POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

A Data-Driven Framework for Probabilistic Optimal Power Flow in Electrical Grids with Renewable Sources

Tesi di Laurea Magistrale in
Mathematical Engineering - Ingegneria matematica

Author: **Marco Vivian**

Student ID: 103494
Advisor: Prof. Giambattista Gruosso
Academic Year: 2024-25

Abstract

In recent times, the rapid growth in usage of photovoltaic (PV) energy and electric vehicles (EV) has been heavily reshaping the operational functioning of electrical distribution networks by increasing the overall uncertainty level. This work provides an analysis of a network characterized by the integration of distributed energy resources (DERs) with the goal of understanding the roles of these units across different load scenarios, assuming they are allowed to adapt their operation to the network state as permitted by recent technical regulations that support or mandate controllability of small-scale generators (such as inverter-based DERs). The methodology adopted for the analysis relies on the Probabilistic Optimal Power Flow (POPF) integrated within the Monte Carlo simulations framework, where probability density distributions of residential loads in the network are estimated under different conditions and then used as inputs of the POPF to determine the optimal response of DERs in time and load-varying scenarios. These results will serve as a reference for the next part of this work, where the focus shifts to estimating load distributions using Kernel Density Estimation (KDE) and copula-based modeling. Synthetic load profiles will be generated with those approaches, making sure they represent real data well enough, and then new DER distributions will be computed as outputs of the same POPF methodology applied previously. The goal is to obtain results which are comparable to those based on real data, thereby enabling future analyses under a wide range of network scenarios while still preserving a high degree of realism.
In conclusion, this work provides a robust foundation for uncertainty-aware grids with DER integration analysis and hints at future development of synthetic and realistic consumption scenarios with the flexibility to incorporate an even wider range of load types.

Keywords: Probabilistic Optimal Power Flow, Monte Carlo simulation, Copula, Distributed Energy Resources, loads estimation

Abstract in lingua italiana

Negli ultimi tempi, la rapida diffusione dell'energia fotovoltaica (PV) e dei veicoli elettrici (EV) ha profondamente trasformato il funzionamento operativo delle reti di distribuzione elettrica, aumentandone il livello di incertezza. Questa tesi propone l'analisi di una rete caratterizzata dall'integrazione di risorse energetiche distribuite (DER), con l'obiettivo di comprendere il ruolo di queste unità in diversi scenari di carico, assumendo che esse possano adattare il proprio funzionamento allo stato della rete, come consentito dalle recenti normative che supportano o impongono la controllabilità dei generatori di piccola scala. La metodologia adottata in questa tesi si basa sull'utilizzo del Probabilistic Optimal Power Flow (POPF), integrato in un framework di simulazione Monte Carlo. In questo contesto, le densità di probabilità dei carichi residenziali presenti nella rete vengono prima stimate e poi impiegate come input del POPF, al fine di determinare in uscita la risposta ottimale delle DER in diversi scenari che differiscono per variabilità temporale e di carico. Tali risultati saranno poi utilizzati come riferimento per la parte successiva di questa tesi, in cui l'attenzione si sposterà sulle distribuzioni dei carichi, le quali verranno stimate tramite tecniche di Kernel Density Estimation (KDE) e modellazione basata su copule. Profili di carico sintetici verranno generati a partire dalle distribuzioni stimate e nuovi output delle DER saranno calcolati utilizzando lo stesso processo di POPF applicato in precedenza. L'obiettivo è quello di ottenere risultati comparabili a quelli ottenuti attraverso dati reali, aprendo così possibilità di analisi future in un'ampia gamma di scenari di rete differenti, pur mantenendo un elevato grado di realismo.

Parole chiave: Probabilistic Optimal Power Flow, Monte Carlo, copula, risorse energetiche distribuite, stima dei carichi

Contents

Introduction

Traditionally, load flow analysis in power systems has been conducted following deterministic approaches, which just rely on worst-case scenarios, typically considering the simultaneous occurrence of maximum load and minimum generation, or vice versa. Recently, Distributed Energy Resources (DERs) such as photovoltaic panels, wind turbines, and battery storage systems have experienced a significant increase in popularity, mainly due to decreasing installation costs and the growing demand for cleaner and more sustainable energy solutions. At the same time, the increasing trend of electricity as one of the main sources for energy generation, alongside more accessible ways to retrieve high-resolution consumption data, has made it possible to model household demand with way better granularity. On the other side, this level of detail also introduces high variability and uncertainty into load forecasting. This poses significant challenges to the planning and control of power systems, considering also that, unlike traditional generators, DERs are often located closer to the final consumers, and they are characterized by an uncertain behavior due to their dependence on external conditions like climate change and user habits. As a result, power system operators and researchers have been increasingly interested in methods that can incorporate uncertainty directly into the decision-making process. It's the case, for example, of [1]: a paper in which the focus is to analyze the steady-state operating conditions of an electrical distribution system with energy generated by wind and photovoltaic sources. Another step in that direction was also made by [2], which extended the analysis to the possibility of having electric vehicles (EV) connected to the network, a new load source that made its appearance only in recent years. This work, instead, takes a quite different approach: rather than controlling the uncertainty of DER energy generation, the elements of the network that are going to be modeled will be the loads. Modeling load uncertainty using real household consumption data ensures that sampled load scenarios reflect plausible and data-driven operating conditions. Moreover, generating synthetic load profiles via copula-based sampling aims to preserve both the temporal patterns and spatial dependencies among different types of consumers. On the contrary, DER generators, which are traditionally considered non-dispatchable, will be supposed to be controllable. This modeling choice allows us to assess the network's optimal response to load variability and to identify the functional composition of the system, namely, which generators are actively involved in reaching the equilibrium. This assumption aligns with the future-oriented vision of smart grids, where DERs are increasingly equipped with advanced control capabilities, in accordance with scientific research like [3] and some recent norms aimed at increasing stability in electrical networks. Under these settings, multiple iterations of an Optimal Power Flow (OPF) are going to be performed through a Monte Carlo (MC) algorithm. The goal is not just to solve for one optimal configuration but to characterize the distribution of optimal setpoints for each controllable generator across thousands of realistic system states. This allows the identification of recurrent patterns in generator usage, such as preferred dispatch regions or saturation effects.

Ultimately, the project aims to help build control strategies and planning policies by revealing how controllable DERs would organize themselves in an optimal, data-driven scenario under uncertainty, highlighting their role in supporting voltage regulation and system flexibility.

This study is organized as follows:

- **Chapter 1:** provides some background elements about Optimal Power Flow and about the Monte Carlo approach to the case study.
- **Chapter 2:** presents the network used for reference and on which all the simulations will be run.
- **Chapter 3:** introduces the dataset of choice, which will be used for estimating load profiles, and discusses all the modeling decisions taken to better split data into different consumption groups and scenarios.
- **Chapter 4:** deals with the actual Optimal Power Flow implementation and every other aspect related to it, as well as the sampling strategy adopted.
- **Chapter 5:** includes information about how the synthetic data of loads are generated, sampled, and used for the Monte Carlo simulation. It also presents the results of the synthetic OPF, comparing the output with what was obtained in Chapter 4 and assessing both positive and negative implications of the results.
- **Chapter 6:** is where, finally, conclusions are drawn and further ideas on how this research could be extended, or analysis improved, are exposed.

1 | Background Framework

This chapter describes the theory behind Optimal Power Flow by first presenting the general problem, describing all its details, and providing its formulations. Secondly, it illustrates how the shift from a deterministic to a stochastic setting is embedded by presenting the state of the art in Probabilistic Load Flow, and lastly, it provides a list of techniques commonly used to solve the numerical problem with a focus on the one used in this project.

1.1. Optimal Power Flow

Introduced in 1962 by Carpentier, **Optimal Power Flow (OPF)** is one of the most researched and fundamental sets of constrained and nonlinear optimization problems in the field of engineering of power systems. Its purpose is to determine the optimal operating conditions of an electrical power grid subject to a set of physical, operational, and economical constraints. As of today, the OPF problem incorporates as its defining hallmark the presence in the equality constraints of the AC (Alternating Current) power flow equations, which are derived from Kirchhoff's laws and model the steady-state behavior of voltages, currents, and power injections at each bus in the network. In the following section, the key points of an OPF formulation are addressed.

1.1.1. Control Variates

Control variates represent a set of quantities that the system can deliberately adjust to move its status to an optimal operating point. These variables are not determined by physical laws alone (like the power flow equations) but rather represent degrees of freedom available for optimization within prescribed limits. They usually are:

Active power output of generators (P^{gen})
Amount of power generated by dispatchable units. It is adjusted to meet demand and minimize generation cost.

Reactive power output of generators (Q^{gen})
Amount of reactive power that a generator injects or absorbs to support voltage regulation and maintain reactive power balance across the network.

Voltage magnitudes at generation buses (V^{gen})
Electric potential levels (measured in volts or per-unit) at the buses where generators are connected.

Slack bus power injection
The slack bus is used to compensate for the mismatch between the total generation and total load plus losses in the system. Its power injection is hence not fixed in advance, but it is determined as output of the OPF solution.

Other control variates like shunt capacitors, phase shifter angles, and tap positions of transformers can be considered. However, the control variables strictly needed for this project are just the ones listed before.

1.1.2. State Variables

State variables are the internal variables that alone are able to describe the steady-state condition of the network.

Voltage magnitudes and angles at all buses
In an electric power system, each bus (node) has a complex voltage, usually written as: $V = |V|e^{j\theta}$ where $|V|$ = voltage magnitude and θ = voltage angle.

Once one knows all voltage magnitudes and angles, he can compute power flows on lines, injections at buses, losses, and constraint violations.

1.1.3. Constraints

Constraints are rules or limits that the solution of an optimization problem must satisfy. They define what is feasible or acceptable in the system.
More specifically, in an OPF problem, constraints make sure that the solution respects:

- Physical laws, like Kirchhoff's laws,
- Operational limits, like generator capacities or voltage bounds,
- Safety margins, like maximum current in lines.

There are two types of constraints:

Equality Constraints:
Active and reactive power balance: These equations ensure that the net real and reactive power injected at each bus matches the real power flow determined by the voltage conditions and network parameters.

$$\begin{aligned} P_i^{\text{gen}} - P_i^{\text{load}} &= P_i^{\text{injected}}, \quad \forall i \in \{1, \dots, N\}, \\ Q_i^{\text{gen}} - Q_i^{\text{load}} &= Q_i^{\text{injected}}, \quad \forall i \in \{1, \dots, N\}. \end{aligned}$$

Which can be more specifically written as:

$$\begin{aligned} P_i^{\text{gen}} - P_i^{\text{load}} &= V_i \sum_{j=1}^{N} V_j \left(G_{ij} \cos(\theta_i - \theta_j) + B_{ij} \sin(\theta_i - \theta_j) \right), \quad \forall i \in \{1, \dots, N\}, \\ Q_i^{\text{gen}} - Q_i^{\text{load}} &= V_i \sum_{j=1}^{N} V_j \left(G_{ij} \sin(\theta_i - \theta_j) - B_{ij} \cos(\theta_i - \theta_j) \right), \quad \forall i \in \{1, \dots, N\}. \end{aligned}$$

where:

- N is the total number of buses.
- V_i, V_j are the voltage magnitudes at buses i and j.
- θ_i, θ_j are the voltage angles at buses i and j.
- G_{ij} is the conductance between buses i and j.
- B_{ij} is the susceptance between buses i and j.
- $P_i^{\text{gen}}, Q_i^{\text{gen}}$ are the generated active and reactive powers at bus i.
- $P_i^{\text{load}}, Q_i^{\text{load}}$ are the active and reactive loads at bus i.

Inequality Constraints:
Inequality constraints limit the operating region by specifying maximum and minimum bounds for various system variables. They are crucial for avoiding line overloads, imposing generator limits, and obtaining acceptable voltage ranges in the solution.

The active and reactive power generation of generators, loads, DC lines, and static generators can be defined as a flexibility for the OPF.

$$
\begin{aligned}
&P_{min,i} \leq P_g \leq P_{max,g},\ g \ \epsilon \ gen \\
&Q_{min,g} \leq Q_g \leq Q_{max,g},\ g \ \epsilon \ gen \\
&P_{min,sg} \leq P_{sg} \leq P_{max,sg},\ sg \ \epsilon \ sgen \\
&Q_{min,sg} \leq Q_{sg} \leq Q_{max,sg},\ sg \ \epsilon \ sgen \\
&P_{max,g},\ g \ \epsilon \ dcline \\
&Q_{min,g} \leq Q_g \leq Q_{max,g},\ g \ \epsilon \ dcline \\
&P_{min,eg} \leq P_{eg} \leq P_{max,eg},\ eg \ \epsilon \ ext\_grid \\
&Q_{min,eg} \leq Q_{eg} \leq Q_{max,eg},\ eg \ \epsilon \ ext\_grid \\
&P_{min,ld} \leq P_{ld} \leq P_{max,ld},\ ld \ \epsilon \ load \\
&Q_{min,ld} \leq Q_{ld} \leq Q_{max,ld},\ ld \ \epsilon \ load \\
&P_{min,st} \leq P_{st} \leq P_{max,st},\ st \ \epsilon \ storage \\
&Q_{min,st} \leq Q_{st} \leq Q_{max,st},\ st \ \epsilon \ storage
\end{aligned}
$$

The network constraints related to bus voltages and branch flows:

$$
\begin{aligned}
&V_{min,j} \leq V_{g,i} \leq V_{min,i},\ j \ \epsilon \ bus \\
&L_k \leq L_{max,k},\ k \ \epsilon \ trafo \\
&L_l \leq L_{max,l},\ l \ \epsilon \ line \\
&L_l \leq L_{max,l},\ l \ \epsilon \ trafo_{3w}
\end{aligned}
$$

Note: These constraints are set according to [4] to preserve coherence in the implementation.

1.1.4. Objective Function

Generally speaking, the objective function is the key element in any optimization problem. It is a mathematical representation of the operational objective to be optimized (often minimized), and, in the context of OPF, it can represent different quantities based on the case study: one can focus on the minimization of active power losses, emissions, or voltage deviation, or, as in this case, one can also be interested in minimizing the economic cost associated with dispatching generation resources.

It is usually represented as $f(x, u)$ indicating that it is a function of state variables x and control variables u of the system. However, in literature, many different types of objective functions have been used; for example, in [5], a multi-objective function approach is used, trying to optimize both the total fuel cost of generation units and the total emission

released from each generation unit. In [6], the minimization of the power transmission loss function is set as the goal to achieve.
However, in conventional OPF formulations, and hence in this work too, the objective function $f(x, u)$ is usually defined as the total generation cost:

$$f(x, u) = \sum_{i \in \mathcal{G}} C_i(P_{G_i})$$

where:

- $\mathcal{G}$ is the set of generators,
- P_{G_i} is the active power output of generator i,
- $C_i(P_{G_i})$ is the cost function associated with generator i.

As regards the function C_i, it can also be modeled in different ways: a first option is to consider it to behave like a **piecewise linear cost function**, that is:

$$C_i(P_{G_i}) = \begin{cases} m_{i1} P_{G_i} + b_{i1}, & \text{if } P_{G_i} \in [P_i^{\min}, P_{i1}] \\ m_{i2} P_{G_i} + b_{i2}, & \text{if } P_{G_i} \in (P_{i1}, P_{i2}] \\ \vdots & \vdots \\ m_{in} P_{G_i} + b_{in}, & \text{if } P_{G_i} \in (P_{i(n-1)}, P_i^{\max}] \end{cases}$$

This expresses each segment as a linear function with explicit slope and intercept.
Equivalently, with a form that is often preferred in numerical implementations, especially when the breakpoints are known and precomputed:

$$C_i(p) = C_\alpha + (p - p_\alpha) \frac{C_{\alpha+1} - C_\alpha}{p_{\alpha+1} - p_\alpha}, \quad \text{for } p \in (p_\alpha, p_{\alpha+1}), \quad \alpha = 0, ..., n-1$$

$$(p_\alpha, C_\alpha) = \begin{cases} (p_0, C_0), & p_0 < p < p_1 \\ \vdots \\ (p_{n-1}, C_{n-1}), & p_{n-1} < p < p_n \end{cases}$$

Note: Piecewise linear interpolation used for generator cost modeling. Adapted from [4]

It is mainly used when generator bidding curves or tariffs are represented by discrete segments; it is common in electricity markets.

A second common option for modeling the function C_i, and the one adopted in this work, is to represent it as a **convex quadratic function**

$$C_i(P_{G_i}) = a_i + b_i P_{G_i} + c_i P_{G_i}^2 \tag{1.1}$$

with coefficients $a_i, b_i, c_i \in \mathbb{R}$, where $c_i \geq 0$ to ensure convexity. This form captures generator fuel costs and efficiency characteristics. Since for this work the choice of a polynomial quadratic function was made as the function to be minimized, the modeling of the cost function translates into the choice of the parameters $a_i, b_i, c_i \in \mathbb{R}$ of (1.1) or of the overall shape of the objective function itself in order to facilitate the convergence.

1.1.5. Problem Formulation

The Optimal Power Flow problem is typically formulated as follows:

$$\begin{aligned} \min_{x,u} \quad & f(x,u) \\ \text{subject to} \quad & g(x,u) = 0 \quad \text{(Power flow equations)} \\ & h(x,u) \leq 0 \quad \text{(Operational constraints)} \end{aligned}$$

where:

- x: State variables of the system (as in 1.1.2)
- u: Control variables (or control variates, as in 1.1.1)
- $f(x,u)$: Objective function (as in 1.1.4)
- $g(x,u) = 0$: Equality constraints representing the AC power flow equations (as explained in detail in 1.1.3)
- $h(x,u) \leq 0$: Inequality constraints (as in 1.1.3)

1.2. Probabilistic Load Flow and State of the Art

The goal of the Probabilistic Load Flow (PLF) analysis is to examine the uncertainties of the inputs in power flow calculations, that is, to capture the variability of all loads and generators in the network by representing them through appropriate probability density functions (PDFs). These PDFs are then used as inputs to the PLF computations that can be addressed numerically by repeatedly sampling from the input distributions and performing a deterministic power flow calculation for each realization, as is done in Monte

Carlo methods. The outputs of the PLF are PDFs too to represent uncertainty.
A lot of studies have been concentrating on load modeling since they constitute one of the main sources of uncertainty. One of the principal choices for aggregated loads PDF is to use a normal distribution; however, many differences arise between different methods: [7] used real-world data for computing the mean value and a fixed 7% of that value for the standard deviation. Others, like [8], used historical data with the role of empirical PDFs. Regarding residential loads, instead, the most common PDFs utilized are the Beta, Gamma or Exponential.
Another option to include uncertainty in the model is through the direct modeling of PV sources: their power variability can also be represented by the clear-sky index, which is defined as the ratio between the actual global horizontal irradiance measured at a given time and location and the theoretical clear-sky irradiance that would be expected under cloudless conditions at the same time and location. In order to model this, typically mixtures of distributions such as, for example, normal, lognormal, or polynomial are implemented, as in [9]. For EV modeling, instead, normal, Poisson distributions or empirical ones are usually implemented.
The last component in power systems analysis which contributes to the overall uncertainty is the correlation between inputs (PV, EV, and loads in each combination of two of them). Generally speaking, the joint variability of two random variables can be measured through the covariance, but even though these methods are widely used for their simplicity, they often fail to accurately capture the dependency structure in more complex systems. For this reason, copulas have been used for magnitude correlations between inputs, like again in [8]. Anyway, a key insight from the literature is that correlation techniques are typically selected independently of the input variable types, highlighting their versatility in modeling dependencies across diverse input configurations.
As anticipated in the introduction, this work takes quite a different approach: following the insight of [10], which in its conclusions hints toward the implementation of new models for PV, EV, or load representation, residential loads will be modeled making use of real-world data to build a synthetic dataset through Copula methods. By modeling correlation, the inputs become way more representative of true data behavior. This procedure, with further developments, could lead to a more reliable and data-driven framework for probabilistic power system analysis. Moreover, while the approach relies on real-world measurements, it does not stay empirical; instead, it builds a probabilistic model capable of simulating diverse, but still realistic, load scenarios.
To summarize, in order to be able to address these particular situations and to avoid treating them with worst-case assumptions, the several approaches used in literature are exposed in Figure 1.1

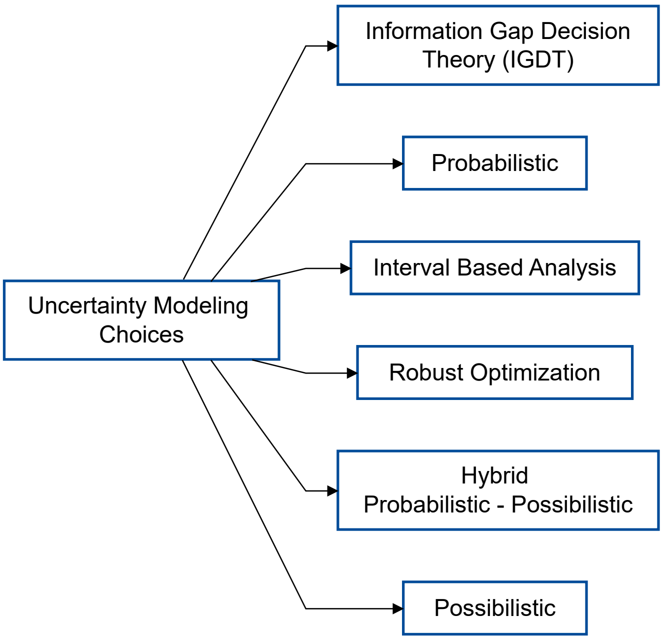


Figure 1.1: Summary of main uncertainty modeling choices, adapted from [11]

1.3. Probabilistic-Optimal Power Flow

In this work, the probabilistic approach is embraced with the introduction of the **Probabilistic Optimal Power Flow (POPF)**. As anticipated in the previous section, it is a further extension of the classic OPF, which accounts for the uncertainties in power systems. Instead of considering deterministic values for loads, with POPF one can model these quantities as random variables (or stochastic processes), each with a probability distribution which is supposed to be known from an algorithmic point of view. In this context, these probability distributions are estimated starting from real data with the goal of accurately capturing the statistical behavior of load variations, thereby enabling the generation of realistic scenarios for the probabilistic OPF analysis.

1.4. Solving Techniques

The main goal of the OPF is to determine the output probability density functions (PDFs) of the variables of interest. Aiming at this, several methods have already been tested, and Figure 1.2, provides a summary of them. In particular, the method that is going to be embraced with this work comes from the numerical family: it is based on iterative and deterministic power flow calculations for each set of new samples.

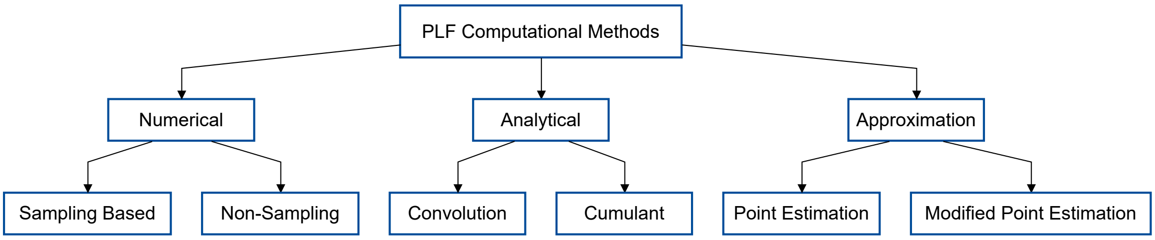


Figure 1.2: Summary of OPF solving techniques

For this purpose, the numerical method called **Monte Carlo Simulation (MCS)** is introduced to repeatedly draw samples: the output PDFs in MCS are obtained by generating multiple samples from the input distributions and solving the power flow problem for each realization. In order to perform this procedure, it is fundamental to introduce a stopping criterion, which is designed to terminate the procedure when a sufficient number of samples have been evaluated to ensure the reliability and statistical accuracy of the estimated output distribution. In this work, it is set to be a threshold on the number of simulations (and hence PF computations) performed.
MCS methods can be further classified based on their sampling method, and, as anticipated before, this work relies on a stratified and dependency-preserving sampling approach based on the classification of data into consumption groups in order to maintain both their empirical proportions and temporal correlations. More specifically, the proposed method can be seen as a temporally extended variant of standard MCS. Rather than performing a single-time snapshot analysis, the OPF is solved at multiple discrete time instants, each of them representative of the behavior of the system during the span of time from one time step to the next one. For each time instant, load PDFs are estimated and samples are drawn. Although temporal evolution is not modeled explicitly, this approach allows to have a wider understanding of the system behavior across short-term periods of time.

2 | Test Network Description

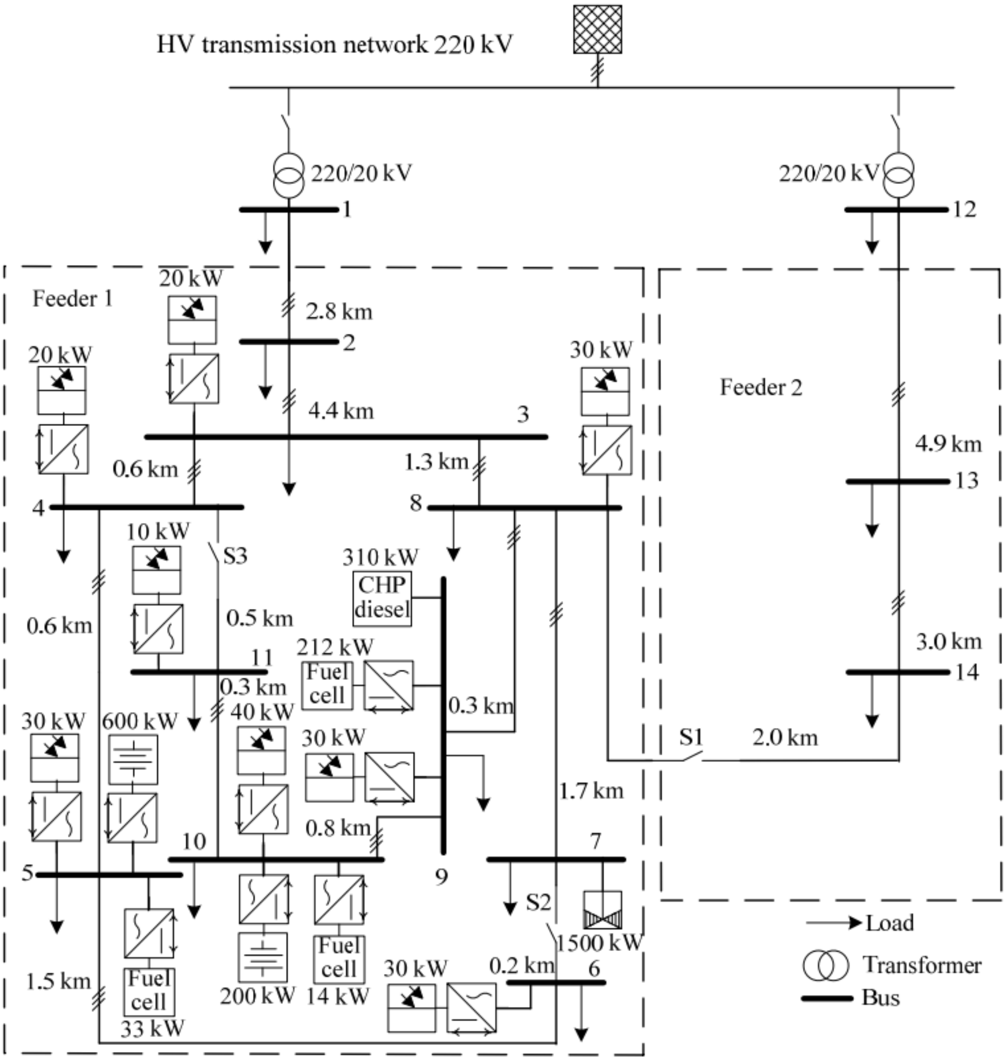


Figure 2.1: Case of study network from [4]

For this study, the network represented in Figure 2.1 was selected from the `pandapower` documentation [4]. Pandapower is an open-source Python-based tool for automated analysis and optimization of power systems. It provides a convenient interface for creating, simulating, and evaluating electrical networks, supporting functions such as power flow, optimal power flow, short-circuit analysis, and state estimation. Among the built-in networks often used as benchmarks, the chosen power grid is a medium-voltage (MV) distribution network with all Distributed Energy Resources (DER). This network is a comprehensive benchmark model developed by the CIGRE Task Force C6.04.02. It is designed to facilitate the analysis and validation of methods for the integration of DER into MV distribution systems, as deeply explained in [12, 13]. The CIGRE medium-voltage (MV) distribution network operates primarily at 20 kV. In the original specification provided, the network is connected to the high-voltage (HV) transmission system through a 220/20 kV transformer. However, in the implementation available within the `pandapower` library, the high-voltage side is modeled as 110 kV instead of 220 kV due to the absence of predefined 220 kV transformer types in the standard component library. This adaptation does not affect the MV topology or the internal structure of the network and is commonly accepted for simulation purposes where detailed HV behavior is not the primary focus. This network is implemented in the Pandapower Python library and can be accessed using the function `create_cigre_network_mv(with_der="all")`. In the context of this work, the network serves as the base model for conducting Monte Carlo-based Probabilistic Optimal Power Flow (POPF) simulations. Synthetic and real load data are applied to its nodes to assess the impact of uncertainty on energy dispatch.

The key structural elements and distributed resources present in this benchmark network are discussed in the following section.

2.1. Network Composition

The components of the case study network can be listed as follows:

- **Buses:** The network consists of 15 buses. Bus 0 is connected to the high-voltage (HV) external grid at 110 kV, serving as the slack bus, establishing the voltage and angle reference. The other buses operate at 20 kV, forming the MV distribution framework.
- **Lines:** The network features 15 transmission lines with various lengths ranging from 0.24 km to 4.89 km, possessing standard electrical parameters defined by resistance (R) and reactance (X). Two different types of lines ("cs" and "ol") are included, each with specified thermal current limits.
- **Transformers:** Two identical transformers ("Trafo 0-1" and "Trafo 0-12") connect the HV and MV systems. Each transformer has a rated power of 25 MVA, with primary and secondary voltages of 110 kV and 20 kV, respectively, a short-circuit impedance of approximately 12%, and no-load losses of 0.16 kW. The transformers

do not have tap-changing capabilities and exhibit a phase shift of 30 degrees.

- **Loads:** The network hosts 18 distinct load points, categorized into residential (Load R) and commercial (Load C). Residential loads vary significantly in magnitude, with the largest (Load R1 and Load R12) each having approximately 15 MW and 3 MVar. Commercial loads are smaller in magnitude and are distributed across various nodes.
- **Switches:** A total of 8 switches are present, consisting of load-break switches (LBS) and circuit breakers (CB). These switches facilitate operational flexibility by enabling network reconfiguration. Three load-break switches (S1, S2, and S3) are notably kept open under normal conditions, ensuring network radiality and providing operational options during faults or maintenance activities.
- **External grid connection:** The network's external grid connection at Bus 0 is characterized by a reference voltage magnitude of 1.03 per unit (pu), a maximum short-circuit power capability of 5000 MVA, and an associated R/X ratio ranging between 0.1 and 0.1. Although initialized with these values, the external grid is dynamically adjusted during the Optimal Power Flow (OPF) simulations conducted in this study.
- **DERs:** The network integrates 15 static generators (sgens) representing DERs, including photovoltaic (PV) panels, wind turbines, battery storage systems, residential fuel cells, a combined heat and power (CHP) diesel unit, and a CHP fuel cell unit. These will be the object of study.
- **Bus geodata:** Each bus is associated with geographical coordinates, facilitating visualization and spatial analysis.

In particular, each component and its initialization values can be obtained by running `create_cigre_network_mv(with_der="all")` in the Python environment, and they can be consulted in Appendix A.

3 | Dataset Preparation

In this chapter, several key aspects of the dataset are addressed to prepare and better understand the input used in later phases of the analysis. First, a general overview of the dataset is provided, including its structure, temporal resolution, and the type of information available for each household. Then, a data cleaning process is described, where non-residential profiles and potentially corrupted or anomalous entries are identified and filtered out to ensure the reliability of subsequent analysis. Next, the dataset is divided into groups through clustering techniques applied to representative consumption features to identify distinct usage profiles among the residential consumers. The resulting clusters are then examined in detail, especially with visual representation highlighting differences in magnitude and variability of energy demand. Finally, the chapter shifts its focus onto the temporal dimension of the data, investigating consumption behavior across seasons and different days of the week.
This analysis provides insight into recurring patterns and lays the foundation for modeling the temporal variability in the subsequent synthetic profile generation chapter.

3.1. Dataset Description

This work employs the dataset introduced by Bendík (2023) [14], which contains high-resolution power consumption profiles for 1,000 households in the Slovak Republic over a full calendar year, including both rural and urban environments. Each household, along with other specifics covered in more depth in the following section, is represented by a time series of active power consumption values, sampled every 15 minutes. This makes it an ideal resource for studies focused on understanding diverse consumption behaviors and for simulating realistic load scenarios in power systems analysis. Each household's energy consumption was measured and recorded at 15-minute intervals, resulting in 96 data points per day and approximately 35,040 data points per year per household. The dataset includes both active and reactive power measurements, which allows for in-depth analyses not only of energy usage but also of power quality and the dynamics of reactive power flow in low-voltage networks. The structure of the dataset is organized into JSON files, one per household, where each file contains daily entries. Additionally, the dataset includes a CSV metadata file that describes the static attributes of each meter, such as its identifier (`meterID`) and the ZIP code of its location (ensuring a level of geographic traceability). From a technical standpoint, the measurements were collected using intelligent metering systems compliant with European accuracy standards (IEC Class 1) and the Measuring Instruments Directive. The meters were installed at the point of common coupling (PCC) for each household, and the data was transmitted and archived remotely on a regular basis.

The dataset is publicly available via Mendeley Data and is released under a permissive Creative Commons Attribution 4.0 International (CC BY 4.0) license, allowing for unrestricted reuse with proper attribution. This dataset is particularly well-suited for probabilistic OPF modeling due to several key characteristics. First, the high temporal resolution (96 data points per day) enables the possibility to model load variations inside days at specific time instances, which is essential for capturing realistic operational conditions of networks. Second, the large number of distinct household profiles provides a statistically diverse sample of consumption behaviors, allowing for robust statistical analysis. More specifically, this diversity supports the development of synthetic data generation strategies, which would be one of the main focuses later on, while preserving realistic spatial correlation across nodes.

The key features present in the dataset are listed below:

- **MeterID:** Unique identifier for each household or metering point.
- **Date:** day of the year.
- **Season:** the corresponding season.
- **Weekday:** the corresponding day of the week.
- **Consumption:** timeseries of 96 values, each of them representing the average active power consumption in kW over 15 minutes on a particular day.
- **ZIP:** postal code.
- **Zone:** the first two numbers of the ZIP code, each of them representing a different region.

3.2. Data Cleaning

In the initial stage of data preparation, following a first quick visual analysis of the dataset, many of them are detected not to represent consumption of households or buildings. These entries show instead distinct characteristics such as highly regular consumption patterns (often limited to nighttime hours), flat load profiles showing minimal variation throughout the day, and anomalously low values in total daily energy consumption. This is the typical behavior of consumes which represent, for instance, public light installation, and since the goal of the project is to model residential demand realistically, a rule-based filtering strategy was applied: a few descriptive statistics were considered, such as the ratio between night and day consumption, total energy, or number of active hours. By applying these criteria, public lighting profiles were identified, ensuring that the remaining dataset was composed exclusively of relevant residential profiles.

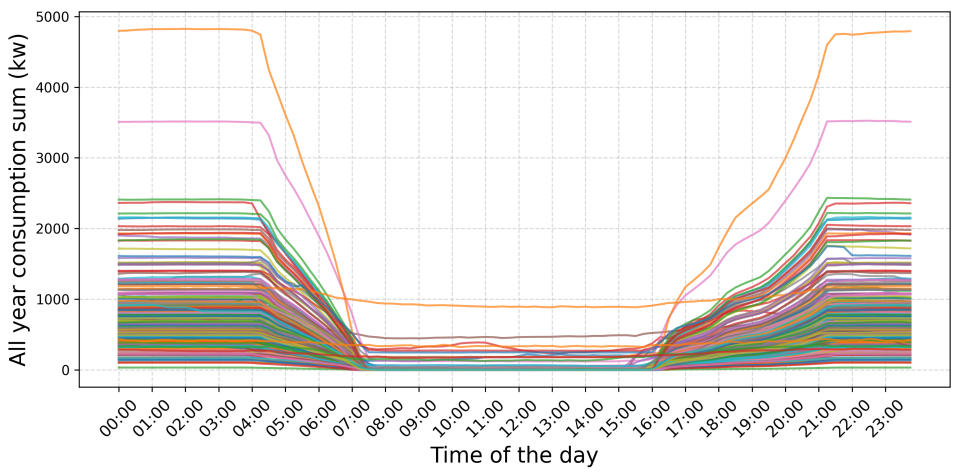
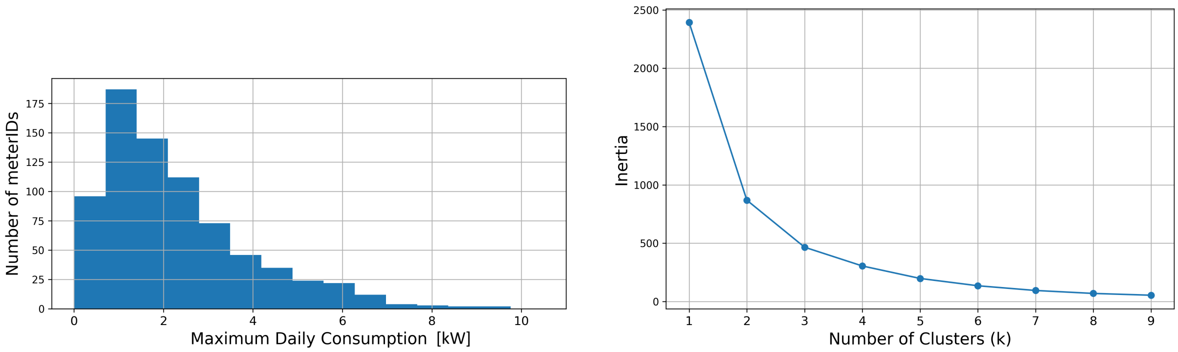


Figure 3.1: All identified time series of public lighting

In Figure 3.1 are shown all the time timeseries which have been identified to belong to the public lightning group. It is evident that none of these profiles correspond to typical household consumption. As a result, for the purpose of this work, all of them have been filtered out of the initial dataset, leaving it with 766 different profiles.

3.3. Data Clustering

To account for the differences in energy usage across different types of households, a clustering approach based on overall consumption levels was applied. This step's aim is to find groups with similar load magnitudes to allow the construction of more representative scenarios and to better capture variability and reduce the risk of aggregating fundamentally different load behaviors. To do so, for each unique `meterID`, the average of the daily peaks in active power consumption over the entire year was computed. The distribution of these values is shown in Figure 3.2a.



(a) Histogram of mean peak values

(b) Inertia vs. number of clusters

Figure 3.2: Exploratory analysis of average daily peak consumption and optimal cluster selection.

The resulting set of average peak consumption values was then clustered using the k-means algorithm. As shown by Figure 3.2b, several numbers of clusters have been tested, and the choice of $k = 3$ as the number of clusters seems the most suitable based on the elbow method. Consequently, labels have been assigned to each unique meterID accordingly.

3.4. Clusters Examination

After choosing an appropriate number of clusters, it's also important to check if the actual groups are meaningful, that is, if the elements within each cluster exhibit different behaviors in terms of electrical consumption magnitude. Before doing that, the total number of unique households cataloged in each group, that is the cluster sizes, are compared in Table 3.1.

Cluster	Number of Profiles
0	437
1	253
2	75

Table 3.1: Number of profiles in each consumption cluster.

In terms of cluster sizes, it's fair to assume that the majority of the residential points of consumption in a country are represented by average households and small residential buildings, while the number of high-consumption households, such as those with intensive appliance usage, is relatively low. These patterns become more evident by looking at

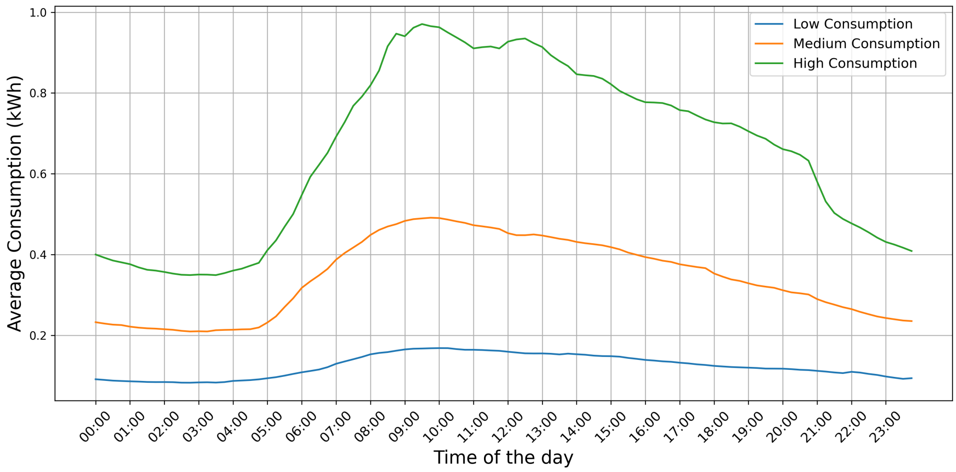


Figure 3.3: Comparison between time series of different consumption groups

Figure 3.3. The average daily consumption patterns differ significantly across the three clusters. Households in the high-consumption group display a sharp rise in consumption in the early hours of the morning followed by a peak between 9:00 a.m. and 10:00 a.m., likely due to increased appliance usage during active hours (e.g., cooking, heating, water usage). This group maintains relatively high consumption throughout the day, gradually decreasing toward late evening. The medium-consumption group follows a similar but more subdued pattern, with a moderate peak around the same morning hours, followed by a gradual decline. In contrast, the low-consumption group maintains a relatively flat profile, with only a slight increase in the early morning and a gentle decline over the afternoon. This behavior may reflect smaller households or minimal daytime appliance use, and it highlights their lower sensitivity to time-of-day effects. These differences confirm that the clustering successfully captures distinct consumer behaviors not only in terms of total energy consumed but also in their temporal dynamics, which is crucial for accurately modeling load uncertainty and dependency across time.

3.5. Time-Based Consumption Patterns

Exploiting the temporal variability of electricity consumption is a key and an essential step for load modeling. In this section, the dataset is analyzed with respect to seasonal effects and weekly usage patterns.

3.5.1. Seasonal Variations

Seasonal trends are certainly one of the main motifs of variations in energy demand. In particular, higher consumption during colder months due to heating is typical of residential households, while increased usage in summer is more commonly observed in medium-consumption buildings, likely due to cooling systems and commercial activity being active during the whole day. All these patterns are clearly presented in Figure 3.4 and can be found also in all the other consumption classes.

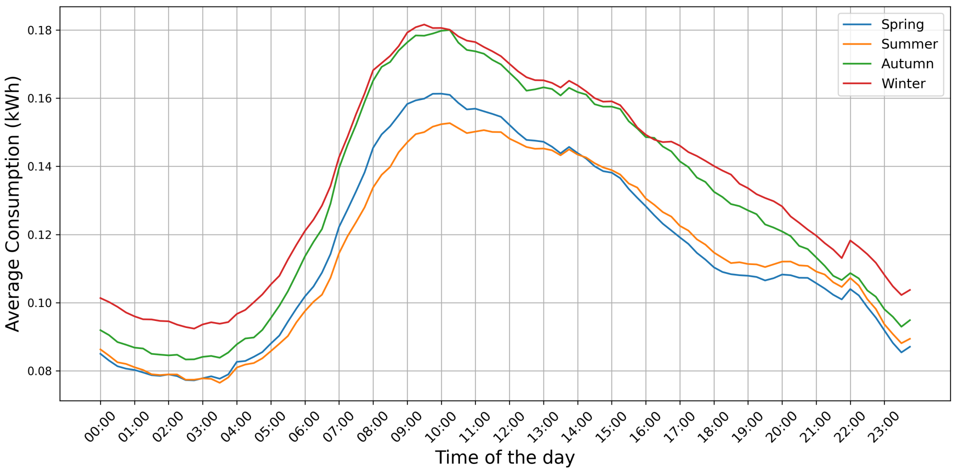


Figure 3.4: Visual analysis of low-consumption profiles by season.

As shown in Figure 3.4, all seasonal profiles share the same general structure, but their amplitude and timing vary. The winter curve displays both higher baseline consumption and a sharper morning peak, likely driven by heating demand and earlier lighting use. On the other side, the summer and the spring profiles are the lowest, as expected from hotter periods of the year.

3.5.2. Weekly Variations

In parallel, the distinction between weekdays and non-working days captures behavioral differences tied to typical work and rest days. These temporal dimensions are incorporated into the modeling process to enhance the accuracy and representativeness of the resulting synthetic load scenarios. The plot below provides a clear visualization of these patterns. (see Appendix B.1 for plots of all level consumption).

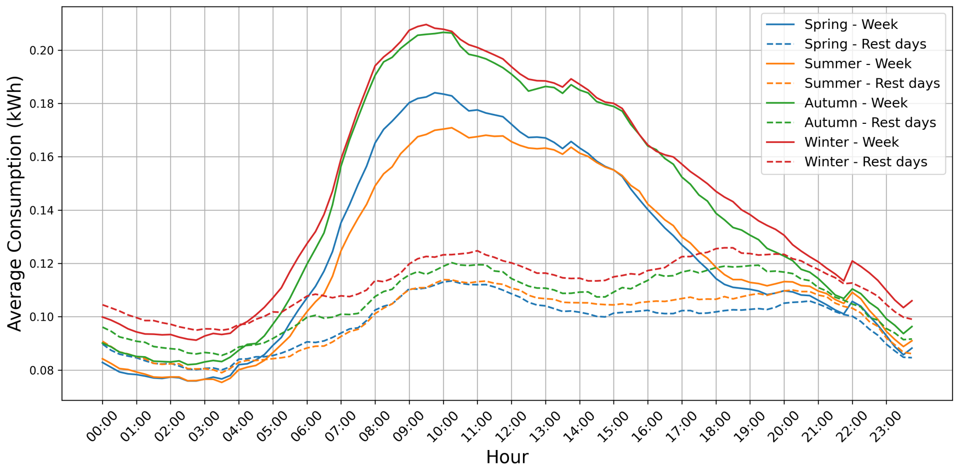


Figure 3.5: Low-consumption patterns across work and rest days.

Figure 3.5 shows the average daily load profiles for the low-consumption cluster, disaggregated by season (spring, summer, autumn, winter) and day type (weekdays vs. rest days). Each curve represents the average hourly consumption in kilowatt-hours (kWh), providing insights into both temporal behavior and seasonal variation among households with lower energy usage. The seasonal dependence is strong and evident: winter weekday profiles show the highest peak, reaching over 0.2 kWh at around 8:00–9:00 a.m. Summer and spring rest days represent the lowest consumption levels, maintaining a relatively flat profile throughout the day. For each season, weekday curves lie above rest day curves, indicating higher average consumption on workdays. On the other hand, rest day profiles tend to be flatter, with lower morning peaks and a more gradual slope.

The higher weekday consumption, even in residential contexts, is likely due to routine-driven behavior: heating systems, lighting, and basic appliances operate in a synchronized way before occupants leave home. Weekend usage is instead spread more diffusely over time and may include partial or full-day absences, reducing average consumption.

4 | Data-Driven Probabilistic Optimal Power Flow

This chapter presents the application of the POPF framework using real household consumption data presented in previous chapters. The goal is to evaluate the behavior and performance of the network under realistic loading conditions, reflecting actual temporal and spatial demand variability.

This step serves as a baseline for comparison with subsequent simulations involving statistically generated consumption profiles and is organized as follows:

1. **Network initialization:** as discussed in Chapter 2.
2. **Setting constraints:** to assure the feasibility of the solution and ensure that all operational limits of the network components are respected.
3. **Assignment of cost functions:** to each generator and external grid.
4. **Start of MC iterations**
5. **Output analysis**

4.1. Setting Constraints

In order to set appropriate constraints for all the runs of the OPF performed during the iterations of the MC algorithm, a pilot run on a limited number of iterations was performed. For this first run, both generators and loads were set free to vary without any too strict limit. As regards the cost functions, all of them, including the ones of the generators as well as the one of the external grid, were set to be identical. In this way none of the components of the network was preferred to another in the optimization process, and the system was free to determine the power dispatch solely based on feasibility and physical network limitations. The obtained results in terms of active power consumption were used to define proper constraints as follows.

Generators Active Power Constraints

To define bounds for the active power output of the generators, lower and upper limits were set based on a statistical margin around the mean expected value. Specifically:

$$P_i^{\min} = \max\{0.05,\, \mu_i - w \cdot \sigma_i\}, \quad i = 1, \ldots, 8$$
$$P_i^{\max} = \mu_i + w \cdot \sigma_i, \quad i = 1, \ldots, 8$$

where:

- μ_i is the mean active power of generator i (computed through the pilot run),
- σ_i is the standard deviation of its expected power (computed through the pilot run),
- w is a scalar width factor which is set to 2.

It's always a good practice to ensure that the $P_i^{\min}$ value never falls below (or really close to) zero in order for the network to be well initialized.

Generators Reactive Power Constraints

Regarding the reactive power limits instead, for each controllable generator, they are set proportionally to its maximum active power output. Specifically, the following bounds are imposed:

$$Q_i^{\min} = -\alpha \cdot P_i^{\max}, \qquad Q_i^{\max} = \alpha \cdot P_i^{\max}$$

where:

- $Q_i^{\min}$ and $Q_i^{\max}$ are the minimum and maximum reactive power outputs of generator i,
- $P_i^{\max}$ is the generator's upper active power limit,

As for what concerns the choice of α, it's used as a proportionality coefficient, and in this work it's set to 0.3 to reflect typical inverter-based capability limits. This corresponds to an operational power factor of approximately 0.95, which is more conservative than the $\text{PF} = 0.9$ often assumed in standards such as Demirok et al. (2011) [15].
To be more clear, the power factor (PF) of a generator is defined as the ratio between active power P and apparent power S:

$$\text{PF} = \frac{P}{\sqrt{P^2 + Q^2}} \tag{4.1}$$

Rearranging 4.1 allows us to express the reactive power Q as a function of P and PF obtaining:

$$Q = P \cdot \tan\left(\arccos(\mathrm{PF})\right) \tag{4.2}$$

For example, assuming a power factor of PF $= 0.9$, the corresponding reactive-to-active power ratio is:

$$Q = P \cdot \tan\left(\arccos(0.9)\right) \approx 0.484 \cdot P$$

This implies that generators operating at PF $= 0.9$ can provide reactive power up to approximately $\pm 48.4\%$ of their active power. The model used in this work hence assumes more restrictive bounds, limiting the reactive power capability to $\pm 30\%$ of the generator's maximum active power P_{max}. This corresponds, following from 4.2, to a power factor of approximately 0.954.

Load Reactive Power Constraints

To reflect realistic behavior while allowing for some flexibility in the power factor of loads, reactive power limits are imposed on each load as a percentage of its nominal reactive power demand. Specifically, the bounds are defined as:

$$Q_{\text{load}}^{\text{min}} = 0.9 \cdot Q_{\text{nom}}, \qquad Q_{\text{load}}^{\text{max}} = 1.1 \cdot Q_{\text{nom}}$$

where Q_{nom} is the initial (nominal) reactive power value assigned to the load in the original dataset or network model (see Appendix A).

This formulation allows for a $\pm 10\%$ deviation from the nominal value, enabling the OPF to slightly adjust the reactive power absorbed by the loads. While loads are generally not controllable in practice, allowing limited flexibility on Q can help achieve feasible OPF solutions and better voltage regulation without significantly distorting demand characteristics.

External Grid Constraints

The external grid, modeled as a slack bus in the network, is configured with both active power and voltage constraints to reflect realistic operational boundaries. The active power exchange with the external grid is limited as follows:

$$P_{\text{ext}}^{\text{min}} = 0 \text{ MW}, \qquad P_{\text{ext}}^{\text{max}} = 60 \text{ MW}$$

These bounds prevent unrealistic import or export values and ensure that local generation is prioritized up to the system's capacity before relying on external supply.

Additionally, voltage magnitude limits are enforced at all buses in the network to ensure secure operation within standard voltage quality thresholds:

$$V_{\text{bus}}^{\text{min}} = 0.95 \text{ p.u.}, \qquad V_{\text{bus}}^{\text{max}} = 1.05 \text{ p.u.}$$

These limits are consistent with typical voltage regulation standards for medium-voltage distribution networks. The external grid thus serves as a voltage reference while participating in active power balancing within constrained, market-aligned conditions.

Transformer and Line Loading Constraints

To avoid thermal overloading and ensure safe operation of the network infrastructure, constraints are imposed on both transformers and distribution lines. Specifically, the maximum allowable loading for each component is set as a percentage of its nominal capacity:

$$\text{Loading}_{\text{max}}^{\text{trafo}} = 95\%, \qquad \text{Loading}_{\text{max}}^{\text{line}} = 95\%$$

This means that neither transformers nor lines are permitted to operate beyond 95% of their rated apparent power. This conservative threshold accounts for measurement uncertainty, thermal margins, and potential safety factors in real-world operation.

By incorporating these loading constraints into the Optimal Power Flow (OPF) problem, the model prevents unrealistic or unsafe dispatch solutions that could compromise the reliability of the distribution system.

4.2. Cost Function Modeling

To define the economic dispatch behavior of the generators and the external grid, quadratic cost functions were assigned using the `pp.create_poly_cost()` function in pandapower. To simulate more realistic scenarios with respect to the previous pilot run, the cost functions were modeled to enforce a strong preference in operating around the expected mean output for the generators and, at the same time, to discourage excessive usage of the external grid.

Cost Function for External Grid

For the external grid connection, a quadratic cost function of the form

$$C(P) = c_2 P^2 + c_1 P + c_0$$

was defined with the following coefficients:

- $c_2 = 0.001$: even if in the real world the energy cost is usually linear, this imposes a small quadratic penalty for deviations from zero, which can help the convergence of the algorithm,
- $c_1 = 80$: a relatively high linear cost to discourage active power import; moreover, the usual market price for energy falls between 50 and 100 €/MWh,
- $c_0 = 0$: no fixed cost.

This configuration makes the use of external grid power economically unfavorable, encouraging the optimization to favor local generation when feasible.

Cost Function for DERs Generators

For each generator, the idea behind the modeling of the cost function was for it to be minimal at the generator's expected mean power output μ_i. This preference was implemented by expressing the cost function in the form:

$$C(P_i) = c_2^{(i)} (P_i - \mu_i)^2$$

which expands to:

$$C(P_i) = c_2^{(i)} P_i^2 - 2c_2^{(i)} \mu_i P_i + c_2^{(i)} \mu_i^2 \tag{4.3}$$

Equation 4.3 was implemented using the following coefficients:

- $c_2 = 2.0$: strength of the quadratic penalty, useful for improving convergence,
- $c_1 = -2c_2\mu_i$: even if by itself is negative, it ensures that the minimum is located at $P_i = \mu_i$,
- $c_0 = c_2\mu_i^2$: fixed cost.

This formulation promotes dispatch solutions that keep each generator's output close to its statistically expected value, improving stability and consistency across MC simulations.

4.3. Overview of the MC Simulation Framework

It is important to note that since the goal of this procedure is to estimate the probability density functions of load consumption, focusing on relatively small time windows is likely the best approach. Based on this consideration, the following analysis will be performed on three specific time intervals, one at a time: 6–9 a.m., 12–3 p.m., and 7–10 p.m. It is to remember that each time interval is, in its turn, divided into thirteen time steps of fifteen minutes each. Now, for all time steps t and for a `n_iteration` times each, a random date is drawn, and loads are sampled for that specific date at that specific time instant. The Optimal Power Flow is then solved for those specific loads, and the results, namely generator and external grid consumption, total cost, bus information, and line loading percentages, are saved in an array. The algorithm, after `n_iteration`, will then proceed to move to the following time step $t + 1$ and repeat itself. After reaching the last time step and performing all necessary iterations, the algorithm will then stop.

4.4. Loads Sampling Strategy

During each iteration of the MC process, loads are sampled for every time instant at specific (but randomly extracted) days. The idea behind the sampling procedure is to make use of real data to reconstruct plausible and temporally consistent load profiles that reflect the diversity of household electricity consumption. Among all loads already present in the network after its initialization (Table A.1), by looking at their active power consumptions, eight were selected to be approximated using real consumption data (nominally the ones at buses 3, 4, 5, 6, 8, 10, 11 and 14). Since their consumption magnitudes are characteristic of aggregated household demand (fig. 3.3), to each selected load has been assigned a different number of profiles, drawn from low, medium, and high consumption (sec. 3.4) in such a way that their proportions reflect realistic load composition patterns during different times of the day. Specifically, three distinct sampling schemes were defined: 60% - 30% - 10% (morning), 40% - 40% - 20% (afternoon), and 20% - 40% - 40% (evening), respectively, for low, medium, and high groups. During every run of the MC process, each load in the network was constructed by selecting one real consumption profile per group from the same day (to preserve natural daily correlations between households), scaling each by the number of users per group, and summing them to obtain the total aggregated demand. In order to maintain statistical consistency, a normalization coefficient was

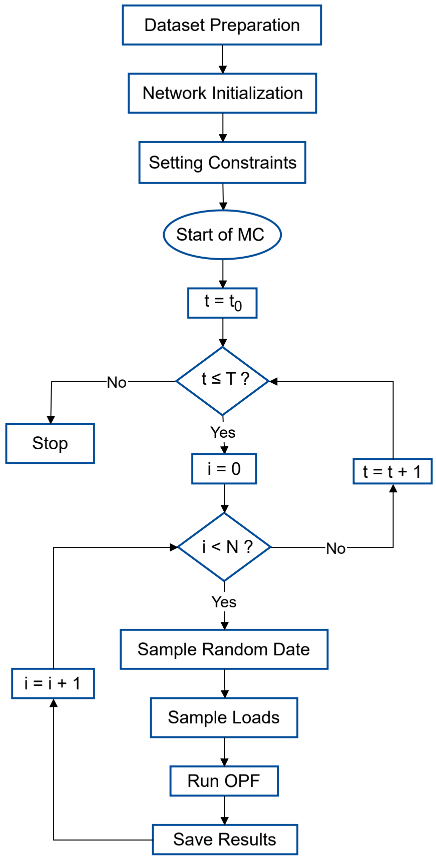


Figure 4.1: Visual representation of MC steps

introduced. By applying a scaling factor k, calculated from the real average total consumption ratios between time slots, the simulated loads are explicitly adjusted so that the aggregate proportions between consumptions in each time slot correspond to those found in the dataset. This approach ensures that the simulated time slots reflect both the diversity of real users and the correct temporal dynamics observed in practice. To better understand the dynamics of the MC process, diagram 4.1 can be consulted.

4.5. Results Discussion

At the end of all the iterations, the output of the MC cycle can be represented with the histograms of the generators' consumptions across the three different time slots. Moreover, the probability density functions for these distributions were computed and plotted with a kernel density estimation approach.

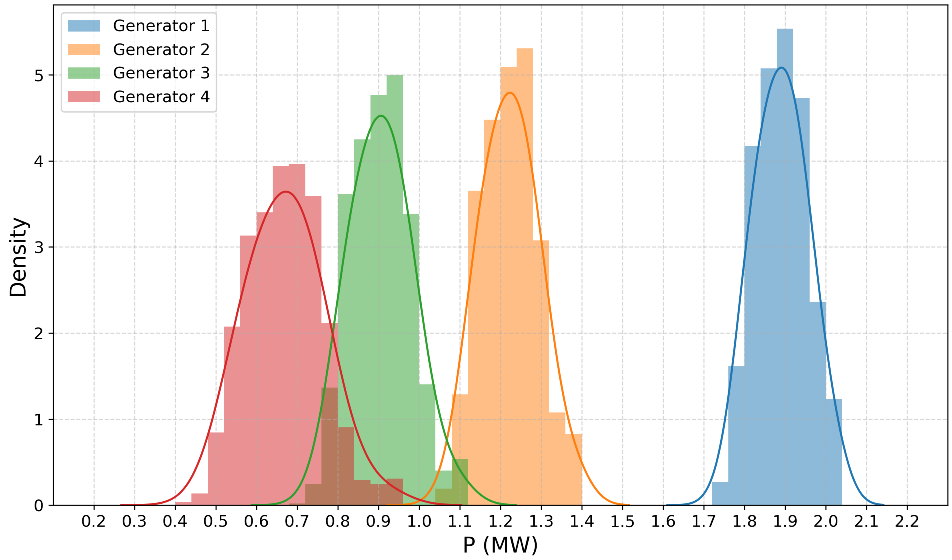
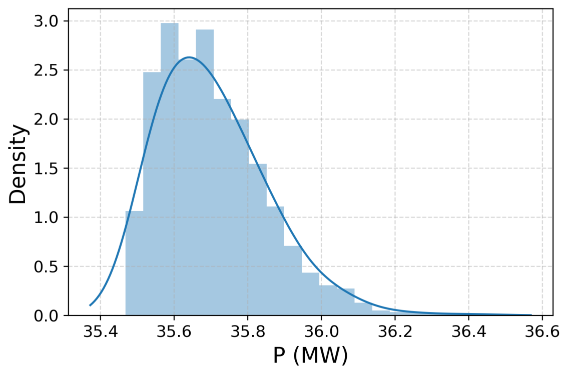
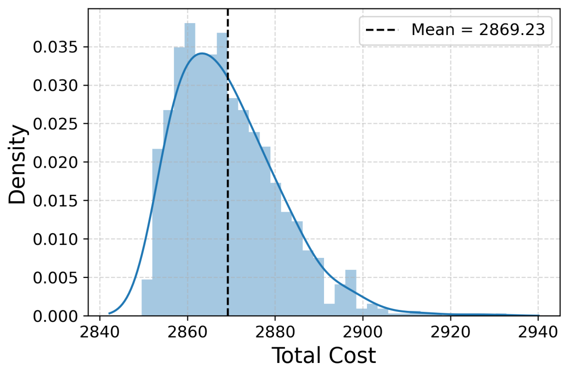


Figure 4.2: Distribution of active power output of the first generators (6-9 a.m.).

With the same thought process, at each iteration, both the external grid consumption and the cost of total energy production are computed, and their histogram as well as their probability density function follow.



(a) External grid power consumption distribution



(b) Total cost distribution

Figure 4.3: Distribution of consumption and costs (6-9 A.M.)

To visualize power generation of all generators across other time slots too, see Appendix B.2. Apart from the lines directly connecting the network to the external grid, which reach loading levels close to the previously set constraint of 95%, all other lines exhibit significantly lower loading percentages, remaining well within the established limits. The same thing can be stated for voltage magnitudes, which always stay in a safe range.

5 | Synthetic Load Modeling and Copula OPF

This chapter presents the ideas and the process behind the creation of synthetic data to be used in the network and to demonstrate how such data can be used for network analysis. The results obtained using synthetic data will then be compared to those derived from real data to understand whether both approaches provide analogous outcomes. This comparison aims at evaluating the reliability of synthetic data as a tool for simulation and decision-making in electrical grids.
Later in this chapter the focus will shift to the analysis of the Monte Carlo simulation results obtained using synthetic data. For illustrative purposes, only the outcomes corresponding to the 6–9 a.m. and 12-3 p.m. time slots will be presented. However, the outputs for the other (evening) time slot exhibit comparable densities with respect to the results obtained in the morning and an overall good quality of fit with respect to their original MC results too. All the left-out results in terms of generator distribution are available in Appendix B.

5.1. Data Generation: PCA and KDE

To create synthetic daily profiles that preserve the main statistical and temporal characteristics of real household data, a two-stage generative method is applied, combining **Principal Component Analysis (PCA)** with **Kernel Density Estimation (KDE)**.

1. **Organize Data by Scenario:**
   To better try exploiting differences in power demand between loads, the three groups of consumption were considered, and for each one of them, a further division of data was performed based on workdays/rest days, and the following procedure was applied separately for each subpart of the whole dataset.

2. **Log-space transformation:**
   Given each dataset $X \in \mathbb{R}^{n \times 96}$, where each row represents one daily profile sampled

at 15-minute resolution, a logarithmic transformation was applied to reduce potential skewness or heavy-tailed data, a phenomenon typical in real-world datasets like energy consumption:

$$X_{\text{log}} = \log(1 + X)$$

Note that electricity consumption data often contains zeros, especially in residential contexts. This transformation safely maps zero-valued consumption observations, improving numerical stability.

3. **Principal Component Analysis (PCA):**
   Perform dimensionality reduction by projecting the log-transformed data onto a lower-dimensional subspace:

   $$X_{\text{PCA}} = \text{PCA}(X_{\text{log}})$$

   This step captures the most significant information while reducing the complexity of the dataset.

4. **Kernel Density Estimation (KDE) and Sampling:**
   It is a non-parametric density estimation technique, and it is applied to approximate probability distributions. Let $X_{\text{PCA}} \in \mathbb{R}^{n \times d}$ denote the matrix of PCA-transformed observations from the previous point, where $d \ll 96$ is the number of retained principal components. The goal is to estimate the continuous multivariate probability density function $\hat{f}_{\text{PCA}}$ over this d-dimensional space. Then, sample n_{samples} new points from this estimated distribution:

   $$\tilde{X}_{\text{PCA}} \sim \hat{f}_{\text{PCA}}$$

5. **Inverse Transformations:**
   Map the synthetic points back to the original feature space:

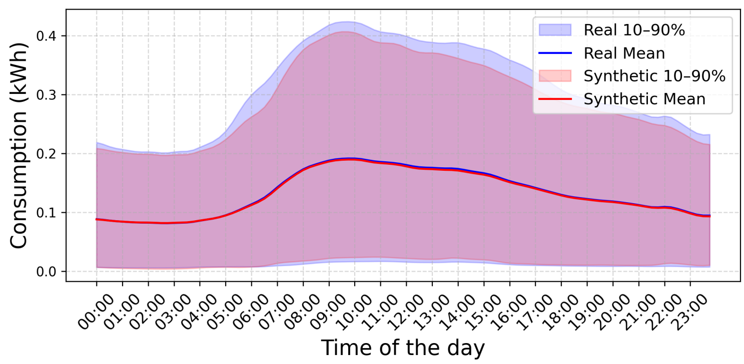
   $$\tilde{X}_{\text{log}} = \text{PCA}^{-1}(\tilde{X}_{\text{PCA}})$$

   Then, apply the inverse of the log transformation: $\tilde{X} = \exp(\tilde{X}_{\text{log}}) - 1$

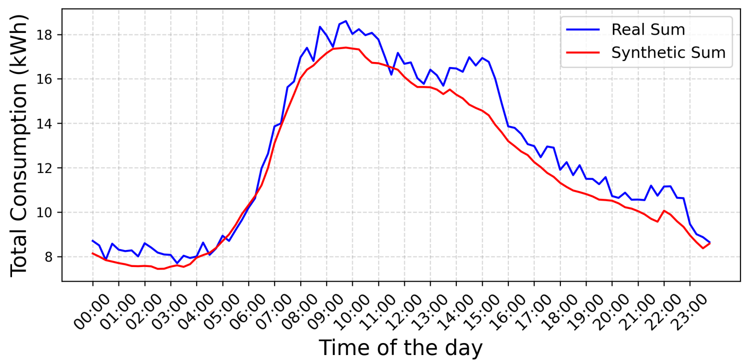
6. **Post-processing:**
   Clip zero values to a small positive threshold since there are not loads that behave like a switch and that can be exactly zero during the course of a day.

   $$\tilde{X}_{\text{final}} = \max(\tilde{X}, 10^{-3})$$

This estimation and sampling procedure provides some good results in terms of effectiveness at capturing marginal distributions and accuracy in representing real data (as can be seen from Figure 5.1). If this data is fed to the MC algorithm, it will provide just slightly worse results compared to what is going to be presented in the next section. The main reason for this is that this procedure lacks one crucial feature: it does not preserve structured dependencies across different consumption groups. However, with this setting and from a computational standpoint, this method still provides acceptable results, but it might not generalize well to situations in which the correlation between case study groups is going to be higher and more definite.



(a) Daily consumption profile of random subset aggregates



(b) Aggregated profiles of 100 households during workdays

Figure 5.1: Visual analysis of a hundred low-profiles' consumption patterns

5.2. Data Generation: Copula

As mentioned, the previously explained data generation process carries some problematic: first, it can be observed that generating data, group by group, is indeed an independent procedure and typically doesn't account for seasonality, day type, or hour-specific behavior unless this is encoded explicitly in preprocessing (which often leads to fragmented or high-dimensional data). Moreover, the effect of other phenomena like particular weather conditions, blackouts etc. won't be taken into account: by generating consumption profiles independently, all these coordinated behaviors of the loads will end up being lost.

The procedure that is now going to be presented instead, and which is based on copulas, aims at constructing a 3-dimensional model, one for each combination of season (hot/cold), day type (weekday/non-working day), and time slot, to capture the joint dependency structure between the low, medium, and high consumption groups at each time instant.

1. **Organize Data by Scenario:**
   The dataset is split into the three consumption groups: low, medium, and high, and further into seasonal and day-type scenarios (e.g., cold weekday, hot weekend). Each group contains real consumption profiles.

2. **Sample Values at Specific Time Steps:**
   For each time instant of interest, a large number of samples are drawn from the three main groups.

3. **Normalize Using Empirical Distributions:**
   Each set of sampled values is converted to the unit interval $[0, 1]$ using its empirical cumulative distribution function (ECDF). This step standardizes the data and prepares it for copula modeling, where marginals must be uniform. A visualization of transformed data $u \sim \mathcal{U}[0, 1]$ follows:

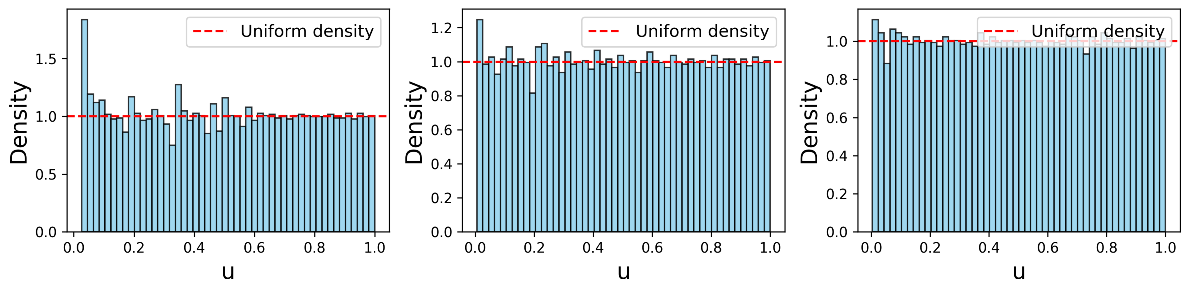


Figure 5.2: Transformed data from each group (from left to right, low-medium-high consumption) during the morning time slot.

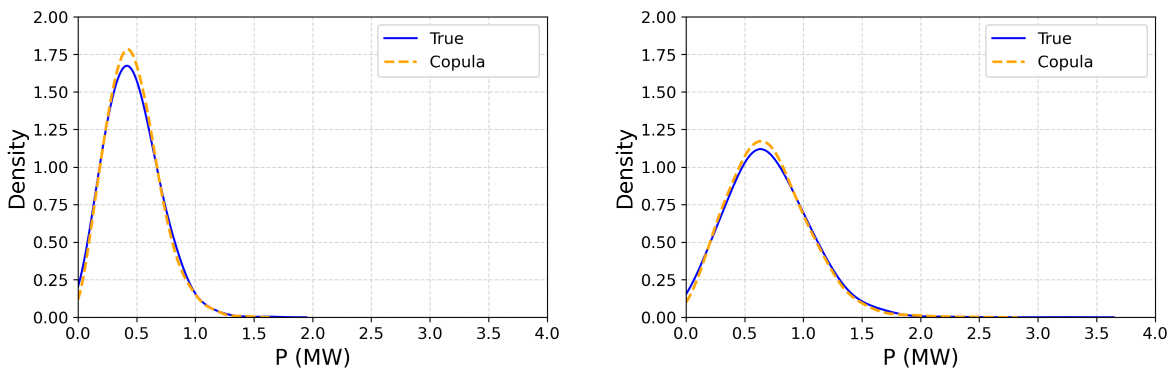
4. **Fit a 3-Dimensional Copula:**
   The three normalized vectors are combined to form a matrix of pseudo-observations, which is used to fit a 3-dimensional vine copula. This copula models the dependency structure between the three groups for that specific time and scenario.

5. **Store Models and ECDF for Later Sampling:**
   Remembering that for each (season, day type) framework, a copula is fit between the low, medium, and high consumption groups, using only the households relevant to that scenario, the fitted copula (and the ECDFs) of each group are stored for each scenario and time step. These can later be used to generate new, dependency-preserving synthetic profiles by first sampling from the copula in $[0,1]^3$ and then inverting the ECDFs to map back to realistic consumption values.

Now it is possible to proceed following what is explained in section 4.3 to set all that is needed to perform a new MC run. The only difference will be that this time the sampling will be performed from a new synthetic dataset. This dataset will be created beforehand of the MC cycle in such a way that at each iteration loads can be sampled, proceeding as it was done in (4.4) but this time each sampled point will initially be from a $\mathcal{U}[0,1]^3$ and only later it will be transformed, to actually represent a consumption value, with the ECDF. Finally, all the constraints will not be changed with respect to section 4.1.

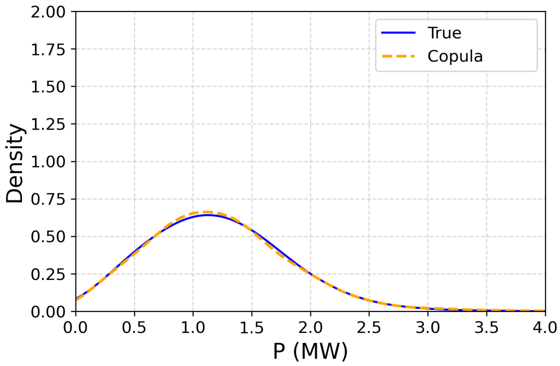
5.3. Comparison with Real Data

Before actually proceeding with the Monte Carlo algorithm that will lead to the estimation of output probability density functions of DER generators, the synthetic data generation process can be tested: by simulating what represents only the initial step of the Monte Carlo algorithm, namely, the load sampling stage (Figure 4.1), and performing it using both real and synthetic datasets, a preliminary assessment of the potential accuracy of the results and of the quality of the data generating process can be made. By examining the time slot 12-3 p.m. which is where the highest variability in terms of loads is expected, we can observe the following results:

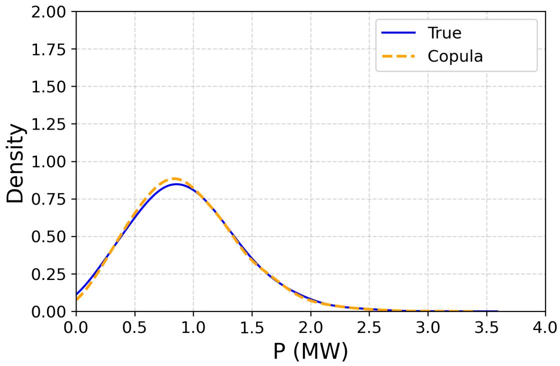


(a) Load 1 density comparison.

(b) Load 2 density comparison.



(c) Load 3 density comparison.



(d) Load 4 density comparison.

Figure 5.3: Density comparison between real and synthetic data for four representative loads in the afternoon time slot.

This time slot was selected for analysis due to its typically high variability in load demand. This period often coincides with midday residential activity, commercial operations, and peak solar generation, increasing the uncertainty. As expected, in it can be found really low-centered loads as well as high-centered ones with a different variability across each one of the consumption patterns. The density plots show a strong connection between the true and copula-based distributions for all the presented loads (the other four of them are omitted just for conciseness but still present the same quality of fit as the ones shown before).

5.4. Results Discussion

This section focuses on the analysis of the Monte Carlo simulation results obtained using synthetic data. For illustrative purposes, only the outcomes corresponding to the 6–9 a.m. and 12-3 p.m. time slots are presented here. However, the outputs for the other time slot exhibit a comparable quality of fit with respect to the original results too.

5.4.1. Morning and Afternoon Consumption Patterns

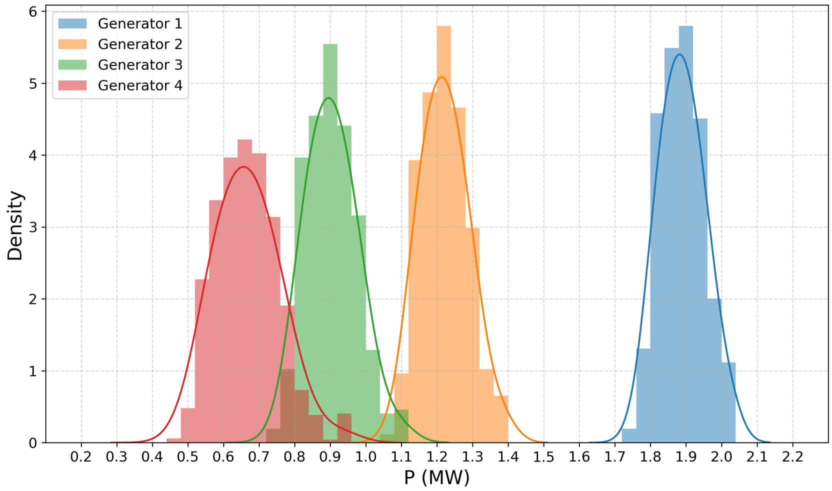


Figure 5.4: Morning consumption patterns.

Figure 5.4 illustrates the power output distributions of four generators during the morning consumption interval. Each generator exhibits a unimodal and well-separated density, indicating distinct dispatch levels. Generator 1 operates at the highest power range (centered around 2.1 MW), while Generator 4 operates at the lowest (around 0.9 MW), suggesting a clear prioritization in generator loading. The limited variance of each distribution denotes consistent behavior across the Monte Carlo realizations.

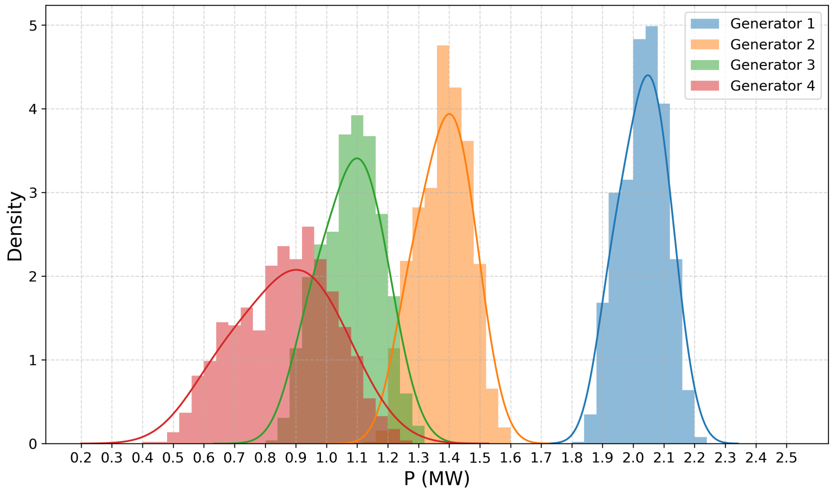


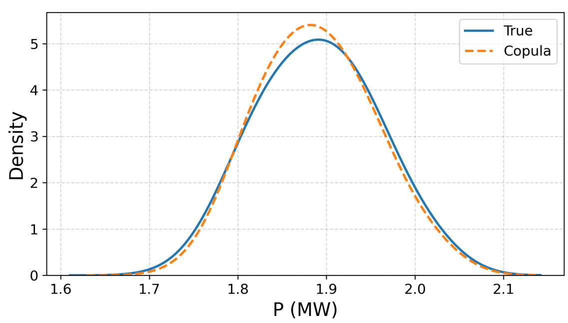
Figure 5.5: Afternoon consumption patterns

Figure 5.5 shows the power output distributions during the 12-3 p.m. interval of the same four generators shown in figure 5.4. Compared to the previous case, all distributions

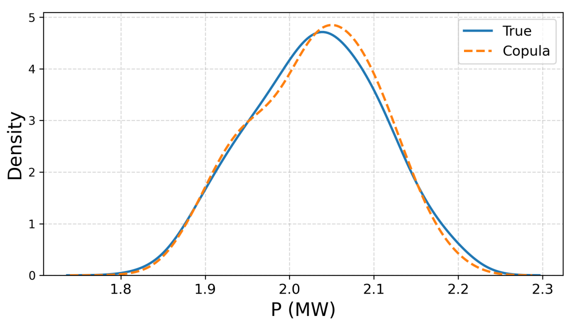
are shifted to the right, indicating an overall increase in generator dispatch levels. Generator 1 continues to supply the largest amount of energy, now peaking around 2.3 MW, while Generator 4 also increases its contribution to approximately 1.1 MW. Despite the higher load, the separation between generators is not as clear as in figure 5.4, but it remains clear, and the distributions retain their narrow shape, suggesting that the system maintains a consistent allocation of generation even under elevated demand conditions.
The distributions show a high concentration of values around the mean, which aligns with the cost function being centered at that point. This indicates that, in most realizations, generators tend to operate close to their optimal setpoints, suggesting a stable and comfortable operating regime. However, the presence of high variance reveals that in certain scenarios, significant deviations from the mean are still required, highlighting residual instability or sensitivity of the system under more demanding conditions.

5.4.2. Real vs Synthetic Loads Outputs Comparison

Figures 5.6 and 5.7 compare the power output distributions of Generator 1 and Generator 4 across two time intervals: 6–9 a.m. (morning) and 12–3 p.m. (afternoon), using both true and copula-based samples. For both generators, the copula-based density approximates the true distribution closely, capturing the general shape in each time window. However, some discrepancies are observable: in the afternoon periods (plots b and d), the copula tends to slightly overestimate the peak density, indicating a tendency to concentrate values more tightly around the mean.

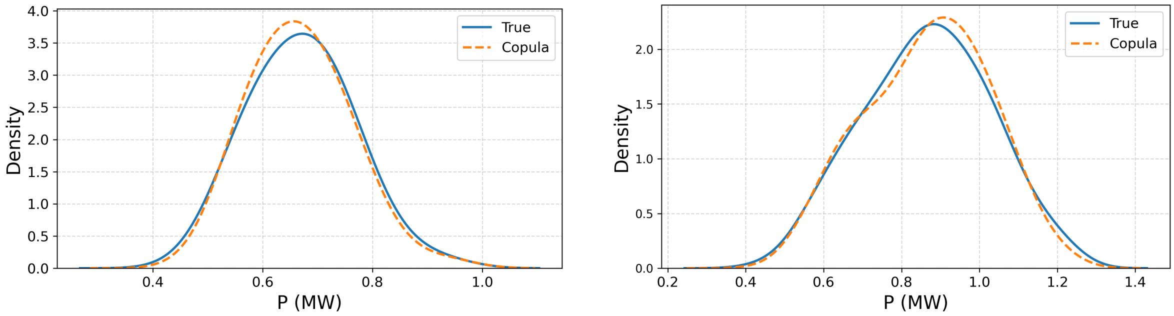


(a) Gen 1 consumption 6–9 a.m.



(b) Gen 1 consumption 12–3 p.m.

Figure 5.6: Density comparison of generator 1 across morning and afternoon: true data vs. copula model.



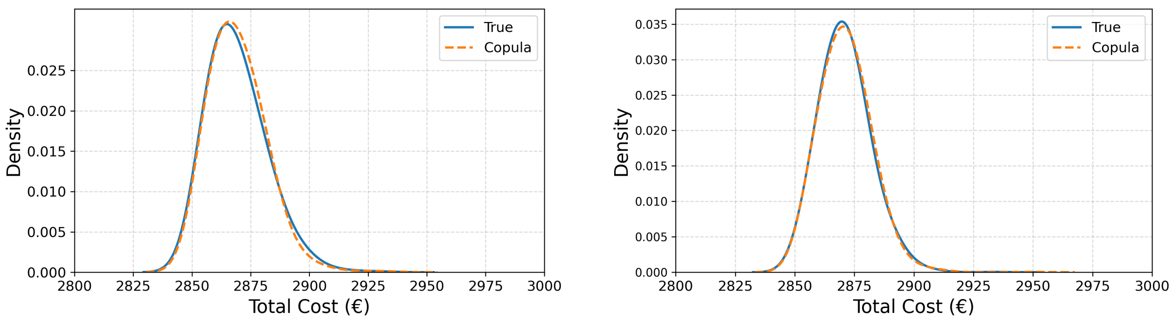
(a) Gen 4 consumption 6–9 p.m.

(b) Gen 4 consumption 12–3 p.m.

Figure 5.7: Density comparison of generator 4 across morning and afternoon: true data vs. copula model.

5.4.3. Morning and Afternoon Cost Distributions

Figure 5.8 presents a comparison between the true and copula-based density estimates of the total operational cost during two time intervals: morning (6–9 a.m.) and afternoon (12–3 p.m.). In both cases, the copula model closely approximates the true cost distribution, and it reproduces the shape, peak location, and spread accurately. The morning distribution (a) displays a slight positive skew, which is well captured by the copula. In the afternoon (b), the cost distribution shifts rightward, indicating higher expected total costs. Overall, the copula model demonstrates robustness in replicating cost variability in each scenario.



(a) Morning cost distribution

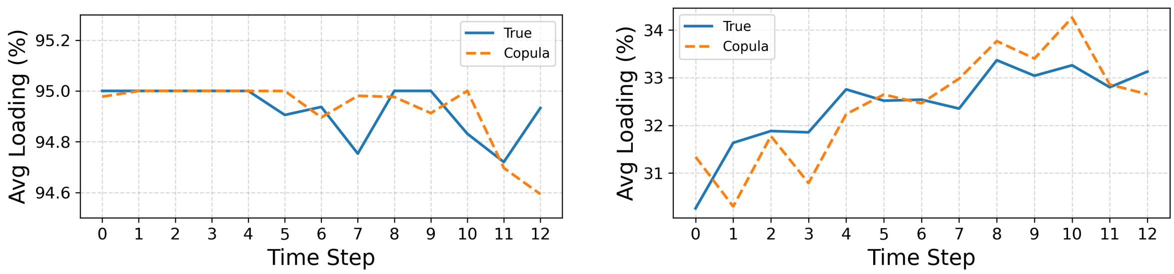
(b) Afternoon cost distribution

Figure 5.8: Density comparison of total costs: true data vs. copula model.

5.4.4. Lines Loading Inspection

In figure 5.9 the trend of line loading percentage for some reference line of the network is also analyzed. Each graph is obtained by averaging the load level of that specific line for each time step and across all the realizations of the MC process that occurred in that specific time instant. It has to be said that, for computational reasons, the number of runs of the MC algorithm for a single specific time is around one hundred. That's the main reason for the discrepancy in the plots, even though the general trend is successfully preserved even by using the copula-generated dataset.

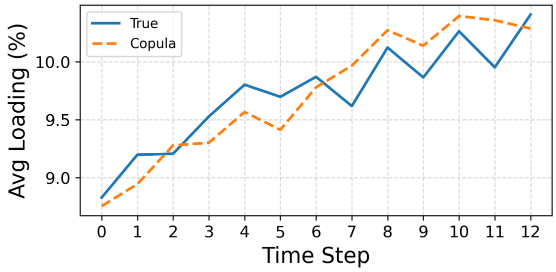
The three specific lines were selected based on their relevance within the network: line 0 is located directly under the external grid connection and acts as the main feeder, consistently aggregating the power flow required by all downstream buses not supplied locally by DERs. As such, it typically operates near its thermal limit, and it plays the role of a natural bottleneck in the network, making it a key indicator of global loading stress. Overall, it never overcomes the boundaries set as constraints, but its pattern, together with the fact that the majority of other lines operate way under their limit, suggests an unbalanced usage of the network's capacity. This behavior can be attributed to the nature of the network: all generators in the system are DERs. As a result, during periods when local generation is insufficient, the external grid becomes the main supply source. Line 9 is instead part of a side branch that creates an alternative path in the radial network. Analyzing it helps understand how the OPF manages power flow across optional or parallel routes. Line 10, finally, is a long line located toward the end of the feeder. Due to its electrical characteristics and physical distance from the substation, it's probably the one more susceptible to voltage drops.



(a) Line 0 loading pattern – morning

(b) Line 9 loading pattern – morning

Figure 5.9: Loading patterns of Lines 0 and 9 during the 6–9 a.m. time slot.



(c) Line 10 loading pattern – morning

Figure 5.9: Loading patterns of Line 10 during the 6–9 a.m. time slot.

5.5. Strengths and Limitations of the Approach

This chapter provides an evaluation of the methodology adopted throughout this work by exposing its primary strengths and inherent limitations.

One of the main strengths of the approach utilized in this work relies on the implementation of a data-driven load modeling procedure: it certainly ensures realism, and it reduces reliance on potential oversimplifying load assumptions. This is even more strengthened by the copula-based sampling, which preserves the spatial dependencies between household groups and the temporal coherence of daily consumption patterns: unlike traditional Monte Carlo methods that sample loads independently, this strategy allows for the generation of synthetic scenarios that remain consistent with observed correlations. By assuming controllable DERs generators (within technically plausible limits), the method also aligns with modern grid scenarios in which decentralized energy resources can participate actively in system regulation. Moreover, the segmentation of the day into time periods with separate load modeling enables a more precise capture of demand variability throughout time, resulting in better-informed operational strategies. Finally, the structure of this work allows for flexible testing of alternative sampling strategies, generator cost functions, and network configurations, making it adaptable to a variety of different research directions or practical scenarios.

As for limitations, some of them are actually noticeable by observing Figures 5.4 and 5.5: it's true that for different time slots in a day, generators tend to stabilize, assuming different values of power outputs, but at the same time, their relative consumption doesn't change much, indicating a consistent ordering in their operational roles. This can be reasonably attributed to the fact that this analysis relies on the modeling of just residential loads in the network; all other consumption sources, across consecutive computations of the OPF, keep their active power consumption fixed to the value to which they were set at the initialization of the network (Table A.1). This way, all the variability of the system relies on the variation of the residential loads, which in a real-world scenario constitute just a fraction of the total energy consumed in a network.

This also raises another relevant consideration: even if the segmentation of load profiles into low, medium, and high consumption groups was functional for the analysis and led to clearly distinguishable categories, the correlation structure between these groups, while partially self-evident and exploited by copulas, is not as informative as, for instance, the correlation between fundamentally different load types such as residential and commercial profiles. Moreover, the temporal dependency is not explicitly modeled: the rescaling factor applied during the sample phase is just a starting point strategy to preserve autocorrelation in the load time series. There are, in fact, more advanced ways to include this aspect in the analysis, such as time series models (e.g., ARMA, Gaussian Processes) or copula-based models extended to the temporal domain.

6 | Conclusions and Future Developments

This thesis presented a data-driven framework enhanced with load correlation modeling for performing Probabilistic Optimal Power Flow (POPF) in distribution networks under realistic load uncertainty. The procedure made use of real household demand data divided into consumption-specific bands. Both copula-based load modeling and sampling strategy were employed to preserve spatial dependencies between consumer groups and maintain temporal coherence, overcoming strong assumptions on either fixed consumption or on the arbitrary choice of load prior distributions for modeling uncertainty. This setting enabled the development of a Monte Carlo approach for the analysis of time-varying scenarios of distribution networks with integrated DERs. Some statistics about POPF's results obtained both with real and synthetic data are summarized in the following tables:

Time Slot	**Avg Total Cost [€]**	**Std Total Cost [€]**	**Avg Ext Grid P [MW]**	**DER Coverage [%]**	**Lines > 90% Loading (avg)**
Morning	2869.225	11.734	35.703	20.908	1.986
Afternoon	2871.017	10.012	35.726	23.537	2.000
Evening	2864.363	9.035	35.644	21.117	1.989

Table 6.1: Summary of real data OPF metrics across different time slots.

Time Slot	**Avg Total Cost [€]**	**Std Total Cost [€]**	**Avg Ext Grid P [MW]**	**DER Coverage [%]**	**Lines > 90% Loading (avg)**
Morning	2869.265	11.224	35.704	20.868	1.989
Afternoon	2871.247	10.058	35.729	23.543	2.000
Evening	2863.751	8.650	35.636	21.046	1.995

Table 6.2: Summary of synthetic data OPF metrics across different time slots.

The comparison between the OPF results obtained using real and synthetic data highlights a strong consistency across key operational metrics. Total costs, external grid power usage, and DER coverage remain remarkably similar, indicating that the synthetic data generation method preserves the statistical properties and structural dependencies of the original dataset. Minor variations in standard deviations, particularly in the evening time slot, reflect slight differences in variability but do not compromise overall system behavior. The stability of the line loading indicators across all time slots further confirms that the synthetic profiles replicate network stress patterns with high fidelity.
Focusing more specifically on the values assumed by each metric, the average total cost is highest in the afternoon, in line with the typical residential load distribution, which tends to peak during this period. The cost variability is also highest in the morning, when the system begins ramping up and load dynamics are more unpredictable. In the evening, variability drops significantly, indicating more regular and controlled behavior. The average power drawn from the external grid, instead, remains relatively stable across all time slots (around 35.7 MW), with a slight drop in the evening, consistent with lower system demand. The DER coverage peaks in the afternoon (23.5%), likely driven by solar PV generation, and drops slightly in the morning (20.9%). This confirms that distributed generators contribute most effectively during mid-day hours. Finally, the number of lines operating above 90% of their thermal limit stays consistently close to two across all time slots. These are likely the lines connecting the external grid to the distribution network. Their high loading is a structural consequence of a system composed entirely of DERs. Since these generators have very low marginal costs, the OPF tends to exploit them as much as possible, resulting in consistently high usage of key transfer lines.

Talking more in general about the implication of these results on the methodology of this work, and in particular regarding loads modeling, results demonstrated that even without assuming any preemptive structure for the inputs, the approach managed to retain flexibility in representing diverse load behaviors. In particular, from a sampling standpoint, the implemented procedure to extract loads that will be later used in the Monte Carlo method ensured that the time autocorrelation between consumption time series was preserved by sampling from one load distribution for each time instant and rescaling it by multiplying by an appropriate weight. Compared to an approach that involves extracting inputs from various households' time series at a common time instant, the implications on the results are significant: in the latter case, the total load curve would be way smoother due to an averaging effect of load distributions. This would translate into DERs likely experiencing milder fluctuations. By performing the sampling as previously mentioned and as explained in more detail in section 4.4, all the sharp load peaks are retained, implying that the followingly implemented POPF needs to work harder to maintain feasibility. This certainly leads to a more precise representation of worst-case scenarios as well as to the opportunity to examine temporal stress propagation of, for example, lines loading, as analyzed in subsection 5.4.4 of the results discussion chapter. At the same time, this brings to the analysis also some negative aspects: due to the higher variability of inputs even in consecutive time steps, each iteration of the OPF needs to be really well constrained, and, especially in time slots with typically a higher demand in terms of households' power consumption is expected, the output might end up being more uncertain. In this work's case, this can be observed in subsection 5.4.1 of the results discussion chapter, more precisely in Figure 5.5 where more relaxed constraints were required in order to guarantee the convergence of the OPF and, consequentially, higher standard deviations in the output PDFs were observed. This is, in fact, a limitation of the method: enforcing a higher degree of realism in the system may lead to the intrinsic creation of more unstable scenarios carrying a higher level of variability and resulting in less precise and harder-to-interpret outcomes, especially in time scenarios of higher energy demand and higher stress of the network. Another factor that has to be taken into account is the time-splitting decision: analyzing the response of the network across multiple different scenarios was one of the main goals of this work. The applied procedure certainly enables extension to other time periods, but at the same time it might break the natural flow of the daily load profiles by losing the ability to model inter-period dynamics (e.g., the capability of explaining how a high load in the morning leads to behavior in the afternoon). However, mainly for computational reasons, it is difficult to perform accurate analysis by considering just one time period which extends to a whole day.

Regarding the computation involved in the applied methodology, that is, the Monte Carlo approach for different time slots, the results demonstrated that the proposed approach is able to produce interpretable and realistic generator dispatch patterns across different time periods with DER generators which adapted to residential load variations with consistent operational roles, highlighting the method's ability to capture temporal variability in network operation. Even if all these factors contributed to proving the method's reliability in handling time-varying uncertainty, one of the main challenges associated with this framework is its computational cost: data was already segmented into one data point every 15 minutes, and time slots were defined with a duration of three hours each; furthermore, the MC method entails iterating the same process (sampling loads and running the OPF) a considerable number of times, resulting in a total computational complexity of $\mathcal{O}\left(T \cdot N \cdot T_{\mathrm{OPF}}(n_{\mathrm{loads}})\right)$ where T is the number of data points in each time slot, N is an arbitrary number indicating the amount of iteration that the MC process will run for each time stamp, and $T_{\mathrm{OPF}}(n_{\mathrm{loads}})$ is the total nonlinear cost of a single OPF run with

n_{loads} being, of course, the total number of loads for which samples are extracted. (Note that the load sampling step contributes with a linear cost of $\mathcal{O}(n_{\text{loads}})$ which is negligible compared to the OPF solve). All of this makes the MC approach one of the most reliable procedures for this kind of analysis, but its increasing complexity must be taken into consideration when evaluating an eventual generalization to longer time intervals.

Finally, copula-based data correlation modeling is maybe the most common approach to modeling dependence between loads: the Gaussian assumptions for data are not required, and marginals can be empirical. This avoids oversimplification and enables more realistic modeling of skewed or multimodal load patterns. Moreover, vine copulas, the one used in this work, are known for their ability to capture extreme co-movements like days when all users spike due to weather, holidays, etc. These are behaviors that are hard to capture with traditional correlation matrices. However, copulas model joint behavior at a single time instant but not how consumption evolves over time; this limits the ability to capture lagged or sequential dependencies unless extended with time-varying copulas or Markov structures.

There are several potential extensions to the analysis that could be considered: the inclusion of more heterogeneous load types among the ones to model is clearly one of the main directions to explore. Introducing commercial and industrial time-varying consumption data would enrich the diversity of load behaviors and allow for more accurate responses of generators while at the same time exploiting better and stronger correlation structures across load categories. It is also important to note that the operational constraints for DER generators were defined using a fixed rule (see 4.1), uniformly applied across all time slots. While this ensures consistency, it may not reflect the actual availability of generation resources, especially in the evening, when solar PV production is no longer active. A more realistic modeling approach could involve time-dependent constraint definitions, allowing the OPF to reflect technology-specific limitations more accurately throughout the day. Another aspect that could be developed even more is the design of cost functions: implementing dynamic cost structures based on market or environmental conditions could improve economic realism and support multi-objective optimization. Another natural extension involves refining the temporal modeling of load uncertainty. While this work considered time slots as independent, introducing time series models or Markov-based approaches could help capture inter-temporal dependencies and improve the realism of dynamic system behavior, helping the generalization and the extension of the input modeling procedure to many other scenarios.

Bibliography

[1] Guido Carpinelli, Pierluigi Caramia, and Pietro Varilone. Multi-linear monte carlo simulation method for probabilistic load flow of distribution systems with wind and photovoltaic generation systems. *Renewable Energy*, 76:283–295, 2015.

[2] Chenxi Wu, Fushuan Wen, Youlin Lou, and Feng Xin. Probabilistic load flow analysis of photovoltaic generation system with plug-in electric vehicles. *International Journal of Electrical Power & Energy Systems*, 64:1221–1228, 2015.

[3] Yasin Zabihinia Gerdroodbari, Abu Bakr Pengwah, Reza Razzaghi, Rahmat Heidari, and Lachlan L.H. Andrew. A method to control distributed energy resources in distribution networks using smart meter data. *International Journal of Electrical Power & Energy Systems*, 153:109293, 2023.

[4] PandaPower documentation. Optimisation problem: constraints and cost function explained.

[5] S. Shargh, B. Khorshid ghazani, B. Mohammadi-ivatloo, H. Seyedi, and M. Abapour. Probabilistic multi-objective optimal power flow considering correlated wind power and load uncertainties. *Renewable Energy*, 94:10–21, 2016.

[6] U. Leeton, D. Uthitsunthorn, U. Kwannetr, N. Sinsuphun, and T. Kulworawanichpong. Power loss minimization using optimal power flow based on particle swarm optimization. In *ECTI-CON2010: The 2010 ECTI International Confernce on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, pages 440–444, 2010.

[7] Nitesh Ganesh Bhat, B. Rajanarayan Prusty, and Debashisha Jena. Cumulant-based correlated probabilistic load flow considering photovoltaic generation and electric vehicle charging demand. *Frontiers in Energy*, 11(2):184–196, 2017.

[8] J.C. Hernández, F.J. Ruiz-Rodriguez, and F. Jurado. Modelling and assessment of the combined technical impact of electric vehicles and photovoltaic generation in radial distribution systems. *Energy*, 141:316–332, 2017.

[9] K. G. Terry Hollands and Harry Suehrcke. A three-state model for the probability distribution of instantaneous solar radiation, with applications. *Solar Energy*, 96:103–112, 2013.

[10] Umar Hanif Ramadhani, Mahmoud Shepero, Joakim Munkhammar, Joakim Widén, and Nicholas Etherden. Review of probabilistic load flow approaches for power distribution systems with photovoltaic generation and electric vehicle charging. *International Journal of Electrical Power & Energy Systems*, 120:106003, 2020.

[11] Morteza Aien, Ali Hajebrahimi, and Mahmud Fotuhi-Firuzabad. A comprehensive review on uncertainty modeling techniques in power system studies. *Renewable and Sustainable Energy Reviews*, 57:1077–1089, 2016.

[12] K. Strunz, Ehsan Abbasi, Robert Fletcher, Reza Iravani, and Géza Joos. *TF C6.04.02 : TB 575 – Benchmark Systems for Network Integration of Renewable and Distributed Energy Resources.* 04 2014.

[13] Krzysztof Rudion, Antje Orths, Z.A. Styczynski, and K. Strunz. Design of benchmark of medium voltage distribution network for investigation of dg integration. page 6 pp., 01 2006.

[14] Jakub Bendík. Dataset of 15-minute values of active and reactive power consumption of 1000 households during single year, 2023.

[15] Erhan Demirok, Pablo Casado González, Kenn H. B. Frederiksen, Dezso Sera, Pedro Rodriguez, and Remus Teodorescu. Local reactive power control methods for overvoltage prevention of distributed solar inverters in low-voltage grids. *IEEE Journal of Photovoltaics*, 1(2):174–182, 2011.

A | Appendix A

A.1. Additional Tables - Network Description

name	bus	P [MW]	Q [MVar]	S_n [MVar]	scaling	in_service
Load R1	1	14.99400	3.044662	15.300	1.0	True
Load R3	3	0.27645	0.069285	0.285	1.0	True
Load R4	4	0.43165	0.108182	0.445	1.0	True
Load R5	5	0.72750	0.182329	0.750	1.0	True
Load R6	6	0.54805	0.137354	0.565	1.0	True
Load R8	8	0.58685	0.147078	0.605	1.0	True
Load R10	10	0.47530	0.119121	0.490	1.0	True
Load R11	11	0.32980	0.082656	0.340	1.0	True
Load R12	12	14.99400	3.044662	15.300	1.0	True
Load R14	14	0.20855	0.052268	0.215	1.0	True
Load CI1	1	4.85400	1.592474	5.210	1.0	True
Load CI3	3	0.22525	0.139597	0.265	1.0	True
Load CI7	7	0.07650	0.047410	0.090	1.0	True
Load CI9	9	0.57375	0.355578	0.675	1.0	True
Load CI10	10	0.06800	0.042143	0.080	1.0	True
Load CI12	12	5.01600	1.648679	5.280	1.0	True
Load CI13	13	0.03400	0.021617	0.040	1.0	True
Load CI14	14	0.33150	0.205445	0.390	1.0	True

Table A.1: Load Data

name	bus	P [MW]	Q [MVar]	S_n [MVar]	scaling	in_service	type
PV 3	3	0.020	0.020	0.020	1.0	True	PV
PV 4	4	0.020	0.020	0.020	1.0	True	PV
PV 5	5	0.030	0.030	0.030	1.0	True	PV
PV 6	6	0.030	0.030	0.030	1.0	True	PV
PV 8	8	0.030	0.030	0.030	1.0	True	PV
PV 9	9	0.030	0.030	0.030	1.0	True	PV
PV 10	10	0.040	0.040	0.040	1.0	True	PV
PV 11	11	0.010	0.010	0.010	1.0	True	PV
WKA 7	7	1.500	0.500	1.500	1.0	True	WP
Res. fuel cell 1	5	0.030	0.030	0.033	1.0	True	Res. fuel cell
CHP diesel 1	9	0.003	0.003	0.003	1.0	True	CHP diesel
Fuel cell 1	9	0.212	0.212	0.212	1.0	True	Fuel cell
Res. fuel cell 2	10	0.014	0.014	0.014	1.0	True	Res. fuel cell

Table A.2: Power sources of the CIGRE MV network and their attributes.

name	from	to	len[km]	R[Ω/km]	X[Ω/km]	I max[kA]	parallel	type
Line 1-2	1	2	2.82	0.501	0.716	0.145	1	cs
Line 2-3	2	3	4.42	0.501	0.716	0.145	1	cs
Line 3-4	3	4	0.61	0.501	0.716	0.145	1	cs
Line 4-5	4	5	0.56	0.501	0.716	0.145	1	cs
Line 5-6	5	6	1.54	0.501	0.716	0.145	1	cs
Line 7-8	7	8	1.67	0.501	0.716	0.145	1	cs
Line 8-9	8	9	0.32	0.501	0.716	0.145	1	cs
Line 9-10	9	10	0.77	0.501	0.716	0.145	1	cs
Line 10-11	10	11	0.33	0.501	0.716	0.145	1	cs
Line 3-8	3	8	1.30	0.501	0.716	0.145	1	cs
Line 12-13	12	13	4.89	0.510	0.366	0.195	1	ol
Line 13-14	13	14	2.99	0.510	0.366	0.195	1	ol
Line 6-7	6	7	0.24	0.501	0.716	0.145	1	cs
Line 11-4	11	4	0.49	0.501	0.716	0.145	1	cs
Line 14-8	14	8	2.00	0.510	0.366	0.195	1	ol

Table A.3: Line data of the CIGRE MV network.

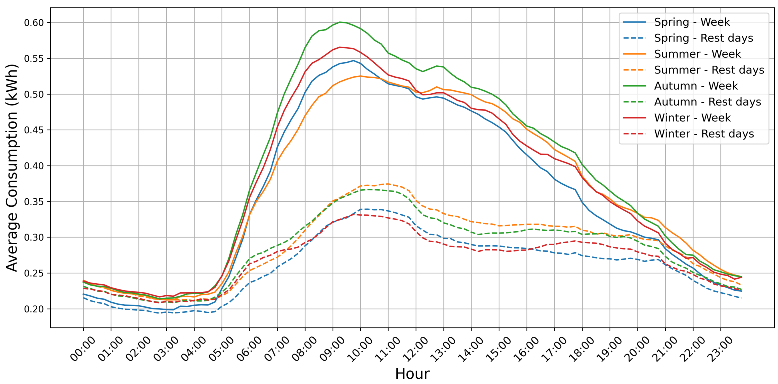
name	V_n [kV]	type	zone	in_service	geo [x, y]
Bus 0	110.0	b	CIGRE_MV	True	[7.0, 16.0]
Bus 1	20.0	b	CIGRE_MV	True	[4.0, 15.0]
Bus 2	20.0	b	CIGRE_MV	True	[4.0, 13.0]
Bus 3	20.0	b	CIGRE_MV	True	[4.0, 11.0]
Bus 4	20.0	b	CIGRE_MV	True	[2.5, 9.0]
Bus 5	20.0	b	CIGRE_MV	True	[1.0, 7.0]
Bus 6	20.0	b	CIGRE_MV	True	[1.0, 3.0]
Bus 7	20.0	b	CIGRE_MV	True	[8.0, 3.0]
Bus 8	20.0	b	CIGRE_MV	True	[8.0, 5.0]
Bus 9	20.0	b	CIGRE_MV	True	[6.0, 5.0]
Bus 10	20.0	b	CIGRE_MV	True	[4.0, 5.0]
Bus 11	20.0	b	CIGRE_MV	True	[4.0, 7.0]
Bus 12	20.0	b	CIGRE_MV	True	[10.0, 15.0]
Bus 13	20.0	b	CIGRE_MV	True	[10.0, 11.0]
Bus 14	20.0	b	CIGRE_MV	True	[10.0, 5.0]

Table A.4: Bus data of the CIGRE MV network.

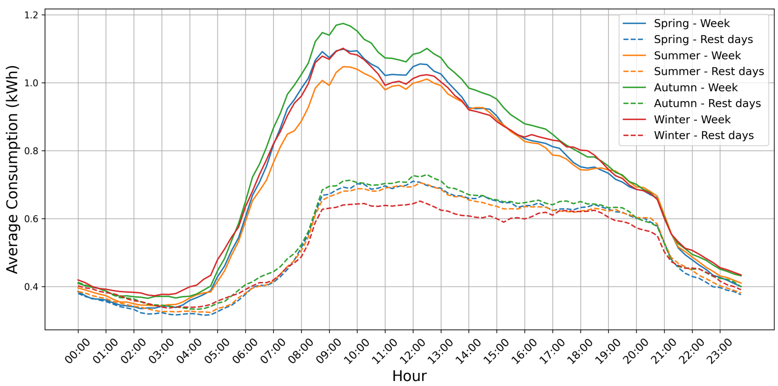
B | Appendix B

This appendix includes additional plots referenced in the main chapters. They provide visual support for different aspects of the analysis.

B.1 Consumption comparison across types of days.



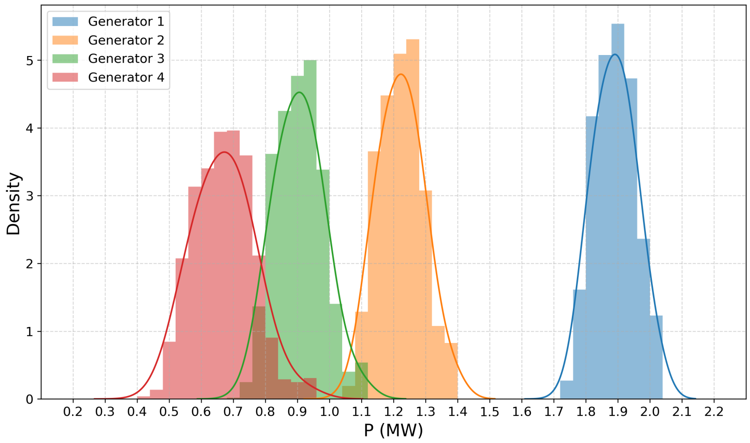
(a) Medium-consumption patterns



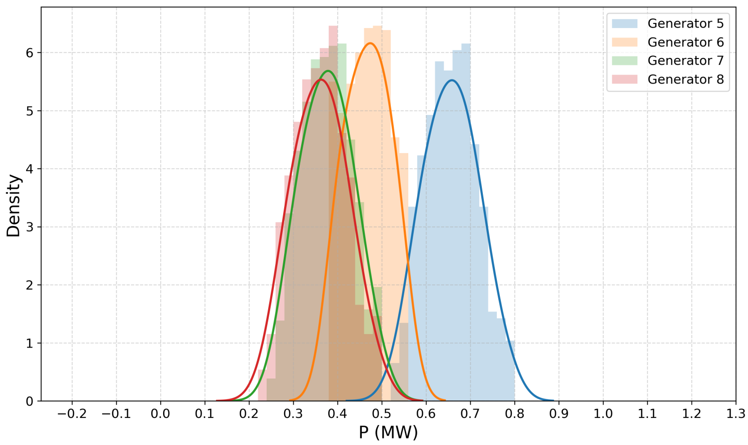
(b) High-consumption patterns

Figure B.1: Comparison of medium and high consumption patterns across type of days

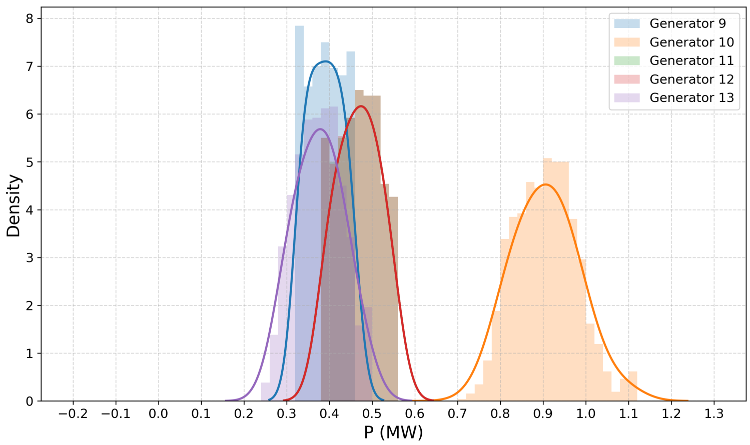
B.2 Power output of all generators (6–9 a.m.)



(a) Generators 1–4



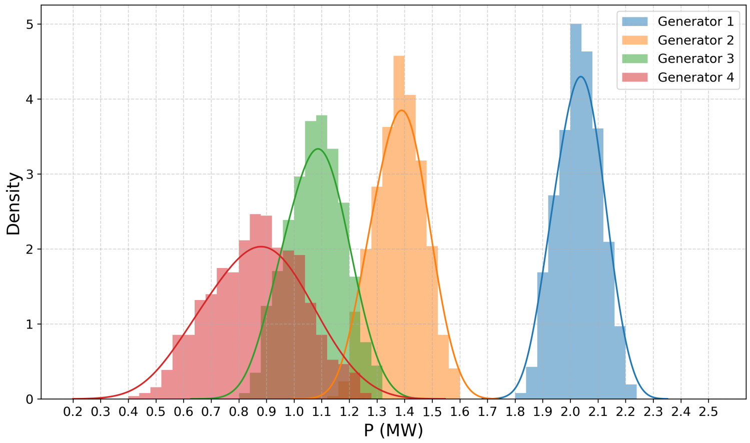
(b) Generators 5-8



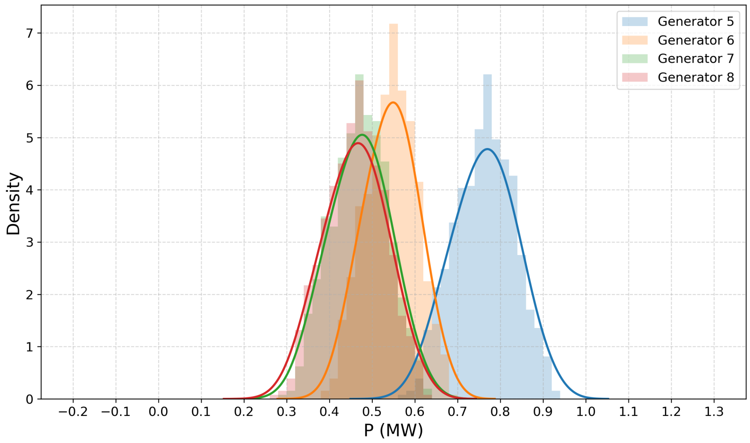
(c) Generators 9–13

Figure B.2: OPF morning results for generators grouped by indices: 1–4, 5–8, and 9–13.

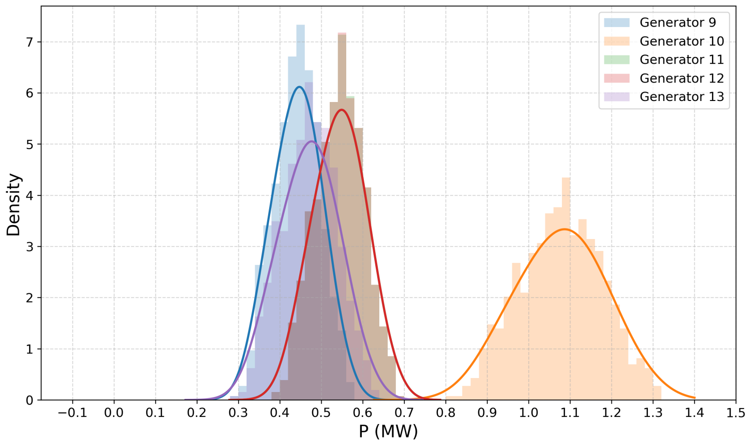
B.3 Power output of all generators (12–3 p.m.)



(a) Generators 1–4



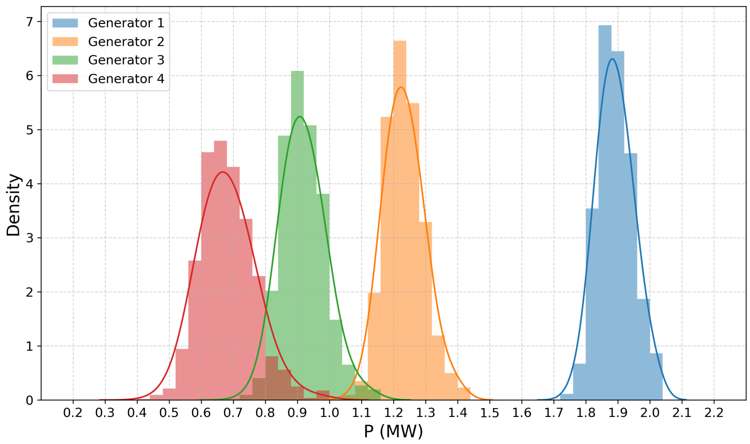
(b) Generators 5-8



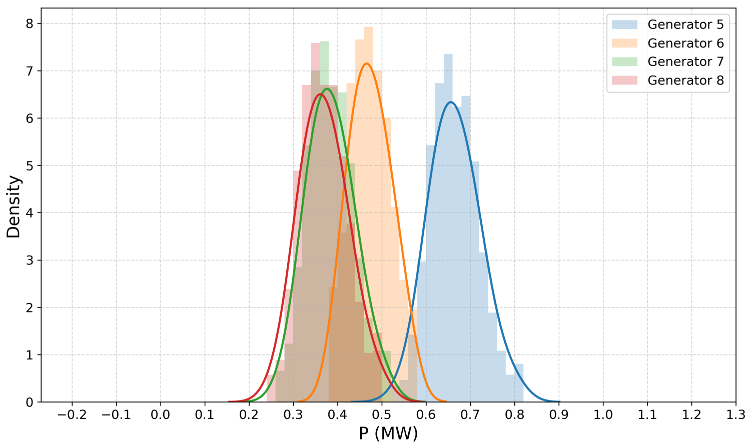
(c) Generators 9–13

Figure B.3: OPF afternoon results for generators grouped by indices: 1-4, 5-8, and 9-13.

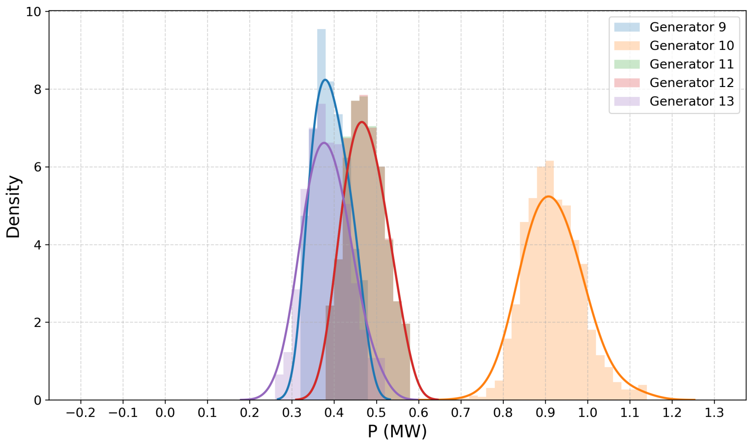
B.4 Power output of all generators (7-10 p.m.)



(a) Generators 1–4



(b) Generators 5–8



(c) Generators 9–13

Figure B.4: OPF evening results for generators grouped by indices: 1-4, 5-8, and 9-13.

List of Figures

List of Tables