purpose of the energy use diagnostics is to assess the energy consumption and associated costs of commercial buildings in this case. To adapt the diagnostics to the methodology scope, some requirements have been considered:

- Data sources. Available data sources such as on-site Building Energy Management Systems and data from utilities or energy suppliers.
- Time intervals. The processes should be frequency adaptable to match different time intervals (hourly, monthly data, etc.).
- Subsystems. Assign different energy use to specific areas and/or subsystems in the building.
- Benchmark. Compare energy consumption with other buildings of similar size, configuration, use and climatic conditions.
- Measures. Identify energy-saving measures and determine their potential.

By applying the developed algorithms in section 3.1, a regression with 10% boundaries is obtained. In this case, data located out of that boundary will be considered as misuse energy and some energy saving measures will be given to match the regression boundaries. These measures include optimizing heating and cooling schedules, upgrading lighting systems and heating/cooling equipment, and improving the building envelope.

3.3. Tool description

The above detailed methodology enabling early building performance diagnostics is accessible via a web-dashboard. To ensure user-friendliness, the API facilitates communication with the algorithms by simplifying the underlying calculations and exposing only the necessary objects to stakeholders without requiring knowledge of the operations that occur behind the scenes. Consequently, the web-dashboard enables users to upload specific building information and datasets to obtain early building energy diagnostics through the API, which accesses the developed algorithms.

3.3.1. Input data

To enhance user-friendliness, the input data has been simplified. While additional data such as bank holidays and HVAC characteristics could enable more functionalities and analysis, simplicity has been prioritized to make the application more appealing to potential users. The minimal dataset needed includes the following information.

- Building characteristics:
    - Building location (city): this information will be used to load the weather file of the indicated city or location. The developed API presents a set of 3 preloaded locations; Dublin (Ireland), Madrid (Spain) and Thessaloniki (Greece).
    - Building size (m^2): the total area of the building will be used to normalize its consumption and to compare it to the reference values.
- Energy data with hourly granularity (electricity, natural gas, or other fuels) in kWh.
- Service provided: it is necessary to indicate whether the energy data loaded corresponds to heating, cooling, or heating and cooling.

https://doi.org/10.52202/069564-0297

- Actual building usage:
    - Opening and closing hours.
    - Working days.

3.3.2. Data analysis

Before proceeding to the analysis of the uploaded data, a pre-processing step is performed. This includes a study that analyses the validity of the uploaded data, eliminates possible outliers and erroneous measurement values, transforms the hourly data to a daily frequency and indicates whether the final data quantity, after cleaning, is sufficient for further analysis. If the pre-processing gives satisfactory results, the analysis of the data is performed to generate the building's energy signature.

Three different analysis processes are performed on the uploaded consumption data:

- Non-working hours consumption analysis: based on the hourly frequency data as well as the information provided on days and times of use of the building, a study of the building's consumption is carried out to determine how much is consumed during non-working periods.
- Benchmarking: if the uploaded data cover a whole year, a comparison of the total consumption with the reference values is made.
- Energy signature: daily frequency data is classified according to whether it is a working or non-working day. A piecewise regression is applied in the working day dataset. The information of whether the provided data corresponds to heating, cooling or both is used to determine the number of breakpoints of the regression.

3.3.3. Output

The different results of the analysis are reflected in a graph where the regression line, the 10% boundaries and the uploaded data are shown. The outputs are listed below:

- Energy misuse in non-working hours.
- Energy consumption comparison against reference values.
- Energy signature of the building.
- Qualitative analysis of the building and its system based on the calculated breakpoint in the piecewise regression.

4. User interface of the tool

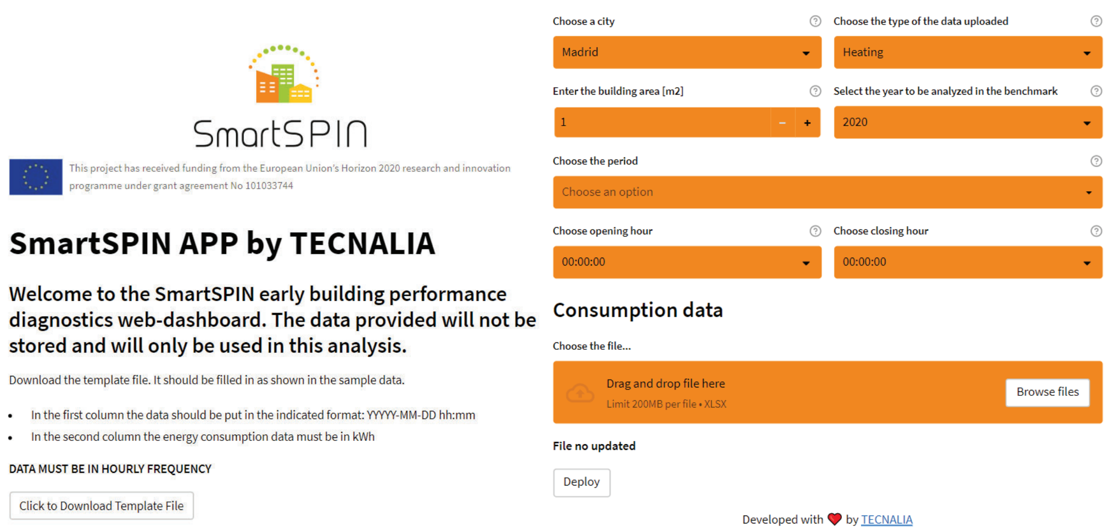

Figure 2. Web-dashboard, serving as interface between users and algorithms.

As a result of the detailed methodology, an API and web-dashboard were developed. The user interface is linked to the developed algorithms via API. Once the calculations are performed, some recommendations are offered to the user.

This tool allows to users to provide data to obtain an early building performance diagnostics web-dashboard. The user enters the building characteristics through the different available options as shown in Figure 2. As stated above, data must be introduced in hourly frequency.

5. Case study

The proposed methodology is applied to a demo site located in Dublin (Ireland), consisting of six floors with a classical façade constructed in reconstituted stone precast concrete panels. The building was constructed in 1996 and was retrofitted in 2014, after a prolonged period of being unoccupied. After the refurbishment, the building was partially occupied in 2015 for office uses in the following portions of floor area as shown in Table 1. The landlord area includes the ground floor reception, stair cores, toilets, basement level, and subbasement level.

Table 1. Floor areas per occupancy.

	Area (%)	Area (m^2)	Occupancy (persons)
Landlord	29.14%	1,809.00	
Tenant 1 (Investment Services)	43.72%	2,714.36	265
Tenant 2 (Private Banking and Asset Management)	12.97%	805.20	65
Tenant 3 (Hedge Fund)	14.18%	880.44	120

A site survey of the case study building was conducted to gather the necessary information required for applying the tool, such as HVAC system components, occupancy, and schedules information, available monitored data, etc. Due to the building use, there are no energy intensive processes associated to the normal operation of the building. Therefore, main consumptions are related to lighting and HVAC systems. The opening times of the offices are from 8h to 18h from Monday to Friday. There is no occupancy during weekends or holidays. The input data of the case study building have been introduced via the developed web-dashboard as shown in Figure 3.

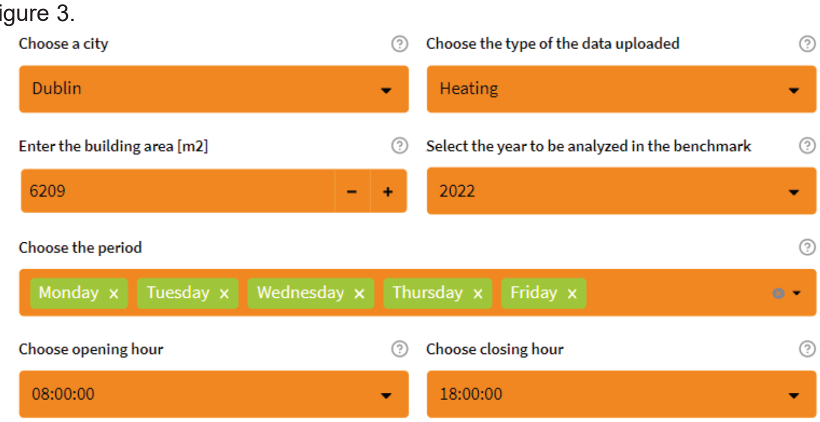

Figure 3. Input data introduced for the Irish demo-site in the web-dashboard.

Apart from input data detailed in Figure 3, heating consumption data (see Figure 4) were also obtained in hourly basis from 26th January 2022 to 14th December 2022 (both included). The boxplot graph shows the statistical distribution of the consumption for each day of the week.

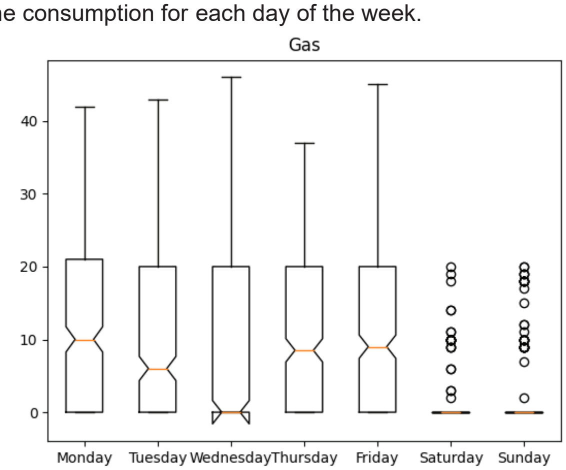

Figure 4. Heating energy consumption distribution during a typical week in the Dublin building.

https://doi.org/10.52202/069564-0297

6. Results

The present section includes the early building diagnostic for a demo site located in Ireland. Once the required input data have been uploaded (see Figure 3), the API provided results analysis as shown in Figure 5. The graphic shown in Figure 5 represents the daily aggregate values of heating consumption (kWh) versus daily average outdoor temperature, where red points belong to weekend and other non-working days. On the other hand, grey points represent heating consumption for working days. As detailed in methodology, 10% upper and lower boundaries have been added to the model and are represented in the graphic with a grey shadow. Data out of this boundary represents energy misuse.

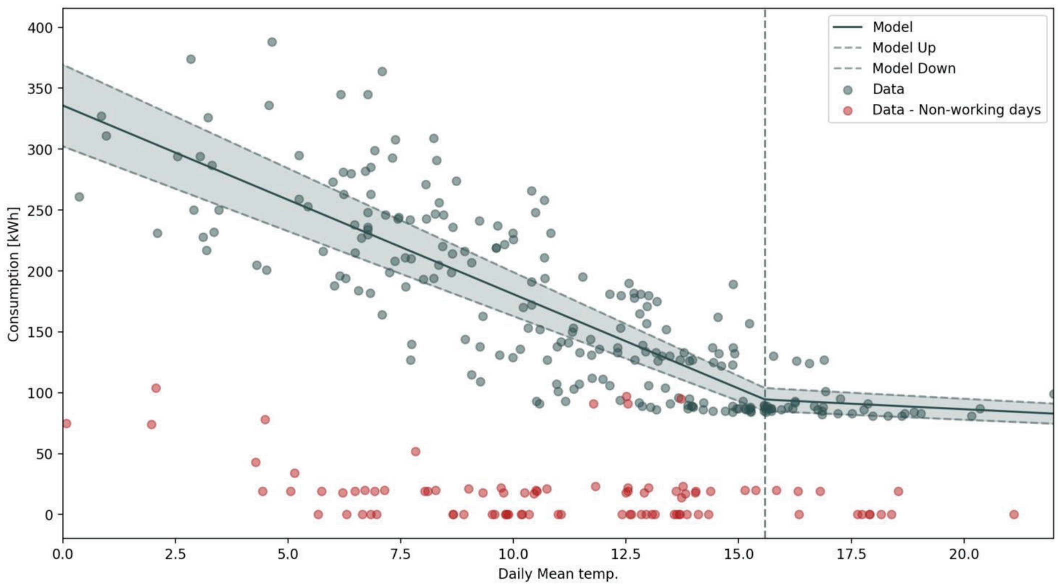

Figure 5. Daily aggregated values of heating consumption (kWh) versus daily average outdoor temperature.

As observed in the figure, heating consumption is weather dependent and increase with low outdoor temperatures. The changepoint corresponding to the regression model take place for a mean outdoor temperature of 15.58ºC. That means that the building heating demand starts for outdoor temperature under 15.58ºC, which could be considered as high. Some consumption peaks have been observed for non-working days (see red dots in Figure 5), which are assumed to be failures on the monitoring system and can be neglected.

7. Conclusions and discussion of results

This paper presents the development of an API and a web-dashboard which allows users for an early building performance diagnostic.

The creation of the API that accesses the developed algorithms highlights the critical role that data availability and synchronization play in this context. The recommended diagnostic procedure heavily relies on utility meters (which are still being delivered with some difficulties) and climate data, which is available from open sources. However, obtaining data from energy sub-meters, building usage, and indoor comfort conditions is still of interest and would significantly enhance performance assessment.

The algorithms created for its implementation in the API have been employed to evaluate the data provided by a commercial demo building in Dublin. The evaluation of this data has been thoroughly presented and analysed in this document to serve as an exemplary case study of the methodology. By application of the developed methodology, it is possible to identify the most significant energy streams in the building using minimal information.

After the application of the regression model, it has been proven that it correctly represents the average performance of the offices during operational days. Data collected in days where the heating consumption is out of a 10% boundary from the model are identified as misuse days and might be separated from non-operational days. Some recommendations can be made based on the analysis performed.

The early building diagnostic outcomes involve the computation of key performance indicators, which enable the automatic identification of energy-saving measures when compared against current building performance databases.

https://doi.org/10.52202/069564-0297

Acknowledgements

This study has been carried out in the context of SmartSPIN project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744. This publication reflects only the authors' views and neither the Agency nor the Commission are responsible for any use that may be made of the information contained therein.

Nomenclature

c	first segment
H	Heavised step function
x	data
y	data

Greek symbols

α	gradient of the first segment
β	change in gradient from the first to second segments
ξ	noise term
ψ	breakpoint position

https://doi.org/10.52202/069564-0297

References

[1] Fraga-De Cal, B.; Garrido-Marijuan, A.; Eguiarte, O.; Arregi, B.; Romero-Amorrortu, A.; Mezzasalma, G.; Ferrarini, G.; Bernardi, A., «Energy Performance Assessment of Innovative Building Solutions Coming from Construction and Demolition Waste Materials», Materials 2021, 14, 1226. https://doi.org/10.3390/ma14051226

[2] «Energy performance of buildings directive». https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en (Accessed on 11th May 2023).

[3] M. Economidou, «Europe's buildings under the microscope», *Building Performance Institute Europe (BPIE)*, otc. 2011, ISBN: 9789491143014

[4] M. Economidou and P. Bertoldi, «Practices to overcome split incentives in the EU building stock», European Council for an Energy Efficient Economy, 2015, ECEEE 2015 Summer Study Proceedings

[5] «ASHRAE Guideline 14-2014: Measurement of Energy, Demand and Water Savings», ASHRAE, 2014

[6] «APMPC ASHRAE Performance Measurement Protocols for Commercial Buildings», ASHRAE, 2010

[7] O. Eguiarte, A. Garrido-Marijuan, R. Garay-Martinez, M. Raud, and I. Hagu, «Data-driven assessment for the supervision of District Heating Networks», *Energy Rep.*, vol. 8, pp. 34-40, dic. 2022, doi: 10.1016/j.egyr.2022.10.212.

[8] A. Garrido-Marijuan, O. Eguiarte, R. Garay-Martinez, M. Raud, and I. Hagu, «Lessons Learnt from Substation Inspection on Low Temperature District Heating Networks», *Environ. Sci. Proc.*, vol. 11, n.º 1, Art. n.º 1, 2021, doi: 10.3390/environsciproc2021011033.

[9] M. Lumbreras *et al.*, «Data driven model for heat load prediction in buildings connected to District Heating by using smart heat meters», *Energy*, vol. 239, p. 122318, ene. 2022, doi: 10.1016/j.energy.2021.122318.

[10] B. Arregi and R. Garay, «Regression analysis of the energy consumption of tertiary buildings», *Energy Procedia*, vol. 122, pp. 9-14, sep. 2017, doi: 10.1016/j.egypro.2017.07.290.

[11] P. Westermann, C. Deb, A. Schlueter, and R. Evins, «Unsupervised learning of energy signatures to identify the heating system and building type using smart meter data», *Appl. Energy*, vol. 264, p. 114715, abr. 2020, doi: 10.1016/j.apenergy.2020.114715.

[12] M. Eriksson, J. Akander, and B. Moshfegh, «Development and validation of energy signature method – Case study on a multi-family building in Sweden before and after deep renovation», *Energy Build.*, vol. 210, p. 109756, mar. 2020, doi: 10.1016/j.enbuild.2020.109756.

[13] L. Belussi and L. Danza, «Method for the prediction of malfunctions of buildings through real energy consumption analysis: Holistic and multidisciplinary approach of Energy Signature», *Energy Build.*, vol. 55, pp. 715-720, dic. 2012, doi: 10.1016/j.enbuild.2012.09.003.

[14] R. Hitchin and I. Knight, «Daily energy consumption signatures and control charts for air-conditioned buildings», *Energy Build.*, vol. 112, pp. 101-109, ene. 2016, doi: 10.1016/j.enbuild.2015.11.059.

[15] FR. Jacobson. «Energy signature and energy monitoring in building energy management systems». *CLIMA 2000 Congress,* Copenhagen, 1985.

[16] G. Nordström, H. Johnsson, and S. Lidelöw, «Using the Energy Signature Method to Estimate the Effective U-Value of Buildings», en *Sustainability in Energy and Buildings*, A. Hakansson, M. Höjer, R. J. Howlett, y L. C. Jain, Eds., en Smart Innovation, Systems and Technologies, vol. 22. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 35-44. doi: 10.1007/978-3-642-36645-1_4.

[17] M. T. Paulus, D. E. Claridge, and C. Culp, «Algorithm for automating the selection of a temperature dependent change point model», Energy Build., vol. 87, pp. 95–104, Jan. 2015, doi: 10.1016/j.enbuild.2014.11.033.

[18] V. Muggeo, «Estimating Regression Models with Unknown Break-Points», *Stat. Med.*, vol. 22, pp. 3055-71, oct. 2003, doi: 10.1002/sim.1545.

[19] C. Pilgrim, «piecewise-regression (aka segmented regression) in Python», *J. Open Source Softw.*, vol. 6, n.º 68, p. 3859, dic. 2021, doi: 10.21105/joss.03859.

[20] S. Seabold and J. Perktold, «Statsmodels: Econometric and Statistical Modeling with Python», presented in Python in Science Conference, Austin, Texas, 2010, pp. 92-96. doi: 10.25080/Majora-92bf1922-011.