Data analysis approach for characterizing residential energy consumption based on statistics of household appliances ownership

Juan P. Chavat and Sergio Nesmachnow

Universidad de la República, Uruguay

E-mail: {juan.pablo.chavat,sergion}@fing.edu.uy;

Abstract. Worldwide, residential electricity demand has increased constantly, expecting to double in 2050 the demand of 2010. Different policies have been proposed to achieve a smart use of electricity. This article presents a data-analysis approach to evaluate the potential household electricity consumption from statistical data. The main axis of the study are statistics of appliance ownership and information of the appliance characteristics, gathered from census surveys and local shops. An index to estimate the electricity consumption is performed. The validation of the proposed index is carried out using real consumption data from the Electricity Consumption Data set of Uruguay and Ordinary Least Square linear regressions. Jupyter notebooks, Python language and well-know libraries such as Pandas and Numpy were used during the implementation. The main results show that administrative regions located on the West/Southwest coastlines present the highest index scores. In turn, census sections/segments on the West/Southwest coastlines of Montevideo performed the highest scores while the lowest scores can be found at the outskirts of the city. The proposed methodology can be applied for electricity consumption estimation in other regions/countries where census data is publicly available.

1. Introduction

Residential electricity demand has increased constantly worldwide, and it is expected to double by 2050 the demand recorded in 2010 [1]. For that reason, several investigations have been carried out with the main goal of applying policies that motivate customers to save energy and reduce the climate impact in factories, buildings and homes [2]. Information technologies play a major role for properly managing electricity demand and consumption at different operation levels [3, 4, 5].

Uruguay is a country with 3.4 million inhabitants, where electricity is provided by a state-owned company, *Administración Nacional de Usinas y Trasmisiones Eléctricas* (UTE). By 2020, electricity was provided to almost 1.5 million customers (e.g., households, commerce, and industries), 90.5% of residential type. Almost half of the inhabitants live in the capital city, Montevideo. According to 2018 information, electrification achieves a rate of 99.8% in the city, with an average consumption of 246 kWh per residential customer.

Techniques that aims to have better use of the energy resources, such as demand management, are based on energy consumption analysis and characterization of the use. A possible approach to these techniques is to encourage behavioural changes in the customers, resulting in savings. Data analysis and the consumption characterization provides accurate information that can be used in the construction of policies, plans and tariffs for promoting behavioural changes.

This article extends the short communication presented at the 1st International Workshop on Advanced Information and Computation Technologies and Systems [6]. The research was developed within the evaluation of the potential residential electricity consumption using a data-based approach to analyze statistics of appliances ownership for different census areas, and information about the characteristics of the appliances [7]. For the validation, the application of Ordinary Least Square (OLS) linear regressions are proposed for case studies in Uruguay and Montevideo, and updated data from the National Household Continuous Survey (ECH) gathered in 2019 is used.

The research is carried out in the context of the project "Computational intelligence to characterize the use of electric energy in residential customers", funded by the National Administration of Power Plants and Electrical Transmissions (Spanish: Administración Nacional de Usinas y Trasmisiones Eléctricas, UTE), and Universidad de la República, Uruguay. The project aims to characterize the different uses of electricity by the residences in Uruguay, by means of computational intelligence and data analytics techniques, and identify the most consuming-impactful appliances and its use patterns in the consumption.

2. Analysis of residential electricity consumption
The analysis is based on appliance ownership information from a national survey by National Statistics Institute (INE), Uruguay, and appliance characteristics information, collected from local shops with a presence on the Internet.

2.1. Question and hypothesis
The work originates from the formulation of the following question: can an index build from appliance ownership statistics model the electricity consumption per area?

Energy-intensive appliances determines, greatly, the final consumption of a household. For example, the presence of air conditioners and electric water heater represent a large part of the final consumption, but are not present in all households. If the ownership of these appliances is quantified by census areas, the potential electricity consumption can be calculated. Thus, this work formulates the follow hypothesis: The most energy-intensive appliances owned, the higher the potential electricity consumption.

2.2. Related work
The related literature brings several possible approaches for the energy consumption analysis, applied in different countries. Many of the works apply statistical tools, such as different type of regressions, and uses information provided by surveys.

Chévez et al. [8], applied clustering to organize 1010 census areas of Gran La Plata, Argentina, into eight groups to analyze energy consumption similarities. The k-means clustering algorithm was applied. The resulting groups were related to sociodemographic variables. The study concluded that electricity demand increased with the number of people per home and per room, that areas with a large number of apartments had lesser electricity consumption, and, instead, the more precarious buildings in the area, the higher the electricity consumption (the less basic needs are met, the higher consumption). In addition, two relevant problems in the electricity sector of the country were identified: i) consumption peaks cannot be satisfied, ii) a poor diversification of the electricity matrix.

Electricity consumption of 3941 households from Ireland was related to socio-economic and demographic variables, and dwelling characteristics, to be analyzed [9]. The analysis concluded that houses with more bedrooms, presence of 36–55 yeas old people, and/or presence of professionals, had higher electricity consumption. Instead, apartments or lower/middle social class households had lower electricity consumption. Households using electricity for cooking or water heating had higher consumption than the rest, and the time of maximum consumption was during the morning for older household response person (HRP) or late in the day for middle-age HRP. Later [10], authors proposed inferring household characteristics from the electricity consumption data, processed using multilevel and logistic regressions. Results showed an accuracy of 60% on the classification of the employment status of the HRP and the feasibility of the approach to infer useful social information.

A study of residential electricity consumption in Brazil during 1985-2013 [11] concluded that the most demanding appliances in the country are the electric shower (19%), refrigerator (28%), lamps (15%), TV (11%), air conditioner and freezer (5% each one). Using linear regressions, elasticity values were obtained from a bunch of explanatory variables (such as the number of households in the country, family income, electricity tariff, etc) and then related to consumption behaviours. Results show that electricity consumption increase 1.53% with a rise of 1% in the number of residences, 0.19% with an increment of 1% in the families income, instead it decreases 0.23% with an increment of 1% in the tariff prices. The models showed a strong relationship between explanatory variables and electricity consumption.

In Uruguay, an extensive work that analyzes the residential electricity consumption based on socioeconomic and dwelling characteristics, energy uses and temperature, was carried on using data of 2994 households [12]. General conclusions stats that owning certain appliances, such as electric water heater and/or air conditioner, directly impacts over the electricity consumption, and thermal comfort appliances are more frequent in households with high electricity consumption. After applying OLS regressions and Quantile Regression (QR), it is concluded, although the income per capita is an influential variable over the electricity consumption, other variables, such as the family composition, dwelling characteristics, or energy uses, must be taken into account.

The analysis of related works allowed finding several published research processing statistics variables and electricity consumption, to extract valuable information about their relationships. In this line of work, this article presents an analysis relating appliance ownership variables to elaborate an index of electricity consumption per areas.

3. The proposed approach for electricity consumption analysis

The approach consists of analyzing electricity consumption through an index based on statistical information about appliances ownership per home. Statistics are obtained from periodical surveys (such as the ECH survey in Uruguay), considering variables that quantify the appliance ownership, the number of residents per home, and household georeferencing data.

Given a type of census area r with m areas, the likelihood of owning each appliance is a matrix $\mathbf{A}^{(r)} \in \mathbb{R}^{mxn}$. Two vectors are computed: $\overrightarrow{c^{(r)}} \in \mathbb{R}^n$ with the mean consumption of each appliance, and $index^{(r)} \in \mathbb{R}^m$ using the frequency of use of each appliance. Using these three elements, the index is calculated as described in Equation 1.

$$index^{(r)} = \mathbf{A}_{m,n}^{(r)} \cdot \overrightarrow{c^{(r)}} \cdot \overrightarrow{f^{(r)}} \cdot \overrightarrow{1} \tag{1}$$

4. Validation of the proposed index: estimation of electricity consumption in Uruguay and Montevideo

A relevant case study is analyzed to validate the proposed index: estimation of electricity consumption in Uruguay and its capital city Montevideo. This is a relevant case study, since electrification achieves a rate of 99.9% in Uruguay, with an average consumption of 246 kWh per household. Furthermore, few works have analyzed residential electricity consumption in Uruguay.

Three datasets were used for the building and validating the index:

(i) *Census data.* The collected information is georeferenced in three levels: departments (Uruguay), and census section, and census segments (Montevideo). Information about appliance ownership, georeferentiation and number of rooms were used.

(ii) *Appliances data*, collected from local shops. Utilization frequency and mean power consumption were properly weighted to determine the effective consumption.

(iii) *Real electricity consumption*, as reported in the Electricity Consumption Data set of UruguaY (ECD-UY) dataset [13] for the total household consumption subset.

5. Data sets and data processing

This section describes the three data sets used for the building and validation of the index.

5.1. Census data

Census data is provided yearly by INE in the format of the ECH, that collects information in several areas from a statistical representative set of households around the country. The collected information is georeferenced in at least three levels: departments, census section and census segments. For the index, only the information about appliance ownership, georeferentiation and number of rooms were used. Due to census sections and segments data are present only for households in Montevideo, the evaluation of the index at this area levels was limited to this city.

The process of preparing the data for the index consisted of multiple steps. First, the columns with Yes/No values were transformed to 0, 1 values to facilitate the multiplications, and the number of different type of notebooks was merged to a single sum. Then, the number of lights were multiplied by the number of rooms, the air conditioner variable was separated from the variable that indicates another electric heating source, and all the ones that indicate the presence of an appliance was multiplied by its corresponding quantity variable. Finally, integrity validations were performed, where, for example, the variables that quantify an appliance were validated to be greater or equal to zero. No integrity errors were found.

5.2. Appliances data

Based on the set of 17 appliances surveyed by the ECH, appliance consumption was collected from local shops with a presence on the Internet. Five different models of each appliance were used to calculate the mean power consumption of each one. Of the 17 appliances, the most energy-intensive are clothes dryer (3154.0 W), shower heater (1810.0 W) and electric water heater (1600.0 W), while the less energy-intensive are DVD/VHS players (10.3 W), lightning (11.8 W) and radio (20.2 W). Since the use intensity affects directly to the consumption of an appliance and based on the authors own experience, a frequency of use was assigned to each one and the mean power was finally weighted by it.

5.3. Real electricity consumption

ECD-UY dataset provides three subsets with real electricity consumption from Uruguayan residences. In this article, aggregate consumption from the *total household consumption* subset is considered. Records were obtained from smart-meters georeferenced and installed in customers

houses (distributed on the main Uruguayan cities) by UTE. The data preparation consisted in filtering the records by the follow conditions: i) records of consumption during the year 2019 (coinciding with the census year), ii) customers that counts with at least 95 % of the corresponding records ($365{\times}24{\times}4$), for the days of the year, hours of a day, and records per hour respectively), and iii) monthly average consumption lower than 5000 kWh, assuming that larger consumption values do not corresponds to a household (e.g., commercial or industrial instead). Filtered data include the monthly average consumption of 8874 customers, distributed on the 19 departments, and, in Montevideo on 25 census sections and 464 census segments.

6. Implementation
The implementation consisted of loading the datasets, matrices/vectors construction for appliance ownership and power demand, processing the index, visualization of results and validation. Data cleansing and monthly average consumption were processed in the National Supercomputing Center of Uruguay (Cluster-UY) [14]. The code was implemented in Jupyter notebooks using Python language and libraries (Pandas, Numpy, Matplotlib and statsmodels). The resulting notebooks and scripts are available for download at `https://bit.ly/3qnoIeC`.

7. Results
Index results at the department level show a difference of up to 1.5 times between the lowest and highest scores. The top-three scored departments are Montevideo (4599), Salto (4156) and Colonia (4155), while the bottom-thee departments are Cerro Largo (3060), Treinta y Tres (3358) and Rocha (3387). Geographically, the departments with higher scores are located on the South/West coastline and progressively decrease towards the Northeast of the country.

In Montevideo, the difference between the lowest and highest scores is 1.5 for sections and 8 for segments. Highest score areas are located over the Southeast coastlines and the further from the Southeast coastline its the area, the lower the resulting score. The section with a higher score (5561) is located in neighbourhoods Carrasco, Malvin, Buceo, Union, Malvin Norte, Punta Gorda, Las Canteras and Carrasco Norte. On the other hand, the section with a lower score (3697) is located in neighbourhoods Ituzaingo, Manga, Toledo Chico, Villa Garcia, Manga Rural, Punta de Rieles, Bella Italia, Villa Española, Manga, Jardines del Hipódromo, Flor de Maroñas, Piedras Blancas and Unión. The segments with highest score (10435) is located in Carrasco neighbourhood, while the one with lowest (1305) is located in neighbourhoods Tres Ombúes and Victoria. Departments, segments and section results are presented in Figure 1.

Monthly average consumption per census area was calculated and then used together with index results in OLS linear regressions to validate the correlation between them (considering the R-squared metric). In the three evaluated areas, the distribution of the georeferenced customers is not homogeneous, resulting in problems of data representation in some sub-areas, and therefore inaccurate average monthly consumption.

Several OLS validations were performed, considering different number of sub-areas, depending on a minimum number of customers data available. For departments, OLS was performed considering the departments with a minimum of customers (from 1 to 350). Many anomalies were detected for some departments. The range of R-squared results was from 0.24 (minimum of one customer, 19 departments) to 0.94 (minimum of 350 customers, 5 departments). A similar behavior was observed for the section level, which was performed considering a minimum of one customer (25 sections), resulting in an R-squared of 0.01, to 300 customers (4 sections), resulting in an R-squared of 0.37. Less accurate results were obtained for segments, due to the lower concentration of customers per census segment. OLS was performed considering segments with a minimum of one customer (464 segments), resulting in an R-squared of 0.0, to 25 customers (18 segments), resulting in R-squared of 0.11.

(a) by departments

(b) by sections

(c) by segments
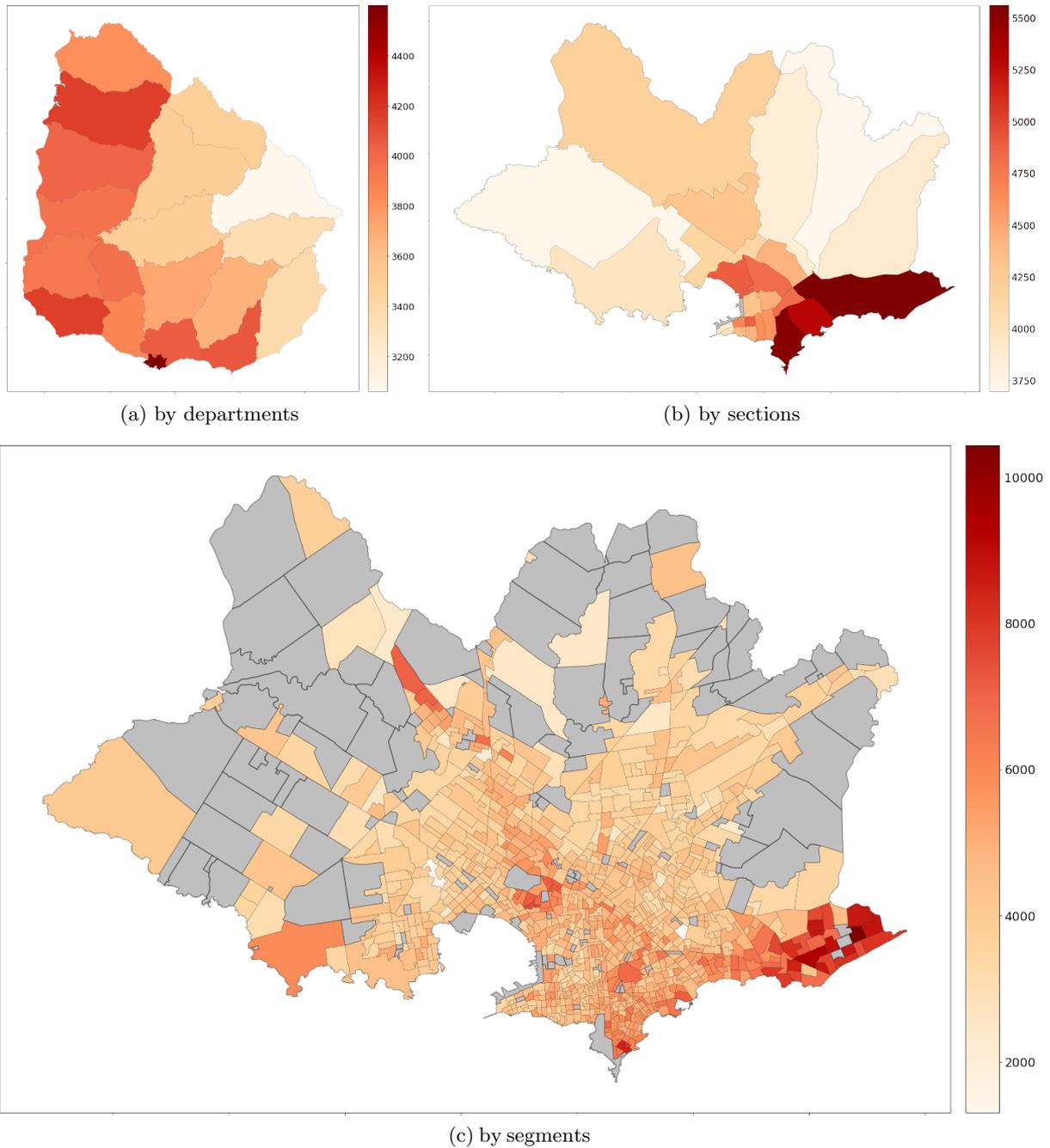
Figure 1. Index score calculated for three census areas

8. Conclusion and future work

This article presented a methodology applying a data-analysis approach for estimating electricity consumption based on publicly available statistics on the ownership of household appliances. Several data sources (census data, appliance characteristics data, and real consumption data) were reviewed and used to define an index for the electricity consumption estimation.

The index was validated on a case study in Uruguay, using data form the ECH survey from year 2019. The proposed index was processed for three census areas in Uruguay and Montevideo.

Results showed that, at department level, Southern/Western coastlines concentrate the most consuming departments in Uruguay, while at section/segment level the most consuming areas in Montevideo are on the Southeast coastline. Index score at segments level showed a difference of up to 8 times between the highest and lowest areas. OLS linear regressions were performed to validate the results. Despite the limited georeferenced real consumption data available, validation instances considering sub-areas with a high number of customers result in higher R-squared values, showing a tendency of correlation between the index values and the real energy consumption. The proposed index is also a useful method to detect anomalies in energy consumption records.

The main lines for future work include studying sociodemographic variables (e.g., income, family size, etc), analyzing the index evolution based on previous ECH surveys and collecting enough data to validate the tendency showed by the R-squared results.

References

[1] Larcher D and Tarascon J 2015 *Nature Chemistry* **7** 19–29
[2] Ford R 2009 *Reducing domestic energy consumption through behaviour modification* Ph.D. thesis Oxford University
[3] Luján E, Otero A, Valenzuela S, Mocskos E, Steffenel L and Nesmachnow S 2019 *Revista Facultad de Ingeniería Universidad de Antioquia*
[4] Orsi E and Nesmachnow S 2017 Smart home energy planning using IoT and the cloud *IEEE URUCON*
[5] Chavat J, Nesmachnow S and Graneri J 2020 *Revista Facultad de Ingeniería Universidad de Antioquia*
[6] Chavat J and Nesmachnow S 2020 Elaboration of an index of electricity consumption based on statistics of household appliances ownership 1st *International Workshop on Advanced Information and Computation Technologies and Systems*
[7] Chavat J and Nesmachnow S 2020 Analysis of residential electricity consumption by areas in Uruguay *Smart Cities* (*Communications in Computer and Information Science* vol 1359) (Springer) pp 42–57
[8] Chévez P, Barbero D, Martini I and Discoli C 2017 *Sustainable Cities and Society* **32** 115–129 ISSN 22106707
[9] McLoughlin F, Duffy A and Conlon M 2012 *Energy and Buildings* **48** 240–248 ISSN 03787788
[10] Anderson B, Lin S, Newing A, Bahaj A B and James P 2017 *Computers, Environment and Urban Systems* **63** 58–67 ISSN 01989715
[11] Villareal M and Moreira J 2016 *Energy Policy* **96** 251–259
[12] Laureiro P 2018 *Serie Documentos de investigación estudiantil, Universidad de la República, Uruguay*
[13] Chavat J, Graneri J, Alvez G and Nesmachnow S 2020 *Scientific Data* (submitted)
[14] Nesmachnow S and Iturriaga S 2019 Cluster-UY: Collaborative Scientific High Performance Computing in Uruguay *Communications in Computer and Information Science* (Springer) pp 188–202