





# Residential Harmonic Injection Models Based on Field Measurements

Pablo Rodríguez-Pajarón , *Student Member, IEEE*, Araceli Hernández , *Member, IEEE*, Hugo Mendonça , *Member, IEEE*, and Jovica V. Milanović , *Fellow, IEEE*

**Abstract**—In recent years, harmonic current levels have increased in residential distribution networks due to the growing penetration of nonlinear loads at customer premises. To properly model the harmonic current injection of residences, probabilistic models that account for the stochastic behavior of domestic loads are needed. As a part of the study reported in this paper, a field measurement campaign was performed in 24 dwellings during a one-week period with a one-minute resolution. Harmonic current magnitudes and phase angles of odd harmonic orders up to the 25th as well as the active and reactive power demanded were recorded. By applying two different unsupervised learning techniques (i.e., nonparametric density estimation and Gaussian mixture models) to these measurements, two probabilistic models of harmonic injection of individual residential sites have been derived. Statistical probability distributions have been determined for the magnitude and phase of each harmonic order segmented in different power demand intervals. The developed models are validated on a test feeder by comparing harmonic voltages caused by the injection of the synthetically generated currents with those obtained from measurements. Both field measurement results and the detailed data defining the probabilistic harmonic current models are made public in an Open Science database repository.

**Index Terms**—Harmonic currents, power quality, residential network, stochastic models.

## I. INTRODUCTION

**H**ARMONIC distortion in distribution networks is becoming an increasing concern for industry. Usually, driven by energy saving policies, a growing number of nonlinear (NL) loads are connected to the network [1]. These NL loads (such as switching-mode power supplies or variable speed drives) inject harmonic currents. At present, their proliferation in residential premises can cause harmonic distortion in the supply voltage and affect other loads connected to the same network [2]–[7]. For

Manuscript received 28 January 2022; revised 29 April 2022 and 25 June 2022; accepted 31 July 2022. Date of publication 2 August 2022; date of current version 24 January 2023. This work was supported in part by MCIN/AEI/10.13039/501100011033 under Grant RTI2018-097424-B-I00 and in part by ERDF A way of making Europe. Paper no. TPWRD-00124-2022. (Corresponding author: Pablo Rodríguez-Pajarón.)

Pablo Rodríguez-Pajarón, Araceli Hernández, and Hugo Mendonça are with the Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, 28006 Madrid, Spain (e-mail: pablo.rpajaron@upm.es; araceli.hernandez@upm.es; hugo.rocha@upm.es).

Jovica V. Milanović is with the Department of Electrical and Electronic Engineering, University of Manchester, M13 9PL Manchester, U.K. (e-mail: milanovic@manchester.ac.uk).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TPWRD.2022.3195962>.

Digital Object Identifier 10.1109/TPWRD.2022.3195962

example, extensive field measurement data collected from various locations in North America show that the harmonic current distortion levels of residential feeders are much higher than those of the industrial and commercial feeders [2]. In the near future, this potential impact is expected to be even more noticeable, due to the increasing penetration of new loads such as electric vehicles (EVs) and electronic-based home appliances. Furthermore, different international standards, such as EN 50160 [8] or IEEE Std. 519 [9], are enforcing compliance with voltage waveform requirements in public low voltage distribution networks.

To analyze the scenarios that may jeopardize voltage distortion levels and compromise compliance with the standards limits in low voltage power distribution systems, it is critical to realistically model the harmonic injection of residential loads. The aim of this research is to develop realistic (based on field measurements) harmonic injection models of residential loads and to incorporate their characteristic stochastic behavior (by using a probabilistic approach). To expand the usability of developed models and to facilitate further research, the developed probabilistic models and the measurements that they are based on, have been made publicly available in an open science repository [10].

The models for harmonic current injection from individual residences, developed in this study, can be used for harmonic propagation studies, assessment of harmonic voltages and verification of compliance with standard limits in any network. The probabilistic nature of the proposed models allows consideration of the uncertainty in harmonic injections and, therefore, quantification of the variation in the expected harmonic voltages by means of probabilistic techniques such as the Monte Carlo [4], [11], [12] method or the most advanced surrogate techniques. The characteristics of the network (and, in particular, its short circuit power level) will influence the voltage distortion reached at network buses and, therefore, the compliance with standard limits such as EN50160 [8] or IEEE Std. 519 [9].

Different approaches have been followed in the past to develop residential harmonic current models. Modeling of residential harmonic injections can be performed by developing appropriate equipment models or aggregated models [13], [14]. The former is based on modeling each of the electric devices that are typically found in at the residential customer site individually. The models for linear loads are usually modeled as a constant power load or a constant admittance load, while models for NL loads, which are the most important regarding harmonic analysis, are more complex. Two main alternatives have been widely used to

represent NL loads of household appliances: time-domain and frequency domain models.

Time-domain models consist of a detailed representation of electronic circuits integrated in the appliance. These models require a deep analysis and knowledge of the device, which is sometimes difficult to obtain and very often not applicable to a similar device with different design characteristics. These models are also highly computationally demanding. Some examples of NL time-domain models can be found in [15]–[17].

Frequency domain models are the most commonly used in electrical simulations. NL loads can be simulated as ideal current sources, either neglecting the influence that distorted voltages may have on injected currents [18] or considering it [19]. NL loads can also be modeled as coupled or uncoupled Norton equivalent models ([20], [21] or [22]–[24] respectively).

The abovementioned equipment models are often too detailed to study harmonic currents or voltages in residential distribution systems and lead to an excessive complexity and detail in the load modeling. In contrast, aggregate models represent the injected currents of a group of components, with no need to model in detail each of them. Depending on the study performed, a different level of aggregation may be needed: from individual household aggregation (models that represent a single residence including all its household devices) to models of several aggregated dwellings.

Harmonic assessment of low voltage residential networks usually requires the level of detail given by individual dwelling aggregated models, whereas studies on medium voltage systems can be sufficiently accurate with models that aggregate harmonic contribution by several dwellings. Harmonic models for several aggregated dwellings have been reported in the past [25]–[27]; however, models of harmonic injection of individual residences are much less frequent. Different techniques can be applied to build these aggregated harmonic injection models. They can be generally classified as bottom-up models and measurement-based models. Bottom-up models consider the residence by modeling equipment or types of devices, which are then combined correctly to form an aggregate model of the whole residence [4], [11], [12], [28], [29]. Measurement-based models are built from real data extracted from the network. They usually analyze measurements with some statistical method to take into account the random behavior of the harmonic injections [26], [27], [30]–[32]. While the accuracy of bottom-up models relies on the correct modelling of multiple uncertain variables, the advantage of the measurement-based harmonic injection models such as the one proposed here is the fidelity of the model to realistic data.

Most of the above models are either valid for an aggregation of several residences [25]–[27], [32] or do not take into account the high variability of residential harmonic currents [4], [12], [30], [31] and consider the harmonic injection of a type of device invariable. Other models statistically analyze some residence measurements to determine their contribution to the total harmonic distortion but do not provide complete probabilistic harmonic characterization of the residential loads [2], [33]. To fill this gap, this study proposes two different models for harmonic injections of individual residences. These models are

probabilistic to account for the stochastic nature of the loads and are based on measurements performed in 24 houses during a one-week period.

The contributions of this study are both methodological and research enabling: i) first, two methodologies, one based on nonparametric density estimation and the other on Gaussian Mixture Models, are developed to obtain probabilistic harmonic models of individual residences based on field measurements performed on 24 dwellings; and ii) second, the measured data as well as the developed statistical injection models are made available to researchers in an open database that can be accessed in [10].

## II. FIELD MEASUREMENTS

### A. Description of the Measurement Campaign

There are several databases that offer data from individual residential sites. However, some of them do not include harmonic injections at all [34], [35] (only related to power consumption), whereas others only offer current total harmonic distortion ( $THD_I$ ) [36], [37], which is not sufficient to build complete models of harmonic emission. Some other databases offer current waveforms but usually for a small number of dwellings, such as in [38] or [39] (including data for 1 and 3 dwellings respectively). None of them offer harmonic current injections for a sufficient number of sites or with sufficient detail to build representative models. This lack of real and available data of residential harmonic injections justifies the measurement campaign that took place during 2020 and 2021 in Madrid, Spain.

The measurements performed include harmonic current magnitudes and phase angles for odd harmonic orders up to the 25th as well as active and reactive power demanded by the individual dwellings [40]. The measurements were performed during a one-week period with a one-minute sampling rate recording one-minute average values of measured quantities. Measurements were performed with a *Fluke 435 Power Quality Analyzer* or a *Fluke 435-II Power Quality Analyzer* and a *Fluke i30s* current clamp. Measurements complied with IEC-61000-4-7 [41]. According to that, harmonic current phase angles refer to the fundamental voltage at the measuring site. All measurements were taken at single-phase supply points.

During the campaign, the measurements were taken at 24 different residential sites. All the measurement data are made public for researchers to use for their own purposes and research. Measurements are available in the database called MEMORIA (MEasureMents Of Residential Injections of hArMonics) which can be freely accessed in [10]. In this database, the measured values can be downloaded together with a data sheet indicating the characteristics (household appliances, squared meters, etc.) of the measured dwelling.

The measured sites include apartments (71% of samples) and single houses or semidetached (29%). Table I includes the percentage of dwellings with different numbers of TVs, computers and refrigerators, as three of the most typical residential devices. Approximately 70% of the refrigerators were purchased before 2005, whereas only 30% were purchased after 2015. When

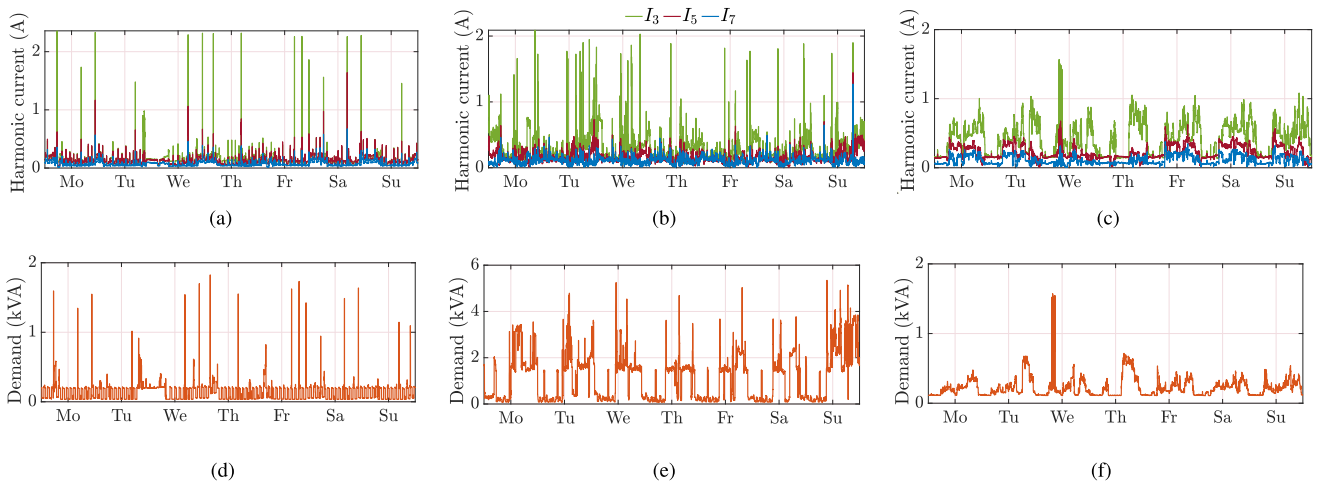


Fig. 1. Weekly profile of the 3rd, 5th and 7th harmonic current magnitudes of three measured sites (left) and their apparent power demand (right). (a) Site 4: harmonics. (b) Site 12: harmonics. (c) Site 14: harmonics. (d) Site 4: power. (e) Site 12: power. (f) Site 14: power.

TABLE I  
SUMMARY OF THE PROPORTION OF DWELLINGS WITH DIFFERENT NUMBERS OF DEVICES

	1 appliance	2 appliances	3 or more appliances
TVs	33%	17%	50%
Computers	8%	33%	55%
Refrigerators	83%	12%	5%

looking at washing machines, a different situation is found, as 89% of dwellings have one and more than half of them were purchased after 2015. The washing machine purchase year is an important variable to consider, as modern devices may include a variable speed drive that has an impact on harmonic injection [42], [43]. Similar consideration should be given to the presence of an induction hob, which was present in 58% of dwellings. The rest of them used a noninduction electric hob. None of the measured sites included PV generation or EV chargers.

### B. Overview of Measured Harmonic Currents

The measured harmonic currents at residential sites show a stochastic behavior over time. This stochastic behavior is a common characteristic of all measured sites, but the shape of the harmonic profiles is very different from one site to another. These two attributes are illustrated in Fig. 1, where harmonic magnitudes for the 3rd, 5th, and 7th currents are plotted on the left for three different dwellings in the dataset (sites 4, 12 and 14) during a week as an illustration of the mentioned variable harmonic profiles. Fig. 1 also shows the power demand (in kVA) for the same selected sites (right part of the figure).

From Fig. 1, two main conclusions can be drawn. First, by comparing the left and right sides of the figure, some similarities can be found between harmonic emission and power. This fact is further discussed in the next section and is one of the starting points in harmonic model development.

Second, harmonic weekly profiles of individual dwellings do not show a clear cyclic behavior, which means that modeling

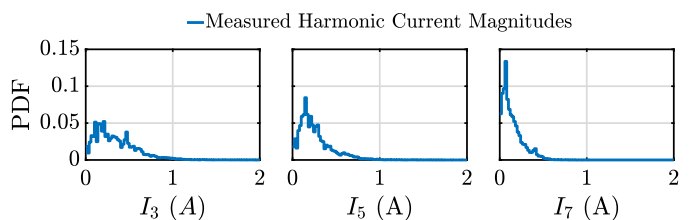


Fig. 2. Probability distribution functions for harmonic current magnitudes of all measured sites.

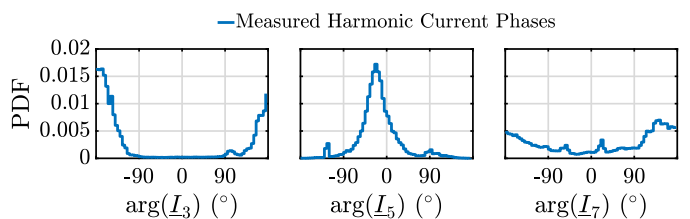


Fig. 3. Probability distribution functions for harmonic current phase angles of all measured sites.

their emissions by means of time series is not a suitable option, as it could be for aggregated residential profiles with smoother shapes [14]. For this reason, the time instant is not considered in the developed harmonic models as the key influential parameter, and the proposed stochastic modeling of harmonics during the week omits this variable. By doing so, the probability distribution functions of harmonic current magnitudes measured during the one-week period can be calculated. These functions are plotted in Fig. 2 for harmonic currents injected by the residences included in the dataset (due to space limitations, only 3rd, 5th and 7th harmonic orders are plotted).

Fig. 3 shows the probability distribution function (PDF) of the phase angles of the main harmonics (3rd, 5th and 7th). The 3rd and 5th harmonic currents exhibit a clearly prevailing phase. In the case of the 7th harmonic, the phase angle distribution is much wider, and different values can be found. Increasing

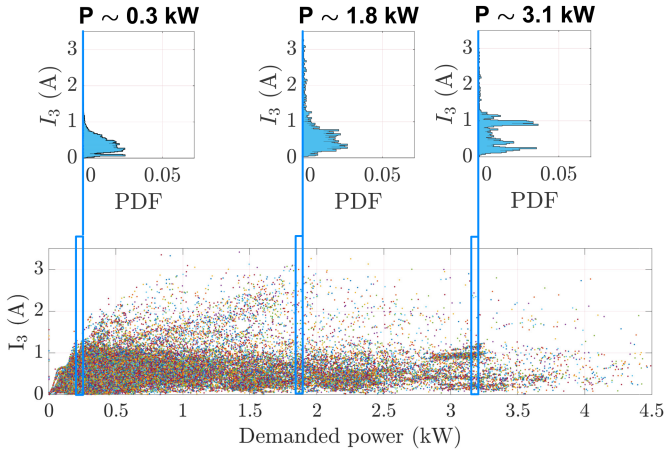


Fig. 4. Example of distributions of harmonic magnitudes for the 3rd order in different segments of the demanded fundamental power.

the harmonic order beyond the 7th makes the phase angle distributions even more disperse.

### III. SEGMENTATION OF HARMONIC MEASUREMENT DATA

#### A. Harmonic Current Magnitudes

A simple model could be developed where each harmonic order is modeled by the corresponding probability distribution function extracted from Figs. 2 and 3, but it would omit important information about the time dependency or correlation between different harmonic orders. On the other hand, a complete model in which time dependency and interdependency among harmonic orders are considered (e.g., by means of multivariate time series) would be very complex because harmonics of single dwellings are very stochastic and do not show clear cyclic behavior. An intermediate solution is adopted in this work where time dependency is not directly included in the model but its effect is incorporated by considering the dependency of different harmonic orders on the demanded fundamental power (i.e., active power demanded at the fundamental frequency).

Indeed, different values of demanded fundamental power lead to different probability distributions on harmonic injections [27], as illustrated in Fig. 4 for the 3rd harmonic. This figure presents all values measured for the 3rd harmonic magnitude at all the time intervals in all measured sites (in y-axis) and the demanded fundamental power in the residence at the same time interval (in the x-axis). Each measured sample is represented as a colored dot. In the upper part of the figure, the probability distribution of the harmonic magnitudes is represented, but only for values in the specified range of demanded fundamental power (approximately 0.3 kW, 1.8 kW and 3.1 kW). The distributions of harmonic currents are different for different demanded powers. The same procedure shown here as an example for 3rd harmonic order applies to other harmonic orders.

With this, the probability distribution proposed in Fig. 2 can be segmented into different distributions according to a specified range of demanded power for each of them. In this study, a total of 15 demanded power intervals are used to segment harmonic magnitudes as a trade-off between the accuracy and size of the

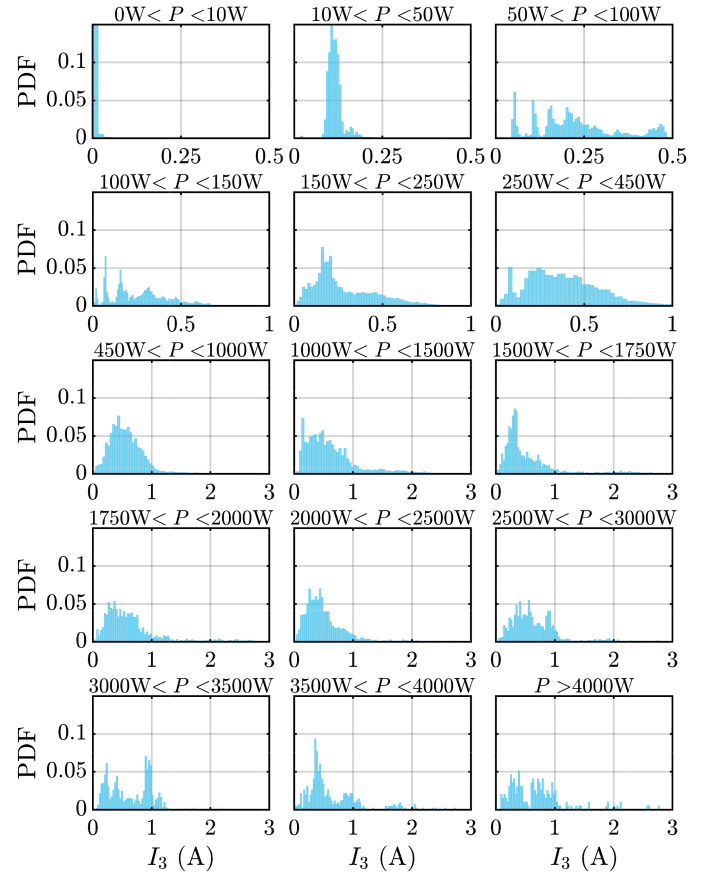


Fig. 5. Segmented probability distribution functions for 3rd harmonic magnitudes.

model. Fig. 5 illustrates this idea for the probability distribution function of the harmonic magnitudes of the 3rd harmonic order. The intervals of power in which distributions are divided are indicated in the upper part of the figures. Note that the probability distribution functions in Fig. 5 are only 3rd harmonic magnitudes, whereas all other harmonic orders up to 15th have been segmented using the same power intervals with a similar procedure.

By adopting harmonic magnitude segmentation according to the demanded fundamental power, time dependency is implicitly considered, as the demanded power can be modeled in time, according to probabilistic models representing the residence occupants' habits (e.g., by the bottom-up approach as reported in [44]).

Segmenting harmonic modeling into different demanded power intervals implicitly accounts for the interdependency between harmonic orders, since probability models of all harmonic orders are developed and referred to the same power segments. For example, if the harmonic injection of a dwelling is to be simulated in a time interval when the demanded power is 800 W, a sample will be taken from the specific probability distribution of the 3rd harmonic current for 800 W demanded power, as well as a sample from the probability distribution of the 5th harmonic current and the probability distribution of the rest of the harmonic orders extracted for the same interval of 800 W.

Probability distributions of individual harmonics for 800 W are different from the probability distributions of the corresponding harmonics for different demanded powers. This approach based on the dataset segmentation offers a good compromise between accuracy and simplicity compared to alternative approaches based on multivariate probabilistic techniques, such as Gaussian Copulas [45], [46] or multivariate Gaussian Mixtures. Its validity will be further analyzed and discussed in Section VI and specifically shown in Fig. 21.

Note that the methodology proposed provides the harmonic currents for each residence as an aggregation of all the devices included in the residence. In the same way, the power indicated refers to the power consumed by the whole residence and not by each individual appliance. Therefore, the previous example refers to the harmonic content of the current absorbed by the house when the total power demanded (by the residence as a whole) was 800 W. This power does not necessarily correspond to the demand of a single device (such as a power electronic converter) but to a combination of switched on appliances in the house with total demanded power of 800 W.

Finally, demanded power from single dwellings can be easily simulated from the multiple models available in the literature (such as [44], [47], [48]) or directly taken from real residences with smart-meter data, which are now widely installed in European distribution systems [49].

### B. Harmonic Current Phase Angles

Harmonic phase angles also show great variability, as seen in Fig. 3. To understand which variable better explains the variability of phase angles, a principal component analysis (PCA) [50] was conducted. PCA is a statistical technique that helps to decide which variables better explain the source of variability in multivariate samples. Geometrically speaking, principal components represent the directions of the data that lead to the largest variance. The larger the contribution to the variance of a variable, the more important is the source of variability influencing that variable. In the analysis conducted to understand where the variability of harmonic angles comes from, the most important variables identified were harmonic magnitudes (with contribution to the variance of 32%), whereas fundamental power demand appears to be significantly less decisive, with contribution to the variance of 19%.

Therefore, according to the performed PCA, it can be concluded that the variability of phase angles is more affected by, or a consequence of, the harmonic magnitudes than the demanded power.

As observed in measurements [10], [38], [39], in general, there is a prevailing harmonic phase angle for high magnitudes and a more dispersed harmonic phase when the harmonic magnitude is low. In Fig. 6, the distribution of phase angles for different segments considered in this work is presented for the 3rd harmonic order. A similar procedure is followed for all the other harmonic orders. For all harmonic phase angles, 6 intervals are used to segment the data. Intervals are selected according to the maximum registered value of the harmonic current of the same order,  $I_h^{(max)}$ . Selected intervals are shown in Table II, and

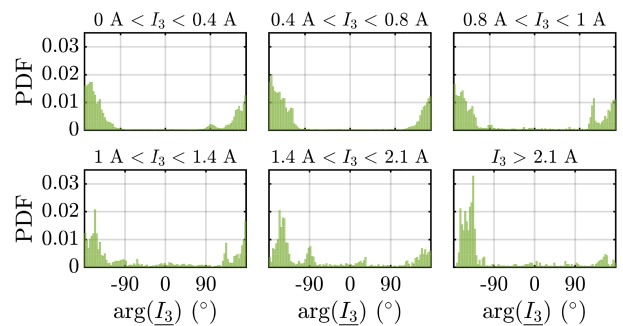


Fig. 6. Segmented probability distribution functions for 3rd harmonic phase angles.

TABLE II  
INTERVALS OF MAGNITUDE FOR SEGMENTATION OF ARGUMENT DATA

Interval	1	2	3	4	5	6
Initial value ( $\times I_h^{(max)}$ )	0	0.1	0.2	0.25	0.33	0.5
Final value ( $\times I_h^{(max)}$ )	0.1	0.2	0.25	0.33	0.5	1

applied for 3rd harmonic order in Fig. 6. The number of intervals is selected to accurately fit the measurements while avoiding the existence of intervals with similar distributions.

### C. Contrast Analysis of Segmented Data Distribution

Previous works for statistical harmonic modeling have modeled harmonic injections of several aggregated residences with different parametric statistical distributions [13], [25], [27], [51]). Parametric distributions assume that sample data come from a population that can be adequately modeled by a probability distribution that has a fixed set of parameters. Most well-known statistical distribution functions are parametric: e.g., constant, normal or lognormal distributions. However, in the case of aggregation of several residences, harmonic injection distributions are smoother than in the case of single residences. In addition, the aforementioned models [13], [25], [27], [51] do not take into account interdependency among different harmonic orders.

In this study, a contrast analysis was performed to find the probability distribution that properly fits the data. In this section, both harmonic magnitudes and phase angles for odd harmonic orders up to the 15th are contrasted with Kolmogorov-Smirnov and Chi-squared tests [52] to test if the distribution follows a known statistical distribution. Normal, lognormal, Weibull, exponential and gamma distributions with parameters automatically obtained with Matlab Statistics and Machine Learning Toolbox [53] for the dataset of each harmonic order and power segment (or magnitude segment) were tested. In total, 735 parametric probability distributions (525 for harmonic magnitudes and 210 for harmonic phase angles) were contrasted with the aforementioned statistical tests. All statistical analysis were performed with a 95% confidence value and are run in MATLAB, through MATLAB Statistics and Machine Learning Toolbox.

The results of these tests have shown that most of the probability distributions for different harmonic orders in both magnitudes and phase angles cannot be characterized by the aforementioned

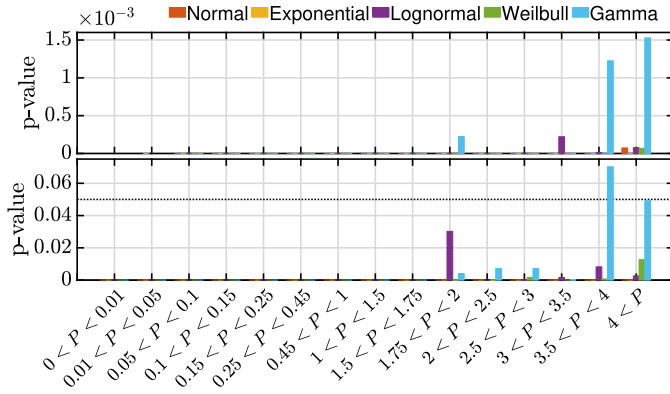


Fig. 7. Contrast analysis p-value for distributions of 3rd harmonic magnitudes with chi-squared (top) and Kolmogorov-Smirnov (bottom) tests. The null hypothesis considers that the data come from the tested distribution.

parametric distributions with the specified confidence level and the tests performed. In particular, only in twelve of the distributions tested (out of a total of 735 distributions) one of the tests concluded that data could be fitted by the parametric distribution, but p-value is close to reject the hypothesis. As an example, this is illustrated in Fig. 7, where p-values of the 15 tests conducted for different distributions of 3rd harmonic magnitudes corresponding to different power segments are shown.

A similar contrast analysis has been performed to distributions of current phase angles, resulting in the same conclusion: with the confidence level and the tests performed, data cannot be characterized by the tested parametric distributions.

Thus, considering that data cannot be fitted with known parametric distributions with sufficient confidence, this paper proposes modeling harmonic injections with nonparametric distributions or Gaussian mixture models, which will be presented and discussed in the next sections.

#### IV. NONPARAMETRIC HARMONIC INJECTION MODELS

To overcome the inaccurate fitting that classical parametric distributions have on the segmented probability distribution of harmonic current injections, this paper proposes modeling them through nonparametric kernel distributions (kNP). Such probabilistic techniques have been previously used for modeling other power quality disturbances [54]–[56] such as voltage sags and voltage fluctuations.

Kernel density estimation is a nonparametric method to smooth histograms by finding multiple local parametric distributions around sampling data. The number, type, and extension of parametric distributions that conform to the kNP distribution (namely, number of points, kernel smoother and bandwidth) are adjusted during the fitting of the kNP distribution [50]. In this work, as is commonly done, a normal distribution is used as smoothing function, so nonparametric density estimation is similar to a Gaussian mixture model, where the number of Gaussians is not predefined before building the model.

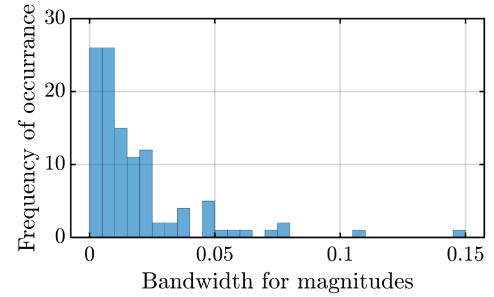


Fig. 8. Frequency distribution of bandwidth parameter “b” used in the kNP models of harmonic magnitude distributions.

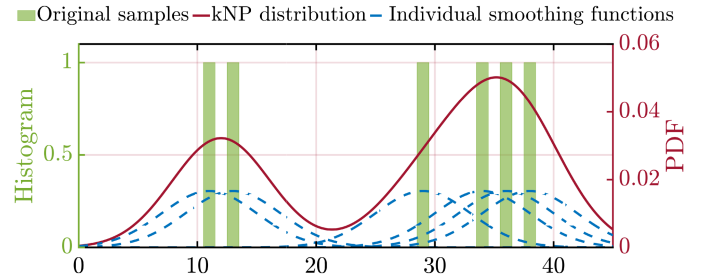


Fig. 9. Example of kNP distribution (red) built from sampling data (green) and composed of multiple normal distributions (dashed blue).

The following expression defines the kNP distribution and analytically describes the explanation given before:

$$\hat{f}_b(x) = \frac{1}{n \cdot b} \cdot \sum_{i=1}^n K\left(\frac{x - x_i}{b}\right) \quad (1)$$

where  $\hat{f}_b(x)$  is the estimated kernel distribution,  $x_i$  with  $i = 1 \dots n$  are random samples of the available data,  $n$  the sample size,  $K(\dots)$  is the kernel smoothing function (in this study, the Gaussian function has been used) and  $b$  is the bandwidth. Estimation of  $\hat{f}_b(x)$  is performed with MATLAB Statistics and Machine Learning Toolbox [53], which finds the optimal  $b$  (using Silverman’s rule [57]) for estimation. As an example, Fig. 8 shows the frequency distribution of value  $b$  for the kNP models built for harmonic magnitudes. These models have been defined for 7 odd harmonic orders segmented in 15 power intervals, i.e. Fig. 7 shows the value of  $b$  in 105 modeled harmonic magnitude distributions. The parameter  $b$  for all the proposed distributions can be found in the open access database [10], where the distributions are provided in Matlab format.

Fig. 9 shows a simple example where six random samples (green histogram) are used to build a kNP distribution (in red). The kernel estimation technique smooths the histogram, and in this example, it is built from six different normal distributions (orange dashed lines), each of which locally models each data sample. The obtained nonparametric distribution describes the finite data in the form of a continuous distribution that can be used to extract new randomly generated synthetic samples.

When this procedure is applied to the segmented measured distribution data presented in Figs. 5 and 6, the kNP distributions

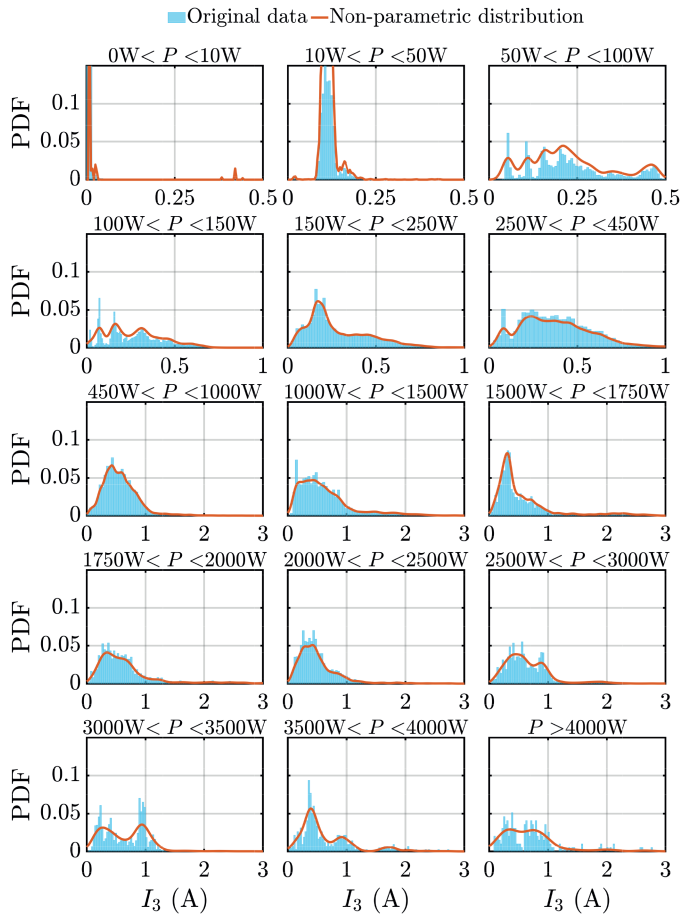


Fig. 10. kNP distribution for segmented PDF of 3rd harmonic magnitudes.

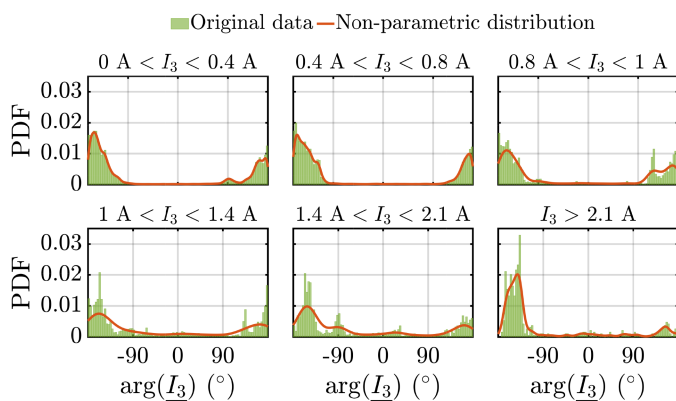


Fig. 11. kNP distribution for segmented PDF of 3rd harmonic phase angles.

shown with a red line in Figs. 10 and 11 are obtained for magnitudes and phase angles, respectively, of 3rd harmonic currents. The similarity of these figures to the original measured data represented in blue or green can be clearly observed.

The same procedure of building the kNP distribution for harmonic currents and phase angles was performed for all the defined data segments and all the harmonic orders included in the measurement dataset.

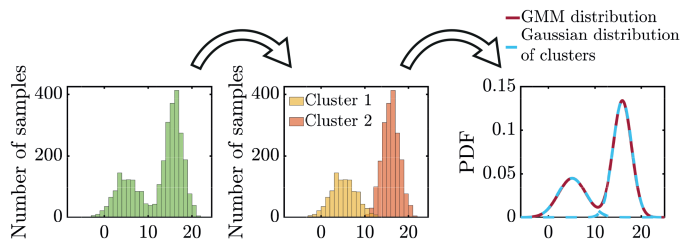


Fig. 12. Example of GMM.

Kernel nonparametric distribution modeling is an automatic way to represent discrete data in the form of probability distributions. Its accuracy will be discussed in Section VI.

## V. GAUSSIAN MIXTURE HARMONIC INJECTION MODELS

The previously presented model based on kNP distribution accurately fits measured data to nonparametric distributions. In addition, this methodology can be automatically applied to any other measurement datasets without supervision. Despite these advantages, it has two main shortcomings. First, it requires high computational effort for both building the model out of data and, more importantly, for obtaining random samples from the built distribution, especially when the number of samples conforming to the model is large. Second, in a nonparametric model, the definition of the probability distribution functions is given by the measured samples and the bandwidth instead of a reduced set of parameters (such as the mean or standard deviation) that would define a classical distribution. This feature makes sharing or applying a pre-defined model difficult, as no parameter can characterize the modeled distribution and the measurement dataset is needed each time the model is used.

To overcome these two issues, in this work, a second unsupervised technique is proposed to model individual residential harmonic currents: Gaussian mixture models. This statistical approach has been used in the past for other power quality indices [58] and analyzed for characterizing harmonics [59], [60].

A Gaussian mixture model (GMM) is a probabilistic method used for data clustering and data mining that assumes that the dataset is generated from a mix of a finite number of Gaussian distributions with unknown parameters [61] that can be used to model the dataset distribution when it does not fit any known parametric distribution. The GMM model also returns the probability that any point belongs to any given cluster. There are different methods to establish the most appropriate number of Gaussian distributions (or components) of the GMM for each dataset. In this model, the silhouette criterion is applied [61]. In this work, Gaussian mixture models (GMMs) are used to cluster the distributions of measured harmonic currents (magnitudes and phase angles) for each harmonic order and each power segment presented in Section III. This means that for each of the aforementioned probability distributions, different groups or clusters are identified, and parameters that define a Gaussian distribution that fits each of the clusters are obtained.

Fig. 12 illustrates the idea of obtaining a Gaussian mixture model. The left part of the figure shows a distribution of data,

described with its histogram. If these data points were to be modeled with a Gaussian mixture model, an identification of groups or clusters would be performed, as shown in the middle plot, where 2 clusters have been identified. With it, each of the clusters is modeled with a univariate Gaussian distribution, as shown by the blue dashed lines in the right part of Fig. 12, which together form the Gaussian mixture model distribution shown by red line.

The ideas presented in Fig. 12 are based on the following procedure.

- Calculation of the number  $N$  of Gaussian components (i.e., the number of clusters in which the data are divided). The number of clusters that define the Gaussian mixture model is a key parameter in this method. In this paper, to find a representative number of clusters the silhouette criterion is used. This criterion compares how similar a sample is to other samples in the same cluster to maximize this similarity by selecting appropriate clusters [62]. Once the number of clusters is decided, the centroid of each of them is selected, usually with the EM (expectation-maximization) algorithm [50], [61]. Expectation maximization is an algorithm that finds maximum likelihood estimators of a distribution (in this case, centroids for clusters) in datasets where not all variables of the model are known.
- Calculation of the weight  $w_i$  of a specific cluster  $i$ . This parameter represents the probability of a sample being in a specific cluster  $i$ . Once the number of components and the center of each component are found, the weight of a sample represents the distance to its center. For instance, if the sample is in the center of cluster  $i$ , the weight will be 1 for cluster  $i$  and 0 for all the other clusters. If it is half-way between two clusters, its weight is divided accordingly [50].
- Calculation of the mean  $\mu_i$  and standard deviation  $\sigma_i$  of cluster  $i$ . These parameters define the Gaussian density function that fits data in cluster  $i$ .  $\mu_i$  and  $\sigma_i$  are computed considering the weights of each sample in cluster  $i$ .

All the aforementioned characteristics have been obtained for the segmented measurement dataset of harmonic magnitudes and phase angles with MATLAB Statistics and Machine Learning Toolbox [53].

Once the Gaussian distributions are characterized, to obtain a random value of the Gaussian mixture model, a similar procedure must be followed: first, one of the clusters is randomly selected by considering its weight, and then a random sample is obtained from the specific normal distribution.

An example of the Gaussian mixture distributions obtained for the 3rd harmonic magnitude and phase are shown in Figs. 13 and 14, respectively. These figures show the same shape as distributions in Figs. 5 and 6, and so are very similar to Figs. 10 and 11, although the method to obtain them is very different. Further comparison of these two models is provided in Section VI. Additionally, figures show how the shape of Gaussian mixture distributions is not as close to real data as it was for KNP.

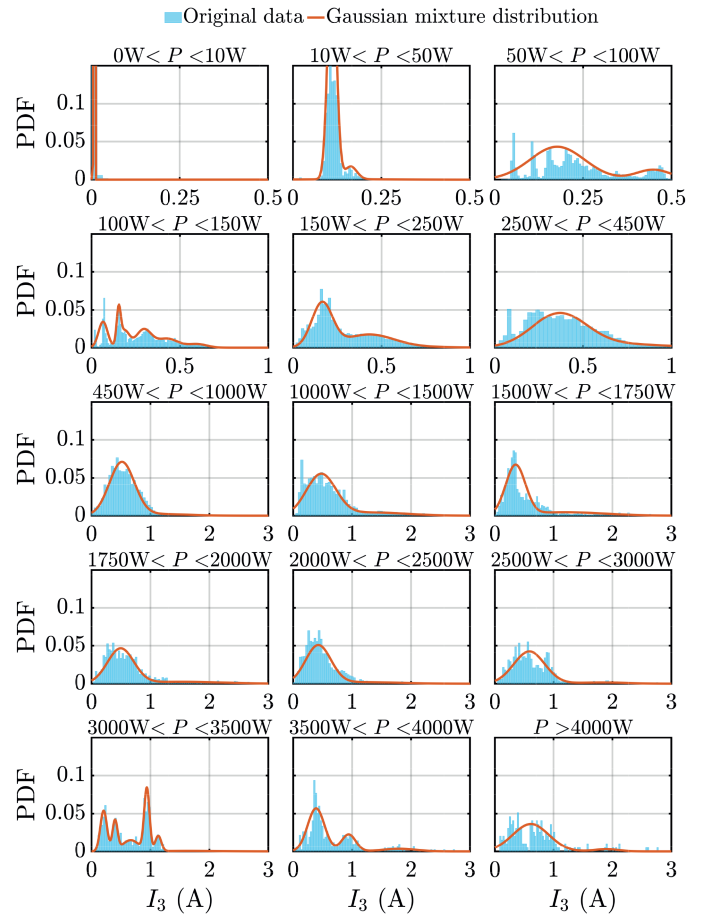


Fig. 13. GMM distributions for segmented PDF of 3rd harmonic magnitudes.

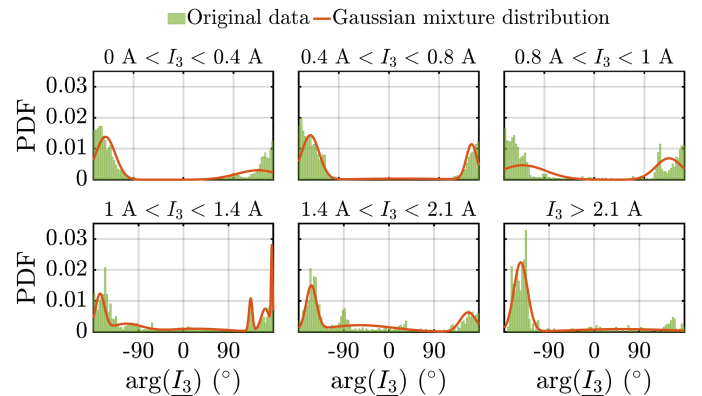


Fig. 14. GMM distributions for the segmented PDF of 3rd harmonic phase angles.

As stated when explaining the nonparametric model, note that for each interval and harmonic order, independent and univariate GMM distributions are built. Each of the distributions that conform to the residential harmonic model is built from all data measured.

For each harmonic order, 15 distributions for magnitudes and 6 for phase angles are obtained in accordance with the segmentation used. Each of the 21 distributions is composed of several Gaussian distributions, so in this model harmonic



currents are represented by a large number of simple classical distributions.

Gaussian mixture model accuracy is discussed in the next section. Obtaining Gaussian mixture models is computationally demanding, as with kNP models. Nevertheless, obtaining a random sample of preexisting Gaussian mixture model distributions is computationally easy and less demanding when compared to the kernel. Additionally, GMMs can be shared and used without the need to have the whole dataset. For instance, for sharing the model for 3rd harmonic magnitude and first demanded power interval built with nonparametric distributions, the whole dataset of this distribution is needed (i.e., 40,768 bytes when treated as a MATLAB array of doubles, with size  $5,096 \times 1$ ), whereas the Gaussian mixture model associated with this distribution is only formed by two weights, two means and two standard deviations (48 bytes when treated as a MATLAB array of doubles, with dimensions of  $3 \times 2$ ). The two developed harmonic injection distribution models have been made available for the research community and can be accessed in [10].

## VI. VALIDATION OF MODELS

### A. Validation of Harmonic Currents

The models proposed in this paper characterize the probabilistic distribution of measured harmonic currents and allow generating synthetically harmonic current values with the same probability as real values.

To validate the proposed methodology, harmonics were measured at 5 additional residences during a period of one week following the same measurement procedure described in Section II. Harmonic currents measured at these residences were not used in model development. The validation process is based on:

- 1) Obtaining the distribution of individual harmonics measured at the 5 additional residences.
- 2) Sampling the probability distributions obtained in the model. Since the modelled harmonic distributions are segmented for different power intervals, the distribution sampled at a particular time instant is allocated to the power interval corresponding to the demanded power measured at the same instant.
- 3) The distribution of all measured values and the distributions obtained from the sampling in the models are compared.

In Fig. 15, the values of the harmonic currents obtained from the 5 measured residences and those obtained from the modeled distributions are illustrated as a PDF for both magnitude and phase angle. In this figure, the magnitudes for the 3rd, 5th and 7th harmonic orders of all the measured (and synthetically generated) sites overlap, as they are very similar. Regarding phase angles, Gaussian mixture model distributions are slightly different, especially for the 7th harmonic order.

A two-sample Kolmogorov-Smirnov test [63] was performed to analyze whether nonparametric models and Gaussian mixture models can be considered as distributions equivalent to real measured data distributions. In Table III, p-values of the test

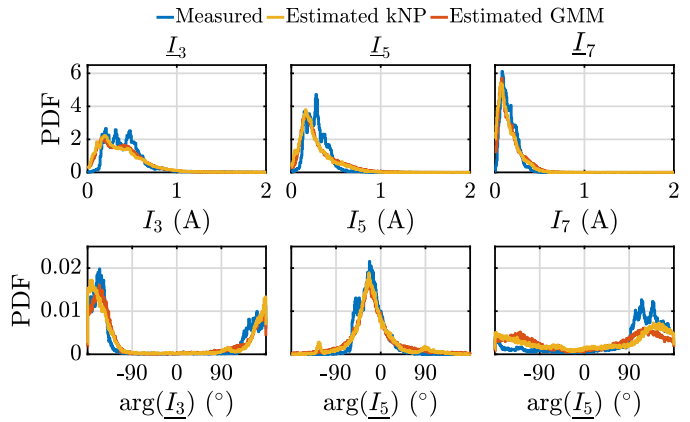


Fig. 15. Probability distribution functions of measured and modeled harmonic currents.

TABLE III  
CONTRAST ANALYSIS P-VALUES BETWEEN MODELS AND MEASURED DISTRIBUTIONS

	Magnitudes		phase angles	
	kNP	GMM	kNP	GMM
3 <sup>rd</sup>	0.14	0.10	0.14	0.06
5 <sup>th</sup>	0.19	0.26	0.09	0.34
7 <sup>th</sup>	0.09	0.34	0.02	0.04

comparing models for the 3rd, 5th and 7th harmonic magnitudes and phase angles are shown. All tests confirm the null hypothesis that models and measurements can be taken as the same distribution with a 95% confidence level (p-value  $> 0.05$ ), except for the phase angle of the 7th harmonic order (p-value  $\geq 0.02$ ), which is close enough to be accepted.

The latter representation did not consider the differences among different residences, as it shows harmonic currents of different measured residences included in the same dataset. To check the performance of the method to model individual residences, Fig. 16 shows a boxplot of the harmonic magnitudes, for the 3rd, 5th and 7th harmonic orders of each measured site and the synthetically generated samples.

The boxplots in Fig. 16 show that the harmonic current magnitudes for individual dwellings are very similar for the two proposed methods and closely match the real measured currents, having a similar variability and range. This proves that the real harmonic distributions of individual residential sites can be properly modeled by the proposed models just by knowing the realistic distribution of the demanded real power of the residence. The same similarities in values and variability are also obtained for phase angles, which are omitted here due to space limitations.

### B. Validation of Harmonic Voltages in a LV Network

Harmonic current models are mainly used to compute harmonic voltages in a distribution network. To validate whether the model provides accurate harmonic voltages, both the measured harmonic currents and the synthetically obtained ones are injected into a low voltage test feeder. The test feeder used is LVN#3 - Feeder 5 [64]. This network consists of

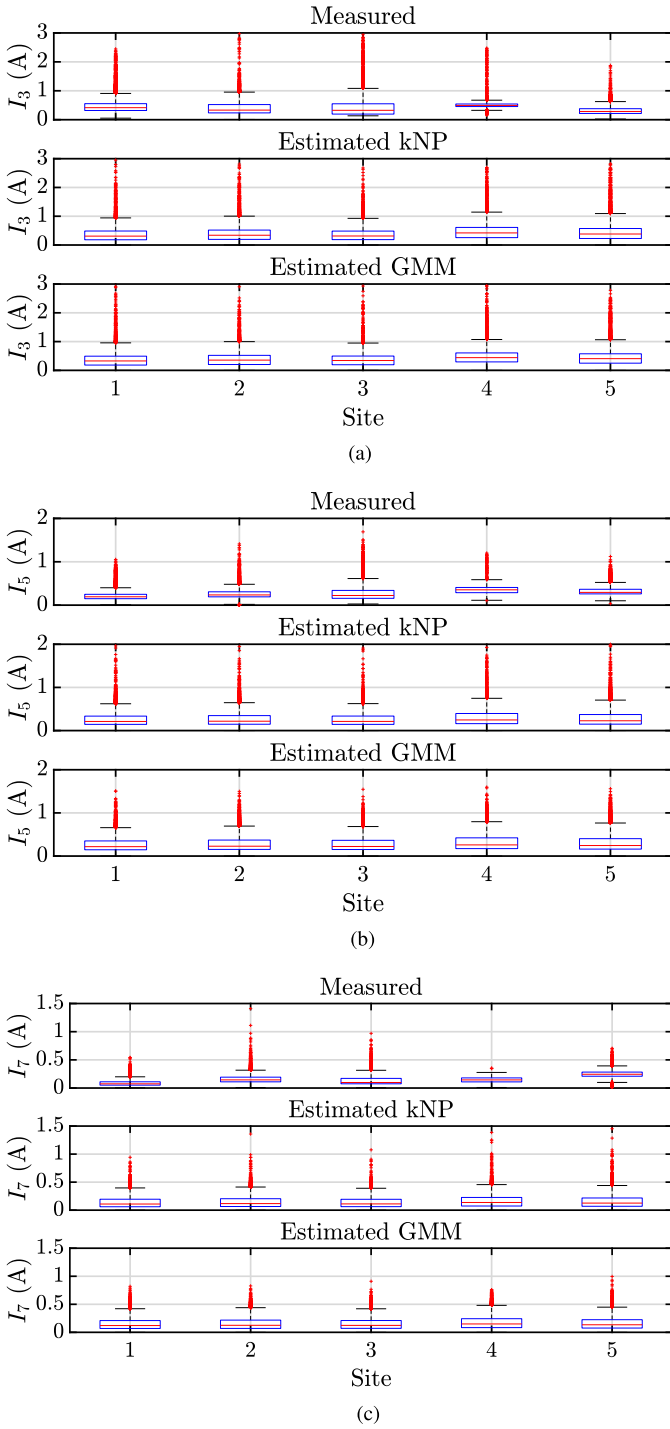


Fig. 16. Boxplot for measured and generated harmonic current magnitudes. (a) 3rd harmonic magnitude. (b) 5th harmonic magnitude. (c) 7th harmonic magnitude.

21 residential consumers connected through a feeder to an 11/0.416 kV transformer, whose rated power has been set to 100 kVA. Each of the consumers is allocated to one phase of the network and is chosen among one of the 24 measurements; that is, consumers are single-phase and the network is three-phase. Fig. 17 shows a single line diagram of the test feeder.

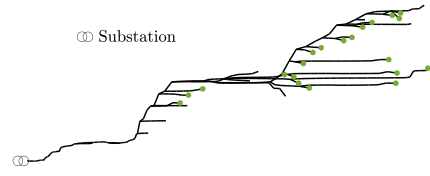


Fig. 17. Representative European Network #3 - Feeder 5 single line diagram.

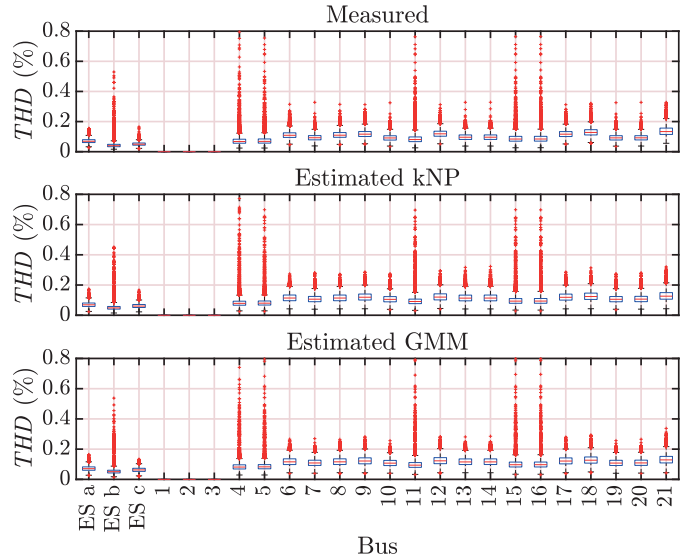


Fig. 18. Voltage  $THD$ .

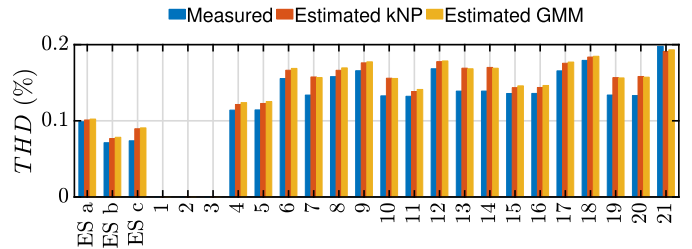


Fig. 19. 95th percentile of voltage  $THD$ .

Harmonic currents are injected into the test feeder, and a current injection problem is solved with OPENDSS simulator and in the MATLAB environment. To do so, residential consumers are considered constant  $P$  and  $Q$  loads at fundamental frequencies and as Norton equivalents for harmonic frequencies.

According to this procedure, Fig. 18 shows the voltage  $THD$  obtained at each of the three phase angles of the secondary bus of the main substation (ES) and at each of the consumer points of common coupling when injecting measured and synthetic harmonic currents. Data are presented in the form of a boxplot. Additionally, Fig. 19 shows the 95th percentile of voltage  $THD$  at network buses, a typical value in voltage harmonic standards [8], which is accurately estimated with both methodologies.

The three plots in Fig. 18 show the same voltage  $THD$ . Synthetically generated harmonic currents lead to the same level of voltage harmonics as measured currents, and only minor differences can be detected at some buses regarding the

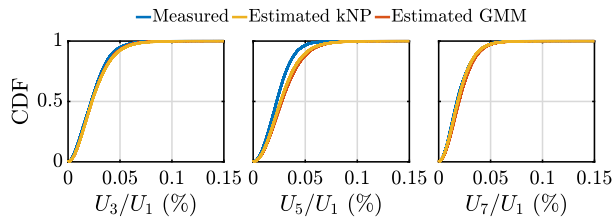


Fig. 20. Cumulative distribution function (CDF) of voltage harmonics at the secondary of substation.

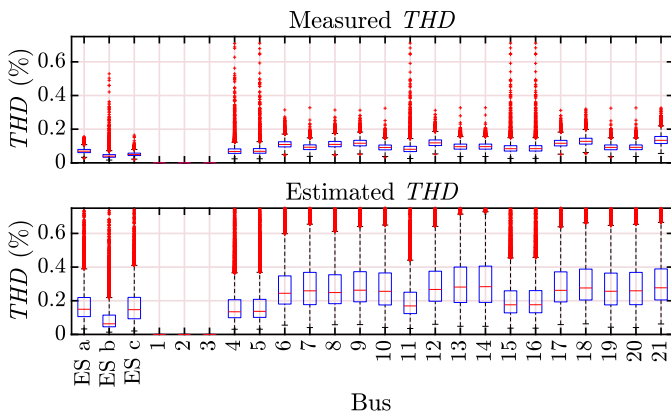


Fig. 21. Voltage  $THD$  obtained with the measured harmonic currents and with kNP distribution functions when a single power demand interval is used.

maximum values of  $THD$ . As previously explained, although the kNP distribution model provides very accurate results, its simulation is very time-consuming, as obtaining random samples for each time interval of the simulated week for every harmonic order requires high computational effort. For instance, sampling harmonic injections every minute in the week (10,080 intervals) requires 3 minutes with Gaussian mixture models and 31 minutes with kNP models when simulated in an AMD Ryzen 3960X CPU @ 3.8-4.5 GHz with 128 GB RAM.

To look closer at individual harmonic voltages, distortion at the substation secondary is assessed. This allows us to study how harmonic models are accurate enough to sum or cancel considering their magnitudes and phase angles as real currents do. In Fig. 20, the 3rd, 5th and 7th harmonic voltages at the secondary bus of a transformer during one week are shown in the form of cumulative distribution functions. A very slight overestimation of the 5th harmonic voltages can be observed, whereas the 3rd and 7th harmonic voltages are very accurately obtained. The kNP model slightly better replicates the real voltage values, although the Gaussian mixture models also provide sufficiently accurate results.

### C. Validation of the Segmentation Procedure

Finally, to justify the method applied based on segmenting distributions into different power ranges, Fig. 21 is presented. It follows the same idea presented in Fig. 18, but in this case, the kNP harmonic model is developed based on only one real power interval (from 0 to 8000 W) to generate synthetic current profiles, i.e., no segmentation of data according to power is applied.

Harmonic voltages are highly overestimated in this case due to the omission of the interdependency between harmonic orders and the time, which is achieved through power segmentation. This results in higher harmonic summations, i.e., the correlation of harmonics is underestimated.

## VII. CONCLUSION

In this paper, different alternatives to model residential harmonic injections have been described. The proposed methods are applied to and validated against harmonic and real power measurements carried out in 31 individual residences over a one-week period with a 1 minute resolution in Madrid, Spain, in 2020 and 2021.

The first proposed method models the distribution of harmonics in each power segment with kernel density estimation, while the second applies Gaussian mixture models to fit the data. To obtain stochastic models, all harmonic data were segmented according to different levels of demanded power, and probability distributions were obtained for each of the intervals for harmonic magnitudes. A similar procedure was followed to model harmonic phase angles, but in this case segmentation was performed based on different values of corresponding harmonic magnitudes. Both models resulted in an accurate fitting of the measured data. Further validation was carried out to compare harmonic voltages across the test feeder caused by the injection of the synthetically generated currents and measured currents. This validation confirmed that the results of kNP models are slightly more accurate, but very close to those of the GMM models which, on the other hand, kNP models are simpler to use and require less memory resources as they parameterize the modelled probability functions.

A comparison between patterns of harmonic injection of individual residential customers from different countries/locations has been identified by the authors as a future line of research.

All developed models and harmonic measurement data are included in an Open Science database and are available for researchers for further study and validation.

## APPENDIX GAUSSIAN MIXTURE MODELS

A Gaussian Mixture Model (GMM) is a function that is comprised of several Gaussians, each identified by  $i \in \{1, \dots, C_N\}$ , where  $C_N$  is the total number of clusters (or components) of the dataset. In a univariate (one-dimensional) case, each Gaussian  $i$  in the mixture is characterized by means of the following classical Gaussian equation:

$$\mathcal{N}(x|\mu_i, \sigma_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right) \quad (2)$$

where  $\mu_i$  is the mean  $\sigma_i$  is the variance of the  $i$ -th Gaussian component.

In addition, the GMM is defined by the mixture of component weights which provide a weight or probability factor  $\phi_i$  that expresses the probability of a data point  $x$  to belong to a certain

component  $i$ -th as follows:

$$p(x) = \sum_{i=1}^{C_N} \phi_i \mathcal{N}(x | \mu_i, \sigma_i) \quad (3)$$

The probability factor  $\phi_i$  of each component must fulfill the following constrain:

$$\sum_{i=1}^{C_N} \phi_i = 1 \quad (4)$$

So that the total probability distribution normalizes to 1.

In order to determine the parameters of GMM models, the Expectation-Maximization (EM) algorithm is used. EM is an iterative method which assumes random components and computes for each data point a probability of being generated by each component of the model. Then, the model parameters are adjusted to maximize the likelihood of the data given those assignments. This is repeated iteratively until converge to a local optimum.

#### REFERENCES

- [1] H. Sharma, M. Rylander, and D. Dorr, "Grid impacts due to increased penetration of newer harmonic sources," *IEEE Trans. Ind. Appl.*, vol. 52, no. 1, pp. 99–104, Jan./Feb. 2016.
- [2] Y. Wang, J. Yong, Y. Sun, W. Xu, and D. Wong, "Characteristics of harmonic distortions in residential distribution systems," *IEEE Trans. Power Del.*, vol. 32, no. 3, pp. 1495–1504, Jun. 2017.
- [3] J. Meyer, A.-M. Blanco, M. Domagk, and P. Schegner, "Assessment of prevailing harmonic current emission in public low-voltage networks," *IEEE Trans. Power Del.*, vol. 32, no. 2, pp. 962–970, Jun. 2017.
- [4] G. G. Ye, M. M. Nijhuis, V. V. Cuk, and J. S. Cobben, "Stochastic residential harmonic source modeling for grid impact studies," *Energies*, vol. 10, no. 3, 2017, Art. no. 372. [Online]. Available: <https://www.narcis.nl/publication/RecordID/oi:library.tue.nl>
- [5] N. R. Watson, T. L. Scott, and S. J. Hirsch, "Implications for distribution networks of high penetration of compact fluorescent lamps," *IEEE Trans. Power Del.*, vol. 24, no. 3, pp. 1521–1528, Jul. 2009.
- [6] P. Rodríguez-Pajarón, A. Hernández, and J. V. Milanović, "Probabilistic assessment of the impact of electric vehicles and nonlinear loads on power quality in residential networks," *Int. J. Elect. Power Energy Syst.*, vol. 129, 2021, Art. no. 106807.
- [7] P. Rodríguez-Pajarón, A. H. Bayo, and J. V. Milanović, "Probabilistic assessment of the influence of transformer rating on power quality indices in future residential networks," *Int. J. Elect. Power Energy Syst.*, vol. 135, 2022, Art. no. 107582.
- [8] *Voltage Characteristics of Electricity Supplied by Public Distribution Networks*, CENELEC, Brussels, Belgium, Standard EN 50160, 2010.
- [9] *IEEE Recommended Practices and Requirements for Harmonic Control in Electrical Power Systems*, IEEE Standard P519, 2015. [Online]. Available: <https://ieeexplore.ieee.org/document/7050220>
- [10] P. Rodríguez Pajarón and A. H. Bayo, "Measurements of residential injections of harmonics - MEMORIA," e-cienciaDatos, V3, 2021. [Online]. Available: <https://doi.org/10.21950/JOW63G>
- [11] R. Torquato, Q. Shi, W. Xu, and W. Freitas, "A Monte Carlo simulation platform for studying low voltage residential networks," *IEEE Trans. Smart Grid*, vol. 5, no. 6, pp. 2766–2776, Nov. 2014.
- [12] D. Salles, C. Jiang, W. Xu, W. Freitas, and H. E. Mazin, "Assessing the collective harmonic impact of modern residential loads—Part I: Methodology," *IEEE Trans. Power Del.*, vol. 27, no. 4, pp. 1937–1946, Oct. 2012.
- [13] A. M. Blanco, A. Greverer, S. Miller, J. Meyer, and P. Schegner, "Stochastic harmonic load model of residential users based on measurements," in *Proc. IEEE Conf. Eindhoven PowerTech*, 2015, pp. 1–6.
- [14] A. Blanco Castañeda, "Stochastic harmonic emission model of aggregate residential customers," Ph.D. dissertation, Fakultät Elektrotechnik und Informationstechnik, KIT, Karlsruhe, Germany, Aug. 2017.
- [15] C. Cresswell, "Steady state load models for power system analysis," Ph.D. dissertation, Univ. Edinburgh, Edinburgh, U.K., 2009.
- [16] A. Mansoor, W. Grady, P. Staats, R. Thallam, M. Doyle, and M. Samotyj, "Predicting the net harmonic currents produced by large numbers of distributed single-phase computer loads," *IEEE Trans. Power Del.*, vol. 10, no. 4, pp. 2001–2006, Oct. 1995.
- [17] L. Sainz, J. Pedra, and J. J. Mesas, "Single-phase full-wave rectifier study with experimental measurements," *Electric Power Syst. Res.*, vol. 77, no. 3–4, pp. 339–351, 2007.
- [18] M. F. McGranaghan, S. Santoso, R. C. Dugan, and H. W. Beaty, *Electrical Power Systems Quality*, 3rd ed., New York, NY, USA: McGraw-Hill, 2012.
- [19] M. J. Ghorbani, M. S. Rad, H. Mokhtari, M. E. Honarmand, and M. Youhannaie, "Residential loads modeling by norton equivalent model of household loads," in *Proc. Asia-Pacific Power Energy Eng. Conf.*, 2011, pp. 1–4. [Online]. Available: <https://ieeexplore.ieee.org/document/5748819>
- [20] A. S. Koch, J. M. A. Myrzik, T. Wiesner, and L. Jendernalik, "Evaluation and validation of Norton approaches for nonlinear harmonic models," in *Proc. IEEE Grenoble Conf.*, 2013, pp. 1–6.
- [21] M. J. Ghorbani, M. S. Rad, H. Mokhtari, M. E. Honarmand, and M. Youhannaie, "Residential loads modeling by norton equivalent model of household loads," in *Proc. Asia-Pacific Power Energy Eng. Conf.*, 2011, pp. 1–4.
- [22] M. Fauri, "Harmonic modelling of non-linear load by means of crossed frequency admittance matrix," *IEEE Trans. Power Syst.*, vol. 12, no. 4, pp. 1632–1638, Nov. 1997.
- [23] Y. Sun, G. Zhang, W. Xu, and J. G. Mayordomo, "A harmonically coupled admittance matrix model for AC/DC converters," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 1574–1582, Nov. 2007.
- [24] Y. Sun, X. Xie, L. Zhang, and S. Li, "A voltage adaptive dynamic harmonic model of nonlinear home appliances," *IEEE Trans. Ind. Electron.*, vol. 67, no. 5, pp. 3607–3617, May 2020.
- [25] M. T. Au and J. V. Milanović, "Development of stochastic aggregate harmonic load model based on field measurements," *IEEE Trans. Power Del.*, vol. 22, no. 1, pp. 323–330, Jan. 2007.
- [26] E. Kazibwe, T. Ortmeier, and M. Hammam, "Summation of probabilistic harmonic vectors (power systems)," *IEEE Trans. Power Del.*, vol. 4, no. 1, pp. 621–628, Jan. 1989.
- [27] X. Xiao, Z. Li, Y. Wang, and Y. Zhou, "A practical approach to estimate harmonic distortions in residential distribution system," *IEEE Trans. Power Delivery*, vol. 36, no. 3, pp. 1418–1427, Jun. 2020.
- [28] G. Ye, M. Nijhuis, V. Cuk, and J. F. Cobben, "Appliance-based residential harmonic load modelling," in *Proc. IEEE Manchester PowerTech*, 2017, pp. 1–6.
- [29] P. Rodríguez-Pajarón, E. Caro, A. Hernández, and M. Izzeddine, "A bottom-up model for simulating residential harmonic injections," *Energy Buildings*, 2022, pp. 1–14. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378778822002742>
- [30] A. Bosovic et al., "Deterministic aggregated harmonic source models for harmonic analysis of large medium voltage distribution networks," *IET Generation, Transmiss. Distrib.*, vol. 13, no. 19, pp. 4421–4430, 2019.
- [31] T. Goetze and W. Wellssow, "A statistical approach to the calculation of harmonics in MV systems caused by dispersed LV customers," in *Proc. Power Ind. Comput. Appl. Conf.*, 1995, pp. 221–227.
- [32] M. Au and J. Milanovic, "Stochastic assessment of harmonic distortion level of medium voltage radial distribution network," in *Proc. Int. Conf. Probabilistic Methods Appl. to Power Syst.*, 2006, pp. 1–6.
- [33] H. E. Mazin, E. E. Nino, W. Xu, and J. Yong, "A study on the harmonic contributions of residential loads," *IEEE Trans. Power Del.*, vol. 26, no. 3, pp. 1592–1599, Jul. 2011. [Online]. Available: <https://ieeexplore.ieee.org/document/5910684>
- [34] A. Reinhardt et al., "On the accuracy of appliance identification based on distributed load metering data," in *Proc. 2nd IFIP Conf. Sustain. Internet ICT Sustainability*, 2012, pp. 1–9.
- [35] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, "Smart\*: An open data set and tools for enabling research in sustainable homes," in *Proc. Workshop Data Mining Appl. Sustainability*, 2012, Art. no. 108.
- [36] Pecan Street Inc. Dataport, 2017. [Online]. Available: <http://www.pecanstreet.org/>
- [37] J. Z. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Proc. Workshop Data Mining Appl. Sustain.*, 2011, pp. 59–62.
- [38] A. Filip et al., "Blued: A fully labeled public dataset for event-based nonintrusive load monitoring research," in *Proc. 2nd Workshop Data Mining Appl. Sustainability*, 2011, Art. no. 2012.

- [39] J. Kelly and W. Knottenbelt, "The U.K.-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five U.K. homes," *Sci. Data*, vol. 2, no. 150007, pp. 1–14, Mar. 2015.
- [40] P. Rodríguez-Pajarón, A. Hernández, R. Castro, and M. Izzeddine, "Statistical analysis of field measurement data of harmonic currents in residential distribution system," in *Proc. 26th Int. Conf. Exhib. Electricity Distrib.*, 2021, pp. 1–6.
- [41] 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," CENELEC, Brussels, Belgium, Standard IEC 61000-4-7, 2018.
- [42] A. M. Jungreis and A. W. Kelley, "Adjustable speed drive for residential applications," *IEEE Trans. Ind. Appl.*, vol. 31, no. 6, pp. 1315–1322, Nov./Dec. 1995.
- [43] F. Zare, F. Blaabjerg, P. Davari, G. W. Chang, and J. Adabi, "IEEE access special section editorial: Power quality and harmonics issues of future and smart grids," *IEEE Access*, vol. 7, pp. 132803–132805, 2019.
- [44] I. Richardson, M. Thomson, D. Infield, and C. Clifford, "Domestic electricity use: A high-resolution energy demand model," *Energy Buildings*, vol. 42, no. 10, pp. 1878–1887, 2010.
- [45] H. Palahalli, P. Maffezzoni, and G. Grusso, "Gaussian copula methodology to model photovoltaic generation uncertainty correlation in power distribution networks," *Energies*, vol. 14, no. 9, 2021, Art. no. 2349.
- [46] J. Widén, M. Shepero, and J. Munkhammar, "Probabilistic load flow for power grids with high PV penetrations using copula-based modeling of spatially correlated solar irradiance," *IEEE J. Photovolt.*, vol. 7, no. 6, pp. 1740–1745, Nov. 2017.
- [47] D. A. Brodín, K. Paridari, and L. Nordström, "Matlab applications to generate synthetic electricity load profiles of office buildings and detached houses," in *Proc. IEEE Conf. Innovative Smart Grid Technol.-Asia*, 2017, pp. 1–6.
- [48] E. McKenna and M. Thomson, "High-resolution stochastic integrated thermal–electrical domestic demand model," *Appl. Energy*, vol. 165, pp. 445–461, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0306261915016621>
- [49] F. Tounquet and C. Alaton, *Benchmarking Smart Metering Deployment in the EU-28*. Luxembourg: Publications Office, 2020.
- [50] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer Series in Statistics Series)*. Berlin, Germany: Springer, 2009. [Online]. Available: <https://web.stanford.edu/hastie/ElemStatLearn/>
- [51] M. T. Au and J. V. Milanović, "Establishing harmonic distortion level of distribution network based on stochastic aggregate harmonic load models," *IEEE Trans. Power Del.*, vol. 22, no. 2, pp. 1086–1092, Apr. 2007.
- [52] C. Heumann and M. Schomaker Shalabh, *Introduction to Statistics and Data Analysis*. Berlin, Germany: Springer, 2016.
- [53] MathWorks, *Statistics and Machine Learning Toolbox User's Guide*. Natick, MA, USA: The MathWorks, Inc., Mar. 2020.
- [54] P. Kuwalek, "The application of kernel density estimation for aided the process of locating sources of voltage fluctuations," *Przegląd Elektrotechniczny*, vol. 95, no. 8, pp. 70–74, 2019.
- [55] Y. Wang, H. Luo, and X.-Y. Xiao, "Voltage sag frequency kernel density estimation method considering protection characteristics and fault distribution," *Electric Power Syst. Res.*, vol. 170, pp. 128–137, 2019.
- [56] Q.-k. Sun, J.-b. Wang, T.-j. Jing, and Y.-h. Guo, "Research on voltage quality analysis system based on kernel density estimation for power grid," in *Proc. IEEE Conf. Energy Internet Energy Syst. Integration*, 2017, pp. 1–6.
- [57] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Evanston, IL, USA: Routledge, 2018.
- [58] J. Meyer and P. Schegner, "Characterization of power quality in low voltage networks based on modeling by mixture distributions," in *Proc. Int. Conf. Probabilistic Methods Appl. Power Syst.*, 2006, pp. 1–6.
- [59] M. Jarkovoi, L. Kítt, and M. Iqbal, "Probabilistic bivariate modeling of harmonic current," in *Proc. Int. Conf. Harmon. Qual. Power*, 2020, pp. 1–6.
- [60] Z. Tang, Q. Shu, F. Xu, and Y. Jiang, "Novelty method for the utility harmonic impedance estimation based on Gaussian mixed model," *IET Generation, Transmiss. Distrib.*, vol. 14, no. 13, pp. 2573–2580, 2020.
- [61] G. McLachlan and D. Peel, *Finite Mixture Models*. Hoboken, NJ, USA: Wiley, Nov. 2000.
- [62] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ, USA: Wiley, 2009, vol. 344.
- [63] F. J. Massey Jr., "The Kolmogorov-Smirnov test for goodness of fit," *J. Amer. Stat. Assoc.*, vol. 46, no. 253, pp. 68–78, 1951.
- [64] V. Rigoni, L. F. Ochoa, G. Chicco, A. Navarro-Espinosa, and T. Gozel, "Representative residential LV feeders: A case study for the North West of England," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 348–360, Jan. 2016.

**Pablo Rodríguez-Pajarón** (Student Member, IEEE) received the B.S. and M.Sc. degrees in industrial engineering and the Ph.D. degree in electrical engineering from the Universidad Politécnica de Madrid, Madrid, Spain, in 2017, 2019, and 2022, respectively.

**Araceli Hernández** (Member, IEEE) is currently an Associate Professor with the Department of Automatics, Electrical and Electronics Engineering and Industrial Computing, Universidad Politécnica de Madrid, Madrid, Spain.

**Hugo Mendonça** (Member, IEEE) received the B.Eng. degree in electrical engineering from the Universidade Federal do Maranhão, São Luís, Brazil, in 2006, the M.Sc. and the Ph.D. degrees in electrical engineering from the Universidad Politécnica de Madrid, Madrid, Spain, in 2012 and 2020, respectively.

**Jovica V. Milanović** (Fellow, IEEE) is currently a Professor of electrical power engineering and the Deputy Head of the Department of Electrical and Electronic Engineering, The University of Manchester, Manchester, U.K.