University of Liège

Doctoral Thesis

Hosting Capacity of Low-Voltage Distribution Networks

Author:
Amina Benzerga

Advisor:
Prof. Damien Ernst

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor in Philosophy in Applied Sciences

Montefiore Institute
Department of Electrical Engineering and Computer Science
October 2024

Happiness can be found, even in the darkest of times, if one only remembers to turn on the light.

J.K. Rowling

Abstract

The power system, one of humanity's most significant engineering achievements, is also among the most complex. With the rapid rise of distributed energy resources (DERs) in distribution systems, networks are increasingly confronted with new challenges, including voltage regulation, power quality, and congestion. To mitigate these issues, it is essential to assess the current state of power networks to implement cost-effective solutions that ensure the reliability and efficiency of power systems. For a given network, its potential to accommodate DERs is referred to as the hosting capacity. More precisely, the hosting capacity is defined as the maximum amount of DERs that can be integrated into a network before operational problems arise. This manuscript presents five research contributions related to the hosting capacity of distribution networks.

The first contribution of this thesis is the development of a network topology identification method, which is the foundation for accurately evaluating HC. Given the evolving nature of distribution networks, understanding the network's structure is crucial for subsequent analysis and planning. The obtained network topology can be used for different simulations and HC studies.

While in the literature the concept of hosting capacity is well-known, it has no formal definition. The second contribution addresses this gap by providing a unified definition of the hosting capacity problem. The presented definition offers a structured framework for HC assessment, enabling consistent evaluations across different contexts. In addition, the paper presents how works from a systematic review related to hosting capacity can be cast in the framework.

The next three contributions present methods to evaluate the hosting capacity on different applications. The third contribution builds on the network topology identification. An individual, i.e., which evaluates only one DER type at a time, HC calculation method is proposed. The case study on which this method is performed is the reconstructed Belgian-inspired network. The aim of this work is to evaluate the distribution of

scenarios over different key performance indices to assess the HC.

The fourth contribution proposes a method that highlights the impact of selecting the phase to connect new DERs, which potentially limits the imbalance, on the HC. The method is evaluated on a Belgian-inspired network and with photovoltaic panels (PVs) then electric vehicles (EVs) as DERs.

Finally, the last contribution of thesis extends the HC framework to a combined application, where PV, EV, and heat pumps are integrated, demonstrating the versatility and comprehensiveness of the unified HC formalism. This combined analysis provides a broad perspective on how different DER technologies interact and influence overall network capacity, offering insights for future power system planning and DER integration strategies.

Contents

List of Abbreviations

ANM	Active Network Management
BESS	Battery Energy Storage Systems
μCHP	Micro Combined Heat and Power unit
DER	Distributed Energy Resource
DG	Distributed Generation
DHC	Dynamic Hosting Capacity
DN	Distribution Network
DOE	Dynamic Operating Envelope
DS	Distribution System
DSO	Distribution System Operator
EHV	Extra High voltage
EV	Electric Vehicle
GO	Global Optimisation
HC	Hosting Capacity
HP	Heat Pump
HV	High Voltage
IEA	International Energy Agency
IQR	Interquartil Range
IS	Individual Selection
KPI	Key Performance Index
LCT	Low Carbon Technology
LV	Low Voltage
MV	Medium Voltage
OpenDSS	Open Distribution System Simulator
OV	Over Voltage
PCC	Pearson Correlation Coefficient
PDF	Probability Density Function
PF	Power Flow
PS	Power System
PV	Photovoltaic
RES	Renewable Energy Source

RMSE	Root Mean Square Error
SM	Smart Meter
THD	Total Harmonic Distortion
TPI	Topological Path Identification
TRC	Transformer Rated Capacity
TSO	Transmission System Operator
THD_v	Total Voltage Harmonic Distortion
UV	Under Voltage
VD	Voltage Deviation
VU	Voltage Umbalance
WP	Wind Profiles
WT	Wind Turbine

Chapter 1

Introduction

The more you know about the past, the better you are prepared for the future.

Theodore Roosevelt

1.1 Motivations

Society has become accustomed to the convenience provided by reliable electricity. National economies depend heavily on electricity for industrial operations, communication, infrastructure, and overall societal functioning. Reliable and consistent access to electricity has become so integrated that its availability is often taken for granted. Power outage can lead to significant social and economic impacts. In some cases, these outages cause billion euros worth of damages [1].

Providing reliable electricity is the main goal of power systems (PS). Powers systems are the interconnected networks responsible for generating, transmitting, and distributing electricity. In Belgium, which serves as the reference system of this thesis, but also worldwide, the electricity generation is, to this day, highly reliant on traditional sources such as fossil fuels. In recent decades, there has been an increasing awareness of the need to reduce this dependence. This shift is mainly driven by three factors.

First, Belgium has almost non-existent domestic reserves of fossil fuels which causes it to be heavily reliant on imports from foreign countries [2]. This creates a precarious situation of geopolitical vulnerability. The inherent drawback of this dependency is the impact on the national economy leaving it exposed to spikes in international fuel prices. A recent example

of the impact of the dependence on foreign countries is the war between Ukraine and Russia, which resulted in a spike in energy prices across Europe[3].

Second, fossil fuels are finite resources, and their depletion over time presents a long-term challenge. Diversifying its energy mix and integrating renewable sources such as wind, solar, biomass, and geothermal energy will lead the Belgian energy system to be more resilient and sustainable.

Finally, the environmental impact of fossil fuel consumption has come under growing scrutiny. The combustion of fossil fuels is a major contributor to greenhouse gas emissions, especially carbon dioxide (CO_2), which accelerates global warming and exacerbates climate change. Reducing reliance on these fuels is not only a national priority but also a critical component of Belgium's commitments to international climate agreements, such as the Paris Agreement [4].

In order to reach these objectives of reducing fossil fuel usage, citizens and companies, sometimes with incentives from governments, change their electrical behaviour. This shift drastically impacts the way power systems are used, challenging their reliability. Power systems, and particularly distribution networks which are sub-parts of power systems, were not designed for a massive penetration of decentralised energy resources (DERs). In Belgium, the penetration of renewable generation has nearly quadrupled from 2010 as can be seen in Fig. 1.1 [5]. Similarly, the penetration of other distributed energy resources such as electric vehicles (EVs) and heat pumps in power systems have also increased [6], [7]. This adoption of DERs is further exacerbated by the drastic decrease of DER prices in the past decade. For instance, the price of PV panels has decreased by 80% in the last decade [8].

To ensure reliability of the network in regards to the threat posed by this massive integration of DERs, their integration might need to be limited. To prevent arbitrary and counter-productive limitations on DER integration, identifying the bottlenecks in the distribution network is crucial. Bottlenecks of given networks are identified by assessing their capacity to host new DER technologies while maintaining their reliability. This thesis focuses on the hosting capacity of distribution networks and presents work related to it.

1.2 Context

To provide context for the work presented in this thesis, the remainder of this chapter explains the traditional operation of power systems and discusses the challenges associated with the evolution of its exploitation.

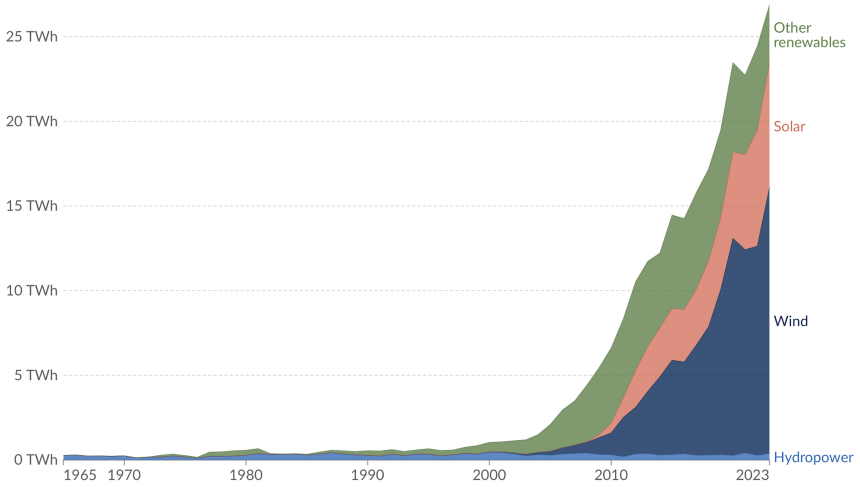


Figure 1.1: Renewable electricity generation in Belgium from 1965 to 2023 (image from [5]). Note that 'Other renewables' refers to renewable sources including geothermal, biomass, waste, wave and tidal.

1.2.1 Power systems

The power system, also referred as the power grid, has for sole objective to provide electricity to end-users. Traditionally, electricity was generated in centralised plants, e.g. nuclear plants, and then transported safely and with as little losses as possible to customers. This conventional power system is depicted in Fig. 1.2.

To minimise the losses, two levels of transportation grids are used: transmissions, transmitting currents at high-voltages (HV) and extra high voltages (EHV), and distribution, conveying currents at medium-voltages (MV) and low-voltages (LV). The voltage ranges per levels are given in Table 1.1. The two transportation grids are managed and operated by different entities: transmission system operators (TSOs) and Distribution system operators (DSOs).

Using different voltage levels reduces Joule losses, which occur as heat when a current I flows through lines with a resistance R. Joule losses are proportional to the square of the current I^2R. Transmitting power at high voltages and low currents minimises these losses. Indeed, electrical power is the product of voltage and current. Therefore, to maintain the same power level while reducing the current, the voltage must be increased. Using lower current significantly reduces the I^2R losses which improves transmission

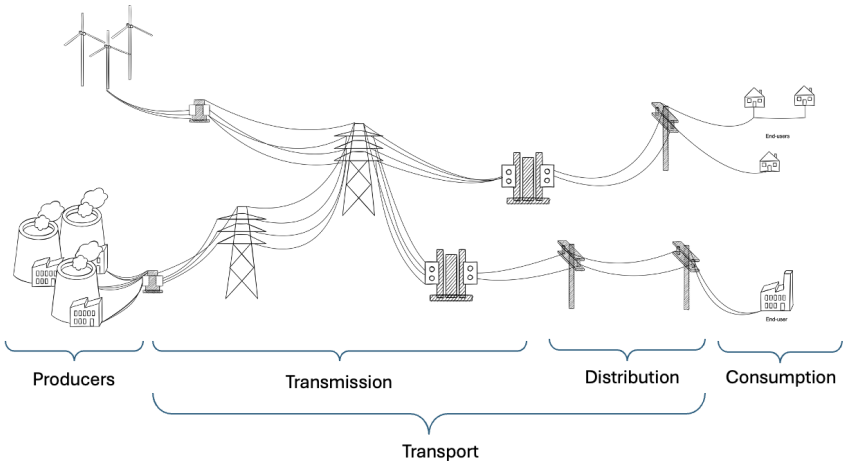


Figure 1.2: Illustration of the traditional organisation of the power system. This illustration is inspired from [9], [10].

Voltage	230V-400V	6kV-15kV	30kV-150kV	220kV - 380kV
Level	Low	Medium	High	Extra high
Operator	DSO		TSO	

Table 1.1: Belgian voltage ranges by levels and the corresponding operator. [11]–[13].

efficiency.

In addition to using different voltage levels, parts of the power systems operate using a three-phase system for both transmission and distribution. In three-phase systems, three separate electrical currents flow through power lines. This structure allows for more efficient and continuous power delivery, ensuring smooth operation of electrical equipment and reducing power losses. However, some distribution networks, particularly in residential areas, use single-phase systems, where only one current flows.

This thesis focuses for the main part on low-voltage distribution networks with residential customers and their DERs.

1.2.2 Distribution networks

Distribution networks (DNs) are typically radial, tree-like, and with simple structures. These networks are fragmented, with multiple DSOs managing

them. In Belgium, for instance, several DSOs operate within the country.

In the past, DN were considered simple because electricity flowed in one direction: from the transmission system to consumers. The simple structure of distribution network led to a lower priority on real-time monitoring and many distribution systems were built decades ago, before modern digital monitoring technologies became widespread. In some cases, DSOs may not have yet fully digitised or updated records of their distribution infrastructure, leaving parts of the network undocumented or poorly mapped. This limits the comprehensive understanding of the system.

Historically, investments in grid management have focused on transmission networks, where power flows are larger and failures have widespread consequences. As a result, distribution networks have received less investment in data acquisition systems, as DSOs have prioritised network expansion to meet the growing demand over installing advanced monitoring technologies. The "fit and forget" doctrine has traditionally governed distribution network management. This approach assumes that once the system is built, it will operate reliably with minimal intervention, aside from occasional maintenance. Infrastructure is often oversized to handle potential power peaks, assuming a reference demand and one-directional power flows. With this strategy, real-time monitoring and complex controls were seen as unnecessary, making distribution networks simpler and less costly to manage. However, with the decentralisation and widespread adoption of various types of DERs, challenges have emerged that necessitate better information and a clearer understanding of the network to improve management and ensure its reliability.

1.2.3 Distributed Energy Resources

Distributed energy resources are small-scale energy assets that generate, store or consume energy. DERs can be categorised in these groups, as shown in Fig. 1.3: generators, e.g. photovoltaic panels, and consumers, e.g. electric vehicles and heat pumps. The categories can be overlapping as for instance some EVs can be used to store energy while their initial intent is to consume energy. Note that decentralised generation (DGs) are not, by definition, restricted to renewable energy sources (RES), for instance, diesel generators are DG but not RES.

Distributed energy resources like PV panels and electric vehicles are rapidly changing the flow of electricity in distribution networks. When DERs were integrated on a small scale, the issues they caused in distribution networks were minimal. However, their widespread adoption has both introduced new challenges and amplified existing problems in the distribution network.

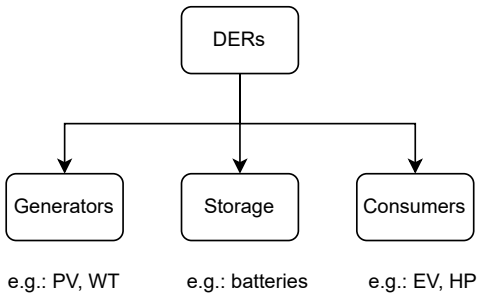


Figure 1.3: Terminology of distributed energy resources with most common examples at distribution level.

Furthermore, what were once simple, one-way networks have become more complex, with electricity sometimes flowing back from consumers into the grid. Indeed, some distributed energy resources, like PV systems, cause *reverse power flow* when their production exceeds local demand, particularly in areas with high DER penetration. This creates new challenges for DSOs, who must now manage bidirectional power flows while having nearly no real-time information about their own network topology.

The integration of DERs and changing consumer behaviours, like the widespread adoption of electric vehicles, has led to shifting load patterns. The "duck curve", originally seen in California [14], illustrates how high solar generation during midday combined with evening consumption peaks can strain the grid. These fluctuations in generation and demand can cause high voltage variations, with voltage rising or dropping outside of acceptable limits in different parts of the network. The Belgian duck curve is shown in Fig. 1.4.

Voltage fluctuations become more frequent with DERs, like solar PV installations, as they produce variable amounts of power based on external conditions, such as cloud cover [15]. When solar irradiance drops suddenly, PV output decreases rapidly, reducing the power injected into the grid. This quick ramping up and down challenges the grid's ability to balance supply and demand, making it harder to maintain stable voltage levels. The grid must constantly adjust to these fluctuations, which can lead to instability, particularly in areas with high DER penetration.

In addition, the growing adoption of EVs adds to this challenge by increasing power demand and causing congestion in distribution networks. Congestion refers to situations where the capacity of the grid to deliver

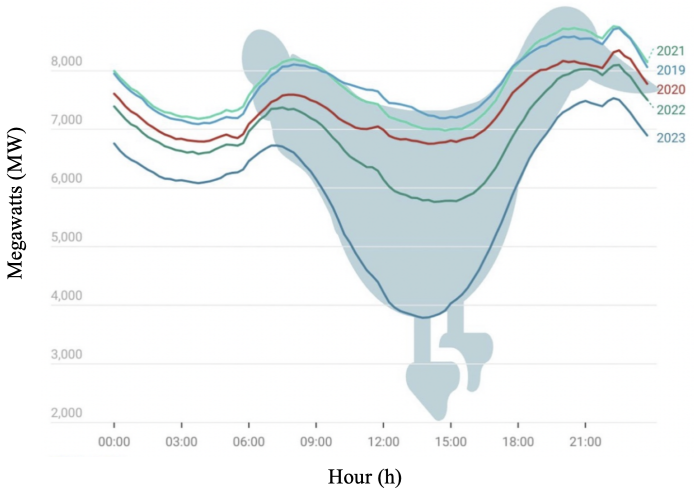


Figure 1.4: Impact of solar capacity in Belgium in June 2023 (from [16]). The pattern created by the midday dip in the net load curve, followed by a steep rise in the evenings when solar generation drops off, looks like the outline of a duck, so this pattern is often called a duck curve [14].

electricity is limited due to high demand, which can occur when many EVs charge simultaneously, especially in residential areas, leading to overloads and reduced reliability.

Finally, the imbalances in the three-phase distribution network is exacerbated by some DERs that connected to the grid using single-phase connections. This means that instead of distributing the load and DERs evenly across all three phases, they are sometimes concentrated on a single phase, causing uneven power distribution in the phases. This further complicates the task of maintaining grid stability and voltage control.

This list is not exhaustive but highlights some of the key issues that DERs bring to modern distribution networks. These problems limit the amount of DERs that can be hosted in the distribution network without making investments, a concept referred as the hosting capacity (HC) which is the focus of this thesis.

1.2.4 Hosting capacity

The hosting capacity refers to the maximum level at which the integration of new distributed energy resources can occur without causing operational issues. Determining this capacity is a complex challenge. They are at least

three main reasons that make it complex.

First, although the concept itself is understood intuitively by the research community, there is no unified definition.

Then, evaluating hosting capacity involves analysing the network, however, data on distribution networks is often scarce. Critical information, such as network topology or load data, is frequently missing. These data are difficult to replicate or forecast, especially given the diversity of customers, their various behaviours and the changing consumption behaviours driven by the increasing adoption of DERs disrupt traditional load profiles.

Lastly, the inherent uncertainties associated with DERs, both in terms of their variable generation and unpredictable installations, add another layer of complexity to capacity evaluations. As the focus of this thesis is the hosting capacity, these challenges are addressed in the following chapters.

1.3 Outline of the manuscript

This thesis contributes to address the challenges posed by the increasing integration of DERs into distribution networks. The manuscript is structured as follows.

Following the presentation of the publication in this chapter, Chapter 2 focuses on the identification of distribution network topology. This chapter introduces a method for obtaining a digital representation of a distribution network, even with only partial availability of smart meter data. Smart meters (SM) are digital devices that measure electricity in real-time and are used to monitor the grid. The presented method is applied to a Belgian inspired network with 40% of customers equipped with smart meters. The results provide a foundation for accurately computing hosting capacity on reconstructed real-world networks.

Chapter 3 focuses on establishing a unified framework for the concept of hosting capacity. This chapter provides a common foundation for addressing hosting capacity challenges and presents two examples where previous HC problems from the literature are re-framed using the presented formalism. Additionally, it offers a comparison of key aspects from relevant studies, highlighting the advantages of creating a common ground for discussions on hosting capacity.

The next three chapters, Chapters 4, 5 and 6, demonstrate practical applications of hosting capacity computation. They proposes methods to compute the HC on different scenarios and settings. Chapter 4 introduces a method for evaluating the hosting capacity of individual DERs, with the approach being applied specifically to two DER types: PVs and EVs. The method provides multiple probabilistic performance indicators that reflect

various potential configurations. The potential configurations result from new DERs randomly added to the network. The network used for the case study is the one reconstructed from the case study of Chapter 2.

Chapter 5 explores the influence of optimal phase connection of DERs on constraints affecting hosting capacity. This approach provides valuable insights into how small and cost-effective network modifications can improve the HC. The chapter operates under the assumption that all DERs, such as photovoltaic systems, can be curtailed when the network encounters issues. This assumption provides a clear cost estimate associated with optimising phase connections, giving DSOs valuable insights into budget considerations for implementing this solution. The presented approach is also applied on the same two DERs individually as Chapter 4, i.e., PVs and then EVs, and the case study is a Belgian inspired network. Note that the approaches in Chapters 4 and 5 were developed prior to the framework presented in Chapter 3 and therefore do not use that framework.

Chapter 6 presents a method for assessing the hosting capacity of combined DERs, focusing on the integration of PVs, EVs, and HPs. This method builds upon the formalism defined in Chapter 3 and is applied on the reconstruction of a real Belgian network for which 65% of customers have smart meters.

Finally, the last chapter of this manuscript, Chapter 7, is dedicated to the conclusion of the presented work. An overall conclusion is presented in Section 7.1 while Section 7.2 suggests future prospects for research in the field.

1.4 List of publications

The scientific publications related to this thesis are:

- A. Benzerga, D. Maruli, A. Sutera, S. Mathieu, and D. Ernst. Low-voltage network topology and impedance identification using smart meter measurements. In *Proceedings of the 2021 IEEE Madrid PowerTech*, 2021 [17].
  Chapter 2 is based on this publication.

- D. Maruli, S. Mathieu, A. Benzerga, A. Sutera, and D. Ernst. Reconstruction of low-voltage networks with limited observability. In *IEEE PES Innovative Smart Grid Technologies Conference Europe*, 2021 [18].
  Chapter 2 is based on this publication.

- A. Benzerga, A. Bahmanyar, S. Mathieu, and D. Ernst. Smart Meter Data Analytics Case Study: Identification of LV Distribution Network

Topology to Design Optimal Planning Solutions. In *Applications of Big Data and Artificial Intelligence in Smart Energy Systems* (pp. 161 - 192). IEEE River Publishers, 2023 [19].
Chapters 2 and 4 are based on this publication.

- A. Benzerga, A. Bahmanyar, G. Derval, and D. Ernst. A unified definition of hosting capacity, applications and review. 2024 [20].
  Chapter 3 is based on this publication.

- A. Benzerga, S. Mathieu, A. Bahmanyar, and D. Ernst. Probabilistic capacity assessment for three-phase low-voltage distribution networks. In *Proceeding of the IEEE 15th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*. IEEE, 2021 [21].
  Chapter 4 is based on this publication.

- A. Benzerga, M. Vassallo, S. Gérard, J. Vandeburie, and D. Ernst. Combined PV-EV-HP Hosting Capacity Analysis of a Belgian Low-Voltage Distribution Network. In *Proceeding of the 34th Australasian Universities Power Engineering Conference (AUPEC)*. 2024.
  Chapter 6 is based on this publication.

- A. Benzerga, S. Gérard, S. Lachi, Q. Garnier, A. Bahmanyar, and D. Ernst. Optimal connection phase selection for single-phase electrical vehicle chargers. In *Proceeding of the 2022 CIRED workshop on E-mobility and power distribution systems (CIRED)*. 2022 [22].
  Chapter 5 is based on this publication.

- A. Benzerga, A. Bahmanyar, and D. Ernst. Optimal Connection Phase Selection of Residential Distributed Energy Resources and its Impact on Aggregated Demand. In *Proceeding of the 11th Bulk Power Systems Dynamics and Control Symposium (IREP)*. 2022 [23].
  Chapter 5 is based on this publication.

During this thesis, other collaborations have led to publications that are not discussed within this manuscript:

- M. Vassallo, A. Benzerga, A. Bahmanyar, and D. Ernst. Fair Reinforcement Learning Algorithm for PV Active Control in LV Distribution Networks. In *Proceeding of ICCEP 2023 conference*. 2023 [24].

- V. Dachet, A. Benzerga, R. Fonteneau, and D. Ernst. Towards CO2 valorisation in a multi remote renewable energy hub framework. In *Proceedings of ECOS 2023 - The 36th International Conference on*

Efficiency, Cost, Optimisation, Simulation and Environmental Impact of Energy Systems. 2023 [25].

- V. Dachet, A. Benzerga, D. Coppitters, F. Contino, R. Fonteneau, and D. Ernst. Towards CO2 valorization in a multi remote renewable energy hub framework with uncertainty quantification. In *Journal of Environmental Management, 363, 121262*. 2024 [26].

- J. Mbenoun, A. Benzerga, B. Miftari, G. Detienne, T. Deschuyteneer, J. Vazquez, G. Derval, and D. Ernst. Integration of offshore energy into national energy system: a case study on Belgium. 2024 [27].

Chapter 2

Network topology identification

An investment in knowledge pays the best interest.

Benjamin Franklin

This chapter is built upon studies carried out in cooperation with a Belgian DSO in response to the challenges it faces for LV network planning and control. It presents a method that is designed to form and update the LV network topology, and to estimate its branch impedances, by relying only on the recorded smart meter data.

Distribution system operators have been upgrading their network over several decades, though not always keeping digital records of all changes. As a result, the operators do not always known exactly how their customers are connected to the network. Some of these customers are equipped with smart meters, providing voltage and current time-series. These measurements can be used to identify the network topology and the line impedances. This chapter presents a method to identify radially operated low-voltage networks which can be applied with limited number of smart-meters. The resulting identified model provides the map of the network and impedances of the inferred lines, allowing to perform subsequent analysis (e.g. power flow). Simulation results on a case study with 128 nodes show an average error 0.69% in computed voltages, while only 40% of the nodes are equipped with smart meters.

2.1 Notations

This section defines the key sets and variables used in this chapter.

Sets

$\mathcal{T}$	Set of observation period
$\mathcal{N}$	Set of nodes
$\mathcal{O}$	Set of nodes observed with smart meters
$\mathcal{H}$	Set of nodes not observed
$\mathcal{E}$	Set of edges
$\mathcal{P}$	Set of phases
$\mathcal{A}_n$	Set of new possible installations type for node n
$\mathcal{M}_{k,n}$	Set of types of installations that are connected to node n in configuration k

Variables

ψ_n	Geographical coordinates $n \in \mathcal{N}$
$\mathbf{Z}_e$	Impedance of edge $e \in \mathcal{E}$
$\mathbf{S}^b_{n,p}$	Initial base power injection of $n \in \mathcal{N}$ in $p \in \mathcal{P}$
$\mathbf{S}_n$	Power injection of $n \in \mathcal{N}$
$\mathbf{S}_{k,n,m}$	Power injection of installations of type $m \in \mathcal{M}$ at $n \in \mathcal{N}$ for configuration k
$\mathbf{I}$	Current
$\mathbf{V}$	Voltage
$\overline{V}$	Over-voltage threshold
$\underline{V}$	Under-voltage threshold

2.2 Introduction

With the increasing use of distributed energy resources and electric vehicles, distribution system operators (DSOs) are encountering serious difficulties in guaranteeing the safety of their Low-Voltage (LV) network in the years to come. Being able to effectively integrate distributed energy resources and electric vehicles is a fundamental step to accelerate the energy transition process. At LV levels, distribution networks are mostly operated radially, and power is distributed through several feeders, i.e. main electrical lines carrying power from the substation to the customer. Most residential loads are connected to the feeder through a single phase and a neutral wire. The phase to which a load is connected may be selected arbitrarily. Power imbalances between phases are expected, leading to a reduced hosting capacity of the system. In order to implement effective preventive or corrective mea-

sures against voltage or congestion issues, DSOs need to be able to assess the system's response to various realistic scenarios. This analysis is usually performed through power-flow studies, but reliable solutions require accurate information about topology of the network and physical characteristics of the lines. DSOs do not always know how households, feeders and other appliances are interconnected. Furthermore, LV networks topology can change over time because of faults, maintenance or reconfiguration, and existing databases can contain outdated or inaccurate information. DSOs can therefore lack possession of a reliable model of their LV network, which can hinder the efficient management and development of their system. Network identification is the mathematical process that allows to deduce this information. Effective network identification methods for LV networks are, thus, essential for the development of smarter grids [28]. This work presents a methodology to retrieve the topology and the cable parameters of an LV network from time-series measurements provided by a limited number of smart meters in the grid.

This chapter is structured as follows. Section 2.3 reviews relevant literature. Then, the distribution network model is introduced in Section 2.4. Section 2.5 states the problem, defining inputs and outputs. Section 2.6 describes the methodology proposed to address the problem. Section 2.7 examines a specific case study to evaluate the potential of such methodology and discuss the results. Finally, Section 2.8 summarises the conclusions and introduces potential future work related to this chapter.

2.3 Literature review

In transmission systems, topology and cable parameter information is usually stored in appropriate databases [28]. Unfortunately, the same does not apply to LV distribution networks. With the increasing importance and complexity of distribution systems, the identification problem of LV networks has gained more attention and it has been recently tackled adopting different approaches. Recent literature, such as [28]–[30], focuses on identifying the topology of the network when limited information is available. Algorithms that aim, as this work does, to identify both network topology and lines parameters at the same time, are presented in [31]–[34]. Authors of [31] use the evaluation of voltage sensitivities with respect to active and reactive power injections and Prũfer sequences to identify the topology of small networks, assuming that only specific cables types and lengths are used for the lines. The identification problem in [32] takes the name of inverse power flow problem, where the system admittance matrix is found by solving an unconstrained least-squares problem. The case with non-

measured nodes in the grid, also referred to as *hidden nodes*, is also tackled, both for meshed and also radial topology, with the assumption that these hidden nodes have zero net current injections. The inverse power-flow problem is extended to poly-phase systems in [33], with the full-observability assumption. Finally, an algorithm to jointly estimate both admittance and topology, assuming that the measurements for all non-zero power injecting nodes are available, is presented in [34]. A summary comparison table can be found in Table 2.1. All the methods that have been examined share the assumption that every node in the network, or at least the power-injecting ones, has a meter attached to it. This chapter presents a methodology to address the network identification problem even when some power-injecting nodes in the grid are not metered.

Table 2.1: Literature review where ○ indicates that the article does not provide feature, ◐ partially provides feature, ● provides feature.

Features | papers	[28]	[29]	[30]	[31]	[32]	[33]	[34]	[∗]
Line parameters estimation	○	○	○	◐	●	●	●	●
Hidden non-injecting nodes	○	○	○	●	●	○	●	●
Hidden injecting nodes	○	○	○	○	○	○	○	●
Required measurements	V	V	E	V, S	V, I	V, I	V, S	V,I
No assumptions on cables	◐	●	●	○	●	●	●	◐
Unbalanced poly-phase	○	○	○	○	○	●	○	●

∗ Proposed methodology
V = Voltage; I = Current; S = Power; E = Energy

2.4 Distribution network model

Let us consider an unbalanced three-phase, four-wire LV radial distribution network such as in Figure 2.1. The network radial topology can be represented by a tree graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is the set of nodes in the network and $\mathcal{E}$ is the set of edges linking these nodes. For each node $n \in \mathcal{N}$, let ψ_n be the two-tuple geographical coordinates of the node. For each edge $e \in \mathcal{E}$, let $\mathbf{Z_e} \in \mathbb{C}^3$ be the three-phase impedance matrix of the edge. **Bold** face variables designate complex values or phasors throughout the chapter. The three phases of the network are denoted by indices a, b and c and the set of phases by $\mathcal{P} = \{a, b, c\}$.

The network is studied over a finite time period discretised in T consecutive time steps with an interval of Δt. The set of time steps in the period is $\mathcal{T}$. The value of a variable x at time-step $t \in \mathcal{T}$ is denoted by x_t. The initial base power injection time series of phase $p \in \mathcal{P}$ of node $n \in \mathcal{N}$ over the study period is denoted by $\mathbf{S}^b_{n,p} \in \mathbb{C}^{|\mathcal{T}|}$. The magnitudes of the voltage and the current injection in phase $p \in \mathcal{P}$ of node n at time step t are denoted by $V_{n,p,t}$ and $I_{n,p,t}$, respectively.

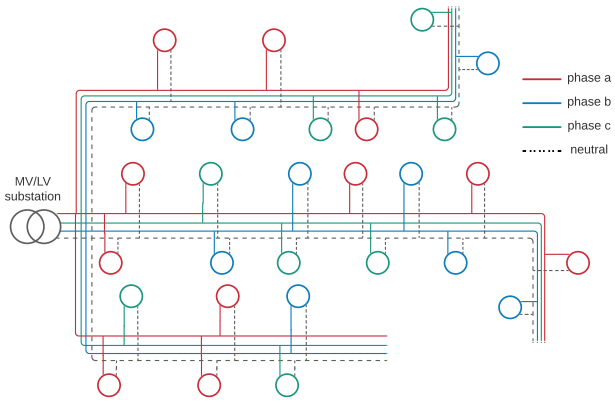


Figure 2.1: A three-phase, four-wire radial distribution network with three feeders.

A subset $\mathcal{O} \subseteq \mathcal{N}$ of nodes are observed with smart meters. Information about ψ_o, $V_{o,p}$, $I_{o,p}$, $\forall\ o \in \mathcal{O}$ are considered as the inputs for the analysis. It is assumed that the smart meters are able to provide the magnitude of the measured variables and not the phasors. It is also assumed that all the customers are connected to the network neutral conductor, however the proposed methods are applicable to other earthing schemes.

2.5 Problem Statement

The topology and parameters of LV electricity distribution networks are, to a large extent, currently unknown. Smart meter data offer a unique opportunity to identify the model of LV networks, but the identification involves several challenges:

- An LV network involves single-phase customers unevenly distributed between three phases.
- The customers' load varies stochastically over time.
- Smart meters are not installed at every customer connection node.

At the LV level, the only reliable sources of information are the smart meters measurements and their geographical location. This section will present a method developed to identify the model of a low-voltage network by analysing smart meters' recorded voltages and powers, and the geographical location of the metered nodes, without relying on any additional information. However, the meters are usually not installed at all the nodes. The network identification problem infers a network model describing:

- Its topology including the nodes (metered or non-metered), their geographical locations and edges between;
- Edges' three-phase impedance;
- Current injection time series of the non-metered nodes

The following assumptions are made:

- During the observation period, the network topology does not change;
- Customer connections to the main feeder are of single-phase two-wire type;
- Customers connection phase is known;
- There is at least one meter connected at every phase of every feeder;
- three-phase measurements of MV/LV substation are available;
- The X/R ratio of cables (γ) is known.

The topology identification has two main steps. In the first step, the goal is to construct three single-phase estimates of the network topology. Then, the algorithm evaluates the identified line-section impedances. Using the result of the first step, the second step of the algorithm merges the three single-phase estimated typologies to form a three-phase feeder model [17].

2.6 Methodology: LV network identification using smart meter data

2.6.1 Single-phase topology identification algorithm

As shown in Figure 2.2, this step considers the three phases of the network graph $\mathcal{G}$ independently. Phase $p \in \mathcal{P}$ of the network can be modelled as a rooted tree $\mathcal{G}_p = (\mathcal{N}_p, \mathcal{E}_p)$[1]. $\mathcal{N}_p$ represents the subset of nodes connected to phase p of the substation, while the set of edges $\mathcal{E}_p$ represents the phase p of the lines connecting those nodes. The root node of $\mathcal{G}_p$ represents phase p of the MV/LV substation.

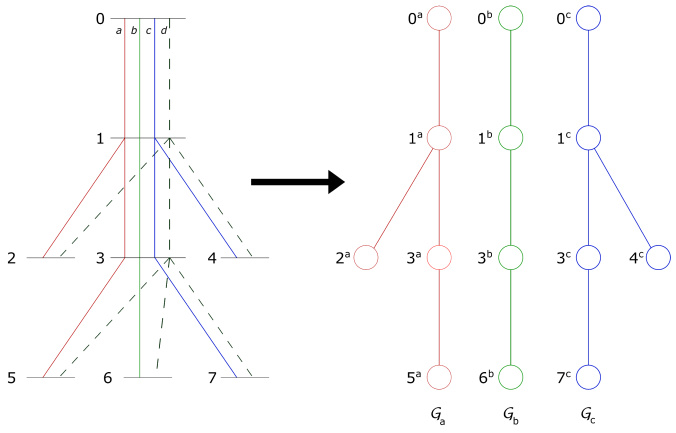


Figure 2.2: Graphical representation of $\mathcal{G}_a$, $\mathcal{G}_b$, $\mathcal{G}_c$ for a simple 8-node network.

The goal of single-phase topology identification is to make an estimation of these three rooted trees $\hat{\mathcal{G}}_p = (\hat{\mathcal{N}}_p, \hat{\mathcal{E}}_p)$ [2]. The algorithm has three main steps: topology estimation, topology validation, and hidden node detection. For each phase $p \in \mathcal{P}$, the steps should be carried out independently.

Topology Estimation

The initial estimation of the set of the network nodes $\hat{\mathcal{N}}_p$ corresponds to the set of observed ones $\mathcal{O}_p$. The first step of the algorithm estimates a rooted tree $\hat{\mathcal{G}}_p$ for each phase $p \in \mathcal{P}$ connecting these nodes. This operation is carried out using both node geographical information to draw the

[1] A rooted tree is a tree in which one node has been designated as the root.
[2] The hat symbol ˆ atop of a variable denotes that its value is estimated.

network map, and correlation analysis on voltage measurements to infer connectivity between the nodes. As proposed in [28], [35], nodes connected together have similar voltage patterns, i.e. high correlation between their measured voltages. However, the patterns of load and production can have a similar profile for different nodes of the LV network, as a result of similar weather conditions and residential occupancy profiles. To avoid this affecting the results of the voltage correlation analysis, as proposed in [28], a pre-processing of the voltage time series is performed by applying a high-pass filter. Pearson's Correlation Coefficient (PCC) is then used to reveal the relationship between the measured voltages.

Let us denote the weighted graph of $\hat{\mathcal{N}}_p$ by $\hat{\mathcal{G}}_{w,p}$, where the edge (i,j) weight, w_{ij}, is equal to the PCC between the filtered voltage time series of nodes i and j. The first estimate of network topology $\hat{\mathcal{G}}_p = (\hat{\mathcal{N}}_p, \hat{\mathcal{E}}_p)$ can be obtained forming the maximum spanning tree[3] on $\hat{\mathcal{G}}_{w,p}$.

Topology Validation

The purpose of the topology validation is to confirm the topology obtained in the first step. This step accepts or rejects the edges by investigating the potential existence of missing nodes or wrong connections in each edge $e = (i,j) \in \hat{\mathcal{E}}_p$. Let $\hat{Z}_{e,p} = \{\hat{Z}_{e,p,1}, \ldots, \hat{Z}_{e,p,T}\}$ denote the estimated impedance magnitude time series for each edge $e = (i,j) \in \hat{\mathcal{E}}_p$, and for all time steps $t \in \mathcal{T}$:

$$\hat{Z}_{e,p,t} = \frac{V_{j,p,t} - V_{i,p,t}}{\hat{I}_{e,p,t}} \tag{2.1}$$

where the estimated current $\hat{I}_{e,p,t}$ flowing in e is given by summing up the load currents of the downstream nodes of edge e.

Edge impedances are assumed constant during the considered period $\mathcal{T}$. Therefore, the edge impedances $\hat{Z}_{e,p}$ estimated for all time intervals should be close to each other. With this assumption, if, for each edge, the relative standard deviation of the impedance time series $\hat{Z}_{e,p,t}$ is less than a tolerance level λ, the edge can be considered valid. It should be noted that in practical conditions, the line impedance is somewhat variable depending on several factors, such as the ambient temperature. Therefore, an appropriate value for the tolerance level λ should be selected to avoid the rejection of correctly identified edges.

[3] A maximum spanning tree is a subtree of a weighted graph which connects all vertices (i.e. a spanning tree) and has the maximum weight among all possible spanning trees

Hidden Node Detection

Let $\mathcal{H} = \mathcal{N} \backslash \mathcal{O}$ be the set of non-metered nodes in the network. These nodes are hidden from the viewpoint of the identification algorithm. Consider Figure 2.3, where there is a hidden node $H \in \mathcal{H}_p$ in the network, i.e., node 5. Since the current of the hidden node $I_{H,p}$ is not measured, it is not considered in the topology validation step. This may result in the rejection of the edges in the path from the hidden node to the substation, as shown in Figure 2.3, due to the inaccurate estimation of the downstream current and consequently the inaccurate estimate of the edge impedance time series. Moreover, whenever a hidden node is between two other nodes, as node 5 is, the topology estimation step identifies an incorrect edge connecting the upstream and downstream nodes, like the edge between node 3 and node 6 in the figure.

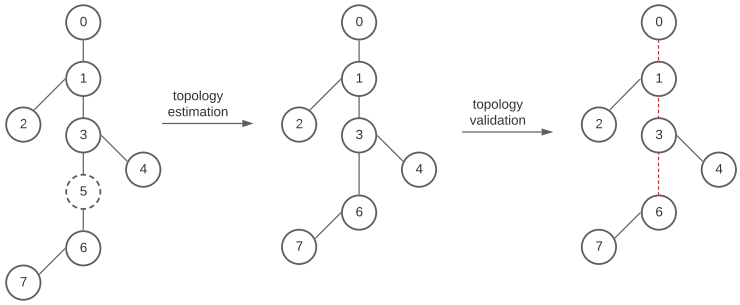


Figure 2.3: Steps for the identification of a network topology with node 5 as a hidden node. Dashed edges are not validated.

Consider the farthest node from the root among all the nodes of the rejected edges (node 6 in Figure 2.3), let us denote this node by $A \in \hat{\mathcal{N}}_p$, and its parent and grandparent nodes by B and C, respectively. As shown in Figure 2.4, there are three topological possibilities to place a hidden node H adjacent to A. As shown, we refer to these topologies by terms *Bridge*, *Leaf*, and *Common parent*.

To make an estimation of a hidden node $X \in \mathcal{H}_p$, the hidden node detection algorithm evaluates these possible topologies to check which one suits the best.

The current $\hat{I}_{BC,p}$ flowing through BC has two portions. $\hat{I}_{A^+,p} = \hat{I}_{AB,p}$ which is the contribution of A and its downstream nodes, and $\hat{I}_{B^+,p}$ which

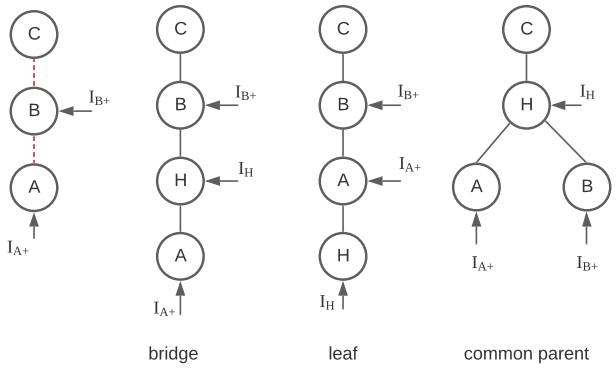


Figure 2.4: Possible locations for a hidden node H adjacent to A.

is the current injected at node B. We have:

$$\hat{I}_{B^+,p} = \hat{I}_{AB,p} - \hat{I}_{BC,p} \tag{2.2}$$

To evaluate the three configurations for the estimation of a hidden node X location, the algorithm solves the following three minimisation problems. These minimisation problems quantify, for each configuration, the level of basic circuit laws violation.

For the *Bridge* configuration:

$$\delta^b = \min \sum_{t \in \mathcal{T}} |V_{B,p,t} - V_{C,p,t} - \hat{Z}_{CB,p}(\hat{I}_{A^+,p,t} + \hat{I}_{B^+,p,t} + \hat{I}_{X,p,t})| \tag{2.3a}$$

subject to, $\forall t \in \mathcal{T}$

$$V_{B,p,t} = V_{A,p,t} - \hat{Z}_{XA,p}\hat{I}_{A^+,p,t} - \hat{Z}_{BX,p}(\hat{I}_{A^+,p,t} + \hat{I}_{X,p,t}) \tag{2.3b}$$

with $(\hat{I}_{X,p}, \hat{Z}_{XA,p}, \hat{Z}_{BX,p}, \hat{Z}_{CB,p}) \in \mathbb{R}^4$.

For the *Leaf* configuration:

$$\delta^l = \min \sum_{t \in \mathcal{T}} |V_{B,p,t} - V_{C,p,t} - \hat{Z}_{CB,p}(\hat{I}_{A^+,p,t} + \hat{I}_{B^+,p,t} + \hat{I}_{X,p,t})| \tag{2.4a}$$

subject to, $\forall t \in \mathcal{T}$,

$$V_{B,p,t} = V_{A,p,t} - \hat{Z}_{BA,p}(\hat{I}_{A^+,p,t} + \hat{I}_{X,p,t}) \tag{2.4b}$$

with $(\hat{I}_{X,p}, \hat{Z}_{BA}, \hat{Z}_{CB}) \in \mathbb{R}^3$.

For the *Common parent* configuration:

$$\delta^{cp} = \min \sum_{t=0}^{T} |(V_{A,p,t} - \hat{Z}_{XA,p}\hat{I}_{A^+,p,t}) - (V_{B,p,t} - \hat{Z}_{XB,p}\hat{I}_{B^+,t,p})| \tag{2.5}$$

with $(\hat{Z}_{XA,p}, \hat{Z}_{XB,p}) \in \mathbb{R}^2$.

Once problems (2.3 - 2.5) have been solved, the configuration with the smallest value amongst $\delta \in \{\delta^b, \delta^l, \delta^{cp}\}$, i.e., the one with the minimum circuit laws violations, is selected as the most probable one.

The set of nodes $\hat{\mathcal{N}}_p$ is updated based on the selected configuration. For the bridge configuration, an additional node X is added to $\hat{\mathcal{N}}_p$. The current injection magnitude of the hidden node $\hat{I}_{X,p}$ and the values of $\hat{Z}_{XA,p}$ and $\hat{Z}_{BX,p}$ are estimated using the solution of Eq. (2.3). The hidden node voltage time series $V_{X,p}$ is computed as follows:

$$\hat{V}_{X,p,t} = V_{A,p,t} - \hat{Z}_{XA,p}\hat{I}_{A^+,p,t} \qquad \forall t \in \{1, \ldots, T\} \tag{2.6}$$

For the leaf configuration, the current injection magnitude of the hidden node $\hat{I}_{X,p}$ can be estimated using the solution of Eq. (2.4). However, the hidden node voltage time series cannot be estimated for the bridge configuration. To consider the effect of the hidden node, the algorithm replaces node A with a virtual node X' with the same voltage as A, but with the current injection calculated as follows:

$$\hat{I}_{X',p} = \hat{I}_{A,p} + \hat{I}_{X,p} \tag{2.7}$$

This allows to modify the current flowing in the upstream of A for the next validation step.

For the common parent configuration, an additional node X is added to $\hat{\mathcal{N}}_p$. The impedances $\hat{Z}_{XA,p}$ and $\hat{Z}_{XB,p}$ can be estimated using the solution of Eq.(2.5). The common node voltage magnitude $\hat{V}_{X,p}$ is estimated by averaging the voltage values computed from both downstream branches:

$$\hat{V}_{X,p,t} = (V_{A,p,t} - \hat{Z}_{XA,p}\hat{I}_{A^+,p,t} + V_{B,p,t} - \hat{Z}_{XB,p}\hat{I}_{B^+,p,t})/2 \tag{2.8}$$

For this configuration, the current injection of the hidden node cannot be computed from Eq.(2.5). The algorithm assumes that node X is indeed a zero-power injecting node. If edge CX is accepted in the next topology validation step, this also validates the assumption. Otherwise, node X will be added in a leaf configuration. Then, the value of $\hat{I}_{CX,p}$ will be estimated by solving the problem of Eq.(2.4).

If more than one hidden node is in the same area, the estimate of the node by solving the problems of Eqs.(2.3) - (2.5) will not correspond to an actual hidden node in the network. Therefore, the algorithm needs to check the validity of the topology formed by the addition of H. If the relative standard deviation of the impedance time series Z_{CX} for *bridge* and *leaf* configurations, or Z_{AX} and Z_{BX} for the *common parent* configuration are below the tolerance level λ, the algorithm accepts the identified hidden node. Otherwise, a new node Y will be added to the set of network nodes $\hat{\mathcal{N}}_p$.

The algorithm assumes that the hidden node Y is in the middle of nodes A and B, connected to them by branches with equal impedances, i.e., $\hat{Z}_{YA,p} = \hat{Z}_{BY,p}$. The impedances are evaluated by multiplying a per-unit-of-length default impedance Z^* by half the distance between the nodes. The voltage and current injection of the new hidden node are then calculated as follows:

$$\hat{V}_{Y,p,t} = V_{A,p,t} - \hat{Z}_{YA,p}\hat{I}_{A^+,p,t} \qquad \forall t \in \mathcal{T} \tag{2.9}$$

$$\hat{I}_{Y,p,t} = \frac{V_{B,p,t} - V_{Y,p,t}}{\hat{Z}_{BY,p}} - \hat{I}_{A^+,p,t} \qquad \forall t \in \mathcal{T} \tag{2.10}$$

The topology validation and estimation steps are repeated until all the edges in $\hat{\mathcal{G}}_p$ are labelled as valid. The impedance magnitude of each validated single-phase edge is set to the average over the three phases of the mean values of $\hat{Z}_{e,p,t}$ for all time steps.

2.6.2 Three-phase topology formation

This step merges the identified single-phase topologies $\hat{\mathcal{G}}_a$, $\hat{\mathcal{G}}_b$ and $\hat{\mathcal{G}}_c$ to form a three-phase model of the network. Each of the single-phase graphs $\hat{\mathcal{G}}_p$ has $\hat{F}_p$ feeders as the paths starting from the root node. To find a three-phase topology, these paths must be merged to form three-phase feeders.

Let us define $\delta_{k,f}$ as the distance between a single-phase feeder f and a node k in another single-phase feeder as follows:

$$\delta_{k,f} = \min\{\|\psi_k - (\rho\psi_i + (1-\rho)\psi_j)\|,\ \forall (i,j) \in f, \rho \in [0,1]\}, \tag{2.11}$$

where i and j are nodes of f with minimum distances to k.

The right part of the subtraction shows the coordinates of the closest point on feeder f between nodes i and j to node k. This distance is illustrated in Figure 2.5 for two single-phase feeders.

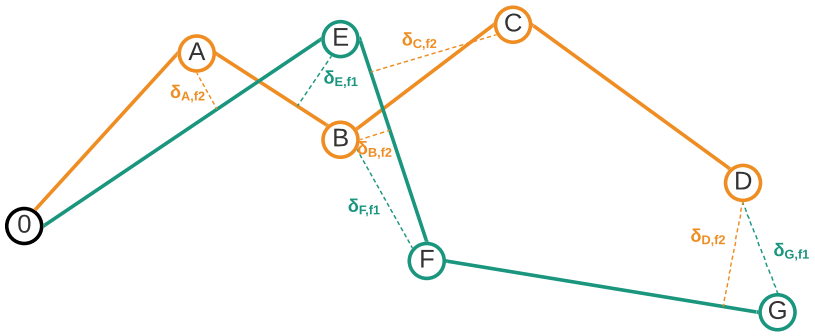


Figure 2.5: Illustration of the distance between single-phase feeders f_1 and f_2 to their nodes

The distance between two single-phase feeders, Δ_{f_i,f_j}, can be defined as follows:

$$\Delta_{f_i,f_j} = \frac{1}{2} \times \left(\frac{\sum\limits_{k \in f_i} \delta_{k,f_j}}{|f_i|} + \frac{\sum\limits_{k \in f_j} \delta_{k,f_i}}{|f_j|} \right) \tag{2.12}$$

where $|f_i|$ and $|f_j|$ are the number of nodes in single-phase feeders f_i and f_j, respectively.

Let $\hat{\mathcal{F}}$ be the set of all the single-phase feeders in the single-phase graphs. To merge the single-phase feeders to form three-phase feeders, we define a binary variable x_{f_i,f_j}, which is equal to 1 if the single-phase feeders f_i and f_j are in the same three-phase feeder, and 0 if they are not. Having the distances between the single-phase feeders, they can be grouped by solving the following optimisation problem:

$$\min_{x \in \mathbb{B}^{|\hat{\mathcal{F}}| \times |\hat{\mathcal{F}}|}} \sum_{f_i \in \hat{\mathcal{F}}} \sum_{f_j \in \hat{\mathcal{F}}} x_{f_i,f_j} \Delta_{f_i,f_j} \tag{2.13a}$$

subject to:

$$x_{f_i,f_j} = x_{f_j,f_i} \qquad \forall (f_i, f_j) \in \hat{\mathcal{F}}^2 \tag{2.13b}$$

$$\sum_{f_j \in \hat{\mathcal{F}}} x_{f_i,f_j} = 3 \qquad \forall f_i \in \hat{\mathcal{F}} \tag{2.13c}$$

$$x_{f_i,f_j} \geq x_{f_k,f_i} + x_{f_k,f_j} - 1 \qquad \forall (f_i, f_j, f_k) \in \hat{\mathcal{F}}^3 \tag{2.13d}$$

Constraint of Eq.(2.13b) ensures the symmetry of x. Constraint of Eq.(2.13c) ensures that every three-phase feeder contains one feeder from each single-phase graph. Constraint of Eq.(2.13d) states that if single-phase feeders f_k and f_j, and single-phase feeders f_k and f_i are in the same three-phase feeder, then the single-phase feeders f_i and f_j are also in the same three-phase feeder.

After grouping the single-phase feeders, the next step is to link the nodes of each group to form the three-phase network feeders. To identify the links between the nodes, the optimisation problem of Eq.(2.14) solves a minimum spanning tree problem involving node depths [36]. The result is an incidence matrix y, defining the edges $\hat{\mathcal{E}}$ of the three-phase network graph $\hat{\mathcal{G}} = (\hat{\mathcal{N}}, \hat{\mathcal{E}})$.

$$\min_{y \in \mathbb{B}^{|\hat{\mathcal{N}}| \times |\hat{\mathcal{N}}|},\ l \in \mathbb{N}^{|\hat{\mathcal{N}}|}} \sum_{i \in \hat{\mathcal{N}}} \sum_{j \in \hat{\mathcal{N}}} y_{i,j}\ W_{i,j} \tag{2.14a}$$

subject to:

$$\sum_{i \in \hat{\mathcal{N}}} y_{i,0} = 0 \tag{2.14b}$$

$$\sum_{i \in \hat{\mathcal{N}}} y_{0,i} = \hat{F} \tag{2.14c}$$

$$\sum_{j \in \hat{\mathcal{N}}} y_{j,i} = 1 \qquad \forall i \in \hat{\mathcal{N}} \tag{2.14d}$$

$$l_0 = 0 \tag{2.14e}$$

$$l_j \geq l_i + y_{i,j} - |\hat{\mathcal{N}}|(1 - y_{i,j}) \qquad \forall (i,j) \in \hat{\mathcal{N}}^2 \tag{2.14f}$$

$$l_j \leq l_i + 1 + |\hat{\mathcal{N}}|(1 - y_{i,j}) \qquad \forall (i,j) \in \hat{\mathcal{N}}^2 \tag{2.14g}$$

$$l_j \geq l_i + 1 \qquad \forall (i,j) \in \hat{\mathcal{E}}_p \tag{2.14h}$$

where $W_{i,j}$ is the Euclidean distance between two nodes i and j, only if they belong to the same phase or the same feeder:

$$W_{i,j} = \begin{cases} \|\psi_i - \psi_j\| & \text{if } (i,j) \in \hat{\mathcal{E}}_p \text{ or if } x_{i,j} = 1 \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.15}$$

Constraints of Eqs.(2.14b) - (2.14d) ensure that the identified graph is a rooted tree with the main substation as its root node 0 and set the expected number of feeders in the topology to $\hat{F}$. Equations (2.14e)-(2.14g) define l_j as the depth of node j and set the root node depth to zero. The constraint of Eq.(2.14h) imposes that if node i is followed by node j in the single-phase graph, the depth of node i in the three-phase graph is greater than the depth of j.

Finally, for each phase of each identified edge $e \in \hat{\mathcal{E}}_p, \forall p \in \mathcal{P}$ in the single phase topology, let $\hat{\Pi}(e)$ be the equivalent path of e in the three phase topology, i.e. a set of consecutive edges in the three phase topology representing e. Let $d_{\hat{\Pi}(e)}$ be the total length of this three phase topology path, the impedance of each edge b of that path $\hat{\Pi}(e)$ is calculated, for each phase, as follows:

$$\hat{Z}_b = \frac{\hat{Z}_e}{d_{\hat{\Pi}(e)}} d_b \qquad \forall b \in \hat{\Pi}(e) \tag{2.16}$$

where $\hat{Z}_e$ represents the single-phase impedance estimated in Section 2.6.1, and d_b the length of the edge b in the three-phase graph $\hat{\mathcal{G}}$.

Assuming that all the edges have the same type of conductors in all three phases, the magnitude of self-impedance of each three-phase edge is calculated as the mean value of the three single-phase impedances $\hat{Z}_{b,p}$ for that edge. The edges' resistance and reactance values are then computed with the hypothesis of a constant X/R ratio (γ).

2.7 Case study

In this section, the discussed methods are applied to a network inspired from a real Belgian LV distribution network and their performances are discussed.

The network is a three-phase, four-wire network with 128 single-phase customers distributed across four feeders. Only 40%, i.e., 52, of the customer nodes are equipped with smart meters. The network can be seen in Figure 2.6, the markers correspond to nodes with smart-meters.

The smart meter data includes power and voltage time series. Load profiles are extracted from residential load readings of the Low Carbon London Project [37]. For voltage time series, the simulations were run using OpenDSS [38] over 30 days at a 30-minute resolution, resulting in 1440 time-steps. Two different cable types were used, their self-impedance values were chosen as $\|Z_1\| = 0.0012\ \Omega/m$ and $\|Z_2\| = 0.0009\ \Omega/m$ for the main feeders and their laterals, respectively. Different power factors, chosen between 0.93 to 0.97, are allocated to each customer to generate the

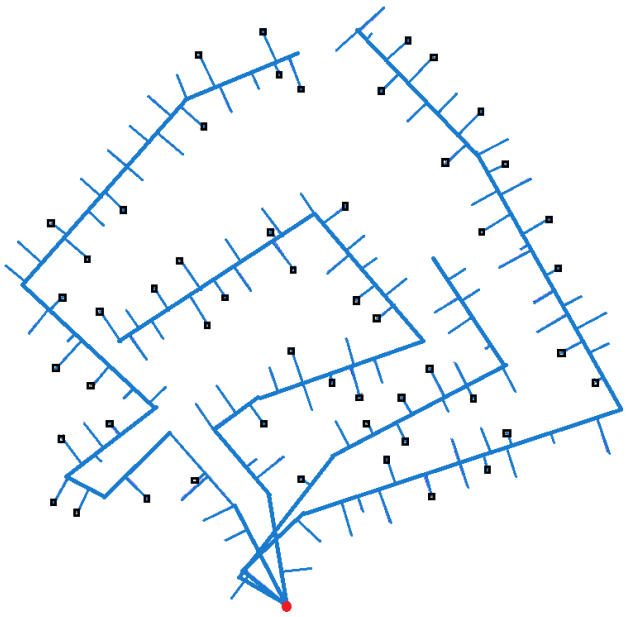


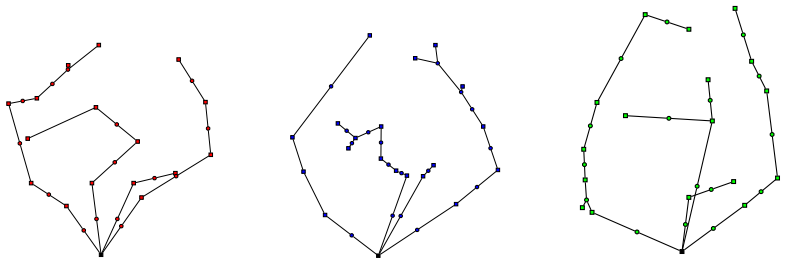
Figure 2.6: Network case study where each marker corresponds to a metered customer.

active and reactive power profiles. The load-flow analysis was also carried out using OpenDss.

As mentioned above, 40% of the customers have smart meters, which are distributed randomly. The identification algorithm inputs are voltages and power injections recorded by smart meters, and their geographical data. In this case study and to be consistent with common smart meter measurements, both voltage and power inputs are magnitude values with a unit power factor. The hyper-parameter settings are the following: $\lambda = 0.1$, $\gamma = 0.1$ and $Z^* = 0.001\ \Omega/m$.

The estimated single-phase topologies are shown in Fig. 2.7. Rectangles represent the nodes with smart-meters while the circles are *hidden nodes*. These single-phase topologies are used for the three-phase topology as described in the methodology. The estimated topology over the real network is shown in Figure 2.8. The algorithm can identify, as shown in the figure, the general shape of the network and it detects the correct number of feeders. Moreover, the metered nodes are correctly associated with their respective feeder.

Figure 2.9 presents the obtained impedances for each branch aggregated



(a) Phase A (b) Phase B (c) Phase C

Figure 2.7: Estimated single-phase topologies. Rectangles represent metered nodes while circle represent hidden nodes.

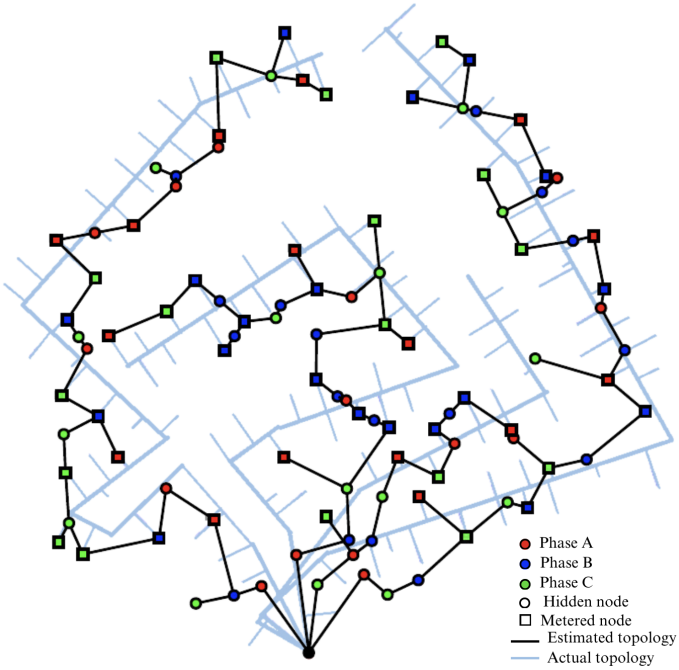


Figure 2.8: Network case study, the actual network is in light blue and the estimated network is in black.

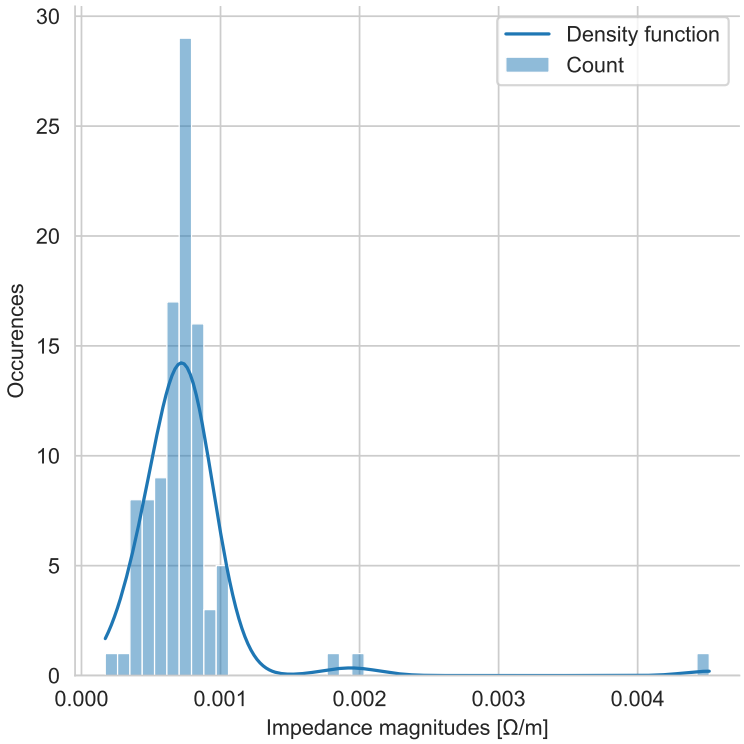
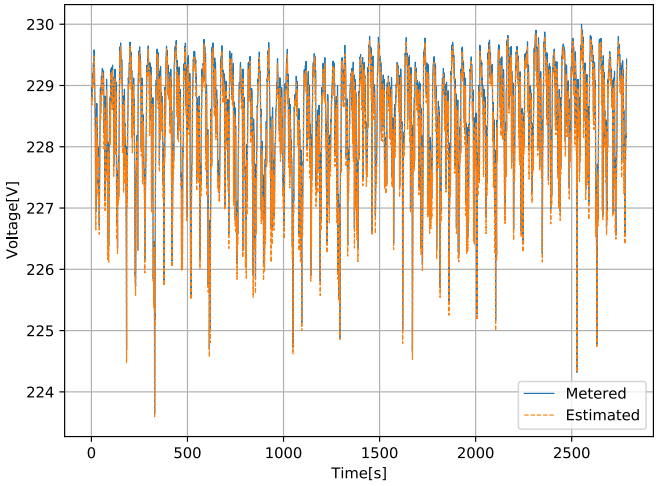


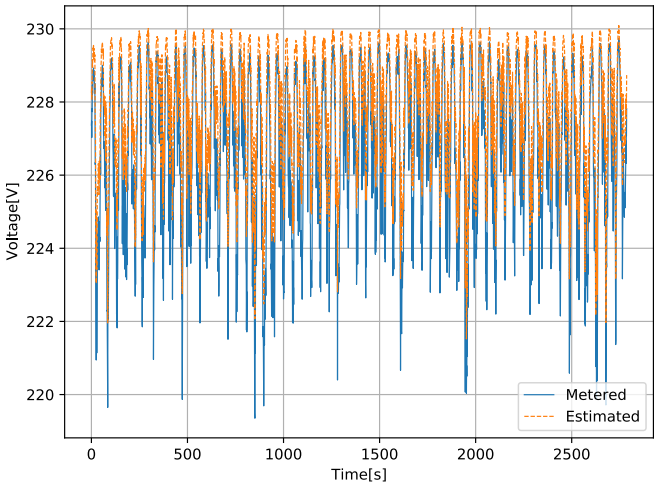
Figure 2.9: Impedance magnitudes.

by occurrences and the probability distribution. The maximum occurrences gravitate around the expected range $[0.0009, 0.0012]$. The mean however is shifted to a lower value of 0.0008 Ω/m with a standard deviation of 0.0005 Ω/m. When removing the 3 outlier points on the right of the figure, the standard deviation decreases to 0.0002 Ω/m. This can indicate that these points might be good locations to add a smart meter.

To assess the accuracy of the results, another load-flow analysis was computed with the resulting topology, impedances and estimated current injections. The voltage time-series provided by the load flow were compared to the input voltage readings. The minimum, median and maximum observed Root-Mean-Square Errors (RMSE) are 0.09%, 0.69% and, 3.53% respectively. These values suggest that despite only 40% of nodes equipped with smart-meters, the outputted model lead to acceptable voltage approximations for network planning. Figure 2.10 shows voltage magnitude of



(a) Node 180



(b) Node 215

Figure 2.10: Metered and estimated voltage magnitude.

two other nodes. While Figure 2.10a depicts a low estimation error, the estimated voltage profile of Figure 2.10b presents several under-voltages corresponding to less severe cases in the correct time-series. This error is due to a localised inaccuracy in the topology estimation that led to under-estimation of the load around the meter.

2.8 Intermediate conclusion

This chapter proposes an algorithm which identifies the topology of a low-voltage network when a subset of nodes, either customers or nodal points of connection, is not equipped with smart meters. This objective is achieved by analysing voltages, currents and geographical data of the metered nodes, without relying on additional information. The performances of the algorithm are evaluated on a case study with 128 customers, 52 of them equipped with smart meters. Despite the low observability of the system, the algorithm is able to produce a model whose topology reflects the structure of the network. A load-flow analysis performed using the inferred model shows that the computed voltage time-series matches the correct values with average RSME of less than 1%.

Future work could focus on how to exploit the solution of additional load-flow analysis using the estimated model to understand which area in the model topology presents larger inaccuracy and to investigate how to improve it. Furthermore, it would be interesting to investigate what the minimum number of meters needed is to obtain relevant solutions and where they should be installed for an optimal estimation of the model. Additionally, further effort could be put into a more exhaustive estimation of the edge parameters, taking into account the evaluation of mutual and shunt impedances. Finally, the model provided by this algorithm could be used for further studies to maximise the integration of renewable generation and electric car connections within LV networks.

Chapter 3

Hosting capacity formalism

I never learned from a man who agreed with me.

Robert A. Heinlein

Following the insights gained from the research on hosting capacity (presented in next chapter), it became evident that comparing various studies was a challenge due to the lack of a consistent and unified definition. The notion of hosting capacity is not trivial, and creating a universal definition requires a certain level of abstraction to encompass all potential scenarios. Over the past decade, while network hosting capacity has been widely studied, no standardised or universally accepted definition has been established. This chapter seeks to address this gap by proposing a unified definition of hosting capacity. The goal is to create a framework that can be applied consistently across different contexts, enabling better comparability and coherence in future research.

A practical, illustrative example is presented to concretely demonstrate the new definition, ensuring clarity and facilitating understanding of its practical application. Moreover, this chapter aligns key literature articles in the field with the proposed definition. Through this comparative analysis, it demonstrates how the proposed definition can be flexibly and effectively applied across a wide range of research contributions. This analysis also highlights the potential for a shared foundational understanding, which can serve as a basis for future studies and guide the determination of hosting capacity across various systems.

3.1 Notations

N	Network
E	Exogeneous data
T	Time
Sets	
$\mathcal{P}$	Set of network issues
	e.g., $\mathcal{P}$ = { over-voltage, overloading }
$\mathcal{C}$	Set of customer nodes
$\mathcal{H}$	Set of types of technologies { PV, EV, HP, ... }
$\mathcal{I}_h$	Set of installation options of technology $h \in \mathcal{H}$
	e.g., $\mathcal{I}_{\text{EV}}$ = { 0kW, 2.3 kW, 3.7 kW, 7.4 kW, 11 kW, 22 kW }
$\mathcal{T}$	Set of time steps, $\mathcal{T} = \{1, ..., T\}$.
$\mathcal{S}$	Set of all scenarios
$\mathcal{S}^c$	Set of considered scenarios
$\mathcal{S}^{\checkmark}$	Set of valid scenarios
$\mathcal{S}^{\times}$	Set of invalid scenarios
$\mathcal{S}^f$	Set of feasible scenarios
$\mathcal{S}^s$	Set of feasible and safe scenarios
$\mathcal{A}$	Set of all possible penetrations
	e.g., if the penetration is defined as the number of customers
$\mathcal{A}^c$	Set of considered penetrations
$\mathcal{A}^{\checkmark}$	Set of penetrations that can be associated to at least one valid scenario
$\mathcal{A}^{\times}$	Set of invalid penetrations
$\mathcal{A}^f$	Set of feasible hosting capacities
$\mathcal{A}^s$	Set of feasible and safe hosting capacity

3.2 Introduction

The ongoing energy transition is significantly changing production and consumption dynamics. The number of decentralised energy producers has been steadily increasing for several years. According to the International Energy Agency (IEA), photovoltaic panel (PV) generation increased by a record 270 TWh (up 26%) in 2022 [39]. On the other hand, high consumption technologies like heat pumps (HPs) and electric vehicles (EVs) are also experiencing growth, with the IEA anticipating penetration rates of at least 25% and 40% for 2050, respectively [39].

Technologies influencing the low-voltage distribution network such as photovoltaic panels, electric vehicles and heat pumps are referred to as

Distributed Energy Resources, DERs. Figure 3.1 illustrates the terminology for common technologies as in Chapter 1. Producers, or distributed generations (DGs), involve for instance PV and wind turbines (WT), while new consumers involve both HPs and EVs. Battery energy storage systems (BESSs) enable energy from renewables to be stored and then released when it is needed most. In some cases, EVs can be considered as BESSs. DERs is the terminology used to address all decentralised technologies, thus all of the aforementioned technologies are DERs. This terminology is used for the remainder of the chapter.

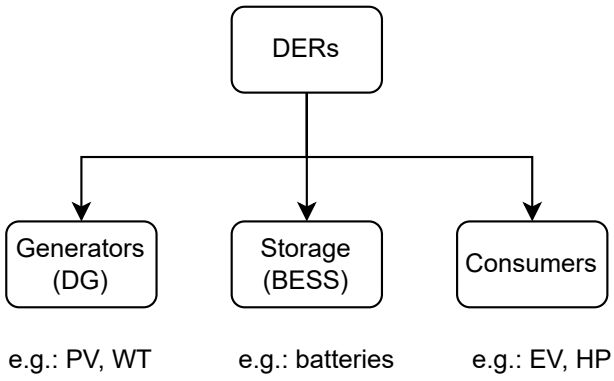


Figure 3.1: Terminology of the most common emerging technologies at distribution level. DG refers to distributed generator and BESS to battery energy storage systems.

Existing distribution networks, predominantly established in the last century, were not originally designed to accommodate such production and consumption transformations. This situation presents new challenges for distribution system operators (DSOs) in maintaining service levels and ensuring network reliability. Therefore, identifying and minimising the associated costs to alleviate these challenges becomes crucial.

DSOs need to identify the quantity of such technologies their network can host. This is complex as the quantity of installations that can be hosted depends on their location, the installed capacity, the consumption profiles, the potential production and many other factors. In order to provide a summarised view of all these factors, the concept of *hosting capacity* was introduced.

Hosting capacity (HC) is a measure of the capacity of a network to accommodate DER installations before encountering any operational issues. This capacity is generally expressed as one or multiple *penetration*, which can be a number of installations, their power, their consumption, etc. Figure 3.2 illustrates the concept of the definition. In this illustration, the hosting capacity is a function of the penetration rate, which represents how much of a new technology is added to the network; and a performance index, which determines whether any operational issue happened.

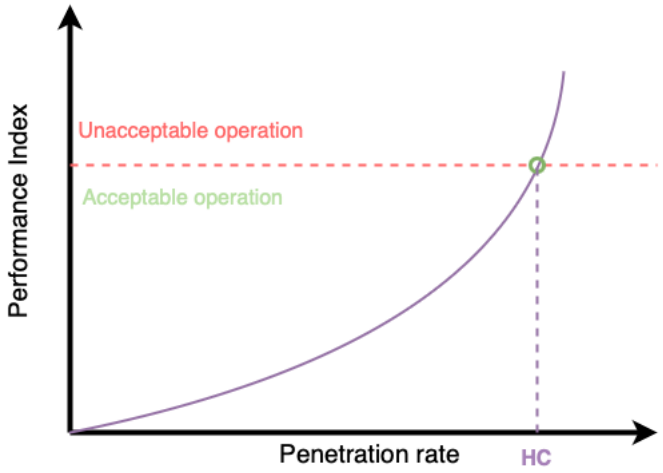


Figure 3.2: One dimension(performance index) representation of the concept of HC.

HC has become a popular and rapidly evolving subject of interest in the academic community; the large amount of literature available on HC in databases like Scopus and Google Scholar demonstrates its growing importance. Multiple variations of the definition of the penetrations and of the HC computation procedure have been experimented on by the community. There currently is, however, no established common ground for the mathematical definition of the HC problem. Given the extensive interest in HC, it is essential to gather and organise the knowledge into a coherent structure. To address this, this chapter aims at providing a unified definition of the hosting capacity problem that encompasses all these different aspects and to show how papers can be re-framed using the mathematical definition. Additionally, a concise review of the predominant literature is presented to highlight the simplicity of organising related work under the lens and terminology of the presented definition. This process will provide

guidance for future research in the area of HC.

The main contributions of this chapter are:

- Presentation of a formal unified definition of HC. This definition lays out the theoretical HC problem and provides common ground of all 'in practice' HC computations;
- Presentation of how predominant researches on HC falls into the generic HC definition;
- Reviewing similar research according to the proposed definition.

The structure of this chapter is as follows: first, Section 3.3 presents a framework that generalises HC. Then, two papers are reviewed as specific cases of the proposed framework in Section 3.4 and Section 3.5 illustrates how related works can now easily be compared using the framework. Finally, Section 3.6 concludes this chapter with a summary and the main outcomes.

Existing hosting capacity reviews

Various reviews have already been published on hosting capacity such as [40]–[48]. While some areas of these reviews naturally overlap, there are notable distinctions from the proposed review.

First, the technologies covered in [40], [41], [43]–[48] primarily focus on distributed generation, with [41], [44], [46], [47] specifically concentrating on PV technologies. On the other hand, the recent study in [42] focuses only on EVs. Then, the review in [45] is dedicated to medium voltage (MV). Enhancement techniques aiming at increasing the hosting capacity are discussed in [43], [45], [47] and in addition to that [43] presents an historical overview. Finally, [41] concludes that a general definition is needed.

This chapter stands apart from the aforementioned by adopting a more inclusive approach, considering distributed energy resources (DERs), and ultimately presenting a unified definition of hosting capacity.

Paper selection criteria

As aforementioned, a large quantity of papers addressing the hosting capacity issue are available. Therefore, the review in Section 3.5 intentionally avoids aiming for comprehensiveness. The aim of this review is to show how related works can easily be compared thanks to the presented framework. The reviewed papers were selected on several bases:

- While the definition is not constrained to low-voltage (LV) networks, and can be extended to medium-voltage networks, the selection of paper mostly focuses, for conciseness, on LV distribution networks.
- All types of DER were accepted;
- There was no filtering on the input of the methods as this is not the scope of this review.

Furthermore, only papers with a substantial number of citations were selected. The number was set to a hundred, thus papers with more than 100 citations on Scopus with the search terms "Hosting Capacity" were selected. This threshold can be judged as high but it is attributed to the abundance of papers on the topic garnering 100 citations. Some papers were selected outside of this criterion as they add an interesting value for the review and were sometimes heavily mentioned by others. Note that the two papers used in Section 3.4 were taken from the selection.

3.3 Definition of hosting capacity (HC)

In this section a unified definition of the hosting capacity is presented. The deterministic definition is first presented in 3.3.1 as well as an example to illustrate it, and then methods to model usual types of stochasticity and integrate them into the definition are presented in section 3.3.2. For the sake of readability, in the remainder of this chapter, a set is written using a calligraphic uppercase letter (e.g., $\mathcal{X}$) and vectors are written as lowercase bold characters (e.g., $\mathbf{x}$).

3.3.1 Deterministic definition

The hosting capacity[1] concept was, according to [43], first introduced in [49]:

> **Definition.** *The HC is the amount of new resources that can be hosted by a network before facing any issues, i.e., compromising its operational limits or violating safety constraints.*

The hosting capacity is computed for a given network. Let N denote a modelisation of a network for which one wants to determine the hosting

[1] It is important to clarify that the definition presented in this section is regarded as deterministic, but not in the sense described in [46].

capacity. This network model represents a topology and can contain, for instance, where all the nodes are, the links between them and also a description of the various electrical elements. The definition of N is left fuzzy on purpose to allow flexible representation of the network depending on, for instance, the software used to tackle the problem or the modelisation options such as considering an AC network or a DC one. Note that network and network model are used interchangeably hereafter. In addition to the network model, exogenous data E are defined, which include factors such as sun irradiation or load profiles.

Inside the network are customers that can withdraw or inject power, install new DERs, or more generically have an impact on the network. The set of such customers, also referred as *connection nodes*, is denoted by $\mathcal{C}$ and an element of this set is referenced using the index c. The set of possible types of DERs that customers can add is denoted $\mathcal{H}$. En element h of $\mathcal{H}$ can be for example EV or PV. When deploying a particular technology, like PVs for instance, multiple installation options are available, allowing for decisions to be made on various aspects, such as the installation capacity and phase configuration, among other considerations. The set of possible installation options for a technology $h \in \mathcal{H}$ is defined as $\mathcal{I}_h$. Note that when a technology $h \in \mathcal{H}$ is not installed for customer $c \in \mathcal{C}$, then the corresponding value $\mathcal{I}_h$ is chosen equal to $\emptyset$.

The hosting capacity is determined over a specific time period, denoted by $\mathcal{T}$, during which installations can be modified and their impact on the overall system can be assessed. This time period can be continuous or discrete but is always bounded. Without loss of generality, 1 is always the first time considered and T is the last. Typically, $\mathcal{T}$ is either $[1, T]$ or $\{1, \dots, T\}$. An element of $\mathcal{T}$ is referenced using the index t.

At each time step, a customer has a state for each technology that is its installed option. This state is written as:

$$s_{c,h,t} \in \mathcal{I}_h, \forall t \in \mathcal{T}, c \in \mathcal{C}, h \in \mathcal{H}. \tag{3.1}$$

Gathering all the states for all customers, time steps and technology types, forms a scenario: a hypothetical evolution of the network over the time period with the given set of technology types. Formally, a scenario is defined as the tuple of states of all customers for all time steps and all technology types:

$$\begin{aligned} \mathbf{s} &= \langle s_{c,h,t} \mid \forall t \in \mathcal{T}, h \in \mathcal{H}, c \in \mathcal{C} \rangle \\ &= \langle s_{c_1,h_1,t_1}, s_{c_1,h_1,t_2}, \dots, s_{c_1,h_2,t_1}, s_{c_1,h_2,t_2}, \dots, s_{c_2,h_1,t_1}, \dots \rangle. \end{aligned} \tag{3.2}$$

By design, only one installation type can be installed per technology and per customer. It is not restrictive as $\mathcal{I}_h$ can accommodate for these options.

For example, the capacity of a PV can vary as well as the connected phases, in which case $\mathcal{I}_{\text{PV}}$ could be defined as :

$$\mathcal{I}_{\text{PV}} = \{\emptyset, (\text{phase 1: 5KW}), (\text{phase 2: 10KW}), \\ (\text{phase 1: 5KW; phase 2: 5kW}), (\text{phase 1: 10KW}), \ldots\}.$$

The set of all possible scenarios is $\mathcal{S}$, formally:

$$\mathcal{S} = \bigtimes_{h \in \mathcal{H}, c \in \mathcal{C}, t \in \mathcal{T}} \mathcal{I}_h. \tag{3.3}$$

Even though only one scenario will realise for a given network, the hosting capacity is independent from this scenario as it aims to evaluate the capacity of the network to host new technologies. Therefore, determining the hosting capacity is subject to several uncertainties. Two categories of uncertainty were defined: aleatory and epistemic [46]. Aleatory uncertainties, alternatively known as certain uncertainties or inherent uncertainties, deal with uncertainties that are known to be stochastic. These includes the new installations' consumption or production, as well as the customer loads. These uncertainties are part of the exogenous data E. Epistemic uncertainties, also referred to as systematic uncertainties or uncertain uncertainties, result from the lack of knowledge or information. These uncertainties are the location of the new installations and their type, the options chosen and the time steps when the installations are added. These uncertainties are taken into account by evaluating the hosting capacity on multiple scenarios.

As the number of possible scenarios could be intractable, a subset of $\mathcal{S}$ is often *considered*:

$$\mathcal{S}^c \subseteq \mathcal{S}, \tag{3.4}$$

where $\mathcal{S}^c$ is the set of considered scenarios, with the superscript c chosen to emphasise this.

For a given scenario, the penetration is a measure that gauges the amount of resources present on the network. It is represented by a vector $\mathbf{a} \in \mathcal{A}$, the set $\mathcal{A}$ being the one of all representable penetrations. A penetration can be, for instance, the number of new DERs. Let the function $\mathbf{g}(\mathbf{s}) : \mathcal{S}^c \to \mathcal{A}$ compute the penetration for a given scenario $\mathbf{s}$. For instance, $\mathbf{g}(\mathbf{s})$ can return a vector composed of both the number of customers with PVs and the total production of the PVs in scenario $\mathbf{s}$. $\mathbf{g}(\mathbf{s})$ is not injective: multiple scenarios can have the same penetration (for example, if the penetration is defined as the number of PVs installed, two scenarios can have the same number of PVs in different places).

The set of *considered* penetrations can thus be defined from the previously defined set of *considered* scenarios:

$$\mathcal{A}^c = \{\mathbf{g}(\mathbf{s}) \mid \forall \mathbf{s} \in \mathcal{S}^c\} \tag{3.5}$$

The hosting capacity depends on the previously mentioned penetrations and is primarily governed by the physical constraints of the network; indeed, the addition of new DERs can lead to network issues. Let the set $\mathcal{P}$ be the set of network issues that can be encountered. An example of issue is nodal over-voltage, i.e. a point in the network having a voltage greater than a defined threshold. Let $f_t(\mathbf{s})$ be a binary function which identifies whether, at time t and for the scenario $\mathbf{s}$, any issues from $\mathcal{P}$ occurs in the network N. To evaluate f_t, both the network N data and exogenous data E are required. The function f_t is formally defined as follows:

$$f_t(\mathbf{s}) = \begin{cases} 1, & \text{if at least one issue } p \in \mathcal{P} \text{ occurs at time } t \in \mathcal{T} \text{ in } N \text{ given } E; \\ 0, & \text{otherwise.} \end{cases} \tag{3.6}$$

Let f be an aggregation of f_t over the time period based on defined conditions. For instance, f could be defined as returning 1 if at least for one time step there was an issue in the network (f_t is equal to one), i.e.:

$$f(\mathbf{s}) = \begin{cases} 1, & \text{if} \quad \bigvee_{t \in \mathcal{T}} f_t(\mathbf{s}) = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.7}$$

Where the symbol $\bigvee$ represents the logical *or*. Usually, papers about hosting capacity differ on the way the scenarios are defined as well as how both f and g are defined. This will be addressed in Section 3.5.

The set of validated scenarios is defined as the set of scenarios where no issues are detected:

$$\mathcal{S}^{\checkmark} = \{\mathbf{s} \in \mathcal{S}^c \mid f(\mathbf{s}) = 0\}. \tag{3.8}$$

The validated penetration set is derived from this set of scenarios. It is the subset of $\mathcal{A}$ which can be associated with **at least** a valid scenario:

$$\mathcal{A}^{\checkmark} = \{\mathbf{g}(\mathbf{s}) \mid \forall \mathbf{s} \in \mathcal{S}^{\checkmark}\}. \tag{3.9}$$

Similarly, the set of non-validated scenarios and the set of invalid penetrations are defined as

$$\mathcal{S}^{\times} = \{\mathbf{s} \in \mathcal{S}^c \mid f(\mathbf{s}) \neq 0\}, \qquad \mathcal{A}^{\times} = \{\mathbf{g}(\mathbf{s}) \mid \forall \mathbf{s} \in \mathcal{S}^{\times}\}. \tag{3.10}$$

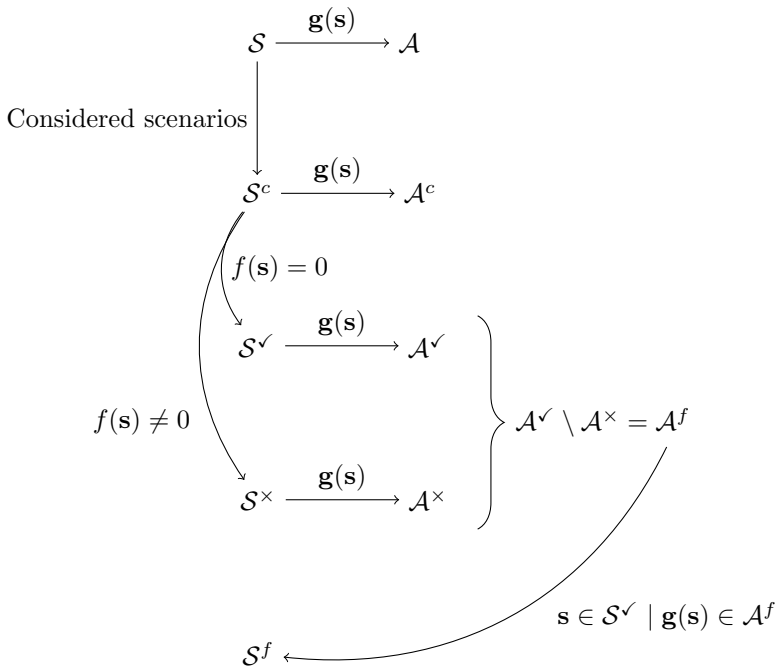
The hosting capacity of a network, referred as feasible penetrations $\mathcal{A}^f$, is defined as the valid penetrations that cannot be linked to an invalid scenario:

$$\mathcal{A}^f = \mathcal{A}^{\checkmark} \setminus \mathcal{A}^{\times}. \tag{3.11}$$

Note that $\mathcal{A}^f, \mathcal{A}^\checkmark, \mathcal{A}^\times \subseteq \mathcal{A}^c \subseteq \mathcal{A}$, and $\mathcal{A}^f \subseteq \mathcal{A}^\checkmark$. Let $\mathcal{S}^f$ be the set of scenarios leading to a penetration in $\mathcal{A}^f$:

$$\mathcal{S}^f = \left\{ \mathbf{s} \in \mathcal{S}^\checkmark \mid \mathbf{g}(\mathbf{s}) \in \mathcal{A}^f \right\}. \tag{3.12}$$

The relations between the sets $\mathcal{S}$, $\mathcal{A}$, and their derivatives are schematically explained below.



Note that in $\mathcal{A}^f$, having a (valid) penetration $\mathbf{a}$ does not imply that all penetrations $\mathbf{a}' \in \mathcal{A}$ that are dominated by $\mathbf{a}$ ($\mathbf{a}' \leq \mathbf{a}$ for some partial comparison operator $<$) are also valid. This means that for instance one scenario with more installations can have no issues while a scenario with fewer installations can have one. This is mainly due to issues potentially being dependent on the topology and the location of the installations. Therefore, one could define stricter limits for the final hosting capacity.

$\mathcal{A}^s$, the *safe* penetrations, is defined as the subset of $\mathcal{A}^f$ which possesses the above-mentioned property, i.e., all dominated penetrations are feasible (recall that $\leq$ is a partial comparison operator between two penetrations):

$$\mathcal{A}^s = \left\{ \mathbf{a} \in \mathcal{A}^f \mid \nexists \mathbf{a}' \in \mathcal{A}^\times : \mathbf{a}' \leq \mathbf{a} \right\}. \tag{3.13}$$

The corresponding scenario set is denoted $\mathcal{S}^s$ and defined by:

$$\mathcal{S}^s = \left\{ \mathbf{s} \in \mathcal{S}^f \mid \mathbf{g}(\mathbf{s}) \in \mathcal{A}^s \right\}. \tag{3.14}$$

The defined scenario sets are illustrated in Figure 3.3.

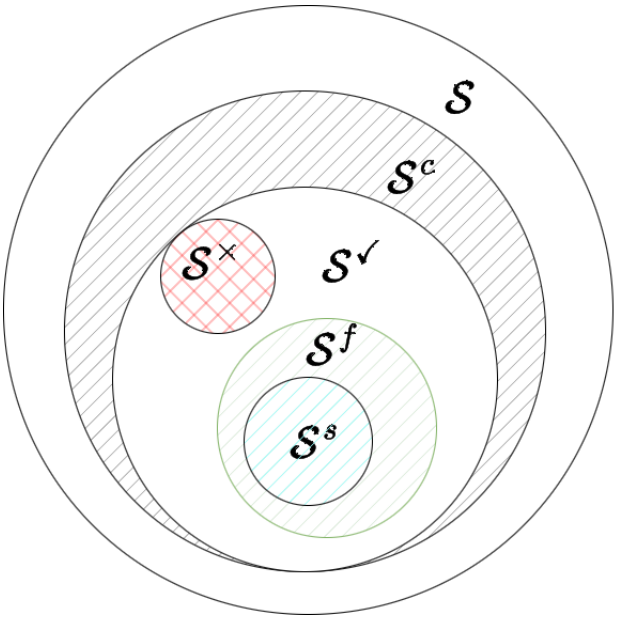


Figure 3.3: Sets of scenarios, where $\mathcal{S}$ is the set of all possible scenarios, $\mathcal{S}^c$ is the set of considered scenarios, $\mathcal{S}^\times$ the set of invalid scenarios, $\mathcal{S}^\checkmark$ of valid scenarios, $\mathcal{S}^f$is the set of feasible scenarios and $\mathcal{S}^s$ is the set of safe scenarios.

For penetrations in one dimension, most papers choose to use $\mathcal{A}^s$ rather than $\mathcal{A}^f$ as their definition for the hosting capacity, as it is a single connected space. Such papers generally define the penetrations $\mathcal{A}$ as $\mathbb{R}_+$ or $\mathbb{Z}_+$ (or subsets), thus making $\mathcal{A}^s$ a continuous or discrete range starting at zero and ending at a maximum value that is generally reported as the hosting capacity. Some papers also report these as probability density functions.

In Figure 3.4, an illustrative example of the different penetration sets defined above is given. All subfigures lie in the plan formed by the set $\mathcal{A}$, here exemplified by $\mathcal{A} = \mathbb{R}^2_+$. The two dimensions of this example can for example represent the number of PVs and EVs in a scenario. The sets $\mathcal{A}^\checkmark$ and $\mathcal{A}^\times$, representing respectively the penetrations reachable by valid scenarios and invalid ones, are shown in yellow and red respectively, as

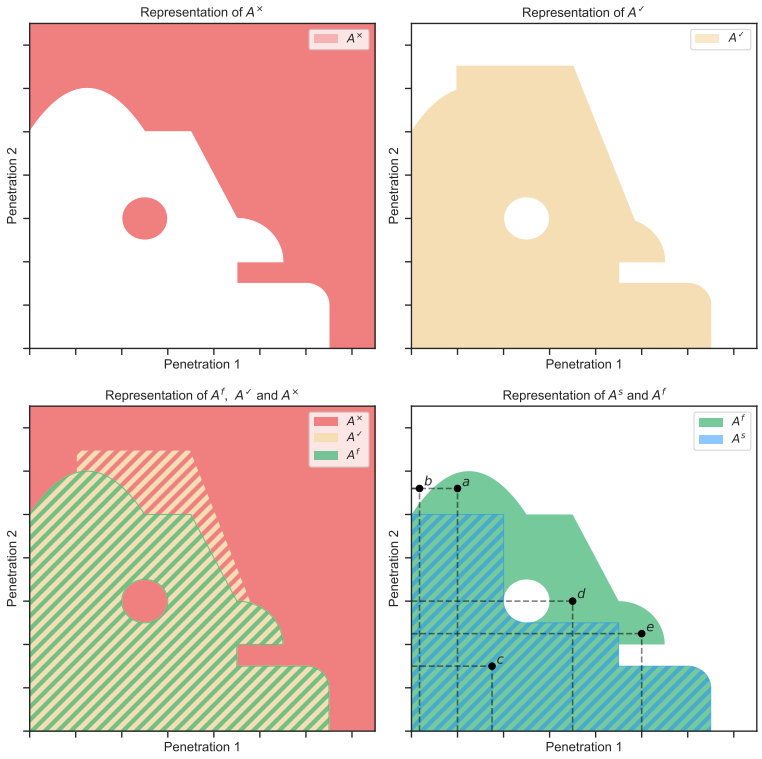


Figure 3.4: Two-dimensional representations of set $\mathcal{A}^f$, $\mathcal{A}^s$, $\mathcal{A}^\checkmark$ and $\mathcal{A}^\times$. The axes are a quantification of the penetration of two installations. The stripes mean that sets intersect.

can be seen in the top subfigures. They can intersect: there can exist two scenarios sharing the same number of PVs and EVs such that one fails and the other is valid, depending, for example, on the location of the PVs/EVs in the network topology.

The feasible penetration set $\mathcal{A}^f$, in green in the lower subfigures, is the the subset of $\mathcal{A}^\checkmark$ (in yellow) that is not part of $\mathcal{A}^\times$ (in red). A feasible penetration $p = \langle p_1, p_2 \rangle$ that is in $\mathcal{A}^f$ can thus be reported (for example, to the DSO) as supported by the network ("Having p_1 PVs and p_2 EVs would only lead to valid scenarios").

This may, however, be seen as abusive in certain contexts as shown in the lower right subfigure. For example, the penetration a is in $\mathcal{A}^f$ and, thus, all scenarios having this penetration are valid. However, point b, which lies

to the left of a (meaning that it can be, for example, the penetration with the same number of EVs but fewer PVs), is not in $\mathcal{A}^f$; all scenarios having penetration b are not valid. In the notations presented above, a is said to be *dominating* b, that is $b \leq a$, using here the standard element-wise vector comparison operator.

To avoid such contradictions about dominated but invalid penetrations ("having 10 PVs and 100 EVs is supported in the network but not 10 PVs and 50 EVs"), one can instead use the set $\mathcal{A}^s$, defined as the set of points where this situation cannot happen: for any penetration in $\mathcal{A}^s$, all penetrations lower than it are also valid. Thus, the penetration a is not in $\mathcal{A}^s$ due to b being a counterexample. However, c is in the set as the rectangle (in this 2-D example) on its lower left is fully included in $\mathcal{A}^f$. The penetration d is not in the set as the rectangle on its lower left encompasses part of the hole in $\mathcal{A}^f$. The same is true for the penetration e.

Hosting capacity

Summary of the definition of hosting capacity:

> The HC is the set of penetrations the network can sustain while not encountering any issues. More precisely, the HC is chosen as the feasible penetrations $\mathcal{A}^f$ which are penetrations that are associated with scenarios with no issues:
>
> $$\mathcal{A}^f = \{\mathbf{g}(\mathbf{s}) | \forall \mathbf{s} \in \mathcal{S} : f(\mathbf{s}) = 0\}.$$
>
> This set, and the hosting capacity, can be reduced to be the safe penetrations $\mathcal{A}^s$ with the constrain on the penetrations $\mathbf{a}$: $\nexists \mathbf{a}' \in \mathcal{A}^\times : \mathbf{a}' \leq \mathbf{a}$.

Up until here, the hosting capacity was introduced as sets ($\mathcal{A}^f$ and $\mathcal{A}^s$) representing all the penetrations that can be hosted. Alternatively, one can work with the frontiers of the sets, which is easier in one dimension.

In one dimension, most studies evaluate penetrations in $\mathcal{A}^s$ while not dealing explicitly with the sets defined above but rather with a scalar: indeed, in one dimension, $\mathcal{A}^s$ is a range between 0 and another scalar, that is sometimes called the hosting capacity as a shortcut. It can be equivalently computed by

$$\max_{\mathbf{a}} : \{\mathbf{s} \mid \mathbf{g}(\mathbf{s}) \leq \mathbf{a}\} \cap \mathcal{S}^\times = 0. \tag{3.15}$$

This is referred to in the literature as *first violation* or *minimal hosting capacity*. Furthermore, some studies also determine the *maximal hosting capacity* (also referred to as *all scenarios with violation*) which is defined

as the minimal penetration for which all scenarios encounter at least one issue:

$$\min_{\mathbf{a}} : \{\mathbf{s} \mid \mathbf{g}(\mathbf{s}) > \mathbf{a}\} \subset \mathcal{S}^{\times}. \tag{3.16}$$

Note that working with penetrations in $\mathcal{A}^f$ would imply several frontiers (in one dimension, multiple ranges).

In more dimensions, giving one value as a summary for the hosting capacity is more complex as the frontiers of the sets are not scalars. The concepts presented above can be extended to higher dimension via the use of partial comparison operators as done earlier in this chapter.

3.3.2 Stochastic definition

In the following section, two ways to introduce stochasticity in the presented definition are addressed. The first manner to introduce stochasticity is related to the considered scenarios, while the second is on the detection of issues. These are not mutually exclusive: one may consider more than one way to introduce stochasticity in determining the hosting capacity.

Stochasticity related to scenarios selection

As previously mentioned, evaluating all scenarios is, most of the time, intractable. Therefore, the subset $\mathcal{S}^c$ is necessary to select a subset of scenarios to consider. Building this set can be done in multiple ways. For example, this set can be constructed by sampling the scenarios in $\mathcal{S}$ using a distribution on $\mathcal{S}$, that may be informed with external information. This distribution, denoted $\mathbb{D}_{\mathcal{S}}$, represents the probability of realisation of each scenario. This allows to built $\mathcal{S}^c \sim \mathbb{D}_{\mathcal{S}}$ with a limited number of scenarios that are nevertheless representative. Note that all scenarios can have the same probability ($\mathbb{D}_{\mathcal{S}}$ being in this case the uniform distribution), which is often the case in the papers reviewed in section 3.5.

Stochasticity related to issues detection

The functions f_t and f were introduced as binary functions, one identifies if an issue $p \in \mathcal{P}$ occurred in the network N given E at time t and the other is aggregating f_t. A function $h_t \in [0, 1]$ and its aggregation $h \in [0, 1]$ can be defined as returning the probability of an issue occurring in N given E. The stochastic of function h can come from E that can itself be stochastic, for instance if E contains a distribution of load time series. Given a threshold

α for the probability, f_t can now be:

$$f_t(\mathbf{s}) = \begin{cases} 1, & \text{if } h_t(\mathbf{s}) > \alpha \\ 0, & \text{otherwise.} \end{cases} \tag{3.17}$$

This translates into f_t indicating that a scenario that has less than $\alpha\%$ probability of having an issue at time t is valid. Or, for all time steps, f can be defined as:

$$f(\mathbf{s}) = \begin{cases} 1, & \text{if } h(\mathbf{s}) > \alpha \\ 0, & \text{otherwise;} \end{cases} \tag{3.18}$$

i.e. having a scenario that has less than $\alpha\%$ probability of having an issue over the whole time range. Multiple variant of this are possible: the framework presented here can accommodate a wide range of settings.

As an example, consider the aggregation f that returns 0 when no issues occur for at most 95% of the time steps of the scenario, i.e.,

$$f(\mathbf{s}) = \begin{cases} 0, & \text{if } \mathop{\mathbb{E}}_{t \in \mathcal{T}} \dfrac{f_t(\mathbf{s})}{|\mathcal{T}|} < 0.95 \\ 1, & \text{otherwise.} \end{cases}$$

3.3.3 Example

The following subsection illustrates, with a small example, how the formalism of the generic definition can be applied to describe a hosting capacity computation. This example showcases the determination of the hosting capacity, taking into account only over-voltage issues, for a small distribution network where customers can add new photovoltaic capacities.

Following the formalism of the definition, the example is presented as follows:

- The **network model** N, represented in Figure 3.5, is composed of 8 buses: 7 low-voltage ones and 1 medium voltage one. Let $\mathcal{B}$ denote the set of buses, $\mathcal{B} = \{b0, b1, b2, b3, b4, b5, b6, b7, b8\}$. The network is modelled using PandaPower [50]. The medium-voltage bus is supplied by an external grid with a voltage set to 1 p.u., then a transformer converts medium-voltage to low-voltage. The transformer has a maximum apparent power handling capacity of 0.4 MVA and operates with a primary voltage of 20 kV and a secondary voltage of 0.4 kV. The transformer was chosen from the PandaPower standard library. All lines (L0 to L6) have the same standard type ("NAYY 4x50 SE" from PandaPower). Feeder lines (L0 and L3) have a length of 800m while lateral lines are 200m long except for L2, which is 400m long.

- Five **customers** are modelled by loads added to buses: $\mathcal{C} = \{b3, b4, b6, b7, b8\}$.
- In addition to the loads, customers $b3$ and $b4$ can install PVs as **installation type**, therefore $\mathcal{H} = \{\text{PV}\}$.
- These customers can choose, as **options**, either to install 50 PV panels of $300W_{peak}$ or to install no PVs at all, thus $\mathcal{I}_{\text{PV}} = \{\emptyset, 50 \times 300W_{peak}\}$.
- The **time period** is composed of one time step ($\mathcal{T} = \{1\}$).
- In this example, the **exogenous data** E, are the PV productions and customer loads. PV installations are at their peak production, i.e. a PV installation produces 15kW during the time frame, while the customers' loads are 1kW except for $b4$, which is 0.8kW. Note that the model only considers active power.
- All the **considered scenarios** are regrouped in Table 3.1. Scenario a is the initial topology as shown Fig. 3.5, scenarios b, c and d are, respectively, presented in Fig. 3.6a, Fig. 3.6b, Fig. 3.6c.
- The **issue** considered is over-voltage ($\mathcal{P} = \{\text{OV}\}$). An over-voltage at time t is flagged when the voltage ($V_{b,t}$) is greater than 1.05 pu for at least one bus b in $\mathcal{B}$, i.e.:

$$f(s) = f_t(s) = \begin{cases} 1, & \text{if } \exists b \in \mathcal{B} : V_{b,t} > 1.05pu; \\ 0, & \text{otherwise.} \end{cases}$$

  The aggregation f is the same as f_t as there is only one time step.

Scenarios in $\mathcal{S}^c$	b3	b4
a	$\emptyset$	$\emptyset$
b	$50 \times 300W_{\text{peak}}$ PVs	$\emptyset$
c	$\emptyset$	$50 \times 300W_{\text{peak}}$ PVs
d	$50 \times 300W_{\text{peak}}$ PVs	$50 \times 300W_{\text{peak}}$ PVs

Table 3.1: Considered scenarios ($a, b, c, d \in \mathcal{S}^c$) for the example for installing PVs with uncertain location. $\emptyset$ means that no installation was added at that customer.

The power flow is computed using PandaPower with the single-phase model of the network. The outputs are given in Table 3.2 for each scenario.

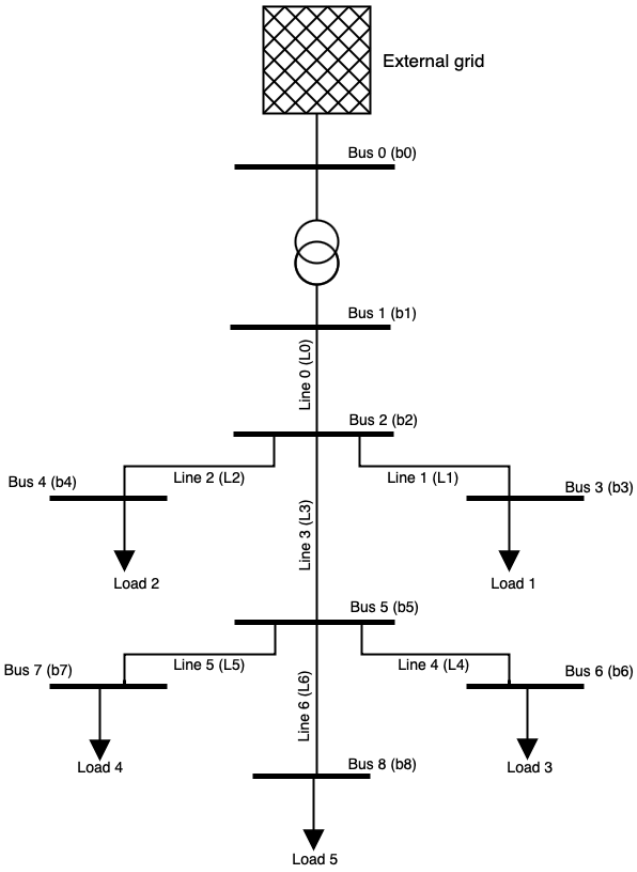


Figure 3.5: Example network with 5 customers represented by 5 loads.

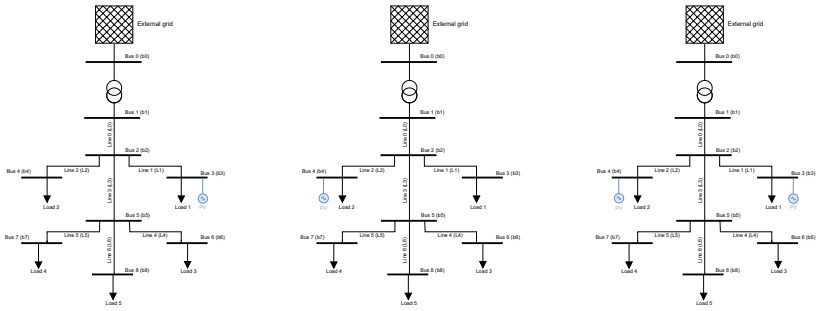
Scenarios c and d have at least an over-voltage at one bus (in V_{b4} for example), while the others have no bus in over-voltage. Therefore,

$$f(a) = f(b) = 0, \text{ and}$$
$$f(c) = f(d) = 1.$$

The function g returns, as a penetration measure, the number of buses with a PV installation:

$$g(s) = |\{b \in \mathcal{B} \mid s_b \neq \emptyset\}|,$$

where s_b designates the installations at bus b in scenario s. Note that $g(s)$



(a) Scenario b (b) Scenario c (c) Scenario d

Figure 3.6: Topologies of scenarios b, c and d in $\mathcal{S}^c$. In blue is the added PV installation.

$s \in \mathcal{S}^c$	V_{b0}	V_{b1}	V_{b2}	V_{b3}	V_{b4}	V_{b5}	V_{b6}	V_{b7}	V_{b8}
a	1	1	0.98	0.98	0.98	0.97	0.97	0.97	0.97
b	1	1	1.03	1.04	1.03	1.02	1.02	1.02	1.02
c	1	1	1.03	1.03	**1.05**	1.02	1.02	1.02	1.02
d	1	1	**1.08**	**1.09**	**1.10**	**1.07**	**1.07**	**1.07**	**1.07**

Table 3.2: Powerflow voltage output in p.u. by scenario. Over-voltages are in **bold**.

is a scalar and thus not written in bold. The penetration for each scenario is given in Table 3.3.

The penetration sets and the related scenario sets, are:

- The set $\mathcal{A}^c$ of all **considered** output of $g(s)$, $\forall s \in \mathcal{S}^c$, is thus $\{0, 1, 2\}$;
- The set of **valid scenarios** is $\mathcal{S}^{\checkmark} = \{a, b\}$ and $\mathcal{A}^{\checkmark} = \{0, 1\}$;
- The set of **invalid** ones is $\mathcal{S}^{\times} = \{c, d\}$ and $\mathcal{A}^{\times} = \{1, 2\}$;
- By definition of **feasible penetration set** $\mathcal{A}^f$, that is the penetrations that includes only valid scenarios, $\mathcal{A}^f = \mathcal{A}^{\checkmark} \setminus \mathcal{A}^{\times} = \{0\}$.

The different sets of scenarios and their corresponding penetration sets are summarised in Table 3.4. The *feasible* $\mathcal{A}^f$ and *safe* $\mathcal{A}^s$ penetration sets are the same, thus the hosting capacity of this example is $\mathcal{A}^f = \mathcal{A}^s = \{0\}$. As this example is one-dimensional, this corresponds to the *first violation* as

$s \in \mathcal{S}^c$	$g(s)$
a	0
b	1
c	1
d	2

Table 3.3: Penetration for each scenario computed using $g(s)$.

defined in Section 3.3. *All violation* HC is 2, as all scenarios with penetration 2 are not valid. Indeed, scenario d, the only scenario with a penetration equal to 2, is not valid.

Scenarios sets	penetration sets
$\mathcal{S}^{\checkmark} = \{a, b\}$	$\mathcal{A}^{\checkmark} = \{0, 1\}$
$\mathcal{S}^{\times} = \{c, d\}$	$\mathcal{A}^{\times} = \{1, 2\}$
$\mathcal{S}^f = \mathcal{S}^s = \{a\}$	$\mathcal{A}^f = \mathcal{A}^s = \{0\}$

Table 3.4: Scenarios sets and their penetrations.

Deterministic versus Stochastic cases

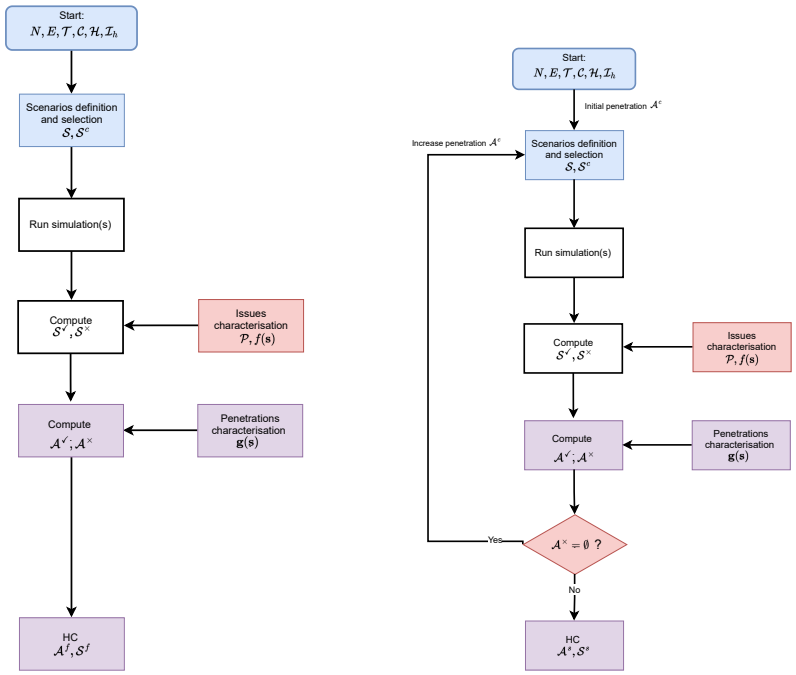
This example is deterministic as no uncertainty is taken into account: all possible scenarios are evaluated and exogenous data are unique. A stochastic approach would either sample scenarios or sample exogenous data to account for uncertainty. For instance, adding stochasticity for scenario selection by randomly sampling scenarios: scenarios a and b could be sampled and thus, $\mathcal{S}^c$ is $\{a, b\}$. In such context, the hosting capacity would be $\mathcal{A}^f = \{0, 1\}$.

3.4 Application of the framework on two papers

In this section, the framework presented in Section 3.3 is applied to two papers to illustrated how the concepts are applied to the literature.

In this framework, how the HC is actually computed is not explained and there are many ways to do so in the literature. Some of them follows directly the natural steps arising from the framework, while others iteratively refine their scenario definitions until they reach a predefined condition. These two workflows are shown in Figure 3.7. Note that this chapter does not

elaborate on how the simulations to determine the physical state of the network for given scenarios are run.



(a) Direct flow (b) Iterative flow

Figure 3.7: Flowcharts of a direct and an iterative HC method. In blue are the system assumptions, in red the issue characterisation, and in violet the HC computation, as presented in sections 3.5.1, 3.5.2 and 3.5.3, respectively. The iterative flow 3.7b computes the *safe* penetrations as the stop condition ensures that all dominated penetrations are feasible. Note that in white are steps not discussed in this chapter.

Both workflows follow this general scheme: first the scenario sets $\mathcal{S}$ and $\mathcal{S}^c$ are build for a network N and exogenous data E. These scenarios encompasses the technology types $\mathcal{H}$ and their options $\mathcal{I}_h$. Then, to construct the sets $\mathcal{S}^{\checkmark}$, $\mathcal{S}^{\times}$ and $\mathcal{S}^f$, the function f is introduced given the sets of considered issues $\mathcal{P}$. Note that most studies do not explicitly construct the invalid $\mathcal{S}^{\checkmark}$, valid $\mathcal{S}^{\times}$ and feasible $\mathcal{S}^f$ sets. Finally, to assess the penetration of a scenario and depending on the set in which the scenario belongs, the sets $\mathcal{A}^{\checkmark}$, $\mathcal{A}^{\times}$ and $\mathcal{A}^f$ are built using a function $\mathbf{g}$. Each of these three steps is highlighted in a different colour in Figure 3.7.

For the sake of conciseness and simplicity, two papers were selected to illustrate how the literature fits in the framework.

3.4.1 Paper 1: "Understanding Photovoltaic Hosting Capacity of Distribution Circuits" by Dubey *et al.* [51]

This paper aims at determining the HC for PVs and the impact of specific factors on the HC. The second part which deals with the impact of several factors will not be addressed in this review. The study first computes a base case which is the existing configuration and then stochastically assesses the HC by simulating several scenarios with different penetrations levels. The details of this part are given below.

The **Network model** N represents, in OpenDSS, an actual three-phase distribution network with 12.47kV supplied by a 24MVA substation transformer. The base case has 1.196 MW of existing PV. The schema of this network can be found in [51]. No further indications about the network were given.

All **customers** can have new installations and the **installation type** is PVs: $\mathcal{H} = \{\text{PV}\}$. The installation size is the only **option**. The size is randomly picked from two distributions represented by probability density functions (PDFs) depending on the type of customer: residential or commercial.

The study follows the paradigm developed by EPRI [52] for steady-state analysis of HC. This means that one time step was used for the **time period** ($\mathcal{T} = \{1\}$). This time step is derived from **exogenous data**. The study uses two sets of exogenous data: minimum and maximum load both obtained from yearly load demand measured in the substation of the network in 2013. Note that PV production is not exogenous data as the size is used as the production rather than time series.

The **considered scenarios** set contains 5000 scenarios. This set is built by increasing the penetration, here defined as the percentage of the total customers equipped with PV systems. The increase is done by steps of 2% for penetration going from 0% to 100% resulting in a total of $|\mathcal{S}^c| = 5000$. Location of PVs are randomly selected. Note that the considered uncertainties are both epistemic as they are the location and the size of the installations.

The **issues** considered are over-voltage (OV), voltage deviation (VD) and voltage unbalance (VU). Theses issues are explained and formulas are given in Section 3.5.2. In this study, the three issues are considered separately as they are not aggregated and thus leads to three HCs: one for OV, one for VD and one for VU. Also, as the study is steady-state, f is the same

as f_t. The thresholds used for each f are 5% more than the nominal voltage for OV, and 3% for both VD and VU. This study used three independent issues to highlight the impact of each for the hosting capacity.

Although penetrations are defined as percentage of the total customers equipped with PV systems for determining the considered scenarios, the penetration output by the function $\mathbf{g}$ and used for the hosting capacity is defined as the total additional PV size in kW.

They consider the *first violation* and *all scenarios with violation* HC as defined in Section 3.3. Additionally, they defined *50% scenarios with violation* HC which is the smallest penetration such that 50% of scenarios encounter an issue, i.e.:

$$\min_{\mathbf{a}} : |\{\mathbf{s} \mid \mathbf{g}(\mathbf{s}) = \mathbf{a}\} \subset \mathcal{S}^{\times}| = |\{\mathbf{s} \mid \mathbf{g}(\mathbf{s}) = \mathbf{a}\} \subset \mathcal{S}^{\checkmark}|. \tag{3.19}$$

The two sets of exogenous data E, the three limitation functions f, the safe penetration set and the three outputs (*first violation*, *50% scenarios with violation* and *all scenarios with violations*) lead to 18 HC values given in Table III of the paper and reproduced in Table 3.5.

Cases		Additional PV Size (kW)	
		Max Load	Min Load
Overvoltage	first violation	9,442	5,454
	50% scenarios with violation	9,578	5,536
	All scenarios with violation	9,659	5,722
Voltage deviations	first violation	1,756	2,776
	50% scenarios with violation	1,834	2,970
	All scenarios with violation	1,909	3,088
Voltage unbalance	first violation	0	5,760
	50% scenarios with violation	0	6,101
	All scenarios with violation	0	6,291

Table 3.5: PV hosting capacity from [51] (copied with authorization).

3.4.2 Paper 2: "Assessing the Potential of Network Reconfiguration to Improve Distributed Generation Hosting Capacity in Active Distribution Systems" by Capitanescu *et al.* [53]

This research explores how to increase DG hosting capacity using network reconfiguration. The study is conducted in two parts: the authors first determined the HC and then studied how to use reconfiguration to improve

it. In the context of the present chapter, only the first part is of interest and is detailed.

The method proposed is applied on a 34-bus 12.66-kV distribution system **network model**. The network has 37 lines, 34 nodes and one feeder, more information about the network can be found in [54]. In this network, a subset of 8 **customers** (G1,..., G8) are selected to have new installations. The choice of using a subset of customers is justified by (i) the fact that tools can be used to determine most suitable locations and; (ii) these locations might be enforced, for instance, by regulatory rules.

Although the methodology is given for general DG, the test case is made using wind turbines as the **installation type** thus, $\mathcal{H} = \{\text{WT}\}$. The considered **option** for the installation is the size. Two sizes, related to two different wind profiles (WP), are available and, for simplicity, the sizes are named following the profiles: $\mathcal{I}_{WT} = \{WP1, WP2\}$.

This study is time variant and the **time period** is composed of 146 time steps ($|\mathcal{T}| = 146$). The **exogenous data** accounted in these time steps are historic demand and wind data.

Table 3.6 gathers a summary of the three **considered scenarios** ($|\mathcal{S}^c| = 3$) of the study.

Scenarios in $\mathcal{S}^c$	G1	G2	G3	G4	G5	G6	G7	G8
A	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	WP2	$\emptyset$	$\emptyset$	WP1
B	WP1	WP1	WP2	WP2	$\emptyset$	$\emptyset$	WP1	WP1
C	WP1	WP1	WP2	WP2	WP2	WP2	WP1	WP1

Table 3.6: Considered scenarios ($A, B, C \in \mathcal{S}^c$) for the [53] for installing WT. The empty set $\emptyset$ means that no installation was added at that customer.

Both voltage and thermal **issues** are considered. The minimum and maximum voltage limits are respectively set to 0.95 p.u. and 1.05 p.u. at all nodes, and the thermal limit of all lines is set to 6.6 MVA. Both limits are further addressed in Section 3.5.2.

Finally, **penetrations** are defined as the sum of the nominal capacity of the DG and g returns the sum of the maximal nominal capacity. The HC for the three scenarios are 3.622MW, 4.161MW and 7.154MW for scenario A, for scenario B and for scenario C, respectively. This study is considered deterministic as no uncertainty is considered.

3.5 Review and application

This section aims at showing how related works on hosting capacity problems can be easily compared using the generic framework.

This section is divided into three parts corresponding to the three steps presented in Section 3.4 :

- First, the **system assumptions** are discussed in Subsection 3.5.1; the methods used to construct the scenarios sets are explored and a quick overview of the network and data each paper uses is done. How papers address uncertainties is also highlighted.

- Second, the **issues characterisation** (often referred as *limiting factors* in the literature [41], [55]); Subsection 3.5.2 identifies the common issues used in various research to populate the set $\mathcal{P}$ and how they trigger and aggregate them (i.e., their definition of the functions f and f_t).

- Finally, **the hosting capacity computation**. Subsection 3.5.3 gathers the most frequent types of penetrations ($\mathcal{A}$) and how they are computed ($\mathbf{g}$).

3.5.1 System assumptions

This subsection details system assumptions, which are the different hypothesises that are used to build the scenario sets $\mathcal{S}$ and $\mathcal{S}^c$. The construction of the system assumptions can be divided into three phases and the remainder of this section will address these:

- First, *network model* has to be constructed and *exogenous data* has to be gathered;

- Then, the *model scope* has to be defined. The scope determines the accepted values for each previously defined sets ($\mathcal{T}$, $\mathcal{C}$, $\mathcal{H}$ and $\mathcal{I}_{\mathcal{H}}$).

- Finally, *uncertainty* needs to be taken into account. Choices made to manage and model the uncertainty restrict both the scenarios $\mathcal{S}$ and exogenous data E spaces by methodology choices.

A summary of some assumptions can be found in Table 3.7. The different categories used in the table are explained below.

Papers	Data (N & E)	Model scope ($\mathcal{T}$, $\mathcal{H}$ and $\mathcal{I}_{\mathcal{H}}$)				Uncertainty ($\mathcal{S}$ & E)
	Empirical Data	Time variant	Technology	Different sizes	Single phase	Considered
[56]	○	●	HP	○	○	○
[53]	○	●	WT	●	●	○
[57]	●	○	EV	●	○	●
[58], [59]	○	○	PV	●	●	●
[60]	○	○	PV	●	?	●
[61]	○	○	PV	●	○	●
[62]	○	●	PV	○	○	●
[63]	○	●	PV	●	○	●
[64]	○	●	PV	●	◑	●
[65]	○	●	PV	●	●	●
[66]	○	●	PV	●	?	●
[67]	◑	●	PV	●	◑	●
[51]	●	○	PV	●	○	●
[68]	●	●	PV	●	◑	●
[69]	●	●	PV	●	?	○
[70]	●	●	PV	○	?	○
[71]	●	○	PV	○	○	●
[72]	●	○	PV	●	○	●
[73]	●	●	PV	●	●	●
[74]	●	●	PV	○	◑	○
[75]	●	●	PV	●	○	●
[76]	●	●	PV	●	○	
[77]	◑	●	PV, EV, HP, μCHP	●	●	●
[78]	◑	●	BESS, [PV]	●	○	○
[55]	●	◑	PV, [EV]	◑	●	●
[21]	◑	●	PV, EV	○	●	●
[79]	○	●	PV, EV	◑	●	●

Table 3.7: References characteristics. Legend: ● means the characteristic is fully, ◑ partially, ○ not implemented and ? not specified. Note that "Single phase" is ticked if all new installations are connected using only one-phase connection.

Network and exogenous data: N and E

The network model N is a representation of a network with a variable degree of sophistication depending on the work and the available data. It

can be made from real-world, empirical data or from synthetic data (for example the IEEE 33-bus or 123-bus feeders [58], [62], [79]), or from a mix of both (see, e.g., [21], [56], [79]). This model also encompasses hypotheses made on the network or on its simulation, e.g., the complexity and details of the power flow computation. Exogenous data E are mainly historical data or probabilistic approximations that are used as an estimate for the future.

Model scope: $\mathcal{T}$, $\mathcal{C}$, $\mathcal{H}$ and $\mathcal{I}_{\mathcal{H}}$

The model scope defines the values that populates each of the sets related to the system assumptions, i.e. the set of time steps $\mathcal{T}$, the set of connection nodes for technologies $\mathcal{C}$, the set of types of technologies $\mathcal{H}$ and the set of technologies' options $\mathcal{I}_{\mathcal{H}}$.

Time invariant vs time variant: $\mathcal{T}$

Electrical systems varies through time. The advantage of using time series, and thus time variant models, is their closeness to reality. Indeed, the model is able to reflect the time related phenomena that occurs in real networks. Nevertheless, time representation is highly dependent on the previous data assumptions and data availability. The model can thus either be time invariant or time variant. Time-invariant modelisations only use one time step. These studies often choose the worst-case time step and are referred to as conservative. Methods using time series varies from one another on:

- The granularity of the time series (i.e., the amount of time between two time steps);
- The considered period. For instance, some papers consider multiple full days, while others consider consecutive hours in a single reference day. Some consider a range of hours in multiple days (e.g., Monday to Friday from 09:00 to 18:00).

Connection nodes: $\mathcal{C}$

Studies can either consider all customer nodes (e.g., [21], [51], [55], [65], [79]) or only a subset of customers (e.g., [56], [58], [62], [66], [68]) as connection nodes.

Technology type: $\mathcal{H}$

As the HC definition is not specific to one technology, it is worth mentioning which paper is studying which technology (i.e., set of types of technologies $\mathcal{H}$). The most common technology is DG and more precisely PVs

as they represent the first technology to have been used to define the hosting capacity. Note that some studies develop methods to compute HC that are not specific on a technology type and then test the methodology on one or more technology types, often independently (e.g., [77]). In [77], these technologies are referred to as Low Carbon Technologies (LCT). They consider micro combined heat and power units (μCHP) as generator.

Technology options: $\mathcal{I}_{\mathcal{H}}$

Options varies from one technology type to another. The two options encountered in the reviewed papers are the size of the technology and the connection type. For the size, studies use one size or allow the flexibility of different sizes. Note that some studies use several sizes but only one size is available per run, e.g., they first attempt a run with $\mathcal{I}_{PV} = \{\emptyset,\ 1\ \text{kWc}\}$ and then another $\mathcal{I}_{PV} = \{\emptyset,\ 2\ \text{kWc}\}$. The connection type can be in one, two or three phases, depending on the network, the available data, the authors' choice and the technology itself. Some studies, such as [68], enforce that if the customer is on a single phase the installation has to also be on a single phase.

Uncertainties

Studies may or may not take uncertainties into account. For the one that does, as previously mentioned, the literature identifies two categories of uncertainty: aleatory and epistemic [46]. The aleatory uncertainties deal with exogenous data (in E) such as the technologies' consumption or production and the customers' consumption. These are inherently stochastic. These are, in practice, sampled when computing f and g.

Epistemic uncertainties deal with the lack of knowledge for, for instance, the type $\mathcal{H}$ and options of the technologies $\mathcal{I}_{\mathcal{H}}$ or their locations $\mathcal{C}$. These uncertainties are considered by evaluating several scenarios $\mathcal{S}^c$.

In both cases, these uncertainties can be accounted for by using, for instance, simple Monte Carlo simulations or more complex statistical methods. These uncertainties are not mutually exclusives: studies can use both. Note that on the other hand, studies neglecting all uncertainties, referred to as deterministic, are restricting both $\mathcal{S}$ and E with defined size and values. In recent papers, there is less and less deterministic methods, as the underlying system is, in reality, stochastic.

Table 3.8 showcase how studies account for uncertainties. Deterministic methods are omitted from the table for conciseness.

Papers	Aleatory		Epistemic		
	Customer	Installation	Installation		
	Loads	Loads or Production	Location	Size	Type
[21], [57], [59], [68], [71]	○	○	●	○	○
[51], [60], [61], [63], [72], [73], [76]	○	○	●	●	○
[55]	○	◑	●	○	○
[58]	○	●	●	●	○
[67]	●	○	●	○	○
[75]	●	○	●	●	○
[62], [79]	●	●	○	○	○
[64]–[66], [77]	●	●	●	●	○

Table 3.8: Uncertainties types for reviewed papers that take into account uncertainties. Legend: ● means the characteristic is fully, ◑ partially, ○ not implemented. Partially means that not all simulations account for that uncertainty.

3.5.2 Issues characterisation

This section discusses the elements inside the set $\mathcal{P}$ and how they are evaluated using the function f, for the papers already analysed above. As a reminder f identifies whether any issues from $\mathcal{P}$ occurs in the network. In some papers, the function f is sometimes not precisely defined or too complex to display here; therefore the focus is put on some readable examples to emphasise the genericity of the formalisation. Motivated readers are encouraged to read the original papers to gather the full definitions of f and $\mathcal{P}$.

There are four main categories of issue [80]: voltage dependant, load dependant, protection and harmonics. These categories are not specific to DER. Papers such as [41], [43], [55], [80], [81] attempt to define all existing issues.

The following subsections are structured using these categories. Each of these subsections showcases how the issues associated with the considered category fit in the definition presented in Section 3.3. Note that there

is no intent to be exhaustive in the list of limiting factors, only the ones encountered in the papers reviewed are mentioned.

Voltage-related issues

Voltage issues are typically defined as voltage unbalance, voltage levels (under- or over-voltage) or deviations:

- **Voltage unbalance** refers to an uneven distribution of voltage magnitudes or phase angles in a three-phase electrical system. It occurs when the three phases of the system have different voltage levels or when the phase angles between the voltages are not equal. This translates Eq (3.6) into:

  $$f_t(s) = \begin{cases} 1, & \text{if } \exists m \in \mathcal{M} : |V_{m,t,\phi_1} - V_{m,t,\phi_2}| > \beta, \forall \phi_1 \neq \phi_2; \phi_1, \phi_2 \in \Phi; \\ 0, & \text{otherwise.} \end{cases}$$

  where $\mathcal{M}$ is the set of monitored elements, V is the complex voltage of an element of $\mathcal{M}$ for the phase ϕ, Φ is the set of phases, β is a given threshold. Note that the difference can be computed taking into account only magnitudes, or the angles, or both.

  In several cases the hosting capacity could be improved by balancing the network [22], [23].

- **Voltage level** issues happen due to an unusual use of the network. In the case of PVs, or any other DG device, when there is greater production compared to consumption, the remaining production is injected in the network. This process is called **reverse power flow** as the power flows in the opposite direction compared to the direction it was originally intended (i.e., from the centralised generation to the distribution as opposed to from the prosumers DG to other consumers). This can cause an over-voltage. In the case of EVs, or any other bigger electricity consumer than first intended as normal network use, the consumption is bigger than that planned. And the consumers, when peak consumption occurs, withdraw too much electricity leading to an under-voltage as the production and the capacity of the cables were not made for such peak consumption. This translates into:

  $$f_t(s) = \begin{cases} 1, & \text{if } \exists m \in \mathcal{M} : \underline{\nu} > V_{m,t} \vee V_{m,t} > \overline{\nu}; \\ 0, & \text{otherwise.} \end{cases}$$

  where $\mathcal{M}$ is the set of monitored elements, V is the voltage of a node and $\underline{\nu}$ and $\overline{\nu}$ are given lower and upper bound thresholds. This

definition is a base that often varies from one paper to another. For instance, the measured element can vary between studies (e.g., $V_{m,t}$ can be $\max_m V_{m,t}$ as in [71]) and some studies may have different thresholds when they have more than one type of measured element (e.g., [68]).

- **Voltage deviation** is the deviation in voltage from the initial network (i.e., with no installation) to a full-penetration network.

$$f_t(s) = \begin{cases} 1, & \text{if } \exists m \in \mathcal{M} : |V^b_{m,t} - V_{m,t}| > \mu; \\ 0, & \text{otherwise.} \end{cases}$$

where $\mathcal{M}$ is the set of monitored elements, V is the voltage of a node and V^b is the voltage of the same node in the initial network, and μ is a given threshold.

Note that the first two issues exist without new installations while the last one is exclusively linked to the addition of new ones. Most researches use standards to define the thresholds (β, ν, μ) for issue detection f. Several standards were used to limit these thresholds, for instance EN-50160 is a European standard, ANSI is an American standard, and VDE-AR-N is a German standard. These standards might have some variations such as BS EN-50160 which is the British version of EN-50160. Table 3.9 gathers the papers by issues and standard. Some papers do not clearly specify the standard, but the closer fit is assumed, and some do not explicitly use these standards, for instance [57] uses a threshold set by the Brazilian government for distribution.

Load-related issues

Load-related issues occur when transformers or cables experience thermal overload, indicating excessive current flow that generates heat, potentially causing damage or failure. Here is an example on how to define f to detect overload by limiting the current:

$$f_t(s) = \begin{cases} 1, & \text{if } \exists m \in \mathcal{M} : I_{m,t} > \alpha \times I^l_m; \\ 0, & \text{otherwise.} \end{cases}$$

where $\mathcal{M}$ is the set of measured elements, I is the current of an element and I^l is the rated capacity of the measured element, and α is a given threshold.

Some studies (e.g., [21], [56], [57], [61], [62]) limit the power which is equivalent to limiting the current under constant voltage. To detect the overload, most papers used two categories of measured element: the

Papers	Unbalance	Levels	Deviation	Standard
[21], [61], [62], [64], [65], [71], [72], [75]	○	●	○	ANSI
[51], [59], [63], [73]	●	●	●	ANSI
[70]	○	○	●	EN50160
[60], [76]	○	●	●	ANSI
[58], [68]	●	●	○	ANSI
[67], [74]	●	●	○	EN50160
[55]*, [79], [56]*, [53], [77]	○	●	○	EN50160
[69]	○	●	○	VDE-AR-N 4105
[78]	●	●	●	Australian Act 1945
[57]	●	●	○	Other
[66]	○	●	○	Other

Table 3.9: List of voltage-related issues by papers. ● means the issue is taken into account, while ○ means it is not. The asterisk * means that the standard was not properly mentioned but corresponds to the given one.

conductor current as in [56], [57], [61], [62], [68], [75] and the transformer load [21], [57], [61], [62], [75]. The threshold varies from one study to another (from 50% [57] to 187% [68]) and several thresholds can be used in the same study for different types of measured element (e.g.[75]). These thresholds are ratios of a rated capacity, but some papers used defined values (e.g., [61], [62]). Also previously defined standards can be used for overload such as EN50160 in [67], [74].

Protection-related issues

Protection-related issues are a vast category. They were defined in [80] as issues taking more time to affect the user level in contrast to voltage and load issues even though both also have direct consequences for the end user.

A first protection issue is rapid voltage magnitude variations [55]. These occur because of variations in production or consumption over a period of less than one minute. These rapid changes can cause flicker. An example of a standard that limits these is IEC 61000-4-30. Standards to limit short-

term (Pst) and long-term(Plt) flicker severity are defined in IEC 61000-4-15 [55].

Another protection issue is used in [58] which restricts the PV power factor, i.e. the ratio of real power (in watts) to apparent power (in volt-amperes), between 0.8 and 1 to contribute to the stability of the power grid, stability in the sense of the ability to maintain a balanced and reliable supply of electricity.

Harmonics-related issues

Harmonics issues occur when the current injections are not sinusoidal. Several standard indices exist [55] for individual harmonics, interharmonics, and for total harmonic distortions (THD) (e.g., IEEE 519, IEC 61000-4-7, IEC 61000-4-30, IEC/TR 61000-3-6).

In [55], superharmonics are mentioned: these are waveform distortion in the frequency range between 2 kHz and 150 kHz and are injected by an increasing number of devices connected to the grid. Unfortunately, [55] considers that using this issue as a limiting factor for the hosting capacity is not feasible as no limits for distortion in this frequency range are set. Furthermore, the main barrier for taking into account the harmonics as a limiting factor is the lack of appropriate calculation models, especially when considering low-voltage and medium-voltage networks [55]. Thus, harmonics are not often used as limiting factors for the hosting capacity.

Issues aggregation

Some studies may consider several issues. They may consider them independently and thus have several definitions of the f function giving a hosting capacity per issue; or they may aggregate them to have one f function. The aggregation of issues can be a complicated formula but most aggregations are a simple OR operator over several issues. Furthermore, time aggregation of f_t is necessary for time-variant studies such as in Eq (3.7). Another example of time aggregation is an average during a time period (e.g., [70], [75]). Most studies do not explicitly give how they aggregate the time and the issues.

Issues summary

In Table 3.10 is a summary of all papers and the issues categories. Voltage deviation issues are the most limiting factor [43]. Therefore, these issues are addressed in the majority of studies as shown in Table 3.10. Both Protection and Harmonics issues are less used and mostly not as the only limitation.

Papers	Voltage	Load	Protection	Harmonics
[51], [59], [60], [63]–[65], [70]–[73], [76], [79]	●	○	○	○
[21], [55], [56], [61], [62], [67], [68], [74], [75], [53], [66], [69], [77], [78]	●	●	○	○
[57]	●	●	○	●
[58]	●	●	●	○

Table 3.10: List of issues by papers. ● means the issue is taken into account, while ○ means it is not.

Most hosting capacity studies are considered as static as they do not take into account the duration of the issues. Taking into account the duration means that the aggregation f evaluates several consecutive time steps f_t. For instance, with a duration of i:

$$f(\mathbf{s}) = \begin{cases} 1, & \text{if} \quad \exists t \in \{1, ..., |\mathcal{T}| - i\} : \\ & f_t(\mathbf{s}) = 1 \wedge f_{t+1}(\mathbf{s}) = 1 \wedge ... \wedge f_{t+i}(\mathbf{s}) = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.20}$$

This means that f reports an issue only if it occurs for i consecutive time steps. Studies that take into account the duration of the issues are referred to as dynamic hosting capacity (DHC) such as in [63], [68], [70], [75], [77].

3.5.3 Hosting capacity computation

This section focuses on the final part of determining the hosting capacity: penetration calculation. It is how the $\mathbf{g}$ function is defined for different studies and the output/penetration format choice ($\mathcal{A}$).

As the penetration is not a defined concept shared among all the reviewed papers, to compute the hosting capacity one has to first define the wanted output of $\mathbf{g}$, i.e., how to quantify the set of penetrations $\mathcal{A}$. This penetration can either be an absolute quantity that reflects the amount of new technologies that are installed or a ratio between this absolute quantity

and a reference. The encountered possibilities for the absolute quantity are, with their abbreviated name given in bold to aid readability:

- The number of customers with a new installation, abbreviated as **NC**;
- The total production or consumption of a new installation, **TP**;
- The maximal capacity of a new installation, **MC**.

The encountered reference quantities are, again with their abbreviated name:

- The total number of customers that can accommodate new installations, **TNC**;
- The total consumption of customers, **TC**;
- The maximal consumption (load) of customers also referred as peak load, **PL**;
- The total capacity that is accepted by the transformer, also referred as transformer rated capacity and thus abbreviated to **TRC**.

Each absolute and reference quantity has is perks depending on the study. For instance, choosing the number of houses equipped with the technology, as an absolute quantity, is intuitive; but from one house to another the loads vary and thus this quantity can lack precision. Also, since peak demand can put significant stress on the grid, evaluating the hosting capacity based on maximum capacity, as a reference quantity, ensures that the grid can handle high-energy demand without issues. Using the maximum capacity allows for unified approach for evaluating hosting capacity, regardless of the specific technology being added. Neither does it require detailed modelling of various DER operational conditions and thus simplifies the planning process. Despite these advantages, it might not fully reflect the actual average or expected output of the DER installation under typical operating conditions.

Table 3.11 regroups all reviewed papers and organises them following their HC computation choices. This table also shows the different output types for each paper. These can be divided in four types:

- The set of penetration $\mathcal{A}^f$ or $\mathcal{A}^s$ as a *range*, for instance the penetration is a number of customers, HC $= \mathcal{A}^f = [0-5;7]$ meaning that all considered scenarios with penetration in this set are feasible;
- The set of penetration $\mathcal{A}^f$ or $\mathcal{A}^s$ as a range with probability density function (PDF). This means having the scenario distribution;

Papers	Absolute quantities	Reference quantities	Units	Types
[55]	NC	-	Customers	PDF
[57]	NC	-	Customers	SV
[21], [77]	NC	TNC	- (%)	PDF
[56]	NC	TNC	- (%)	OV
[74]	NC	TNC	- (%)	SV
[64]	TP	-	Power	Range, SV
[51], [65]	TP	-	Power	SV
[79]	TP	TC	- (%)	PDF
[71]	TP	PL	- (%)	Range
[58]	TP	PL	- (%)	SV
[61], [73]	MC	-	Power	Range
[66]	MC	-	Power	PDF
[62], [76]	MC	-	Power	OV
[53], [69], [70]	MC	-	Power	SV
[78]	MC	-	Energy	SV
[67]	MC	PL	- (%)	SV, Range
[59], [60]	MC	PL	- (%)	SV
[68]	MC	TRC	- (%)	OV
[63]	MC	-, PL	Power, - (%)	Range
[75]	NC, MC	TNC,-	Customers, Power	SV
[72]	NC, MC	TNC,-	Customers, Power	OV

Table 3.11: Summary of papers' definition of hosting capacity computation. Several acronyms are used for the readability of this table: NC for the number of customers, TP means total installation production or consumption, MC for the maximal capacity of a new installation, TNC for the total number of customers that can accommodate new installations, TC for total consumption of customers, PL for peak load, TRC for transformer-rated capacity, OV for one value, SV for several values and PDF for probability density function. One row reads as follows: Paper [55] uses the number of customers with new installations (NC) as the absolute metrics to gauge the penetration. This value is not compared to a reference, hence the "-", thus the hosting capacity is expressed as the number of customers as a unit. They compute the set of penetrations and choose to represent the distribution of scenarios over the penetrations using a PDF.

- A penetration of this set corresponding to one scenario, i.e., *one value*. This would happen for instance in a deterministic study having an iterative workflow (Figure 3.7b) or considering one scenario;
- Several penetrations of this set corresponding to the output of several scenarios, i.e., *several values*.

Note that studies might consider the same scenario with different E and thus still be considered as outputting *one value* (e.g., [56]).

As mentioned in Section 3.3, papers computing the one-dimensional HC and with output type *one value*, choose stricter limits for the hosting capacity: all dominated penetrations are feasible ($\mathcal{A}^s$) and output the *minimal hosting capacity*. Some of these papers also choose to output *maximal hosting capacity* such as in [51], [63], [65], [73]. On the other hand, in any dimension, papers that compute either the set $\mathcal{A}^f$ or $\mathcal{A}^s$, use PDFs to showcase the distribution of scenarios over the penetrations e.g., in [79] with two dimensions.

3.6 Intermediate conclusion

The need to determine network hosting capacity is now widely acknowledged, and numerous researchers have worked on the subject these last years. However, there was, prior to this work, no well-defined formalism.

This chapter introduces a general definition of the hosting capacity problem along with definitions of all the elements necessary for its formulation. First, the deterministic definition has been presented, followed by the exploration of various methods for incorporating uncertainties, thus stochasticity.

The definition is then applied to a small fictitious example and then to two prominent papers from the hosting capacity research domain. These latter papers serve to illustrate the use of the definition in concrete and more complex cases while the small example ensure that the different aspect of the definition are commonly understood.

The central focus of the definition is to establish a common ground for concepts; thus, it does not prescribe how to construct different sets or compute hosting capacity, recognising the diversity in approaches. Nevertheless, the related work section provides several examples from the literature on how these tasks are accomplished. This restricted but systematic review of hosting capacity field demonstrates the ease with which related works can be conducted to compare studies using the presented framework.

Future works stand to benefit from this formalism to clarify terminology for subsequent research and identify with more clarity the gaps in the literature that need addressing.

Chapter 4

Individual hosting capacity assessment

In the end we retain from our studies only that which we practically apply.

Johann Wolfgang Von Goethe

The increase of photovoltaic panels and electric vehicles in low-voltage distribution systems leads to over-voltage, under-voltage, and congestion issues. These issues, related to new installations, add a considerable cost to distribution system operators, and therefore to customers. Distribution system operators want to limit these costs by determining the impact of photovoltaic panel production and electrical vehicle charging. This chapter presents a probabilistic method enabling operators to evaluate the network capacity, defined as the number of new installations that can be added to the network without adapting it to overcome under- and over-voltage. The method provides multiple probabilistic performance indicators reflecting a large number of possible configurations resulting from new installations added to the low-voltage network. The evaluation of this method is done using a case study based on an existing European network. The method provides tangible results as the maximum number of photovoltaic installations or electric vehicle chargers within a defined confidence level. Results, in the test case, show that the individual capacity of the network is evaluated as a 45% penetration rate for photovoltaic installations, or 4% for electric vehicle chargers, with a 5% violation of operational indicators.

4.1 Notations

Sets	
$\mathcal{T}$	Set of time steps
$\mathcal{N}$	Set of nodes
$\mathcal{E}$	Set of edges
$\mathcal{P}$	Set of phases
$\mathcal{I}$	Set of indicators
$\mathcal{S}_{k,n}$	Set of power injections in configuration k at node n
$\mathcal{C}_r$	Set of configurations for a penetration rate r
$\mathcal{K}$	Set of considered configurations for a penetration rate r, $\mathcal{K} \subseteq \mathcal{C}_r$ with $\lvert\mathcal{K}\rvert \leq K$
$\mathcal{A}_n$	Set of new possible installations type for node n
$\mathcal{M}_{k,n}$	Set of installed types for node n in configuration k
Variables	
Δt	Interval between consecutive time steps
r	Penetration rate, number of installations in the network
r_i	Maximum penetration rate tolerated for indicator i
K	Number of configurations
R	Hosting capacity
$\overline{h}_i$	Threshold rate for indicator i
$\overline{F}_i$	Risk tolerance for indicator i
$\overline{V}$	Over-voltage threshold
$\underline{V}$	Under-voltage threshold
$\mathbf{Z}_e$	Impedance of edge $e \in \mathcal{E}$
$\mathbf{S}^b_{n,p}$	Initial base power injection of $n \in \mathcal{N}$ in $p \in \mathcal{P}$
$\mathbf{S}_n$	Power injection of $n \in \mathcal{N}$
$\mathbf{S}_{k,n}$	Power injection at node $n \in \mathcal{N}$ for configuration k
$\mathbf{S}_{k,n,m}$	Power injection of installations of type $m \in \mathcal{M}$ at $n \in \mathcal{N}$ for configuration k
$\mathbf{V}_n$	Voltage of $n \in \mathcal{N}$
$\mathbf{V}_k$	Voltages of all nodes in configuration k
$\mathbf{V}_{k,n}$	Voltage of $n \in \mathcal{N}$ in configuration k
$\mathbf{I}_e$	Edge current of $e \in \mathcal{E}$
$\mathbf{I}_k$	Edge currents for all edges of configuration k
$\mathbf{I}_{k,e}$	Edge current of $e \in \mathcal{E}$ in configuration k

4.2 Introduction

The European Union directives [82] - [83] encourage increasing penetration rates for both Electric Vehicles (EVs) and Photovoltaic (PV) installations. However, this has become a concern for Distribution System Operators (DSOs). Indeed, adding new installations of these types to the Low-Voltage (LV) network leads to several issues such as over-voltages or under-voltages. To prevent these issues, and to plan future corrective actions, DSOs need to quantify the impact of such installations on their LV networks by evaluating the potential network capacity given likely future new installations in the network. The problem that complicates this quantification is that the decision to install these new devices is not made by DSOs but by the customers themselves according to their financial means and needs. For DSOs, this translates into uncertainty in the type of added installations, their capacity, and the order in which they will be connected to its network. These uncertainties turn the network capacity assessment into a complicated, but equally important problem.

One common method to evaluate this capacity is to check if there exists a possibility of under- or over-voltage, with a power flow for a given set of fixed installations. If it is the case, DSOs consider this set of PV and EV installations as unacceptable when trying to prevent network issues and avoid customers complaints. This conclusion is, however, too conservative with respect to what happens in practice. For instance, when a PV installation produces and the PV inverter detects an over-voltage, it temporarily disconnects the installation from the network. This action of the inverter prevents any significant over-voltage from occurring in the network. Therefore, checking the possibility of over-voltage for a fixed set of installations is not an appropriate method to quantify the network capacity. The real impact for customers and DSOs is that energy not produced due to inverter curtailment resulting from network issues. The same reasoning is applied to EVs as when an under-voltage is detected, charging of EVs is interrupted.

This chapter extends the network capacity assessment by (i) determining the number of new PV or EV installations that can be added to the network; (ii) considering the uncertainty related to the position of new installations; (iii) considering time variance by using time-series and (iv) quantifying the capacity in terms of energy. The energy quantification are stochastic functions of the set of installations obtained using Probability Density Functions (PDF).

The rest of the chapter is presented as follows. Related work is presented in Section 4.3. Section 4.4 presents the network capacity assessment problem. Section 4.5 explains a stochastic approach to solve this problem. Section 4.6 presents the results obtained on two applications. Finally,

Section 4.7 concludes the chapter and presents some discussions.

4.3 Literature review

The hosting capacity assessment problem is a stochastic problem as it involves uncertain parameters. These uncertainties are categorised into aleatory and epistemic [46]. The former refers to power consumption and injection that are unknown variables. The epistemic uncertainties relate to the location or size of future installations in the network.

Several researches, reviewed in [43], [46], study the low-voltage network capacity assessment problem. All reviewed methods rely on power-flow analyses and can be divided into three categories [46]: (i) time-invariant deterministic, (ii) time-invariant stochastic, (iii) time-variant deterministic. Time-invariant deterministic methods use one typical value for each node as its injection or consumption representing the worst-case scenario. Such consideration leads to a pessimistic assessment that does not represent the real network operation under time-variant production and consumption patterns. Time-invariant deterministic methods discard aleatory uncertainties. Time-invariant stochastic methods capture the stochastic nature of the hosting capacity assessment problem by considering both types of uncertainties. Epistemic uncertainties are only included in some of the stochastic techniques. Deterministic time-series capacity assessment methods ensure the correlation between injections and consumptions are satisfied and entail modelling aleatory uncertainties.

Paper [55] presents a method that addresses epistemic uncertainties for the capacity assessment of EV and PV installations. Epistemic uncertainties are handled with Monte-Carlo simulations, randomly selecting installation sites. To assess the capacity of PV installation, the lowest consumption of the year is taken as a reference consumption and the maximum total PV production is selected for the installations. For EV assessment, the highest consumption values and various nominal charging powers are considered.

The difference in results between [55], [84] and [85] highlights the importance of considering correlated injection/withdrawal time-series. The authors of [85] conclude on the importance of considering high-resolution time-series, but this can only be done at the cost of a shorter study time window that does not truly enable capturing and quantifying the impacts of uncontrolled EV charging.

Most of the papers reviewed in [43], [46] conclude that the voltage issues are the most important to consider as they occur before any other network issues. Indicators based on these are privileged to quantify the hosting capacity limit. Several papers reviewed in [43] consider certain technical

enhancements to increase this limit.

In the scientific papers reviewed, the capacity assessment problem was not addressed with both time-variant and stochastic methodology. This chapter presents a methodology that uses time-series to model the aleatory uncertainty and stochastic capacity assessment to consider epistemic uncertainty.

4.4 Problem Statement

The notations defined in Section 2.4 are used in this chapter. This highlights that the presented methodology is applicable on both an estimated topology or a real topology.

This study considers an unbalanced three-phase LV network. The network topology is represented by a tree graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ where $\mathcal{N}$ is the set of nodes in the network and $\mathcal{E}$ is the set of edges linking these nodes. The observation period $\mathcal{T}$ of the network is the set of all consecutive time steps denoted t. A variable x followed by a subscript t, that is x_t, refers to the value of x at time t while boldface $\mathbf{x}$ refers to the entire time-series. The set $\mathcal{P}$ denotes the set of phases.

Let $\mathcal{A}_n$ be the set of new possible installation types, e.g. photovoltaic panels (PV) or electric vehicles (EV), at a node n of the network. A specific configuration k of new installations can be defined by selecting a set of installed types $\mathcal{M}_{k,n} \subset \mathcal{A}_n$ and their corresponding power injections $\mathbf{S}_{k,n,m}$, for each node n, and each installation m. The set of power injections in configuration k at node n is:

$$\mathcal{S}_{k,n} = \{\mathbf{S}_{k,n,m} | m \in \mathcal{M}_{k,n}\}. \tag{4.1}$$

These power injections $\mathbf{S}_{k,n,m} \in \mathbb{C}^{3|\mathcal{T}|}$, add up to the base power injections of the nodes:

$$\mathbf{S}_{k,n} = \mathbf{S}_n^b + \sum_{\mathbf{s} \in \mathcal{S}_{k,n}} \mathbf{s}. \tag{4.2}$$

The injected powers of the new installations change the network voltages and currents. By solving the power flow equations G for the network with new power injections, the updated values of network variables can be obtained:

$$\{\mathbf{V}_{k,n} | n \in \mathcal{N}\}, \{\mathbf{I}_{k,e} | e \in \mathcal{E}\} = G_{\mathcal{G}}(\{\mathbf{S}_{k,n} | n \in \mathcal{N}\}). \tag{4.3}$$

A set of indicators $\mathcal{I}$, such as the maximum loading of the main MV/LV substation, evaluates each configuration of installations to identify unacceptable ones. This process can be performed with any set of indicators representing the network safety, reliability or other operational criteria.

Each indicator $i \in \mathcal{I}$ is a function of the voltages for each node in $\mathcal{N}$ and currents of each edge of $\mathcal{E}$ both over each time step in $\mathcal{T}$:

$$h_i(\mathbf{V}, \mathbf{I}) : \mathbb{C}^{|\mathcal{N}||\mathcal{T}|} \times \mathbb{C}^{|\mathcal{E}||\mathcal{T}|} \to \mathbb{R}. \tag{4.4}$$

The DSO can define a threshold rate $\overline{h}_i$ over which a configuration i is not acceptable.

Defining these thresholds $\{\overline{h}_i | i \in \mathcal{I}\}$ enables to determine the number of installations the network can support. For a fixed number of installations r to distribute across the number of nodes $\#\mathcal{N}$, $\frac{\#\mathcal{N}!}{(\#\mathcal{N}-r)!\cdot r!}$ combinations exists. The set of configurations $\mathcal{C}_r$ contains all possible combinations with r installations. Let $f_{i,r} : \mathbb{R} \to \mathbb{R}^+$ be a probability density function (PDF) of h_i over the configurations within the set $\mathcal{C}_r$. Values of indicator i, obtained for the set $\mathcal{C}_r$, are the support of the PDF. The distribution is assumed to be uniform.

The maximum tolerated number of installations r, i.e., network hosting capacity, depends on the accepted level of risk for the DSO. The tolerated number of installations r_i for an indicator i with the level of risk defined as $\overline{F}_i \in [0, 1]$ is as follows:

$$\int_{\overline{h}_i}^{+\infty} f_{i,r_i}(h_i) dh_i \leq \overline{F}_i \tag{4.5}$$

where the integral of the indicator i PDF, from the maximum threshold rate $\overline{h}_i$ to $+\infty$, i.e., the black area in Figure 4.1, gives the probability that h_i violates its threshold.

Having a set of several indicators $\mathcal{I}$, the hosting capacity of the network R is the minimum of r_i values obtained:

$$R = \min\{r_i | i \in \mathcal{I}\}. \tag{4.6}$$

R presents the network hosting capacity for new installations, without any indicator violating the defined threshold $\overline{h}_i$, with a probability equal to or less than the risk tolerance $\overline{F}_i$.

4.5 Hosting capacity assessment algorithm

The methodology takes as inputs the identified network graph $\mathcal{G}$ of Chapter 2 and its impedances $\mathbf{Z}_e$ $\forall e \in \mathcal{E}$, the forecasted values for initial base power injection time series $\mathbf{S}^b_{n,p}$ $\forall (n, p) \in \mathcal{N} \times \mathcal{P}$ and sets $\mathcal{A}_n$ $\forall n \in \mathcal{N}$ of new considered installation types, for each network node n. The maximum number of configurations is limited to $K \in \mathbb{R}$ for computational tractability.

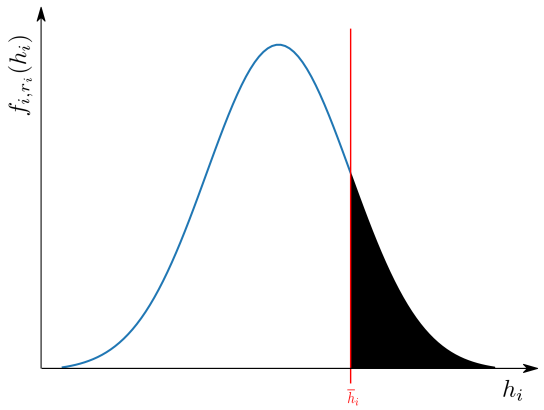


Figure 4.1: PDF of an indicator i for a given penetration rate r_i. The black filled area represents the probability that a configuration violates the threshold.

The algorithm pseudo code is presented in Alg. 1. For each number of new installations r, K configurations are generated. For each configuration, a power flow is performed to find the network currents and voltages, which are used to compute indicators for each configuration. Line 6 computes the ratio of generated configurations having indicators values higher than their accepted limit. These ratios are the probabilities defined in Eq. (4.5) for a configuration which has indicators violating their limit. The probabilities are compared to the corresponding risk tolerance of their indicator $\overline{F}_i$, the first indicator ratio to exceed its limit has the smallest r and is set as the total capacity of the network R [21].

Three indicators are considered. The indicator α denotes the power flowing through the main MV/LV substation transformer, to assess the total impact of installations on the network.

$$h_{\alpha,k} = \max_{(p,t)\in\mathcal{P}\times T} |S_{k,0,p,t}| \tag{4.7}$$

where 0 represents the network main substation node.

The other two indicators evaluate the impact of installation curtailment. Installations such as PV panels, when their connection node voltage exceeds a threshold $\overline{V}$, temporarily disconnect from the network. For a configuration k, the indicator β, represents the energy spilled and is defined as follows:

$$h_{\beta,k} = \sum_{(n,p,t)\in\mathcal{N}\times\mathcal{P}\times T:V_{k,n,p,t}>\overline{V}} |S_{k,n,p,t}| \cdot \Delta t \tag{4.8}$$

Require: K, $G_{\mathcal{G}}(\mathbf{S})$, $\{\mathcal{A}_n | n \in \mathcal{N}\}$
$r \leftarrow 0$
while $r < \sum_{n \in \mathcal{N}} |\mathcal{A}_n|$ **do**
  $\mathcal{K} \leftarrow$ sample K configurations with r installations
  $\{\mathbf{V}_k, \mathbf{I}_k | k \in \mathcal{K}\} \leftarrow \{G_{\mathcal{G}}(\mathbf{S}_k) | k \in \mathcal{K}\}$
  for each $i \in \mathcal{I}$ **do**
    $F_{i,r} \leftarrow \sum_{k \in \mathcal{K}: h_i(\mathbf{V}_k, \mathbf{I}_k) > \overline{h}_i} 1 \ / \ |\mathcal{K}|$
    if $F_{i,r} > \overline{F}_i$ **then**
      $R \leftarrow r - 1$
      return R
    end if
  end for
end while
$R \leftarrow \sum_{n \in \mathcal{N}} |\mathcal{A}_n|$
return R

Algorithm 1: Network capacity assessment algorithm [21].

where Δt is the interval between consecutive time steps in hours.

This indicator represents the sum of the energy spilled over the study interval, due to PVs disconnection because of the over-voltage limit.

When there is an under-voltage detected in the EVs' connection point, the charging might be interrupted. Therefore, for this type of installation, an indicator γ is defined to quantify the energy not served due to an under-voltage:

$$h_{\gamma,k} = \sum_{n,p,t \in \mathcal{N} \times \mathcal{P} \times T : V_{k,n,p,t} < \underline{V}} |S_{k,n,p,t}| \cdot \Delta t \tag{4.9}$$

This indicator represents the sum of the energy not served to charge EVs over the study interval, due to the interruption of their charging, due to under-voltage.

4.6 Applications

This section presents the results for hosting capacity assessment applied on the network model retrieved from Chapter 2. To reduce the factorial growth of the number of considered configurations, K is limited for each penetration rate. The value of K is set to 500 for PVs and 350 for EVs. This limit is chosen according to a trade-off to analyse as many configurations as possible, and still be computationally feasible. For each config-

uration, the number of installations depends on the penetration rate and the installations are randomly positioned. A load flow is performed for each configuration over a period of one week with a 30-minute time step to compute KPIs.

4.6.1 Capacity assessment for photovoltaic panels

As the network is based on an existing Belgian one, the PV panel energy production was extracted from a Belgian provider [11]. The number of panels per installation is set to 13 to fulfil the average household energy consumption in Belgium as described in [86]. To be consistent, the wattage peak of each panel is 290W, which is the common wattage peak of Belgium panel production [86]. Figure 4.2 shows the resulting power production over the considered week for one installation. The same reference PV panel and the same number of panels by installation were used for each new installation.

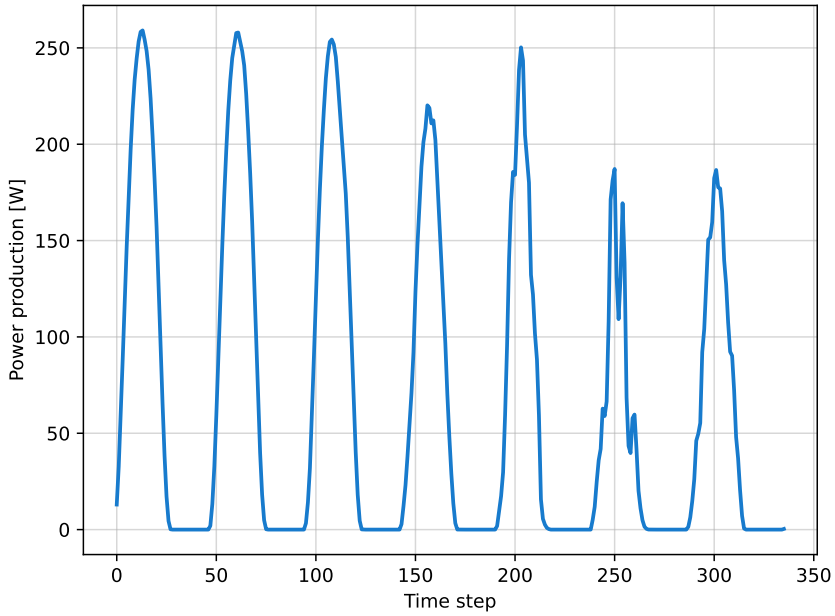


Figure 4.2: Reference PV production over a week (336 time steps of 30 minutes granularity) for the network case study.

The energy spilled computed using Eq. (4.8) is shown in Figure 4.3. On

this figure, the energy spilled is aggregated using the mean at each time step of all 500 configurations for both 43 and 70 installations. The peak PV production induced by the sun cycle occurs at the same period of time, i.e. between 11am and 1:30pm, as the lower consumption period of households, and this results in a peak of energy spilled during that period. Note that for both 43 and 70 installations, the energy spilled has the same pattern as the PV production. As expected, the more PVs added, the more energy is spilled.

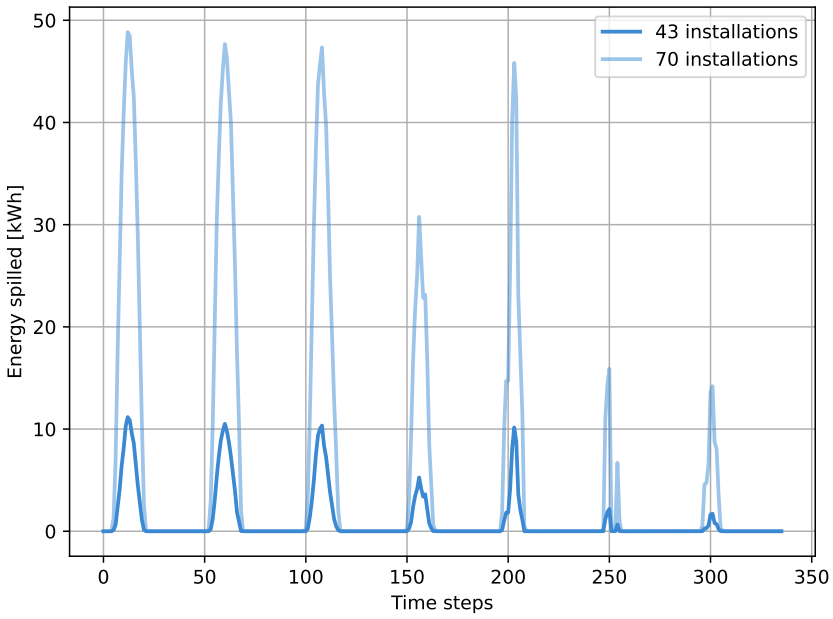


Figure 4.3: Energy spilled by day over 500 configurations and 336 time steps (i.e. a week), with 43 and 73 installations.

The approximated PDFs of the ratio of energy spilled over the potential PV production is shown in Figure 4.4. This is an approximation of the PDF, as not all configurations were considered. The configurations were sampled randomly. This figure allows to visualise how the distribution of energy spilled changes with increasing penetration. The lower and upper arrows represent, respectively, the maximum and minimum values of the PDF, while the thick middle line is the median. The darker area around the median is the typical range, i.e. between percentile one and three. The lighter blue area is the extended range using the interquartile range

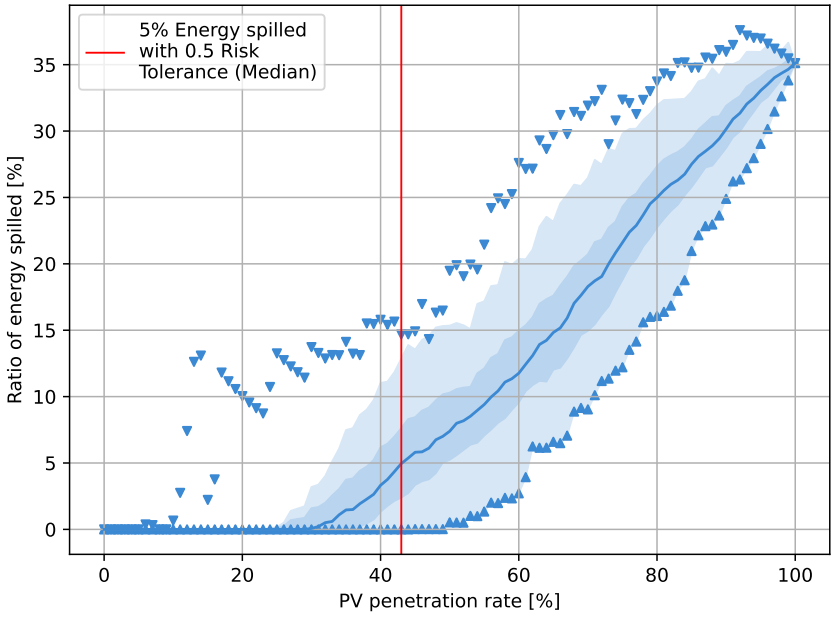


Figure 4.4: Distributions, over 500 configurations of the ratio of energy spilled over potential energy production. The dark-shaded area represent the range between the first and third quartiles, indicating the typical data. The lighter-shaded area extends this range to encompass the interquartile range. The upper and lower triangles mark the maximum and minimum values, respectively. The red vertical line denotes the 5% threshold of allowed energy spilled compared to production based on a 50% risk tolerance (the median of the distribution).

(IQR). Thus, this area is [Q1 - IQR; Q3 + IQR] and highlights unusual extremes in the distribution, for instance the maximum values that are far from this zone. This figure shows that if all customers add PV panels, the energy spilled would represent 35% of the potential installed production over the considered week of computation. If the DSO decides to set the ratio threshold of energy spilled $\overline{h}_\beta$ to 5%, with a risk tolerance $\overline{F}_\beta$ of 0.5, which is the median, the penetration rate allowed is 43% and is represented by the red vertical line in Figure 4.4. The figure shows that using PDF enables more realistic results with respect to worst-case scenario planning. Here, the worst-case would induce a penetration rate of 11%, corresponding to the first occurrence of violating the 5% limit. Any penetration rate below

33%, corresponding to the first penetration rate for which the typical and extended areas in blue cross the 5% mark, is too conservative as only outliers trespass the 5% threshold.

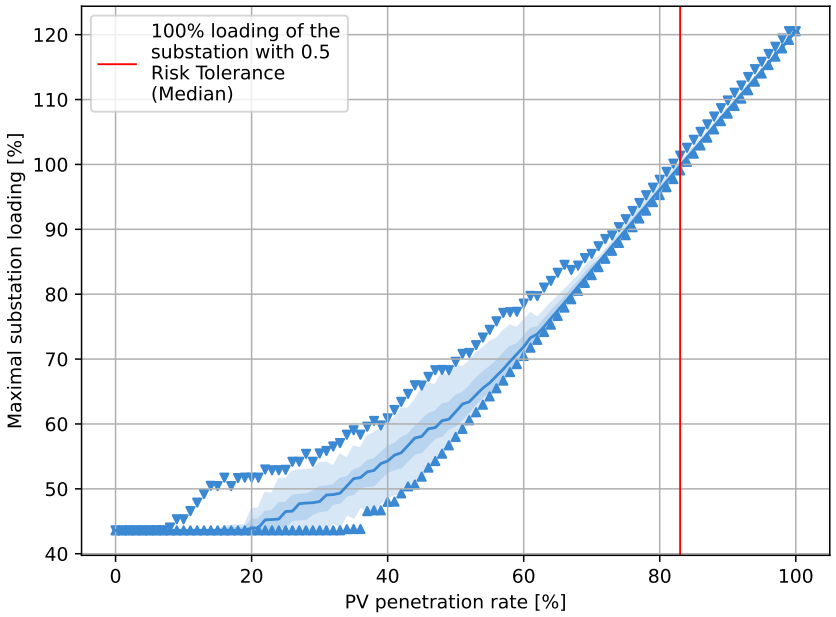


Figure 4.5: Distributions of the medium-voltage substation loading as a function of PV penetration rate, considering only the maximal power flowing through for each configuration. The dark-shaded region represents the typical range of substation loading, spanning from the first to the third quartiles, while the lighter-shaded area extends to the interquartile range. The upper and lower triangles mark the maximum and minimum loading values, respectively. The red vertical line represents the KPI threshold, set at 250 kVA for substation loading.

Figure 4.5 shows the ratio of maximum power flowing through the substation KPI, h_α, over the authorised maximum value set by the DSO. For this case study, the maximum power capacity allowed is 250kVA. The upper and lower bounds for the maximum power passing through the substation for penetration rates over 70% are nearly the same. This results from higher numbers of added PVs leading to less possibilities for connection points (i.e. similar configurations) and thus, similar total load. The tolerance risk in that region of the graph is thus monotonous. The red line represents reach-

ing 100% of the maximum power capacity of the substation (i.e., 250kVA). For this KPI, the allowed penetration rate is 83%. Note that with no installations, i.e., a zero-penetration rate, the power passing through the substation is already at 40% of its capacity.

Table 4.1: The hosting capacity (HC) obtained for 500 configurations and 336 time steps with PVs.

Indicator	Threshold	HC with minimum risk tolerance	HC with median risk tolerance	HC with maximum risk tolerance
Energy spilled	5%	11%	43%	49%
Substation loading	250kVA	82%	83%	84%

As indicated in Eq. (4.6), the stricter KPI limit determines the allowed hosting capacity. Table 4.1 regroups the results obtained for both considered KPIs. As can be seen, for this case study, the allowed penetration rate of PV panels is determined by the energy spilled.

4.6.2 Capacity assessment for electric vehicles

For this case study, a similar EV model was chosen for all the EV charging installations. This EV battery capacity was set to 40kWh, as a common value for available EVs. For the sake of simplicity, only one charging power limit of 3.6 kW was used. Both the battery capacity and the charging powers were sourced from [87]. For each new EV charger installation, the EV charging time series was randomly generated using the probability distribution shown in Figure 4.6. The distribution models drivers commuting during peak hours.

The non-served energy computed following Eq.(4.9) was aggregated by computing the mean of all configurations for each time step to depict the daily trend and is shown in Figure 4.7. The daily trend of non-served energy follows the EV charging probability.

Figure 4.8 shows the non-served energy distribution rationed over the potential energy consumption distribution. For a hundred percent penetration rate, the non-served energy represents more than 65% of the potential consumption. In the case of a non-served energy ration threshold $\overline{h}_\gamma$ of 5% of the potentially consumed energy, the median risk tolerance allows 1% for this case study. The third percentile is not much greater, and the best-case scenario (risk tolerance of 1) only allows for a 14% penetration rate.

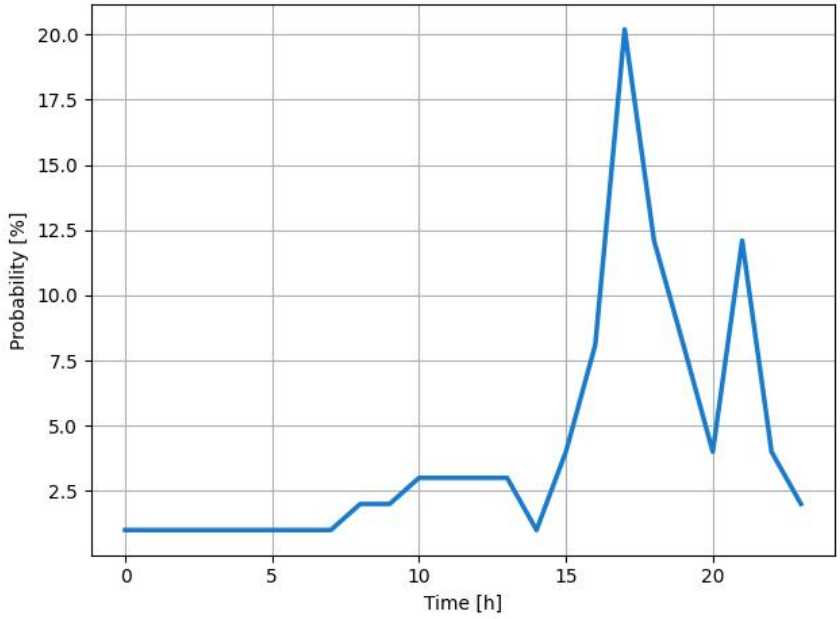


Figure 4.6: Daily probability distribution of EV charging for the network case study.

The results of the ratio of maximum power flowing through the substation KPI over the DSO threshold, of 250KVA, is shown in Figure 4.9. The allowed penetration rate is 23% for this KPI. Table 4.2 regroups the results obtained for both considered KPIs. As can be seen, for this case study, maximum power flowing through the substation KPI is more permissible than the energy non-served.

Table 4.2: The hosting capacity (HC) obtained for 500 configurations and 336 time steps with EV as installations.

Indicator	Threshold	HC with minimum risk tolerance	HC with median risk tolerance	HC with maximum risk tolerance
Energy non-served	5%	1%	1%	14%
Substation loading	250kVA	20%	23%	29%

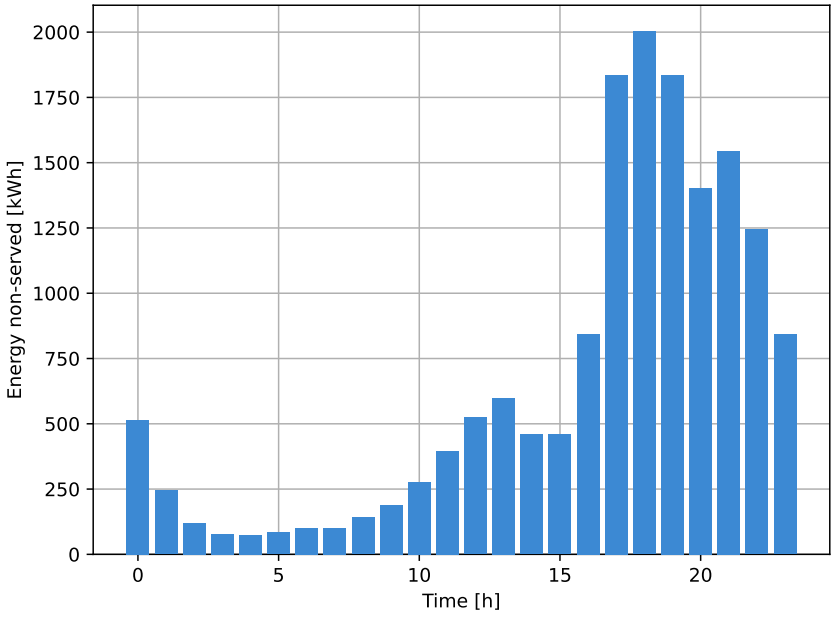


Figure 4.7: The distribution per day of non-served energy with 79 installations over 350 configurations and 336 time steps.

4.7 Intermediate conclusion

This chapter presents a new method to assess the individual network capacity considering uncertainty on installation positions and production or consumption time-series. Two types of installations (DERs) are considered: PV and EV. The impact of adding PV installations to a network is quantified by the energy not produced due to the curtailment by the inverter when detecting an over-voltage. For EVs, a symmetric reasoning is applied; the capacity is defined as the non-served energy due to under-voltage limit curtailment. The proposed method is able to consider the uncertainties in installation location and production or consumption time series, providing stochastic results as a set of KPIs. The simulation studies were performed assuming that the LV network model was not available. Results obtained with a full-size European-based test case from Chapter 2 show an individual capacity of 45% penetration rate for photovoltaic installations and an individual capacity of 4% for electric vehicle chargers, with 5% tolerance for energy spilled and energy non-served, respectively, and 0.5 risk tolerance.

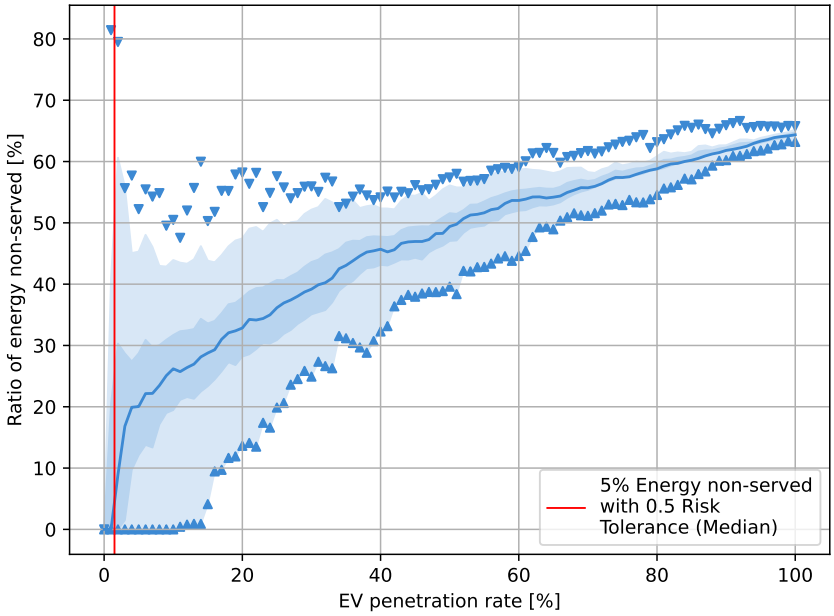


Figure 4.8: Distributions, over 350 configurations of the ratio of non-served energy over potential energy consumption. The dark-shaded area represent the range between the first and third quartiles, indicating the typical data. The lighter-shaded area extends this range to encompass the interquartile range. The upper and lower triangles mark the maximum and minimum values, respectively. The red vertical line denotes the 5% threshold of allowed energy non-served compared to potential consumption based on a 50% risk tolerance (the median of the distribution).

The method is applicable on any distribution system, it has low data requirements, and it provides tangible results which can be of help for DSOs in their real-world practical cases.

This work can be extended along several lines. Improvements should first be focused on optimising the computation time to enable considering both EV and PV installations simultaneously to compute more granular network capacity. Two options are possible to minimise this computation time: decreasing the number of evaluated configurations or the number of time steps simulated. The current number of configurations could be reduced using more-complex scenario-selection methods such as sampling techniques. Selecting a subset of days representing most days of a year

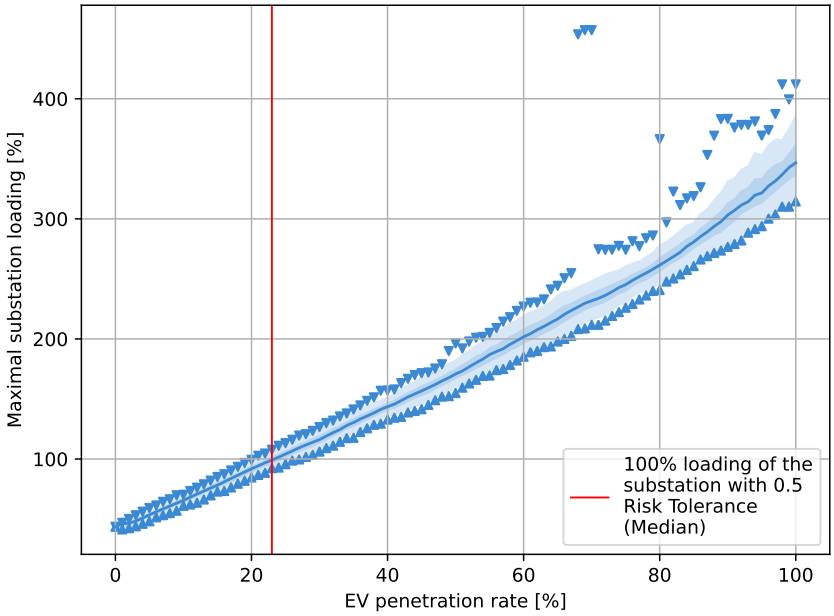


Figure 4.9: Distributions of the medium-voltage substation loading as a function of EV penetration rate, considering only the maximal power flowing through for each configuration. The dark-shaded region represents the typical range of substation loading, spanning from the first to the third quartiles, while the lighter-shaded area extends to the interquartile range. The upper and lower triangles mark the maximum and minimum loading values, respectively. The red vertical line represents the KPI threshold, set at 250 kVA for substation loading.

would significantly decrease the number of time steps evaluated by the method while providing an overview of the network capacity on the entire year. Finally, another extension of this work could be simultaneous capacity assessment for different types of installations. However, this will increase the connection possibilities, and hence the computation time.

Chapter 5

Impact of phase selection on hosting capacity

By failing to plan, you are preparing to fail.

Benjamin Franklin

Networks in distribution are three-phases but most the customers' loads and distributed energy resources (DERs) are single-phase. As the connections are mostly unknown and installed randomly, this causes an imbalance between the phases in the network. The recent major increase in decentralised energy resources such as photovoltaic panels and electric vehicle alters the loading profile of distribution systems (DSs) which exacerbates the imbalance in DS network. Distribution system operators (DSOs) try to manage the deployment of new DERs to decrease the operational costs that reverberates on customers. However, DER location and size are factors beyond any DSO's reach. This chapter presents a practical method to enhance the hosting capacity of a given DS network by minimising its operational costs due to new DER deployments, through optimal selection of their connection phase.

This chapter presents a practical method to minimise the DS operational costs due to new DER deployments, through optimal selection of their connection phase. The method is designed as a simple and practical tool for DSOs and it provides tangible actions to increase the hosting capacity (HC) while lowering the costs with minimum effort. The results obtained on a network based on a real-life Belgian low-voltage distribution network for photovoltaic panels (PVs) and electric vehicles (EVs) individ-

ually show how this simple optimal decision can help DSOs to decrease their operational costs and thus increase the hosting capacity. The impact of such distribution grid management efforts on aggregated demand for higher voltage levels in the case of PVs is also evaluated and discussed in this chapter.

5.1 Notations

Sets

$\mathcal{T}$	Set of observation periods
$\mathcal{N}$	Set of nodes
$\mathcal{E}$	Set of edges
$\mathcal{P}$	Set of phases
$\mathcal{I}$	Set of DER

Variables

ϕ_i	Connection phase of DER $i \in \mathcal{I}$
Z_e	Impedance of edge $e \in \mathcal{E}$
I_e	Edge current of $e \in \mathcal{E}$
$S_{n,p}$	Power injection of $n \in \mathcal{N}$ in $p \in \mathcal{P}$
P_{i,ϕ_i}	Power injection of $i \in \mathcal{I}$ in ϕ_i
$\overline{P_i}$	Maximum production of DER $i \in \mathcal{I}$
$\underline{V_n}$	Voltage of $n \in \mathcal{N}$
$\overline{V}$	Over-voltage threshold
$\underline{V}$	Under-voltage threshold
γ	Power-to-energy conversion coefficient
e_t^{price}	Electricity price at time t
P_t^{NL}	Total network active power loss at t
$P_{i,t}^c$	Production or load of the DER i at time t
P_t^{root}	Power fed to the feeder by the grid at the root node at t
P_t^{input}	Total power fed to the network at t
P_t^{load}	Total loads of the network at t
C_t^{NL}	Cost due to network losses at t
C_t^{DER}	Cost due to curtailment of DER production or load at t

5.2 Introduction

Phase imbalance is not a new phenomenon encountered by distribution system operators. The imbalance usually results from an uneven allocation

of loads, random consumer behaviour, and structural asymmetries [88]. However, with the random connection of emerging decentralised energy resources, such as photovoltaic panels, the phase imbalance may increase considerably [89]. This, in turn, results in frequent voltage and current issues, which are typical limiting factors for DS hosting capacity for renewable energy sources, RES [68]. For instance, over-voltages may lead to curtailment of DG production, which is currently only permitted to avoid network issues and to ensure power security [90]. If these issues are not mitigated, they may considerably influence the operational costs for distribution system operators and limit the network hosting capacity for new DERs.

While DSOs might have to reinforce their network infrastructures to host substantial volumes of new DERs, this represents considerable financial investment. There are several solutions proposed in the literature to mitigate the problem and increase the hosting capacity without reinforcing the grid [91]. In [53], [92] the authors propose a network reconfiguration to maximise the hosting capacity. Network reconfiguration is a major DSO control tool for loss minimisation and post-fault service restoration. However, real-time reconfiguration is not a practical solution at the LV level due to the scarcity of remotely controlled switches at this level. Active network management, ANM, is proposed as another solution to maximise the generation of already installed PV units through the optimal setting of available control variables [93]–[95]. In [96] the authors propose a method for optimal placement and sizing of PV units to reduce active power losses while achieving a high penetration level. However, the location and size of residential DER (EVs and PVs) are usually decided by the unit's owner based on several factors such as available area and financial means. Studies managing the charging of EVs rely on the potential flexibility of the charging time as well as the possibility to charge an EV at a different location. Papers [97]–[99] present methods to rebalance the network by controlling either the location and/or the time of EV charging. Paper [98] uses the battery flexibility of EVs to counter network imbalance. The network control strategies focus mainly on automatic switches [100]–[104]. There are also papers on optimal placement of equipment such as harmonic filters, battery storage units, phase-reconfiguration devices, and voltage regulators to increase the network hosting capacity [105]–[108]. Massive integration of such devices at the LV level demands considerable investment.

This research investigates a simple solution to manage DS operational costs by connecting new DERs, such as PV units and EVs, to optimal phases. A practical framework for identification of the optimal connection phase of new single-phase personal DERs is proposed. The optimal selection of the connection phase of DERs is a less-investigated solution, though

it is a technique which requires the least effort from the DSO's point of view. The framework is composed of two main building blocks: I. a module which calculates a defined cost for a given set of connection phases of network single-phase DERs, II. a derivative-free optimiser that finds the optimal connection phase of network single-phase DERs in conjunction with the module. The methodology is examined through simulation studies on a low-voltage distribution network based on a real-life Belgian network individually on EVs and PVs. Furthermore, the impact of such local decisions on the DS aggregated demand profile is analysed in the case of PVs.

The optimal phase selection method is designed from a practical point of view. Each time the DSO receives a notification of the installation of a new DER by a customer, the proposed method helps to find the correct connection phase of the DER to maximise the network hosting capacity, and to minimise the operational costs due to phase imbalance during the study horizon. In parallel, the method checks if it is cost-effective to switch the connection phase of all DERs already installed in the network. Finally, the method presents the optimal decision to the DSO.

The remainder of this chapter is organised as follows. Section 5.3 presents the proposed methodology for DER optimal connection phase selection. Results obtained on a Belgian network case study for both EVs and PVs are discussed in Section 5.4. Finally, Section 5.5 concludes this chapter.

5.3 Methodology

The proposed solution finds the most appropriate connection phase for new DERs. Each time a new DER needs to be added to the network, i.e., a customer buys PV panels or EV charger and notifies the DSO, a first step of the algorithm finds the connection phase that leads to less over- or under-voltage and less operational costs in a future horizon. Then, a global optimiser finds the set of phases for each previously added DER unit that again minimise the over- or under-voltage and operational costs. As rephasing existing customers with DER means DSOs will incur a labour cost, the global optimal solution is applied only when it is cost-efficient.

Consider an imbalanced three-phase distribution network. The network topology can be represented by a tree graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ where $\mathcal{N}$ is the set of network nodes and $\mathcal{E}$ is the set of edges linking the nodes. The set $\mathcal{P} = \{a, b, c\}$ denotes the set of phases. The impedance of an edge $e \in \mathcal{E}$ is denoted by Z_e, representing the impedance of a three-phase cable and neutral conductor. The node voltages and edge currents are denoted by $V_{n|n\in\mathcal{N}}$ and $I_{e|e\in\mathcal{E}}$ respectively. The network is observed over a period $\mathcal{T}$

with time steps denoted by t and of length δt. For each phase $p \in \mathcal{P}$ of node $n \in \mathcal{N}$, the load is represented by a time-series denoted by $S_{n,p} \in \mathbb{C}^{|\mathcal{T}|}$. The set $\mathcal{I}$ is the set of DERs. Let $\phi_i \in \mathcal{P}$ be the phase on which the DER $i \in \mathcal{I}$ is connected. The time-series of a DER unit $i \in \mathcal{I}$ on phase $\phi_i \in \mathcal{P}$ is denoted by $P_{i,\phi_i} \in \mathbb{R}^{|\mathcal{T}|}$. When the considered type of DER is one that produces electricity, the maximal production of such DER unit is referred to as $\overline{P}_i$.

The network is assumed to have an initial number of DERs and an acceptable level of imbalance. The method is designed to act as a practical tool for DSOs. Each time the DSO receives a request for a new DER installation at a random customer location, the proposed method performs an individual selection (IS) to find the optimal connection phase of the DER unit. The methodology inputs are DERs production or load P_{i,ϕ_i} and network loads $S_{n,p}$ time-series forecast for a time horizon $\mathcal{T}$. The optimum connection phase of the new DER is identified as the one that minimises a cost function. The considered cost function has two main components:

- The first component is the cost directly related to imbalance. Several terms can be considered in the cost due to imbalance, including the equipment ageing cost, additional network investment costs because of the inefficient use of its capacity, extra energy losses, and nuisance tripping. Without losing the generality of the proposed method, only the cost of network losses is considered for this study.

- The second component of the cost function is related to voltage levels and it is the cost of curtailing DER production or load.

The cost of curtailing production or load and the cost due to network losses are denoted by C^{DER} and C^{NL}, respectively. The minimisation problem for a new DER unit i is defined as follows:

$$\min_{\{\phi_i\}} \left(\sum_{t \in \mathcal{T}} (C_t^{NL} + C_t^{DER}) \right) \tag{5.1a}$$

subject to

$$0 \leq P_{i,t,\phi_i} \leq \overline{P}_i, \qquad \forall t \in \mathcal{T}, i \in \mathcal{I} \tag{5.1b}$$

$$\underline{V} \leq |V_{n,t,p}| \leq \overline{V} \qquad \forall n \in \mathcal{N}, t \in \mathcal{T}, p \in \mathcal{P} \tag{5.1c}$$

where the network nodal voltages and edge currents for each time step t are computed by performing a three-phase power flow:

$$\begin{aligned} V_{n,t}, I_{e,t} = \boldsymbol{PF}(\mathcal{G}, P_{i,t}, S_{n,t}, Z_e), \\ \forall n \in \mathcal{N}, e \in \mathcal{E}, t \in \mathcal{T}, i \in \mathcal{I} \end{aligned} \tag{5.1d}$$

Constraint (5.1b) ensures that the DER i is within its production or load limits ($\overline{P}_i$). Equation (5.1c) ensures that nodal voltages, computed through power flow ***(PF)*** (5.1d), respect the voltage minimum and maximum thresholds. When voltage limits are not respected, DER productions or loads are curtailed based on a defined policy. The curtailment ensures the acceptable level of imbalance in each time step. The pseudo code Algo. 2 presents the IS algorithm.

Require: $\mathcal{G}$, $Z_{e|\forall e\in\mathcal{E}}$, $S_{n|\forall n\in\mathcal{N}}$, $P_{i|\forall i\in\mathcal{I}}$
for each $\phi \in \mathcal{P}$ **do**
  for each $t \in \mathcal{T}$ **do**
    $V_{n,t}, I_{e,t} \leftarrow \boldsymbol{PF}(\mathcal{G}, P_{i,t}, S_{n,t}, Z_e)$
    if $V_{n,t|\forall n\in\mathcal{N}} > \overline{V}$ **then**
      $P^c_{i,t}, V_{n,t}, I_{e,t} \leftarrow \text{curtail}(P_{i,t}, V_{n,t}, \mathcal{G}, S_{n,t}, Z_e)$
    else
      $P^c_{i,t} = P_{i,t}$
    end if
    $cost_{\phi,t} += (C^{NL}_t + C^{DER}_t)$
  end for
  $cost_\phi += \sum_{t\in\mathcal{T}} cost_{\phi,t}$
  $p^o = \phi^p \mid cost_p = min(cost_\phi)$
end for
return p^o

Algorithm 2: Individual selection algorithm, where ***PF*** designates the power flow computation function, *cost* is the cost function and *curtail* is the curtailment function.

After installing several DER units in the network, it might reach a condition where modifying the connection phase of all previously installed DERs is cost-efficient. To check this, the method runs a global optimisation (GO) to find the optimal connection phase of all the DER units connected to the network. The optimal set of phases $\phi_{i|\forall i\in\mathcal{I}}$ is the one which minimises the cost function:

$$\min_{\{\phi_{i|\forall i\in\mathcal{I}}\}} \left(\sum_{t\in\mathcal{T}} (C^{NL}_t + C^{DER}_t) \right) \tag{5.2}$$

Global optimisation is performed using a genetic algorithm. Of course, modifying the connection phase of the all previously installed DERs has a cost. This is the cost to change some customers connection phases to coincide with the GO result. Considering this cost aims to avoid frequent changes of connection phases. If the IS cost is greater than the sum of the GO cost and phase-switching cost, the customers connection phase will be switched according to the GO result. Otherwise, the new DER unit will

be connected to the optimum phase selected by the IS. A flowchart of the general architecture is shown in Fig. 5.1.

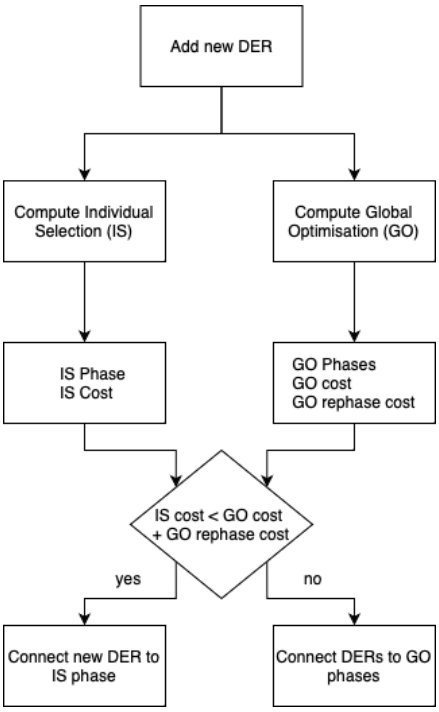


Figure 5.1: General architecture of the proposed method.

As previously mentioned, for the sake of this study, only costs caused by network losses C^{NL} and DER curtailments C^{DER} are considered. For each time step t, the C_t^{NL} is defined as:

$$C_t^{NL} = P_t^{NL} \times \gamma \times e_t^{price}, \tag{5.3}$$

where γ is the power-to-energy conversion coefficient considering a constant power during period δt, e_t^{price} is the electricity price and P_t^{NL} is the total network active power loss defined as:

$$P_t^{NL} = P_t^{input} - P_t^{load}, \tag{5.4}$$

where the considered loads include, when applicable, the added DER load:

$$P_t^{load} = \sum_{\forall i \in \mathcal{I}} P_{i,t} + P_{i,t}^c \tag{5.5}$$

and where P_t^{input} is the power fed to the network:

$$P_t^{input} = P_t^{root} + \sum_{\forall i \in \mathcal{I}} P_{i,t}^c, \tag{5.6}$$

where $P_{i,t}^c$ is the production of DER unit i for time interval t after curtailment and is only considered when the installed DER can produce (e.g. PV) and P_t^{root} is the active power fed to the feeder at the root node, and is calculated from the power flow results.

The curtailments' C_t^{DER} cost is defined as:

$$C_t^{DER} = \sum_{\forall i \in \mathcal{I}} (P_{i,t} - P_{i,t}^c) \times \gamma \times e_t^{price}, \tag{5.7}$$

The curtailment strategy is not a part of the methodology. Rather, it serves as a way to take into account voltage issues in the cost. Therefore, the curtailment policy is not discussed in this section.

5.4 Case studies

The simulation study was conducted on an LV distribution network inspired by a real-life European network [21] shown in Fig. 5.2. The network has 128 customers nodes spread on four feeders and a total of 256 nodes. Each customer is represented by a coloured circle and the three phases are represented by a different colours.

The cost for switching the connection phase of a PV unit owner was arbitrarily set to 100 euros. The price of electricity was set to 0.2702 euros per kWh, in line with average price in Belgium in 2020 [109]. This price was set constant for each time steps.

5.4.1 PVs as DER

This section presents the results for simulation on the presented network with PV as DER.

The number of PVs in one unit is set to 13 as it is the minimal number to accommodate Belgian household load and the watt peak of each PV is 290W [86]. The PV production time-series were retrieved from [11]. The curtailment strategy of GO is the same as that which is considered for IS. The GO is carried out using the genetic algorithm package *geneticalgorithm* [110].

A simple curtailment policy, that curtails all active PV units when an over-voltage occurs, was considered for the sake of simplicity. The over-voltage threshold was set to +5% of the nominal voltage. More sophisti-

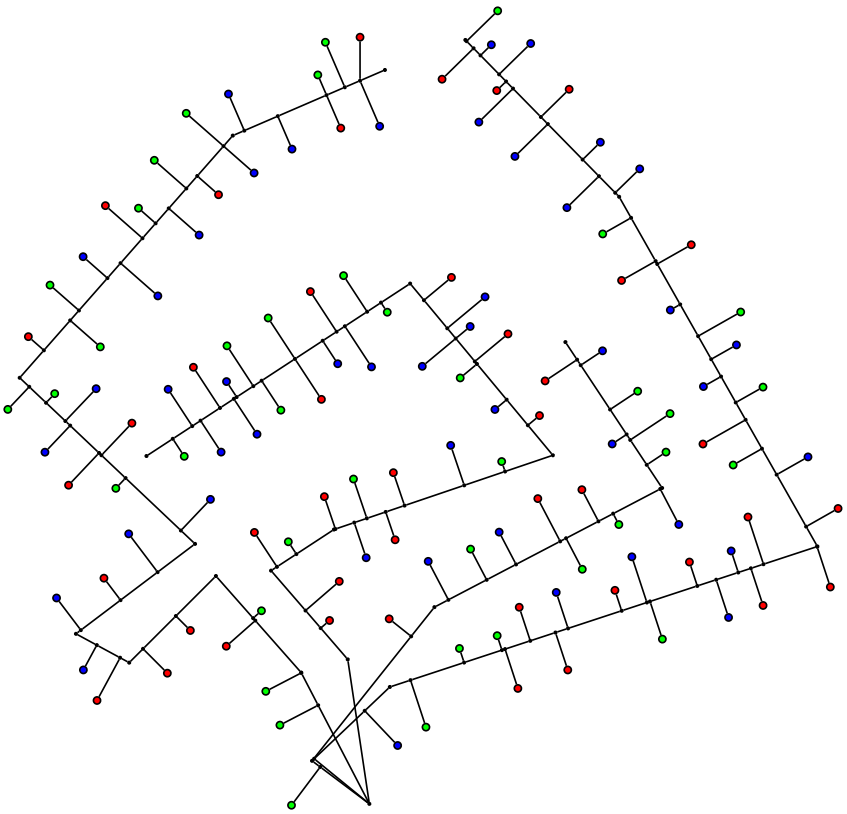


Figure 5.2: Network case study.

cated policies such as the one proposed in [94] can be incorporated in the proposed architecture.

A scenario was designed to test the performance of the proposed method and to show the impact of the DSO decisions on higher voltage levels. In this scenario, 45 customers (i.e., 35% of all the customers) have already installed PVs. The 45 customers were chosen randomly to install PV units at their connection phase. Then, the addition of 5 new PV units is considered sequentially. The cost results for this scenario are shown in Table 5.1. In this table, 'IS cost' for each phase refers to the cost of selecting that phase for the new PV unit without changing the phases of the previously installed PV units. Then, 'GO cost' shows the minimum cost obtained with the optimal phases while 'total GO cost' refers 'GO cost' plus the

Table 5.1: The obtained costs for six new sequentially added PV units on the network with 45 initial PV units. Costs are in euros. The selected option is in bold. IS cost for each phase is the output of IS for the phase. GO cost if the cost of changing several connection phase without the labour cost while the total GO is the sum of the labour cost (number of phase to rephase multiplied by 100€) and the GO cost column.

Total number of PVs	IS cost phase A	IS cost phase B	IS cost phase C	GO cost	Total GO cost (number of customers to rephase)
46	22874	23286	19064	283	**2983** (27)
47	277	**275**	276	273	3973 (37)
48	282	282	**280**	279	2779 (25)
49	**289**	291	290	286	3486 (32)
50	293	295	**292**	287	4087 (38)

labour cost for modifying the connection phase of previously installed PV units. As can be seen, for the first added PV unit, the total GO cost is smaller than all IS costs. Thus, the decision is to modify the connection phase of 27 PV units according to the GO outputted phases. This rephasing enables to considerably decrease the next costs as shown in Fig. 5.3. This figure depicts the difference between using the proposed solution, and connecting PVs to the customer's current connection phase. The cost of curtailment dominates the total cost for the first PV unit. The costs with the proposed solution are the ones in bold in Table 5.1. With the proposed solution and by modifying the connection phases of PV units, the total cost is considerably decreased and is limited to the cost of network losses.

Fig. 5.4 shows the total curtailment without applying the method proposed rephasing action after the first added PV unit. By rephasing 27 PVs, the over-voltages and thus the frequent curtailments are avoided.

Results shown in Fig. 5.3 and Table 5.1 were obtained using accurate time-series for both production and load. However, in practice, such accurate time-series are not available in advance for DSOs. There are several studies, such as [111]–[113], on forecasting the load and PV production time-series. To assess the performance of the proposed method without accurate time-series, the method was also evaluated using inaccurate time-series. The load time-series were generated by random variation of each load time-series within a 30% deviation with normal distribution. The same process generated the PV production time-series within a 5% devia-

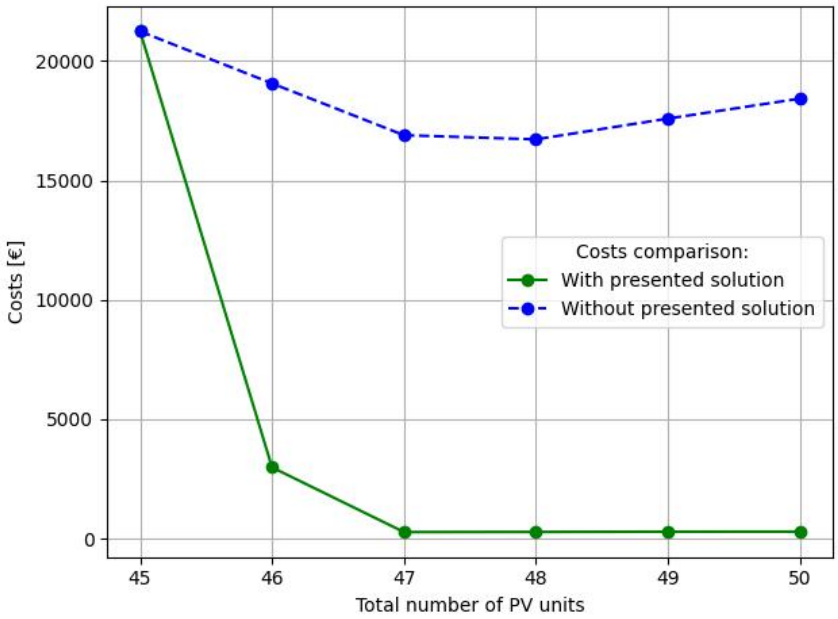


Figure 5.3: Comparison of costs with and without the presented solution. The simulations start with already 45 PV installations. Then, when using the presented solution, at each new installation added both the GO and the IS are preformed before choosing the adequate solution. The costs on the ordinate are the costs of the actions performed when adding the novel installation. When not using the presented solution, the new installation is directly added to the current phase of the customer.

tion with normal distribution. Fig. 5.5 shows the costs for adding the same PVs as previously but with inaccurate time-series and accurate time-series both using the presented method. For the first added PV, the presented method with inaccurate time-series chooses to rephase 31 customers, while 27 customers were rephrased with accurate time-series. For the next added PVs, the curve trend is similar to the case with accurate time-series. Both costs of this figure are still considerably smaller than the cost without the presented method in Fig. 5.3.

Fig. 5.6 shows the effect on the aggregated active power demand. The main impact is that by avoiding frequent PV curtailments, the profile of network demand, particularly during the day, is considerably improved and phases are more balanced. It is interesting to note that the absolute values

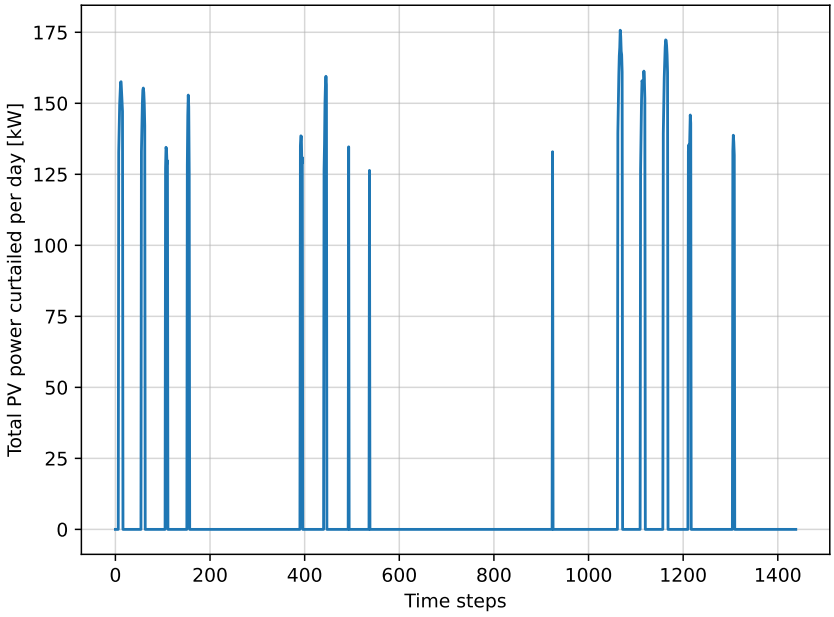


Figure 5.4: Power curtailed over one month period (1440 time-steps) when adding 46 PV units when connecting the PV installations to the current phase of the customer, i.e., without any rephasing.

of demand without the presented solution is lower. This can lead to lower network losses in this case. However, since the cost of curtailment dominates the total cost, the proposed solution maximises the production as shown by the negative values mid-day in the figure. This shows that simple actions taken by DSOs on LV networks can have a considerable impact on higher voltage levels.

The current DSO practice for new PV units is to connect them to the current customer connection phase. This practice may increase the network operational costs due to imbalanced operation, and may also increase the frequency of network issues such as over-voltage. The imbalance can be improved by proper addition of PV units to an optimal phase, but the question is how to choose the optimal phase each time the DSO receives a request for a new PV installation. The other question is what to do if the operational costs for all three connection phases are unacceptable. This chapter presents a simple and practical method designed from the DSO's point of view to provide suggestions for such questions. The results indicate

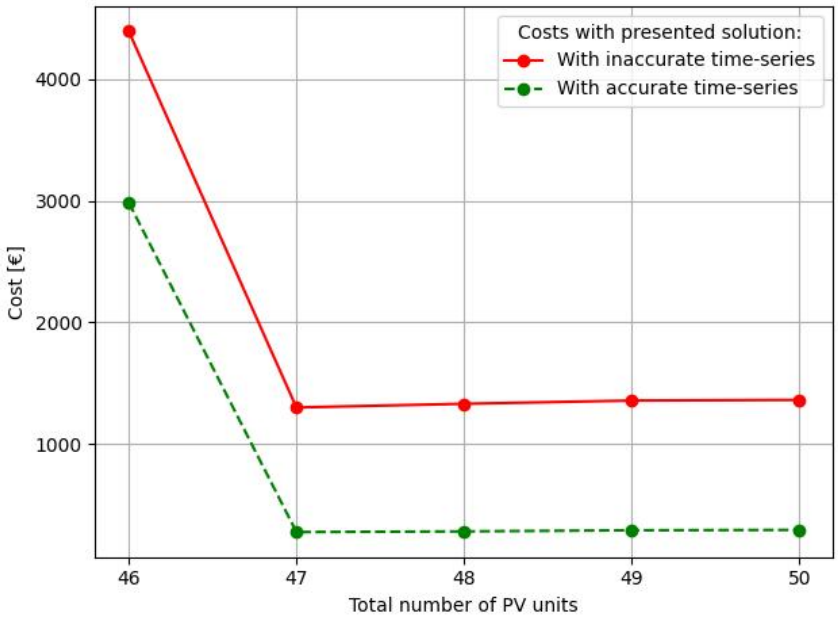


Figure 5.5: Comparison of costs with presented solution using inaccurate or accurate time-series as input. The simulations start with already 45 installations. Then, one by one, a new installation is added and both GO and IS are preformed before choosing the adequate solution. The costs on the ordinate are the costs of the actions performed when adding the novel installation.

that despite the simple procedure, the method can provide good suggestions to considerably reduce the network operational costs.

The proposed method requires a forecast of PV and load time-series to consider the variation of load and production over time. There are methods to predict load and production profiles with a fair accuracy [111]–[113]. The obtained results indicate that the inaccuracies in the forecasted time-series can impact the results of the proposed method. However, the costs with selected phases with inaccurate inputs are still close to the optimal solution and considerably lower than those observed with the current DSO practice.

Considering both stochastic and time variations in an optimisation procedure that repeats the computations over and over may not be computationally tractable. Designing a computationally efficient algorithm to consider stochasticity can be a line for future work. Further studies can

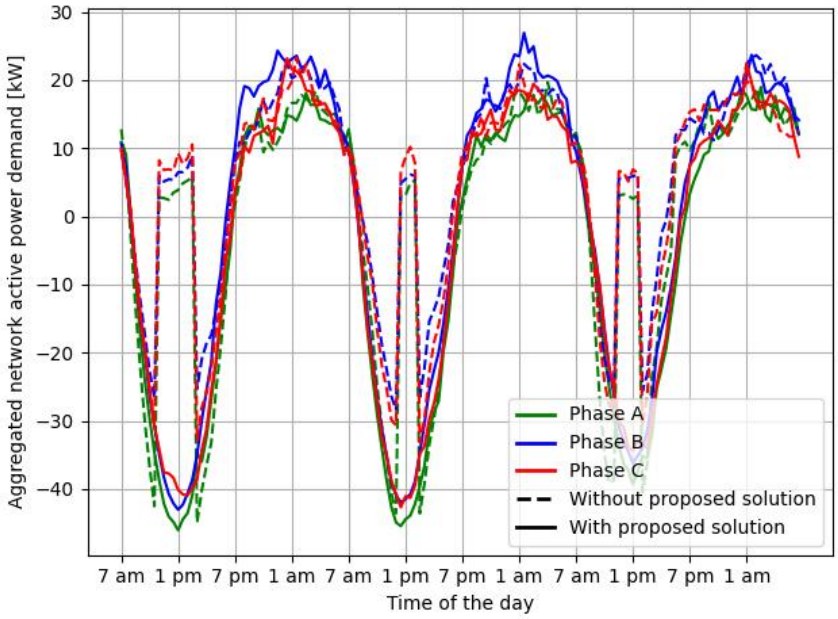


Figure 5.6: Aggregated active power demand by time of the day for three consecutive days.

investigate the study horizon $\mathcal{T}$, electricity price, and phase-switching cost, as they impact the final decision. Moreover, more cost terms could be considered as well as a more sophisticated curtailment policy.

5.4.2 EVs as DER

This section presents the results obtained while applying the methodology to the same test system inspired by an existing Belgian three-phase LV distribution network with EV chargers as DERs. As a reminder, the network has 128 customer nodes spread on four feeders and these nodes are possible EV charger locations. The EV chargers are assumed to be connected to the customer load phase. The EV charging time-series are generated from charging probabilities depicted in Fig. 5.7 [21] which is the same as Fig. 4.6 in Chapter 4. The nominal power of EV chargers considered is 7kW. Two scenarios are considered to highlight the impact of the methodology for EV chargers on the network.

In the first considered scenario, six EV chargers are assumed to be

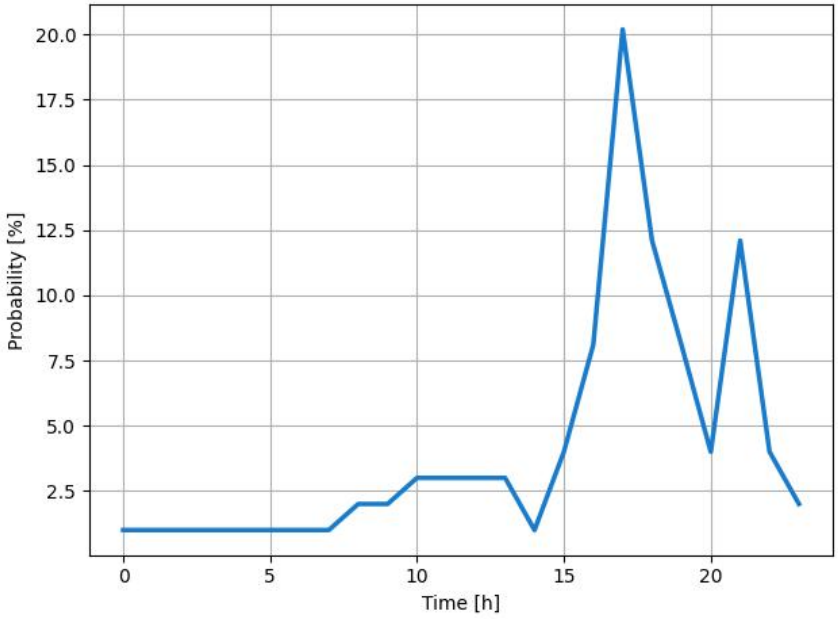


Figure 5.7: EVs charging probability in percent by hours.

Table 5.2: Cost results, in euros (€), obtained for adding sequentially four new EV chargers on the case study network with six initial EV chargers. The selected connection phase is in bold. IS cost for each phase is the output cost of IS for that phase, GO cost is the output of GO without labour cost and total GO cost take into account the labour cost that is 100€times the number of customers to rephase shown in parenthesis.

Total number of EVs	IS cost phase A	IS cost phase B	IS cost phase C	GO cost	Total GO cost (number of customers to rephase)
7	9499	11129	**2484**	2236	2636 (5)
8	**2648**	2651	2677	2207	2807 (6)
9	8502	**3020**	5111	2834	3334 (5)
10	3206	**3159**	3208	2297	3097 (8)

initially connected to the network. Then, sequentially, to mimic several consecutive EV buyers, four chargers were set up according to the hosting

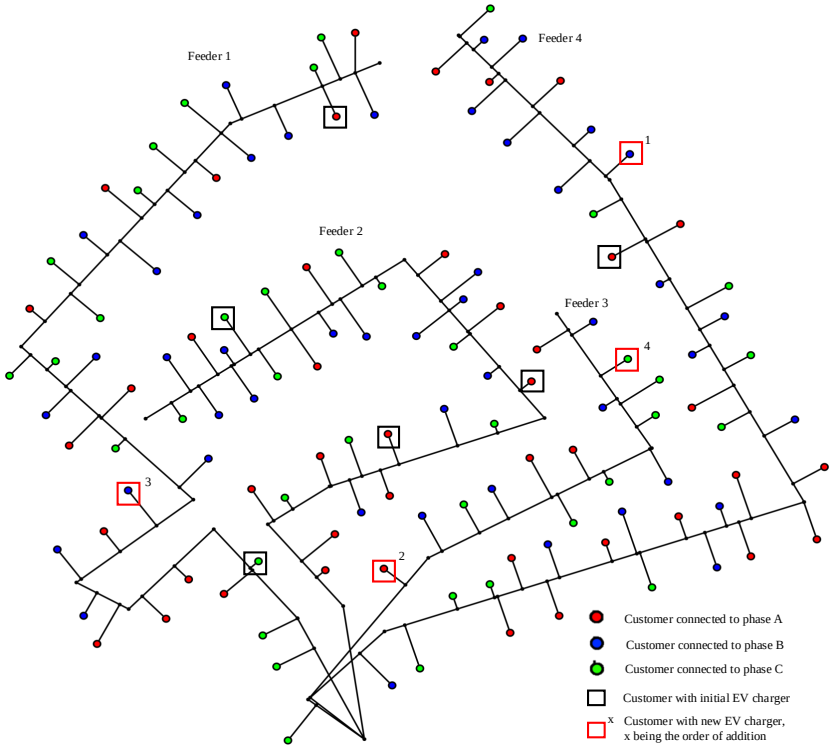


Figure 5.8: Initial network case study for the EV simulations. Each phase has a different colour. The initial EV installations are highlighted with the black rectangles while the new installations have a red rectangle with the around it being their order of installation in the simulation. The phase are the initial ones, i.e., not modified using the proposed method.

capacity found in [21]. The situation on the initial network is shown in Fig. 5.8. The black rectangles are the six initial installations and the four red ones with the order over them are the four chargers that are added consecutively. Table 5.2 shows, by total number of installations, the different costs obtained for the corresponding addition of these four chargers. The first new installation is on feeder 4. On this feeder, an EV charger was initially connected to phase A and most of the customers were connected to phases A and B. For phase B, the customers were mainly at the downstream of the feeder. As expected, and as shown in Table 5.2, phase C is selected. As another example, the third new installation is on feeder 1. On this feeder, two initial EV chargers are connected. One of them is

connected to phase A and is located at the downstream of the feeder while the other one is connected to phase C at the upper section of the feeder. The single-phase customers are almost equally distributed between phases. The new EV charger is, thus, selected to be connected to phase B. The results show that the cost varies depending on the location of the charger. For all the four new EV chargers, the GO finds a set of connection phases for all EV chargers that leads to a smaller operational cost, but the labour cost stops the rephasing from being cost effective.

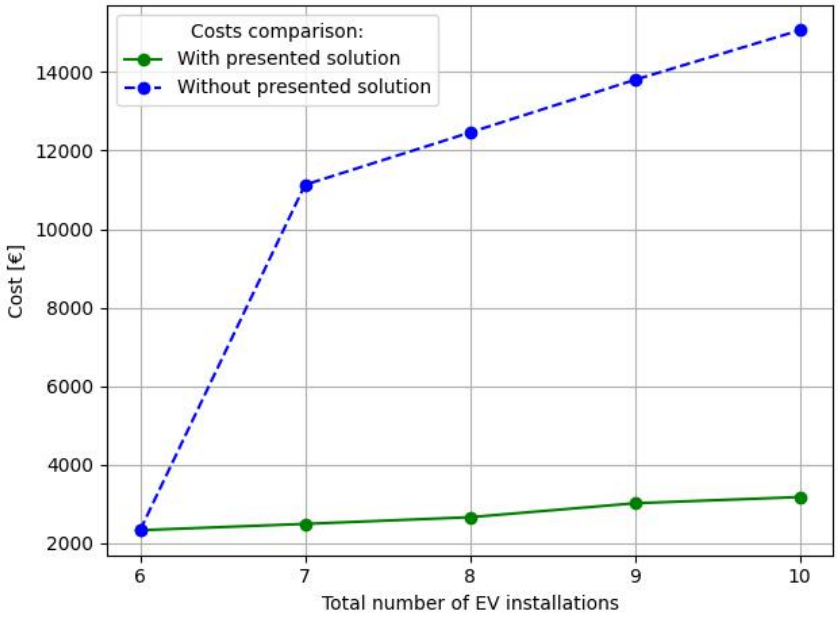


Figure 5.9: Comparison of costs, in euros, with and without the presented solution. The simulations have six initial EV chargers installations. Then, when using the presented solution, at each new installation added both the GO and the IS are preformed before choosing the adequate solution. The costs on the ordinate are the costs of the actions performed when adding the novel installation. When not using the presented solution, the new installation is directly added to the current phase of the customer.

Fig. 5.9 shows the costs evolution with the proposed solution and without it. The cost with the proposed solution is the one in bold shown in Table 5.2. The cost without the proposed solution is the cost while installing the EV chargers at the node customer connection phase. For the first new EV

charger, the connection phase without the proposed solution would be the customer's current phase, phase B, with a cost of 11129€during the considered time horizon. However, the proposed solution proposed to connect it to phase C with a cost of 2482€for the same time horizon. The difference in slop of the two lines shows that the proposed method finds better connection phases. From this figure, the method reduced the costs by up to, on average, 78%.

To highlight the improvement in decreasing the curtailment, a second scenario with eight initial EV chargers is considered. In this case, the total corrective cost, including rephasing costs of four chargers, is 3984€. This cost is smaller than the preventive selection (IS) cost, which is 15637€. Fig. 5.10a shows the curtailment without corrective rephasing (GO) with nine installations and Fig. 5.10b shows the curtailment powers with. The use of the proposed solution can lead to a significant increase of network availability and decrease of curtailed power.

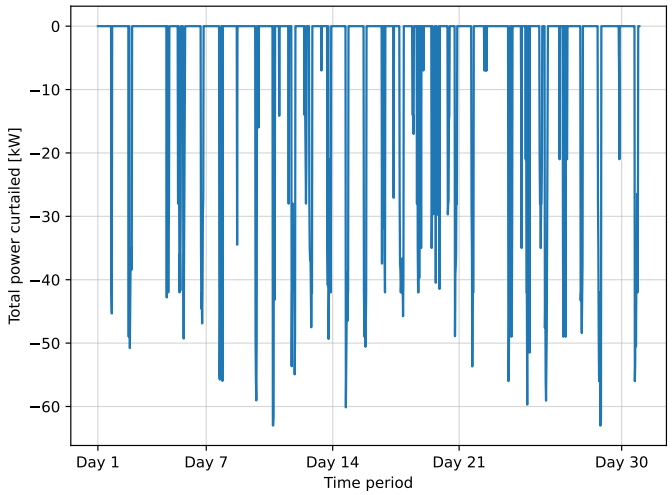
5.5 Intermediate conclusion

With massive integration of DERs, DSOs are facing increasing challenges such as greater voltage imbalance and decrease of power quality in frequency, both leading to higher operational costs. Optimal selection of the connection phase of DERs requires least effort but is a less-investigated solution to this problem. This chapter has presented a two-step method to this end. The method performs an optimal selection of the connection phase of each new DER of a given network to enhance its hosting capacity, and it simultaneously checks if it is cost-effective to modify the connection phase of all the installed network DERs to the optimal phases identified.

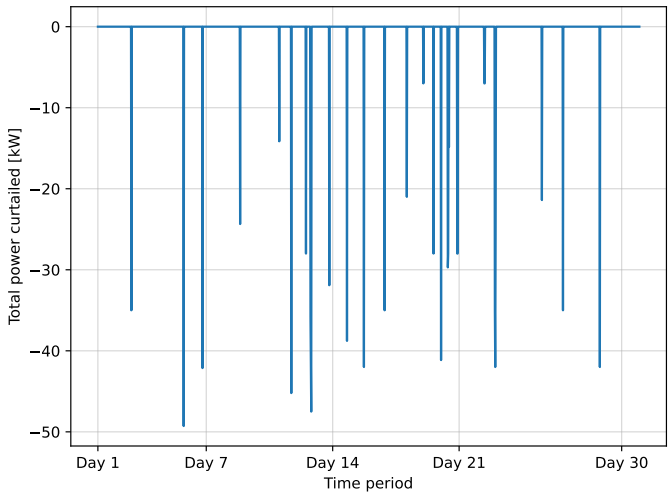
Results on an LV network simulation with PV as a DER shows that using optimal selection of the connection phase helps to considerably decrease the curtailment and the over-voltage, thus increase the hosting capacity while decreasing the DS operational costs. The results also highlighted the impact of LV management with the proposed actions on the aggregated demand. Both the imbalance and the active power demand decrease with the proposed method.

Results obtained show that using optimal selection of the connection phase of EV chargers enables a saving up to 78% in operational cost for a given scenario.

The method has a simple logic and is applicable to any distribution system. It provides actions which can be helpful for DSOs in real-world practical cases. The proposed methodology, as HC studies, is intended as a planning study; therefore, the running time was not investigated.



(a) Curtailment powers for 9 installations without GO



(b) Curtailment powers for 9 installations with GO

Figure 5.10: Simulation curtailment powers for the nine installations without and with GO.

While some possible future works have been addressed in the discussion for the case of PVs, in general, this work can be extended along several lines.

First, a well-thought methodology for careful selection of the study horizon, curtailment, network loss and phase switching costs could be developed, as they affect the final decision. Also, more terms can be considered in the cost function and more sophisticated curtailment strategies or more appropriated flexibility policies should be considered. Second, the tool could take other types of installation into account and the technologies could be evaluated simultaneously. Finally, in this study, future loads and time-series are initially assumed to be known. While some uncertainty for these have been explored for the PV production, further studies should involve the uncertainty in DER characteristics as well as in the customer loads. The method could, thus, be extended to improve its stochasticity.

Chapter 6

Combined hosting capacity assessment

Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.

Marie Curie

This chapter presents an analysis of the combined hosting capacity (HC) of photovoltaic panels (PVs), electric vehicles (EVs), and heat pumps (HPs). The HC problem is defined using a generic formalism previously introduced Chapter 3. It is applied to a reconstruction of a low-voltage Belgian electrical distribution network, which was reconstructed using a topological path identification (TPI) methodology. Exogenous data needed for the HC analysis is provided by an observation tool developed by the Belgian distribution system operator (DSO), RESA. To ensure a realistic representation of the impact of the combined technologies, time-series with a granularity of 15 minutes are used. Results show that under-voltage is encountered rapidly for both EV and HP penetrations higher than 50%. Over-voltage is not faced before 75% of PV penetration. By contrast, even with high penetration rates for all three technologies, the lines of the case study network are not overloaded.

6.1 Notations

N	Network
T	Time
Sets	
$\mathcal{P}$	Set of network issues e.g., $\mathcal{P}$ = { over-voltage, overloading }
$\mathcal{N}$	Set of nodes
$\mathcal{E}$	Set of edges
$\mathcal{C}$	Set of customer nodes
$\mathcal{H}$	Set of types of technologies { PV, EV, HP}
$\mathcal{I}_h$	Set of installation options of technology $h \in \mathcal{H}$ e.g., $\mathcal{I}_{\text{EV}}$ = { 0kW, 3kW, 7kW}
$\mathcal{T}$	Set of time steps, $\mathcal{T} = \{1, ..., T\}$.
$\mathcal{S}$	Set of all scenarios
$\mathcal{S}^c$	Set of considered scenarios
$\mathcal{S}^c_{\mathbf{a}}$	Set of considered scenarios with penetration $\mathbf{a}$
$\mathcal{A}^c$	Set of considered penetrations
$\mathcal{A}^f$	Set of feasible hosting capacities
Variables	
$P_{t,c}$	Power consumption of customer c at time-step t
$P_{t,c,h}$	Power consumption or production at time-step t of technology h installed at customer c
$i_{\mathbf{s},c,h}$	The installation of the technology h at customer c in scenario $\mathbf{s}$
$V_{\max}$	Maximal Voltage limit
$V_{\min}$	Minimal Voltage limit
$V_{t,n}$	Voltage at node n at time-step t
$I_{t,e}$	Current at edge e at time-step t
I^l_e	Nominal current of edge e
$L_{t,e}$	Edge e loading at time-step t
$L_{\max}$	Overloading threshold

6.2 Introduction

The urge for carbon neutrality driven by the European Union (EU) has led to a significant increase in purchase rates in recent years of electric vehicles (EVs), photovoltaic panels (PVs), and heat pumps (HPs). Such increases in low-voltage distribution networks can stress some networks to their safety operation limits. Determining the combined hosting capacity (HC) of PV-

EV-HP systems is a crucial metric for distribution system operators (DSOs) to assess the ability of a power network to accommodate these technologies without adverse effects and plan investments accordingly.

While the topic of HC has been well-researched for several decades, studies involving combined technologies seem to be less mature. There are three categories of methods to evaluate PV-EV-HP HC:

(i) Individually, by evaluating each technology independently;

(ii) Simultaneously, by evaluating combined pairs of technologies;

(iii) Simultaneously, by evaluating all three technologies combined.

The first category (i) is the straightforward increment of the traditional PV HC. For instance, authors in [114] individually compute the HC of the PVs, EVs, and HPs technologies. The second category (ii) accounts for the cumulative influence of a pair of technologies on the network. In [115], the combined effects of the PVs-EVs, EVs-HPs, and PVs-HPs pairs are studied. The third category (iii), which aims to assess the combined impact of all three technologies, is addressed in more recent studies. Authors in [115] consider several penetrations per technology while in [116] equal penetration levels (e.g., 25% for all technologies) are considered. To simultaneously account for multiple technologies, most HC studies focus on the second category (ii) as the third one (iii) is difficult to formalise and complex to implement. To address this problem of formalism, Chapter 3 based on [20] presented a generic formalism for the HC problem with multiple technologies. This formalism enables a comprehensive analysis of all aspects of HC, facilitating clear communication in the field. It also allows for consistent comparison of results and tracking of advancements over time. The present chapter exploits this formalism on a real-life combined HC problem with multiple technologies (PV-EV-HP).

Additionally, the study carries out an HC analysis using several independent penetration rates for each technology. To ensure the accuracy of this analysis, empirical data provided from available information from RESA, a Belgian DSO, is used. Using empirical data provides practical insights into network behaviour under actual operating conditions. Authors in [115], [117] use a real network but, as time-series related to the used network are not available, they use publicly available standard load profiles (SLP) for customer loads. In [116], authors use both a real network and the customer loads associated with it. In this study, a reconstruction of a real low-voltage distribution network as well as the smart meters data for the loads are used.

To summarise, the contributions of this chapter are:

- Adopting the generic formalism [20] for the combined HC problem with three technologies simultaneously (PV-EV-HP) ;
- Using a reconstruction of a real low-voltage distribution network along with its smart meter data;
- Analysing the impact of combined PV-EV-HP on several penetration rates.

The rest of the chapter is structured as follows: Section 6.3 introduces the HC problem. Section 6.4 presents the case study. Section 6.5 develops the implementation to compute the combined HC and Section 6.6 discusses the obtained results. Finally, Section 6.7 concludes the chapter with a summary and future prospects.

6.3 Problem Statement

This chapter follows the formalism defined in [20] and presented in Chapter 3.

Let N be an unbalanced three-phase network. The network, N, is composed of nodes and edges. The sets of all nodes and edges are denoted by $\mathcal{N}$ and $\mathcal{E}$, respectively. The network contains a set of customers $\mathcal{C}$, that is a subset of the nodes: $\mathcal{C} \subset \mathcal{N}$.

The study is conducted over a period of time T and the set of all time-steps is $\mathcal{T} = \{1, ..., T\}$, a time-step of $\mathcal{T}$ is referred to as t. For each time-step t, the consumption of a customer $c \in \mathcal{C}$ is denoted as $P_{t,c}$. In addition to their load, customers can install new technologies from a set of new technologies $\mathcal{H}$. The technologies can have different sizes, and the set of the possible sizes for a technology $h \in \mathcal{H}$ is $\mathcal{I}_h$. The production or consumption of a technology $h \in \mathcal{H}$ of a customer $c \in \mathcal{C}$ at time-step $t \in \mathcal{T}$ is denoted $P_{t,c,h}$.

Given that customers can install different technologies with different options at different time-steps, different possible scenarios are possible. Each scenario represents a combination of installed technologies across the network. The set of all scenarios is referred to as $\mathcal{S}$. A scenario $\mathbf{s}$ from this set, $\mathbf{s} \in \mathcal{S}$, is formally defined as the tuple of installed sizes of each technology at each time-step for each customer:

$$\mathbf{s} = (i_{c,h,t} | \forall c \in \mathcal{C}, h \in \mathcal{H}, t \in \mathcal{T}) \tag{6.1}$$

where $i_{c,h,t} \in \mathcal{I}_h$ is the size of the technology h installed at customer c at time-step t. Note that bold characters are used for tuples.
Given that the set of all possible scenarios is intractable, a subset of it,

defined as $\mathcal{S}^c$, is considered for the analysis. The construction of this subset is further explained in Section 6.4.

Let $\mathcal{A}^c \in \mathbb{R}^{|\mathcal{H}|}$ be the set of all considered penetrations. A penetration $\mathbf{a} \in \mathcal{A}^c$ is a tuple that gauges the amount of new installations in the network for each technology in a given scenario. The function $\mathbf{g}(\mathbf{s}) : \mathcal{S}^c \to \mathcal{A}^c$ computes the penetration tuple for a given scenario $\mathbf{s}$:

$$\mathbf{g}(\mathbf{s}) = (g_h(\mathbf{s}) \mid \forall h \in \mathcal{H}) \tag{6.2}$$

where the penetration of a technology is chosen as the ratio of the number of customers with that technology installed to the total number of customers, and computed using the function $g_h(\mathbf{s})$ defined as:

$$g_h(\mathbf{s}) = \frac{|\{c \in \mathcal{C} \mid i_{\mathbf{s},c,h} \neq \emptyset\}|}{|\mathcal{C}|} \tag{6.3}$$

where $i_{\mathbf{s},c,h}$ designates the installation of the technology h at customer c in the scenario $\mathbf{s}$.

To detect if any issue occurred in scenario $\mathbf{s}$ the function f is defined as follows:

$$f(\mathbf{s}) = \begin{cases} 1, & \text{if } \exists t \in \mathcal{T} : f_t(\mathbf{s}) = 1; \\ 0, & \text{otherwise,} \end{cases} \tag{6.4}$$

where f_t is the function that detects any issue that occurred in the scenario $\mathbf{s}$ at time-step t. This function f_t that detects the considered issues $\mathcal{P}$ is defined as:

$$f_t(\mathbf{s}) = \begin{cases} 1, & \text{if } \bigvee_{p \in \mathcal{P}} f_t^p(\mathbf{s}); \\ 0, & \text{otherwise,} \end{cases} \tag{6.5}$$

where $f_t^p(\mathbf{s})$ is the function that evaluates if the issue $p \in \mathcal{P}$ occurs at time-step t of scenario $\mathbf{s}$. Two voltage level (VL) issues are considered: over- (OV) and under-voltage (UV); alongside the overloading (OL) of the lines. Therefore, the set of considered issues is $\mathcal{P} = \{\text{VL}, \text{OL}\}$.

The voltage level issues are detected by the function $f_t^{VL}(s)$ defined as follows:

$$f_t^{VL}(\mathbf{s}) = \begin{cases} 1, & \text{if } \exists n \in \mathcal{N} : V_{\min} > V_{t,n} > V_{\max}; \\ 0, & \text{otherwise,} \end{cases} \tag{6.6}$$

where $V_{t,n}$ is the voltage of node n at time t and, $V_{\min}$ and $V_{\max}$ are, respectively, the minimal and maximal allowed voltages. The function $f_t^{OL}(\mathbf{s})$ detects lines overload, and is defined as:

$$f_t^{OL}(\mathbf{s}) = \begin{cases} 1, & \text{if } \exists e \in \mathcal{E} : L_{t,e} = \frac{I_{t,e}}{I_e^l}, L_{t,e} > L_{\max}; \\ 0, & \text{otherwise,} \end{cases} \tag{6.7}$$

where $L_{t,e}$ is the edge loading at time-step t, $I_{t,e}$ is the current at edge e at time t, I_e^l is the nominal rated current of edge e and $L_{\max}$ is the threshold for line overloading.

The voltages $V_{t,n}$ at each node and the currents $I_{t,e}$ at each edge are obtained by computing a power flow (PF) on the network with both the customers' loads ($P_{t,c}$) and the installations' production or consumption ($P_{t,c,h}$):

$$\{V_{t,n}, \forall n \in \mathcal{N}\}, \{I_{t,e}, \forall e \in \mathcal{E}\} = \mathrm{PF_N}(P_{t,c}, P_{t,c,h}), \forall t \in \mathcal{T}. \tag{6.8}$$

The HC is the set of penetrations the network can sustain while not encountering any issues. In this study, the HC is chosen as the feasible penetrations $\mathcal{A}^f$, defined in [20], which are penetrations that are associated with scenarios with no issues:

$$\mathcal{A}^f = \{\mathbf{g}(\mathbf{s}) | \forall \mathbf{s} \in \mathcal{S} : f(\mathbf{s}) = 0\}. \tag{6.9}$$

6.4 Case study

This section introduces the case study network along with exogenous data and outlines the scope of the study.

6.4.1 Network

The network considered in this chapter is a reconstruction of a real Belgian distribution network.
The network is reconstructed using the systematic procedure proposed in [118]. The network is obtained by using transformation functions to transform raw data from the DSO into well-defined information. An optimisation algorithm identifies the optimal paths to connect customers to their respective feeder.

Figure 6.1 shows the full network. This is a three-phase unbalanced network, where customers are connected to either only one phase or all three of them.

Given its size and the radial operation of the network, the proposed HC computation will focus only on one substation, but the work can be extended to the whole network. The selected substation is shown in Fig. 6.2.

The substation has two feeders, distinguished by purple (left feeder) and blue (right feeder) lines. Different feeders are only interconnected at the MV/LV substation itself, or through switches that can be opened or closed to control power flow. Table 6.1 details the number of elements per

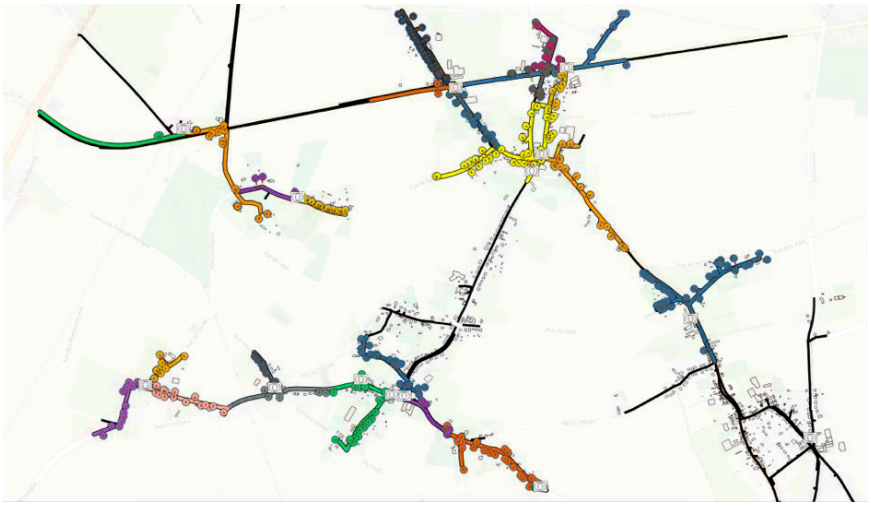


Figure 6.1: Representation of the reconstructed Belgian network. Each colour represents a feeder and the substations are presented using the common two-circle symbol.

Figure 6.2: Considered network which consist of a subpart of the reconstructed Belgian network that have one single substation connected to two-feeders.

category of the considered subnetwork. In the remainder, this subnetwork is referred to as the network.

Table 6.1: Description of the considered single MV/LV transformer network.

Category	Number of elements
customers	23
lines	42
feeders	2
MV/LV substations	1

Among the 23 customers present in the network, 15 of them have smart meters (SM). Both single-phase customers and customers without an SM are randomly connect to a phase. As the time series for customers without SM are missing, an SM time series from a customer with a similar annual consumption is attributed to these customers.

6.4.2 Technologies

The considered technologies are: PVs, EVs and HPs leading to technology set: $\mathcal{H} = \{\text{PV}, \text{EV}, \text{HP}\}$. The penetration tuple defined in Eq. (6.2) is:

$$\mathbf{g}(\mathbf{s}) = \big(g_{PV}(\mathbf{s}), g_{EV}(\mathbf{s}), g_{HP}(\mathbf{s})\big). \tag{6.10}$$

For each technology, the set of options is limited to their size. A single PV installation of size of 20 PVs with a 290-watt peak is available. Two types of EV chargers, 3kW and 7kW, and two heat-pump sizes, 7.5kW and 15kW, are available. The different available sizes per technologies are summarised in Tab. 6.2.

Table 6.2: Technology size per technology

Technology type	Technology size
PV	$\mathcal{I}_{\text{PV}} = \{0,\ 20 \times 290\text{W}_{\text{peak}}\}$
EV	$\mathcal{I}_{\text{EV}} = \{0,\ 3\text{kW},\ 7\text{kW}\}$
HP	$\mathcal{I}_{\text{HP}} = \{0,\ 7.5\text{kW},\ 15\text{kW}\}$

In this study, the size of a given technology for a customer cannot change during the time period and new installations are added at the first time-step ($t = 0$), thus $i_{c,h,0} = i_{c,h,t}, \forall c \in \mathcal{C}, h \in \mathcal{H}, t \in \mathcal{T}$.

6.4.3 Considered scenarios

The considered scenario set $\mathcal{S}^c$ is constructed by first reducing the considered penetration set $\mathcal{A}^c = \{\mathcal{A}^c_h \mid \forall h \in \mathcal{H}\}$: for each technology $h \in \mathcal{H}$, the considered penetrations are $\mathcal{A}^c_h = \{25\%, 50\%, 75\%, 100\%\}$. Since these penetration levels are independent for each technology, the total number of possible combinations is the product of the number of choices for each technology. Therefore, the total number of possible combination in $\mathcal{A}^c$ is $|\mathcal{A}^c| = 4 \times 4 \times 4 = 64$.

Given a tuple $\mathbf{a} \in \mathcal{A}^c$, the size of the subset $\mathcal{S}^c_{\mathbf{a}}$ is m, with $m \in \mathbb{R}^+$. The value m represents the number of scenarios analysed for each penetration tuple $\mathbf{a}$. In this study, m is chosen as 100 as it is a good trade-off between the number of scenarios tested and the total computation burden.

As the number of considered scenarios is restricted, each customer $c \in \mathcal{C}$ is assigned a probability for any technology $h \in \mathcal{H}$. This probability represents the likelihood of the customer c of installing the technology h. When building a scenario $\mathbf{s} \in \mathcal{S}^c_{\mathbf{a}}$, the probabilities are used to decide which customer will install a particular technology. The probabilities are chosen proportional to the household size as considering the household size enables the accounting for the socio-economic aspect of buying a new technology and reflects the link between the household size and the installation of new technologies. Larger households are generally associated with higher income levels, and studies suggest these customers are more likely to adopt these technologies [119].

Table 6.3: Probabilities of installing any type of technology per household size with the corresponding assumption on the number of people per household.

People per Household	**Household size ($\mathbf{m^2}$)**	**Probability (%)**
1	[0,50)	20
2	[50, 75)	40
3	[75, 100)	50
4	[100, 125)	60
5 or +	[125, Inf)	80

Five groups of household size are considered: a person living alone, referred to as isolated; two-people household, three-people household, four-people household, and five- or-more-people household. The distribution of customers in $\mathcal{C}$ in these categories is shown in Fig. 6.3. Table 6.3 gives

the number of people per household size and the associated probabilities of installation. In this work, in the absence of specific supporting data, for a given customer, the probabilities of installing each technology are assumed equal.

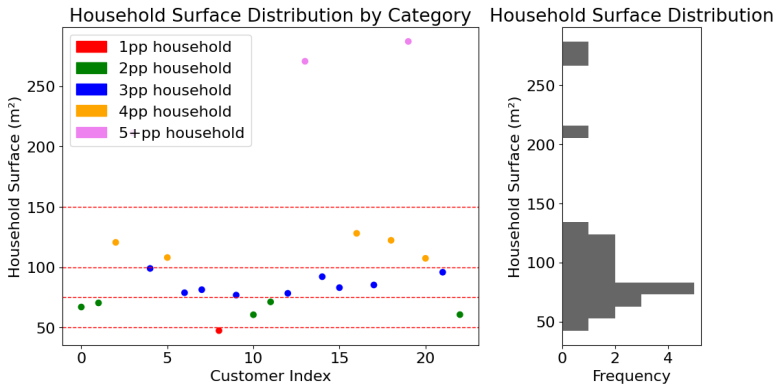


Figure 6.3: Customers to household size distribution.

6.4.4 Exogenous data

Exogenous data gathers all the necessary input data, except the network itself, needed to compute the HC. The main exogenous data is the customers' load profiles and the new technologies' profiles. Both customer load profiles and technology profiles were provided by RESA. Customer load profiles are directly taken from SM data while technology profiles are derived from raw data and predictions done using the Sirius tool [120].

Simulating an entire year for each penetration with a 15-minute granularity for all considered scenarios would be computationally challenging. To reduce this burden while still using time-series to allow one to compute a relevant hosting capacity, one strategy consists of reducing the window of time steps considered in a day, as in [68]. Another approach is to cluster the year into representative days and reduce the evaluated time steps to only these days. As only considering parts of the day does not reduce the workload sufficiently when considering PV-EV-HP, the representative days approach is used in this chapter to reduce the computational workload.

The representative days are the same as the Sirius study [120]. These are either days when the load is completely different from other days and thus do not represent many days in the year but rather an impactful day; or days when the load profile behaviour is highly common in the year.

In total, twelve days are designated as representative days. The number of days that are similar during the year to the considered representative days are presented in Tab. 6.4.

Table 6.4: Number of days in a year of 366 days that are similar to the considered representative days

Day	**# similar days**	**Day**	**# similar days**
1	33	**7**	20
2	42	**8**	42
3	13	**9**	23
4	24	**10**	3
5	51	**11**	10
6	55	**12**	50

6.4.5 Uncertainties

The process of computing the HC depends on future installations, which leads to uncertainties. Two categories of uncertainties for HC were introduced by [46]: epistemic uncertainties which are due to lack of knowledge and aleatory uncertainties which come from elements that are inherently stochastic [20]. In this study, uncertainties from both categories are considered. Table 6.5 gathers the uncertainties by technology type and if they are considered or not. In addition to these and as previously mentioned, the installation type, which is an epistemic uncertainty, is considered.

6.5 Implementation

The iterative implementation of HC analysis for the combined PV-EV-HP technologies of the Belgian network is showcased in the flowchart in Figure

Table 6.5: Summary of the considered uncertainties by technology type.

Installation type	Uncertainties		
	Epistemic		Aleatory
	Localisation	Size	Production/ consumption
PV	✓	✗	✓
EV	✓	✓	✗
HP	✓	✗	✗

6.4. The simulations were run using PandaPower [50].

The voltage thresholds used for the detection of issues are from the European EN50160 standard: $V_{min} = 0.95$ and $V_{max} = 1.05$ in Eq. (6.6). As the time step granularity is 15 min, the duration limit of EN50160 is not considered as it evaluates mean voltage over 10-minute periods. For the lines, an overload (OL) in Eq. 6.7 of at most $L_{\max} = 150\%$ of the nominal rate of the lines is allowed.

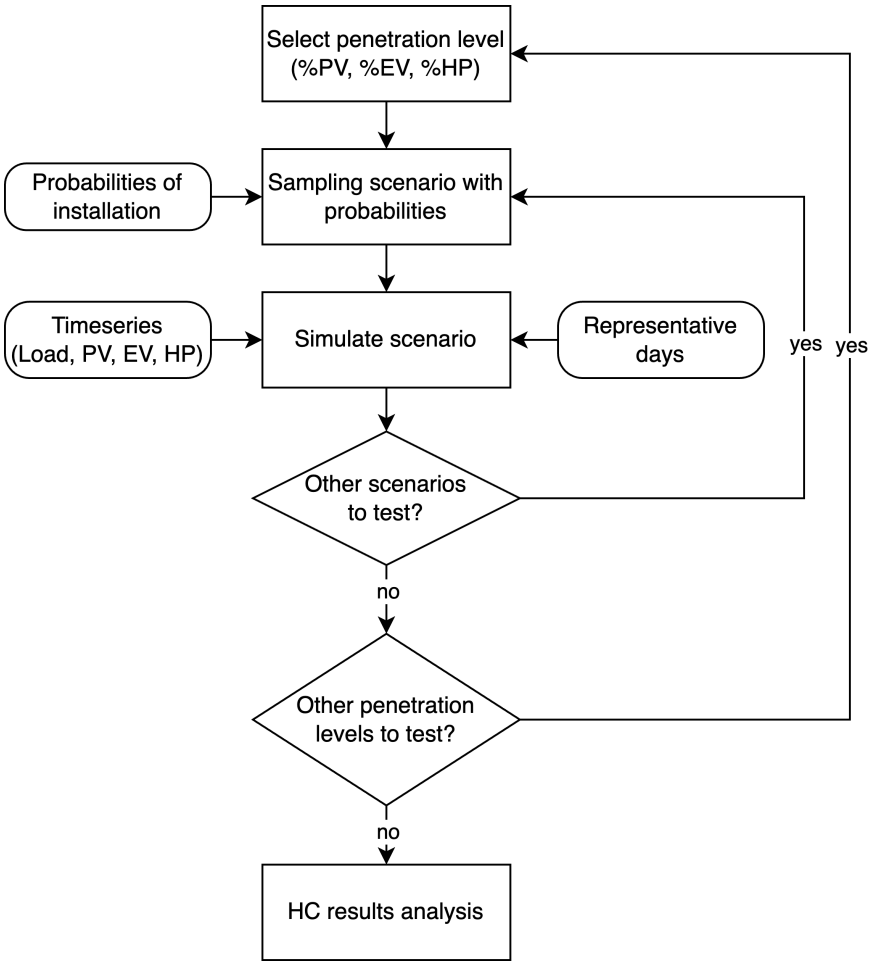


Figure 6.4: Flowchart describing the implementation of the presented HC analysis.

6.6 Results

The HC as defined in Eq. (6.9) is given as follows:

$$\begin{aligned}\text{HC} = \{&(25-50\%, 25\%, 25-100\%),\\ &(25\%, 100\%, 100\%),\\ &(25\%, 50\%, 25-100\%),\\ &(50\%, 75\%, 25-75\%),\\ &(25\%, 75-100\%, 25-50\%),\\ &(50\%, 100\%, 25-50\%),\\ &(50\%, 50\%, 25-100\%)\}.\end{aligned}$$

Recall that the penetration tuple is defined in Eq.(6.10) as (a_{PV}, a_{EV}, a_{HP}).

To analyse the HC and the different penetrations, a graphical approach is used as in [79] as it enables to show the combined penetrations. For readability, all voltage points for all customers and time-steps given a penetration are plot on Fig. 6.5 on a hypothesis space of continuous probability density functions (PDFs). The same is done for line loadings in Fig. 6.6. As anticipated based on the findings of other studies, the lines are not overloaded for all the PV penetrations. There is also no overloading for all EV and HP penetrations, which was not anticipated. It is worth to mention that the transformer load does not even reach 40% of its rated capacity.

The voltage level constraints are, as usual, more limiting. OV is not encountered before 75% of PV penetration. Under-voltage is encountered rapidly for both EV and HP penetrations higher than 50%.

DSOs may decide to accept some scenarios with issues to increase the HC since, from the results, it appears that most tuples are just outside the limits for a small number of scenarios. For example, the tuple $(25\%; 75\%, 75\%)$ presents a small amount of scenarios with a slight UV. Indeed, considering the median or the third percentile, shown in the boxplot, would lead to a bigger hosting capacity set with higher penetrations allowed.

The granularity of 25% for the penetration rate in this case study was arbitrarily chosen to provide a broad overview of the network's hosting capacity. However, this can be further refined with smaller increments to examine specific areas of interest more closely. Additional simulations with finer granularity around key regions can offer deeper insights and allow for more precise assessments where needed.

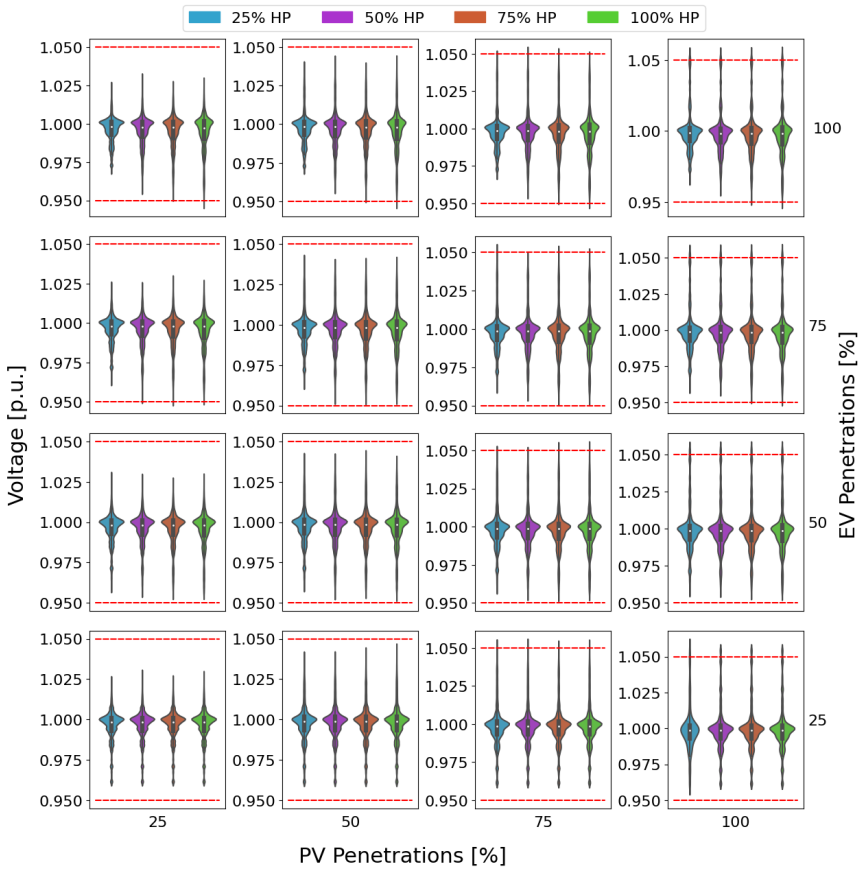


Figure 6.5: HC analysis for bus voltages using violin plots. Each violin plot represents a combination of the three penetration rates (PV, EV, HP). The upper and lower red-dashed lines represent the maximal and minimal voltage thresholds.

6.7 Intermediate conclusion

This chapter presents a combined PV-EV-HP HC analysis, following the formalisation [20] of the HC presented in Chapter 3, on a Belgian network. The analysis conducted with the real data available to the DSO and the usage of probabilities of installation for each technology allowed for a more realistic evaluation of the HC. The results obtained show that the network allows the installation of up to 50% of penetrations for every technology,

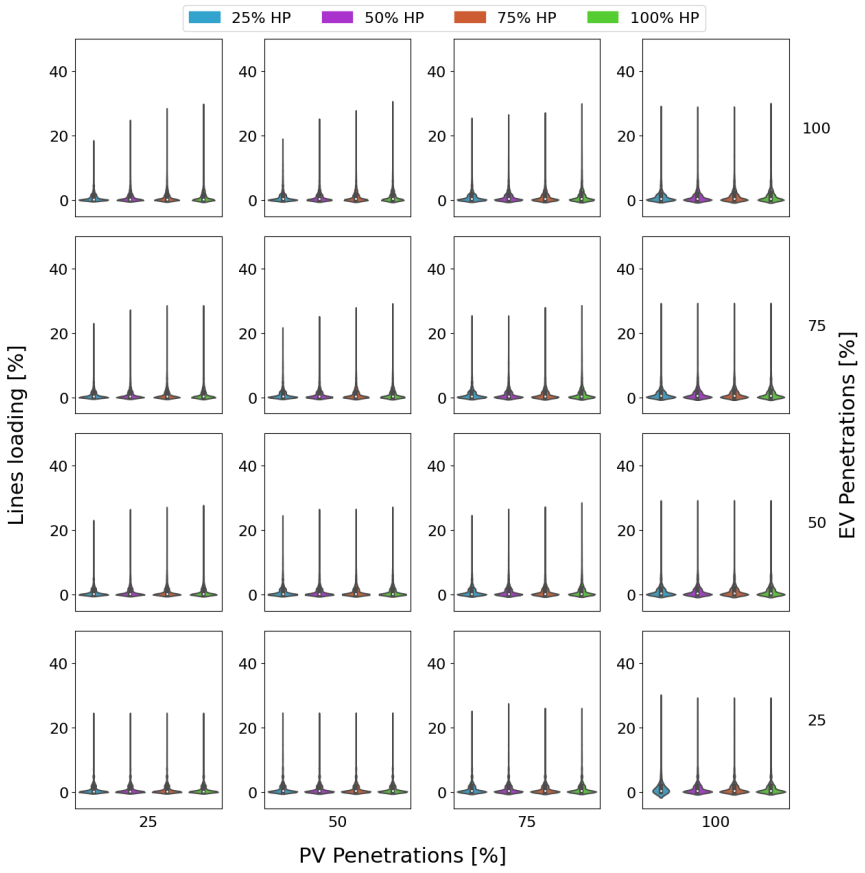


Figure 6.6: HC analysis for line loadings using violin plot. Each violin plot represents a combination of the three penetration rates (PV, EV, HP). The loading threshold is not represented as the values are far from violating it.

while 75% penetration rates are still acceptable if the DSO accepts a number of scenarios with some issues in the network.

To extend this work, considering quasi-static time-series (less the 10 min) would allow one to perform a dynamic hosting capacity analysis, taking into account the duration of the issues. Furthermore, another set of representative days as well as a thoroughly studied motivation behind their selection would lead to a baseline that could validate the obtained results for the considered days. This means carefully considering different factors, such as typical weather conditions, energy consumption patterns, and any

seasonal variations, to ensure that these days reflect a range of real-world situations.

Pairing the present work with network management techniques that are used to enhance the hosting capacity should be considered as an extension. Indeed, this extension could use techniques such as dynamic operating envelopes (DOEs) that manages the operation of distributed energy resources (DERs) [121] or active network management (ANM), which manages the overall network infrastructure [122]. Enhancing the hosting capacity with these techniques would increase the resilience, reliability, and overall efficiency of the distribution networks, with a better energy management in the face of increasing renewable penetration.

Finally, collecting socio-economic data would significantly enhance the ability to model and implement more realistic probabilities when attributing installations to different customer segments. By incorporating factors such as income levels, property types, energy consumption patterns, and regional demographics, it becomes possible to better understand which groups are more likely to adopt distributed energy resources. This data-driven approach allows for the creation of more accurate, granular models that reflect the diversity of customer behaviour and financial capabilities.

Chapter 7

Conclusion

Knowledge has to be improved, challenged, and increased constantly, or it vanishes.

Peter Drucker

7.1 Summary of the contributions

Distribution networks are ageing and the climate transition is right at the corner. As distributed energy resources (DERs), such as photovoltaic units (PVs), electric vehicles (EVs), and heat pumps (HPs), are rapidly deployed, they pose operational challenges for distribution system operators (DSOs), such as voltage regulation and network congestion. The maximal amount of DERs that can be adopted in a network without causing these operational issues is referred to as the hosting capacity (HC). The task of determining the hosting capacity became essential for DSOs to ensure the reliability of the grid.

This thesis focused on addressing the challenges related to the integration of DERs in distribution networks by exploring methods to evaluate and improve the hosting capacity. Five contributions were presented that address these challenges. In Chapter 2, the focus was on topology identification, a crucial step for accurately computing hosting capacity. By identifying the network structure, DSOs are better equipped to assess the potential impacts of new DER installations and determine the hosting capacity more effectively. The results demonstrated that the network model reconstruction algorithm, even with limited observability, could estimate a reliable network model. The voltage time series obtained through load flow

analysis on this estimated model closely matched the values from the initial network model.

Chapter 3 introduced a formal definition of hosting capacity as a unified framework. This chapter established the theoretical foundation for understanding HC and created a standardised approach to measure and compare the network's ability to integrate DERs. The deterministic definition was presented first, followed by methods for incorporating uncertainties to account for stochastic elements in the network. This formalism was complemented by a systematic review of the hosting capacity literature, demonstrating that the framework can accommodate and unify existing approaches. Reviewing past works shows that the definition offers a clear and consistent way to assess and compare hosting capacity across different contexts.

The following three chapters presented methods for computing hosting capacity and explored their applications through case studies. Chapter 4 presented a method calculating the HC for a single DER using probabilistic performance indicators, considering the variability and uncertainty of new DER installations in the network. By analysing numerous configurations, this approach provided a comprehensive assessment of how different DERs could affect network performance. The case study is conducted using the network reconstructed in Chapter 2. The obtained results confirmed that the most limiting aspect of the hosting capacity is the voltage levels, i.e. over- and under-voltage. The proposed method also integrates the risk tolerance on issues caused by DERs accepted by the DSOs and highlights the non-negligible influence of this risk tolerance on the HC.

Chapter 5 presented a two-step method to study the influence of phase selection on hosting capacity. This approach optimised the connection phase for each new DER and assessed the cost-effectiveness of modifying the phase connections of existing DERs. This method was applied on a test case using two types of DERs: PVs and EVs. The LV network simulation results showed that optimal phase selection significantly reduced curtailment and over-voltage issues, lowering operational costs for DSOs. The method also demonstrated a positive impact on imbalance and active power demand, highlighting the benefits of phase optimisation for LV networks.

Finally, Chapter 6 investigated the combined hosting capacity of PV, EV, and HP installations. This method incorporated real Belgian DSO data and probabilistic models to assess the likelihood of different technology combinations. Representative days were used to reduce computational time. This approach provided a more realistic evaluation of the network's hosting capacity, considering the interactions between different types of DERs. The results demonstrated how managing multiple DERs concur-

rently optimised network performance and increased hosting capacity.

7.2 Future works

This section highlights three possible innovative directions to extend the present work. Two of these directions address the issues related to the computational complexity of determining the hosting capacity: one by limiting the time dependency and the other by limiting the total number of scenarios considered. The third direction proposes a practical tool for DSOs, based on the hosting capacity framework, that combines investment strategies with active network management.

7.2.1 Representative days

One area of research that shows great potential is the development of methods to reduce the computational complexity in hosting capacity calculations. A key approach to achieving this is through identifying a smaller, representative set of days from the large volume of time series data. The challenge lies in choosing a reduced subset of days that still accurately reflects the full range of variability in the data.

A critical factor in HC calculations is the temporal correlation between DER outputs and customer loads. Maintaining the temporal correlation ensures that the interactions, such as how demand and generation fluctuate, together or independently, are accurately reflected even after reducing the dataset. For example, customer load may increase during certain hours of the day, while DER output, such as solar energy, peaks during daylight. Understanding and accounting for these time-based interactions is crucial when selecting representative days, as they directly affect the accuracy of HC estimation. Chapter 6 already applied this approach by selecting representative days to reduce computational demands. However, as discussed in the intermediary conclusion, there is still room for improvement in the method used to identify these key days.

Future research could focus on a thorough investigation of existing techniques for selecting representative days, comparing their effectiveness, and analysing how well they maintain the important characteristics of the original data set. Additionally, researches could evaluate whether these methods are applicable not only for HC estimation but also for long-term planning, such as forecasting investments in network upgrades. A valuable tool could be developed to automatically identify these meaningful days and update the selection as new data becomes available, ensuring continued model accuracy.

7.2.2 Socio-economic behaviour

Another area with great potential is the integration of socio-economic studies into HC calculations. By understanding how customers' behaviour influences electricity consumption and DER usage, it would be possible to further reduce computational time without sacrificing accuracy which could reduce the potential number of scenarios to evaluate. A detailed analysis of consumer behaviour patterns could lead to predictive tools capable of forecasting future behaviour shifts. This approach would not only enhance the precision of HC calculations but also enable DSOs to anticipate future challenges and develop more grid management strategies accordingly.

7.2.3 Coupling HC formalism with investment strategies and active network management

The third direction involves leveraging the HC formalism developed in this thesis to create a comprehensive industrial tool that integrates HC calculations with investment strategies and active network management (ANM). This platform could revolutionise how DSOs approach both short-term operational decisions and long-term planning. By coupling HC predictions with investment models, DSOs could optimise grid expansion or reinforcement strategies, ensuring that capital is allocated efficiently. In parallel, incorporating ANM would allow for real-time adjustments to DER integration, enhancing the grid's flexibility and operational efficiency.

Each of these research directions presents significant academic and industrial benefits, offering DSOs innovative tools to improve both the speed and accuracy of hosting capacity evaluations while facilitating more effective grid planning and management.

Bibliography

[1] C. Böttcher, *The cost of blackouts in europe*, 2016. [Online]. Available: https://cordis.europa.eu/article/id/126674-the-cost-of-blackouts-in-europe (visited on 10/09/2024).

[2] IEA, *Belgium energy mix*, 2024. [Online]. Available: https://www.iea.org/countries/belgium/energy-mix (visited on 10/07/2024).

[3] J. F. Adolfsen, F. Kuik, E. M. Lis, and T. Schuler, *The impact of the war in ukraine on euro area energy markets*, 2022. [Online]. Available: https://www.ecb.europa.eu/press/economic-bulletin/focus/2022/html/ecb.ebbox202204_01~68ef3c3dc6.en.html (visited on 10/08/2024).

[4] U. N. F. C. on Climate Change (UNFCCC), *Paris agreement*, Adopted on December 12, 2015, at the 21st Conference of the Parties (COP21) in Paris, France., 2015. [Online]. Available: https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement.

[5] H. Ritchie, M. Roser, and P. Rosado, *Renewable energy*, 07-10-2024, 2020. [Online]. Available: https://ourworldindata.org/renewable-energy.

[6] IEA, *Electric car sales, 2012-2024*, 2024. [Online]. Available: https://www.iea.org/data-and-statistics/charts/electric-car-sales-2012-2024 (visited on 10/07/2024).

[7] E. H. P. Association, "Pump it down: Why heat pump sales dropped in 2023," EHPA, Tech. Rep., 2024.

[8] L. Rodríguez, "Evolution of renewable energy: How energy use has changed over time," *Rated Power*, 2020, 07-10-2024. [Online]. Available: https://ratedpower.com/blog/evolution-renewable-energy/.

[9] A Muir and J Lopatto, "Final report on the august 14, 2003 blackout in the united states and canada: Causes and recommendations," U.S.-Canada Power System Outage Task Force, Ottawa, ON (Canada), Tech. Rep., 2004.

[10] F. Dewangan, S. Siddiqui, M. Biswal, and V. K. Sood, "Smart meters in smart grid," in *Smart Metering*, Elsevier, 2024, pp. 1–37.

[11] Elia, *Production photovoltaïque*, `https://www.elia.be/fr/donnees-de-reseau/production/production-photovoltaique`, 2020.

[12] Cigré, *The electric power system - belgium*, Accessed on Septembrer 30th, 2024, 2014. [Online]. Available: `https://www.cigre.org/userfiles/files/Community/NC/Belgium_The_Electric_Power_System.pdf`.

[13] T. V. Cutsem, *An introductory overview of electric power systems*, 30-09-2024, 2019. [Online]. Available: `https://thierryvancutsem.github.io/home/elec0014/Intro_ELEC0014.pdf`.

[14] R. Bowers, E. Fasching, and K. Antonio, *As solar capacity grows, duck curves are getting deeper in california*, 2023. [Online]. Available: `https://www.eia.gov/todayinenergy/detail.php?id=56880` (visited on 10/07/2024).

[15] R. Ahmed, V. Sreeram, Y. Mishra, and M. Arif, "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization," *Renewable and Sustainable Energy Reviews*, vol. 124, p. 109 792, 2020.

[16] Karel, *Duck curve*, 2023. [Online]. Available: `https://groenerleven.be/2023/08/18/duck-curve/` (visited on 10/08/2024).

[17] A. Benzerga, D. Maruli, A. Sutera, A. Bahmanyar, S. Mathieu, and D. Ernst, "Low-voltage network topology and impedance identification using smart meter measurements," in *Proceedings of the 2021 IEEE Madrid PowerTech*, IEEE, 2021, pp. 1–5.

[18] D. Marulli, S. Mathieu, A. Benzerga, A. Sutera, and D. Ernst, "Reconstruction of low-voltage networks with limited observability," in *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, IEEE, 2021, pp. 1–5.

[19] A. Benzerga, A. Bahmanyar, S. Mathieu, and D. Ernst, "Smart meter data analytics case study: Identification of LV distribution network topology to design optimal planning solutions 7," in *Applications of Big Data and Artificial Intelligence in Smart Energy Systems*, River Publishers, 2023, pp. 161–192.

[20] A. Benzerga, A. Bahmanyar, G. Derval, and D. Ernst, *A unified definition of hosting capacity, applications and review*, 2024.

[21] A. Benzerga, S. Mathieu, A. Bahmanyar, and D. Ernst, "Probabilistic capacity assessment for three-phase low-voltage distribution networks," in *2021 IEEE 15th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, 2021, pp. 1–6. DOI: 10.1109CPE-POWERENG50821.2021.9501226.

[22] A Benzerga, S Gerard, S Lachi, Q Garnier, A Bahmanyar, and D Ernst, "Optimal connection phase selection for single-phase electrical vehicle chargers," in *CIRED Porto Workshop 2022: E-mobility and power distribution systems*, IET, vol. 2022, 2022, pp. 508–512.

[23] A. Benzerga, A. Bahmanyar, and D. Ernst, "Optimal connection phase selection of residential distributed energy resources and its impact on aggregated demand," in *11TH BULK POWER SYSTEMS DYNAMICS AND CONTROL SYMPOSIUM (IREP-2022)*, IREP, 2022, arXiv–2207.

[24] M. Vassallo, A. Benzerga, A. Bahmanyar, and D. Ernst, "Fair reinforcement learning algorithm for PV active control in LV distribution networks," in *2023 International Conference on Clean Electrical Power (ICCEP)*, IEEE, 2023, pp. 796–802.

[25] V. Dachet, A. Benzerga, R. Fonteneau, and D. Ernst, "Towards co2 valorization in a multi remote renewable energy hub framework," *arXiv preprint arXiv:2303.09454*, 2023.

[26] D. Victor, B. Amina, C. Diederik, C. Francesco, F. Raphaël, and E. Damien, "Towards co2 valorization in a multi remote renewable energy hub framework with uncertainty quantification," *Journal of Environmental Management*, vol. 363, p. 121 262, 2024.

[27] J. Mbenoun, A. Benzerga, B. Miftari, G. Detienne, T. Deschuyteneer, J. Vazquez, G. Derval, and D. Ernst, "Integration of offshore energy into national energy system: A case study on belgium," 2024.

[28] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, "Identification of power distribution network topology via voltage correlation analysis," *52nd IEEE Conference on Decision and Control*, pp. 1659–1664, 2013.

[29] Y. Weng, Y. Liao, and R. Rajagopal, "Distributed energy resources topology identification via graphical modeling," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2682–2694, 2016.

[30] S. J. Pappu, N. Bhatt, R. Pasumarthy, and A. Rajeswaran, "Identifying topology of low voltage distribution networks based on smart meter data," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5113–5122, 2017.

[31] K. Soumalas, G. Messinis, and N. Hatziargyriou, "A data driven approach to distribution network topology identification," in *2017 IEEE Manchester PowerTech*, IEEE, 2017, pp. 1–6.

[32] Y. Yuan, S. Low, O. Ardakanian, and C. Tomlin, "Inverse power flow problem," *arXiv preprint arXiv:1610.06631*, 2016.

[33] O. Ardakanian, V. W. Wong, R. Dobbe, S. H. Low, A. von Meier, C. J. Tomlin, and Y. Yuan, "On identification of distribution grids," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 3, pp. 950–960, 2019.

[34] J. Yu, Y. Weng, and R. Rajagopal, "Patopa: A data-driven parameter and topology joint estimation framework in distribution grids," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4335–4347, 2017.

[35] F. Olivier, A. Sutera, P. Geurts, R. Fonteneau, and D. Ernst, "Phase identification of smart meters by clustering voltage measurements," *2018 Power Systems Computation Conference*, 2018.

[36] T. F. Abdelmaguid, "An efficient mixed integer linear programming model for the minimum spanning tree problem," *Mathematics 2018*, 2018.

[37] U. P. Networks, *Smartmeter energy consumption data in london households*, `https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households`.

[38] E. P. R. I. (EPRI), *Opendss*, `https://www.epri.com/pages/sa/opendss`.

[39] IEA, *Renewables 2022*, 2022. [Online]. Available: `https://www.iea.org/reports/renewables-2022` (visited on 07/20/2023).

[40] M. Zain ul Abideen, O. Ellabban, and L. Al-Fagih, "A review of the tools and methods for distribution networks' hosting capacity calculation," *Energies*, vol. 13, no. 11, p. 2758, 2020.

[41] S. Fatima, V. Püvi, and M. Lehtonen, "Review on the PV hosting capacity in distribution networks," *Energies*, vol. 13, no. 18, p. 4756, 2020.

[42] S. Fatima, V. Püvi, M. Lehtonen, and M. Pourakbari-Kasmaei, "A review of electric vehicle hosting capacity quantification and improvement techniques for distribution networks," *IET Generation, Transmission & Distribution*, 2023.

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

[44] A Kharrazi, V Sreeram, and Y Mishra, "Assessment techniques of the impact of grid-tied rooftop photovoltaic generation on the power quality of low voltage distribution network-a review," *Renewable and Sustainable Energy Reviews*, vol. 120, p. 109 643, 2020.

[45] S Mina Mirbagheri, D. Falabretti, V. Ilea, M. Merlo, *et al.*, "Hosting capacity analysis: A review and a new evaluation method in case of parameters uncertainty and multi-generator," in *Proceedings-2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe, EEEIC/I and CPS Europe 2018*, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 1–6.

[46] E. Mulenga, M. H. Bollen, and N. Etherden, "A review of hosting capacity quantification methods for photovoltaics in low-voltage distribution grids," *International Journal of Electrical Power & Energy Systems*, vol. 115, p. 105 445, 2020.

[47] A Rajabi, S Elphick, J David, A Pors, and D Robinson, "Innovative approaches for assessing and enhancing the hosting capacity of PV-rich distribution networks: An australian perspective," *Renewable and Sustainable Energy Reviews*, vol. 161, p. 112 365, 2022.

[48] U. Singh, "The role of hosting capacity study in power system advancements: A review," *arXiv preprint arXiv:2301.04765*, 2023.

[49] M. Bollen and M. Häger, "Power quality: Interactions between distributed energy resources, the grid, and other customers," *Leonardo Energy*, 2005.

[50] L. Thurner, A. Scheidler, F. Schafer, J. H. Menke, J. Dollichon, F. Meier, S. Meinecke, and M. Braun, "Pandapower - an open source python tool for convenient modeling, analysis and optimization of electric power systems," *IEEE Transactions on Power Systems*, 2018, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2018.2829021. [Online]. Available: https://arxiv.org/abs/1709.06743.

[51] A. Dubey, S. Santoso, and A. Maitra, "Understanding photovoltaic hosting capacity of distribution circuits," in *2015 IEEE power & energy society general meeting*, IEEE, 2015, pp. 1–5.

[52] M Rylander and J Smith, "Stochastic analysis to determine feeder hosting capacity for distributed solar PV," EPRI, Tech. Rep. 1026640, 2012.

[53] F. Capitanescu, L. F. Ochoa, H. Margossian, and N. D. Hatziargyriou, "Assessing the potential of network reconfiguration to improve distributed generation hosting capacity in active distribution systems," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 346–356, 2014.

[54] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Transactions on Power delivery*, vol. 4, no. 2, pp. 1401–1407, 1989.

[55] M. H. Bollen and S. K. Rönnberg, "Hosting capacity of the power grid for renewable electricity production and new large consumption equipment," *Energies*, vol. 10, no. 9, p. 1325, 2017.

[56] S. Zhan, T. Gu, W. van den Akker, W. Brus, A. van der Molen, and J. Morren, "Towards congestion management in distribution networks: A dutch case study on increasing heat pump hosting capacity," in *12th IET International Conference on Advances in Power System Control, Operation and Management, APSCOM 2022*, Institution of Engineering and Technology (IET), 2023, pp. 364–369.

[57] B. E. Carmelito and J. M. d. C. Filho, "Hosting capacity of electric vehicles on LV/MV distribution grids—a new methodology assessment," *Energies*, vol. 16, no. 3, p. 1509, 2023.

[58] F. Ding and B. Mather, "On distributed PV hosting capacity estimation, sensitivity study, and improvement," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 3, pp. 1010–1020, 2016.

[59] F. Ding, B. Mather, N. Ainsworth, P. Gotseff, and K. Baker, "Locational sensitivity investigation on PV hosting capacity and fast track PV screening," in *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, IEEE, 2016, pp. 1–5.

[60] F. Ding, B. Mather, and P. Gotseff, "Technologies to increase PV hosting capacity in distribution feeders," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, IEEE, 2016, pp. 1–5.

[61] S. Jothibasu, S. Santoso, and A. Dubey, "Optimization methods for evaluating PV hosting capacity of distribution circuits," in *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*, IEEE, 2019, pp. 0887–0891.

[62] S. Wang, S. Chen, L. Ge, and L. Wu, "Distributed generation hosting capacity evaluation for distribution systems considering the robust optimal operation of oltc and svc," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1111–1123, 2016.

[63] N. Tang and G. Chang, "A stochastic approach for determining PV hosting capacity of a distribution feeder considering voltage quality constraints," in *2018 18th International Conference on Harmonics and Quality of Power (ICHQP)*, IEEE, 2018, pp. 1–5.

[64] O. Ceylan, S. Paudyal, B. P. Bhattarai, and K. S. Myers, "Photovoltaic hosting capacity of feeders with reactive power control and tap changers," in *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, Ieee, 2017, pp. 1–6.

[65] A. Dubey and S. Santoso, "On estimation and sensitivity analysis of distribution circuit's photovoltaic hosting capacity," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2779–2789, 2016.

[66] X. Chen, W. Wu, B. Zhang, and C. Lin, "Data-driven dg capacity assessment method for active distribution networks," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3946–3957, 2016.

[67] A. Arshad, M. Lindner, and M. Lehtonen, "An analysis of photovoltaic hosting capacity in finnish low voltage distribution networks," *Energies*, vol. 10, no. 11, p. 1702, 2017.

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

[69] T. Stetz, F. Marten, and M. Braun, "Improved low voltage grid-integration of photovoltaic systems in germany," *IEEE Transactions on sustainable energy*, vol. 4, no. 2, pp. 534–542, 2012.

[70] P. H. Divshali and L. Söder, "Improving PV hosting capacity of distribution grids considering dynamic voltage characteristic," in *2018 Power Systems Computation Conference (PSCC)*, IEEE, 2018, pp. 1–7.

[71] A. Y. Saber, T. Khandelwal, and A. K. Srivastava, "Fast feeder PV hosting capacity using swarm based intelligent distribution node selection," in *2019 IEEE Power & Energy Society General Meeting (PESGM)*, IEEE, 2019, pp. 1–5.

[72] H. V. Padullaparti, S. Jothibasu, S. Santoso, and G. Todeschini, "Increasing feeder PV hosting capacity by regulating secondary circuit voltages," in *2018 IEEE Power & Energy Society General Meeting (PESGM)*, IEEE, 2018, pp. 1–5.

[73] J Smith and M Rylander, "Stochastic analysis to determine feeder hosting capacity for distributed solar PV," *Electric Power Research Inst., Palo Alto, CA, Tech. Rep*, vol. 1026640, pp. 0885–8950, 2012.

[74] J. Hu, M. Marinelli, M. Coppo, A. Zecchino, and H. W. Bindner, "Coordinated voltage control of a decoupled three-phase on-load tap changer transformer and photovoltaic inverters for managing unbalanced networks," *Electric Power Systems Research*, vol. 131, pp. 264–274, 2016.

[75] A. K. Jain, K. Horowitz, F. Ding, N. Gensollen, B. Mather, and B. Palmintier, "Quasi-static time-series PV hosting capacity methodology and metrics," in *2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, IEEE, 2019, pp. 1–5.

[76] S. Jothibasu, S. Santoso, and A. Dubey, "Determining PV hosting capacity without incurring grid integration cost," in *2016 North American Power Symposium (NAPS)*, IEEE, 2016, pp. 1–5.

[77] A. Navarro-Espinosa and L. F. Ochoa, "Probabilistic impact assessment of low carbon technologies in LV distribution systems," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2192–2203, 2015.

[78] N. Jayasekara, M. A. Masoum, and P. J. Wolfs, "Optimal operation of distributed energy storage systems to improve distribution network load and generation hosting capability," *IEEE Transactions on sustainable energy*, vol. 7, no. 1, pp. 250–261, 2015.

[79] R. Fachrizal, U. H. Ramadhani, J. Munkhammar, and J. Widén, "Combined PV–EV hosting capacity assessment for a residential LV distribution grid with smart EV charging and PV curtailment," *Sustainable Energy, Grids and Networks*, vol. 26, p. 100 445, 2021.

[80] M. Rylander, J. Smith, W. Sunderman, S. Uppalapati, and H. Li, "Distributed photovoltaic feeder analysis: Preliminary findings from hosting capacity analysis of 18 distribution feeders.," EPRI, Tech. Rep. 3002001245, 2013.

[81] O. Lennerhag, G. Pinares, M. H. Bollen, G. Foskolos, and T. Gafurov, "Performance indicators for quantifying the ability of the grid to host renewable electricity production," *CIRED-Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 792–795, 2017.

[82] European Parliament, Council of the European Union, *Directive (EU) no 2009/33/ec*, `https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32009L0033&from=EN`, 2009.

[83] Council of European Union, *Directive (EU) no 2019/1161*, `https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019L1161&from=EN`, 2019.

[84] S. Cundeva, A. K. Mateska, and M. H. Bollen, "Hosting capacity of LV residential grid for uncoordinated EV charging," in *2018 18th International Conference on Harmonics and Quality of Power (ICHQP)*, IEEE, 2018, pp. 1–5.

[85] J. Quirós-Tortós, L. F. Ochoa, A. Navarro-Espinosa, M. Gillie, and R. Hartshorn, "Probabilistic impact assessment of electric vehicle charging on residential uk LV networks," in *CIRED*, 2015, pp. 1–5.

[86] Gysel, Pieter, *Combien de panneaux solaires me faut-il et que vont-ils me coûter ?* `https://blog.eneco.be/fr/soleil/combien-panneaux-solaires/`, 2018.

[87] C. Gaete-Morales, H. Kramer, W.-P. Schill, and A. Zerrahn, "An open tool for creating battery-electric vehicle time series from empirical data: Emobpy," *arXiv preprint arXiv:2005.02765*, 2020.

[88] K. Ma, L. Fang, and W. Kong, "Review of distribution network phase unbalance: Scale, causes, consequences, solutions, and future research directions," *CSEE Journal of Power and Energy systems*, vol. 6, no. 3, pp. 479–488, 2020.

[89] E. N. Silva, A. B. Rodrigues, and M. d. G. da Silva, "Stochastic assessment of the impact of photovoltaic distributed generation on the power quality indices of distribution networks," *Electric Power Systems Research*, vol. 135, pp. 59–67, 2016.

[90] T. Stetz, J. von Appen, F. Niedermeyer, G. Scheibner, R. Sikora, and M. Braun, "Twilight of the grids: The impact of distributed solar on germany?s energy transition," *IEEE Power and Energy Magazine*, vol. 13, no. 2, pp. 50–61, 2015.

[91] V. Klonari, B. Meersman, D. Bozalakov, T. L. Vandoorn, L. Vandevelde, J. Lobry, and F. Vallée, "A probabilistic framework for evaluating voltage unbalance mitigation by photovoltaic inverters," *Sustainable Energy, Grids and Networks*, vol. 8, pp. 1–11, 2016.

[92] M. Alturki and A. Khodaei, "Increasing distribution grid hosting capacity through optimal network reconfiguration," in *2018 North American Power Symposium (NAPS)*, 2018, pp. 1–6.

[93] C. Linn, C. Abbey, H. Gil, and S. Kahrobaee, "Enhancing distribution system hosting capacity through active network management," in *2018 IEEE Conference on Technologies for Sustainability (SusTech)*, 2018, pp. 1–6.

[94] F. Olivier, R. Fonteneau, S. Mathieu, and D. Ernst, "Distributed control of photovoltaic units in unbalanced LV distribution networks to prevent overvoltages," in *2018 IEEE International Conference on Smart Energy Grid Engineering (SEGE)*, 2018, pp. 362–370.

[95] M. Seydali Seyf Abad and J. Ma, "Photovoltaic hosting capacity sensitivity to active distribution network management," *IEEE Transactions on Power Systems*, vol. 36, no. 1, pp. 107–117, 2021.

[96] M. D. Hraiz, J. A. M. García, R. J. Castañeda, and H. Muhsen, "Optimal PV size and location to reduce active power losses while achieving very high penetration level with improvement in voltage profile using modified jaya algorithm," *IEEE Journal of Photovoltaics*, vol. 10, no. 4, pp. 1166–1174, 2020.

[97] A. Bogyrbayeva, S. Jang, A. Shah, Y. J. Jang, and C. Kwon, "A reinforcement learning approach for rebalancing electric vehicle sharing systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 8704–8714, 2021.

[98] S. Chen, Z. Guo, Z. Yang, Y. Xu, and R. S. Cheng, "A game theoretic approach to phase balancing by plug-in electric vehicles in the smart grid," *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 2232–2244, 2019.

[99] N. Jabalameli, M. A. Masoum, and S. Deilami, "Optimal online charging of plug-in electric vehicles considering voltage unbalance factor," in *2017 IEEE Power & Energy Society General Meeting*, IEEE, 2017, pp. 1–5.

[100] M. Spitzer, J. Schlund, E. Apostolaki-Iosifidou, and M. Pruckner, "Optimized integration of electric vehicles in low voltage distribution grids," *Energies*, vol. 12, no. 21, p. 4059, 2019.

[101] A. Kharrazi, V. Sreeram, and Y. Mishra, "Assessment of voltage unbalance due to single phase rooftop photovoltaic panels in residential low voltage distribution network: A study on a real LV network in western australia," in *2017 Australasian Universities Power Engineering Conference (AUPEC)*, IEEE, 2017, pp. 1–6.

[102] C.-H. Lin, C.-S. Chen, H.-J. Chuang, M.-Y. Huang, and C.-W. Huang, "An expert system for three-phase balancing of distribution feeders," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1488–1496, 2008.

[103] G. Grigoraș, B.-C. Neagu, M. Gavrilaș, I. Triștiu, and C. Bulac, "Optimal phase load balancing in low voltage distribution networks using a smart meter data-based algorithm," *Mathematics*, vol. 8, no. 4, p. 549, 2020.

[104] S. Soltani, M Rashidinejad, and A Abdollahi, "Dynamic phase balancing in the smart distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 93, pp. 374–383, 2017.

[105] V. Poulios, E. Vrettos, F. Kienzle, E. Kaffe, H. Luternauer, and G. Andersson, "Optimal placement and sizing of battery storage to increase the PV hosting capacity of low voltage grids," in *International ETG Congress 2015*, 2015, pp. 1–8.

[106] X. Xu, Z. Xu, J. Li, J. Zhao, and L. Xue, "Optimal placement of voltage regulators for photovoltaic hosting capacity maximization," in *2018 IEEE Innovative Smart Grid Technologies - Asia*, 2018, pp. 1278–1282.

[107] S. Sakar, M. E. Balci, S. H. Abdel Aleem, and A. F. Zobaa, "Increasing PV hosting capacity in distorted distribution systems using passive harmonic filtering," *Electric Power Systems Research*, vol. 148, pp. 74–86, 2017.

[108] B. Liu, K. Meng, Z. Y. Dong, P. K. Wong, and w. Wei, "Optimal placement of phase-reconfiguration devices in low-voltage distribution network with residential PV generation," *IEEE Journal of Photovoltaics*, vol. 14, no. 18, pp. 3752–3761, 2020.

[109] *Electricity price statistics*, 2021. [Online]. Available: `httpsec.europa.eueurostatstatistics-explainedindex.phptitle=Electricity_price_statistics#Electricity_prices_for_household_consumers`.

[110] O. Bozorg-Haddad, M. Solgi, and H. A. Loàiciga, *Meta-heuristic and evolutionary algorithms for engineering optimization*. John Wiley & Sons, 2017.

[111] M. Dong and L. Grumbach, "A hybrid distribution feeder long-term load forecasting method based on sequence prediction," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 470–482, 2019.

[112] N. Son and M. Jung, "Analysis of meteorological factor multivariate models for medium-and long-term photovoltaic solar power forecasting using long short-term memory," *Applied Sciences*, vol. 11, no. 1, p. 316, 2020.

[113] S. Ding, R. Li, and Z. Tao, "A novel adaptive discrete grey model with time-varying parameters for long-term photovoltaic power generation forecasting," *Energy Conversion and Management*, vol. 227, p. 113 644, 2021.

[114] A. Navarro-Espinosa and L. F. Ochoa, "Probabilistic impact assessment of low carbon technologies in LV distribution systems," *IEEE Transactions on Power Systems*, vol. 31, pp. 2192–2203, 2016. DOI: 10.1109/TPWRS.2015.2448663.

[115] L. Mehigan, M. Zehir, J. Cuenca, I. Şengör, C. Geaney, and B. Hayes, "Synergies between low carbon technologies in a large-scale MV/LV distribution system," *IEEE Access*, vol. 10, pp. 88655 – 88 666, 2022.

[116] G. Viganò, D. Clerici, C. Michelangeli, D. Moneta, A. Bosisio, A. Morotti, B. Greco, and P. Caterina, "Energy transition through PVs, EVs, and HPs: A case study to assess the impact on the brescia distribution network," in *2021 AEIT International Annual Conference (AEIT)*, IEEE, 2021, pp. 1–6.

[117] C. Edmunds, S. Galloway, J. Dixon, W. Bukhsh, and I. Elders, "Hosting capacity assessment of heat pumps and optimised electric vehicle charging on low- voltage networks," *Applied Energy*, vol. 298, p. 117 093, 2021.

[118] M. Vassallo, A. Bahmanyar, L. Duchesne, A. Leerschool, S. Gerard, T. Wehenkel, and D. Ernst, "A systematic procedure for topological path identification with raw data transformation in electrical distribution networks," in *7th International Conference on Energy, Electrical and Power Engineering (CEEPE 2024)*, CEEPE, Yangzhou, China, 2024.

[119] J. Böning, K. Bruninx, M. Ovaere, G. Pepermans, and E. Delarue, "The effectiveness of future financial benefits on PV adoption - evidence from belgium," *SSRN Electronic Journal*, Jan. 2023. DOI: 10.2139/ssrn.4627140.

[120] J. Vandeburie, T. Wehenkel, and S. Gerard, "Predicting local effects of energy transition through development of a network observation tool," in *27th International Conference on Electricity Distribution (CIRED 2023)*, IET, 2023, pp. 3625–3628.

[121] B. Liu and J. H. Braslavsky, "Robust dynamic operating envelopes for der integration in unbalanced distribution networks," *IEEE Transactions on Power Systems*, 2023.

[122] F. Olivier, P. Aristidou, D. Ernst, and T. Van Cutsem, "Active management of low-voltage networks for mitigating overvoltages due to photovoltaic units," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 926–936, 2015.