Load Profile Analysis in Electrical Systems: The Impact of Electrical Signature and Monitoring Quality in the Energy Digitalization Process

Luca Tari, *Graduate Student Member, IEEE*, Mario Molinara, *Senior Member, IEEE*, and Luigi Ferrigno, *Senior Member, IEEE*

***Abstract*— In the context of energy digitization, Load Profiling can be a useful tool for making decisions about the use and the health of an electrical load and could be adopted for strategic services related to the energy efficiency, characterization, prediction, optimization, and diagnosis of monitored systems. To accomplish this task, it is important a spread of a new generation of smart meters. In this scenario, useful information for the developers of services based on Load Profiling could be the minimum metrological characteristics of smart meters, that is, the monitoring quality (MQ), and the electrical parameters to extract from the electrical signature to have a reliable Load Profiling process, that is, the electrical signature quality (ESQ). This article tries to answer these questions by investigating the impact of the MQ and the ESQ on the performance of Load Profiling. The main results of the article are: 1) generally the Load Profiling process requires metrological characteristics lower than those required for energy billing reducing submetering infrastructure costs; 2) the increasing number of energy features adopted in the Load Profiling not always improves reliability and accuracy; this happens when many of the features considered do not have great sensitivity with respect to changes in the energy states for the considered case study; 3) the use of functions that measure the sensitivity of a feature to the Load Profiling process, such as the considered kernel density estimation (KDE), and a suitable threshold process can delete the parameters with poor sensitivity and can greatly improve the Load Profiling reliability; and 4) the method considered in the article could help to analyze Load Profiling problems related to other physical quantities (i.e., thermal energy profiling or even multiphysical systems), allowing to the definition of a target MQ and selecting a minimum number of useful features to be adopted.**

***Index Terms*— Digitalization, electrical signature analysis, energy management systems, feature selection, load profiling, measurement quality, power system, smart energy, smart monitoring.**

I. Introduction

In THE last years, the International Energy Agency (IEA) [1] is recommending policies and new services that improve energy reliability, affordability, and sustainability. In this scenario, the energy sector is living through a constant transformation in terms of digitization and the development of new energy services based on the smart use of energy information (i.e., Smart Energy).

In accordance also with national and international policies (e.g., COP26, Directive EU 2018/2002–2012/27–etc. [2], [3], [4], [5]), a new paradigm for energy management needs to be introduced to maximize energy efficiency and optimize energy systems. To this end, a proper process of *Smart Monitoring* and *Load Profiling* can help identify the end user's consumption pattern, make consumption forecasts, and define the operating conditions of individual devices. In this way, electricity management can be improved and made more efficient. A Load Profiling process relies on the use of algorithms, typically classification and clustering, so it is crucial to consider a clear concept: if the input data collected do not correctly describe the phenomenon that one wants to analyze, the output of the process will be far from the desired result.

Therefore, in such a process, two parameters play a strategic role: the monitoring quality (MQ) and the electrical signature quality (ESQ) [6]. As regards "MQ," it refers to the minimum accuracy that smart meters have to warrant to accomplish the given task. This parameter is very important since it influences both the energy consumption and energy signature measurements and the overall smart meter network cost [7], [8], [9]. About "ESQ," it refers to the number of parameters extracted from the electrical signature of the device. This is essential because each electrical device has its unique "electrical signature" that distinguishes it from other devices. According to [6], by electrical signature, we mean not the "classical" electrical quantities (P, N, S, I_{rms}, etc.) but the entire harmonic spectrum of voltage and current absorbed by the electrical load, which for the sake of synthesis, can be summarized through the implementation of appropriate computational metrics [10], [11]. Using all or part of these extrapolated parameters makes it possible to bring out "spontaneously" the desired characteristics of the analyzed system, that is, the operating states in the case of Load Profiling. In particular, it is pointed out in [6] that in difficult Load Profiling scenarios, information from the electrical power parameters P, N, and S is not sufficient to identify the operational state of an electrical load, whereas the addition of a few useful parameters selected from the electrical signature can significantly improve

Manuscript received 30 March 2023; revised 2 July 2023; accepted 13 July 2023. Date of publication 7 August 2023; date of current version 16 August 2023. This work was supported in part by "PiattafoRma Integrata per la Manutenzione e gestione Energetica nella fabbrica intelligente, PRIME" within Fondo per la Crescita Sostenibile-Sportello "Fabbrica Intelligente" under Project F/190092/03/X44; and in part by PONIEC 2014–2020 funded by the Italian Ministry of Enterprises and Made in Italy (MISE) under Program F/190092/03/X44–CUP:B36G20001160005 COR:3833298. The Associate Editor coordinating the review process was Dr. Ferdinanda Ponci. *(Corresponding author: Luca Tari.)*

The authors are with the Department of Electrical and Information Engineering "M. Scarano," University of Cassino and Southern Lazio, 03043 Cassino, Italy (e-mail: luca.tari@unicas.it; m.molinara@unicas.it; ferrigno@unicas.it).

Digital Object Identifier 10.1109/TIM.2023.3302387

this process. Similar considerations are made in [12] and [13] with regard to fault diagnosis on electric motors.

With regard to these two points, "MQ" and "ESQ," the purpose of this work is to highlight the importance of the MQ and ESQ in a Load Profiling process and to analyze how the variation of these parameters impacts the Load Profiling performance. Specifically, the analyses reported in this article were conducted for three levels of MQ and three levels of ESQ, respectively, derived from the standard that regulates the commercialization of smart meters and power definitions and measurements. Furthermore, since the performance of the Load Profiling process may also depend on the type and manner of implementation of the profiling algorithms, the analyses just mentioned were repeated three times using three different types of algorithms. This choice was made to have a broader and more general view of the problem and to return a result valid for different architectures and types of algorithms. The analyses reported in this work were first conducted on a *simulated scenario* considering a public energy dataset; subsequently, to obtain more robust and reliable results, a *real scenario* with experimentally acquired electrical loads was considered. In addition, the article proposes a first unsupervised methodology, based on the kernel density estimation (KDE), to select useful features from those related to the electrical signature to improve the Load Profiling process allowing to consider only those features that show good sensitivity to the operative states of a load.

The article is divided into six sections, the structure of which is as follows: the basic concepts of Load Profiling, the most commonly used applications, and techniques are given in Section II. Section III presents the implemented Load Profiling process, types of tests performed, methods, and characteristics in terms of MQ and ESQ. Section IV describes the peculiarities of the used public energy dataset "eLAMI" and the measurement setup for the acquisition of the analyzed real electrical loads. The results of the conducted analysis are reported in Section V. In particular, the performance obtained for three levels of measurement uncertainty and three different combinations of electrical parameters considered are compared. Finally, in Section VI, final considerations are given.

II. Load Profiling: A Brief Introduction

This section briefly introduces Load Profiling, provides an overview of possible applications in various areas of intelligent energy, and describes its potential benefits. In addition, the structure of a Load Profiling process is presented and some of the most commonly used techniques are briefly described.

A. Some Notes About Load Profile

The digitization of the energy sector is part of the modernization and evolution process aimed at tackling current energy and environmental problems, including: using energy smartly and consciously, reducing energy consumption, and developing advanced techniques to maximize energy efficiency. These concepts underpin the energy and ecological transition taking place worldwide [14]. To this aim, a Load Profiling process can be useful to identify consumption patterns, generate forecasts, develop optimal management strategies, and identify each device's operating conditions and nature [15]. Knowledge of the specific consumption of individual devices implies empowering customers with respect to the issues highlighted, making them aware of their impact, their habits, and how much these have contributed to their total energy consumption [16], so that they can adopt energy-saving behaviors. Demand response, load forecasting, and nontechnical loss detection, among other uses, are just some of the many applications of the Load Profiling process [17], [18]. Analyzing data streams collected by smart meters can provide detailed information on demand characteristics that can be used to improve grid operation and planning [19]. Demand–response, a very active area of research around the world, is an efficient strategy to encourage the use of renewable energy sources and minimize the gap between electrical load peaks and valleys [20], [21]. Time-series-based statistical and forecasting studies are frequently used in smart energy.

By monitoring and profiling single loads using smart meters, energy savings can be achieved, as it allows action to be taken on the timing of individual loads and to eliminate unnecessary activities [22], [23]. For example, it is possible to define consumption trends, and performance indices or make evaluations on the operating cycle of the monitored machine (a very interesting application for industries, applicable to industrial production processes) as previously reported [24]. Moreover, downstream of the Load Profiling process, it is possible to construct past and future seasonal trends and define energy efficiency and optimization plans for the monitored system. It should be emphasized that having the "history" of the electrical signature, or a summary thereof, of the monitored load and the corresponding identified operational state has a considerable advantage: the concept of predictive load diagnostics is realized [25]. In particular, by analyzing the trends of the various calculated quantities, it is possible to identify the occurrence of any anomalous trends for each operating state which deviate from the behavior of the same under normal operating conditions [26]. Based on this comparison, *alerts* can be activated if significant deviations occur.

B. Load Profiling Process

The purpose of this section is to describe the structure of a Load Profiling process. Of course, before starting the process, it is necessary to collect data from the system you wish to analyze or have a dataset already acquired. Then, depending on where the process is stopped, several application alternatives can be implemented: 1) the collected data can be processed at regular intervals (dimensional, daily, hourly, etc.) for asynchronous Load Profiling with the monitored system; 2) the data can be stored in a database to be used later for training Load Profiling real-time algorithms; and 3) if already trained algorithms are available, the data acquired from each device can be processed immediately for real-time Load Profiling. The process, illustrated in Fig. 1 and described in detail below, consists of three *steps*: 1) PRELIMINARY STEP; 2) OPERATING STEP; and 3) REAL-TIME STEP.

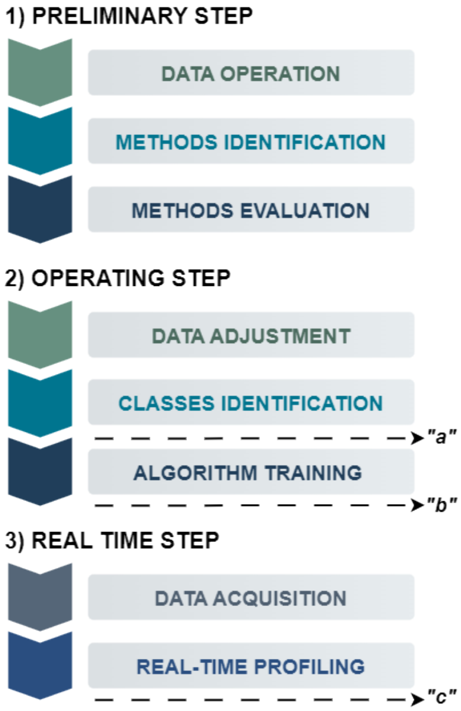

Fig. 1. Block Diagram of the Load Profiling Process composed of three sections: "PRELIMINARY STEP," "OPERATING STEP," and "REAL TIME STEP." The labels "*a*," "*b*," and "*c*" represent the three possible instants of process interruption depending on the desired objective.

1) "PRELIMINARY STEP"—Consists of a preliminary analysis to understand the nature of the data and identify the methods best suited to the scenario analyzed, in terms of Load Profiling and relation to the computational specifications. Thus, the first section consists of.
   a) In "DATA OPERATION," data exploration and data cleaning operations are performed to assess and identify the characteristics of the monitored physical system, relationships and internal distributions, patterns and points of interest, and anomalies or missing data.
   b) In "IDENTIFICATION OF METHODS," different types of techniques for Load Profiling are identified and tested, for example, clustering and classification, possibly combined with feature selection techniques (supervised and unsupervised), as reported in Section II-C.
   c) In "EVALUATION OF METHODS," the identified methods and techniques are evaluated for robustness in relation to the analyzed system. The evaluation can be done by resorting to appropriate metrics, operator experience, knowledge of the analyzed problem, or if available, by resorting to ground truth.

   Once the first step has been executed (iterating if necessary), the next *step* is executed.
2) "OPERATING STEP"—In this section, starting from an unlabeled dataset faithful to the scenario studied in *Step 1*, it is possible to realize the two alternatives "*a*" and "*b*," described earlier or otherwise train the algorithm for alternative "*c*."
   a) In "DATA ADJUSTMENT," the validity of the input data is checked, and, if necessary, corrections are made before proceeding with the subsequent analysis.
   b) In "IDENTIFICATION OF CLASSES," a class search is carried out based on the methods chosen in *Step 1*, for example, by clustering. In this way, the labeling of the considered dataset is obtained. Alternative "*a*" aimed at an asynchronous Load Profiling process with respect to monitoring ends at this point.
   c) In "TRAINING ALGORITHM," the result of the implemented method for class identification is used to train the profiling algorithm, for example, a classifier, which must subsequently work in real time on the monitored system. This is followed by a verification phase of the training on a test data set. At this point, alternative "*b*" ends, aimed solely at training the algorithm for real-time Load Profiling.

   After completing *Step 2*, the Load Profiling tool is ready to be applied to a continuous data stream.
3) "REAL TIME STEP"—If a real-time process is desired, the last operation is the implementation of the algorithm on board the smart meter and the system is ready to run as follows.
   a) In "DATA ACQUISITION," during its operation, the absorption profile of the electrical device being monitored is acquired and processed in real time. Then its electrical signature is extracted, with an accuracy given by the measurement system, and provided as input to the previously trained algorithm.
   b) In "Load Profiling," based on the considerations made in 1)-c) and 2)-c), the operating state of the monitored system is discriminated and the information returned. In this way, the alternative "*c*" is concluded.

The process is thus completed. It must be considered that such a process also applies to other systems of a different physical nature, not only electrical. Of course, the calculated electrical signature, in addition to being processed for Load Profiling purposes, can be used for further analysis, from which information and indices can be extrapolated to be associated with the identified operational state of the monitored load, as reported in Section II-A. In particular, the "REAL TIME STEP" can be extremely useful in predictive diagnostics. Furthermore, for alternatives "*b*" and "*c*," the electrical signature can be stored and reused occasionally to iterate the *Step 2* again to refine and/or update the capabilities of the profiling algorithm. Such an operation is very advantageous in that performance drops in the process are avoided by refining the training over time, especially in the presence of a natural degradation of the monitored system, which would alter the absorbed electrical signature.

C. Typical Load Profiling Techniques

Different types of processing techniques and methods can be implemented to carry out Load Profiling. The scientific literature is particularly active in this field, in fact, on Load Profiling, articles, and several review studies have been published [27], [28]. In general, the most widely used techniques are: 1) clustering and 2) classification algorithms. Clustering is an unsupervised learning technique, unlike classification, which requires a supervised approach. In both cases, both supervised and unsupervised feature selection techniques can be implemented to improve analysis performance. As regards 1) and 2), the basic concepts of each of these techniques are summarized below to provide the reader with a better understanding of the considerations made in Sections II-B and V.

Regarding 1), in [29] and [30], reviews on clustering methods and algorithms are proposed. In [31], three clustering techniques applied to data from meters at different frequencies are analyzed, while in [32], two-stage clustering is proposed. In [33], a review is made of clustering (and classification) methods applied to residential and industrial electricity scenarios, while Chicco et al. [34] report a comparison of clustering techniques for customer identification in the electricity sector. In general, the most widely used *clustering* algorithms are the *k-means*, the generalized Gaussian model (GMM), and the *agglomerative clustering*. The k-means algorithm is a partition method that groups data by separating samples into *n* groups (partitions) of equal variance, minimizing a criterion known as *inertia* or sum of squares within the cluster. The method GMM looks for a set of multidimensional Gaussian probability distributions that can represent the data under consideration. This quantifies the probability with which a data item belongs to a given cluster. Moreover, each cluster is no longer associated with a sphere in space but with a Gaussian pattern; this represents a significant advantage over *k-means*. *Agglomerative clustering* is a hierarchical algorithm strategy that works "bottom-up." In practice, the algorithm starts by considering each sample as a single cluster and, at each subsequent iteration, recursively combines the two most similar clusters until all samples belong to one large cluster. In general, regardless of the type of algorithm implemented, various evaluation criteria and metrics can be used after clustering [35] to define performance and the appropriate number of clusters. For example, *silhouette coefficient* takes into account both separation and cohesion, can vary in the range $[-1, 1]$ where silhouette values close to 1 are preferred while negative values indicate that the cluster observations "fit" better with other clusters instead of their own, that is, that the final clustering is not correct; *Rand index* is a statistical parameter defined as a measure of the similarity between two data clusters (it is assimilated to accuracy); *inertia index* can be recognized as a measure of the internal coherence of clusters; *Davies–Bouldin score* is defined as the average similarity measure of each cluster with its most similar cluster; *elbow method*; *index of separation of a pair of clusters*; and so on.

With regard 2), for example, through a previously trained classification algorithm, a system for real-time load identification can be realized. This is done by identifying the relationships between the different features of the system and the label associated with them using, for example, regression criteria (for continuous data) or classification models (for discrete data) [36]. Reviews on classification methods and algorithms are given in [37].

Along with clustering and classification techniques, feature selection algorithms are often implemented. In general, feature selection has three main objectives: improve the model accuracy, reduce the computational cost, and produce a more interpretable model. Malhi and Gao [38] and Hopf et al. [39] present general reviews on feature selection methods, while Beckel et al. [40] review specialized feature selection techniques for clustering and classification. In [41], the features of greatest interest for identifying load profiles are selected. Often, feature selection techniques are of the supervised type, thus implying prior knowledge of the ground truth of the analyzed system. However, there are also nonsupervised approaches that do not require knowledge of this parameter [42]. In this case, the selection is based on critical data analysis in terms of variability, statistical distribution, trends, and internal correlations, aimed at identifying and selecting those features that potentially have higher information content. Of interest among these is the KDE method [43], which is a nonparametric statistical method useful for identifying patterns through the estimation of metric spaces, that is, the probability density function. In general, the KDE makes it possible to derive the continuous probability distribution function of a discrete quantitative variable.

III. Methodological Approach

This section presents the tests, choices, and methodologies followed during the analysis of the implemented Load Profiling process. The characteristics of the influence parameters considered, that is, MQ and electrical signature, are first illustrated. Subsequently, the tests conducted and the methods used are described.

A. Monitoring and Electric Signature Characteristics

With regard to MQ, in particular, for the definition of the uncertainty levels analyzed, reference was made to a number of regulations and directives related to energy measuring instruments, including: "Directive 2014/32/EU" of the European Parliament and of the Council on the Harmonization of the laws of the Member States relating to the making available on the market of measuring instruments [44], and *EN50470-1* standard of the "European Committee for Electrotechnical Standardization" (CENELEC) on "Electricity metering equipment (a.c.)" [45]. The latter refers to newly manufactured energy meters intended for residential, commercial, and light industrial use and specifies general requirements, tests, and test conditions for the metering apparatus. Specifically, this standard, according to the metrological quality of the apparatus, discriminates meters into three different accuracy classes: *A*, *B*, and *C* (*A* indicates the "worst" class). In this article, for both *scenarios* analyzed, the metrological conditions of classes *A*, *B*, and *C* were replicated, defining three different levels of MQ, denoted by the acronyms "MQ1," "MQ2," and "MQ3," respectively. For example, "MQ1" indicates an MQ condition of level 1, that is, "worst" typical of lesser-performing commercial meters, while "MQ3" represents the condition

TABLE I

QUALITATIVE CHARACTERISTICS OF THE ELECTRICAL SIGNATURE CONSIDERED IN THE THREE LEVELS: "ESQ1," "ESQ2," AND "ESQ3." REPORT: THE CATEGORIES OF ELECTRICAL PARAMETERS CONSIDERED (TYPE OF PARAMETERS); THE NUMBER OF ELECTRICAL PARAMETERS CALCULATED (NUMBER OF PARAMETERS)

ESQ	Type of parameters	Number of parameters
ESQ1	P-S-N	3
ESQ2	IEEE1459	21
ESQ3	IEEE1459 + IEC61000-4-7	123

with lower uncertainty assimilated to meters with high-level performance. Refer to Section IV for further details on data acquisition methods and techniques.

Regarding the quality of the electrical signature, that is, the number of electrical parameters calculated by the meter, three levels of ESQ were considered, respectively, denoted by the acronyms "ESQ1," "ESQ2," and "ESQ3." The number of parameters considered for each level was chosen to replicate three types of commercial meters with different calculation capabilities. Table I summarizes the electrical parameters considered for the three levels of ESQ described below.

1) "ESQ1" refers to meters with basic calculation capability—only the electrical parameters of the load relating to the measurements of active power (P), apparent power (S), and nonactive power (N) are considered. These turn out to be the main electrical parameters extracted from "classical" metering systems. Their measurement does not necessarily require particularly complex and high-performance hardware, so they can be extracted with simplicity and reduced costs.
2) "ESQ2" refers to meters with basic calculation capability but equipped with filters or evolved meters—in addition to the parameters P, S, and N, the remaining parameters introduced by the standard *IEEE-1459* are also considered. These electrical quantities are of particular interest as they do not necessarily require a hardware enhancement of the meters compared to the "ESQ1" case to be extracted during measurement. In fact, although they can be obtained more quickly via fast Fourier transform (FFT), they can also be obtained by applying filters.
3) "ESQ3" refers to a smart meter with high calculation capabilities—parameters related to *IEC 61000-4-7:2009* are considered in addition to the parameters of "ESQ2." The latter is more complex to extract and necessarily requires a frequency analysis by means of FFT, consequently an upgrade at the hardware and firmware level of the meter is necessary, with an inevitable increase in costs.

The metrics and calculation methods implemented in the two case studies, *simulated* and *real*, are the same and allow a large number of electrical parameters to be extrapolated from the monitored system at each measurement interval, including the ground truth of the operational state. All of these parameters provide an electrical signature with a high degree of detail. Table I summarizes these for the three "ESQ" levels considered. These refer only to device-dependent

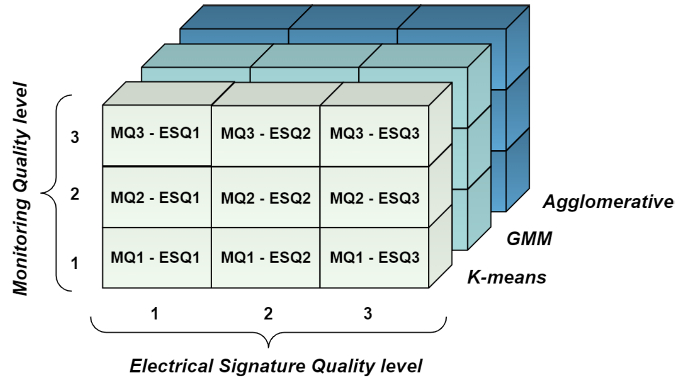

Fig. 2. Tests performed with: three different levels of MQ, three levels of ESQ, and three different Load Profiling algorithms (k-means, GMM, and agglomerative clustering.

electrical quantities, that is, related to the absorbed current spectrum; voltage-dependent parameters alone are neglected as this is imposed in electrical systems. In practice, the power definitions of *IEEE-1459* are implemented, particularly in the single-phase nonsinusoidal case [46], [47]. The electrical parameters calculated according to IEEE1459, for voltage and current, refer to total root mean square (rms), rms of fundamental component, rms of the remaining harmonic content, and dc component. For the power components, they refer to active, apparent, nonactive, and distorted power; power factor and total harmonic distortion parameters. Harmonic behavior of the considered electrical loads has been assessed according to *IEC61000-4-30:2015 "Testing And Measurement Techniques—Power Quality Measurement Methods"* [48] and *IEC61000-4-7:2009 "Assessment of harmonic emissions"* [49]. In detail, in all the harmonic measurements, rms and phase of harmonic groups of voltage and current up to the 50th harmonic order and rms and phase of the harmonic subgroup have been considered, considering a time window of 200 ms (with a frequency resolution of 5 Hz).

B. Test Setup

With reference to the structure of the Load Profiling process illustrated in Fig. 1, this work refers to the case "*a*" described in detail in Section II-B. In practice, an asynchronous Load Profiling process has been implemented with respect to monitoring, that is, the operational state of the considered device is identified at the end of the device acquisition process.

The tests for *simulated* and *real* cases were conducted for three levels of monitoring and ESQ. Furthermore, since the performance of the Load Profiling process may also depend on the type and complexity of the profiling algorithms implemented, the analyses just mentioned were repeated three times using three different types of Load Profiling algorithms: "*k-means*," Gaussian mixture modeling (GMM), and "*agglomerative clustering*." This provided a broader and more general view of the problem. The input data to the clustering algorithms were first appropriately normalized using the *z-score* method [50], that is, mean 0 and standard deviation 1. Fig. 2 summarizes the tests conducted.

For example, considering the two extreme cases, the "worst" and the "best," we have: "MQ1-ESQ1" low quality of

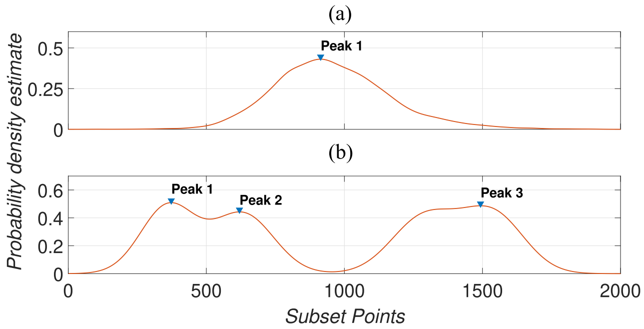

Fig. 3. Application of the KDE method to the "Desk Lamp" load (four operational states). In (a), an example of an unselected feature with a unimodal distribution is shown. In (b), an example of a selected feature is given, the distribution being multimodal and therefore potentially useful for identifying the operational states. (a) Nonactive power (N). (b) Active power (P).

monitoring (high measurement uncertainty) and low quality of electrical signature (few parameters calculated by the meter); "MQ3-ESQ3" high quality (low measurement uncertainty) and high quality of electrical signature (many parameters calculated by the meter). While the opposite extreme cases refer to meters: with high metrological performance but low calculation capabilities ("MQ3-ESQ1"); with low metrological performance but high calculation capabilities ("MQ1-ESQ3"). In between are the remaining intermediate cases. In terms of the methods used, algorithms of different natures and operations were chosen, to highlight the potential and advantages obtained in a Load Profiling process by acting on the different impacting factors.

In addition, as a feature selection method, the KDE described in Section II-C. Being an unsupervised technique, KDE does not require knowledge of the ground truth in the selection phase. The only consideration made is that only multistate loads are considered in this work. Therefore, from the set of starting characteristics defined by the level of "ESQ" considered for each load, features with unimodal distribution were discarded through the KDE method. Fig. 3 shows an application of the KDE methods to the load "Desk lamp" considering two different features associated with two different probability distributions: 1) unimodal and 2) multimodal. Using KDE applied to a subset of the starting samples, only features with a number of peaks in their probability distribution greater than one are selected. The peaks in the distribution represent potential areas of point densities attributable to operational states of operation. A characteristic with a unimodal distribution, therefore, represents a nonsignificant parameter for a multistate load, being almost constant as the operating state changes. It should be considered that the number of peaks shown by a feature is not necessarily exactly the actual number of operating states. For example, Fig. 3 shows a case of unselected and selected features for a load used in this work ("Desk Lamp") which actually has four operational states.

The silhouette coefficient was considered to assess the goodness of clustering [35], which is extremely useful for simultaneously assessing the cohesion and separation of the clusters obtained. The number of identified operational states, hence classes of the system, according to this metric, is the one at the highest silhouette coefficient. The characteristics of the metrics and algorithms are described in Section II-C. All processing operations were performed in the MATLAB[1] and Python[1] environments [51].

TABLE II

SIMULATED LOADS SPECIFICATIONS. IT SHOWS THE IDENTIFIER (DEVICE), THE TYPE OF THE DEVICE (TYPE), THE NUMBER OF OPERATING STATES (N_S), AND THE NOMINAL ACTIVE POWER (P_{Nom})

Device	Type	N_S	P_{Nom} [W]
D1	Fan Heater	4	1900
D2	Desk Lamp	4	10
D3	Smartphone Charger	2	25

IV. ANALYZED SCENARIOS

All the analyses were carried out in simulated and real scenarios. As for the *simulated scenario*, a public dataset has been considered and, in Section IV-A, its main characteristics and how it was used are presented. As for the *real scenario*, in Section IV-B, the characteristics of the chosen devices, the measurement setup, and the process to confirm the acquired data are described.

A. Simulated Scenario

As for the *simulated* case, the dataset used in this work is "eLAMI" whose characteristics are given in [6]. The dataset is publicly accessible at *IEEE-Dataport* [52]. It represents a residential energy dataset with innovative characteristics compared to rivals in the literature today [53]. In particular, it is simulated from experimentally acquired data and with consumption patterns that are true to reality, thanks to an *ad hoc* stochastic model. The main feature of "eLAMI" is the large number of electrical parameters (433) provided for each operational state of each device, with a measurement time of 5 s, as described in [6]. In addition to the calculated quantities, the operational state (ground truth) of each electrical load is also provided. In general, the metrics and calculation methods implemented for the extraction of electrical parameters are based on the implementation of the *IEEE-1459* and *IEC61000-4-7:2009* standards, in accordance with the ESQ levels defined in Section III-A.

Analyses were conducted on the loads present in "eLAMI," for the sake of synthesis, Section V shows the results for three electrical loads in the dataset that were considered most interesting. In particular, the devices shown are: "*D1: Fan Heater*," "*D2: Desk Lamp*," and "*D3: Smartphone charger*." This choice was made, in addition to synthesis, because these were the most interesting devices to be analyzed in a Load Profiling process: D1 and D2 present the largest number of operational states among the loads available in "eLAMI," in D2 and D3 the "off" state corresponds to standby and therefore nonzero absorption that is confused with the state immediately following, and D1 is interesting from the point of view of load type, being comparable to an electrical machine. Table II summarizes the information on the loads just mentioned, extracted from [6].

[1]Trademarked.

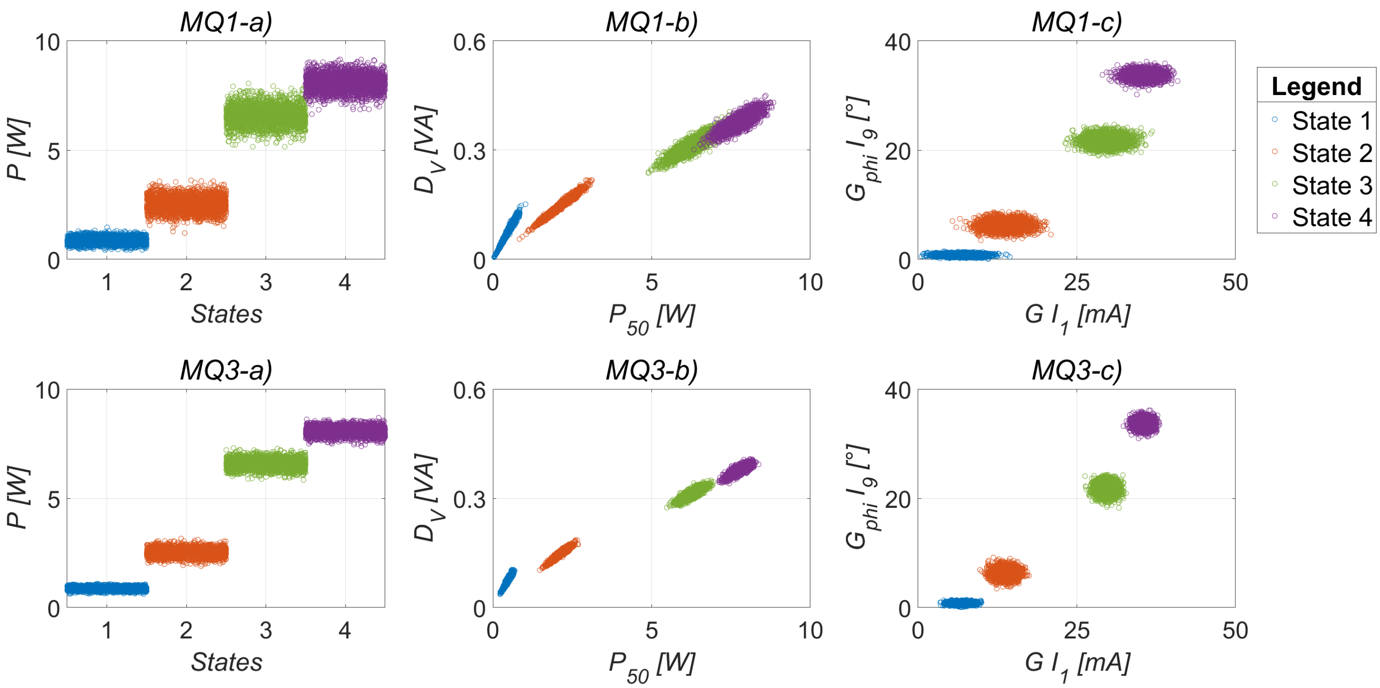

Fig. 4. Simulated electrical load "Desk Lamp." Comparison of two MQ levels and three different sets of features belonging to the three ESQ levels considered. The labels in the legend, from "*State-1*" to "*State-4*," refer to the operating states of the electrical load considered, to which the nominal active power (P) levels correspond, respectively: 0.8, 2.5, 6.5, and 8 W. Each operating state is highlighted in a different color.

For the sake of clarity and to provide a better understanding of the problem, the example in Fig. 4 shows the evolution of certain electrical parameters for different operating states of the "Desk Lamp" load at different levels of MQ and electrical signature. In particular, the quality levels "MQ1" and "MQ3" and three sets of features belonging, respectively, to the levels: 1) "ESQ1," 2) "ESQ2," and 3) "ESQ3" are compared. In detail, for the simulated device considered: *MQ1-a)* and *MQ3-a)* show the absorption profiles in terms of active power (P) with respect to the operational state; *MQ1-b)* and *MQ3-b)* show the voltage distortion power (D_V) as a function of the power factor at the fundamental (P_{50}); *MQ1-c)* and *MQ3-c)* show the phase trend of the ninth harmonic current group ($G_{\text{phi}}I_9$) as a function of the amplitude of the first harmonic current group (GI_1). From the figure, it is possible to appreciate the behavior of the device under the different test conditions considered and the criticalities in Load Profiling. It is evident how the quality of the monitoring system and the natural variability of the device leads in some cases to an overlap between the operating states. At the same time, it is possible to appreciate the "separation" of operating states when more significant features are considered and the quality of monitoring is increased. The other loads considered show conceptually more or less similar behavior, so for reasons of synthesis, the same representation is omitted.

Regarding the analysis interval, to deal with a sufficient number of data, tests were conducted for each load on a number of measurements per operational state equal to 5000 samples. In practice, a balanced subdataset was extracted from "eLAMI" with a small size and an equal number of points per operational state for the different loads. Regarding the quality of monitoring for the *simulated scenario*, as reported by the authors, "eLAMI" in its basic version corresponds to an "MQ2" level defined in this article. As for the other two MQ levels, using the "eLAMI" dataset and with the same methods and techniques implemented in [6], two more datasets were generated, with greater and lesser variability than the basic version, corresponding to "MQ1" and "MQ3."

B. Real Scenario

With regard to the real electrical loads used for the analysis, a bank of dimmer lamps ("*Lamp*") controlled by a silicon-controlled rectifier (SCR) was considered as a sample of the residential case and a single-phase asynchronous motor powered by an inverter (Mot + Inv) as a sample of the light industry case. Thanks to the control offered by the inverter and the dimmer, it was possible to realize an arbitrary number of operating states for both devices. In fact, seven operational states were tested for the "Lamp" case, and eight operational states for the "Mot + Inv" case. All the tests were carried out in the Laboratory of Industrial Measurements (LAMI) of the University of Cassino and Southern Lazio.

Concerning the three MQ levels, "MQ1," "MQ2," and "MQ3," according to Section III-A, three different metrological acquisition conditions were implemented for both real loads treated. Considering a simple measurement chain model of a smart meter, the main hardware elements are represented by the voltage and current transducers and by the data acquisition systems. Their metrological performance defines the performance of all measurement chains. Considering this model, to implement the three "MQ" uncertainty levels, three different voltage and current probes and three different resolutions of the acquisition system were considered. In particular, as for

TABLE III

METROLOGICAL SPECIFICATIONS OF THE SETUP USED TO ACQUIRE THE ACTUAL LOADS FOR THE THREE MQ LEVELS. REPORTS INFORMATION IN TERMS OF THE TYPE OF MEASURED QUANTITY (TYPE), PROBE SCALING FACTOR (K_{probe}), PROBE FULL SCALE (FS_{probe}), PROBE ACCURACY ($\text{ACC}_{\text{probe}}$), AND FULL-SCALE ACQUISITION SYSTEM (FS_{Acq})

MQ	Type	FS_{probe}	$\text{Acc}_{\text{probe}}$	K_{probe}	FS_{ADC}
MQ1	V	1 kV	±2.5 %	1:200 V	±8 V
	I	50 A_{RMS}	±3 %	500 mV/A	±8 V
MQ2	V	1.3 kV	±2 %	1:500 V	±0.8 V
	I	15 A_{RMS}	±2 %	100 mV/A	±2 V
MQ3	V	1.5 kV	±1 %	1:500 V	±0.8 V
	I	30 A_{RMS}	±1 %	100 mV/A	±0.8 V

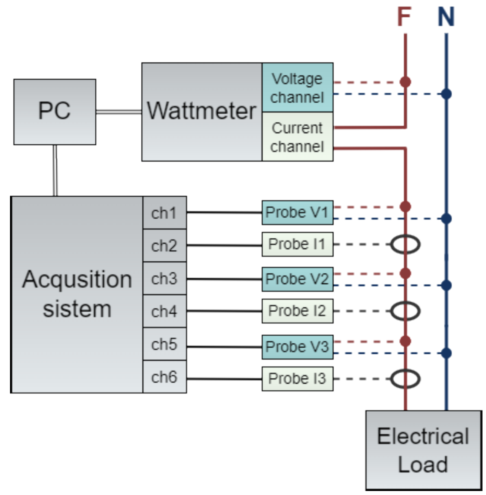

Fig. 5. Block diagram of the data acquisition setup and validation of the acquisition process.

the voltage transducer, a differential probe "Fluke DP120," a differential probe "Tektronix P5200" and a differential probe "LeCroy HVD3106A-6M" (with a "LeCroy TPA10" probe adapters) were, respectively, adopted for "MQ1," "MQ2," and "MQ3." As for the current transducer, a "Siglent CP4050" current probe, a "Tektronix TCP202A" current probe (powered by a "Tektronix-1103" power supply), and a "Teledyne LeCroy CP030A" current probe (with a probe adapter) were, respectively, used for "MQ1," "MQ2" and "MQ3." Acquisitions were carried out using a six-channel data acquisition system to make simultaneous acquisitions from all the aforementioned voltage and current probes. A synchronized data acquisition system based on a "TiePie HS6" and a "TiePie HS5," and a suitable acquisition software developed in the MATLAB environment were adopted for all the experiments. To change the resolution of the data acquisition system, the number of bits (n_{bit}) used was chosen equal to 12, while the full scale of each acquisition channel was set to different values under the three "MQ" conditions. The sampling frequency was equal to 5 kS/s with a measurement observation time of 1 s. As the sampling rate is the same under the three measurement operating conditions, there are no changes in the phase resolution of the measured quantities between "MQ1," "MQ2," and "MQ3." With regard to the acquisition interval, one hour of monitoring was carried out for each tested operating state of the loads considered, for a total of 3600 measurements per tested operating state. All the measurements in the frequency domain were carried out with a frequency resolution of 5 Hz corresponding to a time window of 200 ms to be compliant with the *IEC 61000-4-7:2009* standard.

Table III details all the sampling specifications used, while the block diagram of the experimental setup for data acquisition is shown in Fig. 5.

For the sake of completeness, a summary and general analysis in terms of uncertainty propagation is given below, valid for the voltage and current quantities acquired under the different metrological "MQ" conditions realized. In terms of uncertainty assessment, the traditional Guide to the Expression of Uncertainty in Measurement (GUM) approaches or the Monte-Carlo method can be adopted. In detail, from a theoretical point of view, in relation to (1), the measured value and its uncertainty depend on all the elements of the measurement chain involved: the measurand (X_m), the probe (X_{probe}), the conditioning (X_{cond}), the analog-to-digital converter (X_{ADC}), the environment (X_{env}), and so on [54]. Considering that the environmental conditions are controlled, since all experiments were performed in a reference and notified laboratory, and considering that there is no conditioning system, the uncertainty contributions are reduced to those arising from the probe and the data acquisition system. This reduces the complexity of (1), to a simple case (i.e., a voltage or current probe and a DAQ system). This case has been analyzed by the articles referenced in [55] and [56]. Both articles lead to the conclusion that even if the Monte Carlo approach is generally to be preferred, in this case, the two approaches lead to the same results. For this reason, for the sake of simplicity, in the article, we have preferred to use the traditional GUM approach. Given (1) in explicit form, applying the uncertainty propagation law given in (2), we then derive the two uncertainty contributions that characterize the measurement chain and have the greatest impact on the measurement, namely those related to the probe used (u_{probe}) and the contribution introduced by the data acquisition system (u_{ADC}). Equations (3) and (4) give the definition of the latter. By substituting these expressions with the nameplate data given in Table III, it is possible to obtain the uncertainty values for the voltage and current quantities, which define the metrological quality of the measurement system implemented in the three cases considered. The reported analysis, although trivial and well known in the scientific literature, highlights the effect of the influence parameters considered to be preponderant in the "MQ" and consequently in the analysis process

$$X_{\text{meas}} = f\left(X_m, X_{\text{probe}}, X_{\text{cond}}, X_{\text{ADC}}, X_{\text{env}}, \ldots\right) \tag{1}$$

$$u_{X_{\text{meas}}} = \sqrt{\sum_i \left(\frac{\delta f}{\delta x_{u_{x_i}}}\right)^2 \cdot u_{x_i}^2} \tag{2}$$

$$u_{\text{probe}} = \frac{\text{Acc}_{\text{probe}}}{\sqrt{3}} \tag{3}$$

$$u_{\text{ADC}} = \frac{\text{FS}_{\text{ADC}}}{2^{n_{\text{bit}}} \cdot \sqrt{12}}. \tag{4}$$

TABLE IV

COMPARISON BETWEEN WT3000 AND IMPLEMENTED ACQUISITION SYSTEM FOR ELECTRICAL LOAD "LAMP." THE THREE-MQ-LEVEL COMPARISON IN TERMS OF MEAN VALUE AND STANDARD DEVIATION OF ACTIVE POWER (P)

Load - State	[W]	P_{WT3000}	P_{MQ1}	P_{MQ2}	P_{MQ3}
Lamp - 1	μ	0	0	0	0
	σ	0	0	0	0
Lamp - 2	μ	113.22	114.25	113.98	113.46
	σ	0.31	2.19	1.24	0.68
Lamp - 3	μ	115.45	116.44	115.82	115.51
	σ	0.25	2.98	1.79	0.71
Lamp - 4	μ	100.59	102.41	101.99	101.63
	σ	0.34	3.32	2.37	0.89
Lamp - 5	μ	140.42	142.86	141.93	141.69
	σ	0.29	3.58	2.51	0.93
Lamp - 6	μ	191.52	193.08	192.33	191.73
	σ	0.41	3.97	2.68	1.06
Lamp - 7	μ	59.51	61.39	60.88	60.36
	σ	0.28	3.03	1.95	0.88

To validate and calibrate the considered measurement chain, namely the voltage and current probes, the data acquisition system, and the implemented measurement algorithm, quantities measured on the two considered loads were compared with a reference power meter. In particular, a traceable "Yokogawa Precision Power Analyzer WT3000" reference wattmeter equipped with the harmonic analysis toolbox was adopted, as shown in Fig. 5. The specifications adopted for the WT3000, in relation to the options made available by the manufacturer [57], are as follows: *sampling frequency*—Clock B 189.394 kS/s; *data update rate*—1 s for the P, S, and N power measurements and 200 ms for the harmonic measurements that correspond to a frequency resolution of 5 Hz; *wiring system*—1P2W; *measurement functions*—Type2; *measurement range*—auto. The different quantities calculated by the WT3000 were exported from the instrument via an IEEE-488 bus and a measurement software implemented in the Labview[1] environment. To summarize, Table IV shows the comparison in terms of mean value and standard deviation of the active power (P) extracted from the WT3000 and those processed by the implemented system for the three levels of MQ for each operating state of the monitored load "Lamp." The comparison shows both the compatibility with the reference and the increasing of the measurement uncertainty decreasing the "MQ" level.

Similar to the previous case, also for the *real scenario*, the metrics and calculation methods implemented for the extraction of the electrical parameters, considered in the Load Profiling process, are based on the implementation of the standards *IEEE-1459*, *IEC61000-4-30:2015*, and *IEC61000-4-7:2009*, in accordance with the quality levels of the electrical signatures defined in Section III-A. Furthermore, in analogy to the considerations made in the previous case, for the sake of completeness, Fig. 6 shows the evolution of the absorption profiles of the *real* electrical loads considered "Lamp" and "Inv + Mot." In this case, given the large number of tested operating states of the two devices, only the trend in active power (P) relative to the intermediate monitoring quality level "MQ2" is shown for each. In this way, it is still possible to observe the behavior of the devices considered and the criticality offered in terms of Load Profiling. Given the overlap between the states, it is evident from the figure that it is of fundamental interest to search for further significant and representative parameters of the electrical signature for identification in each of the tested operating states.

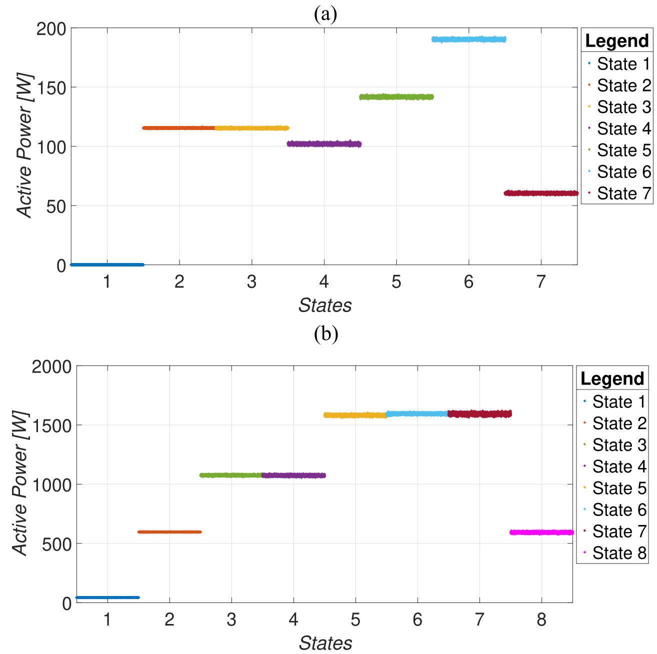

Fig. 6. Active power profile (P) measured at monitoring quality level "MQ2," for both *real* electrical loads considered. For (a), the labels in the legend, from "*State-1*" to "*State-7*," refer to the tested operating states to which the nominal active power levels correspond, respectively: 0, 115, 115.5, 102, 140, 190, and 60 W. Similarly, for (b), from *State-1"* to "*State-8*" correspond to the nominal active power levels, respectively: 43, 595, 1075, 1070, 1580, 1595, 1590, and 590 W. Each operating state is highlighted in a different color.

V. RESULTS

This section reports the results obtained from the tests conducted in terms of Load Profiling, varying the three influencing parameters: MQ, ESQ, and type of profiling algorithm. In summary, the loads considered refer to a simulated and a real scenario. As for the simulated scenario, the selected devices were a "Fan Heater," a "Desk Lamp," and a "Smartphone Charger." They have 4, 4, and 2 operating states, respectively. In the real scenario, a bank of dimmer lamps ("Lamp") and a single-phase asynchronous motor powered by an inverter ("Mot + Inv"), with 7 and 8 operating states, respectively, were considered. The tests were conducted for three MQ levels ("MQ1," "MQ2," and "MQ3"), three ESQ levels ("ESQ1," "ESQ2," and "ESQ3") and three types of Load Profiling algorithms ("*k-means*," GMM, and "*Agglomerative Clustering*"), as shown in Fig. 2. All explanations and descriptions of the choices made are given in Sections III and IV.

For the sake of clarity, the results are reported in tabular form. In analogy to what was done in the previous chapters, they are broken down by the type of *scenario* analyzed. For each table reported, the "$Score_{\mathrm{sh}}$" value (evaluation metrics used), that is, the maximum silhouette coefficient obtained for that test, and the "$Err\ state_{\mathrm{sh}}$" value, that is, the percentage error committed in terms of the operational states identified through the silhouette coefficient with respect to ground truth, are reported within each cell. For the sake of completeness, (5) gives the definition used for the calculation of "$Err\ state_{\mathrm{sh}}$," where "$N_{\mathrm{sh\ state}}$" represents the number of states identified by clustering at the maximum silhouette coefficient, and GT_{state} represents the number of ground-truth operational states of the analyzed load. In this way, it is possible to tell immediately whether the algorithm overestimates, underestimates, or correctly identifies the states of the system. The best condition is obtained when "$Score_{\mathrm{sh}}$" is close to one and "$Err\ state_{\mathrm{sh}}$" is equal to zero. In addition, the cells are colored according to the error committed, so that the behavior obtained from the analysis is immediately apparent

$$\mathrm{Err\ state_{sh}} = \frac{N_{\mathrm{sh\ states}} - GT_{\mathrm{states}}}{GT_{\mathrm{states}}} \times 100. \quad (5)$$

A. Tests Results: "Simulated Scenario"

Starting with the electrical load "Fan Heater," it can be seen from Fig. 7 that at the quality level "ESQ1," the process always tends to underestimate the number of operating states detecting three states instead of four with an error of −25%. Looking at the nominal power of the considered four states [6], it can be highlighted as the power states have nominal values of 0, 15, 925, and 1900 W, respectively. In this case, as shown in the previous chapter, the error is related to incorrect cluster separation of the first and second states even in the presence of low uncertainty levels (i.e., "MQ1"). As the "ESQ" level increases, adding new features to the clustering algorithms, a considerable improvement is highlighted. In addition to correctly identifying the number of states, the silhouette coefficient always assumes a fairly high value for both "ESQ2" and "ESQ3." It follows that for this load, it is more convenient to optimize the quality of the electrical signature than the quality of the monitoring. Therefore, a meter that is less performing from a metrological point of view but slightly advanced from a processing point of view, is sufficient to correctly identify the number of operating states for this analyzed load. In any case, the progressive improvement of the analysis as the "MQ" parameter also increases is evident. Since the same behavior was also obtained for the remaining simulated and real loads when varying the algorithm, only the results of the GMM algorithm, that is, the intermediate case in Fig. 2, are shown below.

Concerning the electrical load "Desk Lamp," from Fig. 8, it can be highlighted the following.

1) For the "MQ1" value, whatever the "ESQ" levels, it is impossible to correctly profile the operating states. In detail, an error equal to −50% is obtained meaning that only two states are recognized. Looking at Fig. 4 ("*MQ1-a)*"), this can be explained considering that states 1 and 2 are very close and the same happens for states 3 and 4.
2) For the "MQ2" value, a correct Load Profiling is obtained only for the "ESQ3" signature level but a low silhouette score is performed. This means that a high number of features partially compensated for the variability due to the uncertainty measurement.
3) For "MQ3" whatever the "ESQ" level, a correct Load Profiling is achieved. When "ESQ3" is experienced the silhouette score reaches a good value that highlights the good shape of the cluster and the reliability of the Load Profiling algorithm. Looking at Fig. 4 ("*MQ3-c)*"), it is possible to appreciate the good shape of the four power states clusters. That is, there is a tendency to obtain clusters characterized by increasingly better cohesion and separation. From these considerations, it follows that for this type of load, in a profiling process, it is not so important to optimize the processing, and thus the computational capabilities of the meter, but it is more advantageous first to improve the quality of the measurement system.

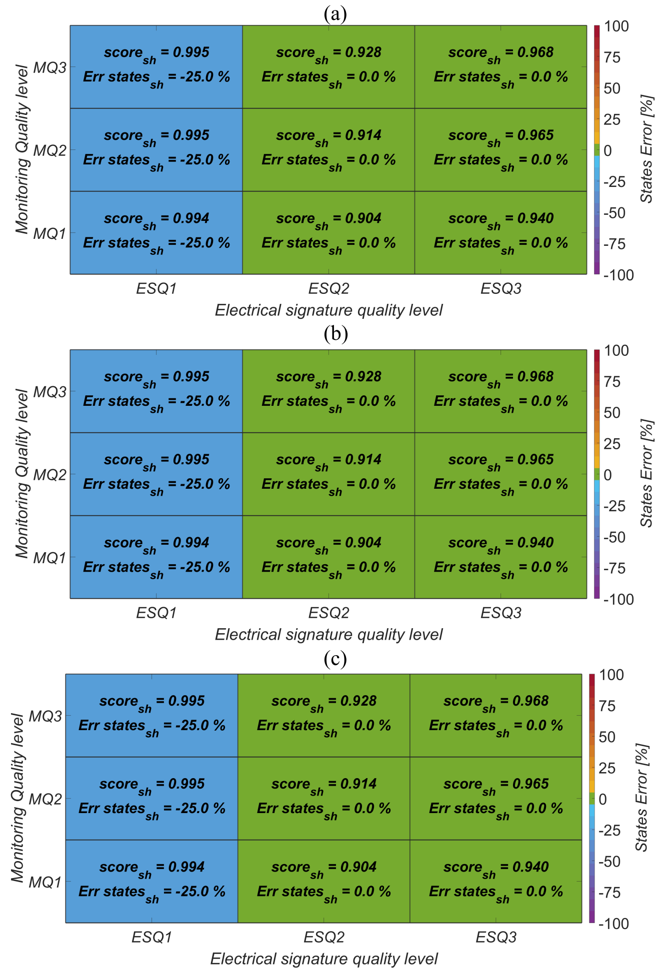

Fig. 7. Results obtained for the "Fan Heater" load with the clustering algorithms: (a) *k*-means, (b) Gaussian mixture model, and (c) agglomerative.

Finally, for the *simulated scenario*, Fig. 9 shows the results obtained for the "Smartphone Charger" load. Given the "simplicity," in terms of the number of states of the analyzed

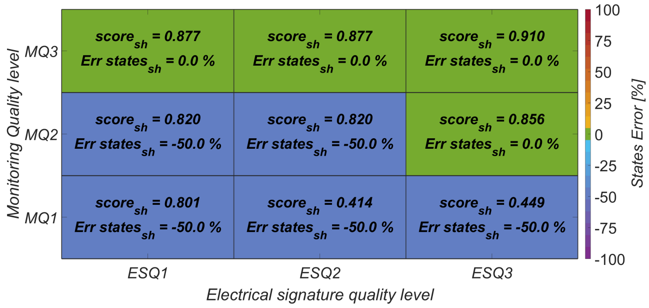

Fig. 8. Results obtained for load “Desk Lamp” by Gaussian mixture model clustering algorithm.

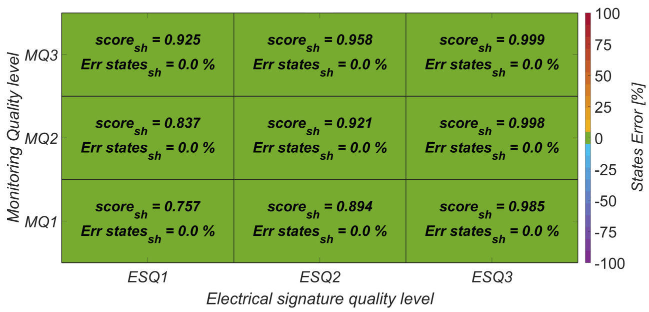

Fig. 9. Results obtained for load “Smartphone Charger” by Gaussian mixture model clustering algorithm.

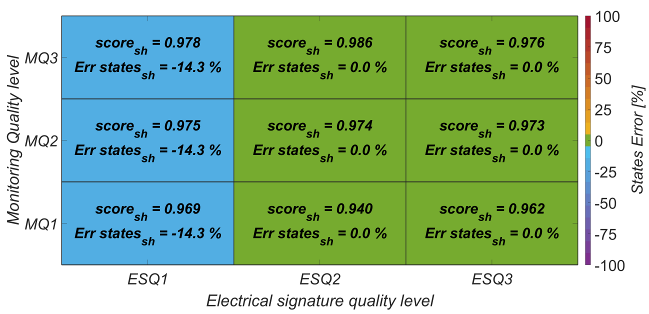

Fig. 10. Results obtained for load “Lamp” by Gaussian mixture model clustering algorithm.

load, the profiling process always returns the correct number of operational states. Moreover, in this case, the effect of the improved quality parameters of monitoring and electrical signature is much more evident than in previous loads. In fact, we go from a silhouette coefficient value of less than 0.8 in the worst case to a value extremely close to 1 in the best case.

B. Tests Results: “Real Scenario”

With regard to the “Lamp” load, that is, the dimmer lamp bank, it is evident from Fig. 10 that the load profile results on this electrical load are similar to those obtained on the “Fan Heater” analyzed in the simulated scenario. In other words, in condition “ESQ1,” the process never identifies the correct number of operating states. At the “ESQ2” and “ESQ3” levels, on the other hand, it is evident that the algorithm always correctly identifies the number of clusters, and thus the operational load states. Furthermore, as explained in the previous subsection, the output obtained at levels “ESQ2” and

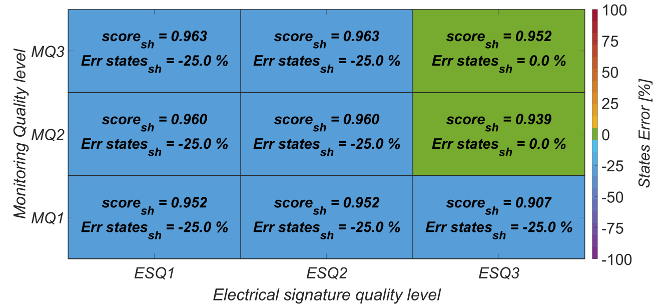

Fig. 11. Results obtained for load “Mot + Inv” by Gaussian mixture model clustering algorithm.

“ESQ3” is characterized by a high value of the silhouette coefficient, which indicates a correct separation and cohesion of the clusters.

Lastly, the results obtained for the electrical load “Mot + Inv” are shown in Fig. 11. This load is characterized by a high number of operating states (8), very close to each other in terms of power as shown in Fig. 6(b) of Section IV-B. It can be seen that at levels “ESQ1” and “ESQ2,” as well as “MQ1,” it is never possible to correctly identify the operating states of the system considered. In fact, for a correct analysis in terms of Load Profiling, it is necessarily only necessary to work at the “MQ2-ESQ3” and “MQ3-ESQ3” levels. That is, it is necessary to have medium- to high-quality monitoring and a representation of the electrical signature as complete as possible. Consider the use of high-performance smart meters. Under these conditions, the high number of parameters considered and the medium and high level of MQ lead to the optimal identification of operational states, characterized by a high silhouette coefficient value.

VI. Conclusion

This article analyses the crucial role played by smart meters in the Load Profiling task considering the influence of the MQ, that is, the measurement uncertainty, and the ESQ, that is, the number of electrical parameters derived from the electrical signature. Some conclusions can be derived from the proposed study.

1) Even if the increasing of the “MQ” improves the LP process, not always an increase in the number of measured parameters (i.e., “ESQ”) brings benefits in the Load Profiling process. This happens when many of the features taken into consideration do not have great sensitivity with respect to changes in the energy states. In these cases, even if a large number of features is calculated, the poor sensitivity could lead the Load Profiling algorithms to wrong decisions leading to an underestimation or an overestimation of the number of actual states of the system under test. In these cases, the high number of not-significant features returns worse results with respect to the use of traditional P, N, and S power indexes.
2) To face this problem, it could be useful to introduce statistical unsupervised methods, such as the proposed

KDE, which allows to eliminate the parameters with poor sensitivity.

3) For the Load Profiling process, in many cases, uncertainties lower than those required by the energy billing (referred to as "MQ3" in the manuscript) are sufficient to correctly assess the task allowing to reduce metering infrastructure costs. In this sense, this work could be useful to the developer since it suggests a method to define the target "MQ" level and the useful features required for the given Load Profiling task.
4) "MQ" and "ESQ" values influence the Load Profiling process in the same way whatever the three clustering algorithms considered in the manuscript.
5) The methodological approach considered in the article could be useful to analyze also Load Profiling problems related to other physical quantities (i.e., thermal energy profiling) or *multiphysical* systems, allowing to user to define a target "MQ" and a minimum number of useful features to be adopted.

REFERENCES

[1] International Energy Agency. *Digitalization and Energy*. Accessed: Mar. 31, 2023. [Online]. Available: https://iea.blob.core.windows.net/assets/b1e6600c-4e40-4d9c-809d-1d1724c763d5/DigitalizationandEnergy3.pdf

[2] COP26. (2022). *COP26: The Glasgow Climate Pact*. [Online]. Available: https://ukcop26.org/wp-content/uploads/2021/11/COP26-Presidency-Outcomes-The-Climate-Pact.pdf

[3] *Directive (EU) 2018/2002 of the European Parliament and of the Council—Of 11 December 2018—Amending Directive 2012/27/EU on Energy Efficiency*. Accessed: Mar. 31, 2023. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018L2002&from=IT

[4] C. Bourelly, D. Capriglione, C. Carissimo, F. Milano, and L. Tari, "Measurement and applications: The role of communication technologies in developing distributed measurement systems and measurement applications," *IEEE Instrum. Meas. Mag.*, vol. 26, no. 4, pp. 19–26, Jun. 2023.

[5] CE. (2022). *Directive 2012/27/CE of the European Parliament And of the Council*. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32012L0027&from=IT

[6] T. Luca, G. Berrettoni, C. Bourelly, G. Cerro, D. Capriglione, and L. Ferrigno, "ELAMI—An innovative simulated dataset of electrical loads for advanced smart energy applications," *IEEE Access*, vol. 10, pp. 91177–91191, 2022.

[7] M. Pau et al., "Design and accuracy analysis of multilevel state estimation based on smart metering infrastructure," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 11, pp. 4300–4312, Nov. 2019.

[8] G. Aurilio, D. Gallo, C. Landi, M. Luiso, and G. Graditi, "A low cost smart meter network for a smart utility," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, May 2014, pp. 380–385.

[9] J. Liu, F. Ponci, A. Monti, C. Muscas, P. A. Pegoraro, and S. Sulis, "Optimal meter placement for robust measurement systems in active distribution grids," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 5, pp. 1096–1105, May 2014.

[10] L. De Oro Arenas, G. de Azevedo e Melo, and C. A. Canesin, "A methodology for power quantities calculation applied to an FPGA-based smart-energy meter," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[11] D. Gallo, C. Landi, N. Pasquino, and N. Polese, "A new methodological approach to quality assurance of energy meters under nonsinusoidal conditions," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 5, pp. 1694–1702, Oct. 2007.

[12] C. Laughman et al., "Power signature analysis," *IEEE Power Energy Mag.*, vol. 1, no. 2, pp. 56–63, Mar. 2003.

[13] A. Bellini, F. Filippetti, G. Franceschini, C. Tassoni, and G. B. Kliman, "Quantitative evaluation of induction motor broken bars by means of electrical signature analysis," *IEEE Trans. Ind. Appl.*, vol. 37, no. 5, pp. 1248–1255, Apr. 2001.

[14] Z. Yan and H. Wen, "Performance analysis of electricity theft detection for the smart grid: An overview," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–28, 2022.

[15] M. Isnen, S. Kurniawan, and E. Garcia-Palacios, "A-SEM: An adaptive smart energy management testbed for shiftable loads optimisation in the smart home," *Measurement*, vol. 152, Feb. 2020, Art. no. 107285. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224119311492

[16] B. Stephen, A. J. Mutanen, S. Galloway, G. Burt, and P. Järventausta, "Enhanced load profiling for residential network customers," *IEEE Trans. Power Del.*, vol. 29, no. 1, pp. 88–96, Feb. 2014.

[17] H.-X. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 16, no. 6, pp. 3586–3592, Aug. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032112001438

[18] M. Negnevitsky, P. Mandal, and A. K. Srivastava, "An overview of forecasting problems and techniques in power systems," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2009, pp. 1–4.

[19] M. Bourdeau, X. Q. Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, "Modeling and forecasting building energy consumption: A review of data-driven techniques," *Sustain. Cities Soc.*, vol. 48, Jul. 2019, Art. no. 101533. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2210670718323862

[20] A. J. Conejo, J. M. Morales, and L. Baringo, "Real-time demand response model," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 236–242, Dec. 2010.

[21] A. Amodei, D. Capriglione, L. Ferrigno, G. Miele, G. Tomasso, and G. Cerro, "A rule-based approach for detecting heartbleed cyber attacks," in *Proc. IEEE Int. Symp. Meas. Netw.*, Jul. 2022, pp. 1–6.

[22] T. S. Zabkowski, M. Bator, and A. Orlowski, "Smart metering—A brief overview of projects, benefits and applications," *Inf. Syst. Manage.*, vol. 1, pp. 72–83, Jan. 2012.

[23] K. Weranga, S. Kumarawadu, and D. P. Chandima, *Smart Metering Applications*. Singapore: Springer, 2014, pp. 115–124, doi: 10.1007/978-981-4451-82-6_6.

[24] A. L. Shah, W. Mesbah, and A. T. Al-Awami, "An algorithm for accurate detection and correction of technical and nontechnical losses using smart metering," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8809–8820, Nov. 2020.

[25] Z. Ren, T. Lin, K. Feng, Y. Zhu, Z. Liu, and K. Yan, "A systematic review on imbalanced learning methods in intelligent fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–35, 2023.

[26] X. Ding and Q. He, "Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 8, pp. 1926–1935, Aug. 2017.

[27] E. Proedrou, "A comprehensive review of residential electricity load profile models," *IEEE Access*, vol. 9, pp. 12114–12133, 2021.

[28] L. Chuan and A. Ukil, "Modeling and validation of electrical load profiling in residential buildings in Singapore," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2800–2809, Sep. 2015.

[29] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999.

[30] A. Saxena et al., "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, pp. 664–681, Dec. 2017.

[31] R. Granell, C. J. Axon, M. Kolokotroni, and D. C. H. Wallom, "A reduced-dimension feature extraction method to represent retail store electricity profiles," *Energy Buildings*, vol. 276, Dec. 2022, Art. no. 112508.

[32] O. Y. Al-Jarrah, Y. Al-Hammadi, P. D. Yoo, and S. Muhaidat, "Multi-layered clustering for power consumption profiling in smart grids," *IEEE Access*, vol. 5, pp. 18459–18468, 2017.

[33] S. Yilmaz, J. Chambers, and M. K. Patel, "Comparison of clustering approaches for domestic electricity load profile characterisa tion—Implications for demand side management," *Energy*, vol. 180, pp. 665–677, Aug. 2019.

[34] G. Chicco, R. Napoli, and F. Piglione, "Comparisons among clustering techniques for electricity customer classification," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 933–940, May 2006.

[35] I. Pauletic, L. N. Prskalo, and M. B. Bakaric, "An overview of clustering models with an application to document clustering," in *Proc. 42nd Int. Conf. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2019, pp. 1659–1664.

[36] A. Amodei, D. Capriglione, G. Cerro, L. Ferrigno, G. Miele, and G. Tomasso, "A measurement approach for inline intrusion detection of heartbleed-like attacks in IoT frameworks," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023.

[37] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: A review of classification and combining techniques," *Artif. Intell. Rev.*, vol. 26, no. 3, pp. 159–190, Nov. 2006.

[38] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 6, pp. 1517–1525, Dec. 2004.

[39] K. Hopf, M. Sodenkamp, I. Kozlovkiy, and T. Staake, "Feature extraction and filtering for household classification based on smart electricity meter data," *Comput. Sci.-Res. Develop.*, vol. 31, no. 3, pp. 141–148, Aug. 2016.

[40] C. Beckel, L. Sadamori, T. Staake, and S. Santini, "Revealing household characteristics from smart meter data," *Energy*, vol. 78, pp. 397–410, Dec. 2014.

[41] P. Nystrup, H. Madsen, E. M. V. Blomgren, and G. de Zotti, "Clustering commercial and industrial load patterns for long-term energy planning," *Smart Energy*, vol. 2, May 2021, Art. no. 100010.

[42] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "A review of unsupervised feature selection methods," *Artif. Intell. Rev.*, vol. 53, no. 2, pp. 907–948, Feb. 2020.

[43] M. S. Uddin, A. Kuh, Y. Weng, and M. Ilic, "Online bad data detection using kernel density estimation," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2015, pp. 1–5.

[44] (2014). *Directive 2014/32/EU of the European Parliament and of the Council of 26, February 2014 on the Harmonisation of the Laws of the Member States Relating to the Making Available on the Market of Measuring Instruments (Recast) Text With EEA Relevance*. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014L0032

[45] *Electricity Metering Equipment (a.c.) General Requirements, Tests and Test Conditions. Metering Equipment (Class Indexes A, B and C)*, Standard BS EN 50470-1:2006+a1, 2018.

[46] *IEEE Standard Definitions for the Measurement of Electric Power Quantities Under Sinusoidal, Nonsinusoidal, Balanced, or Unbalanced Conditions*, IEEE Standard 1459-2010, 2010.

[47] G. Artale et al., "Measurement of simplified single- and three-phase parameters for harmonic emission assessment based on IEEE 1459–2010," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.

[48] *Electromagnetic Compatibility (EMC)—Part 4–30: Testing and Measurement Techniques—Power Quality Measurement Methods*, Standard IEC 61000-4-30, 2015. [Online]. Available: https://webstore.iec.ch/publication/68642

[49] *Electromagnetic Compatibility (EMC)—Part 4–7: Testing and Measurement Techniques—General Guide on Harmonics and Interharmonics Measurements and Instrumentation, for Power Supply Systems and Equipment Connected Thereto*, Standard IEC 61000-4-7, 2002. [Online]. Available: https://webstore.iec.ch/publication/4228

[50] J. Sola and J. Sevilla, "Importance of input data normalization for the application of neural networks to complex industrial problems," *IEEE Trans. Nucl. Sci.*, vol. 44, no. 3, pp. 1464–1468, Jun. 1997.

[51] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 10, pp. 2825–2830, Jul. 2017.

[52] L. Tari, G. Berrettoni, C. Bourelly, G. Cerro, D. Capriglione, and L. Ferrigno, "eLAMI—An innovative simulated dataset of electrical loads for advanced smart energy applications," *IEEE Access*, vol. 10, pp. 91177–91191, 2022, doi: 10.1109/ACCESS.2022.3201960.

[53] P. A. Schirmer and I. Mporas, "Non-intrusive load monitoring: A review," *IEEE Trans. Smart Grid*, vol. 14, no. 1, pp. 769–784, Jan. 2023.

[54] P. da Silva Hack and C. Schwengber ten Caten, "Measurement uncertainty: Literature review and research trends," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2116–2124, Aug. 2012.

[55] S. Sediva and M. Havlikova, "Comparison of GUM and Monte Carlo method for evaluation measurement uncertainty of indirect measurements," in *Proc. 14th Int. Carpathian Control Conf. (ICCC)*, May 2013, pp. 325–329.

[56] G. M. Mahmoud and R. S. Hegazy, "Comparison of GUM and Monte Carlo methods for the uncertainty estimation in hardness measurements," *Int. J. Metrol. Quality Eng.*, vol. 8, p. 14, Jan. 2017, doi: 10.1051/ijmqe/2017014.

[57] Yokogawwa. (2022). *WT3000 Yokogawa*. [Online]. Available: https://tmi.yokogawa.com/solutions/discontinued/wt3000-precision-power-analyzer/

Luca Tari (Graduate Student Member, IEEE) was born in Cassino, Italy, in 1997. He received the M.S. degree (cum laude) in electrical engineering from the University of Cassino and Southern Lazio, Cassino, in October 2021, where he is currently pursuing the Ph.D. degree with the Department of Electrical Engineering.

His current research interests concern electrical and electronic measurements in the field of smart energy and energy digitization. In particular, the design of innovative smart meters and the development of methods and techniques for monitoring, data analysis, optimization and energy efficiency, fault detection, and predictive diagnostics.

Mario Molinara (Senior Member, IEEE) received the Laurea degree (Hons.) in computer science from the University of Sannio, Benevento, Italy, in 1999, and the Ph.D. degree in computer science and telecommunication from the University of Salerno, Salerno, Italy, in 2003.

In 2004, he joined the Dipartimento di Automazione, Elettromagnetismo, Ingengeria dell'Informazione e Matematica Industriale, University of Cassino, Cassino, Italy, where he is currently a Researcher in computer science. His research interests include image analysis and interpretation, classification techniques, statistical learning, biomedical imaging, neural networks, optical character recognition, and map and document processing.

Dr. Molinara is a member of the International Association for Pattern Recognition (IAPR).

Luigi Ferrigno (Senior Member, IEEE) has been a Full Professor of electric and electronic measurement and the Scientific Manager of the Industrial Measurements Laboratory, University of Cassino, Cassino, Italy, since 2004. In 2008, he was a Founding Member of the University spin-off, Spring Off (University of Salerno), Fisciano, Italy. He is an NDE4.0 Ambassador for the Italian Association of Non-Destructive Evaluation and Test (AiPnD) in the EFNDT WG10. He has coordinated and participated in several national and international research projects. His current research interests include the NDT4.0, novel learning sensors and measurement systems for smart city, the Internet of Things (IoT), automotive, smart energy, and environment.

Open Access funding provided by 'Università degli Studi di Cassino e del Lazio Meridionale' within the CRUI CARE Agreement