High Reporting Rate Smart Metering Data for Enhanced Grid Monitoring and Services for Energy Communities

Mihai Sănduleac, *Member, IEEE*, Irina Ciornei, *Senior Member, IEEE*, Lucian Toma, *Senior Member, IEEE*, Radu Plămnescu, *Student Member, IEEE*, Ana-Maria Dumitrescu, *Member, IEEE*, and Mihaela M. Albu, *Senior Member, IEEE*

***Abstract*—Capturing the dynamic behavior of the power distribution grids, especially under high penetration of renewables, is of high interest for grid operators. The distribution power grids are not fully observable due to lack of sufficient metering infrastructure, especially downstream of medium-voltage substations. Therefore, fusion of data recorded at significantly different reporting rates is proposed to increase the situational awareness of the system with non-negligible effect on the accuracy of the monitoring tool. Higher reporting rates are possible for next generation smart meters (SM), but they raise higher concerns about data privacy, already an issue for SM rollout. This article proposes a framework for knowledge extraction from high reporting-rate SM data. The process takes place at SM level and with low computation and communication costs and preserving user privacy, with the scope to increase the accuracy of the monitoring tools for distribution power grids. The methodology makes use of statistical metrics able to capture system dynamics relevant for network diagnosis. The proposed approach is validated on a three-phase low-voltage power flow model applied to a realistic testbed microgrid and real field measurements synchronized at 1 s.**

Manuscript received March 15, 2021; revised June 7, 2021; accepted June 24, 2021. Date of publication July 7, 2021; date of current version February 18, 2022. This work was supported in part by the European Union's Horizon 2020 Research and Innovation Program under Grant 857801 (WEDistrict) and Grant 739551 (KIOS CoE), in part by the Romanian Ministry of Education and Research, CNCS—UEFISCDI, under Grant PNIII-P4-ID-PCE-2020-2876, EMERGE, and in part by the Government of the Republic of Cyprus from European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation under Project INTEGRATED/0916/0035. Paper no. TII-21-1219. (*Corresponding author: Irina Ciornei.*)

Mihai Sănduleac, Lucian Toma, Radu Plămnescu, Ana-Maria Dumitrescu, and Mihaela M. Albu are with the MicroDERLab Research Group of Politehnica, University of Bucharest, 030018 Bucharest, Romania (e-mail: mihai.sanduleac@upb.ro; lucian.toma@upb.ro; radu.plamanescu@upb.ro; anamaria.dumitrescu@upb.ro; albu@ieee.org).

Irina Ciornei is with the KIOS Research and Innovation Center of Excellence, University of Cyprus, 1678 Nicosia, Cyprus, and also with the MicroDERLab Research Group of University Politehnica of Bucharest (UPB), 030018 Bucharest, Romania (e-mail: ciornei.irina@ucy.ac.cy).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2021.3095101.

Digital Object Identifier 10.1109/TII.2021.3095101

***Index Terms*—Data privacy, dynamic behavior of power grids, high reporting rate smart meters (HRRSM), quality of supply (QoS), technological knowledge extraction.**

NOMENCLATURE

Variables

p	Net active power (kW) defined as difference between self generation and consumption at prosumers' nodes; defined over a time interval, usually reported at 1 s.
q	Reactive power (kvar).
u	Voltage amplitude (line to neutral) (V).
i	Line current, rms value (A).
$p^t_{l_k}$	Active power absorbed from the grid reported at time t and measured by the k^{th} meter.
$p^t_{pv_k}$	Active power injected into the grid reported at time t by the k^{th} meter.
$p^t_{s_k}$	Net-active power exchanged with the grid reported at time t by the k^{th} meter.
$\boldsymbol{X}$	Tuple of all discrete random variable vectors, as a time series of quadrant measurands, associated with each smart meter (SM).
$x_{k,t}$	t^{th} sample of the time series vector $\boldsymbol{X}_k$ for data reported by the k^{th} SM.
y_{w_a}	Aggregated value of the 1-s time-series subset, $[\boldsymbol{X}]_{w_a}$, in the range of the aggregation time window, w_a.
$y^{dSN}_{w_a}$	Aggregated time-series for one day over the aggregation time window w_a.
$\boldsymbol{S}$	Set of all metering units, including SM.
s_k	Element of the set $\boldsymbol{S}$.
SM	Set of high reporting rate SM (HRRSM).
sm_k	kth SM in the set of all HRRSM.
$\boldsymbol{W}$	Set of time windows under analysis.
w_a	Element of the time window set $\boldsymbol{W}$.
$\overline{X}_{p95}$, $\underline{X}_{p95}$	Upper and lower bounds of the 95-percentiles of the measurand $x \in \boldsymbol{X}_k$(1-s data).
$[\hat{x}_{lb},\ \hat{x}_{ub}]_{w_a}$ $p95\%$ or $p99\%$	Estimated lower, $\hat{x}_{lb}$, and upper, $\hat{x}_{ub}$, bounds of the 95-percentile, p95% or of the 99-percentile, p99%, respectively, of the

	distribution of the averaged measurand output, x, on the time window w_a.
Functions	
$\phi(\cdot)$	Aggregation function for the specific time window.
$\Delta(\cdot)$	Signal decomposition function.
$F(x)$	Cumulative distribution function (cdf) of a random variable x, with $x \in \boldsymbol{X}_k$.
$f(t)$	Probability density function of a continuous random variable $x(t)$.
Indices	
t	Time index or element position in time-series.
k	Index indicating the meter in the sets $\boldsymbol{S}$ or **SM**.
a	Index denoting the time window (e.g., w_{60}, or $a = 60$ means 1-min time window, while w_{900}, means the 15-min time window).
N	Total number of SMs.
dSN	Total number of 1-s samples in daily time series (daily samples number).
π_p	percentile of a discrete random variable.

I. Introduction

THE POWER distribution grids are facing major structural and operational transformations due to increasing deployment of distributed energy resources, especially renewable energy resources (RES) at low-voltage (LV) feeders on top of changes in the typical LV load profiles (LP). These are associated with higher dynamical change in the operation of the network [1], [2]. The tool at hand for the distribution system operators (DSOs) for monitoring and operation of this part of the power gird is the distribution management systems, which has as core subsystem, meter data management system (MDMS) collecting and analyzing data from advanced metering infrastructures (AMI) and from smart meters (SMs) [3]. This data is not synchronized. On the contrary, it comes at significantly different time intervals and with different quality of associated information (e.g., some network automation sensors may report as fast as 30 samples per second, while SMs often report every 1 h or at best every 15 min) [3], [4]. The MDMS provides data to be further analyzed by other operational submodules such as the enterprise resource planning, transformer load management, outage management system, or the mobile workforce management [3], for which enhanced functionalities were proposed based on the information coming from SMs [5]. Among these, one can mention voltage and frequency support, fault detection and localization [4], electricity thefts detection [5], or demand side response [6], [5]. While these analytical tools commonly rely on the data coming from the SMs, the information is in fact filtered information on the energy delivered in—sometimes unspecified—time intervals rather than, for example, a much useful, power profile. In the following, we denote as "instantaneous" active power at a given moment t the active power evaluated over several periods ($T = 20$ ms in Europe) of the voltage and current signals and reported at the end of the time interval (e.g., every 1 s). Despite the acknowledged need of the DSO for closer observability of the dynamics in each node of the LV network [1], [7], the challenge arise when mitigating 1) costs (e.g., expensive solution to deploy AMI at each MV/LV node to ensure system observability); 2) low quality information from already available infrastructure (e.g., data from SM, if available, comes asynchronously in aggregated form every 60 or 30 min, thus limiting the observability of quantities such the rms values of voltage and current, "instantaneous" active and reactive power [7], [8] reported every 1s; and 3) privacy and cyber security concerns of grid users in providing access to high resolution data collected from SM [9].

The main scope of large uptake of SM infrastructure was to automate the readings with a reporting rate higher than the billing period [5]. However, with the advancement of communication and data analytics capabilities of cheap edge computing extensions, SMs are becoming an important source of information for which increasingly sophisticated data analytics are used to enhance network services such as 1) extracting load signatures using nonintrusive aggregation of measurements gathered with high resolution data acquisition systems (1 s sampling rate) [10]; 2) estimate the network states [7]; 3) load profiling [11]; 4) improve forecasting of both load and generation [12]; and 5) peak-shaving incentives for bills and network costs reductions [13], among many others. For most of these applications, reporting rates of 1 h or half an hour were usually considered as sufficient because the amount of RES penetration was relatively low. However, increased spatial granularity of the energy exchange in modern distribution networks accommodating large number of prosumers and dispersed generation requires monitoring of the power transfer instead of the energy balancing [13]. To achieve this, it is necessary to estimate the system states with resolution as low as 1 min (and even below in case of microgrids). Skewness techniques [7] or fusion of higher resolution data from MV substations metering equipment (e.g., 1 min or below) and asynchronous data from LV SMs were proposed to assess this need [7], [8], [14], [15]. However, despite an improvement in the state estimator models at theoretical level, the obtained virtual measurements for LV nodes still preserve the smoothness effect of averaging measurement values (e.g., on 30 min time windows). This might hinder undetected issues ranging from operation stress of the power assets (e.g., LV transformers and cables), abnormal operation of protection devices, to poor quality of electricity supply. While those phenomena are acknowledged by few reports [16], [17], quantification of their severity is yet unexplored.

A next-generation concept for SM, able to process measurement information made available at 1-s intervals, was recently demonstrated as part of a European project [18]. Further, several past and more recent works looked into nonintrusive load monitoring techniques using aggregated measurements at load level which were gathered with very high reporting rates, down to 1 s [10], [19]. While these applications may serve the user, the grid, or other third parties (e.g., aggregators), they still require special techniques for data handling in order to avoid large volumes of data to be collected and processed from many advanced SM [19], [20].

The specific aim of this article is to enhance the monitoring tools of the LV network operators in terms of system dynamics observation (as it can be captured down to 1 s), using information available, and yet unused, from high reporting rate smart meters (HRRSM). The design of the methodology proposed is shaped by the following research challenges: 1) to capture the footprint of the system dynamics within the industry practice reporting rates; 2) to avoid any increase in cost on both the device and for data transmission (e.g., use the current processing unit of the SM, and undergo the same encryption and communication infrastructure already deployed); and 3) to address the data privacy and cybersecurity concerns of the end-user. To this end, the major contributions of this article are as follows:

1) Propose a general framework for technological knowledge extraction from HRRSM data to enhance the monitoring tools of the DSOs, while still preserving the privacy of the user. Technological knowledge is defined as relevant information ready to be used by network operators for monitoring or diagnosis of network assets.
2) Propose meaningful and simple statistical metrics that allow capturing the system dynamics footprint at several sections of the power network. These metrics are dynamically computed and updated within the time window between two consecutive SM's reporting moments, and they are processed at the SM level with low computational cost.
3) Validate the model framed by these metrics using real field measurements from HRRSM and a three-phase LV load-flow model applied to a realistic testbed microgrid with more than 50% renewables installed on the LV feeders.

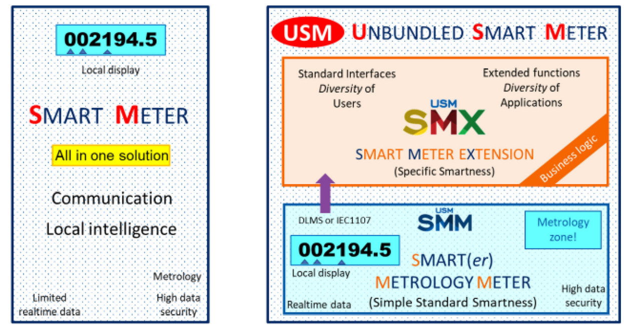
The rest of this article is organized as follows. Section II presents the innovative technology of HRRSM (up to 1 s). Section III presents the problem and the proposed methodology for grid diagnosis using steady-state models for grid operation and real HRRSM data. It also details a methodology for knowledge extraction and quantification for the severity of possible operation or quality of supply (QoS) phenomena encountered in some grid sections. Section IV presents the use cases and critically analyzes the results. Finally, Section V concludes this article.

II. High Reporting Rate Smart Meters

One of the main areas of research in SM deals with increasing their functionalities on top of its metrology scope, accuracy, and reporting rate for billing purposes. In this article, data achieved with the new generation SMs, called unbundled smart meters (USM) [18], were used.

A. USM Components

Fig. 1 shows a comparison between a classical SM and the USM, the latter incorporating one upper layer components, called the smart meter extension (SMX). Thus, in the USM architecture, the smart metrology meter (SMM) is similar to the classical SM and it implements the so-called real-time functions securing the measurement information. Its data cannot be modified from outside, and it is kept unchanged over the whole life of the meter, legally preserving the measurements of energy for billing purposes.



(a) (b)

Fig. 1. SM types. (a) Traditional. (b) Unbundled smart meter.

As a difference from traditional SM, in the upper part of the USM, called SMX, a variety of local intelligence and functions related to smart grid functionalities, as well as the different communication protocols and cyber-security capabilities are implemented. This part is expected to support advanced functionalities that enable new energy services, thus being prone to upgrades during the lifetime of the meter.

B. Applications and Cyber-Security at the SMX Layer

The unbundling into two parts preserves high accuracy and metrology aspects covered by the SMM with the high flexibility needed for a faster electricity market (down to 15 min instead of 1 h) and smart grid requirements in the SMX. In this way, the large investment in an SM infrastructure is done for a usual lifetime also ensuring flexibility for new control modes. With access to the real-time data and with enhanced cyber-security functionality at physical layer of the SMX (e.g., using physical unclonable function) [20], myriad of functionalities is possible. Among the numerous potentials uses for the enhanced data from the SMX, it is worth mentioning the light power quality initiatives such as voltage level assessment and control, high resolution LP, or increasing individual energy awareness [21]. In this article, we are focusing on using high resolution datasets of real-time measurements such as voltage, current, and active power (in rms values) to allow defining new indices for assessing the network capacity and its voltage smoothness from the usual utility reporting rates, which currently are at 15, 30, or 60 min.

III. Problem and Methodology

The primary goal of this article is to provide technological insight on grid operation using HRRSM data on top of already reported temporally aggregated information for billing purposes while preserving the privacy of data to the electricity users. By technological insight, we understand changes of the system operating conditions within the time window between consecutive reporting of SMs, which could have a significant impact on proper operation or the lifetime of the network assets.

A. Problem and Related Works

Generally, dynamic load models are used for assessing the behavior of the network in normal and abnormal operation conditions [2]. The research so far is focused on transmission networks (TN), or at the boundaries between the transmission and distribution [1], [22]. These models use measurements reported every 2 min or even up to 10 ms (phasor measurement units), corroborated with full observability of the network [23]. At LV, the state-of-the-art for estimating the system dynamics (e.g., to take appropriate operation and control actions) is based on estimated load profiling. The latter is determined using aggregated data, such as deriving mean active power from energy consumption recorded at time intervals much higher than TN (e.g., 60 or 30 min) [7]. However, averaging information over such large time windows is not appropriate to model the energy transfer in both directions within the power distribution networks incorporating significant RES-based generators [15]. Methods such as spatio-temporal aggregation were proposed [9] to improve the modeling accuracy. However, there are still two concerns with this approach: 1) they might miss useful information, such as QoS in each LV distribution nodes; and 2) on top of still preserving the smoothing effect on LP, the aggregation process takes place outside the home area network (HAN) of the electricity customer, which might be a source of cyber vulnerability [24]. High resolution SM data, with reporting rates down to 1s have been proposed for non-intrusive load monitoring as a cost-effective solution for appliance level behavior of consumers [10], or to enhance energy management in microgrids [19]. Compared with previous similar works, this paper proposes to capture the dynamics of the LV nodes by appropriate statistical metrics able to provide useful technological insights for the system operator. These metrics could be directly calculated at the functional layer of the meter with similar computational needs, and preserving privacy and cyber security within the SM itself.

B. Methodology

The methodology assumes as baseline scenario the ideal case when the network operator has access to 1-s data. For comparison purposes, three use cases are investigated against this baseline scenario. Two of them refer to the current utility practices, when only temporal-aggregated energy data is transmitted either every 1-h (case A) or down to every 15-min (case B). The third use case deals with the 1-min reporting rates (case C). This is a theoretical but useful case for microgrids operation, similar to TN. The scope is to extract knowledge from HRRSM data which could be used as system dynamic footprint while still preserving user privacy. By *data* we mean the series of recorded measurement values, while by *information* we define the network behavior in terms of voltage variations (including values lower than the minimum of higher than the maximum *a-priori* selected thresholds) and variability of the power/current (rms values) profiles (e.g., spikes).

In order to critically assess this problem, a general knowledge extraction framework (Fig. 2) is proposed. The framework has four layers: (**L1**) ***data collection layer*** which includes the LV level metering sources and an optional data collection and synchronization process for information coming from external sources; (**L2**) ***temporal aggregation layer*** for the local data integrating an expert knowledge formalism; (**L3**) ***descriptive analytics layer*** for both streams of information from the previous two layers; and (**L4**) ***prescriptive analytics and knowledge extraction layer***. The upper layers implement functionalities where data is translated into model parameters and further used in a set of operation specific applications. Two such applications are detailed in the next section of the article.

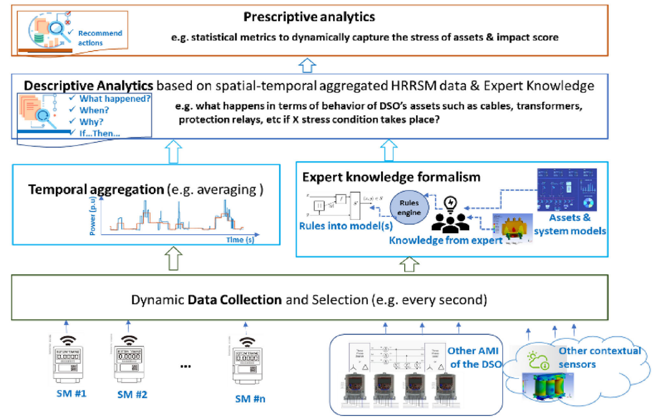


Fig. 2. Framework for knowledge extraction from high-resolution dynamical SM data.

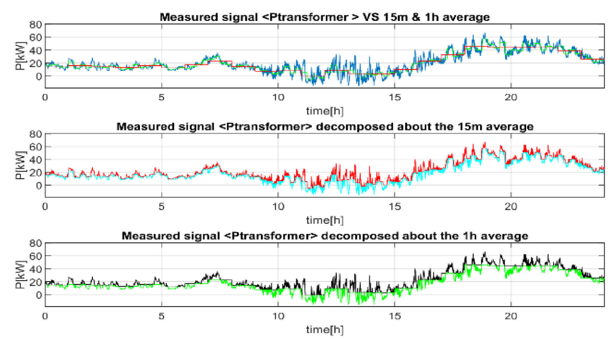


Fig. 3. Decomposition of active power, P at 1 s, on MV/LV transformer against 15- and 60-min time windows.

At the physical layer (**L1**), data is collected from a set of smart measurement units $\boldsymbol{S} = \{s_1, s_2, \ldots, s_N\}$, delivering information on several measurands (e.g., *p, q, u, i*) via measurement results provided with high (1-s) granularity, while the optional information from external contextual sources measurands is retrieved at different time intervals (e.g., every 30 or 15 min for price signals, several hours for handling expected peak time loading, etc.). The second layer (**L2**) consists of two components: 1) the aggregation processor for temporal aggregation of the SM data on 1, 15, 30, and 60 min time windows $w_t \in \boldsymbol{W}$, respectively; and 2) the expert knowledge formalism to be processed on the specific time window.

Let $\boldsymbol{SM} = \{sm1,\ sm2,\ \ldots sm_N\}$ be the subset of HRRSM installed at the local community level, where SM $\in \boldsymbol{S}$. Each SM, sm_k, is measuring a four-quadrant (Q4), subset of measurands $\{p_k, q_k,\ u_k,\ i_k\}$ with a reporting rate of 1 s. Our analysis investigates the behavior of the network following a full diurnal cycle (e.g., total number of Q4 samplings being $T = 86\,400 = 3600$ s $\times$ 24 h). Thus, for each sm_k, we define a daily time-series quadrant matrix $\boldsymbol{X}_k^T = \{x_{k,t}\}$, where $x_{k,t} = \{p_k^t, q_k^t, u_k^t, i_k^t\}$ is the four-quadrant of measurands recorded at the time interval *t* (equal to 1 s). The synchronized data coming from all the SM of the network (from the set ***SM***) forms the tuple $\boldsymbol{X} = \{\boldsymbol{X_1^T}, \boldsymbol{X_2^T}, \ldots, \boldsymbol{X_k^T}, \ldots, \boldsymbol{X_N^T}\}$.

Depending on the scope of the analysis and the network section of interest, only a subset of the tuple $\boldsymbol{X}$ of measurands might be needed. Thus, for the diagnosis purposes of an MV/LV transformer, only *p* and *q* are of interest to obtain another the apparent power, *s* (in kVA) which expresses the loading conditions of the transformer. This is directly proportional with the level of thermal stress of this network asset:

$$s = \Sigma_{\mathrm{t}=1}^{T_{trafo}} \Sigma_{k=1}^{\mathrm{N}} (s_k^t),\text{ where } s_k^t = \sqrt{p_k^{(t2)} + q_k^{(t2)}}(\mathrm{kVA}). \tag{1}$$

The measurement uncertainties for derived quantities are ignored in this article for simplicity purposes. However, they could be easily computed from the SM information, and added later on, without affecting the methodology flow.

In some cases, the SM might be able to report two separate values for the active power in the considered point of common coupling (PCC): the load consumption ($p_{l_k}^t$) and the local (e.g., from PV) power production ($p_{pv_k}^t$). In this case, the net power exchange with the grid ($p_{s_k}^t$) at PCC is defined as

$$p_{(s_k)}^t = p_{(l_k)}^t - p_{(pv_k)}^t(\mathrm{kW}). \tag{2}$$

At the data collection layer, there is also a selection process for the measurand(s) of interest. For simplicity, but without loss of generality, let us assume that one out of the four measurands is of interest for a specific analysis (e.g., the phase-to-neutral voltage in a specific node of the distribution grid, $x_{k,t} = u_k^t$). While rms values of voltage, current, or active power variations above or below the levels imposed by norms (e.g., related to power quality and/or electromagnetic compatibility or by the thermal ratings of the equipment) might be accepted for short periods of time (e.g., several seconds), when they occur frequently might have a detrimental effect on the asset lifetime and normal operation [25]. Therefore, a statistical metric of their frequency of occurrence and severity within the aggregation window are useful information, similar to the assessment of voltage dips. Such metric is the cumulative distribution function (cdf), $F(x)$, of a continuous or discrete random variable, $x(t)$:

$$F\ (x) = P\ (X \leq x) = \int_{-\infty}^{x} f\,(t)\,dt\ ,\text{for } x \in \mathbb{R}. \tag{3}$$

The 100th percentile $(0 \leq a_p \leq 1)$ of a probability distribution with cdf *F*(*x*) is the value π_p, such that

$$F\ (\pi_p) = P\ (X \leq \pi_p) = a_p. \tag{4}$$

In the case of electrical current and active power, ensuring a reliable and safe grid operation requires that their values in the defined (standardized) time intervals lie within limits in more than 95% of their occurrences [26]. Therefore, (3) and (4) become

$$0.95 = F(\bar{P}_{p95}) = \int_{-\infty}^{\bar{P}_{p95}} f\,(p\,(t))\,dt \tag{5}$$

$$0.95 = F(\bar{I}_{p95}) = \int_{-\infty}^{\bar{I}_{p95}} f\,(i\,(t))\,dt\ . \tag{6}$$

In the case of voltage, power quality acceptability levels similarly impose that the rms voltage values estimated on standardized time intervals should be above the minimum level in 95% of occurrences. Therefore, the following applies:

$$0.95 = F(\underline{U}_{p95}) = \int_{\underline{U}_{p95}}^{+\infty} f\,(u\,(t))\,dt. \tag{7}$$

Because some of these bounds are not symmetric relative to the standardized (e.g., nominal or optimal) operating values, we first apply a decomposition of the signal (1-s measurements) to count the positive (upwards) and negative (downwards) exceeds against the to-be-reported SM aggregated value.

Thus, we define the aggregation function, $\phi(\cdot)$, applied to the 1-s signal values, $\boldsymbol{X}_k$, on the aggregation time window, w_a, as the mean value of all 1-s samples, i.e.,

$$\phi_{w_a}\ (\boldsymbol{X}_k) = (\Sigma_{t=1}^{\mathrm{w_a}}\, x_{k,t})/w_a. \tag{8}$$

The proposed aggregation process is able to capture the asymmetrical variations of the signal. This is important because in some diagnostic applications the upward limits might be different than the downwards limits. This approach defers from the power quality aggregation procedure using the quadratic mean (insensitive to the asymmetrical variation of the signal).

For a daily collection of 1-s measurements, the aggregation function will be successively called dSN/w_a times, thus obtaining the aggregated signal, $y_{w_a}^{dSN} = [\phi_{w_a}\ (\boldsymbol{X}_k)]_{dSN}$.

The decomposition function, $\Delta(X_k^{dSN}, y_{w\ a}^{dSN})$, gets the original 1-s signal into a positive component, Δ_+, and a negative component, Δ_-, respectively

$$\Delta_+ = X_k^{dSN} - y_{w\ a}^{dSN} \tag{9}$$

$$\Delta_- = -\left|y_{w\ a}^{dSN} - X_k^{dSN}\right|. \tag{10}$$

An example of this decomposition is presented in Fig. 3.

Then, on each aggregation window, a recursive discrete calculation is made for the percentiles of interest in the case of each measurand (e.g., for active power we apply (5), for current we apply (6), and (7) for voltage, respectively). An illustrative example is provided in Fig. 4. Several percentile bounds, also called confidence bounds, per type of aggregation time window could be created for the rules guiding the prescriptive analytics in Layer **L4** (Table I).

IV. Use Cases and Evaluation of Results

To prove our proposed approach, a realistic LV microgrid testbed with more than 50% RES-based prosumers has been selected. The behavior of this network captures energy transfers governed by lower time constants than the ones assumed traditionally by the DSOs (with still low percentages of RES) [15]. The analysis considers two distinct sets of daily power profiles

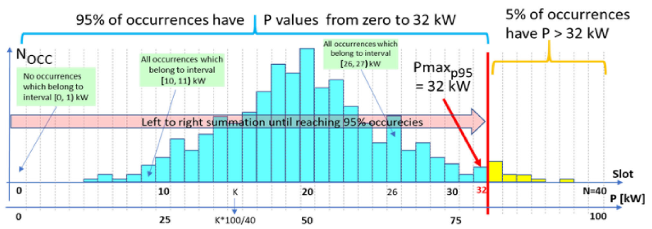


Fig. 4. Illustrative calculation of the 95-percentile for active power (aggregation time window $w_a = 15$ min, and Pmaxp95 = 32 kW.

TABLE I
SIGNAL SAMPLING VERSUS PROBABILITY DISTRIBUTION BOUNDS

#samples 1s (86400)	#samples 1-min (1440)	#samples 15-min (96)	#samples 1h (24)
$[\hat{x}_{lb}, \hat{x}_{ub}]_{1s}$ $p95\%$	$[\hat{x}_{lb}, \hat{x}_{ub}]_{1m}$ $p95\%$	$[\hat{x}_{lb}, \hat{x}_{ub}]_{15m}$ $p95\%$	$[\hat{x}_{lb}, \hat{x}_{ub}]_{1h}$ $p95\%$
$[\hat{x}_{lb}, \hat{x}_{ub}]_{1s}$ $p99\%$	$[\hat{x}_{lb}, \hat{x}_{ub}]_{1m}$ $p99\%$	$[\hat{x}_{lb}, \hat{x}_{ub}]_{15m}$ $p99\%$	$[\hat{x}_{lb}, \hat{x}_{ub}]_{1h}$ $p99\%$

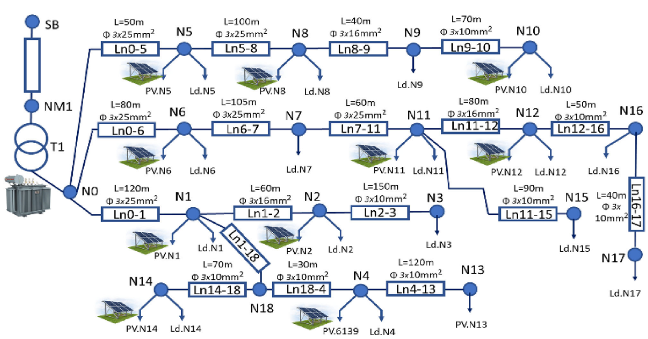


Fig. 5. Topology and specific data of the studied LV grid.

TABLE II
MV/LV TRANSFORMER—PARAMETERS

Rated power	Phases/ coupling	Winding numbers	Primary voltage	Secondary voltage	Core losses	Winding losses
50 kVA	3, Δ/Y	2	20 kV	0.4 kV	145 W	1100 W

based on meteorological conditions: 1) ***Case 1*** is a cloudy day, with low PV production and seldom net-power peaks; 2) ***Case 2*** is a sunny day with many short-time shadows impacting the PV power profile. These two days were selected out of the ten days for which real measurements, fully synchronized 1-s recorded data, were available from 18 distinct locations (Austin database from the PecanStreet data-portal) [27]. The preselection process used a simple algorithm for selecting two most distinct daily net-LP out of the ten available days (e.g., largest root mean square distance). All 18 locations (represented by the nodes $N0$ to $N17$) are connected to the same MV/LV distribution substation. Ten of them are prosumer nodes (with roof-top PV installations), while the other eight are pure loads (Fig. 5).

A network situation without PV production is expected to have lower dynamics and may be subject of other investigations.

All the relevant line and node parameters of the network are provided in the diagram, while the parameters of the MV/LV transformer are given in Table II.

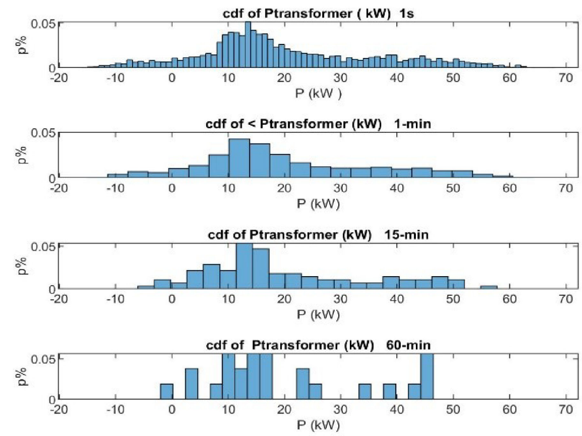


Fig. 6. Cumulative Distribution (cdf) of active power measurements at MV/LV transformer, recorded at 1 s, 1, 15, and 60 min.

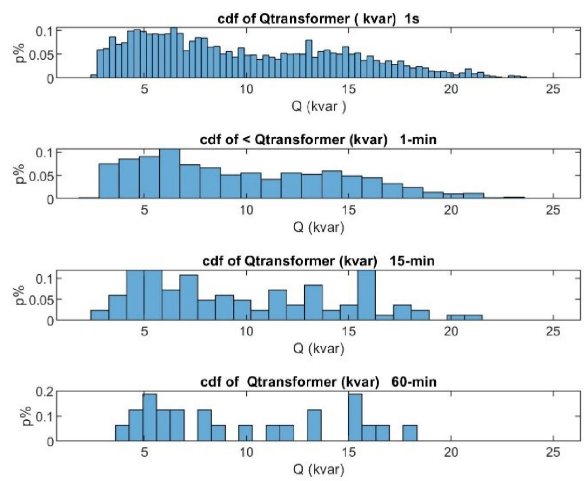


Fig. 7. Cumulative Distribution (cdf) of reactive power measurements at MV/LV transformer, recorded at 1 s, 1, 15, and 60 min.

For simplicity, but without loss of generality, we present the analysis for the balanced three-phase load flow option of the OpenDSS because the scope of the study is to quantify the impact of information loss due data averaging on the accuracy of the considered distribution system model and its network diagnosis capability. It is worth mentioning that the analysis itself does not change in case we use an unbalanced power flow model, besides to be applied on each phase independently.

Three distinctive situations were studied, considering the daily profiles for 1) the active and reactive power on the MV/LV transformer, for capturing the loading level, loading variability, and thermal stress (Figs. 6 and 7); 2) the voltage level on a remote network node (node $N17$, at the end of the LV feeders) to capture the QoS (Fig. 8); and 3) the current on one of the critical lines, such that to assess possible needs for network expansion, or to advice the need for different setups of feeder protection relays in emerging microgrids (Fig. 9). Below we present the cumulative distribution probabilities of several measurands of interest for Case 1 (sunny daily net-LP) as they result from running the load flow model.

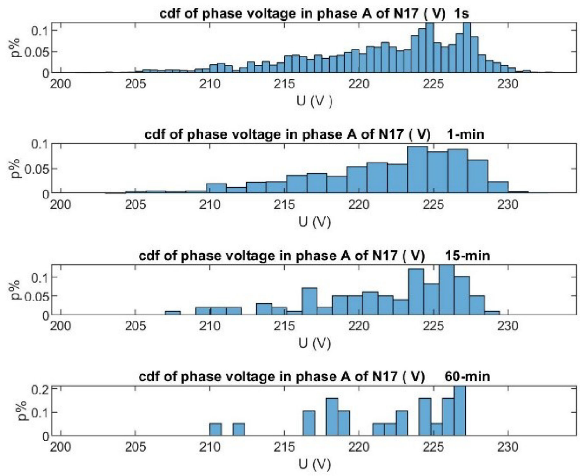


Fig. 8. Cumulative distribution (cdf) of the phase voltage of phase A in node $N17$, recorded at 1 s, 1, 15, and 60.

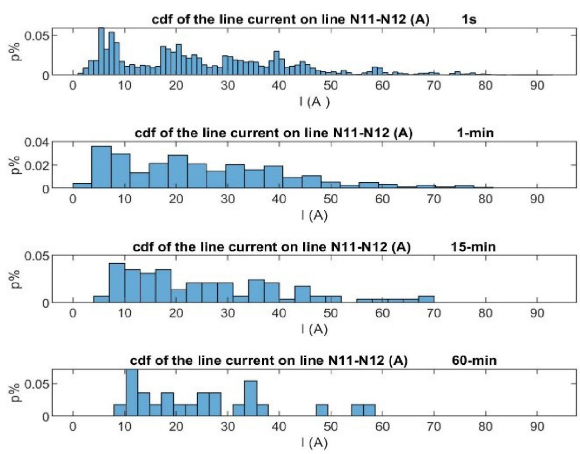


Fig. 9. Cumulative distribution (cdf) of line current ($N11$–$N12$), recorded at 1 s, 1, 15, and 60 min.

1) Active and reactive power (transformer thermal stress).
2) Voltage at $N17$ (QoS).
3) Current on the $N11$–$N12$ line—conductor thermal stress and protection coordination (are needed, e.g., in micro-grids).

Table III gives the percentiles (p95 and p99) of the active power at the MV/LV transformer against the highest recorded value for the active power ($P_{\max}$) for the baseline scenario and each of the three aggregation time windows (1, 15, and 60 min), respectively.

The severity and the percentage of the lost information in terms of dynamic change of the loading conditions of the transformer are given by two terms: (1-p99) and (1-max), respectively. Thus, the difference between p100 (denoted as 1) of the baseline scenario and p99 of the compared scenarios (e.g., 1, 15, 60 min) is denoted in the table as (1-p99). Similarly, the difference between the p100 in the baseline scenario and the maximum recorded value of the measurand of interest in the compared scenarios is demoted as (1-max).

TABLE III
ACTIVE POWER THRESHOLDS IN kW (UPPER BOUNDS) FOR P95-, P99-, AND P100- (MAX), RESPECTIVELY, FOR THE MV/LV TRANSFORMER, BASED ON THE UPWARD DISCRETE INTEGRATION OF THE CORRESPONDING cdf

Time resolution	p95 [kW]	p99 [kW]	max [kW]	1-p99 relative to the 1s information	1-max relative to the 1s information
Case 1 (24 hours, day 1 mix of load and PV production LPs)					
1 second	39.5	46.0	57.5	0 %	0 %
1 min.	39.5	44.5	50.4	3.3 %	12.4 %
15 min.	38.0	42.5	42.4	7.6 %	26.2 %
60 min.	36.5	36.5	36.7	20.7 %	36.1 %
Case 2 (24 hours, day 2 mix of load and PV production LPs)					
1 second	50.5	57.5	67.1	0 %	0 %
1 min.	50.0	57.0	62.8	0.9 %	6.5 %
15 min.	49.0	56.9	56.9	0.9 %	15.3 %
60 min.	45.0	46.5	46.5	19.1 %	31.1 %

TABLE IV
PHASE VOLTAGE THRESHOLD (LOWER BOUNDS) FOR THE REMOTE NODE $N17$ FOR P95-, P99-, AND P100- (MIN), RESPECTIVELY

Time resolution	p95 [V]	p99 [V]	**min** [V]	p99 relative to the 1s information [V]	min relative to the 1s information [V]
Case 1 (24 hours, day 1 mix of load and PV production LPs)					
1 second	210.5	206.0	201.3	0	0
1 min.	210.5	206.0	203.9	0	2.6
15 min.	210.6	206.5	207.1	0.5	5.8
60 min.	211.5	210.3	210.7	4.3	9.4
Case 2 (24 hours, day 2 mix of load and PV production LPs)					
1 second	211.8	208.8	204.3	0	0
1 min.	211.8	209.3	206.0	0.5	1.7
15 min.	212.3	210.0	210.4	1.2	6.1
60 min.	213.5	211.8	212.4	3.0	8.1

It can be seen that in the case of the cloudy day, the quality and quantity of the information lost in terms of system (slow) dynamics are higher than in the case of the sunny day. For Case 1, the loss of information regarding the maximum active power during the day quantified as high as 36.1% when the assessment is based on 60-min LPs and of 26.2% in case of 15-min LPs. Furthermore, in 1% of the situations (described by 1-p99 relative to 1 s information), the quantified loss of information is as high as 20.7% in the case of 60-min time window, and 7.6% in the case of 15-min time window (fifth column). These large deviations emphasize the need for reconsidering the accuracy reached by theoretical state-estimators with pseudo measurements coming from meters with reporting rates above 15 min. They also affect real-time decisions related to power quality, coordination of short time reacting protections, or design of smart devices in distribution networks.

Table IV gives the lowest threshold of the phase voltage values in a remote network node, $N17$, as calculated based on the 95- and 99-percentiles, respectively. The minimum voltage value for the baseline scenario is denoted by **min**. All the above are calculated based on the load flow model run in OpenDSS for each of the use cases (baseline scenario and each of the three averaging time windows 1, 15, and 60 min, respectively). It can be noticed that a significant percentage (>5%) of the voltage dips at the MV/LV transformer pass undetected for reporting rates higher than 15 min (in case no PQ analyzer is installed).

TABLE V
PROBABILITY OF OCCURRENCE P95%, P99% FOR THE CURRENT ON A SELECTED LINE, BASED ON UPWARD INTEGRATING DPS

Time resolution	p95 [A]	p99 [A]	max [A]	1-p99 relative to the 1second information	1-max relative to the 1second information
1 second	59.3	75.3	92.3	0 %	0 %
1 min.	58.7	74.7	79.9	0.9 %	13.5 %
15 min.	57.7	69.3	69.4	8.0 %	24.8 %
60 min.	56.0	56.6	56.6	24.8 %	38.6 %

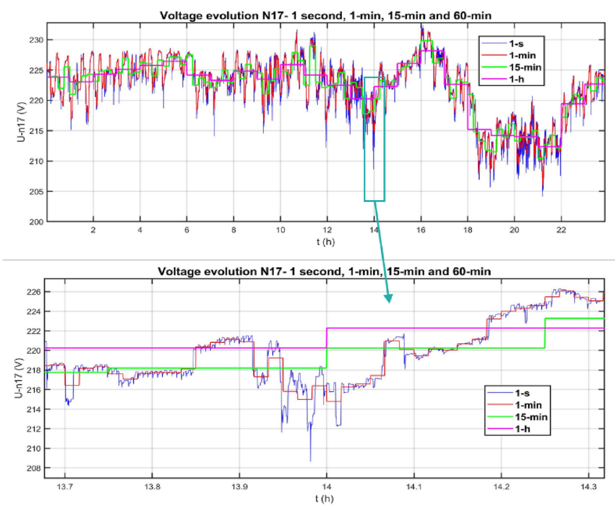


Fig. 10. Voltage profile in *N*17 from measurements with 1 s, 1-, 15-, and 60-min resolution.

Contrary to other impact studies of the RES-based generation where this phenomenon was emphasized only for sunny days (when large reverse power flows occur), it seems that it may actually occur for both cases (sunny and cloudy day).

Traditionally recorded 15 and 60 min discretized LPs give significantly lower p95 and p99 values because they do not capture the dynamics characterized by high spikes, thus hindering QoS issues as well as abnormal loading conditions for MV/LV transformers.

Table V shows the same analysis on the line current measurand on one of the critical line sections of the network. Critical line is defined as the line where the highest power flow could take place. The simulations results show that the thermal stress on this conductor may pass almost undetected above 15 min reporting rates, leading to underestimation of the risks related to conductor failure or to less sensitive setups of the associated protection relays.

Fig. 10 shows the voltage (rms) profiles at one of the critical nodes (*N*17) of the simulated grid. This node particularly emphasizes the significant voltage changes which are not at all visible anywhere above 15 min reporting rates.

In all situations, it can be seen that the high granularity data obtained from HRRSMs could be statistically processed at the SM level to reveal additional information that warn about the variability of the state quantities in emerging LV grids and microgrids (e.g., where available system inertia is inherently low), while still preserving privacy (accurate time series reconstruction from this statistical indices is not possible, without additional information [29]). The fact that boundary information such as p95, p99, and **min/max** values are already useful for a much better understanding of the grid operation aspects allow in case of privacy-sensitive situations (e.g., power profiles for private entities) to still have access to the reality through offline information. Moreover, the p95 and p99 can be also calculated at SM level by end user application, thus being able to send to external actors such as DSO only statistical data which is useful for network operation and planning while preserving desired levels of privacy.

V. CONCLUSION

This article shows the benefits of extracting relevant technological information from high-reporting rate SM using simple to calculate statistical metrics able to preserve user privacy, while observing, offline, the system behavior within the same reporting rates as per current industry standards. This article proposed a general knowledge extraction framework, with a focus on a statistical based methodology which aims to mitigate dual constraints coming from SM data owners (privacy and cyber security concerns) and from the needs of the DSO to enhance the network situational awareness down to the most remote parts of its LV network in order to ensure reliability and high quality of service. Compared to other state of the art approaches where sophisticated data analytics, prone to heavy computational, communication and data handling needs, our proposed approach follows two critical design constraints: to have minimum computational and communication costs and to avoid sending any high resolution large volumes of data outside of the HAN of the user. The proposed statistical metrics to satisfy these constraints are the percentiles (e.g., p95 and p99) and the cumulative probability function (cdf). They are simple enough to be processed and encrypted at SM level using the same processing unit, while capturing relevant system dynamics within the time window between two consecutive reporting moments as per the current industry standard. Our approach was validated using a state of the art three-phase power flow model of a realistic LV microgrid with more than 50% RES. Using real data from a public database, it has been shown, for three different measurands of interest (power on MV/LV transformer, voltage in critical nodes, and current on vulnerable line sections), that there is a significant loss of information regarding the real boundaries of the 95- and 99-percentiles of the grid cases which translates into underestimation of thermal stress on sections of the grid or their critical assets (e.g., transformers) to possible abnormal operation of protection devises, or of the quality of service issues. The proposed methodology is suitable for offline network and assets diagnosis, prescriptive and post-contingency analysis, among others. While our methodology could not be directly applied for online type of operation applications such as state estimation in LV power grids, the proposed statistical metrics could provide additional information to improve the current models (enhance the accuracy of virtual measurements). This will be a future direction of research.

References

[1] E. Proedrou, "A comprehensive review of residential electricity load profile models," *IEEE Access*, vol. 9, pp. 114–133, 2021.

[2] A. Arif, Z. Wang, J. Wang, B. Mather, H. Bashualdo, and D. Zhao, "Load modeling—A review," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5986–5999, Nov. 2017.

[3] A. Ghosal and M. Conti, "Key management systems for smart grid advanced metering infrastructure: A survey," *IEEE Commun. Surv. Tut.*, vol. 21, no. 3, pp. 2831–2848, Jul.–Sep. 2019.

[4] W. Luan, J. Peng, M. Maras, J. Lo, and B. Harapnuk, "Smart meter data analytics for distribution network connectivity verification," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1964–1971, Jul. 2015.

[5] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019.

[6] R. Deng, Z. Yang, M. Chow, and J. Chen, "A survey on demand response in smart grids: Mathematical models and approaches," *IEEE Trans. Ind. Informat.*, vol. 11, no. 3, pp. 570–582, Jun. 2015.

[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] B. Ritt *et al.*, "A distribution system state estimation for an efficient integration of electric vehicle charging infrastructure into low-voltage grids," in *Proc. Int. ETG-Congr. ETG Symp.*, 2019, pp. 1–6.

[9] L. Zhang, J. Zhang, and Y. H. Hu, "A privacy-preserving distributed smart metering temporal and spatial aggregation scheme," *IEEE Access*, vol. 7, pp. 28372–28382, 2019.

[10] M. Dong, P. C. M. Meira, W. Xu, and C. Y. Chung, "Non-intrusive signature extraction for major residential loads," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1421–1430, Sep. 2013.

[11] N. Uribe-Pérez *et al.*, "State of the art and trends review of smart metering in electricity grids," *Appl. Sci.*, vol. 6, no. 3, Mar. 2016, Art. no. 3.

[12] Y. Yang, W. Li, T. A. Gulliver, and S. Li, "Bayesian deep learning-based probabilistic load forecasting in smart grids," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4703–4713, Jul. 2020.

[13] K. Mahmud, J. Ravishankar, M. J. Hossain, and Z. Y. Dong, "The impact of prediction errors in the domestic peak power demand management," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4567–4579, Jul. 2020.

[14] A. Alimardani, L. Hadjidemetriou, and E. Kyriakides, "Distribution system state estimation based on nonsynchronized smart meters," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2919–2928, Nov. 2015.

[15] A. Kotsonias, M. Asprou, L. Hadjidemetriou, and E. Kyriakides, "State estimation for distribution grids with a single-point grounded neutral conductor," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 8167–8177, Oct. 2020.

[16] B. Bayer *et al.*, "The German experience with integrating photovoltaic systems into the low-voltage grids," *Renewable Energy*, vol. 119, pp. 129–141, Apr. 2018.

[17] D. Sidorov *et al.*, "A dynamic analysis of energy storage with renewable and diesel generation using volterra equations," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3451–3459, May 2020.

[18] "NobelGrid—EU horizon 2020 project 2015–2018," [Online]. Available: https://nobelgrid.eu

[19] H. Çimen, N. Çetinkaya, J. C. Vasquez, and J. M. Guerrero, "A microgrid energy management system based on non-intrusive load monitoring via multitask learning," *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 977–987, Mar. 2021.

[20] Sanduleac *et al.*, "Next generation real-time smart meters for ICT based assessment of grid data inconsistencies," *Energies*, vol. 10, no. 7, Jun. 2017, Art. no. 857.

[21] M. Albu, M. Sănduleac, and C. Stănescu, "Syncretic use of smart meters for power quality monitoring in emerging networks," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 485–492, Jan. 2017.

[22] Y. Ji, E. Buechler, and R. Rajagopal, "Data-driven load modeling and forecasting of residential appliances," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2652–2661, May 2020.

[23] V. Vignesh, S. Chakrabarti, and S. C. Srivastava,"Power system load modelling under large and small disturbances using phasor measurement units data," *IET Gener. Transmiss. Distrib.*, vol. 9, no. 12, pp. 1316–1323, 2015.

[24] M. M. Hasan and H. T. Mouftah, "Cloud-centric collaborative security service placement for advanced metering infrastructures," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1339–1348, Mar. 2019.

[25] R. Godina *et al.*, "Effect of loads and other key factors on oil-transformer ageing: Sustainability benefits and challenges," *Energies*, vol. 8, no. 10, pp. 12147–12186, 2015.

[26] *IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers*, IEEE Standard C57.104-2019, 2019.

[27] "Pecan street data port." Accessed: Jul. 10, 2020. [Online]. Available: www.pecanstreet.org/dataport

[28] "OpenDSS." Accessed: Jul. 10, 2020. [Online]. Available: http://smartgrid.epri.com/SimulationTool.aspx

[29] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, 2002, Art. no. 174102.

Mihai Sănduleac (Member, IEEE) received the M.Sc. degree in electrical engineering and the Ph.D. degree in expert systems for power networks from the Polytechnic Institute of Bucharest, Bucharest, Romania, in 1985 and 1998, respectively.

He was with the Energy Research and Modernizing Institute ICEMENERG Bucharest in various positions, also is the Head of Research with Power Systems Laboratory. He had various activities in the electrical energy sector and is currently an Associate Professor with the Department of Electrical Power Systems, University Politehnica of Bucharest, Bucharest. He acted as a Technological Manager in H2020 Nobel Grid and Storage4Grid projects, while being involved in several other European projects, such as SUCCESS, WISEGRID, and WEDISTRICT. He has authored/coauthored various papers and coauthored books in the energy field.

Dr. Sănduleac is Member of CIGRE and the recipient of the 2011 IEEE-PES Romania Chapter Outstanding Engineer Award.

Irina Ciornei (Senior Member, IEEE) received the M.Sc. degree in energy and environment from the Technical University of Iasi, Iasi, Romania, in 2003, and the Ph.D. degree in electrical engineering from the University of Cyprus, Nicosia, Cyprus, in 2012.

She is currently a Research Associate with the KIOS CoE, University of Cyprus, working on various European-funded projects related to smart power grids (e.g., H2020-Smart5Grids, Electron, and CybPhys) serving as task or work package leader. She has authored or coauthored more than 30 articles in highly prestigious scientific journals and international conferences. Her research interests include data-driven analytics, modeling, operation, and control of cyber-physical systems including smart grids and microgrids, power quality and quality of electricity supply, optimization methods based on artificial intelligence, and analytical methods.

Dr. Ciornei was the recipient of a prestigious Marie Curie Individual Fellowship (DCNextEve), 2016, hosted by University Politehnica of Bucharest with a subject on dc microgrids (design, operation, and control). She often volunteers as a reviewer for a series of the *IEEE Transactions and Conferences*.

Lucian Toma (Senior Member, IEEE) received the B.Sc. and Ph.D. degrees in electrical power engineering from the University "Politehnica" of Bucharest, Bucharest, Romania, in 2002 and 2010, respectively.

He is currently an Associate Professor with the Department of Electrical Power Systems, University "Politehnica" of Bucharest. He has coauthored five books, and more than 90 papers published in journals and conference proceedings. He was engaged in ten international (e.g., H2020-RESERVE, H2020-Storage4Grid, and H2020-NobelGrid) and 20 national research projects. His research interests include power system dynamics, power electronic applications in power systems, smart grids, and electricity markets.

Dr. Toma has been the Chair of the IEEE Romania Section and the Chair of the IEEE PES Romania Chapter, since 2020. He was the recipient of the 2017 anniversary award "Remus Rădult" presented by the Romanian Academy of Technical Sciences, the 2013 year award presented by AGIR, and IEEE Outstanding Engineer Awards presented by IEEE PES Romania Chapter in 2016.

Radu Plămnescu (Student Member, IEEE) received the M.Sc. degree in power engineering in 2018 from the University Politehnica of Bucharest, Bucharest, Romania, where he is currently working toward the Ph.D. degree in synchronized measurements and high-reporting rate PQ equipment to develop new protection and control algorithms for microgrids.

He is currently an Assistant Professor with UPB and a Research Assistant with MicroDERLab Group, UPB. He also spent a leave in 2019 at Colorado State University as a Fulbright Visiting Scholar. His research interests include high reporting rate voltage and current measurements.

Mr. Plămnescu is part of the International Committee of Microgrid Symposium Series as a recognition for the work in the microgrid field.

Ana-Maria Dumitrescu (Member, IEEE) received the engineering, M.Sc., and Ph.D. in electrical engineering from the University Politehnica of Bucharest, Bucharest, Romania, in 2003, 2005, and 2012, respectively.

She is currently an Associated Professor with the University Politehnica of Bucharest. She has worked on national and international research projects and has coauthored papers in the electrical engineering field. She is currently working on deriving high-resolution load curves for residential consumers and on integration of cybersecurity tools dedicated to EPES. Her research interests include control strategies, electrical drives, power quality, microgrids, smart grids and smart cities, electrical measurements, and embedded systems.

Mihaela M. Albu (Senior Member, IEEE) received the Ph.D. degree in electrical power systems from the Department of Measurements, Electrical Engineering Faculty, University Politehnica of Bucharest (UPB), Bucharest, Romania, in 1998.

She is currently a Professor with the Department of Measurements, Electrical Engineering Faculty, University Politehnica of Bucharest (UPB), Bucharest, Romania. She is the founder of MicroDERLab group at UPB and coordinates several research teams working in projects funded by national and international grants, with results published in more than 100 papers in journals and conference proceedings. She was on leave with Arizona State University as a Fulbright Fellow in 2002–2003 and in 2010. Her research interests include synchronized measurements, smart metering technologies, power quality, and their applications to microgrids.

Prof. Albu has been a Distinguished Lecturer of the IEEE Instrumentation and Measurement Society, since 2016. She is a Member in several (IEC and IEEE) standardization groups and CIGRE committees.