Electricity Consumption in Portuguese Households: Unveiling Key Determinants through Surveys and Smart Metering Data

Abstract – The present study provides a deep analysis and characterization of household electricity consumption in mainland Portugal through a cross-analysis of electricity metering data from January 2020 to December 2023 and responses to a 53-question energy survey. A sample of around 2,000 households was evaluated through different statistical approaches (multiple linear regression, machine learning, and hierarchical analysis). Key consumption factors were identified, such as contracted electrical power, net surface area, and presence or absence of electrical equipment. The results provide essential insights for defining targeted approaches and measures for enhancing energy efficiency and improving household energy consumption behaviour.

Keywords: Household electricity consumption, smart metering, energy surveys, energy efficiency and flexibility measures, Portugal.

I. Introduction

To provide a deep and fast energy transition to mitigate the multidimensional impacts of climate change, the electricity sector is a crucial target for fundamental shifts [1], particularly related to the reduction of fossil fuels use, sharp variations in the tariffs, and instabilities in the supply chain, highlighting the need for higher energy efficiency and penetration of renewables, along with a more effective electrification process [2]. In this context, one of the first steps to conduct these deep changes is characterizing electricity consumption profiles and patterns in different sectors while identifying the main drivers associated with variations in consumption values, which are important for defining specific approaches to enhance energy efficiency, applying more energy flexibility practices, and optimizing the integration of distributed renewable systems [3]. As the second-largest electricity-consuming sector in the European Union in 2022, accounting for 708 TWh and surpassed only by the industrial sector [4], conducting this analysis in the residential sector is a key aspect of the current decarbonisation process.

With the emerging scenarios of real-time pricing and an increasing number of on-grid self-consumption systems, the electrical grid needs to incorporate new technologies and facilities to maintain energy supply reliability and security. From this standpoint, smart meters emerge as allies for obtaining information regarding electricity consumption, enabling the identification of patterns and profiles through (near) real-time data monitoring [5]. They are already widely applied in European countries such as Portugal (80% of the low-voltage consumers in December 2023) [6]. This pattern mining process between households is important to identify groups and variables that may strongly impact electricity consumption in a household environment, mainly when information about behavioural conditions, dwelling characteristics, occupant aspects, and electrical equipment is available [3].

Under these circumstances, energy surveys can be considered important instruments for obtaining a deeper understanding of smart metering data, adding value to a more comprehensive understanding. This cross-analysis can lead to a more detailed segmentation of households and identification of possible main factors that determine consumption values [7].

Therefore, this study aims to deepen our understanding of household electricity consumption in Portugal and analyse the most significant drivers and their associated impacts on electricity consumption. Section II presents the utilized methodology for the descriptive analysis of the survey responses, calculation of electricity consumption, and importance analysis of the factors and predictor variables, while in Section III we explore the results and respective discussions. Finally, Section IV shows the main conclusions of the study.

II. Methodology

Using different approaches (multiple linear regression, hierarchical importance, and Random Forest methods), the methods were applied to the responses of a 53-question energy survey and the associated daily electricity smart metering data of around 2,000 Portuguese households in mainland Portugal, an important initial step for defining future adapted energy efficiency and flexibility strategies, storage facilities, and consumption reduction.

The energy data regarding energy surveys and electricity consumption were obtained through online responses from 1,985 households with smart metering facilities and clients from the regulated Portuguese electricity service provider. A total of 53 questions in the developed survey can be divided into questions about household electrical appliances (14), heating and cooling systems (9), dwelling characteristics (7), electrical parameters of the installation (3), inhabitants' profile (3), typical household occupancy (3), windows (3), thermal insulation (2), domestic hot water equipment (2), food preparation (2), housing

979-8-3315-1278-1/25/$31.00 ©2025

location (1), and general questions (4). Smart metering data for the survey participants were collected from January 2020 to December 2023.

This electricity consumption cross-analysis can be divided into three main parts: Part A (descriptive analysis of survey responses), Part B (treatment and analysis of electricity consumption), and Part C (inferential statistical analysis between survey variables and consumption data). All the analysis were conducted using the *R software* and *Microsoft PowerBI*.

A. *Part A: Descriptive analysis of survey responses*

The first step of Part A of the study was a data trimming process, which aimed to remove consumers who lacked electricity consumption data. Since survey responses will be part of a cross-analysis with the respective smart metering data, participants without available consumption information have a negligible impact on further analysis. This process resulted in the exclusion of 19 participants (1,966 remaining).

Then a descriptive analysis of the remaining participants' survey responses was conducted. The focus was on obtaining statistical data through the cross-analysis of different variables to identify and analyse significant correlations.

B. *Part B: Treatment and analysis of electricity consumption*

A total of 402,135 daily meter readings associated with the remaining 1,966 participants were analysed through CSV files. As previously done in Part A, it was identified the need to provide a data trimming process in Part B, due to four main complications with the obtained smart metering readings: a) consecutive non-cumulative readings; b) reading value discrepancies related to electricity meter exchanges (identified between the other readings with the “code 21”, defined by the electricity service provider); c) null readings between cumulative readings; and d) consecutive repeated readings.

It is important to note that, for most of the participants, it was available only some daily readings per month, where the lack of data for all days of the analysed period or even many months without any reading values were critical challenges for the analysis. In this context, 2023 was the year with the highest number of non-null and cumulative readings.

In this context, the data trimming process was conducted based on the following steps:

- Exclusion of participants with only null readings, readings with the same value, or readings that do not allow calculating even one annual value (participants with only two reading values where one is null): 65 removed participants, 1,901 remaining.
- Removal of non-cumulative readings (analysis of the following reading and identification of a possible new non-cumulative reading).
- Normalization of readings associated with meter exchanges (obtention of cumulative readings).
- Assignment of a single valid reading per month per participant (oldest or only valid reading of the month).
- Calculation of the annual consumption value, based on the subtraction of the reading value of the next analysed year and the current analysed year, both for the same month.

Aiming to remove the less possible number of participants of the study in a scenario of considerable lack of reading data for different months, it was defined that the annual base was the most reliable and adapted form of calculating electricity consumption values, defining three periods of analysis (P1: 2020-2021; P2: 2021-2022, and P3: 2022-2023). It was possible to calculate, at least for one period of analysis, an annual electricity consumption for the 1,901 participants.

C. *Part C: Inferential statistical analysis between survey variables and consumption data*

The last part of the study aimed at identifying the main factors that significantly affect household electricity consumption and estimating their impact.

The starting point was to verify the pattern of the annual electricity consumption results obtained for each participant. As a result, it was observed that the consumption data were not normalized, exhibiting a skewed distribution, thus a logarithmic normalization was conducted. Subsequently, the analysis focused on evaluating possible linear relations between the predictor variables and its factors (namely the questions and possible responses of the survey, respectively) and the response variable (annual electricity consumption). Similar studies suggested multiple linear regression approaches for comparable scenarios – multiple linear regression models employed for electricity consumption forecasting in Italy [8], and for annual electricity and energy consumption prevision in the United Kingdom [9]. Briefly, multiple linear regression is a method for analysing linear relations between one response variable and more than one predictor variable [10]. In this step of Part C, this method was applied to investigate possible linear relations of the predictor variables and relations with higher or lower electricity consumption values.

Given the study’s numerous predictor variables and the multidimensional nature of the factors affecting household electricity consumption, another article review was conducted to identify methods for capturing nonlinear relations between the variables. Machine learning techniques, such as the Random Forest method, were applied to similar studies in the literature, such as electricity tariffs previsions and price trends of the energy market [11], and the creation of statistical models for characterising and forecasting electricity consumption in high urban density areas [12]. This method is based on creating decision trees, training and testing the model according to the

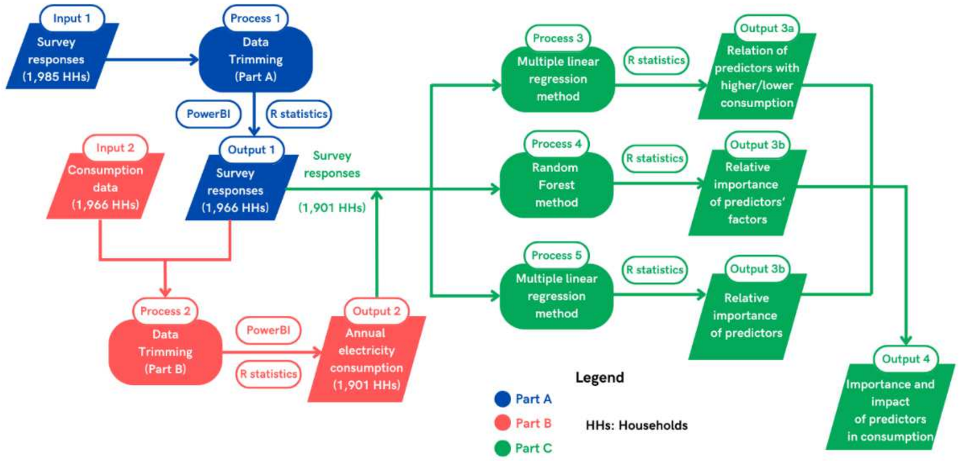

Figure 1: Flowchart of the utilized methodology.

database, and providing the most significant predictor variables in the model about the response variable [13]. In the present study, the Random Forest method was applied to obtain the relative importance of the predictors' factors (responses to the survey questions) in household electricity consumption, through testing and training the model based on the database of survey responses and annual electricity consumption, defining the most significant factor of the predictor variables in the response variable, and calculating the importance of the other factors in relation to the most significant one.

A third method was analysed: the hierarchical analysis of variance variation can identify the most important predictor variables in the response variable, already applied in similar studies [14]. The results of the hierarchical analysis were combined with the results of relative importance obtained for the predictors' factors through the Random Forest method and their respective impact (direct or indirect) of each factor on the electricity consumption obtained through the multiple linear regression approach, leading to a deep understanding of the response variable according to each of the predictor variables and its respective factors. A flowchart representing the three parts of the utilized methodology can be seen in Figure 1.

III. Results and Discussion

In the present session, the obtained results of all three parts of conducted study will be presented. According to the NUTS II regions of mainland Portugal [15], 22.5% of the respondents were in Centro, followed by 21.0% in Alentejo, 21.0% in the Lisbon Metropolitan Area, 20.3% in Norte, and 15.2% in the Algarve – in comparison with the populational distribution for 2023, 37.26% of the Portuguese residents live in the Lisbon Metropolitan Area, 36.2% in Norte, 16.7% in Centro, 4.8% in Algarve, and 4.7% in Alentejo [16].

In In addition, the majority of the participants contracted electrical power between 3.45 kVA and 6.90 kVA (45.4%), followed by contracted power values between 10.35 kVA and 20.70 kVA (27.6%), between 1.15 kVA and 2.30 kVA (25.8%), and between 27.60 kVA and 41.40 kVA (only 1.2% of the participants). From the descriptive analysis of the survey responses of 1,966 participants, the most relevant results obtained were:

a) Inhabitants' profile: 1,088 participants (55.3%) reported having at least one person over the age of 65 in their household, along with only 135 respondents (6.9%) that affirmed having one person requiring special care or having medical conditions that necessitate the use of electrical medical equipment in the habitation.

b) Dwelling characteristics: Most of the participant dwellings were constructed between 1991 and 2005 (636 respondents, 32.4%), while most of the houses had not undergone a rehabilitation process (889 respondents, 45.2%). It was also observed that there were more house units than apartments (1,105 houses, 56.2%), along with the average usable area of the habitations between 50m^2 and 100m^2 (724 respondents, 36.8%). During business days, most dwellings have day and night occupation (1,501, 76.4%), and the same pattern is observed on weekends (1,786, 90.8%).

c) Thermal insulation: Most respondents (1,002, 51.0%) do not have thermal insulation in their walls. The same pattern was observed for thermal insulation on the roof (942, 47.9%).

d) Windows: The greater part of the participants affirm having aluminium windows without thermal break (1,068, 54.3%). In addition, most of the participant dwellings have double-glazed windows (1,150, 58.5%) with solar protections (1,713, 87.1%).

e) Domestic hot water equipment: The three most utilized domestic hot water equipment were water heaters (1,108, 56.4% of the households), storage water heaters (325, 16.53% of the households), and boilers (254, 12.9%). Regarding the type of energy used for this purpose, nearly 60% (1,158, 59.2%) of all participant households affirmed that they use natural gas.

f) Heating and cooling systems: A substantial number of respondents confirmed that they use climatization systems exclusively for heating (940, 47.8%). January, December, and February are the months with the most frequent use of heating systems. In opposite directions, June, September, and July are the months with the most frequent use of cooling systems. The most used energy for climatization among participants was electricity (1,203, 63.5%), with 53% of households having between 2 and 5 climatization equipment.

g) Household electrical appliances: The majority of participants have only one refrigerator (1,540, 78.3%), and most are aged between 5 and 10 years. Regarding chest freezers, most respondents do not have one at home (1,020, 51.9%). Most households have one dishwasher (1,280, 65.1%), and most of that equipment is between 5 and 10 years of age. A similar pattern is observed for washing machines (nearly 84% of all households have one equipment, commonly aged between 5 and 10 years).

h) General questions: A significant proportion of households (1,756, 89.3%) do not have installed photovoltaic systems for self-consumption. In addition, nearly all respondents do not charge an electric vehicle at home.

Regarding the analysis of the electricity consumption data, as previously mentioned, it was possible to obtain at least one yearly electricity consumption value for 1,901 households. The number of households for which we were able to get an annual consumption value per period is presented in Table I.

TABLE I. NUMBER OF HOUSEHOLDS WITH AN ANNUAL ELECTRICITY CONSUMPTION VALUE.

Period of Analysis	Nº of households with an annual electricity consumption	%
P1 (2020-2021)	1,608 households	84.6%
P2 (2021-2022)	1,752 households	92.2%
P3 (2022-2023)	1,877 households	98.7%

Table II presents the national annual results, while the regional values according to NUTS II regions are presented in Table III. According to a study published in July 2021 by the National Institute of Statistics, a typical Portuguese household consumes about 3,300 kWh/year of electricity [17]. Therefore, our sample's values align with the national average values, highlighting the potential relevance of the obtained results. The observed pattern of values slightly below the expected may be associated with inconsistent readings – non-cumulative or null – excluded during the data trimming phases and disturbances resulting from normalizing readings related to meter exchanges.

Concerning the inferential statistical analysis, the combination of the three utilized methods – multiple linear regression, the Random Forest method, and the hierarchical analysis of variance variation – resulted in the identification of the most significant factors and their respective relation with electricity consumption values, as well as the most important predictor variables on the variation of the variance.

TABLE II. NATIONAL AVERAGE ANNUAL ELECTRICITY CONSUMPTION

Period of Analysis	Average annual electricity consumption (kWh/year)
P1 (2020-2021)	3,070
P2 (2021-2022)	2,934
P3 (2022-2023)	2,874

TABLE III. CALCULATED REGIONAL AVERAGE ANNUAL ELECTRICITY CONSUMPTION PER PERIOD OF ANALYSIS.

Period of Analysis	Average annual electricity consumption (kWh/year)				
	Alentejo	Algarve	Centro	Lisbon	Norte
P1 (2020-2021)	3,214	3,223	2,865	3,007	3,513
P2 (2021-2022)	2,994	3,078	2,439	2,856	2,829
P3 (2022-2023)	2,773	2,864	2,297	2,680	2,626

It is important to note that the results of importance obtained through the Random Forest method are relative, so the most important factor is 100% importance, and the subsequent factors are relative importance as a percentage concerning the most significant one. Table IV to Table VI presents the results of the most important factors (the Random Forest method) and their respective impact (multiple linear regression).

TABLE IV. RELATIVE IMPORTANCE AND IMPACT IN CONSUMPTION FOR THE PERIOD OF ANALYSIS P1 (2020-2021).

Factor	Relative Importance (%)	Relation with consumption
Contracted electrical power	100.00	Direct
Usable area	13.65	Direct
Absence of dishwasher	4.35	Inverse
Natural gas (food preparation)	3.18	Inverse
Absence of chest freezer	2.09	Direct
Isolated house unit	1.97	Direct
Absence of washing machine	1.71	Inverse
Inefficient window solution*	1.63	Inverse
Absence of drying machine	1.59	Direct
Standalone stove	1.56	Inverse

* Inefficient window solution (0-1 total points), where wooden or aluminium windows (0 points), aluminium with thermal break or polyvinyl chloride (PVC) (1 point); simple glass (0 points), double glass (1 point); absence (0 points) or presence of solar protection (1 point).

Related to the results obtained through the combination of the Random Forest method and the hierarchical analysis, it was observed that the most important predictor variables, for all the three analysed periods (P1, P2, and P3) are, respectively, contracted electrical power, usable area, position of the dwelling, presence of elderly persons, presence of dishwasher, type of energy for food preparation, and presence of a person requiring special care or having medical conditions that necessitate the use of electrical equipment in the household.

TABLE V. RELATIVE IMPORTANCE AND IMPACT IN CONSUMPTION FOR THE PERIOD OF ANALYSIS P2 (2021-2022).

Factor	Relative Importance (%)	Relation with consumption
Contracted electrical power	100.00	Direct
Usable area	11.83	Direct
Natural gas (food preparation)	3.85	Inverse
Absence of dishwasher	3.01	Inverse
Absence of chest freezer	1.99	Direct
Isolated house unit	1.91	Direct
Cooktop or oven	1.81	Direct
Absence of washing machine	1.78	Inverse
Water heater (natural gas)	1.73	Direct
Inefficient window solution*	1.48	Inverse

TABLE VI. RELATIVE IMPORTANCE AND IMPACT IN CONSUMPTION FOR THE PERIOD OF ANALYSIS P3 (2022-2023).

Factor	Relative Importance (%)	Relation with consumption
Contracted electrical power	100.00	Direct
Usable area	11.85	Direct
Absence of dishwasher	2.99	Inverse
Natural gas (food preparation)	2.95	Inverse
Cooktop or oven	2.88	Direct
Absence of chest freezer	2.48	Direct
Isolated house unit	2.03	Direct
Water heater (natural gas)	1.97	Direct
Absence of washing machine	1.86	Inverse
Standalone stove	1.85	Inverse

In summary, it becomes clear that almost the same factors have the highest levels of relative importance for all the three analysed periods, with a particular highlight to the contracted electrical power and usable area – logically, higher contracted power is strongly linked to higher installed capacity and, logically, more electricity consumption, as well as larger house units have strong connections with more occupants and more electrical equipment, leading to higher consumption levels. Strong relations between lower electricity consumption levels and usage of natural gas for food preparation are also presented in opposite directions of households using induction cooktops and electric stoves. The obtained results also confirm the absence of equipment linked to considerable electricity consumption levels being connected to lower consumption.

IV. CONCLUSIONS

Through this study, it is possible to understand the most significant drivers related to higher or lower values of household electricity consumption, highlighting the importance of smart metering facilities and characterization of consumption profiles as a first step for defining targeted approaches to enhance energy efficiency and avoid irrational consumption.

REFERENCES

[1] Gupta, A., Davis, M., & Kumar, A. (2021). An integrated assessment framework for the decarbonization of the electricity generation sector. Applied Energy, 288, 116634. https://doi.org/10.1016/j.apenergy.2021.116634

[2] Tian, X., An, C., & Chen, Z. (2023). The role of clean energy in achieving decarbonization of electricity generation, transportation, and heating sectors by 2050: A meta-analysis review. Renewable and Sustainable Energy Reviews, 182, 113404. https://doi.org/10.1016/j.rser.2023.113404

[3] Kim, H., Park, S., & Kim, S. (2023). Time-series clustering and forecasting household electricity demand using smart meter data. Energy Reports, 9, 4111–4121. https://doi.org/10.1016/j.egyr.2023.03.042

[4] European Commission. (n.d.). Electricity and heat statistics. Eurostat. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Electricity_and_heat_statistics (accessed on 5 February 2025).

[5] Batalla-Bejerano, J., Trujillo-Baute, E., & Villa-Arrieta, M. (2020). Smart meters and consumer behaviour: Insights from the empirical literature. Energy Policy, 144, 111610. https://doi.org/10.1016/j.enpol.2020.111610

[6] E-redes. (2023, December 12). *E-redes reaches 5 million integrated smart meters*. E-redes. Available at: https://www.e-redes.pt/en/news/2023/12/12/e-redes-reaches-5-million-integrated-smart-meters (accessed on 5 February 2025).

[7] Gouveia, J. P., Seixas, J., & Long, G. (2018). Mining households' energy data to disclose fuel poverty: Lessons for Southern Europe. Journal of Cleaner Production, 178, 534–550. https://doi.org/10.1016/j.jclepro.2018.01.021

[8] Bianco, V., Manca, O., & Nardini, S. (2013). Linear Regression Models to Forecast Electricity Consumption in Italy. Energy Sources, Part B: Economics, Planning, and Policy, 8(1), 86–93. https://doi.org/10.1080/15567240903289549

[9] Liu, Y., & Li, J. (2022). Annual Electricity and Energy Consumption Forecasting for the UK Based on Back Propagation Neural Network, Multiple Linear Regression, and Least Square Support Vector Machine. Processes, 11(1), 44. https://doi.org/10.3390/pr11010044

[10] Uyanık, G. K., & Güler, N. (2013). A Study on Multiple Linear Regression Analysis. Procedia - Social and Behavioral Sciences, 106, 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

[11] González, C., Mira, J. M., & Ojeda, J. A. (2016). Applying Multi-Output Random Forest Models to Electricity Price Forecast. https://doi.org/10.20944/preprints201609.0053.v1

[12] Pesantez, J. E., Li, B., Lee, C., Zhao, Z., Butala, M., & Stillwell, A. S. (2023). A Comparison Study of Predictive Models for Electricity Demand in a Diverse Urban Environment. Energy, 283, 129142. https://doi.org/10.1016/j.energy.2023.129142

[13] Jiang, X., Abdel-Aty, M., Hu, J., & Lee, J. (2016). Investigating macro-level hotzone identification and variable importance using big data: A random forest models approach. Neurocomputing, 181, 53–63. https://doi.org/10.1016/j.neucom.2015.08.097

[14] Lim, H., Kim, S., Kim, Y., & Son, S. (2021). Relative Importance Analysis of Safety Climate Evaluation Factors Using Analytical Hierarchical Process (AHP). Sustainability, 13(8), 4212. https://doi.org/10.3390/su13084212

[15] Medeiros, E., Valente, B., Gonçalves, V., & Castro, P. (2022). How Impactful Are Public Policies on Environmental Sustainability? Debating the Portuguese Case of PO SEUR 2014–2020. Sustainability, 14(13), 7917. https://doi.org/10.3390/su14137917

[16] Instituto Nacional de Estatística, I.P. (2024). Estimativas Provisórias de População Residente - estimativas pós-censitárias assentes nos resultados dos Censos 2021 [Provisional Estimates of Resident Population - Post-Census Estimates Based on the Results of the 2021 Census]. Available at: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&contecto=pi&indOcorrCod=0012903&selTab=tab0 (accessed on 9 February 2025).

[17] Instituto Nacional de Estatística, I.P. (2021). Inquérito ao Consumo de Energia no Sector Doméstico - 2020. Instituto Nacional de Estatística. [Survey on Energy Consumption in Households - 2020. Statistics Portugal]. Available at: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=512530049&DESTAQUESmodo=2&xlang=pt (accessed on 6 February 2025).