Conditional Multivariate Elliptical Copulas to Model Residential Load Profiles from Smart Meter Data

Salazar, Mauricio; Vergara Barrios, P.P.; Nguyen, Phuong H. ; van der Molen, Anne; Slootweg, J.G.

DOI
10.1109/TSG.2021.3078394

Publication date
2021

Document Version
Accepted author manuscript

Published in
IEEE Transactions on Smart Grid

Citation (APA)
Salazar, M., Vergara Barrios, P. P., Nguyen, P. H., van der Molen, A., & Slootweg, J. G. (2021). Conditional Multivariate Elliptical Copulas to Model Residential Load Profiles from Smart Meter Data. *IEEE Transactions on Smart Grid*, *12*(5), 4280-4294. Article 9425537. https://doi.org/10.1109/TSG.2021.3078394

Conditional Multivariate Elliptical Copulas to Model Residential Load Profiles from Smart Meter Data

Mauricio Salazar, *Student Member, IEEE,* Pedro P. Vergara, *Member, IEEE,* Phuong H. Nguyen, *Member, IEEE,* Anne van der Molen, *Member, IEEE,* J.G. Slootweg, *Senior Member, IEEE*

***Abstract*—The development of thorough probability models for highly volatile load profiles based on smart meter data is crucial to obtain accurate results when developing grid planning and operational frameworks. This paper proposes a new top-down modeling approach for residential load profiles (RLPs) based on multivariate elliptical copulas that can capture the complex correlation between time steps. This model can be used to generate individual and aggregated daily RLPs to simulate the operation of medium and low voltage distribution networks in flexible time horizons. Additionally, the proposed model can simulate RLPs conditioned to an annual energy consumption and daily weather profiles such as solar irradiance and temperature. The simulated daily profiles accurately capture the seasonal, weekends, and weekdays power consumption trends. Five databases with actual smart meter measurements at different time resolutions have been used for the model's validation. Results show that the proposed model can successfully replicate statistical properties such as autocorrelation of the time series, and load consumption probability densities for different seasons. The proposed model outperforms other multivariate state-of-the-art methods, such as Gaussian Mixture Models, by one order of magnitude in two different distance metrics for probability distributions.**

***Index Terms*—Multivariate copulas, load modeling, stochastic modeling, Gaussian Mixture Model.**

I. Introduction

THE energy transition comes with increasing penetration of low carbon technologies in the electrical distribution grid such as photovoltaic (PV) and wind generation [1]. This transition has also created environmental awareness on household energy consumption, prompting changes in home appliances, like the swap from gas to electricity for cooking and residential heating, and the use of electric vehicles as a mobility solution. This transformation creates high volatility and uncertainties in the residential load consumption profiles (RLPs), which introduces more challenges to the distribution network operation.

Mauricio Salazar and Phuong H. Nguyen are with the Electrical Energy Systems (EES) Group, Eindhoven University of Technology, Eindhoven 5612AE, The Netherlands, emails: {e.m.salazar.duque, p.p.vergara.barrios, p.nguyen.hong}@tue.nl.

Pedro P. Vergara is with the Electrical Energy Systems (EES) Group, Eindhoven University of Technology, Eindhoven 5612AE, The Netherlands, and also with the Intelligent Electrical Power Grids (IEPG) group, Delft University of Technology, Delft 2628CD, The Netherlands, emails: p.p.vergara.barrios@tue.nl, p.p.vergarabarrios@tudelft.nl.

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

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

Modeling RLPs in low voltage (LV) distribution networks has been an active field of research. Proper stochastic modeling of load consumption is required in different types of studies such as modern grid planning [2], quantification of the impact of low carbon technologies [3], [4], finding secure levels of penetration of PV generation [5], [6] and LV state estimation [7]. The accuracy that these studies can provide relies heavily on the quality of the stochastic models that can capture the variability of the consumption patterns, making the residential load modeling an essential task for making well-founded decisions.

In Europe, the smart meter data is protected due to privacy concerns [8]. Stochastic models have the benefit that the original data's statistical behavior is kept, reflecting it in the simulated profiles, without including any individual measured data. The future distribution grids will have millions of smart meters installed. The capability to describe the consumption patterns with a few parameters in a probabilistic representation is desired to compress large volumes of data in compact models. A stochastic model can also generate training databases of arbitrarily larger size, useful for machine learning tasks [9], and Monte Carlo methods [10], [11].

A. Literature Review and Contributions

The research on RLPs modeling can be grouped into two main approaches:

Bottom-up approach implements a Markov chain model that simulates the dwellers' behavior inside the households and their interaction with home appliances [12]. Usually, these methods are based on social demographic data [13], or appliances characteristics and consumption duration [14]–[17]. Bottom-up approaches have good results and usually are used for testing demand-side management applications. However, their main drawback is that they require the number of dwellers in the households and very detailed information about the use of the appliances. Such modeling is infeasible for the distribution network operators (DNOs) for network analysis to scale up to tens or hundreds of households because of its modeling intensity and privacy concerns.

Top-down approach are more interesting for grid operators since these models can be built based on existing smart meter measurements. The main purpose of the top-down approaches is to capture the statistical properties of the actual measurement data set. There are two main modeling methods used for this approach: Markov chain models, and probabilistic models that use parametric families of probability distribution functions (PDFs). In the Markov chain models, a

transition matrix is developed over discretized bins of power consumption for each time step, and then the model is sampled using a random walk to create the RLPs [18], [19]. The work in [20] presents a Markov chain model that also considers power consumption changes due to seasonality. In [21], load profiles from a smart meter dataset are clustered, and for each group, a Markov model is created, improving the quality of the generated profiles. Nevertheless, the main disadvantage of these models is that the range of power consumption is not accurately modeled due to the required discretization at the consumption levels, creating blind gaps on the edges of the discretized bins. This can be a critical issue as the model might not capture the whole range of consumption values adequately, causing high power consumption underestimation, as shown in [21] and [22].

In the probabilistic models, each time step of the RLP is considered as a random variable, and the load profile is modeled as a joint multivariate probability distribution. Different families of PDFs have been tested to approximate load consumption, e.g., Log-normal, Weibull, Generalized Extreme Value [23], [24]. However, [25] shows that load consumptions do not follow specific PDFs, and it suggested using more flexible techniques to represent different types of load distributions appropriately. One multivariate probabilistic technique is the Gaussian Mixture Models (GMMs) [26], extensively used in the literature due to its flexibility to adapt to unknown multi-modal and multivariate distributions. The work in [27] uses a GMM to capture temporal correlations between time steps of the aggregated RLPs for medium voltage (MV) to LV distribution transformer loading. However, no analysis was presented to model RLPs at the LV level. In general, GMMs are very flexible, but they are limited by the assumption that any multivariate joint distribution can be constructed with elliptically symmetric Gaussian probability densities. As shown later, GMMs can not properly model the complex correlation between time steps of RLPs at the LV level.

A different multivariate probabilistic technique is the use of copulas, which have been introduced in the context of energy systems for clustering [28], wind power and solar irradiance generation modeling [29]. Most of the load modeling that employs copulas is for applications on planning and secure operation on transmission grids [30]–[32], which only use aggregated residential load profiles and not for individual households, making the application for distribution systems relatively new. Moreover, the above-mentioned application approaches deal exclusively with few variables, e.g., wind, PV, and load-generation for few buses on the grid. In this sense, in the technical literature, there has been little effort to model high-dimensional dependent stochastic variables for time step correlations from smart meter data, e.g., a problem with 96 stochastic variables representing a daily profile of 15 minutes resolution. The advantage of copulas is that it does not model the joint distributions assuming elliptical distributions; instead, it focuses on the correlations between the marginal distributions that are modeled independently. As a result, multivariate copula models can be more flexible than GMMs for modelling complex correlations between variables. Additionally, the copula modeling of RLP considers each time step as a continuous random variable of active power consumption, overcoming the problem of power discretization required by the Markov models. The use of copulas to capture temporal correlation on RLPs has been tested before in [33], focused on the modeling of marginal distributions for each time step using GMMs before applying the copula correlation. Unlike this, the model shown in this paper does not use any parametric family distribution over the marginals, giving more flexibility for the copula model to capture complex correlation between time steps. A more advanced technique involving copulas applied to RLPs is the use of vines copulas [28], [34]. The disadvantage is that the model's complexity increases exponentially with the number of variables [35]. Model selection is problematic due to the vines' hierarchical nature, and sampling techniques over one or more dependent variables (conditioned probability model) are not simple [36]. In contrast, our proposed approach keeps using the simplicity of multivariate elliptical copulas (i.e., multivariate Gaussian (MVG) and multivariate t-distribution (MVT) copulas), which is practical to sample and the conditional probability has an analytical solution.

In most power systems applications, the Gaussian copula became the default approach for calculating stochastic variables correlation and scenario generation [29], [31], [37], [38] without any further consideration on the type of data to be modeled. Nevertheless, the MVT copula can benefit the individual RLPs modeling due to its ability to capture high values variations [39]; this property has not been explored for the highly volatile RLPs. In this paper, the modeling approach jointly evaluates the multivariate elliptical copulas for modeling high dimensional temporal correlations, which can be applied for both; individual and aggregated RLPs due to its general mathematical formulation.

Most importantly, no particular attention has been given in the technical literature to take advantage of conditioning the joint probability distribution that the copula models, e.g., to simulate processes when one or more variables are known. In this paper, we have particularly focused on the simulation of load profiles conditioned to an annual energy consumption. Nevertheless, the developed multivariate elliptical copula models can also be use to simulate RLPs conditioned to weather data e.g., temperature, solar irradiation. In this sense, our modeling approach gives and an extra tool for DNOs to evaluate such possible scenarios for LV networks. In summary, the main contributions of this paper are as follows:

- A proposal for a new top-down modeling approach for RLPs based on multivariate elliptical copulas that (i) can capture the high-dimensional temporal correlation between time steps and annual energy consumption; and (ii) can reproduce the high volatility of residential demand accurately. The proposed approach unifies the consumption modeling for MV and LV levels, simulating active power consumption scenarios at 15, 30, and 60 minutes resolution for a whole year.
- A new multivariate elliptical copula-based probability distribution model that simulates RLPs conditioned according to an annual energy consumption and daily weather profiles such as solar irradiance and temperature.

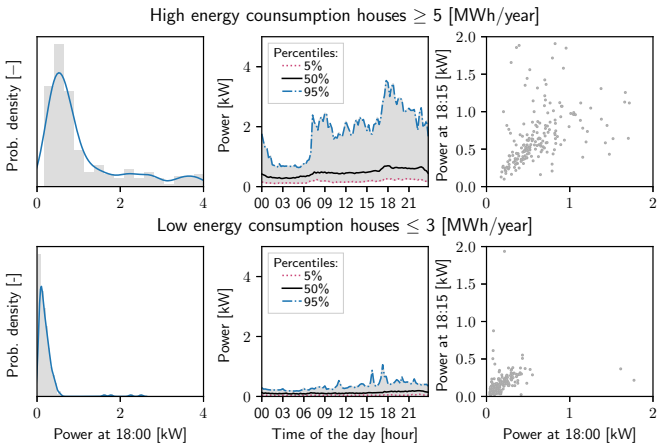

Fig. 1. Example of active power consumption for weekdays in June, for ten households with high annual energy consumption (first row), and ten households with low annual energy consumption (second row).

The remaining paper is organized as follows: Section II describes the statistical modeling problem for RLPs and the mathematical formulation of the proposed algorithms. Sections III presents a case study where the model's effectiveness is tested in a comprehensive case study for modeling individual and aggregated levels. For this, four different smart meter datasets at different time resolutions were used. Section IV summarizes and concludes the main results.

II. Statistical Modeling of Residential Load Profiles

A. Preliminary Analysis of RLP characteristics

The Fig. 1 shows an example of the difference in the statistical properties of RLPs between houses with high and low energy consumption. The left column of Fig. 1 shows the probability density of the active power consumption of one time step of the day for ten houses with high and low energy consumption. As can be seen, the density distributions have a different shape between the low and high energy consumption houses, mostly positively skewed, with longer tail in the high energy consumption case. This difference can also be seen in the central column of Fig. 1, which shows the complete profiles of the houses, in which the 5%, 50%, and 95% percentiles of the distribution densities are highlighted. Additionally, The right column of Fig. 1 shows a scatter plot of active power consumption between two consecutive time steps i.e., 18:00 and 18:15. Two important observations can be drawn: First, that there is a dependency structure between consecutive time steps; and second, that the dependency structure between time steps in the RLPs also depends on the annual energy consumption. For the same time of the day, high energy consumption households have a higher concentration of power values in the lower-left corner of the plot, and a heteroskedastic dispersion in the upper right corner. Furthermore, low energy consumption households have active power values more concentrated and less dispersed in the lower-left corner. Therefore, the annual energy consumption also influences the dependency structure between the time steps, and it should be considered as an extra variable in the statistical modeling.

B. Statistical Modeling

In general, a daily RLP of a household with annual energy consumption W can be discretized into T time steps. Each time step has an active power consumption value, considered a continuous random variable $X_i \; \forall \; i = 1, ..., T$. In this paper, the use of capital letters is for random variables and small letters for observed values. i.e., we let $\boldsymbol{X} = (X_1, ..., X_T)$ denote the random variable and $\boldsymbol{x} = (x_1, ..., x_T)$ its observed realization for active power consumption. Similarly, w as an annual energy realization of the random annual energy consumption variable W. The goal of the proposed model is to find a probability distribution function

$$F(x_1, \ldots, x_T, w), \tag{1}$$

that captures the dependencies between all the random variables that defines the residential load profile, i.e., X_i and W, knowing that each random variable has a different marginal distributions function, i.e., $\{F(x_1) \neq \ldots \neq F(x_T) \neq F(w)\}$. The expression in (1) can be seen as a generative model that can be sampled to simulate plausible load profiles for a household with a random annual energy consumption.

The probabilistic model in (1) can be conditioned to a specific value of annual energy consumption $\hat{w}$ as

$$F(x_1, \ldots, x_T | W = \hat{w}). \tag{2}$$

The condition $\hat{w}$ modifies the dependency structure of the time steps transitions according to the annual energy value $\hat{w}$. The conditioned model in (2) should match statistical properties of a smart meter dataset, consisting of N tuples of actual smart meter measurements of active power consumption for households with different annual energy consumption

$$\mathcal{D} = \{(x_{1,n}, \ldots, x_{i,n}, w_n)\}_{n=1}^{N}, \tag{3}$$

where the sub-index n is used to indicate the instance number in the data set $\mathcal{D}$.

Four statistical criteria are desired for the model in (2) that should match the actual data set $\mathcal{D}$. These are: (i) the density distribution of active power consumption during the year, (ii) the density distribution of the active power rate change between time steps, which is crucial for studies where the temporal behavior or net deviations are important, e.g., demand response management [40]; (iii) density distribution of the active power for each season of the year, divided by weekends and weekdays; and (iv) the average of the autocorrelation of the daily load profiles.

The complex dependency between random variables seen in Fig. 1 makes the modeling a difficult task. The finite mixture modeling based on Gaussian distributions [26] is a popular and flexible option to model (1). These models, known as GMMs, can also be conditioned as showed in [41] and [27] to model (2). Based on this, our proposal is referenced against a conditioned GMM.

C. Multivariate Elliptical Copula Modeling and Selection

For notation simplicity, the set of random variables in (1) are substituted as

$$\{x_1, \ldots, x_T, w\} = \{x_1, \ldots, x_d\},$$

defining sub-index d as $d = T + 1$.

The Sklar's theorem [42] shows that a multivariate joint distribution of random variables X_i can be described by the distribution function of its marginals $F_i(\cdot)$ and a copula $C(\cdot)$ for $i = 1, \ldots, d$. The copula models the dependency between the marginal uniform random variables $[U_1, \ldots, U_d] = [F_1(X_1), \ldots, F_d(X_d)]$. Formally, a function $C(\cdot) : [0,1]^d \rightarrow [0,1]$ is a copula described by

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

In general, $F_i(\cdot)$ is the transformation function from the smart meter measurement space $\mathcal{X}$ to a uniform space $\mathcal{U}$ on which the copula is modeled. The projection transformation is described as $\mathcal{X} \rightarrow \mathcal{U}$. Here, the marginal distributions are not assumed to belong to any parametric probability distribution model. Therefore, $F_i(\cdot)$ in (4) is the Probability Integral Transform (PIT) using an empirical distribution function (EDF), described as

$$F_{\Pi_i}(x_i) = \frac{1}{N+1} \sum_{n=1}^{N} \mathbb{1}_{\{x_{i,n} \leq x_i\}} \; \forall \, x_i \in \mathcal{D}, \tag{5}$$

where $\mathbb{1}$ is the indicator function.

In literature, there are multiple multivariate copula models available. The most common classes are the Archimedean, and the multivariate elliptical copulas. These last are derived from the MVG and MVT probability distributions. Archimedean copulas have only one or two parameters of dependence for their marginal distributions, limiting its applications for multivariate cases [43]. Alternatively, multivariate elliptical copulas offer the possibility to assign different values of dependence for all the pairs of random variables in (4), which is embedded in the correlation matrix of the multivariate elliptical functions. Due to this, multivariate elliptical copulas will be used here.

1) Multivariate Gaussian (MVG) Copula [44]: The MVG copula can be constructed based on (4) using a multivariate normal cumulative distribution function $\Phi_d(\cdot)$, with zero mean vector and correlation matrix $\mathbf{\Sigma} \in \mathbb{R}^{d \times d}$, described as

$$\begin{aligned} C(u_1, \ldots, u_d) &= \Phi_d(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d);\ \mathbf{\Sigma}) \\ &= \Phi_d(z_1, \ldots, z_d;\ \mathbf{\Sigma}), \end{aligned} \tag{6}$$

where $\Phi^{-1}(\cdot)$ is the inverse function of the univariate standard normal cumulative distribution. The inverse function can be seen as a projection from the uniform space to the standardized elliptical distribution space, i.e., $\mathcal{U} \rightarrow \mathcal{Z}$. The corresponding MVG copula density can be expressed as

$$c(u_1, \ldots, u_d; \mathbf{\Sigma}) = \frac{\mathcal{N}_d(z_1, \ldots, z_d;\ \mathbf{\Sigma})}{\prod_{i=1}^{d} \phi(z_i)}, \tag{7}$$

where $\phi(\cdot)$ is the univariate standard normal density distribution function, and $\mathcal{N}_d(\cdot; \mathbf{\Sigma})$ the multivariate normal density distribution function. The linear correlation between variables described by matrix $\mathbf{\Sigma}$ has a known relation with Kendall's tau [44], denoted by τ, which is a rank-based dependence measurement between variables x_i in the dataset ($\mathcal{D}$). This relation is described as

$$\rho_{(k,l)} = \sin(\frac{\pi}{2}\, \tau_{(k,l)}). \tag{8}$$

The subscript (k, l) describes the element position in the matrix $\mathbf{\Sigma}$. Therefore, the parameter estimation $\mathbf{\Sigma}$ for the MVG copula is given by the relation in (8), and it is referred as $\hat{\mathbf{\Sigma}}$.

2) Multivariate t-distribution (MVT) Copula [39]: Similarly, the MVT copula can be constructed using a multivariate cumulative t-distribution function $T_d(\cdot)$, with zero mean vector, scale matrix $\mathbf{\Sigma}$, and $\nu > 0$ degrees of freedom, described as

$$\begin{aligned} C(u_1, \ldots, u_d) &= T_d(T^{-1}(u_1; \nu), \ldots, T^{-1}(u_d; \nu);\ (\mathbf{\Sigma}, \nu)) \\ &= T_d(z_1, \ldots, z_d;\ (\mathbf{\Sigma}, \nu)). \end{aligned} \tag{9}$$

where $T^{-1}(\cdot; \nu)$ is the inverse cumulative distribution function of the univariate t-distribution with $\nu > 0$ degrees of freedom, and serves as the projection function $\mathcal{U} \rightarrow \mathcal{Z}$ for the MVT copula model. The MVT copula density is defined as

$$c(u_1, \ldots, u_d; (\mathbf{\Sigma}, \nu)) = \frac{t_d(z_1, \ldots, z_d;\ (\mathbf{\Sigma}, \nu))}{\prod_{i=1}^{d} t_\nu(z_i; \nu)}, \tag{10}$$

where $t_\nu(\cdot; \nu)$ is the univariate standard density, and $t_d(\cdot; (\mathbf{\Sigma}, \nu))$ the multivariate density t-distribution functions.

The relation in (8) for the parameter estimation of $\mathbf{\Sigma}$ can be extended to the MVT copula [45]. Therefore, $\hat{\mathbf{\Sigma}}$ is the same for the MVG and MVT copulas.

The parameter estimation of ν for the MVT copula is computed using a maximum pseudo-likelihood estimation (MPLE) [46], as the model uses a non-parametric approach over the marginal distributions. The MPLE maximizes the log-likelihood of the MVT copula density (10) over the N uniform pseudo-observations $u_{i,n}$, which are obtained applying the transformation (5) on the smart meter dataset ($\mathcal{D}$). The optimization problem to find the optimal $\hat{\nu}$ is defined as

$$\hat{\nu} = \arg\max_{\nu} \quad \prod_{n=1}^{N} c(u_{1,n}, \ldots, u_{d,n};\ (\hat{\mathbf{\Sigma}}, \nu)). \tag{11}$$

The model selection between MVG and MVT copulas that better describes the dataset $\mathcal{D}$ is made using the Bayesian information criterion (BIC), defined as

$$BIC \equiv -2 \ln(\ell(\mathcal{D}; \hat{\boldsymbol{\theta}})) + \ln(N)\ p, \tag{12}$$

where $\ell(\mathcal{D}; \hat{\boldsymbol{\theta}})$ is the log-likelihood of the multivariate elliptical copula on the dataset $\mathcal{D}$, $\hat{\boldsymbol{\theta}}$ are the fitted parameters which define the multivariate elliptical copula, and p is the number of parameters of the copula. The BIC balances the model goodness of fit, measuring the log-likelihood of the multivariate elliptical copula on the dataset penalizing the model complexity by the number of samples N, and the number of parameters on the model p. The model with the lowest BIC value is selected, meaning that the model is simpler, i.e., explaining the actual data with fewer parameters. The log-likelihood for both multivariate elliptical copulas densities are

$$\ell(\mathcal{D}; \hat{\boldsymbol{\theta}}) = \prod_{n=1}^{N} c(u_{1,n}, \ldots, u_{d,n};\ \hat{\boldsymbol{\theta}}). \tag{13}$$

Here, $\hat{\boldsymbol{\theta}}$ represents the correlation matrix $\hat{\mathbf{\Sigma}}$, for the MVG copula, or $(\hat{\mathbf{\Sigma}}, \hat{\nu})$, for the MVT copula. A summary of the

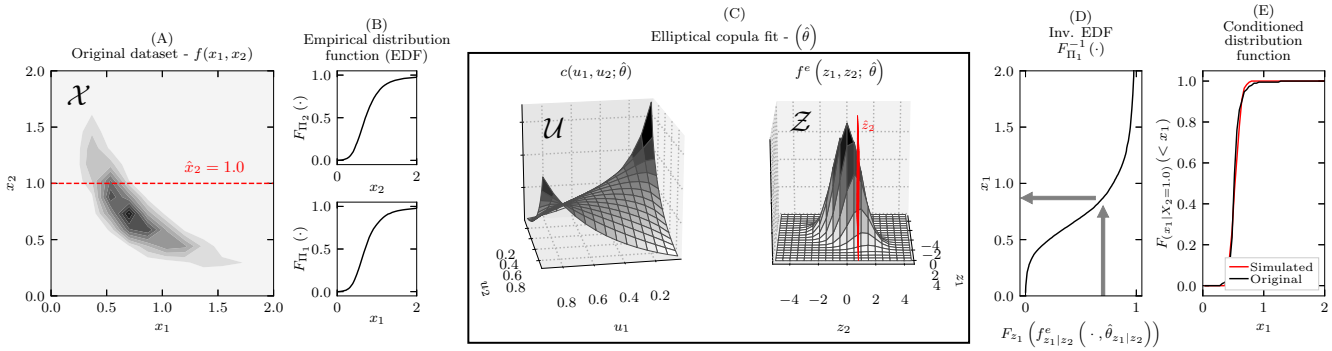

Fig. 2. Example of the conditional elliptical copula approach in a bivariate data set. (A) The original data set with a joint probability density $f(x, y)$ is conditioned in a value $\hat{x}_2 = 1.0$, depicted by the red dotted line. (B) The data set is projected to the uniform space $\mathcal{X} \to \mathcal{U}$ using the EDF. (C) The parameters $\hat{\theta}$ are fitted for a multivariate elliptical copula. The multivariate elliptical copula is conditioned in the $\mathcal{Z}$ space, using the projection of $\hat{x}_2 \to \hat{z}_2$ (vertical red line). (D) The conditioned copula is sampled and projected back to the actual units $\mathcal{Z} \to \mathcal{U} \to \mathcal{X}$. (E) The samples follow the original conditioned distribution function $F(x_1|X_2 = \hat{x}_2)$.

multivariate elliptical copula parameter fitting and model selection procedure can be found in Algorithm 1.

A visual overview of the multivariate elliptical copula fitting process discussed in this section can be seen in Fig. 2. Steps (A) to (C) show an example of the projection of a bivariate dataset to the uniform space using the PIT in (5). Step (C) shows the elliptical distribution in $\mathcal{Z}$ that fits the multivariate elliptical copula density in the uniform space $\mathcal{U}$. Sampling from the fitted elliptical distribution density in $\mathcal{Z}$ space, and projected back to the $\mathcal{X}$ space, creates RLPs with random annual energy consumption that follows the marginal distribution $F(w)$. In order to control the specific generation of RLP for a household with given annual energy consumption, the multivariate elliptical copula model needs to be conditioned. This conditioning can be done as discussed next.

Algorithm 1
Multivariate elliptical copula parameter fitting and model selection.

1) Transform the smart meter dataset from the smart meter measurement space $\mathcal{X}$ to the pseudo-observation uniform space $\mathcal{U}$, using (5).
2) Compute Kendall's tau $\tau_{(k,l)}$ between variables $x_{i,n}$ of the smart meter dataset $\mathcal{D}$.
3) Compute scale matrix $\hat{\mathbf{\Sigma}}$ using relation in (8).
4) Compute the numerical optimization in (11), fixing scale matrix $\hat{\mathbf{\Sigma}}$, and finding optimal $\hat{\nu}$.
5) Compute **BIC** for MVG and MVT copulas using $\hat{\mathbf{\Sigma}}$ and $\hat{\nu}$ in (12).
6) Select the multivariate elliptical copula model with the lowest **BIC**.

D. Conditioned Copula Model

Here, the expression defined in (2) is modeled to simulate RLPs for a household with a specific annual energy consumption $\hat{w}$. The following variable vectors notation is used: $\boldsymbol{x_1} = [x_1, \ldots, x_T]^\intercal$, $x_2 = w$ and $\boldsymbol{u}_1 = [F_{\Pi_1}(x_1), \ldots, F_{\Pi_T}(x_T)]^\intercal$, $u_2 = F_{\Pi_w}(w)$. To condition (2), the Sklar's theorem is extended to its conditional form as

$$F(\boldsymbol{x}_1|x_2) = C(\boldsymbol{u}_1|u_2) = C^e(\boldsymbol{z}_1|z_2;\ \hat{\boldsymbol{\theta}}_{1|2}), \qquad (14)$$

where $\hat{\boldsymbol{\theta}}_{1|2}$ is the conditioned parameters for the multivariate elliptical copula discussed in Section II-C, and $C^e(\cdot)$ refers to any of the multivariate elliptical copula models, i.e., either MVG or MVT copula. Based on this, to condition the multivariate elliptical copula model, the annual energy value $\hat{w}$ should be projected from the smart meter space to the elliptical distribution space $\mathcal{X} \to \mathcal{Z}$. Fig. 2, step (C), shows and example of this projection, which is represented as a red vertical line. The projection $\mathcal{X} \to \mathcal{U} \to \mathcal{Z}$ depends on the copula model selected by the BIC and is expressed as

$$\hat{z}_{\hat{w}} = \begin{cases} T^{-1}(F_{\Pi_w}(\hat{w}); \hat{\nu}) & \text{if } C^e(\cdot) \text{ is a MVT copula,} \\ \Phi^{-1}(F_{\Pi_w}(\hat{w})) & \text{if } C^e(\cdot) \text{ is a MVG copula.} \end{cases} \qquad (15)$$

The annual energy value projected condition the elliptical distribution parameter to $\hat{\boldsymbol{\theta}}_{1|2}$, using $Z_2 = \hat{z}_{\hat{w}}$. The details of parameter conditioning can be found in the Appendix. The conditioned elliptical distribution function is sampled to generate $\hat{\boldsymbol{z}} = [\hat{z}_1, \ldots, \hat{z}_T]^\intercal \in \mathbb{R}^T$, which should be projected to active power units in order to obtain an RLP. The projection $\mathcal{Z} \to \mathcal{U} \to \mathcal{X}$ is done by

$$\hat{x}_i = \begin{cases} F^{-1}_{\Pi_i}(T(\hat{z}_i; \hat{\nu})) & \text{if } C^e(\cdot) \text{ is a MVT copula,} \\ F^{-1}_{\Pi_i}(\Phi(\hat{z}_i)) & \text{if } C^e(\cdot) \text{ is a MVG copula.} \end{cases} \qquad (16)$$

An example of the projection (16) in the bivariate case can be visualized in Fig. 2 in step (D), indicated by the arrows. All the simulated values follow the original conditioned distribution function, as shown in (E). A summary of the steps to simulate RLPs from the conditioned model is described in Algorithm 2.

Algorithm 2
Profile simulation from the conditional multivariate elliptical copula.

1) Project the annual energy value ($\hat{w}$) to the multivariate elliptical copula function space $\mathcal{Z}$ using (15).
2) Condition the parameters of the multivariate elliptical copula $\hat{\boldsymbol{\theta}}_{z_1|Z_2=\hat{z}_{\hat{w}}}$ using (20) and (21).
3) Draw N samples from the conditioned elliptical distribution $\{\hat{\boldsymbol{z}}_n\}_{n=1}^N$.
4) Transform the N samples into power units using (16).

TABLE I
SMART METER DATASETS FOR THE CASE OF STUDY

Country	Total houses	Annual Energy consumption [MWh/year] (min) - (max)	Time Resolution
Netherlands (NL) [47]	77	(1.00) - (11.17)	15 min.
United States (USA) [48]	25	(6.29) - (19.62)	15 min.
United Kingdom (UK) [49]	300	(0.03) - (8.10)	30 min.
Australia (AUS) [50]	300	(1.16) - (8.89)	30 min.

III. CASE OF STUDY

In order to assess the effectiveness of the proposed copula models, an assessment is performed by comparing the statistical properties discussed in Section II for actual RLPs measurements and simulated RLPs for different case of study. Additionally, simulations of RLPs obtained from the conditional GMM are used as a benchmark. These cases of study consists of four applications to analyze the performance of the Algorithms 1 and 2, described as:

1) Individual and aggregated residential data modeling at 15-minutes resolution. Results are presented in terms of parameter estimation analysis and model selection (Sections III-A and III-B).
2) Modeling RLPs at 15, 30, and 60-minute resolutions (Section III-C).
3) Modeling over different consumption load profile patterns, testing for multiple smart meter datasets from different countries (Section III-D). The open dataset sources used for the different tests in this case study are summarized in Table I.
4) The conditional copula modeling is extended to include weather variables i.e., solar irradiance and temperature. The conditional elliptical copula's flexibility to model daily power consumption profiles under different daily weather conditions is analyzed (Section III-E).

The first test is divided in two cases: (i) aggregated RLPs, which represents MV/LV distribution transformer loading for residential areas; and (ii) individual RLPs, which describes individual household consumption. RLPs for the two cases have different correlation characteristics, and the purpose is to evaluate the effectiveness of the methodology in such scenarios. The individual case data set corresponds to smart meter measurements of active power consumption for 77 households in the Netherlands (NL) [47], with a 15 min resolution for one year. The data set for the aggregated case consists of 100 MV/LV distribution transformers, with the same time resolution and period as the individual case.

Figure 3 shows the actual daily RLPs for the weekdays in June. The top left plot shows the aggregated RLPs, which has a range of annual energy consumption between 123 and 160 [MWh/year]. The top right plot shows the individual RLPs with has annual energy consumption between 11.17 and 1.00 [MWh/year]. The bottom row of Fig. 3 shows

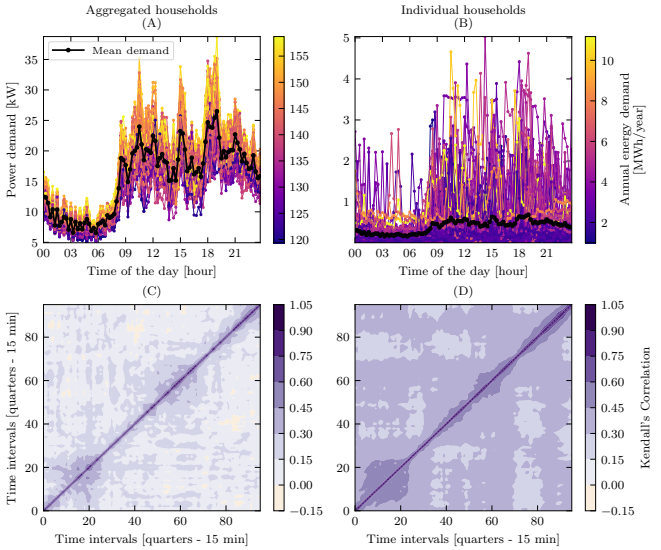

Fig. 3. RLPs for the weekdays in June. The first column shows the aggregated RLP, and the second column shows the individual RLP. The bottom row shows Kendall's tau correlation matrix in a heat map visualization.

a heat map of Kendall's tau correlations coefficients between variables (time steps) of the day. The heat map shows that the covariance between subsequent time steps is stronger in the individual case, which varies between (0.52 - 0.75), compared to the aggregated case that varies between (0.4 - 0.63). Also, the correlation values vanish quickly in the aggregated case for shorter time windows, compared to the individual case, which has correlation values of 0.6 to 0.75 for one hour apart. Based on this, we can see that aggregated and individual cases have different correlation structure and uncertainties.

RLPs for both cases (aggregated and individual) have an intrinsic seasonal trend during the year and have different consumption patterns between weekdays and weekends. Hence, for this case of study, the datasets are split into separate disjoint groups, dividing them into weekdays and weekends, and for each month of the year, creating 24 smaller datasets for each case. Thus, the copula models and the GMM used for the benchmark are fit for each of these datasets to simulate a daily RLPs for different months of the year. A bootstrap technique [51] was applied for building the models and test the performance evaluation. To this end, 70% of the dataset was used for parameter estimation and model selection, while the remaining 30% was reserved as the *original* dataset for evaluation purposes. The bootstrapping was repeated 1000 times to compute the probability distance metrics named the Energy (ED), Kolgomorov-Smirnov (KS), and Wasserstein distances (WD). Algorithms 1, 2, and conditional GMM were implemented in Python 3.8 and run on an Intel i7 @2.8 GHz PC with 8 cores and 32 GB of Memory.

A. Parameter Estimation and Model Selection

The results presented here are related to the modeling procedure of the expression in (1) using Algorithm 1. Additionally, the building procedure of the GMM based on [26] is also presented. The GMM is used for comparison purposes.

Figure 4 shows the parameter estimation results for the GMM and the multivariate elliptical copula models for the

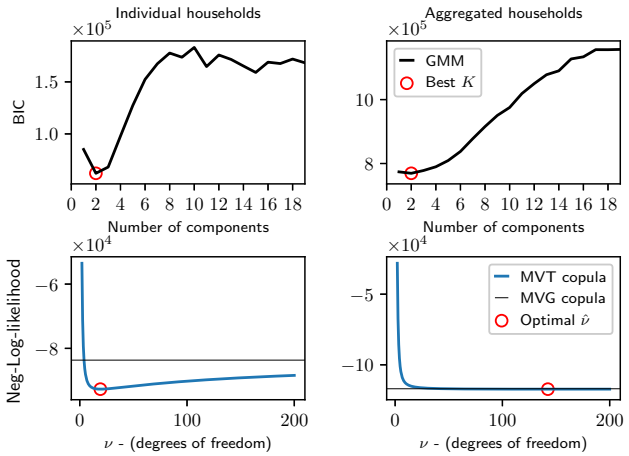

Fig. 4. Parameter fitting for the GMM and multivariate elliptical copula models. The top row is the best number of components for the GMM. The bottom row shows the negative likelihood values for the multivariate elliptical copulas. The first column is for individual residential consumption, the second column is for aggregated residential consumption.

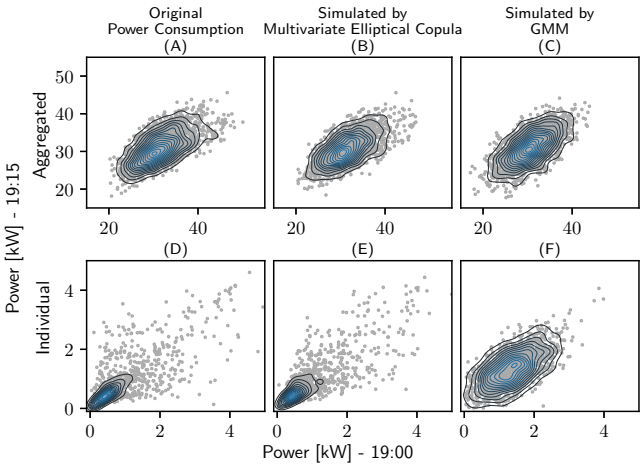

Fig. 5. Comparison between power values of the original dataset and N=1500 simulations for GMM and multivariate elliptical copula models, for the time step transition between 17:00 and 17:15 a weekday in November. The top row is the aggregated case, the bottom row is the individual case.

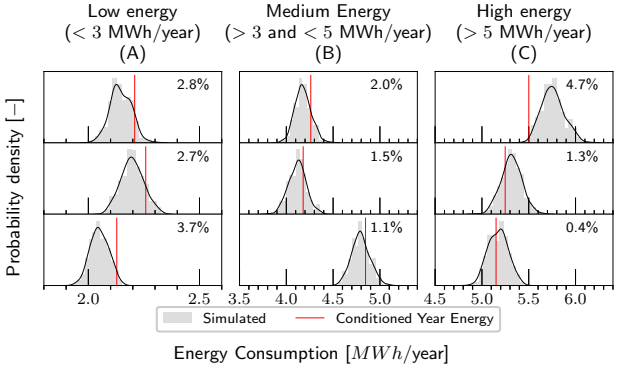

Fig. 6. Simulation results of the conditioned multivariate elliptical copulas for nine households randomly selected and grouped by different annual energy consumption ranges. The red vertical line represents the annual energy consumption value were the models are conditioned. Percentage shows the error between the mean of the simulations and the conditioned energy value.

weekdays of November. The top row shows the number of components for the GMM, defined after running the expectation-maximization algorithm using 1 to 20 components and deciding the optimal number using the BIC. As can be seen in Fig. 4, the optimal number of components that best describes the data set is equal to $K = 2$, for both the individual and aggregated case. The bottom row of Fig. 4 shows the multivariate elliptical copula parameter estimation. In the individual case, the negative log-likelihood curve for the MVT copula has a minimum in $\hat{\nu} = 19.38$ with a negative log-likelihood value lower than the MVG model. The BIC value of the MVT copula was also computed, giving a value of $-151\,376$, and the BIC for the MVG copula is $-133\,325$, selecting the MVT copula model for the individual case. Results show similar results for the rest of the months, where the MVT copula is selected for all individual cases.

In the aggregated case, which is shown in the second column of Fig. 4, the negative log-likelihood of the MVT copula has a flatter behavior than the individual case; this trend is also seen for the rest of the months. The MVT copula has an optimal $\hat{\nu} = 144.44$ with BIC of $-188\,069$, and the BIC of the MVT copula is $-187\,524$, which is a difference of BIC of less than 0.3%. It should be recalled that as $\nu \to \infty$, the MVT distribution tends towards an MVG distribution, which means that both copula models are almost identical. Even though the BIC in the MVT copula is lower than the MVG copula, there is no substantial difference between both types of copula models when the degrees of freedom is high [52]. Results show that for values of $\nu > 200$, both elliptical copula models are indistinguishable.

In order to visually assess the capability to reproduce the complex correlations seen in Fig.1, all the fitted models are sampled to simulate RLPs for both cases. The results are presented in Fig. 5 for the time step transition between 17:00 and 17:15 for one weekday in November. For the individual case, the dependency structure between time steps on the original smart meter measurements (D), can be modeled by the multivariate elliptical copula (E). However, the GMM has a poor representation due to the restriction to fit only Gaussian-shaped distributions (F). This difference highlights the flexibility of the copulas, which can model complex dependence structures seen on RLPs. For the aggregated case, a simpler correlation is seen in the original dataset (A), with a more Gaussian-like distribution. In this case, the multivariate elliptical copula and GMM perform similarly.

B. Simulations with Conditioned Multivariate Elliptical Copula Models

The results presented here are related to the modeling procedure of expression in (2) using the conditioned multivariate elliptical copula model shown in Algorithm 2. Additionally, the building procedure of a conditioned GMM based on [41] is also presented. To quantify the differences between the active power values simulated by the developed copula models and the original dataset, the Energy Distance (ED) [53] and Kolgomorov-Smirnov (KS) distances are used as a probability distance metric.

The multivariate elliptical copula and GMM models are conditioned to each household's annual energy consumption and transformer for the individual and aggregated case, respectively. Simulations are executed for $N = 300$ annual

TABLE II
SUMMARY OF RESULTS FOR INDIVIDUAL RESIDENTIAL DATA - NETHERLANDS DATASET

		Energy Distance		Kolgomorov-Smirnov Distance		Autocorrelation Root Mean Squared Error (RMSE) [%]	
Day of the week	Season	Conditional Mult. elliptical copula	Conditional GMM	Conditional Mult. elliptical copula	Conditional GMM	Conditional Mult. elliptical copula	Conditional GMM
Weekday	Winter	0.012	0.401	0.016	0.342	3.26	5.71
	Spring	0.016	0.357	0.020	0.327	2.80	4.66
	Summer	0.015	0.364	0.024	0.369	2.38	5.64
	Autumn	0.017	0.227	0.022	0.256	2.99	7.52
Weekend	Winter	0.031	0.477	0.024	0.393	3.93	3.15
	Spring	0.036	0.301	0.031	0.280	3.40	5.09
	Summer	0.024	0.510	0.030	0.457	2.69	6.50
	Autumn	0.029	0.313	0.022	0.287	3.70	6.26

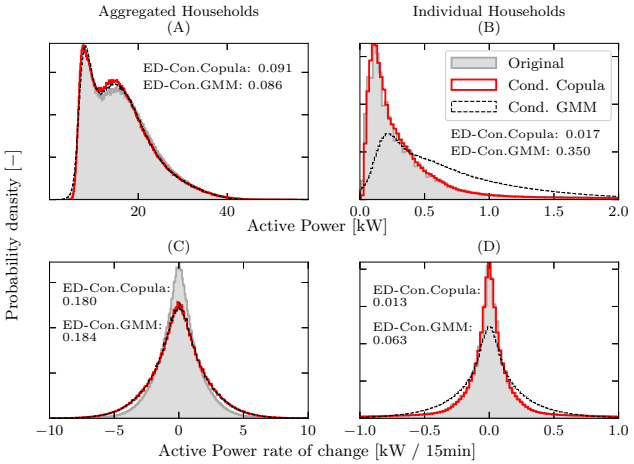

Fig. 7. Density distributions and energy distance metrics of the active power consumption every 15 minutes, for one year for the original data set, and the simulated profiles using conditional multivariate elliptical copula and conditional GMM.

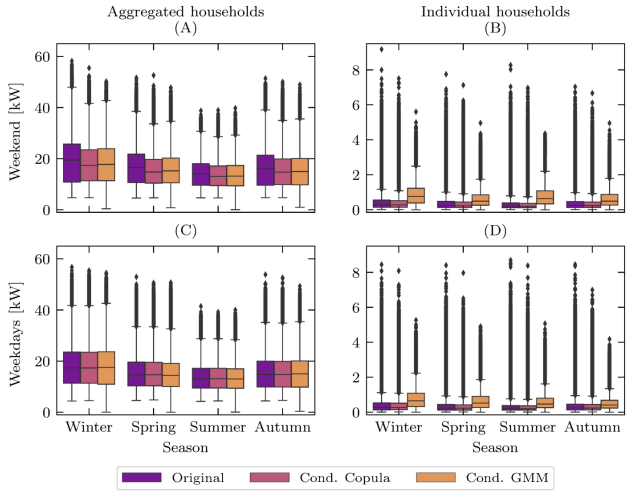

Fig. 8. Load consumption of one year split into seasons, weekends, and weekdays. The conditional multivariate elliptical copula can model the original data set for both aggregated and individual residential consumption cases successfully.

scenarios. Fig. 6 shows simulation results with the elliptical copula models for nine randomly selected households with different annual energy consumption values. The average errors between the conditioned annual electricity consumption values and the mean of the simulations are 4.9%.

The top row in Fig. 7 shows the distribution density of active power values for all the houses (individual case) and transformers (aggregated case) for a year. The results in (A) and (C) show that both techniques have similar performance in the aggregated case, with maximum ED differences of just 4.7%. However, for the individual case in (B) and (D), the conditional copula performs better, improving the ED by 20 times, which means that our proposal outperforms the GMM significantly.

The original dataset and simulation results are split into seasons, weekdays, and weekdays. The results are shown in Fig. 8. The box plots in (A) and (C), which are for the aggregated case, shows that the RLPs generated by the conditional elliptical copula has similar distributions based on the first and third quantiles. However, the GMM tends to underestimate the lower power consumption in the aggregated case, which can be seen in the lower whiskers. For the individual case, (B) and (D), the proposed model outperforms the GMM for all seasons. In (D), the GMM underestimates the high power consumption, simulating just 70% of the highest consumption values, which can be seen in the boxplot's upper fliers. This could be critical for network planning studies, as the equipment could be undersized, resulting in lower grid security. The differences between the density distributions for Fig. 8 are summarized in Table II. Based on the distance metrics values, the multivariate elliptical copula models RLPs accurately, outperforming the GMM by one order of magnitude.

The simulated profiles for the weekdays in June are displayed in Fig. 9. (A) and (B) show that the conditional elliptical copula generates profiles that keep household consumption behavior volatility. The conditional GMM is more conservative, underestimating the high power consumption spikes, as discussed in Fig. 8. Subplots (C) and (D) of Fig. 9 show a heatmap of Kendall's correlation matrix of both models' simulated profiles. The heatmaps should be similar to the original shown in Fig. 3 (D), similar structure means that models can capture the correlation between time steps in the generated profiles. The correlation matrix of the conditional elliptical copula shows similarity to the correlation matrix of the original dataset, with a mean difference of 4%. In contrast, the conditional GMM has a poor structure and a maximum difference of 60%.

It is also important to note that a comparable correlation matrix does not necessarily imply a similar probability distribution for each time step transitions. A two-dimensional Wasserstein distance is computed to quantify the similarity of the probability distributions of time step transitions between the original and simulated datasets. The tests were carried out, showing the results as a heatmap in the subplots (E) and (F) in Fig. 9. The color bar scales for the WD heatmaps show that the conditional elliptical copula is almost one order of magnitude smaller than the conditional GMM. The largest WD values

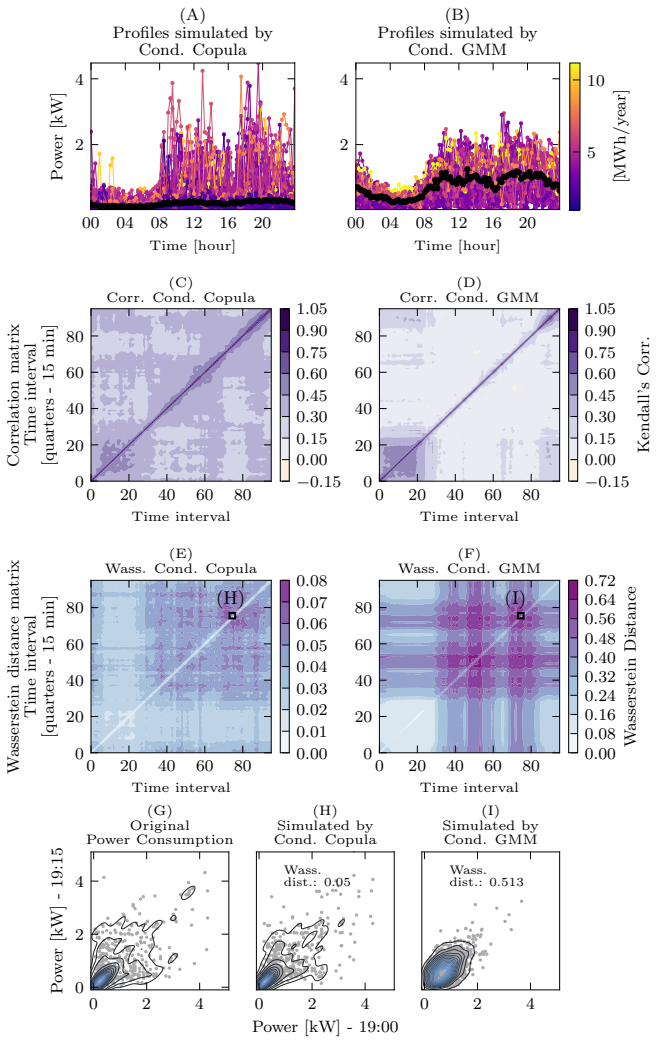

Fig. 9. Simulation results of the conditional multivariate elliptical copula and the conditional GMM for the 15-min resolution models over the NL dataset. (A)-(B) daily profiles simulated from both models. (C)-(D) Kendall's correlation matrix on the simulated profiles. (E)-(F) Show the 2-D Wasserstein distance between the simulated profiles and the original profiles, (G)-(I) Highlight one example of the cross-plot that shows a time transition used to compute the Wasserstein metric.

are 0.074 and 0.688 for the conditional elliptical copula and conditional GMM, respectively. The WD heatmaps highlight one example of active power transition between 19:00 and 19:15, for the original dataset (G), simulated by conditional copula (H) and conditional GMM (I). In (H) and (I), in the upper right corner is shown the WD metric for the specific time step transition on the heatmap. The conditional elliptical copula can simulate the consumptions seen in the tails of the original dataset, e.g., consumptions above 4 kW, which agree with the findings from Fig. 8.

The daily consumption profile is a time series that can be characterized by an autocorrelation plot. Figure 10 shows the autocorrelation signals' averaged value for the weekdays in June between all the houses in the NL dataset for the original and simulated RLPs. The autocorrelation plot shows how is the dependency structure of consumption between the current and past demand values. The plot indicates that the past 20 and 10 times steps are the most significant values for the

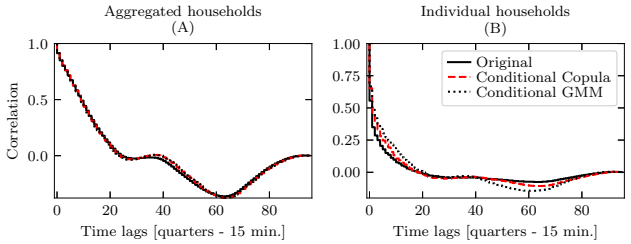

Fig. 10. Averaged daily autocorrelation plots for the aggregated case (A), and the individual case (B) for one year of data.

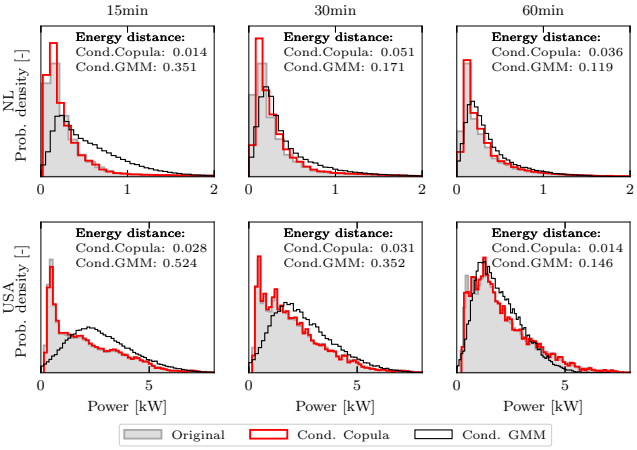

Fig. 11. Probability distribution comparison between original dataset and simulated power profiles with Cond. Copula and Cond. GMM, for profiles with different time resolutions

aggregated and individual cases, respectively. It is also shown that all the simulated values, e.g., the 96-time step vector sampled from the probability models, of the simulated RLPs from the conditional elliptical copula models have the same time series structure of the original profiles for the aggregated and individual cases. The root mean squared error (RMSE) is used to quantify the similarity of the simulated RLPs and the original datasets. The RMSE values in Table II show that the conditional copulas have an average error of only 3.1%, and the GMM models are 5.6%, meaning that the proposed model almost halved the error.

C. Modeling at Different Time Resolutions

Smart meters deployed in field can gather data at different temporal resolutions, e.g., 15, 30, and 60 minutes. For the second case of study, this subsection analyses the model performance at different time resolutions using the NL and USA datasets. Both datasets originally consists of energy data (Wh) at 15 min resolution. An down-sampled for 30 and 60 minutes is done, using a sum of the active power of the corresponding time intervals and then converted to power units (kW).

Figure 11 shows the active power consumption in June for both datasets at different time resolutions. The conditional elliptical copula shows a consistent small ED across time resolutions. The conditional GMM shows an improvement when the resolution is decreasing, reducing the ED by 70%. Two observations are noticed at lower resolutions: The volatility of

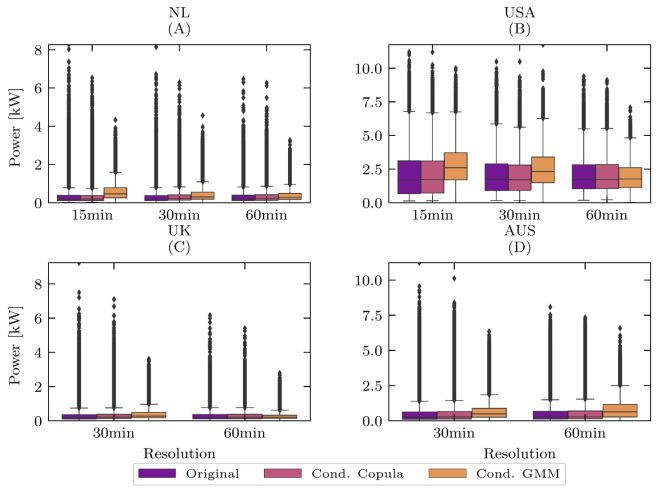

Fig. 12. Box plot comparison between original datasets and simulated active power consumption for different datasets and time resolutions, for the smart meter readings in Fig. 13

power peaks is reduced (lower probability distribution tails), and the median of consumption is shifted to higher values. The conditional GMM has an improvement when the load profile is less uncertain. This can also be seen in Fig. 12, subplots (A) and (B), on which the first and third quantiles of the boxplot for all the models have similar values at 60 minutes resolution. Nevertheless, the conditional elliptical copula shows a consistent good behavior on modeling the high consumption peaks for all time resolutions.

Simulated profiles with the proposed conditioned multivariate elliptical copula at higher resolutions, e.g., 1 and 5-minutes, showed an accuracy decay. At higher data frequencies, an actual RLP shows a "squared-wave" time-series profiles [54] since the use of home appliances is more evident, i.e., it is visible with the devices on/off switching. Additionally, at higher time resolutions, the time of use of the home appliances becomes an important variable. Bottom-up modeling approaches, discussed in Sec. I, can capture such dynamic behaviors of the household profile for relevant demand-response applications.

D. Modeling Different Smart Meter Datasets

The third case of study is analyzed in this subsection, where the effectiveness of the model is quantified for different consumption habits across countries. Fig. 13 from rows (1) to (4) in column (A) shows the difference in profiles for one day in June for the smart meter measurements in Table I. The USA dataset has households with the highest energy consumption per year and has the most evident pattern showing a peak consumption around 6:00 p.m. More volatile patterns are seen in the AUS readings, with higher power values than the NL and UK, which can be observed in Fig. 12. It should be noted that these high consumption peaks are explained because of the winter season (in June) in AUS.

All profiles have a different correlation dependency structure, as shown in Fig. 13 from rows (5) to (8) in column (A). The profile simulations and the correlation matrices can be seen in columns (B), (D), (E), and (F). For all datasets, the

TABLE III
PROBABILITY DISTANCE METRICS FOR ALL DATASETS

		ED		KD		WD	
Resolution	Country	Cond. elliptical copula	Cond. GMM	Cond. elliptical copula	Cond. GMM	Cond. elliptical copula	Cond. GMM
15 min.	NL	0.014	0.351	0.024	0.352	0.038	0.376
	USA	0.028	0.524	0.022	0.261	0.242	1.187
30 min.	NL	0.051	0.171	0.084	0.231	0.062	0.173
	USA	0.031	0.352	0.016	0.184	0.217	0.764
	UK	0.032	0.158	0.058	0.208	0.041	0.197
	AUS	0.030	0.243	0.044	0.261	0.085	0.412
60 min.	NL	0.036	0.119	0.058	0.190	0.052	0.263
	USA	0.014	0.146	0.014	0.067	0.198	0.500
	UK	0.016	0.048	0.030	0.091	0.033	0.186
	AUS	0.022	0.218	0.030	0.168	0.080	0.368

MVT copula is selected by Algorithm 1 as the best model. In general, the conditional elliptical copulas closely replicates the correlation structure with a mean error over all correlation matrices of 6.8%, while the conditional GMM has 10.9%. For the same month, the profiles are down-sampled to 60 minutes resolution, using the same procedure as in Sec. III-C, and the results are shown in the columns (D) to (F) in the same Fig. 13. The up-sampling has a smoothing effect on the correlation heatmaps and reduces the volatility of the load profiles. The underestimation of peaks from the conditional GMM can be seen in the simulation profiles of UK, i.e., subplots (3C) and (3F), in which power values above 2 kW are rarely seen. All probability distance metrics are summarized in Table III, which shows that both models perform better at a lower resolution. Nevertheless, the conditional copula keeps the best scores by one order of magnitude for all the datasets at different resolutions. Finally, Fig. 12 shows that the conditional elliptical copulas models can simulate all the range of power peaks for all 60-minutes resolution cases.

E. Modeling Including Weather Variables

The previous subsections focused on the modeling of the residential profiles conditioned to specific annual energy consumption. Load consumption profiles' changes due to weather factors such as temperature and irradiance, were implicit in the modeling when the datasets were split into 24 disjoint groups. This was done to cope with the seasonal changes during a year. This subsection extends the model in (1) to consider the weather variables into one explicitly joined dataset modeling. To accomplish this, the multivariate copula modeling now includes the continuous random variables irradiance, $Q \in \mathbb{R}^r$, and temperature, $O \in \mathbb{R}^s$, with random realizations q and o:

$$F(x_1, \ldots, x_T, w, q_1, \ldots, q_r, o_1, \ldots, o_s), \tag{17}$$

where r and s represent the index of the time step discretization of the irradiance and temperature profiles, respectively. e.g., for 1-hour resolution of temperature data $s = 24$. The dataset in (3), used to compute the extended model in (17) with Algorithm (1), is also extended to consider meteorological measurements, which were collected at the same time as the active power consumption measurements. Thus, the model (17) is then conditioned using Algorithm 2, based on the energy, temperature, and irradiance variables, such as

$$F(x_1, \ldots, x_T, |W = \hat{w}, \boldsymbol{Q} = \hat{\boldsymbol{q}}, \boldsymbol{O} = \hat{\boldsymbol{o}}), \tag{18}$$

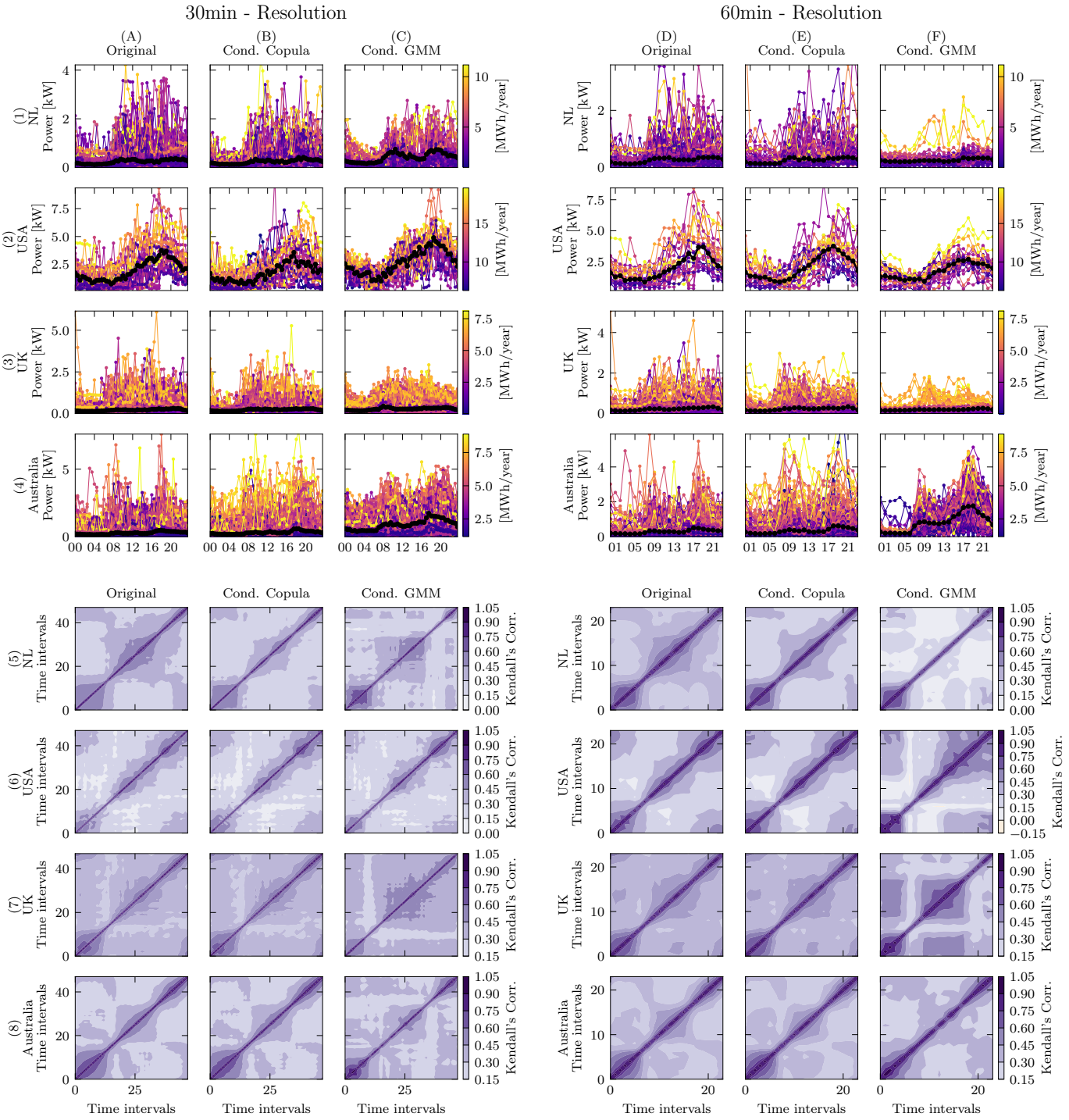

Fig. 13. Simulation results for the smart meter datasets of Table I. Columns (A) and (D) show the original load profiles from one day in June at 30 and 60 minutes resolution. Thick black lines show the median RLP for each dataset. Simulations from the conditional elliptical copula models are shown in columns (B) and (E). Simulations form cond. GMM are in columns (C) and (F). The heat maps of the correlation matrix from (1) to (5) show correlation structures, meaning different consumption profiles patterns from each dataset.

where $\hat{\boldsymbol{q}} = (\hat{q}_1, \ldots, \hat{q}_r)$ is the daily profile of irradiance, and $\hat{\boldsymbol{o}} = (\hat{o}_1, \ldots, \hat{o}_s)$ is a daily profile of temperature. To validate this extended model, an original dataset is used consisting of active power readings of 97 distribution transformers, 71 smart meters (both at 15-minute resolution), and meteorological measurements (at 1-hour resolution) for one year. The number of solar irradiance variables was reduced to those time steps with significant sunlight (8:00 - 18:00). Therefore, the number of dimensions for irradiance variables is $s = 10$. Thus, the models in (17) and (18) have a total of 131 variables.

Figure 14 summarizes the conditioned copula model's simulation results, compared to the original dataset, for multiple scenarios using the bootstrapping method. Due to the model's high dimensionality and heterogeneity of the dataset, the

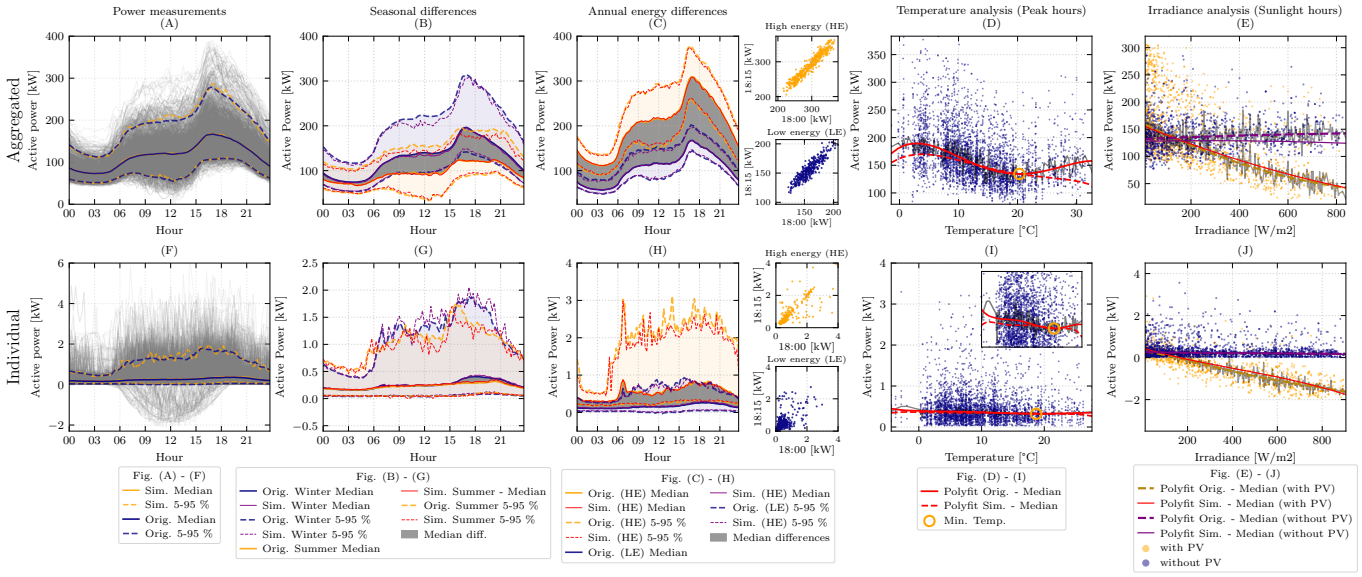

Fig. 14. Summary of the analysis of the aggregated and individual daily profiles changes seen in (A) and (F), considering different factors such as: Winter and summer seasons (B) and(G), high and low annual energy consumption (C) and (D), temperature change during peak hours, i.e., 16:00 - 19:00, (D) and (I), and active power change due to irradiance during sunlight ours (8:00 - 18:00) for meters with and without PV installations (E) and (J). The datasets correspond for loading profiles from January until July. The red dashed lines in all subplots are the quantiles' results from the simulations from the conditional elliptical model that considers weather variables as shown in (17).

conditional GMM could not be computed for this case. In Fig. 14, the meter readings on (A) and (D) have combined measurements of service areas (aggregated level) and households (individual level) with and without PV installations. The different profiles could have different correlations with the irradiance variable, e.g., in sunlight hours, reverse power flow may exist into the grid, visible as a duck curve in the profile. Figure (B) and (G) segregates the profiles based on seasons for winter and summer. The differences over the median between seasons are highlighted in grey. These differences are assumed to be mainly caused by temperature changes and global irradiance profiles between both seasons. A significant difference is seen in the peak hours for the aggregate and individual cases, while the consumption remains nearly the same in the nighttime. This means that changes in weather conditions do not equally change the load profile in a linear way for all the time steps in the daily profile. For the aggregated case, the sunlight hours also significantly differ in the medians due to the transformers with high PV penetration. The maximum error between the medians of the simulated profiles and the original dataset is 4%.

The subplots (C) and (H) on Fig. 14 also highlight in grey the median differences between measurements coming from service areas and households with high and low annual energy consumption. It is clear from these subplots that the annual energy consumption (w) plays a notable role in changing the loading profiles, affecting all time steps during the day, compared to the seasonal factor. In order to further analyze the impact of temperature in the load profiles, subplots (D) and (I) show the active power consumption from peak hours (16:00 to 19:00) versus temperature, which are the most affected hours according to subplots (B) and (G). The power consumption is inversely proportional to the temperature until a minimum point, depicted by an orange circle in subplots (D) and (I), in which the power consumption starts to be directly proportional. This behavior could be attributed to installed cooling systems, on which the minimum point could be the average of the cooling devices' temperature setpoints.

The simulated power at higher temperatures starts to diverge (dashed red line on subplots (D) and (I)). The divergence is explained by the fact that there is a bimodal behavior in the correlation between the two marginals, e.g., at 18:00, the power could have the same power consumption based on two completely different temperatures. One mode is negative, and the other is positively correlated. This means that for high temperatures, e.g., heat waves, a single conditional elliptical copula could underestimate the power consumption, limiting the simulated scenarios under those circumstances. The conditioned elliptical copula model has an error of 5% between the medians of the original and simulated profiles for the temperature analysis.

The changes in power due to the global irradiance are shown in subplots (E) and (J) for the serviced areas/households with and without PV installations. The simulation has a maximum error of 7% of the aggregated case with no PV subplot (E). The simulations from the model diverge at high irradiance due to the same reason of the temperature analysis. Irradiance and temperature are highly correlated, e.g., on Fig 15, conditioned weather column. Meaning that the temperatures are also high at higher irradiance values, on which the cooling systems start to work, and consumption climbs up; this creates a slight divergence at the end between the original and simulated medians. The effect of bimodality is reduced by the active power measurements affected by the PV installations. Higher irradiance values mean higher PV energy production, which lowers the active power consumption from the grid. The model over the data with PV installations has a maximum error of 1.1%.

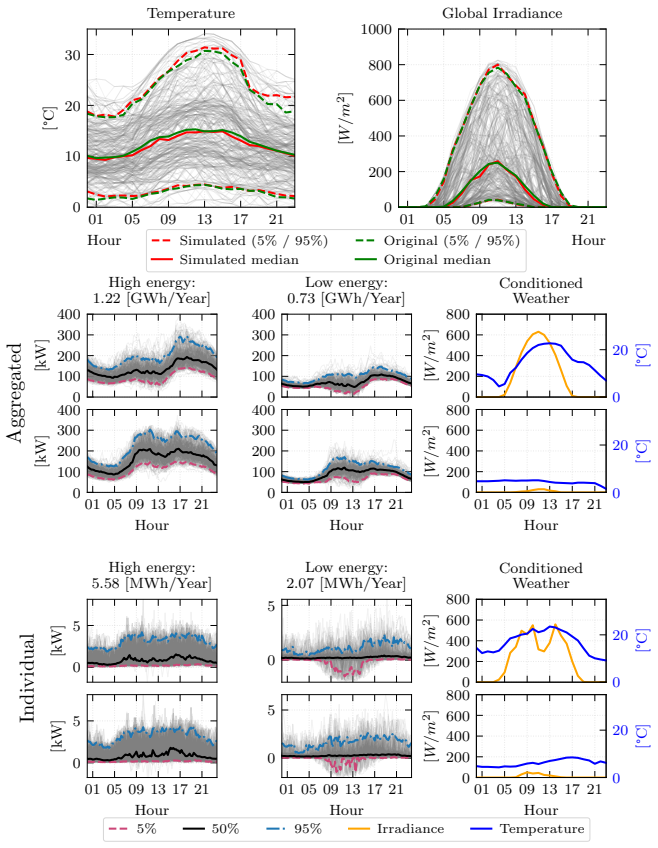

Fig. 15. Simulations of the weather variables from model (17) (top row). Active power simulation results from the conditioned elliptical copula model in (18), for a day in winter and summer at two different annual energy consumption values, for the aggregated and individual case (mid and bottom row).

Figure 15 shows the simulation results for a year of the weather variables from the generative model in (17), on which the 0.05, 0.5, and 0.95 percentiles are overlayed between the original and simulated profiles. The maximum difference on the median is about 1.5% for temperature and 2.3% for irradiance. The mid and bottom row on Fig. 15 shows 300 scenarios for active power consumption with the conditioned model in (18) for one a sunny in summer (high irradiance) and one day in winter (low temperature), for the cases of a household/serviced area with high and low energy consumption. In the aggregated case, it is observed the increase of consumption at lower temperatures and the decrease of the duck curve during the sunlight hours when the irradiance is low. In the individual case, the low annual energy consumption household with PV installation shows a higher generation in sunlight hours in summer than winter. In the case of the high annual energy household, the median of the consumption increases due to the lower temperatures. The simulation examples show that the conditional elliptical copula model can generate profiles consistent with multiple weather conditions.

It should be emphasized that the model in (17) could have both weather profiles at different time resolutions. Suppose the weather variables are increased due to a higher sampling resolution, e.g., 15-min resolution. In that case, the model's dimensionality increases, which can cause an ill-conditioned covariance matrix $\hat{\boldsymbol{\Sigma}}$. A numerical approximation for the nearest correlation matrix can be used to overcome the problem [55] but could potentially decrease the accuracy of the conditional elliptical copula modeling.

IV. Conclusion

In this paper, a new top-down approach based on multivariate elliptical copulas was presented. The proposed approach builds a probabilistic model that is able to capture the statistical properties of any smart meter measurement data set. The model is used to simulate RLPs specifying different annual energy consumptions and different daily weather conditions of temperature and solar irradiance.

Different from conventional top-down approaches based in Markov models, the proposed model does not require active power consumption discretization for each time step. Additionally, a benchmark against a GMM for two cases: aggregated and individual consumption, was also presented. Results showed that the GMM had a fair representation of the true probability distribution of the smart meter dataset at the aggregated level. However, the heteroskedastic dependency structure seen for individual RLP makes the GMM technique less flexible for modeling individual households. Due to this, the multivariate elliptical copula outperforms the GMM in one order of magnitude in the Energy and Kolgomorov-Smirnov distance metrics, and also was found to be 1.8 times better on the RMSE metrics for the autocorrelation plots. On the aggregated case, special preference was seen for the conditional MVG copulas, different from the individual case that the conditional MVT copula models had better fit. Five different smart meter datasets at different time resolutions had been tested, showing the general application of the presented algorithms. Finally, the proposed model is fully flexible in order to capture and simulate the complex correlations and changes caused by temperature and irradiance fluctuations to the daily profiles' time steps.

Appendix

The conditioned elliptical distributions functions are defined using the following notation:

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad (19)$$

where $\{\boldsymbol{X}, \boldsymbol{\mu}\} \in \mathbb{R}^{d_1+d_2}$ for $d_1 + d_2 = d$, and block matrices $\Sigma_{11} \in \mathbb{R}^{d_1 \times d_1}$, $\Sigma_{22} \in \mathbb{R}^{d_2 \times d_2}$, $\Sigma_{12} = \Sigma_{21}^{\intercal} \in \mathbb{R}^{d_1 \times d_2}$.

An elliptical density distribution function $f^e(\cdot, \theta)$, named MVT distribution or MVG distribution, conditioned as $\boldsymbol{X}_1$ given $\boldsymbol{X}_2 = \boldsymbol{x}_2$ is another elliptical distribution $f^e(\cdot; \boldsymbol{\theta}_{1|2})$, defining $\boldsymbol{\theta}_{1|2} = (\boldsymbol{\mu_{1|2}}, \boldsymbol{\Sigma}_{2|1})$ for the MVG distribution and $\boldsymbol{\theta}_{1|2} = (\boldsymbol{\mu_{1|2}}, \boldsymbol{\Sigma}_{2|1}, \nu_{1|2})$ for the MVT distribution. The conditional mean vector and covariance matrix are

$$\begin{aligned} \boldsymbol{\mu}_{1|2} &= \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2) \\ \boldsymbol{\Sigma}_{1|2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \end{aligned} \quad (20)$$

For the conditional MVT distribution, the conditioned mean $\boldsymbol{\mu}_{1|2}$ is the same as the MVG distribution. The conditioned scale and degrees of freedom for the MVT are given by

$$\begin{aligned}\boldsymbol{\Sigma}_{1|2} &= \frac{\nu + (\boldsymbol{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_{22}^{-1} (\boldsymbol{x}_2 - \boldsymbol{\mu}_2)}{\nu + d_1}(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}) \\ \nu_{1|2} &= \nu + d_1.\end{aligned} \tag{21}$$

REFERENCES

[1] I. E. Agency, "World Energy Outlook 2019," 2019.

[2] F. Pilo, G. Celli, E. Ghiani, and G. G. Soma, "New electricity distribution network planning approaches for integrating renewable," *Wiley Interdisciplinary Reviews: Energy and Environment*, vol. 2, no. 2, pp. 140–157, 2013.

[3] 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, 2016.

[4] R. A. Verzijlbergh, M. O. Grond, Z. Lukszo, J. G. Slootweg, and M. D. Ilic, "Network impacts and cost savings of controlled EV charging," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1203–1212, 2012.

[5] T. R. Ricciardi, K. Petrou, J. F. Franco, and L. F. Ochoa, "Defining Customer Export Limits in PV-Rich Low Voltage Networks," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 87–97, 2019.

[6] M. Kolenc, I. Papič, and B. Blažič, "Assessment of maximum distributed generation penetration levels in low voltage networks using a probabilistic approach," *Int. J. Electr. Power Energy Syst.*, vol. 64, pp. 505–515, 2015.

[7] A. Angioni, T. Schlösser, F. Ponci, and A. Monti, "Impact of pseudo-measurements from new power profiles on state estimation in low-voltage grids," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 1, pp. 70–77, 2016.

[8] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Security and Privacy*, 2009.

[9] I. Konstantelos, M. Sun, S. H. Tindemans, S. Issad, P. Panciatici, and G. Strbac, "Using Vine Copulas to Generate Representative System States for Machine Learning," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 225–235, 2019.

[10] R. Torquato, Q. Shi, W. Xu, and W. Freitas, "A monte carlo simulation platform for studying low voltage residential networks," *IEEE Transactions on Smart Grid*, vol. 5, no. 6, pp. 2766–2776, 2014.

[11] 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, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148120308569

[12] X. Liu, Y. Yang, R. Li, and P. S. Nielsen, "A stochastic model for residential user activity simulation," *Energies*, vol. 12, no. 17, pp. 1–17, 2019.

[13] M. Nijhuis, M. Gibescu, and J. F. Cobben, "Bottom-up Markov Chain Monte Carlo approach for scenario based residential load modelling with publicly available data," *Energy and Buildings*, vol. 112, pp. 121–129, 2016.

[14] Z. Guo, Z. J. Wang, and A. Kashani, "Home appliance load modeling from aggregated smart meter data," *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 254–262, 2015.

[15] Y. Ji, E. Buechler, and R. Rajagopal, "Data-Driven Load Modeling and Forecasting of Residential Appliances," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2652–2661, 2020.

[16] W. Kong, Z. Y. Dong, D. J. Hill, J. Ma, J. H. Zhao, and F. J. Luo, "A Hierarchical Hidden Markov Model Framework for Home Appliance Modeling," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3079–3090, 2018.

[17] D. Fischer, A. Härtl, and B. Wille-Haussmann, "Model for electric load profiles with high time resolution for German households," *Energy and Buildings*, vol. 92, pp. 170–179, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.enbuild.2015.01.058

[18] F. McLoughlin, A. Duffy, and M. Conlon, "The Generation of Domestic Electricity Load Profiles through Markov Chain Modelling," *3rd Int. Scientific Conference on Energy and Climate Change*, 2010.

[19] C. Wagner, C. Waniek, and U. Hager, "Modeling of household electricity load profiles for distribution grid planning and operation," *2016 IEEE Int. Conference on Power System Technology, POWERCON 2016*, pp. 1–6, 2016.

[20] T. Zufferey, D. Toffanin, D. Toprak, A. Ulbig, and G. Hug, "Generating stochastic residential load profiles from smart meter data for an optimal power matching at an aggregate level," *20th Power Systems Computation Conference, PSCC 2018*, pp. 1–7, 2018.

[21] W. Labeeuw and G. Deconinck, "Residential electrical load model based on mixture model clustering and markov models," *IEEE Trans. Ind. Informat.*, vol. 9, no. 3, pp. 1561–1569, 2013.

[22] D. Gros, P. Wiest, and K. Rudion, "Comparison of stochastic load profile modeling approaches for low voltage residential consumers," in *2017 IEEE Manchester PowerTech, Powertech 2017*, 2017.

[23] M. Uhrig, R. Mueller, and T. Leibfried, "Statistical consumer modelling based on smart meter measurement data," in *2014 International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2014 - Conference Proceedings*, 2014.

[24] W. Labeeuw and G. Deconinck, "Customer sampling in a smart grid pilot," in *IEEE Power and Energy Society General Meeting*, 2012.

[25] R. Singh, B. C. Pal, and R. A. Jabr, "Statistical representation of distribution system loads using Gaussian mixture model," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 29–37, 2010.

[26] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake, "Finite Mixture Models," *Annual Review of Statistics and Its Application*, vol. 6, no. 1, pp. 355–378, 2019.

[27] Q. Gemine, B. Cornélusse, M. Glavic, R. Fonteneau, and D. Ernst, "A Gaussian mixture approach to model stochastic processes in power systems," *19th Power Systems Computation Conference, PSCC 2016*, pp. 1–7, 2016.

[28] 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, 2017.

[29] G. Papaefthymiou and D. Kurowicka, "Using copulas for modeling stochastic dependence in power system uncertainty analysis," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 40–49, 2009.

[30] B. Klöckl and G. Papaefthymiou, "Multivariate time series models for studies on stochastic generators in power systems," *Electric Power Systems Research*, vol. 80, no. 3, pp. 265–276, 2010.

[31] A. Lojowska, D. Kurowicka, G. Papaefthymiou, and L. Van Der Sluis, "Stochastic modeling of power demand due to EVs using copula," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1960–1968, 2012.

[32] M. T. Bina and D. Ahmadi, "Stochastic Modeling for the Next Day Domestic Demand Response Applications," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 2880–2893, 2015.

[33] R. Bernards, J. Morren, and H. Slootweg, "Statistical modelling of load profiles incorporating correlations using copula," in *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe, ISGT-Europe 2017 - Proceedings*, vol. 2018-January. IEEE, 2017, pp. 1–6.

[34] K. Aas, C. Czado, A. Frigessi, and H. Bakken, "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and economics*, vol. 44, no. 2, pp. 182–198, 2009.

[35] O. Morales Napoles, *Counting Vines*. Singapore: World Scientific Publishing, 2011, pp. 189–218.

[36] R. M. Cooke, D. Kurowicka, and K. Wilson, "Sampling, conditionalizing, counting, merging, searching regular vines," *Journal of Multivariate Analysis*, vol. 138, pp. 4–18, 2015.

[37] H. Park and R. Baldick, "Optimal capacity planning of generation system integrating uncertain solar and wind energy with seasonal variability," *Electric Power Systems Research*, vol. 180, p. 106072, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378779619303918

[38] S. Nan, M. Zhou, G. Li, and Y. Xia, "Optimal scheduling approach on smart residential community considering residential load uncertainties," *Journal of Electrical Engineering & Technology*, vol. 14, no. 2, pp. 613–625, Mar 2019. [Online]. Available: https://doi.org/10.1007/s42835-019-00094-0

[39] S. Demarta and A. J. McNeil, "The t copula and related copulas," *International statistical review*, vol. 73, no. 1, pp. 111–129, 2005.

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

[41] N. Gilardi, S. Bengio, and M. Kanevski, "Conditional Gaussian mixture models for environmental risk mapping," in *Neural Networks for Signal Processing - Proceedings of the IEEE Workshop*, vol. 2002-January, 2002, pp. 777–786.

[42] A. Sklar, "Fonctions de reprtition an dimensions et leursmarges," *Publications de l'Institut Statistique de l'Université de Paris*, 1959.

[43] R. B. Nelsen, *An introduction to copulas*. Springer Science & Business Media, 2007.

[44] C. Czado, *Multivariate Distributions and Copulas*. Cham: Springer International Publishing, 2019, p. 31. [Online]. Available: https://doi.org/10.1007/978-3-030-13785-4_1
[45] F. Lindskog, A. Mcneil, and U. Schmock, "Kendall's tau for elliptical distributions," in *Credit Risk*. Springer, 2003, pp. 149–156.
[46] I. Kojadinovic and J. Yan, "Comparison of three semiparametric methods for estimating dependence parameters in copula models," *Insurance: Mathematics and Economics*, vol. 47, no. 1, pp. 52–63, 2010.
[47] Liander, "Open Data." [Online]. Available: https://www.liander.nl/partners/datadiensten/open-data/data
[48] Pecan Street, "Pecan Street Project." [Online]. Available: https://dataport.pecanstreet.org/
[49] UK Power Networks, "Low carbon London project." [Online]. Available: https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households
[50] Australian Government, "Smart Grid Smart City (SGSC) Project." [Online]. Available: https://data.gov.au/dataset/ds-dga-4e21dea3-9b87-4610-94c7-15a8a77907ef/details
[51] B. Efron, "Bootstrap methods: another look at the jackknife," in *Breakthroughs in statistics*. Springer, 1992, pp. 569–593.
[52] R. Mashal and A. Zeevi, "Beyond Correlation: Extreme Co-movements Between Financial Assets," *SSRN Electronic Journal*, 2005.
[53] G. J. Székely and M. L. Rizzo, "The Energy of Data," *Annual Review of Statistics and Its Application*, vol. 4, no. 1, pp. 447–479, 2017.
[54] Z. Guo, Z. J. Wang, and A. Kashani, "Home appliance load modeling from aggregated smart meter data," *IEEE Transactions on power systems*, vol. 30, no. 1, pp. 254–262, 2014.
[55] N. J. Higham, "Computing the nearest correlation matrix—a problem from finance," *IMA journal of Numerical Analysis*, vol. 22, no. 3, pp. 329–343, 2002.

Edgar Mauricio Salazar Duque (M'17) 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 the Kungliga Tekniska Högskolan (KTH), Stockholm, Sweden, and the Technical University of Eindhoven, in the Netherlands in 2018. He is currently working towards a Ph.D. degree in the electrical energy systems group at the Technical University of Eindhoven. His research focuses on data analysis, and applications of machine learning techniques on power distribution grids for planning and operation.

Pedro P. Vergara (M'19) was born in Barranquilla, Colombia in 1990. He received the B.Sc. degree (with honors) 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, UNICAMP, Campinas, Brazil, in 2015. In 2019, he received his Ph.D. degree from the University of Campinas, UNICAMP, Brazil, and the University of Southern Denmark, SDU, Denmark, funded by the Sao Paulo Research Foundation (FAPESP). In 2019, he joined Eindhoven University of Technology, TU/e, in The Netherlands as a Postdoctoral Researcher. In 2020, he was appointed as Assistant Professor at the Intelligent Electrical Power Grids (IEPG) group at Delft University of Technology, also in The Netherlands. His main research interests include the development of methodologies for control, planning, and operation of electrical distribution systems with high penetration of low-carbon energy resources (e.g, electrical vehicles, PV systems, electric heat pumps) using optimization and machine learning approaches. Dr. Vergara has received the Best Presentation Award at the Summer Optimization School in 2018 organized by the Technical University of Denmark (DTU) and the Best Paper Award at the 3rd IEEE International Conference on Smart Energy Systems and Technologies (SEST), in Turkey, in 2020.

Phuong H. Nguyen (M'06) received the Ph.D. degree from the Eindhoven University of Technology (TU/e), the Netherlands in 2010. During his one-year sabbatical leave in 2019, he took up a group leader position of the Sustainable Energy Systems (SES) group of the Luxembourg Institute of Science and Technology (LIST). Since January 2020, he has been back to TU/e as an associate professor in the Electrical Energy System (EES) 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 of interests includes data analytics with deep learning, real-time system awareness using (IoT) data integrity, as well as predictive and corrective grid control functions.

Anne van der Molen (M'00) received his M.S in Electrical Engineering from Twente University (1997). He is presently working for Dutch Distribution System Operator Stedin where he is engaged with smart grids strategy- and technology planning. His areas of interest include system/market operations, operational technology and flexibility. Next to that, Mr. van der Molen is part-time research associate at Eindhoven University of Technology in the area of intelligent energy systems. He is also 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 member of the technology committee of the association of European Distribution System Operators (E.DSO).

J.G. (Han) Slootweg (M'00; SM'19) received the M.Sc. degree in electrical power engineering in 1998 (cum laude) and the Ph.D. degree in 2003, both from Delft University of Technology, Delft, The Netherlands. He also received the M.Sc. degree in business administration. He is currently Director of the Asset Management Department of Enexis Netbeheer B.V., 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. Han also holds a professorship in Smart Grids at the Electrical Energy Systems group at the Eindhoven University of Technology. He has (co-)authored more than 200 papers, covering a broad range of various aspects of electrical power systems.