IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 30, 2021, accepted June 11, 2021, date of publication June 15, 2021, date of current version June 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3089443

Monthly Electricity Demand Patterns and Their Relationship With the Economic Sector and Geographic Location

JOAQUIN LUQUE, (Senior Member, IEEE), ENRIQUE PERSONAL, (Member, IEEE), ANTONIO GARCIA-DELGADO, AND CARLOS LEON, (Senior Member, IEEE)
Departamento de Tecnología Electrónica, Universidad de Sevilla, 41011 Sevilla, Spain

Corresponding author: Joaquin Luque (jluque@us.es)

This work was supported in part by the Fondo Europeo de Desarrollo Regional (FEDER)/Ministerio de Ciencia e Innovación–Agencia Estatal de Investigación, Government of Spain, under Project RTI2018-094917-B-100, and in part by the ''Centre for the Development of Industrial Technology (CDTI),'' Government of Spain, through the Project ''Eternal Energy.''

ABSTRACT In a highly competitive and liberalized energy market, where the retail of electricity is open to many potential companies, it is essential to have tools that help make decisions and guide the design of marketing strategies. In this sense, it is essential for retailers to know the behavior of their customers to correctly define their commercial strategies. One of the most commonly used methods for this is the characterization of their consumption profiles. Fortunately, for regulatory reasons, in some countries, the monthly electricity demand of each customer is openly available to any competitor. This paper explores whether this information, especially the economic sector and geographic location of a client, is useful for determining the client's demand profile. Specifically, data on electricity demand in Spain from more than 27 million users and for a period of 3 years are analyzed. For this purpose, the electricity consumption of every client is grouped by month and normalized. The resulting demand profiles are later clustered according to different criteria. The main finding of the research is that the combined information on economic activity and location definitely enables prediction of the demand profile. Additionally, profile quality metrics are defined and obtained for the entire dataset. The resulting profiles have a mean dispersion of 10% and a confidence interval of ±17%. To clarify the use of these metrics, several examples are detailed, showing how this profile information can be used to improve the marketing decision-making process for electricity retailers.

INDEX TERMS Energy demand, customer profiling, data engineering, big data applications, statistical learning, pattern analysis.

I. INTRODUCTION

Currently the price of energy is the result of a complex market structure [1], [2]. The liberalization of the electricity sector resulted in many different actors playing various roles in providing energy to end users. In all countries, at the beginning of the electricity supply chain, there are generally a handful of large producer companies that own and manage generation plants. Then a few, or even a single state-owned company, i.e., the transmission system operator (TSO), is responsible for the high-voltage long-range transmission of energy. Finally a few distribution system operator (DSO) companies, usually in an oligopolistic market, are in charge of medium- and low-voltage distribution to the final customer locations [3].

In most countries, all of these activities are strongly regulated and there is not much room for competition at those levels. However, at the end of this supply chain, electricity should be sold to every client, which opens a space for new players (electricity retailers) who carry out marketing, billing, maintenance and customer service activities. In European countries it is common to have several dozen nationwide electricity retailing companies, several hundred of which operate in regional, local or sectoral markets [4]. In this scenario, the retailers buy energy to generators in advance (traditionally one day-head) based on a forecast of their customers' demand. Therefore, any deviation in this estimation

The associate editor coordinating the review of this manuscript and approving it for publication was S. Ali Arefifar.

would translate into an increase in the energy sale price or even into economic losses for retailers. The energy price is usually the key factor for a customer to choose an electricity retailer [5]. Therefore, to maintain competitiveness, these companies have to make many decisions [6], focusing mainly on improving their pricing strategy [7], [8], either by reducing their costs, adjusting tariffs to time-varying rates, promoting demand-side response [9], [10] and/or finding their most profitable customers (niche markets).

A clear example of these niche markets can be found in the current trend of generation deployment at the consumer level (with the consumer taking on the prosumer role). However, consumers find it difficult to profitably sell their surplus energy; moreover, national regulation often does not allow individual clients to compensate for energy demand and production for periods of more than one month. This scenario is especially inappropriate for consumers with a highly seasonal demand, which opens up the opportunity for niche retailers, who can buy this excess nonsalable energy, to fulfill the demands of other customers, while reducing the retailer's purchasing needs.

However, for this economic strategy to be optimized through long-term energy purchases, retailers should keep their overall demand as balanced and stable as possible throughout the year. Therefore, it is essential that retailers fully understand the energy profile of each of their customers, as well as the typical consumption (or generation) profile of potential customers, to fill the gap and find the optimal energy pool.

In this sense, this paper analyzes the influence of the economic activity sector and the geographic location of a customer on its monthly electricity demand profile. Additionally, this relationship is used to cluster customers with the goal of developing more informed marketing strategies in the increasingly competitive electricity retailer arena. The billing information for Spain is used as a case study.

After this introduction justifying the importance of the problem to be addressed and its scientific and technical context, the paper describes the state of the art regarding customer profiling (section II) the structure of the dataset used in the research (section III) and the methodology employed (section IV). The results of profiling and its use to define marketing strategies are presented with several examples in section V. Then, in section VI there is a discussion about the validity of the results and the usability of the economic sector and location to predict the demand profile. Finally, the main findings of the research are presented in the conclusions section.

II. STATE OF THE ART

Profile studies are common in the electricity sector. In particular, retail companies can base their strategies on complex data-driven models [11]. Specifically, to customize prices or address specific offers in all market segments, it is crucial to properly characterize and profile electricity demand [12]. Obviously, to describe these behaviors, it is essential to have historical information from clients. Fortunately, power sector players currently have the enormous advantage of having large amounts of data at almost negligible acquisition cost, allowing the use of data analysis techniques for many purposes [13].

Thus profiling efforts are mainly based on the hourly demand curve, which is currently provided by meter reading equipment [14]. As shown above, profiling customer demand is a field of high interest, has a long tradition and has been used for many other purposes such as customer segmentation [15], detection of nontechnical losses [16] or demand forecasting [17], [18]. The recent pandemic has also been analyzed using changes in electricity demand patterns [19], [20].

Most authors characterize energy demand based solely on the load curves [21], although some other studies also incorporate the influence of climatic variables [22]–[24], sociodemographic factors [25], aggregated [26] or disaggregated [27] global economic activity, and electricity usage. The influence of the economic sector is also considered in some works [28]. The combined influence of various factors on the forecasting of energy demand in the short, medium or long term is analyzed in [29], [30].

The analysis of the electricity demand profiles is mainly based on hourly data obtained through automatic meter readings [31]. Monthly and seasonal analyses are also common [32], while some time-multiscale studies have also been reported [33].

III. DATASETS

To encourage competition in the energy market, Spanish regulation requires DSO companies to share their information on customer consumption. The CNMC (National Commission for Markets and Competition, in Spanish) is in charge of collecting datasets that are not fully public but are available to any registered energy retailing company. Therefore, every stakeholder may know the consumption of every customer across the country. The CNMC dataset contains two files: one to describe the features of each client, and another to specify the energy meter readings [34]. Additionally, two more ancillary files coding the location and the economic activities are incorporated into the dataset. The details of these files are as follows:

A. CUSTOMER FILE

The customer file contains 27,296,335 records, one for each electricity client in mainland Spain (excluding the Balearic and Canary Islands, Ceuta and Melilla). In its original CSV (comma separated values) format, the file size is 8 GB. Each record includes 57 fields, but for the purposes of this paper, only three of them are relevant. The key used for each record is the CUPS (Universal Supply Point Code, in Spanish), an alphanumeric code containing 20 or 22 characters that uniquely identifies each customer. The other two relevant fields are NACE ((European Classification of Economic Activities, in French, a five-character alphanumeric code), indicating each client's activity sector, and the location (a five-digit numeric code) detailed at the municipality level.

Unfortunately, the CNMC has declared the NACE field as optional; thus, retailers are not legally required to provide this information. As a consequence, only 14,945,768 records (55% of the total) have the NACE field reported. In this paper, when the demand profile is related to the location, all records can be used, but if the profile has to be related to the NACE or to an NACE-location pair, only half of the records are usable.

B. READING FILE

The reading file contains 924,338,435 records, one for each electric meter billing reading of each client in the customer file. In its original CSV format, the file size is 106GB. Each record includes 24 fields but for the purposes of this paper only 15 of them are relevant. The key used for each record is a combination of 3 fields: the client's CUPS and the initial and final dates (day-month-year format) of a reading period. In each record, the active energy consumption (W-h, watts-hour) during the specified period is disaggregated into 6 numeric fields, corresponding to the six rates (P1 to P6) in the Spanish time-of-use (TOU) rate structure [35]. Another 6 numeric fields are available for the corresponding reactive energy consumption (var-h, volt-ampere reactive-hour).

For the purposes of this research a unique demand profile, with no hour-based discrimination is required. Therefore, the consumption corresponding to each rate is summed in a single aggregated value.

In this file each CUPS has multiple records, one for each billing reading, containing information about the demand for the last 3 years. Each period covers approximately 30 days but may have a different duration. The relative frequency for the lengths of the periods between two readings is depicted in Fig. 1. The peak for a 30-day period shows that most clients are billed monthly. A lower peak in the 62-day period represents the much less frequent occurrence of bimonthly billing.

C. NACE FILE

To relate energy consumption to the economic sector, the NACE codes are used [36]. They can be specified at 4 different levels of detail: activity (21 codes, identified by a letter); division (88 two-digit codes); group (272 three-digit codes); and class (629 four-digit codes). If different levels of detail are combined, then up to 1011 different NACE codes are possible.

A downloadable electronic version of the NACE codes can be found in [37].

D. LOCATION FILES

To verify the hypothesis that electricity demand profiles also depend on the geographic location, this information has to be incorporated into the research. Three different levels of geographic entities are used: regions (17 two-digit codes) [38]; provinces (52 two-digit codes) [39]; and municipalities (8131 five-digit codes, with the first 2 digits identifying the province and the subsequent 3 indicating the municipality in that province) [40]. If different levels of detail are combined, then up to 8203 different location codes are possible.

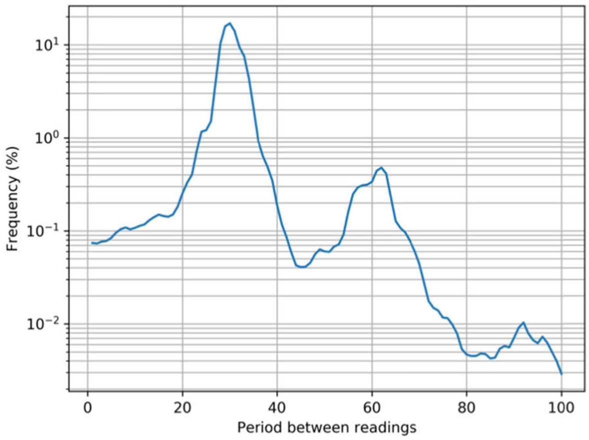

FIGURE 1. Relative frequency (% in log scale) for the length of the period between two readings.

IV. METHODOLOGY

Before analyzing the electricity demand profiles, the dataset described in the previous paragraph must be cleaned and pre-processed. For this purpose, and for the remaining algorithms described in this paper, several Python 3 scripts have been developed.

Although 100 GB databases cannot be properly defined as "big data", they certainly pose additional difficulties to profile analysis that must be considered carefully. For this reason, optimized binary files and speed-focused programming techniques have been extensively used. The overall methodology is depicted in Fig. 2; the number of processes is introduced and explained in the remaining section.

A. PROCESSING THE CUSTOMER FILE

Every record of the customer file has been preserved although only 3 relevant fields are kept: CUPS, NACE and municipality codes. This information has been converted (process #1 in the algorithm diagram) to binary format resulting in a final 0.9 GB file, an approximately 90% reduction in size. Every record is accessed using CUPS (record key) and by employing hashing techniques to identify its position (implemented using Python dictionaries).

B. PROCESSING THE READING FILE

To leverage electricity billing information four different challenges must be faced:

1) The readings for a particular customer are not in a single record, but spread over many records. To address this problem the reading file is transformed (process #2) into a temporary binary file with a very different structure. It has one record for each client containing the number of valid billing periods and up to 50 of them, a number large enough to accommodate 3-year

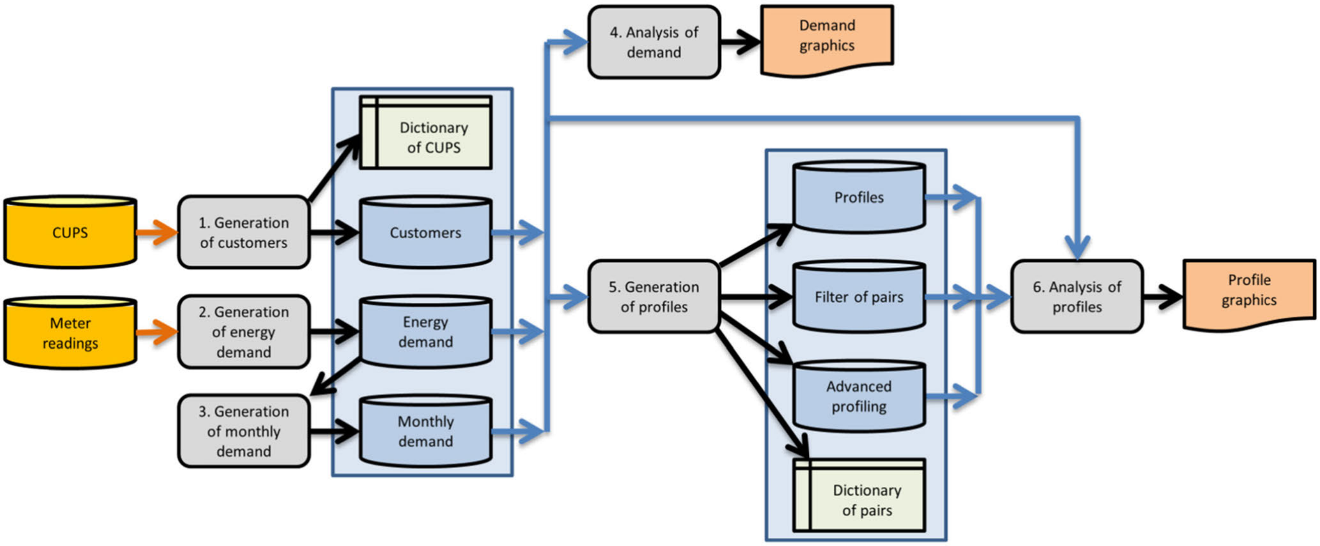

FIGURE 2. Diagram of the energy demand profiling process.

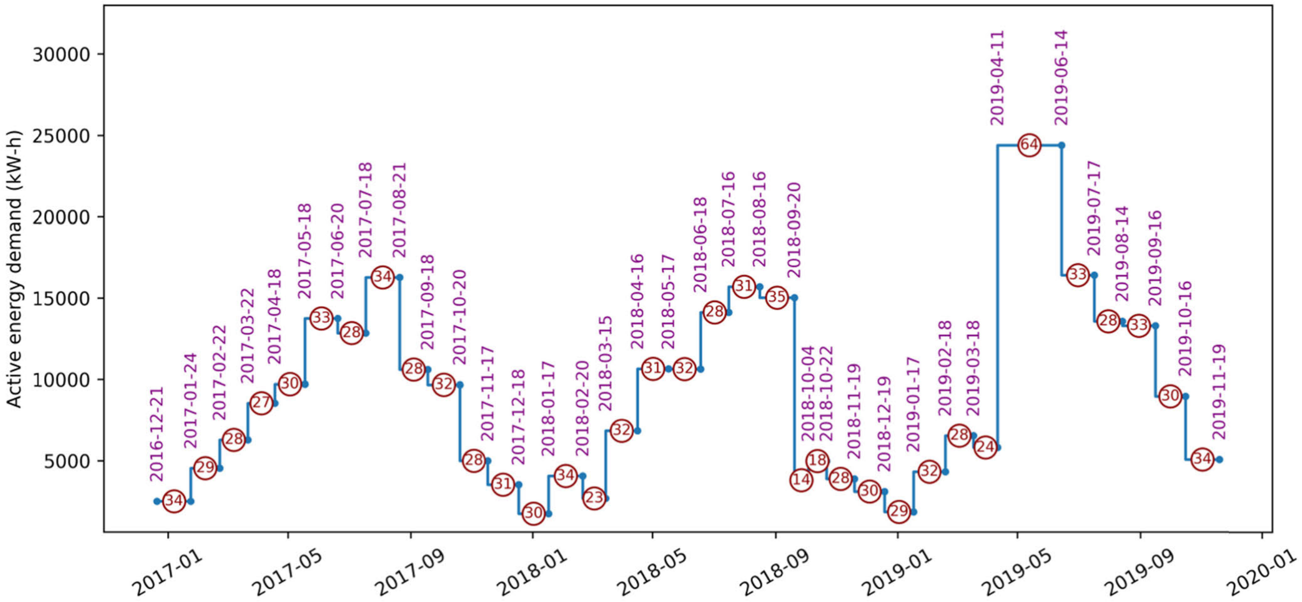

FIGURE 3. Example of active energy demand (kW-h) meter readings for a particular CUPS. The reading dates and time periods between readings (presented in each circle) are also shown.

monthly readings and some spare fields for abnormal reading periods.

2) The energy readings correspond to different time periods. Fig. 3 depicts an example of the active energy meter readings corresponding to a given CUPS (omitted for confidentiality reasons). In addition to periods of approximately 30 days, which are the majority, there are some others as low as 14 days or as high as 64 days. Therefore, instead of the energy consumption E (W-h), it is better to consider the mean power P (W) demanded during a certain reading period of T days, obtained as $P = E/(24 \cdot T)$. The result for the same CUPS is shown in Fig. 4.

3) Reading periods do not exactly fit months, as they rarely start on the first of each month and finish on the last day. Therefore, if the mean power over a given period spans two months, it should be split proportionally into two parts, and each should be assigned to the corresponding month. Subsequently, contributions from two or more periods to the same month must

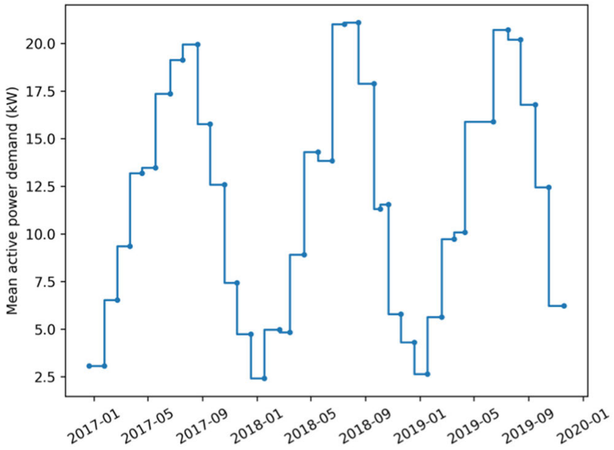

FIGURE 4. **Mean active power demand during each reading period. Example of a particular CUPS.**

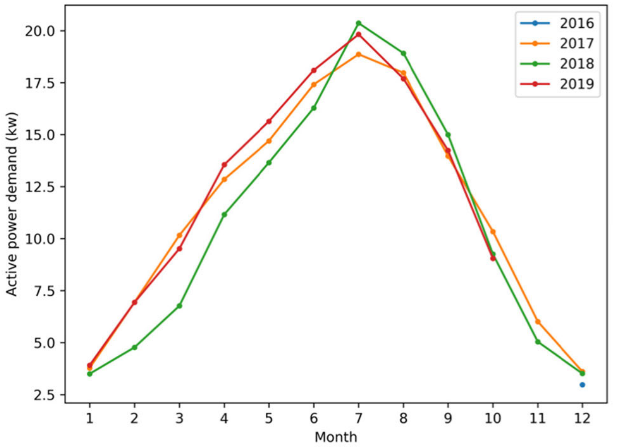

FIGURE 5. **Mean monthly active power demand. Example for a particular CUPS.**

be added. The resulting mean monthly power demand for the CUPS used as an example is depicted in Fig. 5.

4) Eventually, to correctly compare the demand profiles of different CUPSs with the same NACE and/or location, the monthly power demand must be normalized. For this purpose each monthly demand value is divided by an average value. Because some CUPSs present abnormal demand values, instead of calculating the average as the mean value, it is better to use the median (11.16 kW for the CUPS in Fig. 5) since it avoids the extreme influence of outliers.

The final result of this four-stage process is dumped (process #3) into a binary file with a record for every electricity client that contains the client's normalized monthly power demand over a 5-year range. This binary file is 13 GB in size, 12% of the original value, and its information is used to produce different reports and graphics regarding a single customer demand (process #4).

C. RELATING PROFILES TO NACE AND LOCATION

Once a normalized demand profile has been obtained for every CUPS, it is time to analyze its dependence on the NACE and location information.

For this purpose, each NACE-location pair is identified by a code made up of the NACE code followed by a semicolon (";") and the location code. Thus, for instance, the economic sector "hotels" (NACE code "I5510") in the city of Alicante (location code "03014") is coded by "I5510;03014". Specifically, this unique pair identifies up to 67 individual CUPSs, that is, for each month 67 power demand values are available. Then the profile for this pair is defined by the median values at each month.

More formally, let us consider a certain NACE-location pair containing n CUPS. Let us call $P = [P_1, P_2, \cdots, P_{12}]$ the power demand profile of a pair, where P_j is the normalized power corresponding to the j-th month. Let us call $C^{(i)}$ the normalized power consumption corresponding to the i-th CUPS in the pair, where $i \in [1, n]$. This consumption is defined by a matrix

$$C^{(i)} = \begin{bmatrix} C_1^{(i)[1]} & C_2^{(i)[1]} & \cdots & C_{12}^{(i)[1]} \\ C_1^{(i)[2]} & C_2^{(i)[2]} & \cdots & C_{12}^{(i)[2]} \\ \vdots & \vdots & \ddots & \vdots \\ C_1^{(i)[m]} & C_2^{(i)[m]} & \cdots & C_{12}^{(i)[m]} \end{bmatrix}, \tag{1}$$

where $C_j^{(i)[k]}$ represents the normalized power consumption corresponding to the i-th CUPS during the j-th month of the k-th year, in which m is the total number of years with available readings and $k \in [1, m]$. The value of P_j is obtained as

$$P_j \equiv \hat{C}_j = \underset{i,k}{\text{median}}\, C_j^{(i)[k]}. \tag{2}$$

The result is depicted in Fig. 6 where the shadows indicate the 25th to 75th percentiles. The RMSD (root median square distance) and CI (confidence interval) values in the figure title are two metrics of the profile quality which are explained in the next section.

For fast access to the profiles associated with an NACE-location pair, a specific binary file is generated (process #5), along with some other ancillary information. The profile file has 8,293,233 records (1011 × 8203), one for each pair code. Each record contains the percentiles of the profile for each month, with a resolution of 1%. This binary file is 79 GB in size, and its information is used to produce different reports and graphics about the demand of clients identified by the NACE and/or location (process #6).

The processes that generate new files (those numbered 1, 2, 3 and 5 in Fig. 2) have to deal with very large datasets (more than 100 GB). The computing time required by these processes is approximately 100 hours using a modern laptop computer equipped with a solid-state disk. This is a significant but affordable computing effort, as these generation processes are executed only when a new dataset is available (usually once every few months). On the other hand,

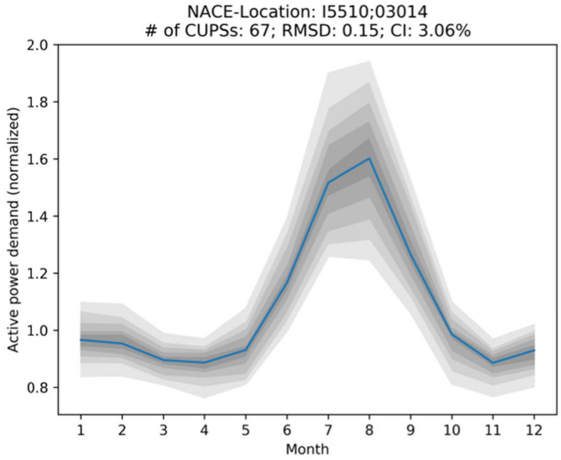

FIGURE 6. Active power demand profile of hotels in the city of Alicante. The blue line indicates the median value in each month. The gray shadows depict the 25th to 75th percentiles.

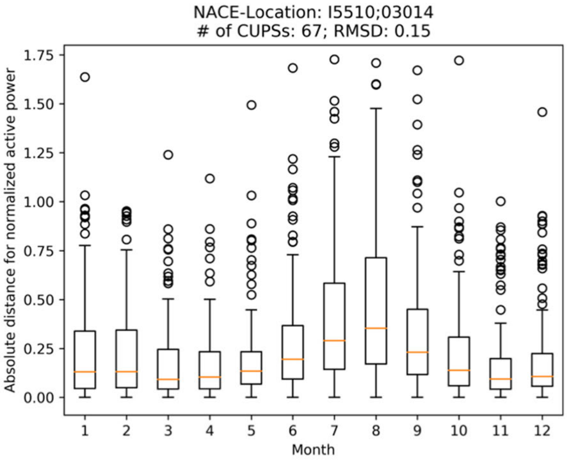

FIGURE 7. Statistical distribution of the absolute distances to the active power demand profile of the hotels in the city of Alicante.

the processes that analyze the energy demand or profiles (those numbered 4 and 6 in Fig. 2) run in real time and usually finish in less than a few seconds.

D. ASSESSING THE DEMAND PROFILES

To evaluate the previously obtained profiles two metrics are used. First, the dispersion of values is considered. In simple terms, the dispersion measures the width of the shaded areas in Fig. 6. More formally, the distance between a given consumption $C_j^{(i)[k]}$ and its profile P_j is defined as

$$D_j^{(i)[k]} \equiv C_j^{(i)[k]} - P_j. \tag{3}$$

The set of distances for the j-th month is defined as

$$D_j \equiv \left\{ D_j^{(i)[k]} \right\}, \quad \forall i \in [1,n], k \in [1,m]. \tag{4}$$

The statistical distribution of the absolute value for the example pair is depicted in Fig. 7. It can be seen that the distances in summer have higher values, corresponding to wider shadows in Fig. 6.

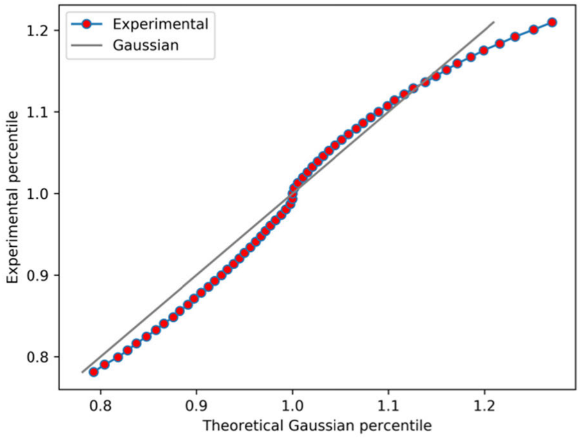

FIGURE 8. Q-Q plot comparing percentiles for a Gaussian distribution and the experimental power demand. Central percentiles between 20 and 80 are shown.

To obtain a single metric for the profile dispersion, the RMSD of a pair is defined as

$$RMSD \equiv \sqrt{\operatorname*{median}_{i,j,k}\left[\left(D_j^{(i)[k]}\right)^2\right]}. \tag{5}$$

The second metric used to evaluate profiling is the confidence interval (CI). More formally, let us consider the set of consumptions during the j-th month, defined as

$$C_j \equiv \left\{ C_j^{(i)[k]} \right\}, \quad \forall i \in [1,n], k \in [1,m]. \tag{6}$$

These consumptions can be regarded as if they were generated by a statistical distribution with a population median $\hat{\mu}_j$. Their sample median is the profile $P_j = \hat{C}_j$. For a $1-\alpha$ certain confidence level (usually chosen as 95%), the confidence interval (CI_j) of its population median is defined as

$$\text{Prob}\left[P_j - CI_j \le \hat{\mu}_j \le P_j + CI_j\right] = 1 - \alpha. \tag{7}$$

Two approaches have been used to determine the values of CI_j, one assuming a certain statistical distribution and another based on numerical simulation. First, suppose that the consumption values C_j are normally distributed which is not very different from the experimental results, and that at least once, the outliers are discarded. Fig. 8 shows a Q-Q (quantile-quantile) plot comparing experimental and theoretical Gaussian data for the central quantiles (between 20 and 80).

Therefore, for normally distributed data, it is a well-known result [41] that the confidence interval of the median (and the mean) is

$$CI_j = t^* \frac{s_j}{\sqrt{N}}, \tag{8}$$

where $N = n \times m$ is the sample size, s is its sample standard deviation, and t^* is the value of the t-Student distribution with $\nu = N - 1$ degrees of freedom, with T_ν such that

$$\text{Prob}\left[|T_\nu| > \alpha\right] = t^*. \tag{9}$$

The second approach uses numerical bootstrapping techniques [42]. Briefly, the N-size C_j set is transformed into

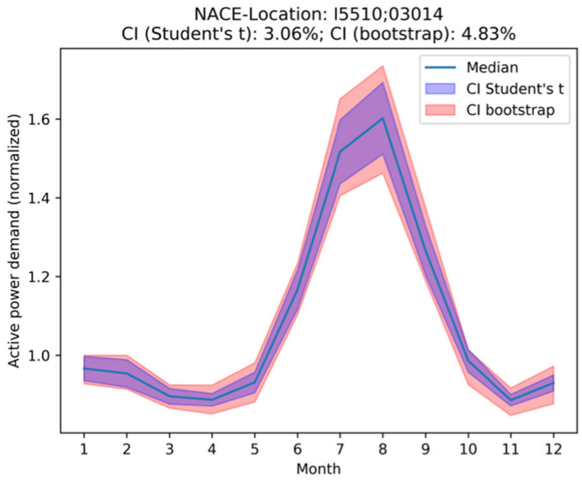

FIGURE 9. Confidence interval of the active power demand profile (median value) of the hotels in the city of Alicante. The computations were carried out using Student's t and bootstrapping.

B different sets where the u-th set $R_j^{\{u\}}$ is obtained by randomly sampling-with-replacement N elements in C_j. Later, the median for each resampled set $\hat{R}_j^{\{u\}}$ is computed. Then the confidence interval (CI_j) is obtained as the value such that

$$\text{Prob}\left[P_j - CI_j \leq \hat{R}_j^{\{u\}} \leq P_j + CI_j\right] = 1 - \alpha. \tag{10}$$

Applying both methods to the example profile, with $B = 10000$, the results shown in Fig. 9 can be obtained.

The confidence interval has a different value for each month. To obtain a single metric CI for the entire profile, the median of CI_j is used, generally expressed as a percentage of the profile, that is,

$$CI \equiv 100 \operatorname*{median}_j \left(\frac{CI_j}{P_j}\right). \tag{11}$$

The overall CI values are also displayed in the title graphic. It can be seen that both methods offer similar results. It has been proved that bootstrapping tends to oversize the confidence intervals [43], which explains the wider bootstrap region in the example. For that reason, and considering that bootstrapping is more computationally demanding, the method used in the rest of the article to calculate the confidence intervals is based on Student's t.

E. LABELING PROFILES

As shown, a power demand profile is defined by a sequence of 12 values, $P = [P_1, P_2, \cdots, P_{12}]$. However, for easy identification of the different profile types, it is useful to assign a single value to every profile. This goal can be achieved by using clustering techniques in which profiles are assigned to a finite (and generally reduced) set of clusters. By numbering clusters, a single value describing each profile is obtained (commonly an integer).

A different approach has been followed in this work, based on dimensionality reduction techniques. Applying principal component analysis (PCA) [44], the original 12 components profile have been transformed into another 12 new components that: a) are orthogonal; and b) minimize the variance in the first components. The results of applying PCA to the active power demand profile are shown in Fig. 10.

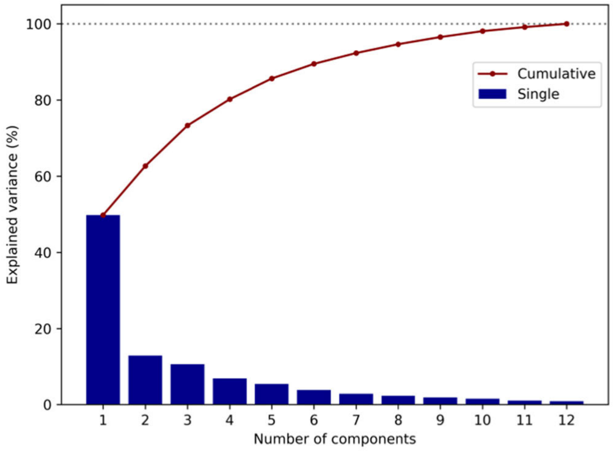

FIGURE 10. Ratio of the explained variance after applying PCA to the active power demand profiles.

In the original profile P, every element P_j explains 1/12 (8%) the total variance. After PCA transformation the variance explained by the first component is 50% of the total.

Therefore, by keeping this component and discarding the others, the original profile can now be approximately represented by a single value P' (a real number). Later, this value is rescaled in the range $[-100, 100]$, resulting in a normalized profile identifier (PID) that can be used to label profiles. The PCA transformation can be written as

$$P \xrightarrow{\text{PCA}} P' \xrightarrow{\text{Scaling}} PID. \tag{12}$$

F. SEPARATING PROFILES

Once the profiles have been obtained, assessed and labeled, it is time to explore whether they have been properly separated using the NACE and/or location. That is, we should explore some type of correlation between PID labels on one side, and the NACE and/or location on the other. For two numeric variables, several correlation metrics have been widely used, for instance, Pearson's correlation coefficient. Unfortunately, in our research, both the NACE and location codes are not numerical but rather categorical variables.

For this case the analysis of variance (ANOVA) test can be employed [45]. Considering for instance the NACE, profile labels are categorized into as many groups as the number of different NACE codes. The null hypothesis H_0 that ANOVA verifies is the following: each group has been generated by the same statistical distribution. The ANOVA test computes the probability of H_0, rejecting it below a certain level of confidence α (usually 5%). The lower Prob $[H_0]$ is, the more useful the NACE code will be in separating all profiles.

ANOVA can also be applied to check whether the NACE code can separate two groups of given profiles. For n NACE codes, up to $n\,(n-1)$ ANOVA tests must be calculated. The

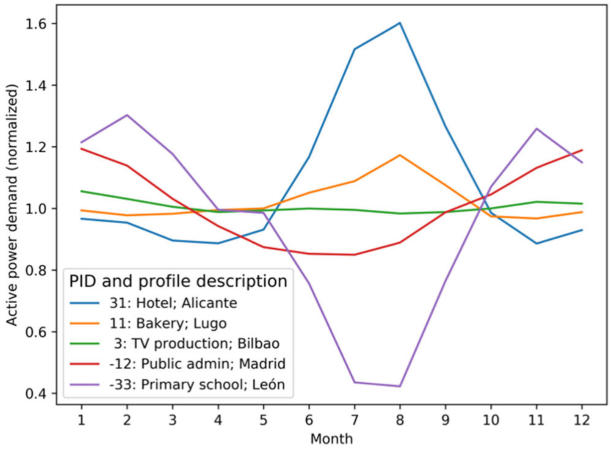

FIGURE 11. Active power demand profiles for five combinations of NACE and location.

higher the ratio of the combinations that reject H_0, the more useful the NACE code will be in separating two profiles.

The ANOVA test makes certain assumptions about the probability distribution of the variables: independence, normality and homoscedasticity (equal variance in each group). If these conditions are not fulfilled, the equivalent less demanding Kruskal-Wallis test can be used [46].

V. RESULTS

A. OBTAINING PROFILES FOR A GIVEN NACE AND LOCATION

Using the datasets and the methodology described in the previous sections, power demand profiles can be obtained for any combination of NACE and location. Up to five examples of these active power profiles are depicted in Fig. 11, indicating the economic sector, the location and their corresponding profile identifier (PID). As shown in this figure, positive PIDs correspond to higher demand in summer, while negative PIDs indicate higher demand in winter.

B. EXPLORING PROFILES AT A GIVEN LOCATION

Many electricity retailers have customers concentrated in a certain geographic area. However, even nationwide retailers usually define their marketing strategies in geographically segmented territories. For these reasons, it is useful to explore how the demand profiles relate to the economic sectors at a given location. Let us imagine an example case of a retailer defining its strategy for the city of Madrid, and let us consider that they are looking for customers with stable demand throughout the year. Then, they should look for economic sectors in Madrid with an almost flat demand profile, that is, with an almost zero PID.

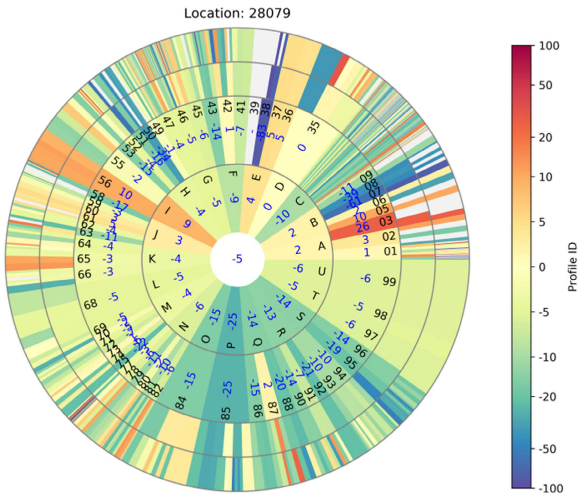

FIGURE 12. Donut chart with the PID values of each economic sector in the city of Madrid.

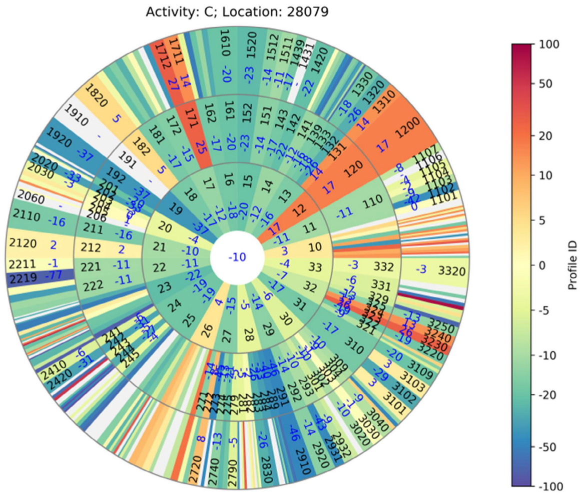

FIGURE 13. Detailed donut chart with the PID values of each industrial economic sector in the city of Madrid.

To search for these customers a donut chart, as depicted in Fig. 12, can be drawn. From the inner to the outer circle the four details of economic sectors (NACE) are described: activity, division, group and class. The letters and numbers in black correspond to the activity and division NACE codes (the rest of the codes are omitted due to lack of space in the graph). The blue number and the color of each sector indicate the value of the metric represented in the figure, in this case the PID in the range [−100, 100]. Gray-colored sectors correspond to NACE codes with no CUPS at that location. A side color bar shows the nonlinear correspondence among colors and PID values.

Continuing with the same example, the retailer may decide to focus on activity "C" (industrial sector) and its dependent subsectors, showing a moderate PID. A detailed donut chart for this industry sector is shown in Fig. 13 (again, some numbers are omitted due to lack of space in the graph), where it can be seen that the pharmaceutical industry (NACE "C2120") has a $PID = 2$ (close to 0). Then, the retailer may decide to focus its marketing strategy on this sector and location.

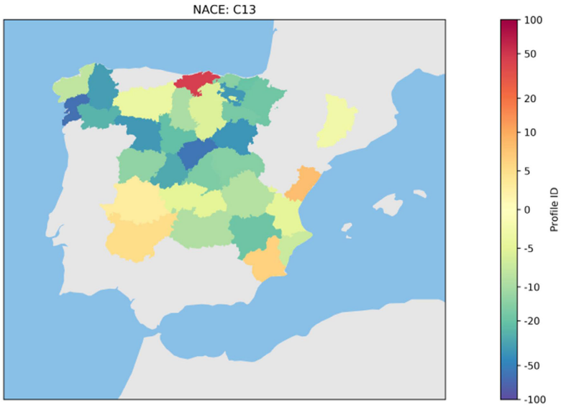

FIGURE 14. Choropleth map with the PID values of the active demand profiles corresponding to textile industries in each Spanish province.

C. EXPLORING PROFILES FOR A GIVEN NACE

A similar situation occurs for retailers that focus on certain economic niches. As another example, let us now consider a company trying to sell electricity to nationwide textile industries (approximately 3000 CUPSs with that definition, although approximately 5000 clients can be estimated in that category if unreported NACE codes are considered). Let us imagine that the most convenient customers are those demanding more energy in winter than in summer, that is, customers with a negative PID. To look for these customers, a choropleth map is drawn in Fig. 14, where the darkest blue province (more negative PID) corresponds to Pontevedra (northwestern Spain), with a $PID = -62$. The areas in gray on the map are provinces with no textile industry (or with an unknown NACE).

D. GLOBAL EXPLORATION OF PROFILES

If the retail company has no restrictions on the NACE or location, then a global search can be performed by looking at each NACE-location pair. Let us now consider the example of a nationwide global-sector retailer who wants to address customers demanding more energy in summer than in winter, that is, customers with a positive PID. For this purpose an NACE-location PID matrix is depicted in Fig. 15, using economic details at the division level and geographical details at the province level. Most of the darker red cells (more positive PIDs) correspond to economic sectors A01 (agriculture), E36 (water treatment), I (hotels), and R93 (sports activities), while the most demanding area in summer corresponds to the Valencian Community. Those should be the main marketing targets for the retail company in the example. Again, the gray area means that the NACE information is absent.

VI. DISCUSSION

A. SPARSITY OF THE NACE-LOCATION PROFILE MATRICES

In the previous section, the PID metric was used to build a matrix using locations as rows and NACE codes as columns. Analogous matrices with the same structure can be defined using different profile metrics: number of CUPSs, dispersion (RMSD), and confidence interval. These profile matrices can be designed using different levels of detail.

Indeed, as previously explained, NACE codes can be defined using 5 different levels of disaggregation: unspecified (1), activity (21), division (88), group (272), and class (629). Analogously, each location code may be defined using 4 levels of disaggregation: unspecified (1), regions (17), provinces (52), and municipalities (8131). Therefore, up to 20 (5×4) different levels of detail can be used to build profile matrices. A metric of the detailing or disaggregation level can be the number of cells in the matrix, that is, the number of NACE-location pairs.

For the most disaggregated matrices, many pairs (cells) are empty; that is, no CUPS fulfills the NACE-location definition of the corresponding pair: in small villages or even in medium-size cities, there are no CUPSs for many sectors of activity. It is clear that the more details regarding the NACE and/or location, the greater the number of empty cells. For the 20 combinations of disaggregation levels, the sparsity of the profile matrices (ratio of empty cells) is depicted in Fig. 16.

High levels of detail affect not only the sparsity of the matrices but also the average number of CUPSs in each matrix cell: the more details there are regarding the NACE and/or location, the lower the average number of CUPSs in each nonempty cell. The results obtained for the 20 combinations of disaggregation levels are depicted in Fig. 17.

B. QUALITY OF PROFILING

The number of CUPSs that define a profile has two opposite effects on its quality parameters. First, dispersion statistics such as the RMSD are biased and, even for a given variance of the population, depend on the number of samples. For instance, let us consider a normally distributed variable x with mean μ and standard deviation σ.

For an n-size sample of x, the sample standard deviation can be defined as

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad (13)$$

where $\bar{x}$ is the sample mean. It can be proven [47] that s is a biased estimation of σ. The relative bias can be computed as

$$b \equiv \frac{E[s]-\sigma}{\sigma} = \sqrt{\frac{2}{n-1}}\,\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} - 1. \qquad (14)$$

In that expression, the bias has a negative value, which means that s underestimates σ, and tends to 0 as n grows. Due to this effect, we should expect the RMSD values to be lower in the profiles obtained with a reduced number of CUPSs. This occurs when the average RMSD is computed for the 20 levels of disaggregation of the profile matrices. The result is shown in Fig. 18.

These experimental results are compared with those derived by (14), showing a good fit, as depicted in Fig. 19.

The second metric for the quality of profiling is its confidence interval (CI). According to (8), the confidence interval

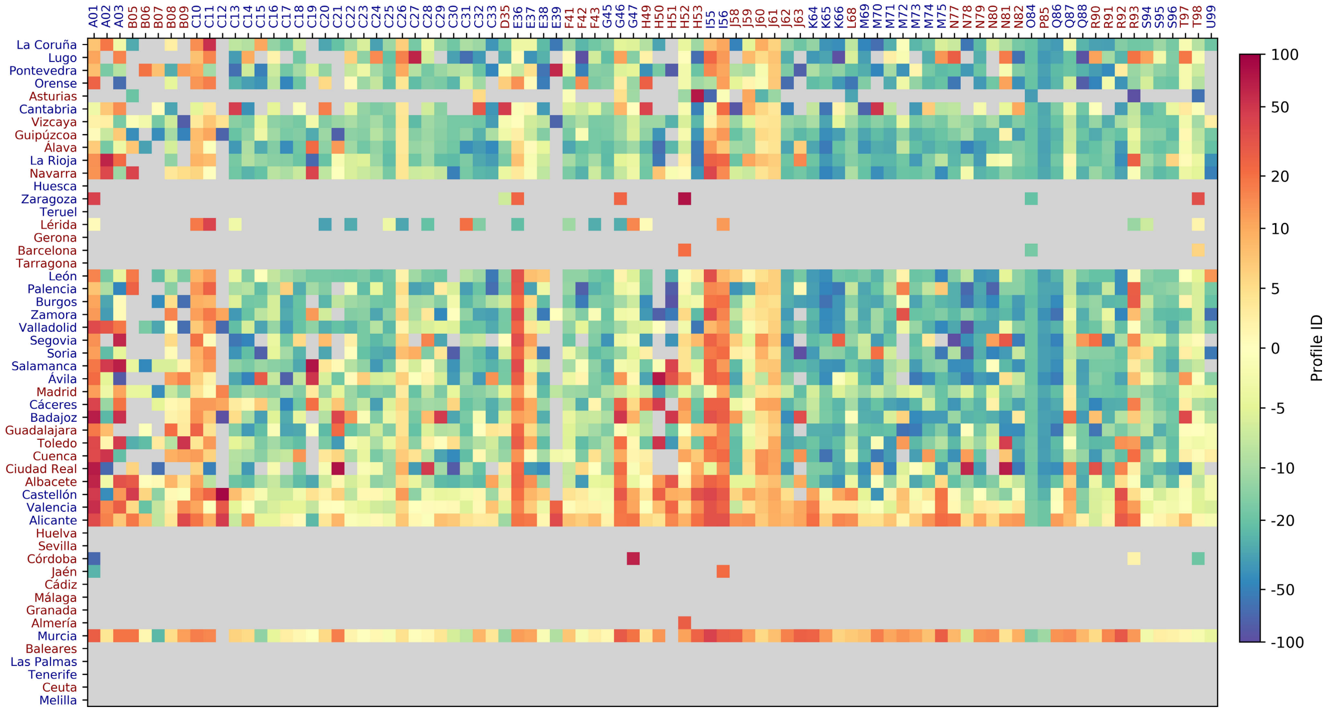

FIGURE 15. NACE-location matrix with the values of active demand profile identification (PID).

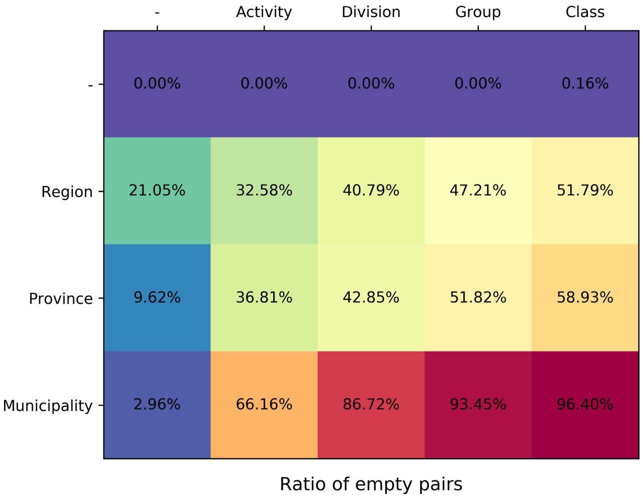

FIGURE 16. Sparsity of profile matrices. For each level of disaggregation, the ratio of empty cells (NACE-location pairs with no CUPS) is given.

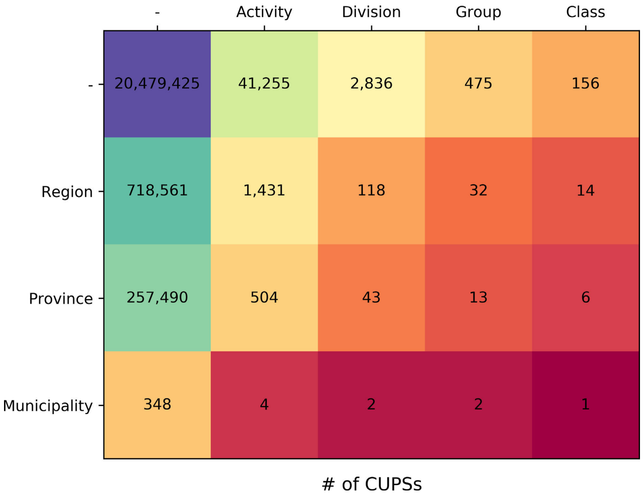

FIGURE 17. For each level of disaggregation, the average number of CUPSs in nonempty cells (NACE-location pairs) is given.

of a profile depends not only directly on the number of CUPSs used to compute it but also indirectly on the standard deviation, the bias of which is related to the number of CUPSs, following (14).

The overall effect on the average CI is computed for the 20 disaggregation levels of profile matrices. The result is shown in Fig. 20.

These experimental results are compared to those derived from the theoretical analysis of a Gaussian distribution, showing a good fitting, as depicted in Fig. 21.

Furthermore, the accuracy of the data on the economic sector and the geographic location of the client is crucial in relation to the reliability of the proposed algorithm and the quality of the resulting profile. This accuracy could be improved using robust state estimation methods such as those described in [48]–[50].

C. SEPARABILITY OF PROFILES BY THE NACE AND LOCATION

To analyze whether the power demand profiles can be separated using the NACE and/or location, it should first be

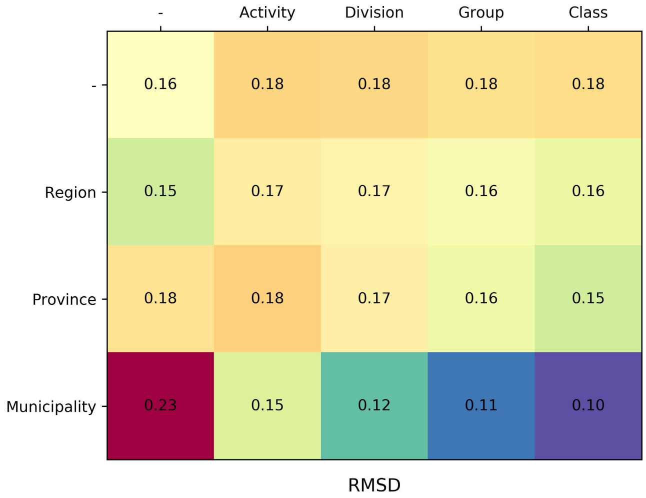

FIGURE 18. For each level of disaggregation, the average profile dispersion (RMSD) is given.

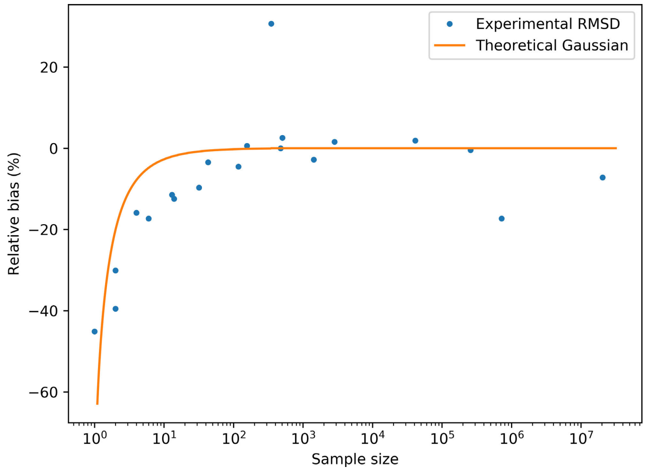

FIGURE 19. Relative bias of two dispersion statistics: experimental average profile dispersion (RMSD) for each level of disaggregation and standard deviation for a theoretical Gaussian distribution.

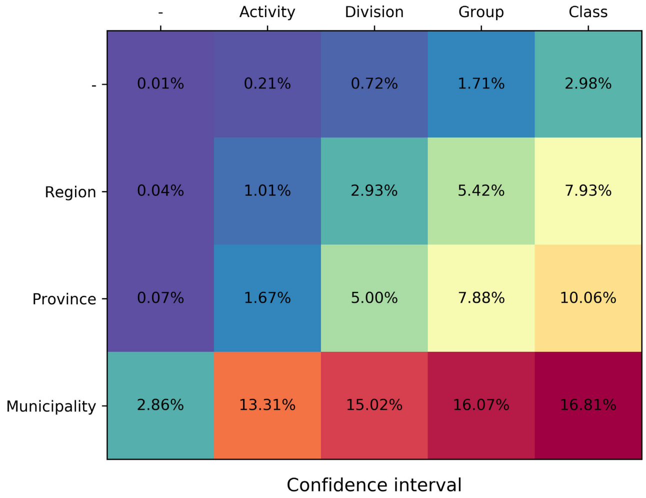

FIGURE 20. For each level of disaggregation, the average profile confidence interval (CI) is given.

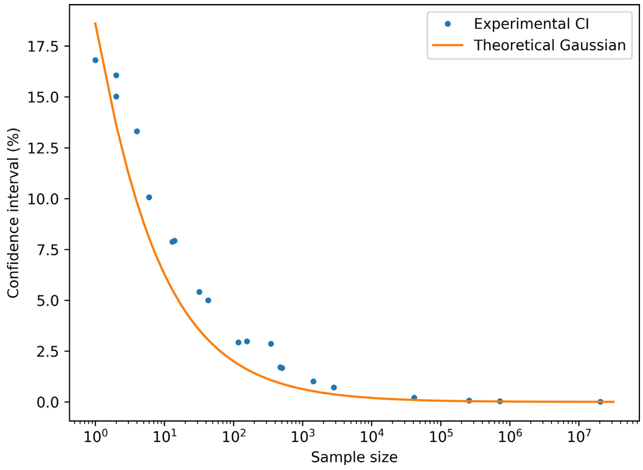

FIGURE 21. Confidence interval in two cases: experimental average profile CI for each disaggregation level and its theoretical values for a Gaussian distribution.

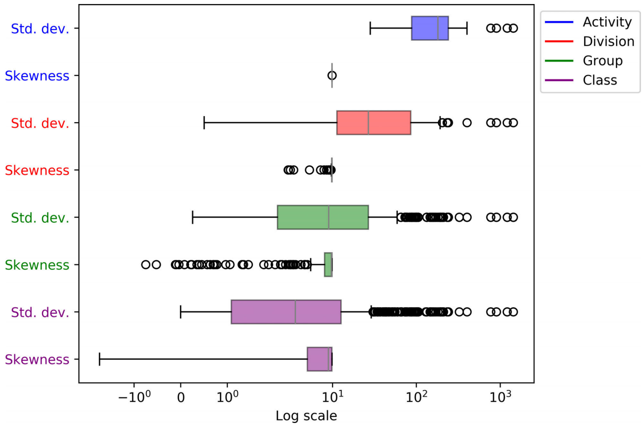

FIGURE 22. Sample standard deviation and skewness of profiles grouped by the NACE code for each level of disaggregation.

determined if the ANOVA test conditions are fulfilled. Let us begin grouping profiles according to the NACE codes. Then, computing the standard deviation and the skewness of each group, the results depicted in Fig. 22 are obtained, where several levels of NACE disaggregation are considered.

It can be seen that the skewness is approximately 10, whereas it should be 0 for a normal distribution. Thus, the normality condition is not fulfilled. Additionally, it can be observed that the standard deviation spans a wide range of values, also violating the homoscedasticity condition. A similar result is obtained if the profiles are grouped by the location code. For these reasons, the ANOVA test was discarded, and the more general Kruskal-Willis test was employed.

With profiles grouped by either the NACE or location codes, the null hypothesis (H_0: every group is generated with the same statistical distribution) has been tested. The resulting probability obtained by the Kruskal-Wallis test is depicted in Fig. 23. It can be seen that for a confidence level $\alpha = 5\%$, the null hypothesis can be rejected, except in the case of using region codes to separate profiles.

A similar analysis can be applied to each pair of groups. The ratio of these pairs achieving a given probability in the Kruskal-Wallis test is depicted in Fig. 24, where the profiles

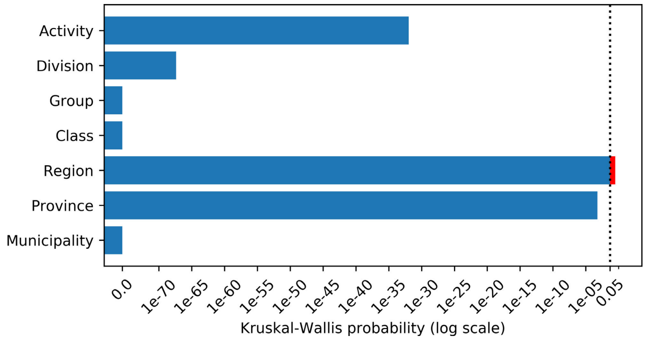

FIGURE 23. Probability that profiles grouped by NACE or location code (at different disaggregation levels), were generated by the same statistical distribution (Kruskal-Wallis test).

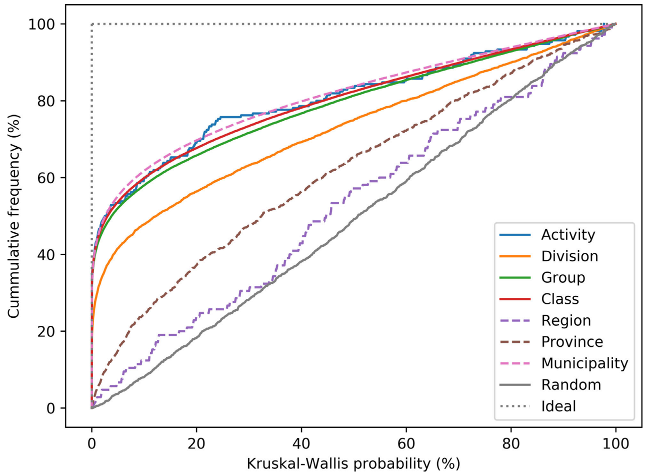

FIGURE 24. Ratio of separable pairs of profiles according to a given probability of the null hypothesis in the Kruskal-Wallis test. The profiles are grouped using only NACE or location codes at different levels of disaggregation.

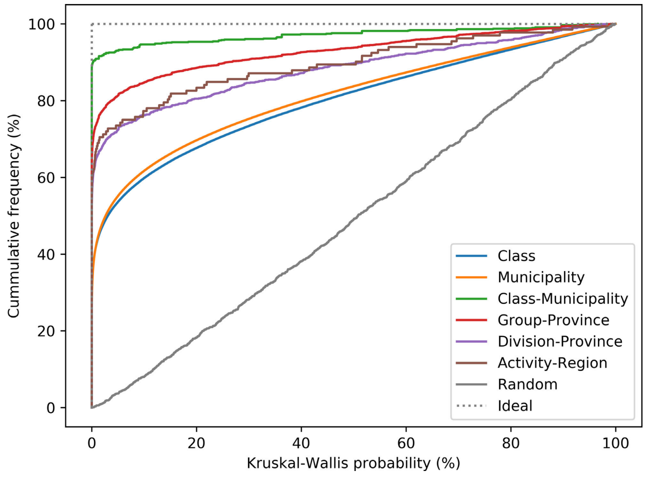

FIGURE 25. Ratio of separable pairs of profiles according to a given probability of the null hypothesis in the Kruskal-Wallis test. The profiles are grouped using both NACE and location codes at different levels of disaggregation.

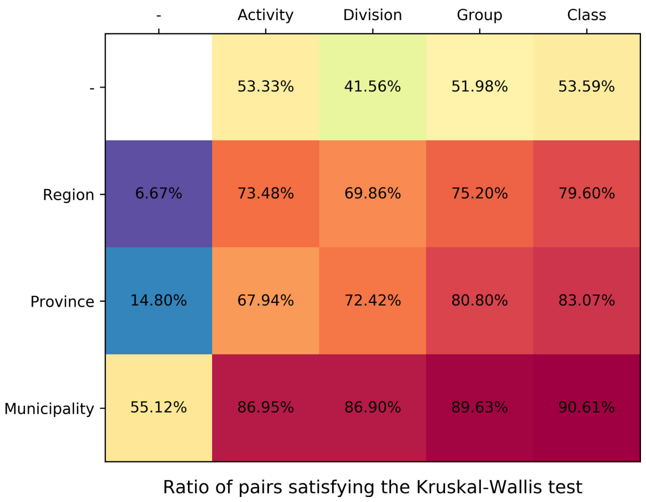

FIGURE 26. For each level of disaggregation, the ratio of separable pairs of profiles is shown, according to the Kruskal-Wallis test with a confidence level of 5%.

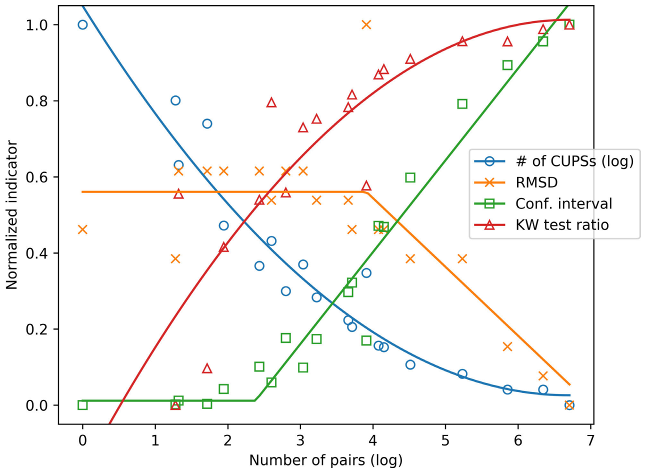

FIGURE 27. Impact of the level of disaggregation of NACE and location on the quality of profiling.

are grouped using only the NACE or location code. The two extreme cases are also drawn: randomly grouped profiles (continuous gray) and profiles grouped in perfectly separable pairs (dashed gray). It can be seen that grouping by the NACE is usually better than that by the location code (except if the municipality code is used). Indeed, grouping profiles by region or province is only slightly better than random grouping.

If both NACE and location codes are used, the result is as shown in Fig. 25, where the best results of using just one code (class or municipality) are also depicted. It can be seen that combining the two codes always offers better results than those obtained using only a single code. Indeed, the grouping of profiles by a combination of class and municipality yielded very good results, with more than 90% of the pairs satisfying the Kruskal-Wallis test with a confidence level of 5%.

The combination of the NACE and location codes at different levels of disaggregation yields the ratio of separable pairs of profiles, as depicted in Fig. 26.

While in [29], [30] the authors use the global size of the economy to forecast total demand, this research uses the economic sector to predict the monthly demand profile.

D. THE ROLE OF THE PROFILE MATRIX DISAGGREGATION LEVEL

In the previous sections, it has been shown that the level of disaggregation of the NACE and location codes, that is, of the profile matrix, impacts several metrics: average number of CUPSs per profile, dispersion, confidence interval and separability. To summarize this effect, the level of disaggregation is measured as the number of cells in the profile matrix, that is, the number of NACE-location pairs, the result of which is depicted in Fig. 27.

It can be seen that higher levels of disaggregation improve the separability of the profiles and their confidence intervals. They also improve (decrease) the average dispersion (RMSD), although this does not actually mean better performance but rather a degradation of the metric due to high bias when it is computed for a small number of samples (CUPSs).

VII. CONCLUSION

This paper has explored the suitability of information on the economic sector and/or the location of an electricity client for the prediction of his/her monthly demand profile. It has been shown that the combined use of both data at the highest available detail offers the best results. Using the economic class and the municipality code to cluster clients, more than 90% of any pair of groups are separable.

The power demand profiles thus obtained have an average confidence interval of ±17% and a dispersion of 10% (which increases to 18% if the bias of the dispersion metric is corrected).

Several examples have also been detailed, showing how this profile information can be used to improve the marketing decision-making process for electricity retailers.

For further liberalization of electricity markets and to foster competitiveness, policy makers should consider mandatorily requiring energy retailers to provide information on the economic sector of their customers. Additionally, they should consider collecting and sharing not only monthly but also hourly energy demand, which will allow for more accurate customer profiling.

ACKNOWLEDGMENT

The authors thank the TSO Company for their collaboration and support.

REFERENCES

[1] A. Obushevs, I. Oleinikova, M. Syed, A. Zaher, and G. Burt, "Future electricity market structure to ensure large volume of RES," in *Proc. 14th Int. Conf. Eur. Energy Market (EEM)*, 2017, pp. 1–6.

[2] I. J. Pérez-Arriaga, J. D. Jenkins, and C. Batlle, "A regulatory framework for an evolving electricity sector: Highlights of the MIT utility of the future study," *Econ. Energy Environ. Policy*, vol. 6, no. 1, pp. 71–92, Jan. 2017, doi: 10.5547/2160-5890.6.1.iper.

[3] M. Vagliasindi and J. Besant-Jones, *Power Market Structure: Revisiting Policy Options*. Washington, DC, USA: International Bank for Reconstruction and Development/The World Bank, 2013.

[4] *Report on the Performance of European Retail Markets in 2018 CEER Report Monitoring Retail Markets WS of Customers and Retail Markets WG*, ACER/CEER, Ljubljana, Slovenia, 2019.

[5] S. C. Littlechild, "Why we need electricity retailers: A reply to Joskow on wholesale spot price pass-through," Judge Inst. Manage. Stud., Cambridge, U.K., Tech. Rep. WP 21/2000, 2000.

[6] J. Yang, J. Zhao, F. Luo, F. Wen, and Z. Y. Dong, "Decision-making for electricity retailers: A brief survey," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4140–4153, Sep. 2018, doi: 10.1109/TSG.2017.2651499.

[7] J. Yang, J. Zhao, F. Wen, and Z. Y. Dong, "A framework of customizing electricity retail prices," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2415–2428, May 2018, doi: 10.1109/TPWRS.2017.2751043.

[8] M. Mulder and B. Willems, "The dutch retail electricity market," *Energy Policy*, vol. 127, pp. 228–239, Apr. 2019, doi: 10.1016/j.enpol.2018.12.010.

[9] S. Wang, S. Bi, and Y.-J.-A. Zhang, "Demand response management for profit maximizing energy loads in real-time electricity market," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6387–6396, Nov. 2018, doi: 10.1109/TPWRS.2018.2827401.

[10] K. Verpoorten, C. De Jonghe, and R. Belmans, "Market barriers for harmonised demand-response in balancing reserves: Cross-country comparison," in *Proc. 13th Int. Conf. Eur. Energy Market (EEM)*, Jun. 2016, pp. 1–5, doi: 10.1109/EEM.2016.7521327.

[11] Y. Liu, D. Zhang, and H. B. Gooi, "Data-driven decision-making strategies for electricity retailers: Deep reinforcement learning approach," *CSEE J. Power Energy Syst.*, vol. 7, no. 2, Mar. 2021, doi: 10.17775/cseejpes.2019.02510.

[12] J. Yang, J. Zhao, F. Wen, and Z. Dong, "A model of customizing electricity retail prices based on load profile clustering analysis," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3374–3386, May 2019, doi: 10.1109/TSG.2018.2825335.

[13] F. V. Scheidt, H. Medinová, N. Ludwig, B. Richter, P. Staudt, and C. Weinhardt, "Data analytics in the electricity sector—A quantitative and qualitative literature review," *Energy AI*, vol. 1, Aug. 2020, Art. no. 100009, doi: 10.1016/j.egyai.2020.100009.

[14] N. Mahmoudi-Kohan, M. P. Moghaddam, and M. K. Sheikh-El-Eslami, "An annual framework for clustering-based pricing for an electricity retailer," *Electr. Power Syst. Res.*, vol. 80, no. 9, pp. 1042–1048, Sep. 2010, doi: 10.1016/j.epsr.2010.01.010.

[15] F. Biscarri, I. Monedero, A. García, J. I. Guerrero, and C. León, "Electricity clustering framework for automatic classification of customer loads," *Expert Syst. Appl.*, vol. 86, pp. 54–63, Nov. 2017, doi: 10.1016/j.eswa.2017.05.049.

[16] J. A. Meira, P. Glauner, R. State, P. Valtchev, L. Dolberg, F. Bettinger, and D. Duarte, "Distilling provider-independent data for general detection of non-technical losses," in *Proc. IEEE Power Energy Conf. Illinois (PECI)*, Feb. 2017, pp. 1–5, doi: 0.1109/PECI.2017.7935765.

[17] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 911–918, Mar. 2015, doi: 10.1109/TSG.2014.2364233.

[18] P. Pelka and G. Dudek, "Pattern-based long short-term memory for midterm electrical load forecasting," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.

[19] S. García, A. Parejo, E. Personal, J. Ignacio Guerrero, F. Biscarri, and C. León, "A retrospective analysis of the impact of the COVID-19 restrictions on energy consumption at a disaggregated level," *Appl. Energy*, vol. 287, Apr. 2021, Art. no. 116547, doi: 10.1016/j.apenergy.2021.116547.

[20] I. Santiago, A. Moreno-Munoz, P. Quintero-Jiménez, F. Garcia-Torres, and M. J. Gonzalez-Redondo, "Electricity demand during pandemic times: The case of the COVID-19 in Spain," *Energy Policy*, vol. 148, Jan. 2021, Art. no. 111964, doi: 10.1016/j.enpol.2020.111964.

[21] K. Zhou, C. Yang, and J. Shen, "Discovering residential electricity consumption patterns through smart-meter data mining: A case study from China," *Utilities Policy*, vol. 44, pp. 73–84, Feb. 2017, doi: 10.1016/j.jup.2017.01.004.

[22] F. Apadula, A. Bassini, A. Elli, and S. Scapin, "Relationships between meteorological variables and monthly electricity demand," *Appl. Energy*, vol. 98, pp. 346–356, Oct. 2012, doi: 10.1016/j.apenergy.2012.03.053.

[23] C.-L. Hor, S. J. Watson, and S. Majithia, "Analyzing the impact of weather variables on monthly electricity demand," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 2078–2085, Nov. 2005, doi: 10.1109/TPWRS.2005.857397.

[24] R. Obringer, S. Mukherjee, and R. Nateghi, "Evaluating the climate sensitivity of coupled electricity-natural gas demand using a multivariate framework," *Appl. Energy*, vol. 262, Mar. 2020, Art. no. 114419, doi: 10.1016/j.apenergy.2019.114419.

[25] M. Hayn, V. Bertsch, and W. Fichtner, "Electricity load profiles in europe: The importance of household segmentation," *Energy Res. Social Sci.*, vol. 3, pp. 30–45, Sep. 2014, doi: 10.1016/j.erss.2014.07.002.

[26] T.-H. Yang, R. Sun, Q. Wei, and Y. Gao, "Saturated demand forecast of regional power grid based on amended self-adaptive logistic model: A case study of east China," *IEEE Access*, vol. 9, pp. 1190–1196, 2021, doi: 10.1109/ACCESS.2020.3046105.

[27] R. Sánchez-Durán, J. Luque, and J. Barbancho, "Long-term demand forecasting in a scenario of energy transition," *Energies*, vol. 12, no. 16, p. 3095, Aug. 2019, doi: 10.3390/en12163095.

[28] J. I. Guerrero, I. Monedero, F. Biscarri, J. Biscarri, R. Millan, and C. Leon, "Non-technical losses reduction by improving the inspections accuracy in a power utility," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1209–1218, Mar. 2018, doi: 10.1109/TPWRS.2017.2721435.

[29] K. Gaur, H. Kumar, R. P. K. Agarwal, K. V. S. Baba, and S. K. Soonee, "Analysing the electricity demand pattern," in *Proc. Nat. Power Syst. Conf. (NPSC)*, Dec. 2016, pp. 1–6, doi: 10.1109/NPSC.2016.7858969.

[30] A. A. Mir, M. Alghassab, K. Ullah, Z. A. Khan, Y. Lu, and M. Imran, "A review of electricity demand forecasting in low and middle income countries: The demand determinants and horizons," *Sustainability*, vol. 12, no. 15, p. 5931, Jul. 2020, doi: 10.3390/su12155931.

[31] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," *Energy*, vol. 42, no. 1, pp. 68–80, Jun. 2012, doi: 10.1016/j.energy.2011.12.031.

[32] R. Sánchez-Durán, J. Barbancho, and J. Luque, "Solar energy production for a decarbonization scenario in spain," *Sustainability*, vol. 11, no. 24, p. 7112, Dec. 2019, doi: 10.3390/su11247112.

[33] J. Luque, D. Anguita, F. Pérez, and R. Denda, "Spectral analysis of electricity demand using Hilbert–Huang transform," *Sensors*, vol. 20, no. 10, p. 2912, May 2020, doi: 10.3390/s20102912.

[34] CNMC. *Databases on Consumers and Supply Points (SIPS) for Gas and Electricity*. Accessed: Feb. 11, 2021. [Online]. Available: https://sede.cnmc.gob.es/en/tramites/energy/databases-consumers-and-supply-points-sips-gas-and-electricity

[35] Y.-C. Hung and G. Michailidis, "Modeling and optimization of Time-of-Use electricity pricing systems," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4116–4127, Jul. 2019, doi: 10.1109/TSG.2018.2850326.

[36] *Statistical Classification of Economic Activities in the European Community, Rev. 2*, NACE, Eurostat, Luxembourg City, Luxembourg, 2008.

[37] Instituto Nacional de Estadística. *Estructura Completa de la CNAE 2009*. Accessed: Feb. 11, 2021. [Online]. Available: https://www.ine.es/daco/daco42/clasificaciones/cnae09/estructura_cnae2009.xls

[38] Instituto Nacional de Estadística. *Relación de Comunidades y Ciudades Autónomas Con Sus Códigos*. Accessed: Feb. 11, 2021. [Online]. Available: https://www.ine.es/daco/daco42/codmun/cod_ccaa.htm

[39] Instituto Nacional de Estadística. *Relación de Provincias Con Sus Códigos*. Accessed: Feb. 11, 2021. [Online]. Available: https://www.ine.es/daco/daco42/codmun/cod_provincia.htm

[40] Instituto Nacional de Estadística. *Relación de Municipios, Provincias, Comunidades Autónomas y Sus Códigos*. Accessed: Feb. 11, 2021. [Online]. Available: https://www.ine.es/daco/daco42/codmun/codmun20/20codmun.xlsx

[41] G. Cumming, *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. London, U.K.: Routledge, 2013.

[42] B. Efron, "Bootstrap methods: Another look at the jackknife," in *Breakthroughs in Statistics*. New York, NY, USA: Springer, 1992, pp. 569–593.

[43] P. Good, *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. Cham, Switzerland: Springer, 2006.

[44] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, Jul. 2010, doi: 10.1002/wics.101.

[45] R. R. Hocking, *Methods and Applications of Linear Models: Regression and the Analysis of Variance*. Hoboken, NJ, USA: Wiley, 2013.

[46] G. W. Corder and D. I. Foreman, *Nonparametric Statistics: A Step-by-Step Approach*. Hoboken, NJ, USA: Wiley, 2014.

[47] E. E. Cureton, "The teacher's corner: Priority correction to 'unbiased estimation of the standard deviation,'" *Amer. Stat.*, vol. 22, no. 3, p. 27, 1968, doi: 10.1080/00031305.1968.10480477.

[48] Y. Chen, Y. Yao, and Y. Zhang, "A robust state estimation method based on SOCP for integrated electricity-heat system," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 810–820, Jan. 2021.

[49] Y. Chen, J. Ma, P. Zhang, F. Liu, and S. Mei, "Robust state estimator based on maximum exponential absolute value," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1537–1544, Jul. 2017.

[50] Y. Chen, F. Liu, S. Mei, and J. Ma, "A robust WLAV state estimation using optimal transformations," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 2190–2191, Jul. 2015.

JOAQUIN LUQUE (Senior Member, IEEE) received the M.S. and Ph.D. degrees in industrial engineering (electrical engineering) and the M.S. degree in philosophy from the University of Seville, Spain, in 1980, 1986, and 1994, respectively. He also has extended managing experience at the University of Seville, where he was the Head of the Department, from 1993 to 2000, and Rector, from 2008 to 2012. He was a Visiting Scholar with the University of California Berkeley, USA, in 2018, and an Invited Professor with the University of Genoa, Italy, in 2019. He is currently a Full Professor of electronic engineering with the University of Seville, with more than 30 years of teaching and research experience in different computer engineering disciplines mainly related to basic electronics, communications and control.

ENRIQUE PERSONAL (Member, IEEE) received the degree in industrial electronic engineering and the degree in automatic control and industrial electronic engineering from the University of Seville, Spain, in 2006 and 2009, respectively, and the Ph.D. degree in industrial computer science, in 2016. He is currently an Associate Professor with the Department of Electronic Technology, University of Seville. His main research interests include smart grids, fault location methods, power systems, demand-side management, and flexibility.

ANTONIO GARCIA-DELGADO received the B.S. degree in electronic physics from the University of Seville, in 1982. Since 1984, he has been a Professor of electronic engineering with the Department of Electronic Technology, University of Seville, being coordinator of its electronic instrumentation and digital signal processing lines. His research interests include smart grids, electricity markets, power devices, and fault location methods for power lines.

CARLOS LEON (Senior Member, IEEE) received the B.Sc. degree in electronic physics and the Ph.D. degree in computer science from the University of Seville, Seville, Spain, in 1991 and 1995, respectively. He is currently a Full Professor of electronic engineering and computer science with the University of Seville and the Head of the Telefonica Chair. He has been the director or a principal investigator of more than 70 research projects, mainly in collaboration with companies. His research interests include knowledge-based systems, computational intelligence, big data analytics, blockchain, edge computing, the cyber-physical IoT systems, and machine learning, focusing on utility system management. On these topics, he is the author of more 200 articles and conference contributions. He is a Senior Member of the IEEE Power Engineering Society.