Risk-Aware Operating Regions for PV-Rich Distribution Networks Considering Irradiance Variability

Duque, Edgar Mauricio Salazar; Giraldo, Juan S.; Vergara, Pedro P.; Nguyen, Phuong H.; van der Molen, Anne; Slootweg, J. G.

DOI
10.1109/TSTE.2023.3281890

Publication date
2023

Document Version
Final published version

Published in
IEEE Transactions on Sustainable Energy

Risk-Aware Operating Regions for PV-Rich Distribution Networks Considering Irradiance Variability

Edgar Mauricio Salazar Duque, *Member, IEEE*, Juan S. Giraldo, *Member, IEEE*, Pedro P. Vergara, *Member, IEEE*, Phuong H. Nguyen, *Member, IEEE*, Anne van der Molen, *Member, IEEE*, and J. G. Slootweg, *Senior Member, IEEE*

***Abstract*—This article proposes a framework to identify, visualize, and quantify risk of potential over/under voltage due to annual energy consumption and PV generation growth. The stochastic modeling considers the following: (i) Active and reactive power profiles for distribution transformers, dependent on annual energy consumption and activity in the serviced areas. (ii) Variable solar irradiance profiles that allow a broader range of PV generation scenarios for sunny, overcast, and cloudy days. The proposed framework uses multivariate-t copulas to model temporal correlations between random variables to generate synthetic scenarios. A probabilistic power flow is computed using the generated scenarios to define critical static operating regions. Results show that classical approaches may underestimate the maximum PV capacity of distribution networks when local irradiance conditions are not considered. Moreover, it is found that including annual energy consumption growth is critical to establishing realistic PV installation capacity limits. Finally, a sensitivity analysis shows that taking a 5% of overvoltage risk could increase up to 15% of the PV installed capacity limits.**

***Index Terms*—Multivariate copulas, stochastic modeling, load modeling, irradiance modeling, hosting capacity.**

I. Introduction

EUROPEAN policies continue encouraging the adoption of low-carbon technologies at a rapid pace. Domestic customers have continuously adopted new low-carbon (LC) technologies driven by new subsidies and environmental consciousness. Such technologies can be commonly found in households as electric heat pumps (EHPs) for thermal needs and adopting of electric vehicles (EVs) as a cleaner transport alternative. This implies that a rampant increase in electricity consumption is expected in the incoming years, increasing loading in the electrical distribution network.

Distributed energy resources (DERs), such as photovoltaic systems (PV), are becoming omnipresent in the distribution network. Solar irradiance, the energy source for PV generation, is intermittent by nature due to geographical location, temperature, and cloudiness conditions. These characteristics make PV generation an irregular and highly variable source of energy. Load growth and intermittency in the DERs have imposed new challenges to operating distribution networks within technical limits [1]. Increased PV installed capacity may bring operational challenges, such as voltage quality problems [2], [3], reverse load flows [4], and load unbalance [5], among others.

Distribution network operators (DNOs) must develop tools to help to assess and quantify the impacts and risks of violating technical limits considering the uncertainties. Specifically, the framework introduced in this article aims to assess the under/overvoltage and overloading conditions for a medium voltage (MV) distribution network. The framework allows DNOs to make a risk-informed decision about the technical violations before committing to a network enhancement or expansion. *Two* main topics of interest for the DNOs are considered and evaluated in this work: The technical impacts of local PV installed capacity growth (under stochastic power generation conditions) and time-varying loads considering the increase in annual energy consumption.

The data used in this article is commonly available for modern DNOs, i.e., historical load consumption, meteorological data, and distribution network topology. Each topic covered in this work has extensive literature, and it is a field of research on its own. The following two subsections present the state-of-the-art with relevant examples for each topic, followed by our contribution, which holistically combines new approaches.

Manuscript received 19 September 2022; revised 6 January 2023 and 16 April 2023; accepted 22 May 2023. Date of publication 1 June 2023; date of current version 20 September 2023. Paper no. TSTE-00965-2022. *(Corresponding author: Edgar Mauricio Salazar Duque.)*

Edgar Mauricio Salazar Duque and Phuong H. Nguyen are with the Electrical Energy Systems (EES) Group, Eindhoven University of Technology, 5612 AE Eindhoven, Netherlands (e-mail: e.m.salazar.duque@tue.nl; p.nguyen.hong@tue.nl).

Juan S. Giraldo is with the Energy Transition Studies Group, Netherlands Organization for Applied Scientific Research (TNO), 2595 DA Amsterdam, Netherlands (e-mail: juan.giraldo@tno.nl).

Pedro P. Vergara is with the Intelligent Electrical Power Grids (IEPG) group, Delft University of Technology, 2628 CD Delft, Netherlands (e-mail: p.p.vergarabarrios@tudelft.nl).

Anne van der Molen is with the Eindhoven University of Technology, 5612 AZ Eindhoven, Netherlands, and also with Stedin, 3011 TA Rotterdam, Netherlands (e-mail: a.e.v.d.molen@tue.nl).

J. G. Slootweg is with the Eindhoven University of Technology, 5612 AZ Eindhoven, Netherlands, and also with Enexis, 5223 MB 's-Hertogenbosch, Netherlands (e-mail: j.g.slootweg@tue.nl).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSTE.2023.3281890.

Digital Object Identifier 10.1109/TSTE.2023.3281890

A. Related Works

1) Quantifying Impacts From PV Installations: Maximizing the amount of PV installed capacity in the distribution network is a priority for DNOs. PV hosting capacity (PVHC) is the maximum allowable PV capacity in the distribution network without violating its technical limits [6]. Multiple techniques have been proposed in the literature to quantify its value and operational issues [7]. The methods described in the literature can be assessed and categorized into three groups:

i) *Deterministic* methods such as [8] analyze maximum PVHC by doing a value sweep of predefined values of system parameters, e.g., network data, PV capacities, and installation locations. This approach is practical, fast, and helped to determine practical limits to quantify hosting capacities in different countries, as summarized in [9]. Nevertheless, ignoring intrinsic characteristics of uncertain variables such as PV location, panel sizes, load, and irradiance variability results in PVHC values that can be over or underestimated.

ii) *Stochastic* methods are the most common in the literature. The work in [10] assigns random PV installation capacities at random connection points in the low voltage distribution network. The mathematical models to determine the PVHC are designed to be run as probabilistic power flow (PPF), in which PVHC limits are based on technical violation counts. Nevertheless, these approaches do not consider the uncertainty of temporal correlation and the coincidence of PV generation and customer power consumption. Additionally, load profiles are often simulated with representative load consumption profiles, which are standardized and deterministic. Therefore, stochastic approaches also tend to underestimate PVHC. Observations in [10] note that an increase in the loading increases the PVHC, and the matter should be explored in order to quantify the increased PVHC margin. This article includes the annual energy consumption growth, which increases the values of the load profiles, changing the maximum PV installed capacity estimation.

iii) *Time-series* methods are more precise in identifying and quantifying technical violations but are computationally demanding. Such techniques are used to study the effectiveness of multiple controller types, e.g., tap changers, capacitor banks, and solar panel inverters [11] which can enhance PVHC. The current work on those methods does not incorporate load profile changes due to increased load growth or different irradiance scenarios, e.g., cloudiness conditions. The reason is that simulations for different high-resolution scenarios, e.g., 1-sec, 1-min, can become computationally challenging or infeasible [12]. Nevertheless, the work in [13] highlights the need for time-series simulation to *allow realistic concurrency between generation and production*, which is more realistic for effectively analyzing control options that enhance PVHC. The focus of our framework is not to speed up or analyze the impacts of different PVHC enhancing technologies. However, it is built upon the fact that time series simulations bring an objective and accurate perspective of the potential distribution network problems, which limits the amount of PV installed capacity.

2) Probabilistic Load Modeling and Electricity Consumption Growth: A probabilistic time-series load modeling requires consumer characterization to represent the different activities that are serviced by distribution networks, e.g., residential, commercial, and office areas. Probabilistic clustering techniques can simultaneously find customer groups while building a probabilistic model. The probabilistic clustering models can be used as generative models to simulate scenarios for a PPF formulation. The most relevant techniques which use probabilistic models on electricity consumption data are Gaussian mixture models (GMM) [14], Dirichlet process mixture models [15], or more flexible techniques such as C-vine copula mixture models [16]. The modeling methods mentioned above accurately capture each consumer group's statistical properties. However, modifying the generative model to simulate profiles corresponding to a specific annual energy consumption value is not straightforward, not allowing simulations that include load growth. Nonetheless, the recent work in [17], which uses a multivariate-t (MVT) copula, proposed a flexible generative model to simulate profiles conditioned to annual energy consumption. In this article, we extend the work of [17] with clustering techniques to have a flexible tool to model time series load consumption for different serviced areas activities, taking into account different annual energy consumption values to simulate load growth.

3) Probabilistic Solar Generation Modeling: A common practice used in PPF studies is to loosely model the variability of PV generation with a normal or beta distribution [18], which does not characterize realistic variability on solar irradiance. Characterization and modeling solar energy intermittency are critical to quantify realistic reliability indexes for the future distribution network [19]. Intermittency levels of solar irradiance data can be measured using variability indexes metrics (VIMs), which groups days based on cloudiness conditions, e.g., clear sky, cloudy, and overcast [20]. The VIMs can be used as a parameter to classify solar irradiance data in order to be used in simulations [21]. However, determining the VIMs parameter thresholds for the type of day classifications can be vague [22].

Aside from using VIMs for irradiance variability characterization and grouping, probabilistic clustering techniques can also be applied to recognize and build probabilistic models for different daily irradiance groups [23], [24], [25], [26], [27]. Nevertheless, the generative models from the clustering can not be used in the same way as for load profiles. The samples will not represent the natural physical irradiance process because solar irradiance has different sunrise/sunset times and intensity parameters due to geographical location, making the modeling challenging.

Additionally, previous studies in generative modeling for irradiance profiles also use Markov models [28], [29] or Dirichlet distribution [30] as flexible probability distribution solutions to capture the wide uncertainties created by the cloudiness conditions. Nevertheless, those approaches lack the flexibility to control the generation of profiles with specific cloudiness scenarios, i.e., different percentages of sunny, cloudy or overcast days.

The irradiance phenomena' high variance and non-linear properties require flexible mathematical probability techniques

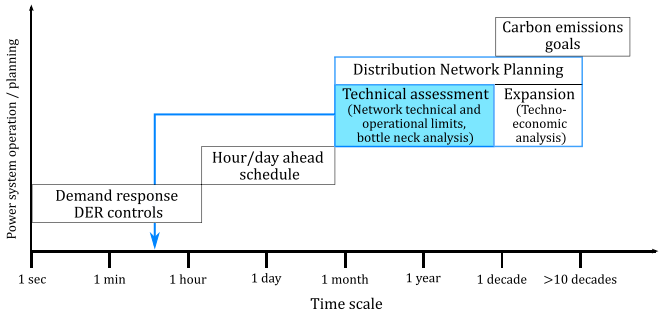

Fig. 1. Framework is designed for the technical assessment step (blue rectangle) to identify networks that require a further expansion study (techno-economic analysis) for future topology changes or grid enhancements.

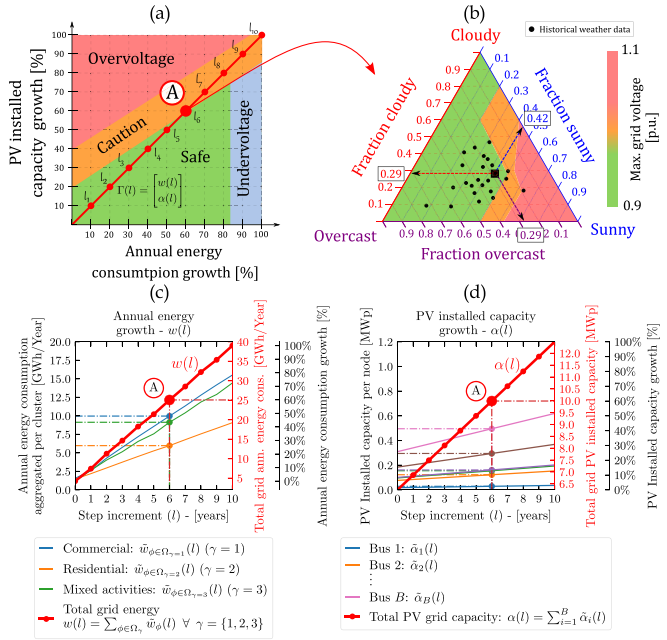

Fig. 2. Visualization of static regions for the voltage technical limits in the network. (a) Four different areas are determined by the distribution network's annual energy consumption and PV installed capacity. (b) Further inspection on the caution regions is done with the ternary plot. Caution region is influenced by the type of irradiance days for the period of study i.e., percentage of sunny, cloudy, and overcast days. Contour colors guide to visualize the maximum voltage over the network. Subplot (c) shows an example of total grid annual energy consumption growth $w(l)$ for ten years, which is composed of the sum of the different area activities growth functions $\tilde{w}(l)$. Total PV installed capacity growth curve $\alpha(l)$ is shown in (d), including PV growth functions on each node in the grid $\tilde{\alpha}(l)$. The x-axis of (a) is the *total* grid annual energy consumption in percentage values shown in the secondary axis in subplot (c). Similarly, the y-axis of subplot (a) is the *total* grid PV installed capacity in percentage values shown as a secondary axis in (d). Simulations with total grid annual energy consumption and PV installed capacity functions define the colors in the static regions in (a). The linear parametric path $\Gamma(l)=[w(l),\alpha(l)]^\intercal$ (red line) in subplot (a) is the result of following the total growth of (c) and (d). Point **A**, highlighted by the white circle, is an example of where is located a specific single step of the growth curves (c) and (d); in the growth path $\Gamma(l)$ in (a). Here, we use year six (l_6). i.e., $\mathrm{A}=\Gamma(l_6)=[w(l_6),\alpha(l_6)]^\intercal$.

to model such patterns. To obtain such flexibility, this article proposes a new generative model for irradiance profiles based on a mixture of MVT copulas, which has the following properties: (i) It follows sunrise and sunset times. (ii) The solar irradiation is upper bounded by the maximum possible physical irradiance value based on the geographical location. (iii) It can be flexible in selecting the number of days with different cloudiness conditions, i.e., clear sky, cloudy, and overcast. (iv) It follows natural variability behaviors measured by VIMs. Modeling different cloudiness (hence intermittency) scenarios allows us to precisely quantify the risks of overvoltage events and the times when the distribution transformer/lines can be overloaded.

Analysis of the impact of the PV generation considering load, PV growth, and irradiance variability at the same time is challenging not only for the modeling but also from the visualization perspective. In this work, we propose a framework that is used in the early stages of the distribution network planning (highlighted in blue in Fig. 1) with the objective is to identify early the distribution networks prone to have a technical problems based on a prognosis of load, PV growth, and irradiance variability conditions in the serviced area. Then, the identified networks require further techno-economical analysis for grid enhancement/expansion, which is out of the scope of this research.

It should be emphasized that even though the framework is for an initial phase of distribution network planning, the simulations are conducted in a 15 min resolution time (blue arrow in Fig. 1) to respect the concurrency between the PV generation and consumption to provide *accurate* and quantifiable technical violation results.

The concepts of static operating regions or static secure regions can be found in the literature since the '80s [31], and they have evolved for different applications and time frames of power system operations; for dynamic and voltage security assessment in transmission systems [32] or assessing the steady-state security in distribution networks with high penetration of distributed energy sources [33]. Unlike previous approaches, the work in [34] includes distributed energy sources uncertainties and correlations for risk assessment. Nevertheless, the cited works aim to find the secure regions for a short-time operational perspective (hours to days) applied to the distribution system base of N-1 security criterion and contingencies, and their focus is not intended in the mid-term operation on the distribution (considering load and PV installed capacity growth). The previous works use the term *secure regions* in the context of reliability analysis. Here we use the term *secure regions* to identify the safe PV and load growth values where the network is still within technical constraints, according to a hypothesis of load and PV installed capacity growth curves for a time horizon of years. Our framework also incorporates the novelty of simultaneously considering the local irradiance variability conditions, load, and PV growth. These are essential factors in quantifying the risk of overvoltage due to PV generation.

Nomograms are figures that express the safe states of operation in a geometrical form, and they are tools that are easy to use for operators to make a decision. Here, we introduce a simple nomogram and a ternary plot to interpret the results of the framework. Fig. 2 is the sketch of the proposed visualization. It helps to understand and quantify the changes in three components (load growth, PV growth, and irradiance variability), including generation, consumption, and irradiance uncertainties.

B. Our Contribution

The framework presented in this article shows a holistic solution to model the following uncertainties: active and reactive time-series consumption adjusted by annual energy consumption growth, and PV generation due to different cloudiness conditions (irradiance variability). The framework uses only one flexible multivariate probabilistic modeling technique: MVT copula. More precisely, the main contributions of this article are as follows:

- A framework is proposed to support DNOs' monitoring process for MV distribution networks to assess the under/overvoltage impacts of load and PV growth under stochastic conditions. The model exploits the MVT copula's flexibility to create scenarios that capture temporal correlation for PV generation, annual energy, and active and reactive power consumption, to generate accurate daily profiles for different area activities serviced by the distribution grid.
- A new irradiance model based on MVT copulas that simulates irradiance profiles for different cloud conditions, e.g., clear sky (sunny), cloudy and overcast. The model has the statistical properties of irradiance variability, sunrise and sunset times which is adaptable to any period of the year in any geographical position.
- The concept of critical static operating regions is introduced as a new representation and visualization tool used to understand and assess the results of the probabilistic load flow, for simulations that includes load, PV growth, and local irradiance variability conditions. Additionally, the proposed nomograms and ternary plots are used as versatile tools to help the DNO steer the load, and PV growth plans to maintain technical limits in the distribution network.

II. Proposed Methodology and Visualization

Our approach focuses on MV distribution networks, which service different activity areas, e.g., residential, commercial, and offices, via MV to low voltage (LV) distribution transformers. DNOs calculates the expected load growth, computed as annual energy consumption increase per year, and PV panel installed capacities for the different serviced areas based on techniques like [35]. It is also assumed that each area's PV installed capacity growth schedule does not overcome the PVHC on the LV side. Fig. 2(c) and 2(d) shows an example of total grid growth for annual energy consumption and PV installed capacity, respectively. In this example, which shows the total network annual energy consumption growth function $w(l)$ for ten years (red line Fig. 2(c)), which is the sum of the energy growth consumption functions for three serviced areas with different activities, denoted as commercial, residential and mixed activities, e.g., offices, mixture commercial and residential. Variable γ labels each type of area. It should be noted that the areas (which are groups of MV grid buses) could have different growth rate functions $\tilde{w}_\gamma(l)$. More precisely, the total grid annual energy growth function is defined as

$$w(l) = \sum_{\phi \in \Omega_\gamma} \tilde{w}_\phi(l) \quad \forall \quad \gamma = \{1, \ldots, \Xi\}, \tag{1}$$

where $\phi = \{1, \ldots, B\}$ is the node number of a grid with B nodes, Ω_γ represents the set of buses that belongs to the area activity γ, and Ξ is the total number of type of area activity, which in this particular example is three. The choice for grouping the grid nodes by areas is for daily profile modeling. For example, the daily load profile for a node with 3 [GWh/Year] energy consumption for a commercial area differs from a residential area, even though they have the same annual energy value. Separating the type of consumption by groups allows us to create specific types of daily profiles, which is critical to maintain the concurrency between load consumption and PV generation, to quantify the overvoltage risks accurately.

Similarly, the total grid PV installed capacity growth function $\alpha(l)$ is the sum of the grid PV installed capacity functions per node $\tilde{\alpha}(l)$, which is expressed as

$$\alpha(l) = \sum_{i=1}^{B} \tilde{\alpha}_i(l). \tag{2}$$

It is *critical* to notice that the percentages in the axes of Fig. 2(a) do not mean a homogeneous growth on the grid for total annual energy consumption or PV installed capacity because each area/node has its growth rate as defined by functions (1) and (2). The percentage denotes the values between the minimum and maximum for the total grid growth and total grid PV installed capacity. For instance, in Fig. 2(c), 0% means a total grid annual energy consumption of 4.6 [GWh/year], and 100% is 39.02 [GWh/year] (total grid annual energy consumption values are in the secondary axis highlighted in red). The same rationale applies to the percentages of PV installed capacity growth (Fig. 2(d)), which corresponds to the y-axis of Fig. 2(a).

A colored nomogram depicted in Fig. 2(a) is the proposed descriptive quantification of the maximum/minimum MV network voltage magnitude for all the different combinations of scenarios of *total network annual energy consumption* versus *total network PV installed capacity* growth (in percentage), which also includes cloudiness conditions (irradiance variability). The nomogram consists of four distinctive regions. (i) Safe region (green): the network will not have any voltage problems in the respective combination of PV installed capacity and annual energy consumption growth. (ii) Caution (orange): these combinations could potentially cause voltage magnitude problems in the network, which could be triggered by PV generation, and it is related to seasonal irradiance conditions, i.e., different types of day combinations of clear sky (sunny), cloudy, and overcast days. The caution frontier begins for voltages that are $\geq V_{\text{caution}}$. (iii) Over-voltage (red) depicts the combinations in which the network might have voltage magnitude violations, which also depends on the local irradiance variability conditions. The frontier begins for voltages that are $\geq \overline{V}$. (iv) The plot's under-voltage (blue) region corresponds to an overloaded network in which voltages magnitude fall below the lower technical limit, and the region starts for voltage magnitudes that are $\leq \underline{V}$.

Next, we analyze the path created by the load and PV installed capacity growth curves on the colored nomogram. The growth functions $w(l)$ and $\alpha(l)$ create a parametric curve [36] on the static regions nomogram defined as

$$\Gamma(l) = \begin{bmatrix} w(l) \\ \alpha(l) \end{bmatrix}, \tag{3}$$

where the values of $w(l)$ and $\alpha(l)$, in percentages, determine the curve's x and y coordinate values. We refer to the parametric curve as a *growth path* throughout the document. In our example, the growth path is shown as a straight line (no curvature) in Fig. 2(a). The nomogram is used to assess the year when the grid starts to have technical violations. In the figure's example, the grid enters a caution region at year six. The benefit of having the parametric representation of the growth path in the nomogram is that the DNO can analyze other potential paths that stay within technical limits. This topic is discussed in the section VI-F.

The static regions are affected by irradiance characteristics of the particular geographical area in which the MV distribution network is located, and also by the level of risk that DNOs is willing to take on the network grid operation (more details in Section VI-E). Thus, depending on each case, the regions can be further inspected using ternary plots, as shown in Fig. 2(b) for the point highlighted with letter A. The ternary plot relates the type of day's fraction for each month in a season, and the fractions sum up to 100%. For instance, one particular summer month in the Netherlands might have fractions of 42% clear sky (sunny), 29% cloudy, and 29% overcast days. The fractions of days in this particular example is marked by a black square on the ternary plot in Fig. 2(b). The colored contour regions in the ternary plot determine if that month does not have an overvoltage violation for the combination of total annual energy consumption and PV installed capacity. A ternary plot can be constructed for *each* combination of PV installed capacity and annual energy consumption growth from the nomogram.

The data used to generate the plots described in Fig. 2 comes from extensive scenario simulations using a PPF formulation. Fig. 3 gives an overview of the framework, describing the data processing and modeling needed to generate the scenarios for the simulations. The following sections explain the approach in detail.

III. Distribution Transformer Load Modeling

This section describes the probabilistic modeling for load profiles given a historical data set. Here, a load profile refers to an MV/LV distribution transformer loading, which is connected to an MV-distribution network.

A. Areas Characterization by Clustering

Here we use the active power profile of the distribution transformer as the data that represents the activity for each serviced area. Let the matrix $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N] \in \mathbb{R}^{M \times N}$ be the distribution transformer active power loading. M is the total number of data points recorded during a determined period, and N is the number of distribution transformers. Each vector

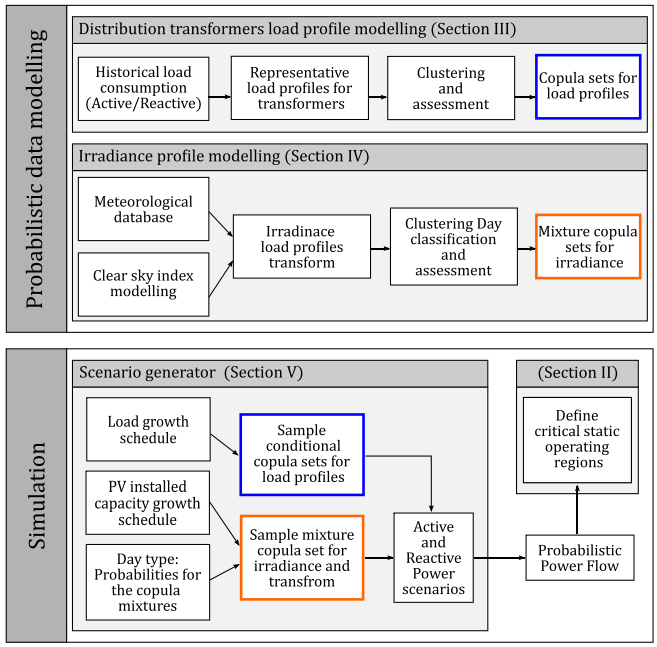

Fig. 3. Proposed approach for the study of MV distribution network, under stochastic conditions.

$\boldsymbol{p}_n = [p_{n,1}, \ldots, p_{n,M}]^\intercal \in \mathbb{R}^M$ is the data of the n-th distribution transformer for $n = \{1, \ldots, N\}$. The daily profile for the distribution transformer is characterized by averaging every T data points in $\boldsymbol{p}_n$, creating a representative load profile (RLP) vector $\bar{\boldsymbol{p}}_{\boldsymbol{n}} = [\bar{p}_{n,1}, \ldots, \bar{p}_{n,T}] \in \mathbb{R}^T$. i.e., for a daily profile at 15-min resolution $T = 96$. An RLP matrix is constructed as $\bar{\boldsymbol{P}} = [\bar{\boldsymbol{p}}_1, \ldots, \bar{\boldsymbol{p}}_N] \in \mathbb{R}^{T \times N}$. A clustering algorithm determines the number of clusters Ξ that best describe the data set in $\bar{\boldsymbol{P}}$.

The distribution transformer clustering is an unsupervised learning problem without ground truth, i.e., there is no previous labeling to know the correct number of clusters (Ξ) in advance. Therefore, cluster validation is done using internal evaluation metrics to guide the correct number of clusters. Such indices are the modified Dunn index (MDI), Davies-Bouldin index (DBI), Caliński-Habarasz index (CHI), and Silhouette index (SI). The clustering result assigns a label to each of the N-RLPs. i.e., distribution transformers, with a number from the label set $\gamma = \{1, \ldots, \Xi\}$. Thus, we define the set of transformers $\Phi\gamma$ that belongs to cluster γ.

B. Probabilistic Load Modeling by Multivariate-t Copulas

The data set for each transformer n is characterized by a tuple of data $\mathcal{M}_n = \{(\boldsymbol{p}_i, \boldsymbol{q}_i, \boldsymbol{w}_i)\}_{i=1}^D$, where vectors: $\boldsymbol{p} \in \mathbb{R}^T$ is the active power, $\boldsymbol{q} \in \mathbb{R}^T$ reactive power, and $w \in \mathbb{R}$ is annual energy consumption. The data set of the transformer is for D days. The data set for the *cluster* γ is $\mathcal{D}_\gamma = \{\mathcal{M}_{n \in \Phi\gamma}\}$. We use the following multivariate random variable expression for the load profile modeling: $\boldsymbol{x} = (\boldsymbol{p}, \boldsymbol{q}, w) \in \mathbb{R}^d$ where $d = 2T + 1$. Using a multivariate copula, we create a multivariate cumulative distribution function for *each cluster* that models the non-linear dependency between random variables, i.e., time steps and

annual energy consumption. For clarity, we drop the γ cluster index. The copula model allows us to construct the multivariate probabilistic model based on the random variables' marginal distributions as

$$\begin{aligned} F(x_1, \dots, x_d) &= C(F_1(x_1), \dots, F_d(x_d)) \\ &= C(u_1, \dots, u_d). \end{aligned} \tag{4}$$

Here, we do not assume a parametric probability function for the marginals $F_j(x_j) \mapsto u_j \; \forall \, j = 1, \dots, d$. Therefore, the transformation to uniform values of the marginals is done using an empirical distribution function,

$$F_{\Pi_j}(x_j) = \frac{1}{D+1} \sum_{r=1}^{D} \mathbb{1}_{\{x_{j,r} \leq x_j\}} \; \forall \, x_j \in \mathcal{D}, \tag{5}$$

In this work, we approximate $C(\cdot)$ with a multivariate-t copula,

$$\begin{aligned} C^{\mathrm{t}}(u_1, \dots, u_d) &= T_d(T^{-1}(u_1; \nu), \dots, T^{-1}(u_d; \nu);\; (\boldsymbol{\Sigma}, \nu)) \\ &= T_d(z_1, \dots, z_d;\; \boldsymbol{\theta}). \end{aligned} \tag{6}$$

The term $T^{-1}(\cdot; \nu)$ is the inverse cumulative distribution function of the univariate t-distribution with ν degrees of freedom, $T_d(\cdot; (\Sigma, \nu))$ is the d-dimensional multivariate-t distribution with covariance matrix Σ, and $\boldsymbol{\theta} = (\Sigma, \nu)$. The procedure to fit the parameters on $C^{\mathrm{t}}(\cdot)$, denoted as $\hat{\boldsymbol{\theta}} = (\hat{\Sigma}, \hat{\nu})$ can be found in [17]. It should be noted that the parameter Σ in the MVT copula model describes the correlation between the random variables in the form of a matrix. In our case, it relates to the temporal correlation between the time steps of the load profile. The benefit of using an MVT copula is that it captures tail dependency, which means that the model can simulate extreme events of high/low active and reactive consumption.

The multivariate joint distribution (4) can be conditioned with respect to one variable using (6) as

$$F(\boldsymbol{x}_1 | x_2) = C(\boldsymbol{u}_1 | u_2) \approx C^{\mathrm{t}}(\boldsymbol{z}_1 | z_2;\; \hat{\boldsymbol{\theta}}_{1|2}), \tag{7}$$

In our application, the meaning of the variables is $\boldsymbol{x}_1 = (\boldsymbol{p}, \boldsymbol{q}) \in \mathbb{R}^{2T}$ and $\boldsymbol{x}_2 = w \in \mathbb{R}$, where w is the annual energy consumption of the MV distribution transformer. The expression (7) allows us to create a conditional probability distribution with a specific energy level $\hat{w}$, i.e., GWh per year, which allows us to simulate profiles for different annual energy consumption values.

The projection from the annual energy consumption value ($\hat{w}$) to the copula domain values are made by

$$\hat{z}_2 = T^{-1}(F_{\Pi_w}(\hat{w}); \hat{\nu}). \tag{8}$$

The value $\hat{z}_2$ is used to condition the MVT copula parameters $\hat{\boldsymbol{\theta}}_{1|2}$ (for more details, the reader is referred to the appendix on [17]). The conditioned copula in (7) can be sampled to get a vector $\hat{\boldsymbol{z}}_1 \sim C^{\mathrm{t}}(\boldsymbol{z}_1 | z_2 = \hat{z}_2;\; \hat{\boldsymbol{\theta}}_{1|2})$.

The sampled vector values are projected to power units, e.g., kW, as

$$\hat{x}_i = F^{-1}_{\Pi_i}(T(\hat{z}_{1,i}; \hat{\nu})) \; \forall \, i = 1, \dots, 2T. \tag{9}$$

The use of conditional copula modeling is used as a generative model to create the profiles of active and reactive power of the transformers based on a specific annual energy value, which maintains the non-linearity and statistical properties of the electrical load profile data set. The simulation of the profiles of active and reactive power for the *cluster* γ is summarized as

$$(\hat{\boldsymbol{p}}, \hat{\boldsymbol{q}})_\gamma \sim F^{\mathrm{pg}}_\gamma(\boldsymbol{p}, \boldsymbol{q} | w = \hat{w}; \hat{\boldsymbol{\theta}}_{(p,q)|\hat{w}}) = F_\gamma(\boldsymbol{x}_1 | x_2 = \hat{w}; \hat{\boldsymbol{\theta}}_{1|2}). \tag{10}$$

Sampling from the conditional MVT copula (10) results in a time-correlated profile for active and reactive power for an annual energy consumption $\hat{w}$. For instance, if the modeling is at 15 min resolution, sampling (10) generates a profile of $96 * 2 = 192$ time steps, the first 96 corresponding to active and the rest for reactive power.

IV. Irradiance Modeling

The challenge of creating a probabilistic model that generates irradiance profiles during the year is that the model should be able to handle the maximum irradiance according to geographical position, and it should also adhere to the sunrise and sunset times on different days of the year. In this section, we formulate the irradiance profile modeling of the framework shown in Fig. 3.

A. Generative Irradiance Profile Model

Let the matrix $\boldsymbol{G} = [\boldsymbol{g}, \dots, \boldsymbol{g}_D]$ be the data set describing the irradiance data for D days, with $\boldsymbol{g} \in \mathbb{R}^T$ describing the daily irradiance profile. The process of building a probabilistic model for the data set $\boldsymbol{G}$ is described in Fig. 4. This procedure can be summarized in three steps: (i) Irradiance normalization and common time frame transformation, (ii) Irradiance clustering and probabilistic modeling, and (iii) Irradiance profile generation by a copula mixture sampling.

The first step consists of setting a common time reference ($\hat{t}$) and normalization of the measured solar irradiance for the following two reasons: (i) It makes the clustering meaningful as it decouples the two sources of natural variation of solar irradiance (Fig. 4(a)), which are the different sunset/sunrise hours throughout the year, and different irradiance intensities at noon [25]. (ii) The advantage of having the same time reference is that it has a fixed number of time steps (random variables) where the MVT copula model can be created.

The normalization of the irradiance is done using a global horizontal irradiance (GHI) model, which estimates the amount of irradiance on the earth's surface at a specific geographical location. The GHI works as a theoretical upper bound of maximum possible radiation.

The time reference conversion is done by approximation with a linear interpolation function per day $g_{i,\hat{t}} = h_i(\hat{t}) \; \forall i \in \Omega_D = \{1, \dots, D\}$ and $\hat{t} \in \Omega_{\hat{t}} = \{1, \dots, \hat{T}\}$, with $\hat{T}$ as the maximum count of sunlight time steps for the year (summer solstice). Variable $g_{i,\hat{t}}$ is the irradiance value for the sunlight in the common time reference. Similarly, the irradiance of the GHI model should be converted to the common time reference in order to create the normalized profile named clear sky index

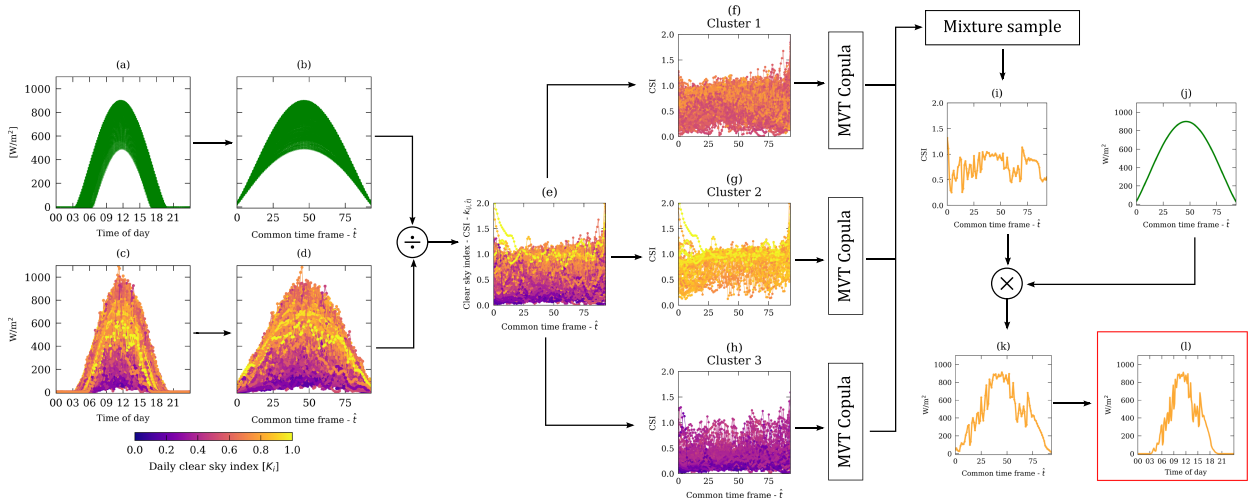

Fig. 4. Irradiance modeling approach. The global irradiance model (a) and the measured irradiance time series (c) are transformed into a common time frame as shown in (b) and (d). The computed clear sky index (CSI) shown in (e) is clustered into three groups: (f) Cloudy, (g) Overcast, and (h) Sunny. An MVT copula model is fitted for each cluster to generate CSI profiles. The models are sampled according to a mixture π_k, generating a CSI profile (i), which is transformed back to irradiance units by multiplying it with the model GHI profile (j) for one day of the year, creating the profile in (k). The synthetic irradiance profile is transformed from the common time domain to actual time units producing the final result in (l).

(CSI) profile, defined as

$$k_{(i,\hat{t})} = \frac{g_{(i,\hat{t})}}{\text{GHI}_{(i,\hat{t})}} \qquad \forall\, \hat{t} \in \Omega_{\hat{t}},\ i \in \Omega_D. \tag{11}$$

Additionally, the daily clear sky index (DCSI), named K_i, is defined as the ratio of the sum of values for daily irradiance and GHI. The DCSI is a metric to characterize the daily solar conditions used to label the type of cloudiness in the day.

The second step consists of clustering the CSI profiles in Fig 4(e) into three groups. The number of clusters is fixed to three to indicate the sunny, cloudy and overcast groups, and the label of the clusters are denoted by κ. The clustering methods and validation indices are the same already discussed in Section III-A. A visual inspection of the three resulting clusters, guided by the DCSI, determines the naming convention for sunny, cloudy, and overcast. High values of DCSI in the cluster indicate that the days have a clear sky, meaning a sunny day. On the other hand, lower values of DCSI mean overcast days, i.e., less irradiance. The remaining group is cloudy, characterized by a significant variation of DCSI values.

Each of the grouped CSI profile data sets is used to create an MVT copula model as outlined in Section III-A and named $F_\kappa^{\text{csi}}(k_{\hat{t}})$. The complete probabilistic model for the CSI profile set (Fig 4(e)) is the mixture of copulas

$$F_{\boldsymbol{\pi}}^{\text{irr}}(k_1, \ldots, k_{\hat{T}}) = \sum_{\kappa=1}^{3} \pi_\kappa F_\kappa^{\text{csi}}(k_1, \ldots, k_{\hat{T}}), \tag{12}$$

where $\sum_{\kappa=1}^{3} \pi_\kappa = 1$, and $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)$.

In the third step, the procedure to generate a CSI profile from the mixture copula model, i.e., $\boldsymbol{k} \sim F^{\text{irr}}(\cdot)$, as follows: (i) Generate a sequence of labels κ (cluster number) based on the multinomial distribution with three categories $\kappa \sim \text{Multi}(D, \boldsymbol{\pi})$ for D days. (ii) Sample $\boldsymbol{k} \sim F_\kappa^{\text{csi}}(\cdot)$ from the selected κ copula model, based on the label sequence created in the previous step. It is *critical* to notice that modifying the mixture of $\boldsymbol{\pi}$ controls the type of day that the model is generating. e.g., Mixture $\boldsymbol{\pi} = (0, 0, 1)$ generates profiles from cluster 3 all the time, or $\boldsymbol{\pi} = (1/3, 1/3, 1/3)$ samples equally on all types of days. Parameter $\boldsymbol{\pi}$ indicates the fraction of the type of days that are simulated to generate the plots in Fig. 2.

The sampled values are brought to irradiance units by doing the inverse process, using (11) to generate irradiance units, and using the interpolation function $f_i(\cdot)$, which is the inverse mapping of $h_i(\cdot)$. In that way, we reference the irradiance values to actual time. The simulation of the profiles or irradiance for a mixture $\boldsymbol{\pi}$ is summarized as

$$\begin{aligned} \hat{\boldsymbol{k}}_i &\sim F_{\boldsymbol{\pi}}^{\text{irr}}(\boldsymbol{k}) \\ \hat{\boldsymbol{g}}_i &= f_i(\hat{\boldsymbol{k}}_i \odot \text{GHI}_i) \qquad \forall\, i \in \Omega_D, \end{aligned} \tag{13}$$

where $\odot$ is the Hadamard product.

The main advantage of this model is that it follows precisely sunset and sunrise times. Also, it is upper bounded with the maximum solar irradiance, which depends on the geographical location of the distribution network under study. This makes the *concurrency between consumption and PV generation more accurate*. Additionally, controlling the copula mixture brings the versatility on which the DNOs can simulate different scenarios of cloudiness, i.e., irradiance variability conditions.

B. Variability Index Metrics for Irradiance Profiles

The generated irradiance profiles are validated using irradiance variability index metrics (VIMs). In this way, we ensure that the statistical properties of the synthetic irradiance profiles compared to the actual measurements are retained. Details of the VIMs metrics computation can be found in the appendix.

All VIMs are computed for every single profile generated by the model from section IV, creating a probability distribution per VIM. The probability distributions are compared against the original irradiance's VIMs probability distributions. The probability distance metrics used to compare are Kolmogorov–Smirnov (KS) and Wasserstein (WD) distances.

V. Scenario Generator and Simulation

An MV distribution network has a set of buses $\phi = \{1, \ldots, B\}$. Based on the historical consumption of each node, one can assign each bus to a cluster that belongs to the γ cluster via a supervised learning algorithm as in [16]. Alternatively, social demographic data of the area can be used to predict the cluster assignment for each node [37], which it can be more accurate using commercial datasets [38]. It is essential to notice that the cluster assignment to the node specifies which is the load profile model (10) is used for the PPF analysis.

The set of buses in the γ cluster is defined as $\Omega\gamma$. The matrix $\hat{\boldsymbol{W}} \in \mathbb{R}^{B\times L}$ defines all buses' annual energy consumption growth. Each matrix element is $\hat{w}_{(\phi,l)}$, where the index (ϕ, l) represents the row and column for the matrix. Index l is the one-step increment in the annual energy consumption. For instance, if bus three has a base annual energy consumption of 1.0 GWh per year, then $\hat{w}_{(\phi=3,l=0)} = 1.0$. If the same bus has a expected growth of 10%, then the one-step increment is $\hat{w}_{(\phi=3,l=1)} = 1.1 \cdot \hat{w}_{(\phi=3,l=0)}$, and so on. It should be recalled that the framework does not require that the annual energy consumption be linear or proportional through the network. Each node could have its own annual energy consumption growth function based on the step increments, i.e., $\hat{w}_{(\phi,l)}(l)$.

Similarly, matrix $\hat{\mathbf{A}} \in \mathbb{R}^{B\times \mathrm{PV}}$ with elements $\alpha_{(\phi,\mathrm{pv})}$ defines the PV installed capacity at each bus. Sub indexes (ϕ, pv) represent the row and column of $\hat{\mathbf{A}}$, and pv is the one-step increment in the PV installed capacity.

It is assumed that all PVs work at a unity power factor. The MV network is not excessively large, so all PV panels are exposed to approximately the same solar irradiance. Additionally, set Ω_Π is the set of mixture combinations of $\boldsymbol{\pi}$, which simulates the impact of different variable irradiance conditions. The active and reactive power used for *each node* in *cluster* γ for the PPF formulation is defined by

$$\hat{\boldsymbol{P}}^{\phi}_{(\hat{w}_{(\phi,l)},\mathrm{pv},\boldsymbol{\pi})} = \hat{\boldsymbol{P}}^{\gamma}_{\hat{w}_{(\phi,l)}} + \alpha_{(\phi,\mathrm{pv})}\hat{\boldsymbol{G}}_{\boldsymbol{\pi}} \tag{14}$$

$$\hat{\boldsymbol{Q}}^{\phi}_{(\hat{w}_{(\phi,l)},\mathrm{pv},\boldsymbol{\pi})} = \hat{\boldsymbol{Q}}^{\gamma}_{\hat{w}_{(\phi,l)}}$$

$$\forall \quad \phi \in \Omega_\gamma, \quad l \in \Omega_L, \quad \mathrm{pv} \in \Omega_{\mathrm{PV}}, \quad \boldsymbol{\pi} \in \Omega_\Pi. \tag{15}$$

Matrices $\hat{\boldsymbol{P}}^{\phi}$ and $\hat{\boldsymbol{Q}}^{\phi}$ are of size $\mathbb{R}^{\mathrm{MC}\times T}$, on which MC is the number of Monte Carlo simulations. Matrices $\hat{\boldsymbol{P}}^{\gamma}$ and $\hat{\boldsymbol{Q}}^{\gamma}$ come from sampling (10) MC times conditioned in annual energy value $\hat{w}_{\phi,l}$. Matrix $\hat{\boldsymbol{G}}_{\boldsymbol{\pi}}$ results from sampling (13) MC times using the mixture $\boldsymbol{\pi}$. This means that we have $|\Omega_L| \cdot |\Omega_{\mathrm{PV}}| \cdot |\Omega_\Pi| \cdot$ MC daily scenarios per node. The power flow technique used for the PPF is described in [39], which is highly robust and efficient for radial network systems.

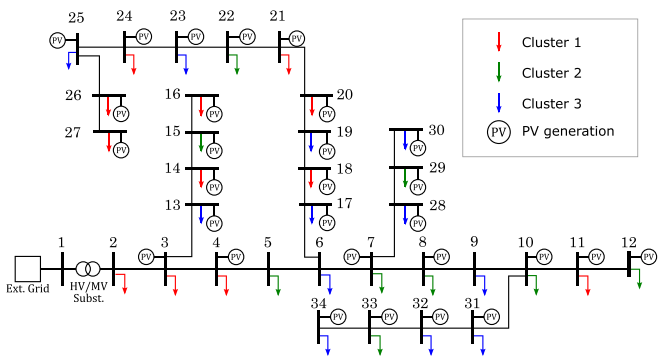

Fig. 5. MV distribution test system with distributed PV generation.

TABLE I
Load Profiles Clustering Metrics and Transformer Count for Three Clusters

	Clustering metrics					Transformer count per cluster		
Algorithm	**CHI**	**SI**	**DBI**	**MDI**	**RI**	**C1**	**C2**	**C3**
KMeans	**611.33**	**0.413**	0.847	0.828	1.000	166 (37%)	117 (26%)	160 (36%)
Spectral	515.00	0.384	**0.812**	**0.768**	0.633	144 (33%)	79 (18%)	220 (50%)
Kmedoids	609.98	0.412	0.848	0.840	0.980	119 (27%)	159 (36%)	165 (37%)
GMM	608.94	0.411	0.850	0.819	0.958	158 (36%)	170 (38%)	115 (26%)
HC(Ward)	486.77	0.361	0.887	1.141	0.454	165 (37%)	172 (39%)	106 (24%)
Birch	490.93	0.341	0.922	1.128	0.698	163 (37%)	174 (39%)	106 (24%)

VI. Case Study

The test case is based on an 11 kV MV-distribution network shown in Fig. 5 [40]. The substation voltage was set to 1.0 p.u. on the secondary side since it is a symmetrical set point and allows a fair assessment, as it allows significant amounts of PV installed that could surpass the technical limits $\overline{V} = 1.05$ p.u. and drops for peak loads below $\underline{V} = 0.95$ p.u. The limit to define the caution frontier is set to $V_{\mathrm{caution}} = 1.045$. In our case study, it was found that the value of $V_{\mathrm{caution}} = 1.045$ was a sensitive threshold to define the caution frontier, detecting scenarios that may lead to an overvoltage level. However, this specific threshold value is not mandatory for the functioning of the framework, and the only condition is that $V_{\mathrm{caution}} < \overline{V}$. The current limit of the transformer feeder is set to $I_{\mathrm{max}} = 230$ A. The loads in Fig. 5 are highlighted by color according to the clustering results in Section VI-A.

A. Clustering Validation Results

The data set of load profiles consists of 543 MV/LV distribution transformers from a municipality in the Netherlands for the summer season, at 15 min resolution. Six clustering techniques are used to group the load profiles as they have different algorithmic approaches to search for the best grouping. A cluster parameter sweep was done from two to fourteen clusters, as shown in Fig. 6. All techniques are evaluated on the indexes discussed in Section III-A. The CHI index indicates that three is the best number of clusters and it is also clear the different load patterns of each group (Fig. 7). A closer inspection of the metric values shown in Table I shows that K-Means and Spectral clustering are the top performers. The highlighted values in Table I are the best clustering metric values for SI, DBI, MDI (lower is better) and CHI (higher is better). However, K-means

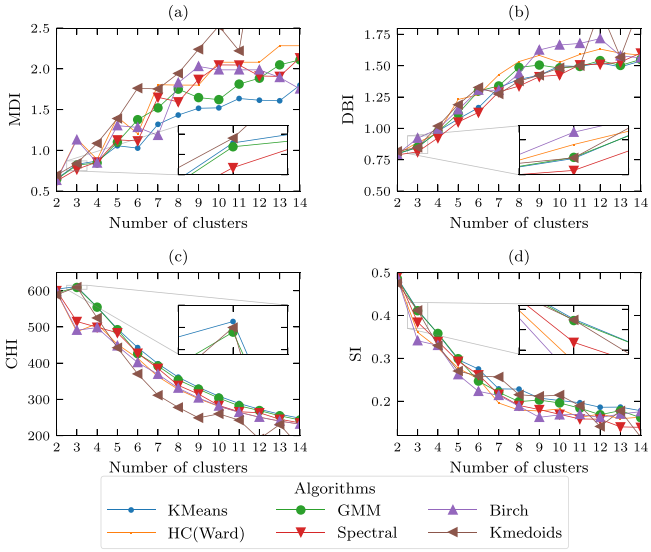

Fig. 6. Clustering validation metrics for different algorithms. For (a) MDI, (b) DBI, and (d) SI lower values are better, and for (c) CHI high values are better.

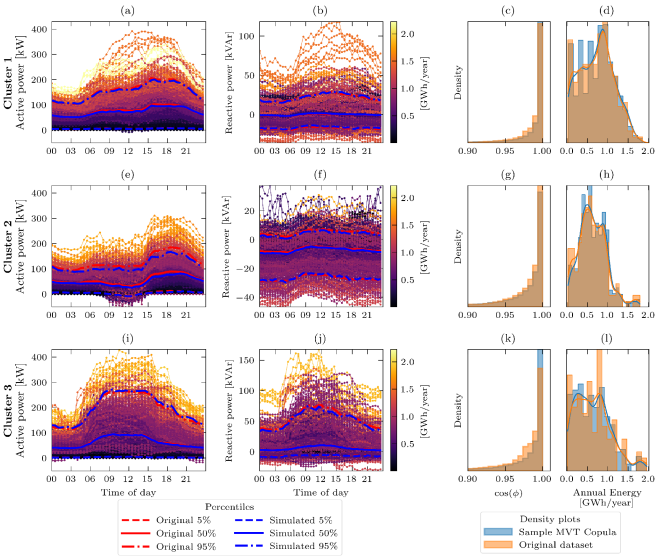

Fig. 7. Load profiles are clustered by the K-means algorithm. Each row corresponds to data of the same cluster. The first column is active power, and the second column is reactive power. The third column is the histogram of the power factor. The fourth column is the distribution of the annual energy consumption. The 95% confidence intervals of the original and simulated profiles are highlighted in columns one and two.

has more balanced groups compared to the Spectral algorithm. K-Means labels are used for load modeling.

B. Load Modeling Simulation Validation

Fig. 7 shows that clusters 1 and 2 are groups of residential consumption characteristics. Cluster 1 contains transformers with higher active power consumption values between 9:00 and 3:00 p.m. than cluster 2. Also, cluster 1 has the highest reactive power consumption among residential clusters. Cluster 3 corresponds to profiles of office/commercial areas, with higher and flatter power consumption between 7:00 a.m. and 5:00 p.m.

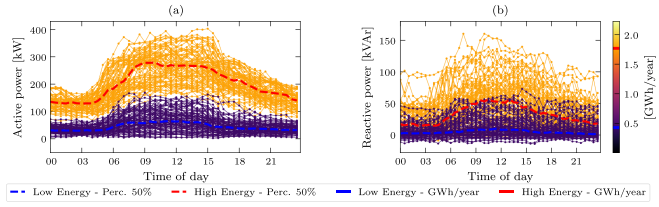

Fig. 8. Example of one hundred daily load profiles generated from sampling the conditioned MVT copula for cluster 3 for high and low annual energy consumption values. (a) Active power. (b) Reactive power.

The 95% percentiles between the original and simulated data sets using (4) show a high agreement between the profiles. It should also be noticed that the distribution of the power factor (Fig. 7(c,g,k)) and annual energy consumption (Fig. 7(d,h,l)) it is also maintained between clusters for the original data and simulated profiles.

An example of profiles generated using the conditioned copulas in (10) is presented in Fig. 8. One hundred profiles are generated for two different annual energy values. One for high energy consumption $\hat{w}_1 = 1.77$ [GWh/year], where the mean values of the profile for active and reactive power are highlighted by a red dotted line. The second set of simulated profiles has $\hat{w}_2 = 0.43$ [GWh/year] for the low energy consumption, and the means are highlighted with a blue dotted line. It should be noted that consumption values per time step does not follow a linear growth, i.e., increasing annual energy consumption ∼4 times does not necessarily increase all power values in the load profile by the same factor. This is a significant benefit of the copula as it captures the non-linear change of each time step according to the annual energy consumption. The accuracy of the load profile scenarios is evaluated using the Kolmogorov–Smirnov (KS) and Wasserstein (WD) distance metrics, following the research in [17].

C. Irradiance Modeling Validation

The irradiance data set comes from the closest meteorological station in the same municipality where the loading data was acquired. The resolution is 15-min. The Haurwitz solar irradiance model [41] is used as GHI. The Haurwitz irradiance model is referenced to the exact geographical coordinates of the pyranometer that acquired the irradiance data. The irradiance data set is transformed to CSI values using (11) as shown in Fig. 4(e). The groups for the CSI data set were partitioned by the K-Means algorithm based on the analysis of the clustering indexes. The groups are shown in Fig. 4(f-h). Clusters 1, 2, and 3 correspond to cloudy, sunny, and overcast days, respectively, according to the DSCI values. In order to justify the need for MVT copula as a generative model for the CSI, we compare its performance against two probability distributions to model $F_{\kappa}^{\text{csi}}(\cdot)$ in (12): (i) A multivariate Gaussian (MVG) and (ii) univariate Normal (UN) distributions, i.e., one Normal distribution per time step.

All three models (MVT copula, MVG, UN) are tested to check if they maintain the irradiance variability statistical properties of an actual irradiance profile, i.e., $\boldsymbol{Q}_{\boldsymbol{\pi}^*} \approx \hat{\boldsymbol{Q}}_{\boldsymbol{\pi}^*}$, where $\boldsymbol{\pi}^* = (0.44, 0.26, 0.3)$ is the original data set mixture. The VIMs

TABLE II
PROBABILITY DISTANCE METRICS FOR VARIABILITY INDEXES

		Cluster 1 (Cloudy)			Cluster 2 (Overcast)			Cluster 3 (Sunny)		
Index	Prob. dist.	MVG	MVT	Norm.	MVG	MVT	Norm.	MVG	MVT	Norm.
SDI	KS	0.407	**0.288**	0.636	0.508	**0.229**	0.698	0.620	**0.423**	0.521
	WD	0.033	**0.022**	0.054	0.032	**0.017**	0.053	0.048	**0.031**	0.039
VI	KS	0.420	**0.270**	0.702	0.338	**0.273**	0.776	0.466	**0.302**	0.573
	WD	0.778	**0.620**	1.583	0.623	**0.444**	1.713	0.907	**0.583**	1.508
MI	KS	0.443	**0.304**	0.869	0.323	**0.311**	0.896	0.399	**0.222**	0.772
	WD	0.035	**0.026**	0.101	0.024	**0.017**	0.096	0.032	**0.021**	0.087
FD	KS	0.421	**0.250**	0.732	**0.234**	0.326	0.814	0.270	**0.221**	0.645
	WD	0.137	**0.093**	0.276	**0.072**	0.088	0.266	0.067	**0.057**	0.179
CSI	KS	0.146	**0.061**	0.139	0.103	**0.064**	0.093	0.062	**0.044**	0.058
	WD	0.053	**0.035**	0.051	0.044	**0.028**	0.043	0.029	**0.025**	0.028

discussed in Section IV-B are computed for each of the irradiance profiles generated by the models (using $\boldsymbol{\pi}^*$) and the original data set. The difference between the probability distribution of each VIM from the generated profiles is measured against the original VIM distribution of the original profiles using the KS and WD probability distance metrics, as described in section IV. In order to have a significant statistical result, a bootstrapping method with 1000 repetitions is performed, and the means of the probability distance metrics are reported in Table II. The highlighted values in Table II are the lowest probability distance metrics for the three clusters (lower is better). It is shown that in almost all VIMs, the MVT copula has superior performance, especially in modeling sunny days, which are the critical profiles for the worst-case overvoltage scenario for high PV generation.

Fig. 9 shows the irradiance profiles for different $\boldsymbol{\pi}$ values for the irradiance mixture model (12). The first row of subplots shows the original data set, which has a mixture of $\boldsymbol{\pi}^*$. The second row of subplots shows data from sampling (13) with mixture $\boldsymbol{\pi}^*$ for the same amount of days as the original irradiance data set. In extreme scenarios, the subplots on the third-row show samples from a mixture that generates 100% of overcast days and the last row 100% percent of sunny days. The last column's heat maps show that irradiance profiles follow the sunrise and sunset times for all the seasons of the year, and the intensity of the heatmaps corresponds to the desired generative process based on the mixtures (bright yellow for clear sky and dark purple for low irradiance/overcast days).

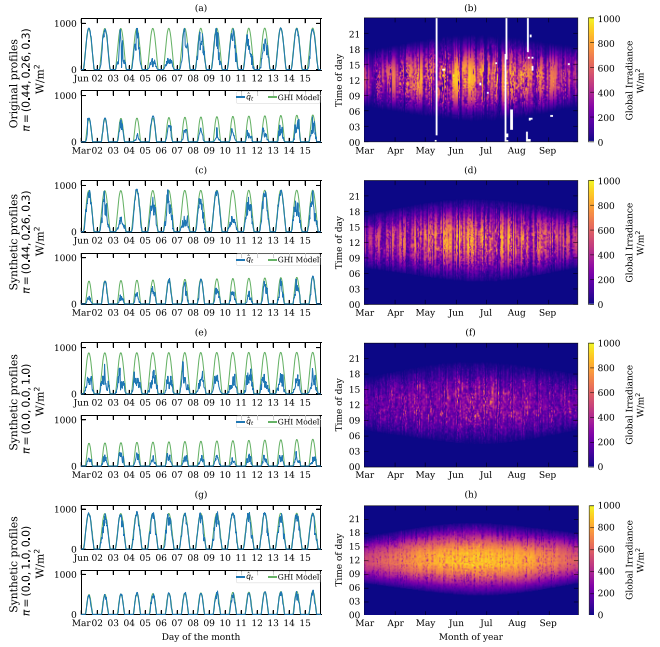

Fig. 9. Sampling example of the probabilistic irradiance model using the mixture of copulas following the procedure in Fig. 4. Subplots in the first row correspond to the meteorological station's measured irradiance data (original data set). The subplots in the second-row show synthetic generated profiles using the mixture values of the original irradiance data set, i.e., $(\pi_1 = 0.44, \pi_2 = 0.26, \pi_3 = 0.3)$. Being π_1 cloudy, π_2 sunny and π_3 overcast fractions. The third row of subplots are the generated profiles for a mixture that only produces overcast days, i.e., $(\pi_1 = 0.0, \pi_2 = 0.0, \pi_3 = 1.0)$. The last row of subplots shows profiles of sunny days the whole year, i.e., $(\pi_1 = 0.0, \pi_2 = 1.0, \pi_3 = 0.0)$. Image plots in the last column show that the generated irradiance profiles follow the year's corresponding sunrise and sunset times.

D. Estimating Voltage Magnitude Risk

The critical static operating regions were identified using a PPF formulation via Monte Carlo simulations. For simplicity, annual energy consumption for each area activity is linearly incremented as percentages above an initial base annual energy consumption per area. In the same way, the PV installed capacity growth is linearly increased for each bus in the distribution grid. Fig. 2(c) and 2(d) show the different annual energy consumption growth rates used for each cluster and PV installed capacities growth rates per node. The combinations of total annual energy consumption and PV growth are in steps of 10% from 0% to 100%, for a total of $|\Omega_{\mathrm{L}}| \cdot |\Omega_{\mathrm{PV}}| = 121$ combinations. Similarly, the mixtures of the types of days ($\boldsymbol{\pi}$) have a resolution of 0.1 per mixture component. More precisely, the set of mixture $\boldsymbol{\pi}$ is $\Omega_{\Pi} = \{(0.0, 0.0, 1.0), (0.0, 0.1, 0.9), (0.0, 0.2, 0.8), \ldots, (1.0, 0.0, 0.0)\}$. The count of allowable combinations, i.e., $\sum_{\kappa}^{3} \boldsymbol{\pi}_{\kappa} = 1$, is $|\Omega_{\Pi}| = 66$. The time step for time series simulations is at 15-min resolution (T=96), and a total of MC=1000 scenarios for all possible combinations are performed. This results in $\cong$ 766M of power flows.

The maximum voltage magnitude in the network for each annual energy consumption and PV installed capacity growth combination is analyzed with the heat map shown in Fig. 10. Each row's figures come from voltage magnitudes simulated from a specific mixture value of irradiance profiles. The first row subplots are for an instance of mixture values for cloudy/overcast days only. i.e., $(\pi_1 = 0.4, \pi_2 = 0.0, \pi_3 = 0.6)$ specified in the title of the heat map's subplot (Fig. 10(a)), the middle row is for a more typical summer mixture, and the bottom row has extreme values of 100% sunny days. It should be reminded that π_1, π_2, and π_3 correspond to cloudy, sunny, and overcast proportion mixture (indicating the fractions for each type of day). The color of the rectangles in each heat map refers to inset plots for four specific cases. The profiles in the insets represent the daily maximum (black) and minimum (grey) network voltage magnitude profiles for a specific percentile, which is the 90% percentile in this example (or 10% risk). The maximum peak voltage magnitude for the 90% percentile gives the *color intensity in the heat maps*.

The solid black contour line in the heat maps, first column subplots Fig. 10 (a,f,k), represents a frontier in which the

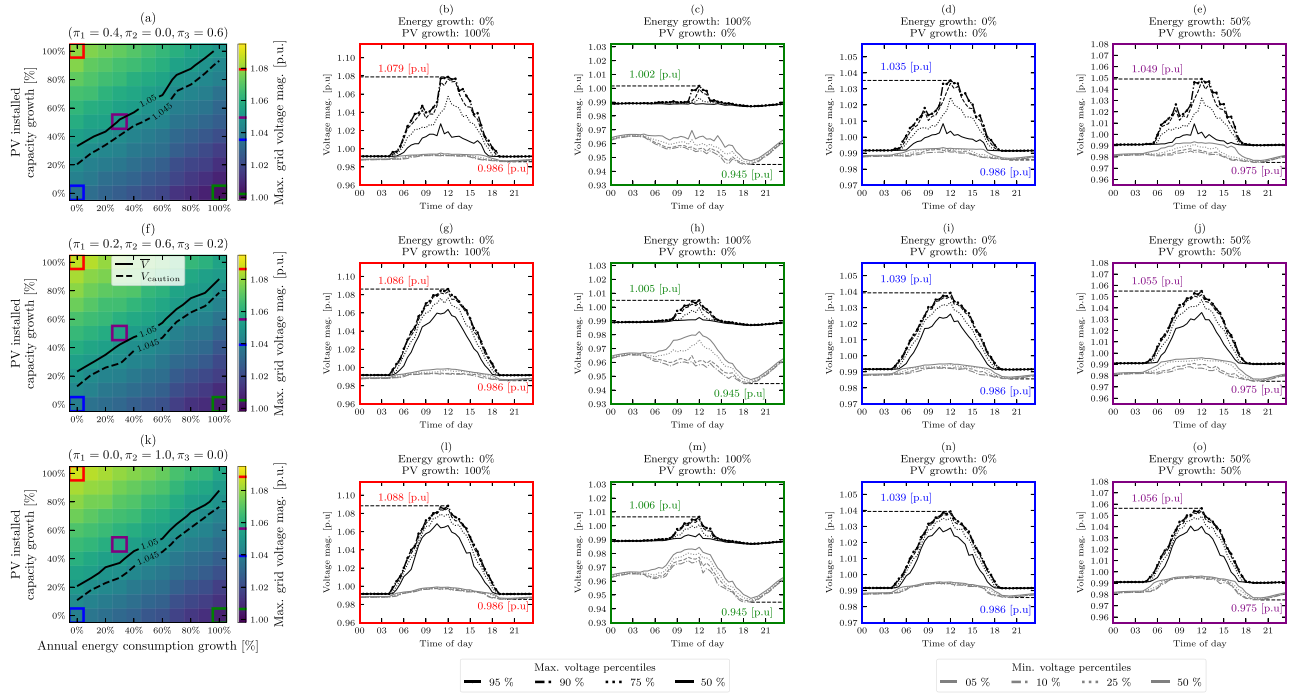

Fig. 10. Analysis of maximum and minimum network voltage magnitude. Heat maps subplots (a,f,k) correspond to 90% percentiles (10% risk) of maximum network voltage magnitude for different annual energy consumption and PV installed capacity growth combinations. Each of the rows shows different irradiance scenarios. First row of subplots (a)-(e) for predominant cloudy/dark days, second row (f)-(j) typical summer mixture, and third row (k)-(o) for 100% sunny irradiance scenarios. The inset rectangles colors on each heat map represent the row inset subplots by color. The inset subplots are the percentiles profiles for the network's maximum (black) and minimum (grey) voltage magnitude values. In the profiles subplots the maximum 90% and minimum 10% voltage magnitude percentiles are highlighted. The maximum peak voltage magnitude for the 90% percentiles gives the color intensity in the heat maps.

maximum 90% percentile of $\overline{V}$ is reached in the network. For instance, the purple rectangle of the upper left in Fig. 10(a), with 30% annual energy consumption growth and 50% PV installed capacity growth lies close of the contour line for voltage limit $\overline{V}$ (seen in the purple inset in the upper-right plot. i.e., Fig. 10(e) with a value of 1.049 [p.u.]). In this example, annual energy consumption and PV growth values below the contour line in the heat map represent safe combinations on which the voltage stays within the network's technical limits with a 90% probability. The same logic applies to the V_{caution} frontier shown in the heat maps as dotted lines.

It is critical to notice that the frontier is not a static value, and it depends on the irradiance conditions of the summer ($\overline{V}$ and V_{caution} contour lines move down in the heat maps). The frontier lowers when the fraction of sunny days increases (π_2 is higher) as higher irradiance scenarios increase the percentiles of maximum voltage profiles. The 90% percentile means a 10% risk of having a day of summer falling outside technical limits. Similar plots like Fig. 10 can be constructed for different levels of risk, e.g., 2% risk (98% percentile), 5% risk (95% percentile), and so on.

All the inset plots have the same lower network voltage magnitude for different irradiance mixtures (minimum voltage magnitude on the grey profiles in Fig. 10). Irradiance conditions do not influence the minimum voltage magnitude on the network. The only factor that affects its value is the increase in annual energy consumption growth. The green insets (Fig. 10(c), (h), and (m)) shows that there is an under voltage magnitude violation at the maximum annual energy consumption growth, and the critical value for the under voltage frontier $\underline{V}$ is found when the annual energy consumption growth is 88%. A similar analysis can be done for overloading in the HV/MV transformer.

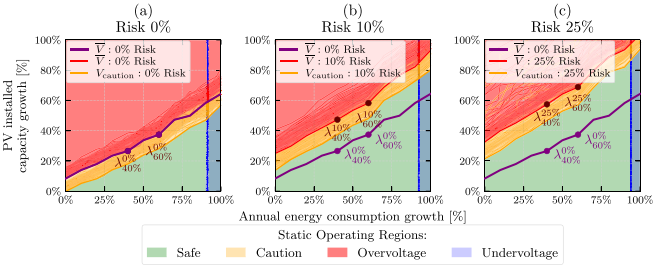

Fig. 11. Overlay of collection of critical frontiers generated by different irradiance conditions for different levels of risk: (a) 0% risk and (b) 10% risk and (c) 25% risk. Purple frontier is the same for all subplots and represents the 0% risk scenario.

E. Static Secure Regions

The collection of voltage contour lines (frontiers of $\overline{V}$) for *all irradiance conditions*, i.e., all $\boldsymbol{\pi}$ mixtures, are plotted in Fig. 11 with red lines that define the overvoltage region. Similarly, the collection of V_{caution} contour lines for all $\boldsymbol{\pi}$ mixtures is the lines in orange, which defines the caution region. There is a total of $2 \cdot |\Omega_{\Pi}|$ lines for each subplot, each set for $\overline{V}$ and V_{caution}. The purple line represents the worst and most stringent case

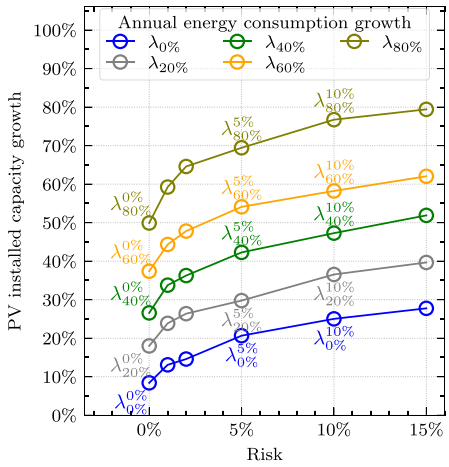

Fig. 12. PV installed capacity over-voltage frontier value versus risk. Color in the lines represents different annual energy consumption growth values. Superscript in λ means risk, subscript means annual energy consumption growth. e.g., $\lambda^{5\%}_{80\%}$ is the PV installed capacity growth over-voltage frontier value for 5% risk and 80% of annual energy consumption growth.

for an overvoltage violation, i.e., 100% sunny days with 0% risk of overvoltage (all simulated scenarios within the maximum voltage magnitude limits). Here, the safe region means that there is a chance of having an overvoltage given the *chosen level of risk*. The 0% risk scenario is the typical PV capacity value computed by deterministic approaches discussed in Section I- A, which is a conservative PV network capacity estimation.

Using our framework, we can now quantify how much PV capacity can be allowed to increase as a function of the allowed risk. The purple frontier highlighted in Fig. 11 are the same for reference (stringent case 0% risk). The lambda values are instances over the frontiers for their respective risks. For instance, $\lambda^{10\%}_{60\%}$ is the PV growth limit for a 60% annual energy consumption growth with a 10% risk for all sunny days simulation (first overvoltage frontier in red). The group of λ points for different cases is shown in Fig. 12. It is noticeable that there is a substantial increase in PV growth limits between $0-5\%$ of the risk. For instance, the PV limits increase from 26.5% to 42.3% for the 40% annual energy growth case. In average, there is a 15% PV installed capacity limits increase for all annual energy consumption growth for the 5% risk. Such small risk can be covered using technologies or services, e.g., PV droop control, tap changer settings, or flexibility markets, allowing a higher level of PV installations.

F. Growth Paths on the Static Operation Regions

Another advantage of the static operation regions nomogram is that the DNO can analyze multiple growth path scenarios. The static operation regions computed in Fig 11 are used as a template for the growth path analysis. For instance, Fig. 13 shows three sets of growth curves (each row of subplots are the curves that define each parametric growth path $\Gamma(l)$ defined in (3)). The nomogram used in this example is for the 10% risk scenario shown in Fig. 11(b). The sub-indexes label each growth path as A, B, and C. It should be noted that each dot in each path

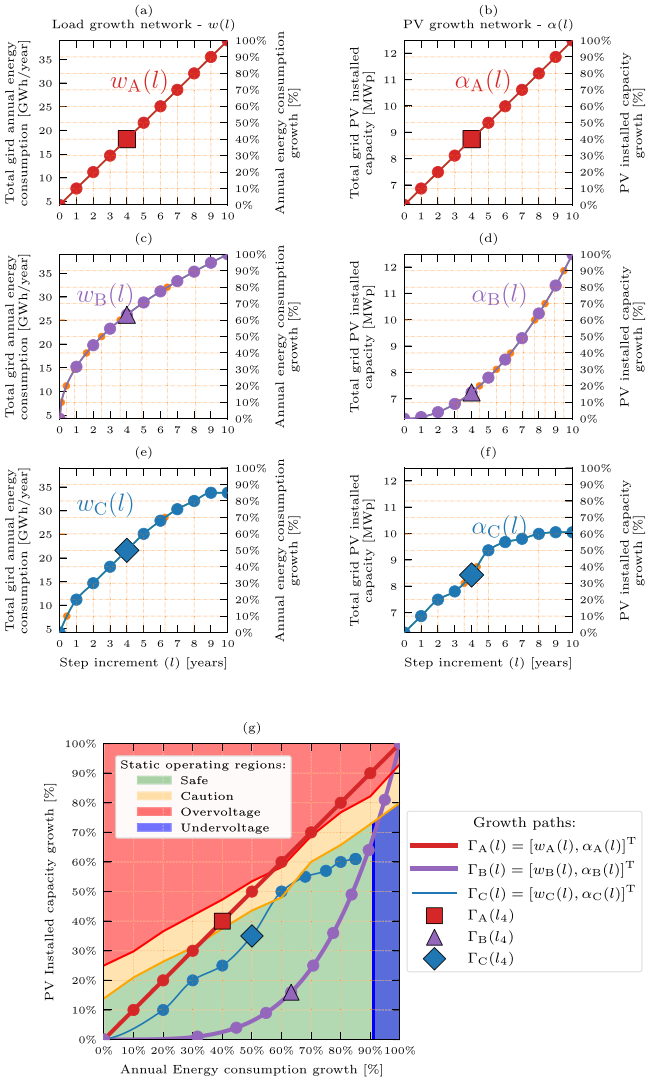

Fig. 13. Example of three different growth function and growth paths for the total grid annual energy consumption and PV installed capacity for a static operating region with 10% risk. Different growth functions, i.e., $w(l)$ and $\alpha(l)$, create different paths $\Gamma(l)$. Linear growth path $\Gamma_{\mathrm{A}}(l) = [w_{\mathrm{A}}(l), \alpha_{\mathrm{A}}(l)]^{\mathrm{T}}$ is composed of the linear function $w_{\mathrm{A}}(l)$ show in (a) and the linear function $\alpha_{\mathrm{A}}(l)$ shown in (b). Non-linear growth functions $w_{\mathrm{B}}(l)$ and $\alpha_{\mathrm{B}}(l)$, shown in (c) and (d), creates a non-linear path shown in purple in (f). The DNO adjusts the growth functions like (e) and (f) aided by the static operating region while considering the risk and maintaining a safe operation. Subplot (g) shows the three different growth paths created by the three sets of growth functions. The square, triangle, and diamond markers in the subplot (g) show the growth path location for year 4 for the different growth functions.

(Fig. 13(g)) corresponds to one step increment l (in years) of the growth functions $w(l)$ and $\alpha(l)$. The growth path $\Gamma_{\mathrm{A}}(l)$ is the initial plan used for the PPF; therefore, it defines colors in the nomogram. In year four of $\Gamma_{\mathrm{A}}(l)$ (highlighted by a red square marker), the grid enters the caution region, corresponding to annual energy consumption and PV installed capacity growth of 40%. Then, it reaches the overvoltage region in year six. A different growth path in which most of the years in the growth horizon are in the safe regions is shown as $\Gamma_{\mathrm{B}}(l)$, where later, at year eight, the grid enters an undervoltage region.

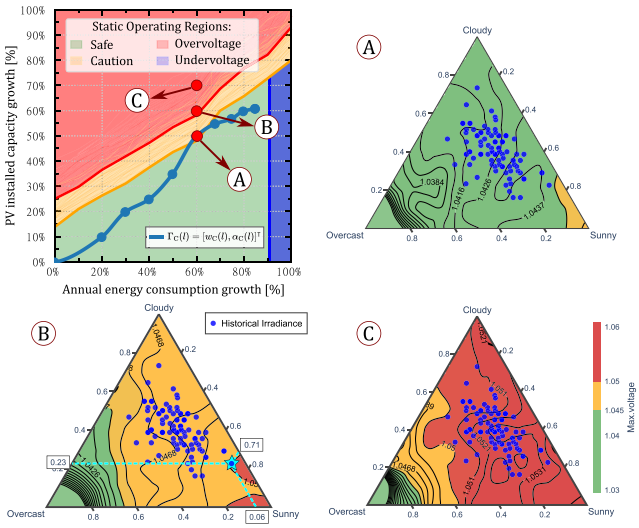

Fig. 14. Example of voltage fraction static operating regions with 10% risk. Ternary plots guide the assessment of the caution and danger regions which are variable due to irradiance conditions. Three cases of annual energy consumption and PV installed capacity growth combinations (A, B, C) are shown with their respective ternary plot. The blue line in the upper left plot is the growth path $\Gamma_C(l) = [w_C(l), \alpha_C(l)]^\intercal$ from Fig. 13(g). Blue scatter dots on the ternary plots show the historical $\boldsymbol{\pi}$ mixtures for the summer months in the last 20 years for the particular geographical area in the case study.

Nevertheless, the DNO can design a path within the safe region, as shown in $\Gamma_C(l)$. A natural question may arise about the validity of the static regions when a different growth path is sketched in the nomogram. In Appendix B, we show which are the restrictions and mathematical proof explaining why the solutions are still correct. It is essential to notice that the paths in the nomogram decouple the time horizon and the PV and energy percentages values, and the points in the paths can be closer to each other for similar combinations. e.g., years 6-9 in $\Gamma_C(l)$. This is a versatile tool for DNOs to decide and steer the original growth plan $\Gamma_A(l)$ into a different one that stays within technical limits. Technologies, services, techniques, or control mechanisms to reach a new growth path are out of the scope of this article, but it is of interest for future research.

The frontiers of the operating regions in Fig. 11 and 13(g) are built using a wide range of combinations of types of days ($\boldsymbol{\pi}$ mixtures). In reality, all irradiance scenarios do not have the same chance of occurrence in the distribution network, as it depends on the characteristics of the geographical location. Each location has different characteristics that affects cloud coverage, such as topography, air humidity, and temperature.

A ternary plot is used to discern and navigate through the frontier lines in the regions (Fig. 14). The ternary plot means that for each annual energy consumption and PV growth combination, the maximum network voltage magnitude for all combinations of mixtures $\boldsymbol{\pi}$ (cloudiness conditions) can be precisely evaluated. For instance, the upper-left plot in Fig. 14 highlights three combinations of annual energy consumption and PV installed capacity growth (**A**, **B**, and **C**) with their respective ternary plots. The blue scatter dots, overlaid in the ternary plot, represent the irradiance mixture for each summer month in the past 20 years on the grid's location under study. For instance, the combination that lies in the frontier between the safe and caution region, highlighted by point **A**, is year four in the growth path $\Gamma_C(l)$ shown in Fig. 13(g), and it is still between technical limits as the historical irradiance data lies below the V_{caution} value (shown in upper right ternary plot in Fig. 14). In the frontier of the over-voltage region, highlighted by point **B** in the upper-left plot (case where all days are perfectly sunny), the likelihood of having an overvoltage is still low because the location never reaches a completely clear sky in summer, i.e., there is not a historical blue dot in the lower right corner of the ternary plot **B**. The historical worst case month in summer has a mixture of 71% sunny, 23% cloudy, and 6% overcast days (highlighted by the blue star in ternary plot **B**), indicating that most of the historical cases are below $\overline{V}$ and the PV limit can potentially be increased even further. Point **C** is always in over-voltage as all historical points lie in the over-voltage region.

VII. Discussion

The current framework operates under the assumption of constant irradiance in the area serviced by the MV distribution, which may not be the case for rural areas. A future research direction is the inclusion of spatial correlation for irradiance models such as in [42], [43], or models with satellite-aided data like in [44].

Reliability analysis adds an extra layer of complexity, as not only over and undervoltage technical limits are included, but also overloading on lines and transformers in the event of a failure. The current framework does not support an "N-1" reliability analysis. Nevertheless, similar nomograms can be proposed to analyze grid components' overloading restrictions. This exciting analysis is left for a future research direction.

The case study presented in this work relies on historical consumption data on serviced areas. However, the load profiles could have new shapes due to disruptive technologies such as electric vehicles and heat pumps. Additionally, the non-heterogeneous changes in the profile can cause the node to switch clusters during the load growth. For instance, one node could change from cluster two, which are residential profiles with low daily consumption, to cluster two in a year to come, which is a residential profile with higher consumption during the daylight, or change to cluster three due to a change on the land use for commercial activities. Nevertheless, the framework can be fed with the possible scenarios based on the possible likelihood of technology adoption [35] or activity change, keeping the uncertainty quantification with the proposed framework.

Another application of the framework is to quantify the benefits of different control technology strategies, e.g., classical PV droop controls [45], data-driven methods [46], using active or reactive power compensation. For this matter, the "probabilistic power flow" step in Fig. 3 should be modified to run the simulations with the control mechanisms to determine each technology's effectiveness. In that way, risk reduction can be quantified due to different control mechanisms.

Including control mechanism comes with a computational challenge. Our framework relies on a computationally fast formulation of the power flow [39] to execute the probabilistic power flow in a laptop with average specifications. Including the actions of the control could increase the computational time dramatically. Therefore, future research should also be focused on effectively including the control mechanism in the probabilistic power flow in a reasonable computational time.

VIII. Conclusion and Future Work

This article proposed a complete probabilistic modeling framework using MVT copulas for load profiles (including area activity) and variable solar irradiance profiles to quantify the voltage magnitude impact of annual energy consumption and PV installed capacity growth in a distribution network. The nomogram and ternary plots can help DNOs to diagnose potential problems in the network, considering different levels of risk and local historical irradiance data. Results show that local irradiance conditions can change the maximum PV installed capacity limits on the network, and calculation of maximum PV capacity using conventional approaches can severely underestimate the PV generation potential on the network, restraining the penetration of clean energy generation. The case study shows that taking 5% of risk could increase in average 15% of the PV installed capacity limits. Additionally, it is shown that the static operating regions are an effective tool for the DNOs to steer the PV installed capacity and annual energy consumption growth to maintain the safe operation of the grid for a pre-defined growing horizon.

The framework relies on existing consumption data sets to model the annual energy consumption growth. Nevertheless, future loads comes with different energy profiles due to EVs, heat pumps, batteries, and thermal storage. A vast body of research has been done on this topic and is out of this article's reach. Future research direction is to incorporate those uncertainties from new technologies into the load profiles, which extended with proposed irradiance modeling can bring an even more precise decision tool for DNOs.

Appendix

A. Variability Index Metrics (VIMs)

The four VIMs used in this work are the following:

1) Fractal Dimension (FD) [21]: The metric uses fractal analysis to characterize the variability of the the irradiance profile for the day. It uses a rectangle as the structuring element to cover the irradiance signal with the area defined as

$$S(\Delta\tau) = \sum_{t=0}^{T-1} \Delta\tau \, |(g_{t+\Delta\tau} - g_t)|. \tag{16}$$

A sweep of different $\Delta\tau$ steps is done to calculate the area for the total signal and calculate the fractal dimension (FD) using the following linear regression

$$\ln\left(\frac{S(\Delta\tau)}{\Delta\tau^2}\right) \cong \text{FD} \cdot \ln\left(\frac{1}{\Delta\tau}\right) + \text{constant} \tag{17}$$

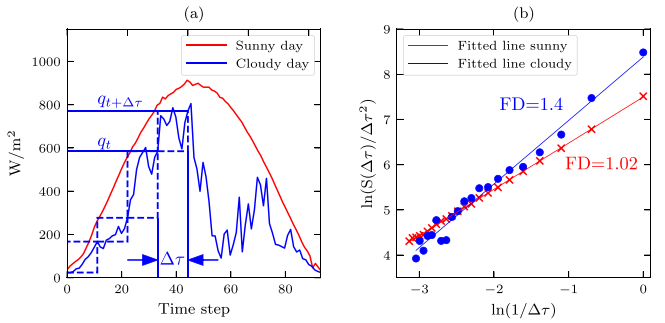

Fig. 15. Example of the fractal dimension (FD) calculation for a sunny and cloudy day. (a) Cloudy irradiance curve covered by rectangles with area computed by (16). (b) The fractal dimension value is the slope of the fitted line by (17). Lower FD values corresponds to sunny days and higher values for cloudy days.

Fig. 15 shows an example for computing the fractal dimension metric for a cloudy and sunny day.

2) Mean of Increments (MI) [20]: Quantifies the average of the absolute ramp values of irradiance during the day, and it is expressed as

$$\text{MI} = \frac{1}{T-1} \sum_{t=1}^{T-1} |\Delta g_t|. \tag{18}$$

3) Standard Deviation of Increments (SDI) [47]: Measures the standard deviation of the changes in irradiance and around its mean, and it is described as:

$$\text{SDI} = \sqrt{\frac{\sum_{t=1}^{T-1} (|\Delta g_t| - \text{MI})^2}{T-1}} \tag{19}$$

4) Variability Index (VI) [48]: It quantifies the variability or the irradiance signal using the ratio of the envelope of the measured irradiance and the GHI model. The relation is the following:

$$\text{VI} = \frac{\sum_{t=1}^{T-1} \sqrt{(\Delta g_t)^2 + (\Delta t)^2}}{\sum_{t=1}^{T-1} \sqrt{(\Delta \text{GHI}_t)^2 + (\Delta t)^2}}. \tag{20}$$

Higher values of VI means that measured signal is highly stochastic, and values closer to 1 are days that are have a clear sky.

B. Growth Paths for Annual Energy and PV Installed Capacity

The growth path **A** over the static operating region in Fig. 13(g) is defined by the parametric curve

$$\Gamma_{\text{A}}(l) = \begin{bmatrix} w_{\text{A}}(l) \\ \alpha_{\text{A}}(l) \end{bmatrix}, \tag{21}$$

where the total annual energy consumption growth $w_{\text{A}}(l)$ consists of the total sum of the annual energy growth functions $w_{i,\text{A}}(l)$ for all nodes, on a grid with B number of nodes, for $i = \{1, \ldots, B\}$. The growth is assumed to have a finite horizon H, i.e., $l \in [0, H]$. Also, each node has a minimum $((\underline{w})_i)$ and maximum $(\overline{w}_i)$ annual energy consumption for the growth

horizon. More precisely, it is expressed as

$$w_{\mathrm{A}}(l) = f_{1,\mathrm{A}}(l)(\overline{w}_1 - (\underline{w})_1) + (\underline{w})_1 + \ldots + f_{B,\mathrm{A}}(l)(\overline{w}_B - (\underline{w})_B) + (\underline{w})_B$$

$$= \sum_{i=1}^{B} \big(\underbrace{f_{i,\mathrm{A}}(l)(\overline{w}_i - \underline{w}_i) + \underline{w}_i}_{\tilde{w}_{i,\mathrm{A}}(l)} \big) \tag{22}$$

$$= \sum_{i=1}^{B} (f_{i,\mathrm{A}}(l)\Delta w_i + \underline{w}_i) \tag{23}$$

Similarly, the total PV installed capacity growth ($\alpha_{\mathrm{A}}(l)$), with minimum ($\underline{pv}_i$) and maximum ($\overline{\mathrm{pv}}_i$) installed capacity in the same horizon, is defined as

$$\alpha_{\mathrm{A}}(l) = \sum_{i=1}^{B} \big(\underbrace{h_{i,\mathrm{A}}(l)(\overline{\mathrm{pv}}_i - \underline{\mathrm{pv}}_i) + \underline{\mathrm{pv}}_i}_{\tilde{\alpha}_{i,\mathrm{A}}(l)} \big). \tag{24}$$

The functions $f_{i,\mathrm{A}}(l)$ and $h_{i,\mathrm{A}}(l)$ are *normalized growth functions* with a range of values between [0,1] and strictly monotonically increasing. e.g., linear, squared, root growth functions in the domain $l \in [0, H]$. It should be noted that the functions $\tilde{w}_{i,\mathrm{A}}(l)$ and $\tilde{\alpha}_{i,\mathrm{A}}(l)$ are just curve de-normalization operations for each node for their respective normalized growth function $f_{i,\mathrm{A}}(l)$ and $h_{i,\mathrm{A}}(l)$.

The static operating regions computed by the growth path functions $\Gamma_{\mathrm{A}}(l) = [w_{\mathrm{A}}(l), \alpha_{\mathrm{A}}(l)]^{\intercal}$ can be used for a new growth path curve $\Gamma_{\mathrm{B}}(l) = [w_{\mathrm{B}}(l), \alpha_{\mathrm{B}}(l)]^{\intercal}$ if it satisfies the relations

$$\frac{\tilde{w}_{i,\mathrm{A}}(l)}{w_{\mathrm{A}}(l)} = \frac{\tilde{w}_{i,\mathrm{B}}(l^*)}{w_{\mathrm{B}}(l^*)}, \tag{25}$$

$$\frac{\tilde{\alpha}_{i,\mathrm{A}}(l)}{\alpha_{\mathrm{A}}(l)} = \frac{\tilde{\alpha}_{i,\mathrm{B}}(l^*)}{\alpha_{\mathrm{B}}(l^*)} \quad \forall \quad i = \{1, \ldots, B\}. \tag{26}$$

Without loss of generalization for the total PV installed capacity growth relation in (25b), we show that for (25a), there is a unique solution l^* on which the ratio is satisfied by the new curve at $w_{\mathrm{B}}(l^*)$ assuming: (i) all *normalized growth functions* are of the same type for the total growth function. i.e., $f_{1,\mathrm{A}}(l) = \ldots = f_{i,\mathrm{A}}(l)$ for $w_{\mathrm{A}}(l)$ and $f_{1,\mathrm{B}}(l) = \ldots = f_{i,\mathrm{B}}(l)$ for $w_{\mathrm{B}}(l)$; (ii) both total growth functions have the same value at the solution point. i.e., $w_{\mathrm{A}}(l) = w_{\mathrm{B}}(l^*)$. In that case, relation (25a) becomes

$$\tilde{w}_{i,\mathrm{A}}(l) = \tilde{w}_{i,\mathrm{B}}(l^*)$$

$$\sum_{i=1}^{B} (f_{i,\mathrm{A}}(l)\Delta w_i + \underline{w}_i) = \sum_{i=1}^{B} (f_{i,\mathrm{B}}(l^*)\Delta w_i + \underline{w}_i)$$

$$f_{i,\mathrm{A}}(l) = f_{i,\mathrm{B}}(l^*), \tag{27}$$

with the unique solution

$$l^* = f_{i,\mathrm{B}}^{-1}(f_{i,\mathrm{A}}(l)), \tag{28}$$

where we use the fact that *normalized growth functions* are strictly monotonically increasing, which has an inverse with one-to-one mapping, meaning that l^* is unique.

As a numerical example, Fig. 16 shows three monotonically increasing load growths which follow relation (25a), for a grid with three nodes. In this particular example, it is evident that

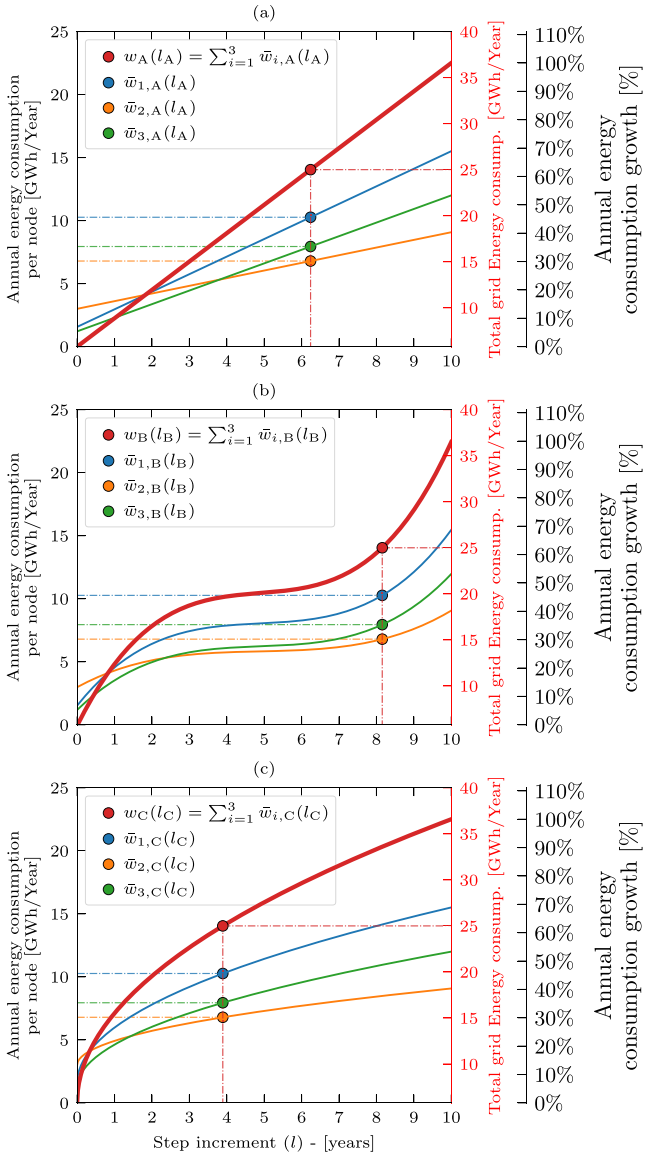

Fig. 16. Example of three growth curves that satisfies the relation (25a) for the normalized growth curve functions: (a) lineal $f_{i,\mathrm{A}}(l) = (l/H)$, (b) normalized spline function $f_{i,\mathrm{B}}(l) = \mathrm{spline}(l/H)$, and (c) square root $f_{i,\mathrm{B}}(l) = \sqrt{(l/H)}$. All highlighted red points are for the total energy consumption of 25 [GWh/year] (secondary y-axes in red). The ratio between the energy values for each individual node and the total energy growth is preserved for all cases (green, blue and orange points have the same values in the left-hand y-axes).

all the annual energy consumption values for *each node* (blue, green, and orange markers) have the same values in the left y-axis for the same total annual energy consumption (the red marker value shown in the right-hand axis in red).

The relation (25) implies that any path $\Gamma_{\mathrm{B}}(l)$ with strictly monotonically increasing growth functions $w_B(l)$ and $\alpha_{\mathrm{B}}(l)$ can be sketched on the static operating regions nomogram (Fig. 2(a) and Fig. 13(g)), and the computed solutions by $\Gamma_{\mathrm{A}}(l)$ (colored zones in the static regions) are still valid. In other words, the total grid annual energy consumptions simulated in the lineal case exists in the other two cases (in reference to Fig. 16). The *node* energy values *preserve the same ratios* with respect to the total energy consumption across the growth horizon, meaning

that the energy values per node used as an input for model (14) and (15) are the same, and therefore the PPF results are equal (same static operating regions). The only difference between the different growth curves is that the total energy values of the lineal case happen sooner or later on the other new growth curve plans. In the example shown in Fig. 16, the total annual energy consumption of 25 [GWh/year] occurs in $l_{\mathrm{A}} = 6.2$ in the lineal case; for the spline growth function at $l_{\mathrm{B}} = 8.2$; and the root function at $l_{\mathrm{C}} = 3.9$ years.

REFERENCES

[1] A. Navarro-Espinosa and L. F. Ochoa, "Probabilistic impact assessment of low carbon technologies in LV distribution systems," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 2192–2203, May 2016.

[2] F. Olivier, P. Aristidou, D. Ernst, and T. V. Cutsem, "Active management of low-voltage networks for mitigating overvoltages due to photovoltaic units," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 926–936, Mar. 2016.

[3] S. Wang, Y. Dong, L. Wu, and B. Yan, "Interval overvoltage risk based PV hosting capacity evaluation considering PV and load uncertainties," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2709–2721, May 2020.

[4] H. Mortazavi, H. Mehrjerdi, M. Saad, S. Lefebvre, D. Asber, and L. Lenoir, "A monitoring technique for reversed power flow detection with high PV penetration level," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2221–2232, Sep. 2015.

[5] R. Torquato, D. Salles, C. O. Pereira, P. C. M. Meira, and W. Freitas, "A comprehensive assessment of PV hosting capacity on low-voltage distribution systems," *IEEE Trans. Power Del.*, vol. 33, no. 2, pp. 1002–1012, Apr. 2018.

[6] T. Stetz, F. Marten, and M. Braun, "Improved low voltage grid-integration of photovoltaic systems in Germany," *IEEE Trans. Sustain. Energy*, vol. 4, no. 2, pp. 534–542, Apr. 2013.

[7] E. Mulenga, M. H. Bollen, and N. Etherden, "A review of hosting capacity quantification methods for photovoltaics in low-voltage distribution grids," *Int. J. Elect. Power Energy Syst.*, vol. 115, Feb. 2020, Art. no. 105445.

[8] A. Hoke, R. Butler, J. Hambrick, and B. Kroposki, "Steady-state analysis of maximum photovoltaic penetration levels on typical distribution feeders," *IEEE Trans. Sustain. Energy*, vol. 4, no. 2, pp. 350–357, Apr. 2013.

[9] S. M. Ismael, S. H. A. Aleem, A. Y. Abdelaziz, and A. F. Zobaa, "State-of-the-art of hosting capacity in modern power systems with distributed generation," *Renewable Energy*, vol. 130, pp. 1002–1020, Jan. 2019.

[10] M. S. S. Abad, J. Ma, D. Zhang, A. S. Ahmadyar, and H. Marzooghi, "Probabilistic assessment of hosting capacity in radial distribution systems," *IEEE Trans. Sustain. Energy*, vol. 9, no. 4, pp. 1935–1947, Oct. 2018.

[11] M. U. Qureshi, S. Grijalva, M. J. Reno, J. Deboever, X. Zhang, and R. J. Broderick, "A fast scalable quasi-static time series analysis method for PV impact studies using linear sensitivity model," *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 301–310, Jan. 2019.

[12] J. Deboever, X. Zhang, M. Reno, R. Broderick, S. Grijalva, and F. Therrien, "Challenges in reducing the computational time of QSTS simulations for distribution system analysis," Sandia Nat. Lab., Albuquerque, NM, and Livermore, CA, USA, Tech. Rep. SAND–2017-5743, 1367462, 653822, Jun. 2017.

[13] M. J. Reno, J. Deboever, and B. Mather, "Motivation and requirements for quasi-static time series (QSTS) for distribution system analysis," in *Proc. IEEE Power Energy Soc. Gen. Meetings*, 2017, pp. 1–5.

[14] S. Haben, C. Singleton, and P. Grindrod, "Analysis and clustering of residential customers energy behavioral demand using smart meter data," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 136–144, Jan. 2016.

[15] R. Granell, C. J. Axon, and D. C. Wallom, "Clustering disaggregated load profiles using a Dirichlet process mixture model," *Energy Convers. Manage.*, vol. 92, pp. 507–516, Mar. 2015.

[16] M. Sun, I. Konstantelos, and G. Strbac, "C-vine copula mixture model for clustering of residential electrical load pattern data," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2382–2393, May 2017.

[17] E. M. S. Duque, P. P. Vergara, P. H. Nguyen, A. van der Molen, and J. G. Slootweg, "Conditional multivariate elliptical copulas to model residential load profiles from smart meter data," *IEEE Trans. Smart Grid*, vol. 12, no. 5, pp. 4280–4294, Sep. 2021.

[18] K. N. Hasan, R. Preece, and J. V. Milanović, "Existing approaches and trends in uncertainty modelling and probabilistic stability analysis of power systems with renewable generation," *Renewable Sustain. Energy Rev.*, vol. 101, pp. 168–180, Mar. 2019.

[19] J. Yin, A. Molini, and A. Porporato, "Impacts of solar intermittency on future photovoltaic reliability," *Nature Commun.*, vol. 11, no. 1, Dec. 2020, Art. no. 4781.

[20] R. Blaga and M. Paulescu, "Quantifiers for the solar irradiance variability: A new perspective," *Sol. Energy*, vol. 174, pp. 606–616, Nov. 2018.

[21] S. Harrouni, A. Guessoum, and A. Maafi, "Classification of daily solar irradiation by fractional analysis of 10-min-means of solar irradiance," *Theor. Appl. Climatol.*, vol. 80, no. 1, pp. 27–36, Feb. 2005.

[22] B. Hartmann, "Comparing various solar irradiance categorization methods–A critique on robustness," *Renewable Energy*, vol. 154, pp. 661–671, Jul. 2020.

[23] C. S. Lai, Y. Jia, M. D. McCulloch, and Z. Xu, "Daily clearness index profiles cluster analysis for photovoltaic system," *IEEE Trans. Ind. Inform.*, vol. 13, no. 5, pp. 2322–2332, Oct. 2017.

[24] A. A. Munshi and Y. A. -R. Mohamed, "Photovoltaic power pattern clustering based on conventional and swarm clustering methods," *Sol. Energy*, vol. 124, pp. 39–56, Feb. 2016.

[25] G. Chicco, V. Cocina, and F. Spertino, "Characterization of solar irradiance profiles for photovoltaic system studies through data rescaling in time and amplitude," in *Proc. IEEE 49th Int. Universities Power Eng. Conf.*, 2014, pp. 1–6.

[26] M. Gastón-Romeo, T. Leon, F. Mallor, and L. Ramírez-Santigosa, "A morphological clustering method for daily solar radiation curves," *Sol. Energy*, vol. 85, no. 9, pp. 1824–1836, Sep. 2011.

[27] T. Soubdhan, R. Emilion, and R. Calif, "Classification of daily solar radiation distributions using a mixture of dirichlet distributions," *Sol. Energy*, vol. 83, no. 7, pp. 1056–1063, Jul. 2009.

[28] Z. Ren, W. Yan, X. Zhao, W. Li, and J. Yu, "Chronological probability model of photovoltaic generation," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1077–1088, May 2014.

[29] B. Ngoko, H. Sugihara, and T. Funaki, "Synthetic generation of high temporal resolution solar radiation data using Markov models," *Sol. Energy*, vol. 103, pp. 160–170, May 2014.

[30] A. Frimane, J. M. Bright, D. Yang, B. Ouhammou, and M. Aggour, "Dirichlet downscaling model for synthetic solar irradiance time series," *J. Renewable Sustain. Energy*, vol. 12, no. 6, Nov. 2020, Art. no. 063702.

[31] F. Wu and S. Kumagai, "Steady-state security regions of power systems," *IEEE Trans. Circuits Syst.*, vol. 29, no. 11, pp. 703–711, Nov. 1982.

[32] F. C. B. Almeida, J. A. Passos Filho, J. L. R. Pereira, R. M. Henriques, and A. L. M. Marcato, "Assessment of load modeling in power system security analysis based on static security regions," *J. Control Automat. Elect. Syst.*, vol. 24, no. 1/2, pp. 148–161, Apr. 2013.

[33] O. F. Avila, J. A. P. Filho, and W. Peres, "Steady-state security assessment in distribution systems with high penetration of distributed energy resources," *Electric Power Syst. Res.*, vol. 201, Dec. 2021, Art. no. 107500.

[34] M. de Jong, G. Papaefthymiou, and P. Palensky, "A framework for incorporation of infeed uncertainty in power system risk-based security assessment," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 613–621, Jan. 2018.

[35] R. Bernards, J. Morren, and H. Slootweg, "Development and implementation of statistical models for estimating diversified adoption of energy transition technologies," *IEEE Trans. Sustain. Energy*, vol. 9, no. 4, pp. 1540–1554, Oct. 2018.

[36] J. A. Thorpe, *Elementary Topics in Differential Geometry*. New York, NY, USA: Springer, 1979.

[37] M. Salazar, I. Dukovska, P. H. Nguyen, R. Bernards, and H. J. Slootweg, "Data driven framework for load profile generation in medium voltage networks via transfer learning," in *Proc. IEEE/PES Innov. Smart Grid Technol. Europe*, 2020, pp. 909–913.

[38] D. Vercamer, B. Steurtewagen, D. Van den Poel, and F. Vermeulen, "Predicting consumer load profiles using commercial and open data," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3693–3701, Sep. 2016.

[39] J. S. Giraldo, O. D. Montoya, P. P. Vergara, and F. Milano, "A fixed-point current injection power flow for electric distribution systems using laurent series," *Electric Power Syst. Res.*, vol. 211, Oct. 2022, Art. no. 108326.

[40] LaPSEE power system test cases repository UNESP, "downloads / sistemas testes," Oct. 2020. [Online]. Available: https://bit.ly/2IdSn91

[41] B. Haurwitz, "Insolation in relation to cloudiness and cloud density," *J. Meteorol.*, vol. 2, no. 3, pp. 154–166, Sep. 1945.

[42] J. Widen, 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.

[43] J. Widén, M. Shepero, and J. Munkhammar, "On the properties of aggregate clear-sky index distributions and an improved model for spatially correlated instantaneous solar irradiance," *Sol. Energy*, vol. 157, pp. 566–580, Nov. 2017.

[44] C. O. Inacio and C. L. T. Borges, "Stochastic model for generation of high-resolution irradiance data and estimation of power output of photovoltaic plants," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 952–960, Apr. 2018.
[45] P. P. Vergara, M. Salazar, T. T. Mai, P. H. Nguyen, and H. Slootweg, "A comprehensive assessment of PV inverters operating with droop control for overvoltage mitigation in LV distribution networks," *Renewable Energy*, vol. 159, pp. 172–183, Oct. 2020.
[46] P. P. Vergara, M. Salazar, J. S. Giraldo, and P. Palensky, "Optimal dispatch of PV inverters in unbalanced distribution systems using reinforcement learning," *Int. J. Elect. Power Energy Syst.*, vol. 136, Mar. 2022, Art. no. 107628.
[47] M. Lave, J. Kleissl, and J. Stein, "Quantifying and simulating solar-plant variability using irradiance data," in *Solar Energy Forecasting and Resource Assessment*. Amsterdam, The Netherlands: Elsevier, 2013, pp. 149–169.
[48] J. Stein, C. Hanse, and J. M. Reno, "The variability index: A new and novel metric for quantifying irradiance and PV output variability," Sandia National Laboratories, Albuquerque, NM, and Livermore, CA, USA, Tech. Rep. SAND2012-2088C 2012.

Edgar Mauricio Salazar Duque (Member, IEEE) received the B.E. degree in electrical and electronic engineering from the Universidad de Los Andes, Bogotá, Colombia, in 2008, the M.Sc. degree (*cum laude*) in smart electrical grids and systems from Kungliga Tekniska Högskolan (KTH), Stockholm, Sweden, and the Technical University of Eindhoven, Eindhoven, The Netherlands, in 2018. He is currently working toward the Ph.D. degree with the Electrical Energy Systems Group, Technical University of Eindhoven, Eindhoven, The Netherlands. His research focuses on data analysis, and applications of machine learning techniques on power distribution grids for planning and operation.

Juan S. Giraldo (Member, IEEE) received the B.Sc. degree in electrical engineering from the Universidad Tecnológica de Pereira, Pereira, Colombia, in 2012, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Campinas, Campinas, Brazil, in 2015 and 2019, respectively. From October 2019 to May 2021, he was a Postdoctoral Fellow with the Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands. Later, from June 2021 to August 2022 he was a Researcher with the Mathematics of Operations Research Group, University of Twente, Enschede, The Netherlands. He is currently a Researcher with the Energy Transition Studies Group, Netherlands Organization for Applied Scientific Research (TNO), Amsterdam, The Netherlands. His research interests include the optimization, planning, and control of energy systems, energy transition pathways, and machine learning applications to energy systems.

Pedro P. Vergara (Member, IEEE) was born in Barranquilla, Colombia, in 1990. He received the B.Sc. degree (with Hons.) in electronic engineering from the Universidad Industrial de Santander, Bucaramanga, Colombia, in 2012, and the M.Sc. degree in electrical engineering from the University of Campinas, Campinas, Brazil, in 2015, and the Ph.D. degree from the University of Campinas, Brazil, and University of Southern Denmark, Odense, Denmark, funded by the Sao Paulo Research Foundation (FAPESP), in 2019. In 2019, he joined the Eindhoven University of Technology, TU/e, Eindhoven, The Netherlands, as a Postdoctoral Researcher. In 2020, he was appointed as an Assistant Professor with the Intelligent Electrical Power Grids Group, Delft University of Technology, Delft, The Netherlands. His main research interests include the development of algorithms for the control, planning, and operation of electrical distribution systems with high penetration of low-carbon energy resources (such as electrical vehicles, PV systems, and electric heat pumps) using optimization and machine learning approaches. Dr. Vergara was the recipient of the Best Presentation Award at the Summer Optimization School in 2018 organized by the Technical University of Denmark and Best Paper Award at the 3rd IEEE International Conference on Smart Energy Systems and Technologies, in Turkey, in 2020.

Phuong H. Nguyen (Member, IEEE) received the Ph.D. degree from the Eindhoven University of Technology (TU/e), Eindhoven, The Netherlands, in 2010. During his one-year sabbatical leave in 2019, he took up a group leader position of the Sustainable Energy Systems Group, Luxembourg Institute of Science and Technology, Hautcharage, Luxembourg. Since January 2020, he has been back to TU/e as an Associate Professor with the Electrical Energy System Group. Dr. Phuong Nguyen has committed his research effort to realize synergies of advanced monitoring and control functions for the distribution networks along with emerging digital technologies. This distinctive combination of competences allows him to develop a research pathway crossing over various domains of mathematical programming, stochastics, data mining, and communication networks. His research interests include data analytics with deep learning, real-time system awareness using (IoT) data integrity, and predictive and corrective grid control functions.

Anne van der Molen (Member, IEEE) received the M.S. degree in electrical engineering from Twente University, Enschede, The Netherlands, in 1997. He is presently working for Dutch Distribution System Operator Stedin where he is engaged with smart grids strategy- and technology planning. His research interests include system/market operations, operational technology, and flexibility. Mr. van der Molen is part-time Research Associate with the Eindhoven University of Technology, Eindhoven, The Netherlands, in the area of intelligent energy systems. He is also the Chair of the Dutch Network Operators association's working group on flexibility and storage, which connects the activities of the Dutch network operators and which works closely together with government, research institutes and industry on capability- and technology development. Mr. van der Molen is also a Member of the Technology Committee of the Association of European Distribution System Operators.

J. G. Slootweg (Senior Member, IEEE) received the M.Sc. degree (*cum laude*) in electrical power engineering and the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 1998 and 2003, respectively, and the second M.Sc. degree in business administration. He is currently the Director of the Asset Management Department, Enexis Netbeheer B.V., 's-Hertogenbosch, The Netherlands, one of the largest Distribution Network Operators of the Netherlands. Its spearheads are the strategic goals of Enexis: accelerating the transition towards a more sustainable energy supply and excellent, state of the art network operation. He also holds a professorship in smart grids with the Electrical Energy Systems Group, Eindhoven University of Technology, Eindhoven, The Netherlands. He has co-authored more than 200 papers, covering a broad range of various aspects of electrical power systems.