

Future Distribution Grids Using Geo-Referenced Synthetic Network Topologies

Abhilash Bandam

Energie & Umwelt / Energy & Environment

Band / Volume 647

ISBN 978-3-95806-790-5

Forschungszentrum Jülich GmbH
Institute of Climate and Energy Systems (ICE)
Jülicher Systemanalyse (ICE-2)

Future Distribution Grids Using Geo-Referenced Synthetic Network Topologies

Abhilash Bandam

Schriften des Forschungszentrums Jülich
Reihe Energie & Umwelt / Energy & Environment

Band / Volume 647

ISSN 1866-1793

ISBN 978-3-95806-790-5

Bibliografische Information der Deutschen Nationalbibliothek.
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte Bibliografische Daten
sind im Internet über <http://dnb.d-nb.de> abrufbar.

Herausgeber
und Vertrieb: Forschungszentrum Jülich GmbH
 Zentralbibliothek, Verlag
 52425 Jülich
 Tel.: +49 2461 61-5368
 Fax: +49 2461 61-6103
 zb-publikation@fz-juelich.de
 www.fz-juelich.de/zb

Umschlaggestaltung: Grafische Medien, Forschungszentrum Jülich GmbH

Druck: Grafische Medien, Forschungszentrum Jülich GmbH

Copyright: Forschungszentrum Jülich 2024

Schriften des Forschungszentrums Jülich
Reihe Energie & Umwelt / Energy & Environment, Band / Volume 647

D 82 (Diss. RWTH Aachen University, 2024)

ISSN 1866-1793
ISBN 978-3-95806-790-5

Vollständig frei verfügbar über das Publikationsportal des Forschungszentrums Jülich (JuSER)
unter www.fz-juelich.de/zb/openaccess.



This is an Open Access publication distributed under the terms of the [Creative Commons Attribution License 4.0](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Distributed solar rooftop photovoltaic generation continues to be the most viable and sustainable electricity generation for reducing greenhouse gas emissions in the power sector. As the potential to leverage distributed generation, energy-efficient distributed demands such as battery electric vehicles and heat pumps are gaining traction for their optimal use of the energy generated by distributed generators. With the integration of decentralized technologies rising substantially, the design and operation of distribution grids will be impacted in the future. Consequently, the future role of distributed demand and generation in distribution grids remains ambiguous. Hence, evaluating the readiness of current distribution networks for the imminent energy transition requires examining the impact of future distributed demand and generation on distribution grids. Because of the lack of access to real-world networks, evaluating the integration of demand and generation units into distribution grids is a pertinent obstacle to overcome. Thus, the purpose of this work is to develop geo-referenced synthetic networks to investigate the impacts of future distributed demand and generation on distribution grids. Moreover, this study also provides potential solutions for the identified issues, within the context of reducing the German energy system's carbon dioxide (CO₂) emissions by 95% by 2050.

This study examines how distributed demand, and generation might affect distribution system components and network vulnerability in German low-voltage distribution grids for the year 2050. This process led to the development of a classification model for classifying building types. The classification model, when applied to buildings in Germany which are retrieved from OpenStreetMap, revealed that 19,747,802 buildings out of 29,497,992 were classified as residential buildings with a percent error of 3.4%. In addition, 500,000 geo-referenced synthetic electrical low-voltage networks for Germany were synthesized using the classified residential buildings. Further, various operational limiting factor violations were recognized through the deployment of future distributed demand and generation with a base case scenario and scenario variations.

In terms of thermal overloading of the transformers, 75% of the networks operate within their design limits, while 25% of the low-voltage networks require additional attention. More than 75% of the networks experience reverse power flows, while 25% of 500,000 networks were identified as critical. Additionally, 11% of the power lines of 500,000 low-voltage networks are thermally overloaded. Undervoltage violations were identified in 34% of the networks' nodes. Nonetheless, these results are associated with the base case scenario. According to the sensitivity analysis, it is concluded that transformer overloading is a result of high heat pump penetration, particularly during the winter. Reverse power flows are primarily caused by high rooftop PV penetration, regardless of distributed demand, most notably in the spring and summer. These findings indicate that some of the 500,000 low-voltage networks will record certain violations with respect to future distributed demand and generation. However, a potential solution involves employing grid reinforcement measures for transformers and power lines, such as installing voltage regulating transformers in the networks with a high rate of undervoltage violations and reinforcing the affected 11% of power lines with new lines. Nevertheless, demand-side management options such as controlled battery electric vehicle charging, controlled heat pump utilization, and battery storage can help mitigate the need for these grid reinforcement measures.

Kurzfassung

Die dezentrale Stromerzeugung durch Dachflächenphotovoltaik ist eine der praktikabelsten und nachhaltigsten Arten der Stromerzeugung zur Verringerung der Treibhausgasemissionen im Energiesektor. Um die Nutzung von dezentralen Erzeugungspotenzialen zu ermöglichen, gewinnen energieeffiziente, dezentrale Nachfragen zum Beispiel durch batteriebetriebene Elektrofahrzeuge oder Wärmepumpen zunehmend an Bedeutung, da sie die dezentral erzeugte Energie optimal nutzen können. Die zunehmende Integration dezentraler Technologien wird sich in Zukunft auch auf die Auslegung und den Betrieb von Verteilnetzen auswirken. Um zu beurteilen, ob die derzeitigen Verteilnetze für die bevorstehende Energiewende gerüstet sind, müssen die Auswirkungen der künftigen verteilten Nachfrage und Erzeugung auf die Verteilnetze untersucht werden. Die Bewertung der Integration von dezentralen Nachfragern und Erzeugern wird allerdings durch den fehlenden Zugang zu realen Netzdaten erschwert. Gegenstand dieser Arbeit ist deshalb die Entwicklung georeferenzierter synthetischer Netze zur Untersuchung des Einflusses dezentraler Erzeugung und Nachfrage auf Verteilnetze. Darüber hinaus bietet diese Arbeit auch Lösungsansätze für die identifizierten Probleme im Zusammenhang mit der Reduzierung der Treibhausgasemissionen im deutschen Energiesystem um 95 % bis 2050.

Diese Arbeit untersucht, wie sich die dezentrale Nachfrage und Erzeugung auf die Komponenten des Verteilsystems und die Netzanfälligkeit in deutschen Niederspannungsnetzen im Jahr 2050 auswirken könnten. Dafür wurde ein Klassifizierungsmodell für die Einteilung von Gebäudetypen entwickelt. Die Anwendung dieses Modells auf die Gebäude in Deutschland basierend auf OpenStreetMap-Daten ergab, dass 19.747.802 von 29.497.992 Gebäuden als Wohngebäude klassifiziert wurden. Der prozentuale Fehler lag bei 3,4 %. Darüber hinaus wurden 500.000 georeferenzierte, elektrische Niederspannungsnetze für Deutschland anhand der klassifizierten Wohngebäude synthetisiert. Basierend auf einem Basisszenario und mehreren Sensitivitätsanalysen wurden verschiedene betriebliche Grenzwertverletzungen durch den zukünftigen Einsatz verteilter Nachfrage und Erzeugung festgestellt.

In Bezug auf die thermische Überlastung von Transformatoren zeigt es sich, dass 75% der Netze innerhalb ihrer Auslegungsgrenzen arbeiten, während 25% der Niederspannungsnetze zusätzliche Aufmerksamkeit erfordern. Bei mehr als 75% aller Netze kommt es zu einer Umkehr des Lastflusses und ein Viertel der 500.000 Netze wird als kritisch eingestuft. Darüber hinaus werden 11% der Leitungen thermisch überlastet. Bei 34 % der Netzknoten werden Störungen aufgrund von Unterspannung festgestellt. Die Sensitivitätsanalysen zeigen, dass die Überlastung der Transformatoren eine Folge des hohen Wärmepumpenanteils ist, der sich insbesondere in den Wintermonaten bemerkbar macht. Die Lastflussumkehr ergibt sich primär durch einen hohen PV-Anteil und ist unabhängig von der verteilten Nachfrage im Frühling und im Sommer zu beobachten. Die Ergebnisse zeigen, dass im Kontext der zukünftigen Nachfrage- und Erzeugungssituation in einigen der 500.000 Niederspannungsnetze technische Störungen zu erwarten sind. Potenzielle Lösungsansätze beinhalten Netzverstärkungsmaßnahmen für Transformatoren und Leitungen, beispielsweise die Installation von spannungsregulierenden Transformatoren und die Verstärkung der betroffenen 11% der Leitungen. Andererseits können Ansätze des Demand-Side-Managements wie das kontrollierte Laden von Elektrofahrzeugen, die kontrollierte Nutzung von Wärmepumpen und der Einsatz von Batteriespeichern dabei helfen, den Bedarf an Netzverstärkungsmaßnahmen zu verringern.

Acknowledgment

First and foremost, I would like to express my gratitude to Prof. Detlef Stolten for his significant assistance, support, and encouragement throughout my PhD journey. His insights were important in shaping the direction of this thesis, and I am truly grateful for his unwavering motivation and belief in my capabilities.

I would also like to extend my appreciation to Jülich Systems Analysis (ICE-2), formally Techno-economic Systems Analysis (IEK-3), for providing the essential infrastructure and resources that facilitated this research. The institute has played a significant role in fostering a protective environment for learning, exploration, and innovation, and for that, I am very grateful.

I would further express my gratitude for the support and assistance of Dr. Jochen Linssen, whose inputs and recommendations significantly enhanced the quality of this thesis. His guidance and constant support have been vital in the successful completion of my thesis.

On a personal note, I would like to express my heartfelt appreciation to my parents for their support and unconditional love. Your support and belief in me have been the greatest support in this journey, for which I am eternally grateful.

Finally, I would like to extend my heartfelt gratitude to my wife, Bhagya Sri Cheerla, whose love, patience, and understanding have been a source of strength throughout this process. Your support has been indispensable, and I could not have completed this thesis without your unwavering encouragement and sacrifices.

I really appreciate everyone for being an integral part of this achievement.

Table of Contents

Abstract	i
Kurzfassung	iii
Acknowledgment	v
Table of Contents	vii
1 Introduction	1
1.1 Motivation	1
1.2 Research Question and Objectives.....	3
1.3 Structure.....	4
2 Challenges and Opportunities in Distribution Networks	9
2.1 Distributed Demand and Generation in Distribution Networks.....	9
2.1.1 Operational Limiting Factors.....	11
2.1.2 Methods and Tools for Assessing Limiting Factors	22
2.1.3 Network Topologies for Operational Limiting Factors Assessment.....	28
2.2 Synthetic Electrical Power Network Models	32
2.3 Supplementary Literature for Future Demand and Generation Deployments in Synthetic Networks.....	36
2.4 Summary	37
3 Distribution Networks Characteristics	41
3.1 Graphical Characteristics	41
3.1.1 Graph Structure.....	43
3.1.2 Node Specific Characteristics	44
3.1.3 Edge Specific Characteristics	47
3.2 Electrical Characteristics.....	48
3.2.1 Power Specific Characteristics	48

3.2.2	Electrical Network Component Characteristics	49
3.3	Real-world Graphical and Electrical Network Characteristics	51
3.3.1	Network Indicators	52
3.3.2	Reliability of Supply Indicators	57
3.3.3	Analysis on Network and Reliability Indicators	59
3.4	Summary	62
4	Open and Synthetic Data for Modeling.....	65
4.1	OpenStreetMap Data Acquisition	66
4.2	OpenStreetMap Data Analysis	70
4.3	Building Type Classification	80
4.3.1	Data Pre-processing	82
4.3.2	Classification Model.....	90
4.3.3	Statistical Validation.....	93
4.4	Synthetic Load Profiles for the Classified Buildings.....	97
4.5	Summary	103
5	Estimation of Synthetic Distribution Network Topologies.....	105
5.1	Geo-referenced Synthetic Low-voltage Networks	106
5.1.1	User Input	106
5.1.2	Modeling Methodology.....	108
5.2	Geo-referenced Synthetic Medium-voltage Networks	123
5.2.1	User Input.....	124
5.2.2	Modeling Methodology.....	127
5.3	Geo-referenced Synthetic High and Extra-high Voltage Networks	135
5.3.1	User Input and Methodology.....	136
5.4	Validation of Geo-referenced Synthetic Low-voltage Networks	138

5.4.1	Statistical Validation.....	139
5.4.2	Operational Validation	144
5.4.3	Geographical Validation.....	145
5.5	Summary	146
6	Distribution Grid Model and Scenario Simulations	149
6.1	Scenarios.....	150
6.2	Methodology Demonstration on an Exemplary Geo-referenced Synthetic Network 155	
6.3	Summary	160
7	Results and Discussion.....	163
7.1	Transformer Violations.....	164
7.2	Power Line Violations	182
7.3	Voltage Violations	187
7.4	Sample Demand Side Management using Controlled BEV Home Charging	191
7.5	Summary	193
8	Conclusions	197
9	Summary	201
	Appendix.....	205
A.	Distribution Grid Operating Voltage Levels for EU Nations.....	205
B.	Corine Land Cover (CLC) Features.....	206
C.	Uncertainties in OSM Building Labels.....	207
D.	Renewable Energy Sources (Open Data)	208
E.	Geo-referenced Synthetic Low-voltage Networks	211
F.	Standard Medium-voltage Cable Types	214
G.	Figures.....	215

H. Tables	221
I. Abbreviations	223
J. References	224

1 Introduction

Increased greenhouse gas emissions (GHG) from the electricity system, transportation, and buildings are a few primary challenges that will further deteriorate global warming, climate change, and general well-being in the coming decades. The three major sectors mentioned above account for approximately 50% of the global GHG emissions. Therefore, to ensure that future generations are not adversely affected by climate change, 184 nations jointly agreed to cut down 89% of global emissions, as outlined in the Paris Agreement [2]. To accomplish the objectives drawn in the Paris Agreement, each nation established its own aims and objectives. In this context, the European Union (EU) drafted its objective of reducing GHG emissions by 80% to 95% relative to 1990 levels by 2050 [3].

To facilitate the reduction of GHG emissions, particularly in the electricity system, various energy transition policies that emphasize the integration of energy-efficient and climate-neutral technologies into power systems have been adopted. Consequently, electrical power networks are attracting more climate-friendly and carbon-neutral energy generation and demand technologies. Electricity generation by renewable sources entails a focus on solar photovoltaic (PV), wind, and biomass to generate clean energy, progressively decentralizing the power grid. To leverage clean electricity generation, which has been increasingly gaining traction, battery electric vehicles (BEVs) are being integrated into the power system to maximize clean energy usage while minimizing greenhouse gas emissions that are associated with combustion engine vehicles in the transportation sector. Similar to the battery electric vehicles in the transportation sector, heat pumps (HPs) seem to be a favorable solution to utilize clean electricity and minimize greenhouse gas emissions from typical heating systems in the building sector. However, these distributed demand and generation (DDG) technologies are specifically introduced into the distribution networks, which enables distribution networks to play a significant role in modern power systems. Therefore, there is a need to address the challenges associated with DDG in the distribution networks. The next sections will elaborate upon the underlying motivation for this thesis, along with the main research question and contributions.

1.1 Motivation

The electrical power grid has been evolving and improving over the decades while supplying electricity to end-users from spatially distributed generation stations via underground cables and overhead lines by establishing networks. When constructing these networks, grid operators consider the circumstances in which operational fuels, such as coal, oil, and natural gas, are abundant and dominating for generation stations. Thus, Figure 1-1 depicts the standard structured electrical power grid stacking from generation stations to end-users with varying voltage levels such as extra-high voltage (EHV) of 380-220 kV, high-voltage (HV) of 110 kV, medium-voltage (MV) of 10/15/30 kV, and low-voltage (LV) of 0.4 kV. However, it is important to note that voltage levels differ by country. For instance, in Germany, extra high voltage refers to the transmission grid, whereas high voltage, medium voltage, and low voltage belong to the distribution grid. However, some other countries regard high and extra high

voltage to be transmission networks, which will be explained in more detail in the following sections. It is necessary to note that multiple networks operate at varying voltage levels within a power grid and each such network has its own unique set of properties. Consequently, limited networks were constructed in consideration of the current circumstances as well as future perspectives.

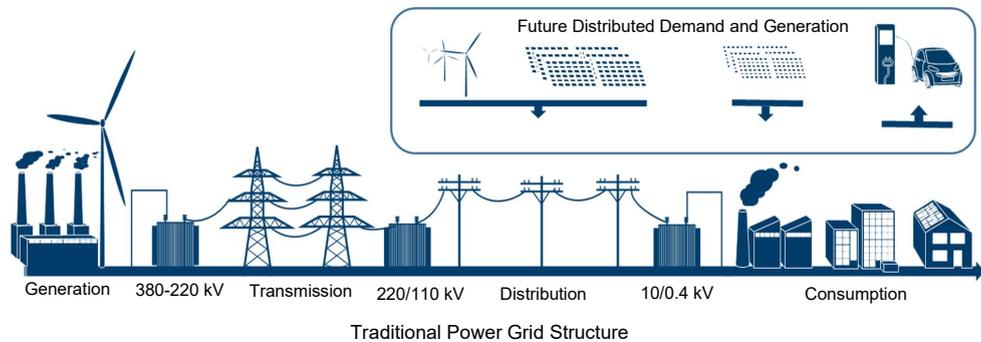


Figure 1-1 Standard power grid structure from generation to demand

In addition to the network structure, integrating Renewable Energy Sources (RES), such as solar photovoltaic, wind, biomass, and hydro into existing grids becomes necessary due to the political decisions to reduce greenhouse gas emissions, the depletion of fossil fuel resources, and technological advancements. Additionally, battery electric vehicles and heat pumps are rapidly gaining traction due to their ability to reduce emissions in the transportation and building sector, respectively. The rapid integration of renewable energy sources, battery electric vehicles, and heat pumps introduce significant changes in demand and generation, resulting in positive and negative peaks in power. Therefore, the introduction of the new distributed generation and demand impairs the existing network's power quality and operating conditions since these networks are not designed for large-scale distributed demand and generation deployment. Moreover, the increased integration of distributed demand and generation also increases the network's size leading to additional technical obstacles. Thus, models and methods that introduce complex problem-solving strategies are necessary to analyze the problems.

However, the ability to explore the impacts of distributed demand and generation in distribution networks is hampered by a scarcity of data on real-world networks for testing and validating established models. Therefore, limited available real-world network topologies were used to test and validate the various algorithms and models that were developed. Currently, security concerns prevent public research from accessing real-world networks. Nevertheless, a possible approach is to employ tests and typical network topologies to navigate the situation. However, the small network sizes of these test networks do not accurately represent the behavior of large networks covering a national supply system. As a result, it becomes necessary to integrate synthetic distribution networks that can reproduce the entire distribution network. Moreover, synthetic network topologies should be generated with geo-referenced

characteristics to replicate real-world networks precisely, which is an important solution. However, geo-referenced synthetic distribution networks are not real but can emulate the attributes of real-world networks. Hence, this work is motivated by the impact analysis of distributed demand and generation in the distribution networks using geo-referenced synthetic distribution networks. The research questions and objectives are listed in the following section:

1.2 Research Question and Objectives

This research focuses on the development of synthetic distribution networks utilizing geo-referenced open data to investigate future distributed demand and generation. This work contributes to solving two major challenges faced by the distribution networks: a lack of real network topologies for planning and research, and the estimation of future distribution demand and generation in distribution networks. More specifically, the research problems that are addressed in this work are as follows:

Are the distribution networks adequate for the anticipated increase in distributed demand and generation in the future?

In the future, which technical violations in distribution grids will emerge through large-scale integration of distributed demand and generation, and how could they be addressed?

How can the impacts of distributed demand and generation on distribution networks be studied despite insufficient public information about real-world networks?

To address these overarching questions, the existing models that can generate synthetic distribution network topologies are first examined. However, the models for establishing synthetic topologies are limited and lack certain features, with only a few models considering geographical locations, while others are limited to a specific place or require external data from system operators. Therefore, this thesis proposes new methods for estimating geo-referenced synthetic distribution network topologies based entirely on publicly available data. Apart from developing geo-referenced synthetic distribution network topologies, addressing problems associated with the integration of future distributed demand and generation is discussed by implementing a distribution grid model.

To summarize, the study's primary objectives to solve the above-mentioned overarching research problems are as follows:

- A complete examination of distribution network parameters, including graphical and electrical characteristics, is essential. This analysis will help understand the topological and electrical characteristics of complex networks and determine whether a generic algorithm may help estimate geo-referenced synthetic distribution networks for European countries.
- An analysis of the open data such as OpenStreetMap (OSM), census, and CORINE land cover data and the development of a model to classify building types is necessary.

The classification model should classify buildings into residential buildings and non-residential buildings. Moreover, in order to associate appropriate load profiles with the buildings, the residential buildings must be classified as single-family houses, multi-family houses, and apartments.

- The development of novel algorithms for estimating geo-referenced synthetic network topologies based on publicly available data. In particular, the methods should be accommodated for constructing low- and medium-voltage network topologies. Furthermore, the estimated geo-referenced synthetic network topologies should be evaluated statistically using complex network analysis.
- The model for predicting future distributed demand and generation in the distribution networks should be developed, incorporating estimated geo-referenced synthetic network topologies and a comprehensive examination of the challenges associated with future distributed demand and generation. Additionally, solutions to issues that may underlie future DDG integration into the distribution networks should be presented.

1.3 Structure

Following the discussion of the objectives, this work is organized to address the research question and to achieve the objectives in the manner depicted in Figure 1-2.

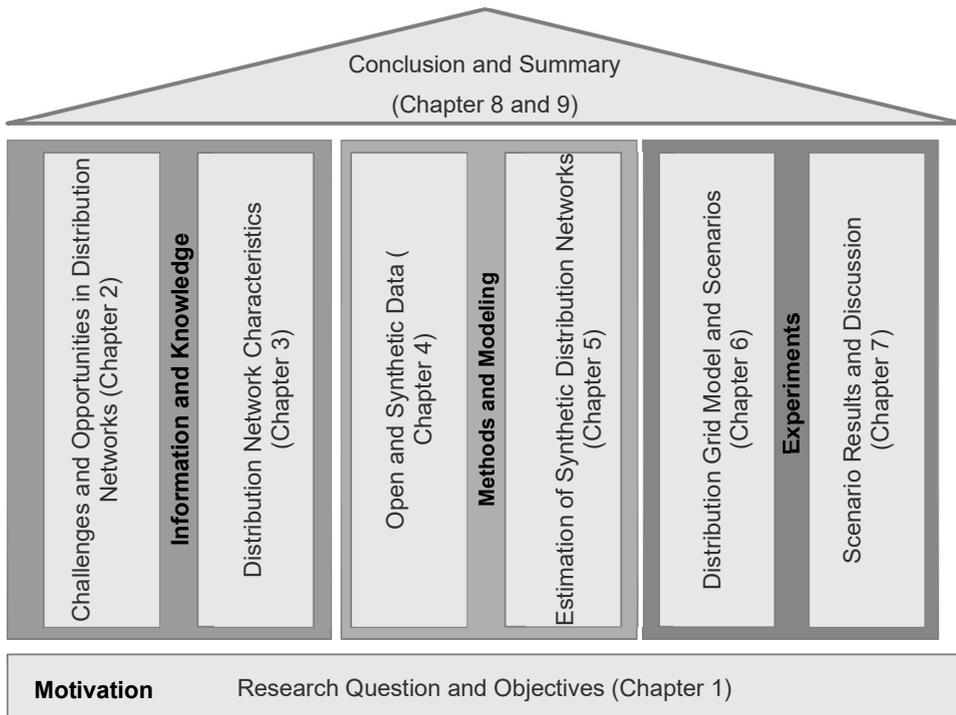


Figure 1-2 Overview of thesis

This thesis is divided into 9 chapters including the introduction. Following the introduction, Chapter 2 of this thesis will introduce the challenges and opportunities surrounding the distribution network. The chapter begins with a discussion on how distributed demand and generation will play a decisive role in the distribution networks. Following that, the primary operational constraint affecting the power system's supply will be examined. In addition to operational limiting factors, the methodologies and tools necessary for computing and analyzing the operational limiting factors will be addressed in the context of future demand and generation integration. Subsequently, available network topologies and their limitations will be presented along with the shift towards geo-referenced synthetic network topologies. Further, the synthetic power network models are addressed in detail, emphasizing transmission and distribution grids. Finally, the literature necessary for estimating geo-referenced synthetic network topologies will be introduced elaborately.

In Chapter 3, the characteristics of the distribution networks are discussed, while emphasizing their graphical and electrical characteristics. In this context, various graphical aspects will be explored in detail, including graph structures, node-specific features such as network size, node degree, betweenness centrality, and clustering coefficient, as well as edge-specific characteristics, including the number of edges, total length, average shortest path, and so forth. Besides the graphical characteristics, electrical characteristics essential for estimating geo-referenced synthetic distribution networks will be covered, including power demand, power supply, line utilization, and transformer utilization. Furthermore, several network and system reliability indicators for European countries derived from distribution system operator observatory data, such as total network length, the share of low-voltage and medium-voltage network length, and the total number of low-voltage transformers are analyzed. Finally, this assignment is then utilized to accurately determine whether developing a generic algorithm can estimate geo-referenced synthetic distribution networks across Europe.

Chapter 4 discusses the open data available for predicting geo-referenced synthetic network topologies in detail. Its quality will be explored using the best-fitting data set, OpenStreetMap (OSM). In this context, the building tags are examined as they are crucial for estimating geo-referenced synthetic distribution networks. Through this analysis, building types are categorized through classification tasks by using various machine learning techniques before settling on a suitable approach for predicting building types. Additionally, missing value imputation and class imbalance will be included while modeling due to the characteristics of the OSM data. The model predictions will be validated using publicly available data. During model development, the classification of building tags by machine learning algorithms is published as:

Bandam, Abhilash, et al. "Classification of building types in Germany: A data-driven modeling approach." Data 7.4 (2022): 45.

Further, synthetic load and generation profiles for households, battery electric vehicles, heat pumps, and rooftop solar PVs will be provided along with the open data.

Chapter 5 describes a novel technique for estimating geo-referenced synthetic distribution networks. As a starting point, required inputs for predicting network topologies are examined. Following that, detailed algorithms for graphical and electrical low-voltage networks are developed and discussed in detail. Apart from the low-voltage network algorithms, algorithms developed for geo-referenced synthetic low-voltage networks are used to describe additional algorithms employed to develop geo-referenced synthetic medium-voltage networks. Accordingly, Figure 1-3 illustrates the modeling approach used to construct the entire distribution network. The process begins with the collection of OSM data for buildings and road infrastructure. The established algorithms are then used to generate geo-referenced synthetic low-voltage networks. Next, medium-voltage networks are constructed using low-voltage networks and OSM data associated with the locations of medium-voltage transformers and non-residential buildings excluding garages and attachments. High-voltage and extra-high voltage networks are provided with medium-voltage transformer placements and OSM data, high-voltage transformer locations, and high- and extra-high-voltage power lines. During the process of model development, a data-driven methodological approach for estimating geo-referenced synthetic low-voltage networks is included in the publication:

Abhilash, Bandam, et al. "Geo-referenced synthetic low-voltage distribution networks: A data-driven approach." 2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe). IEEE, 2021.

Finally, to validate the predicted networks, a complex network analysis approach is employed. This helps compare them to some of the indicators discussed in Chapter 3.

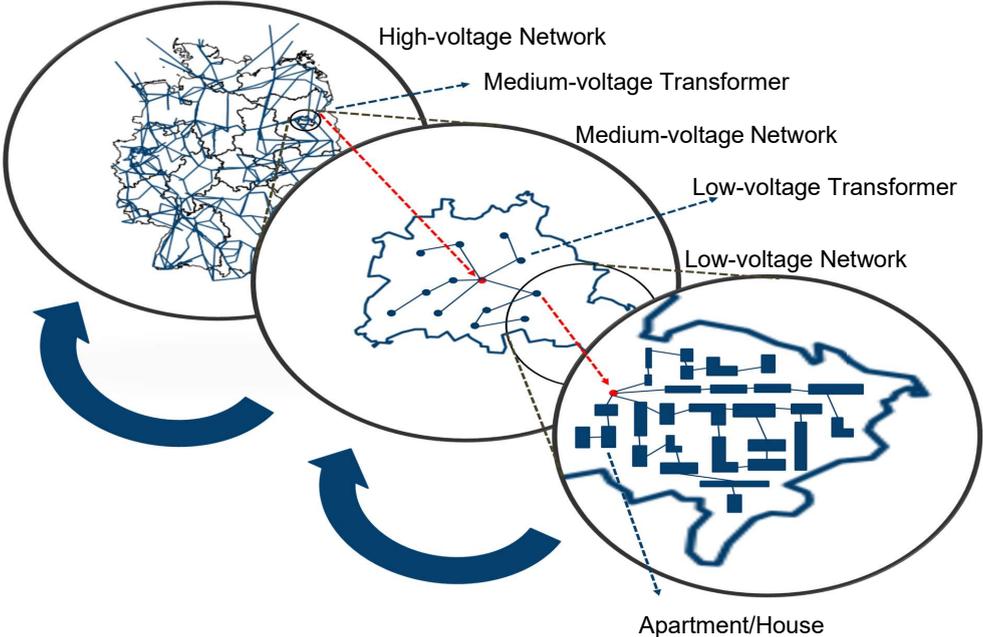


Figure 1-3 Schema for estimating geo-referenced synthetic distribution network topologies

Chapter 6 demonstrates how future distributed demand and generation might be accommodated through estimated geo-referenced synthetic network topologies. This chapter also introduces several scenario-based simulations in line with the published scenario for Germany with 95% carbon dioxide (CO₂) emission reduction. Further, a sensitivity analysis is conducted through varying penetration levels of these distributed demands and generation.

Finally, Chapter 7 discusses the evaluation of future distributed demand and generation within low-voltage distribution networks using geo-referenced synthetic low-voltage networks. In this chapter, various operation limiting factors, and their violations will be examined in detail. Further, measures for reducing future demand and generation related violations are also analyzed.

Chapter 8 summarizes the conclusion and contributions, with a particular emphasis on achieving the thesis objectives.

Finally, Chapter 9 will include a summary of each chapter, with a focus on the accomplishment of the objectives so that the reader can quickly acquire the gist of each chapter.

Each chapter, however, is structured in a manner wherein a set of questions are posed at the beginning and the answers or solutions are provided in the summary. Additionally, each chapter will conclude with a set of key messages.

2 Challenges and Opportunities in Distribution Networks

What are the challenges involved in distribution networks?

Why is it necessary to estimate geo-referenced synthetic distribution network topologies?

Topics covered:

-  Operational limiting factors associated with the integration of distributed demand and generation in the distribution networks
-  Methods and tools required to assess the operational limiting factors
-  Limitations of real, test, IEEE, and typical network topologies
-  Limitations of available synthetic network models
-  Literature for estimating synthetic network topologies

Introducing modern distributed power generation and demand technologies into the distribution networks will lead to significant challenges, including technical constraint violations, which must be addressed. Therefore, fundamental information concerning operational limiting factors, assessment methods, tools, and available electrical network topologies is considered. In this chapter, information is disbursed sequentially, beginning with the challenges that may plague distribution networks due to distributed technologies, and gradually progressing to study the opportunities using geo-referenced synthetic networks. Distributed demand and generation in the distribution networks are first introduced before providing a comprehensive overview of the various operational limiting factors in the distribution networks and their associated performance indicators. Further, various combinations of distributed demand and generation in the distribution grids at both low- and medium-voltage levels are presented, as explored in the literature, particularly emphasizing limiting factors, techniques to study the limiting factors, and tools utilized for the analysis. A summary of the network topologies considered in the literature is also provided in addition to the tools and techniques. Considering the limitations of the existing network topologies, this chapter subsequently covers the existing models for generating synthetic electrical network topologies. Finally, after identifying the limitations of the existing models, the chapter provides the literature, models, algorithms, and data sources required for estimating geo-referenced synthetic distribution network topologies. In this way, the information and knowledge required for answering the main research questions will be established.

2.1 Distributed Demand and Generation in Distribution Networks

Energy-related challenges associated with greenhouse gas emissions are being conquered through the adoption of new ecologically friendly solutions, including the extensive use of renewable energy sources, such as solar photovoltaics, wind, and biomass, to generate electricity. Besides renewable energy sources, end-consumers should adopt heat pumps and

battery electric vehicles to make use of the electricity provided by renewable energy. By adopting such measures, emissions from electrical, buildings, and transportation sectors will decrease significantly. Thus, distributed demand and generation units are being integrated at a high rate in the distribution networks, which include low-voltage, medium-voltage, and high-voltage networks. However, the introduction of distributed demand and generation units presents new obstacles to system operation. These challenges only emerge if the incoming capacities surpass the capacity of the current network infrastructure. Therefore, it is equally vital to analyze operational limiting factors and gain a deeper understanding of the issues in order to establish a realistic plan to avoid technical violations in the future.

The most typical operational limiting factors in the distribution networks, such as voltage limitations, grid component thermal loading, and power flow constraints will also be discussed. Once the limiting factors have been thoroughly analyzed, the measures that can be undertaken to assess them, such as deterministic, probabilistic, and optimization techniques, will be reviewed. However, the emphasis of this study is primarily on existing approaches in the literature in order to determine the effects of distributed demand and generation in the distribution networks. After the techniques are established, the benefits and drawbacks of each technique will be discussed in detail, following which a suitable method can be selected to fulfill the purpose of this work. The tools for modeling electrical networks and for computing limiting factors are also covered in this chapter. In addition to the methods of analysis and knowledge about technical factors, network topologies play an essential role in accommodating the demand and generation in the distribution networks. Depending on the network properties, different network topologies may function independently, and the effects of demand and generation may differ from one network topology to another. Thus, network topologies are extremely important to the overall process, right from integrating distributed technologies to understanding the operational limitations of the distribution networks. Thus, the discussion delves into the electrical network topologies that were considered when assessing the limiting factors, including test, typical, Institute of Electrical and Electronics Engineers (IEEE), and real-world network topologies. Nevertheless, to present the distributed demand and generation in the distribution networks, this section is organized according to the diagram depicted in Figure 2-1.

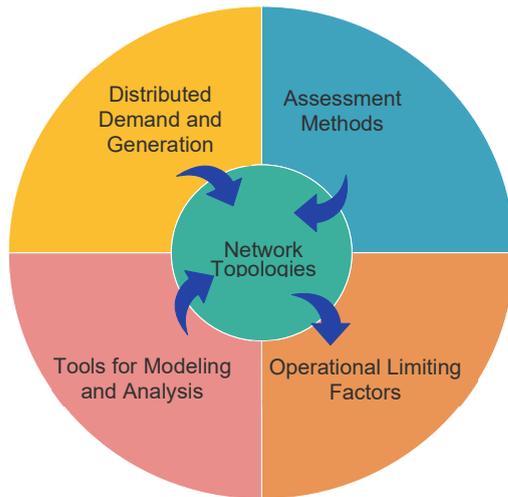


Figure 2-1 Overview of distributed demand and generation in distribution networks

2.1.1 Operational Limiting Factors

Operational limiting factors describe the distribution network properties that are influenced by the integration of distributed demand and generation [4]. In fact, the operational uncertainty of distributed demand and generation affects the standard system operation. The limiting factors include voltage stability, voltage deviation, grid component loadings, and reverse power flows, all of which are the key limiting factors that distribution system operators (DSOs) consider before introducing distributed demand and generation units into the distribution networks. Within this context, many researchers have been examining these limiting factors to address network vulnerabilities and hosting capacity assessments. Therefore, it is worthwhile to present these factors elaborately prior to analyzing the future impacts of distributed demand and generation on distribution networks. As already mentioned, Table 2-1 summarizes the significant limiting variables that have been identified and investigated by the researchers in the literature.

Table 2-1 Summary of operational limiting factors accompanied by performance indicators

Operational limiting factors	Description	Indicators
Voltage stability	Voltage stability refers to the ability of an electric power network to maintain an acceptable voltage on each bus in the network under all circumstances [5].	Voltage stability index
Voltage deviation	Voltage deviation is defined as the deviation of a bus's voltage magnitude from its nominal value [6].	Voltage deviation index Voltage violation limits
Voltage unbalance	Voltage unbalance is the phenomenon that occurs when there is a difference in the magnitudes of phase and line [7].	Voltage unbalance factor Voltage violation limits
Voltage dips	Voltage dip is the phenomenon that occurs due to a sudden reduction of nominal voltage between the ranges of 10% to 90% followed by a voltage recovery after a tiny duration [8].	Frequency of voltage dips Cost of voltage dips
Harmonics	A voltage or current harmonics is a multiple of the fundamental frequency of a system produced by non-linear loads [9].	Total harmonic distortion index
Grid component thermal loadings	Grid components, when operated beyond their limits, cause thermal overloading [10].	Percentage loading
Reverse power flow	Reverse power flow (RPF) occurs when power flows in the direction opposite to its normal course. In fact, power flows from a lower voltage level to a higher voltage level [11].	Reverse power flow factor
Power losses	Losses in electric power networks are technical losses due to heat and corona losses [12].	Power losses

However, the key operational limiting factors studied in this thesis are classified as voltage limiting factors, grid component loads, and power flow constraints. Voltage stability and voltage deviations are two categories under voltage limiting factors. Additionally, the transformer and line loading are the major grid component loadings, and power flow constraints detail reverse power flow and power losses.

Figure 2-2 depicts the classification of relevant operational limiting factors to address the research question of this thesis.

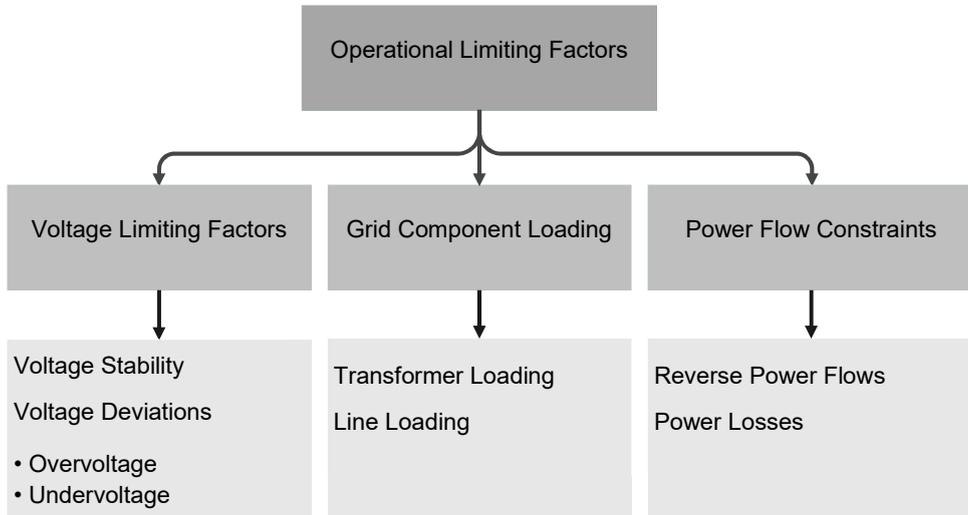


Figure 2-2 Classification of operational limiting factors in the distribution networks

Voltage limiting factors, such as voltage stability and voltage deviations, occur more frequently in distribution networks [13]. As stated in Chapter 1, the power system modernization leads to the disruption of the traditional transmission of active and reactive power from higher to lower voltage levels by distributed generation at lower voltage levels, converting the network from passive to active. However, it is the network's generation and load that governs changes in power flows and voltages. In principle, in any network, the voltage at the transformer node should be greater than the voltage at the consumer node to transmit power from the transformer to the consumer. In addition to the operating principle, when the reactance of the distribution overhead power lines/underground cables toward the resistance ratio is lower than the transmission lines, it leads to a voltage drop from the transformer to the customer. Therefore, the voltage drop at the consumer node rises with the increase in the load or the distance between the nodes (transformer to consumer node).

Adding distributed generators to the network will also influence voltage at the nodes. According to the working principle, the voltage at the generator node must be greater than the transformer node in order to facilitate the transfer of the power generated by the distributed generator [14]. Consequently, the voltage at the generator nodes is higher than the transformer node, resulting in a voltage rise. Due to the radial structure of distribution networks, particularly low voltage and medium voltage, voltage drops and rise change along the feeder. The rise and drop of the voltage along the feeder are depicted in Figure 2-3, where V_n , V_{max} , and V_{min} denote the nominal voltage, maximum permitted voltage, and minimum allowable voltage, respectively. As a result of these ambiguities in the voltage magnitude, violations concerning voltage stability and voltage deviation emerge.

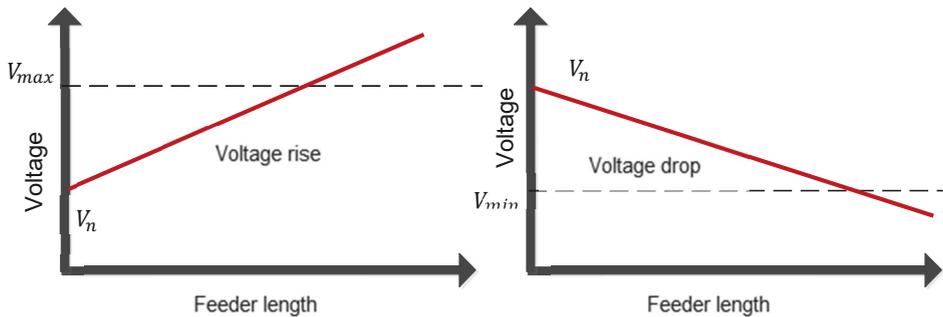


Figure 2-3 Voltage magnitude along the feeder in case of generation and load (left), and only load (right). Adapted from [15]

When it comes to the operation of the electric power system, it is necessary to maintain voltage stability at each node of the network. Voltage stability refers to the power network's capacity to maintain an acceptable voltage on each bus regardless of the circumstances [5, 16]. In fact, for any distribution network, the voltage stability depends on the time of operation. There are two types of instabilities: transient and steady state. Voltage instability may result from undesired voltage levels in any of these states. The most noticeable circumstance is a high network load combined with a lack of reactive power assistance in the steady state. In this circumstance, the voltage drops abruptly, resulting in voltage collapse, which subsequently causes a sudden grid collapse. An outage in the electrical network can lead to network malfunctions, resulting in failures in devices, overheating, and significant power losses. Additionally, it may cause blackouts in a power network. When studying network stability, Modarresi et al. [16] categorized different types of voltage stability indices, including line voltage stability indices, bus voltage stability indices, and overall voltage stability indices. However, the voltage stability index (VSI) [13] is one of the indicators that is typically used to identify a node that is on the verge of a voltage collapse.

Another operational limiting factor under voltage limiting factors is voltage deviation, which is the deviation of a node/bus voltage magnitude in the distribution network from the nominal voltage [6]. The voltage deviation is mainly influenced by loads, generators, and distance from the primary or secondary transformers [13]. As mentioned earlier, the distribution networks are nearly radial in structure, which is uniform across nations, and the drop or rise of the voltage magnitude in a feeder, as shown in Figure 2-3, depends on the load and generation in the network. As a result, it is challenging to ensure consistent voltage magnitude at the nodes. To estimate and address voltage deviations, the voltage deviation index will be utilized. The voltage deviation index provides the deviation in the voltage magnitude at each node. Equation 2-1 is used to calculate the voltage deviation index for each node in the system. To ensure that the system remains under control, the voltage deviation index at each node must be kept to a minimum. If the voltage deviation index is higher than anticipated, the node voltage should be treated with techniques that enable it to maintain normal operating condition.

$$\text{Voltage deviation} = \frac{V_{\text{nominal}} - V_{\text{measured}}}{V_{\text{nominal}}}$$

2-1

As discussed above, a rise and drop in the voltage depend on the network's load and generation. In this context, significantly stronger integration of decentralized power generation through solar PV, wind, and biomass will increase the voltage at the nodes [17]. However, the increase in voltage corresponds to voltage rise should fall within the permitted limits. Overvoltage violations occur if the voltage exceeds the upper limit, while the bus voltage will drop along the feeder if the load requirements are higher than the rated load capacity in the network. In fact, the higher integration of battery electric vehicles [18-20] and heat pumps is the apparent cause of voltage drop. Additionally, this also happens in the presence of distributed generators, in instances where the demand for battery electric vehicles or heat pumps is high, and generators such as PV generate power at the lowest capacity. On the other hand, the decrease in voltage corresponds to a voltage drop causing Undervoltage violations when striking the lower limit.

The overvoltage and undervoltage instances that occur over a short period are compensated by voltage regulators deployed on the network. However, the continuous operation of voltage regulators reduces their lifetime and impairs their efficiency [21]. Therefore, the system should work within the predefined upper and lower voltage limitations to avoid disruptions. In Europe, the voltage standards must adhere to the European standards EN 50160 norms [22, 23]. According to Reference [23], the end consumer voltage should not deviate over $\pm 10\%$ of the nominal voltage for 95% of data measured in one week and $+10\%$, and -15% of the nominal voltage for the total time. Thus, if the system node encounters a voltage rise that exceeds the specified limits, then the system node is said to be under overvoltage violation. Similarly, if the system node encounters a voltage drop that exceeds specified limits, then it is said to be under Undervoltage violation. The voltage limits specified in EN 50160 [22, 23] are listed in Table 2-2.

Table 2-2 Voltage limits according to European standards EN 50160 [22, 23]

Voltage constraint	Limits
Voltage magnitude	$\pm 10\%$ of the V_n for 95% of data measured in a week and $+10\%$ and -15% of V_n for total time.
Voltage unbalance	$+2\%$ for 95% of the week

However, the limits defined in Table 2-2 are shared between medium- and low-voltage networks. Figure 2-4 illustrates the division of voltage range between medium- and low-voltage networks when using standard distribution transformers in a low-voltage network. Thus, each network will obtain a share of $\pm 10\%$ depending on the transformer regulating the bandwidth. However, to utilize the entire voltage band in both medium- and low-voltage, it would be optimal to use a Voltage Regulating Distribution Transformer (VRDT) in all low-voltage networks [24]. Voltage regulating distribution transformer essentially operates by decoupling medium- and low-voltage networks. In addition, the introduction of a voltage regulating distribution

transformer redistributes the voltage band to medium- and low-voltage networks in accordance with the EN 50160 standards. As a result, both medium- and low-voltage networks have access to the full $\pm 10\%$ of available bandwidth. In all parts of the grid in medium-voltage, the distribution of the voltage band between medium-voltage and low-voltage networks with voltage regulating distribution transformer is illustrated in Figure 2-5. Here, each voltage level that is medium voltage and low voltage will have full voltage bandwidth of $\pm 10\%$

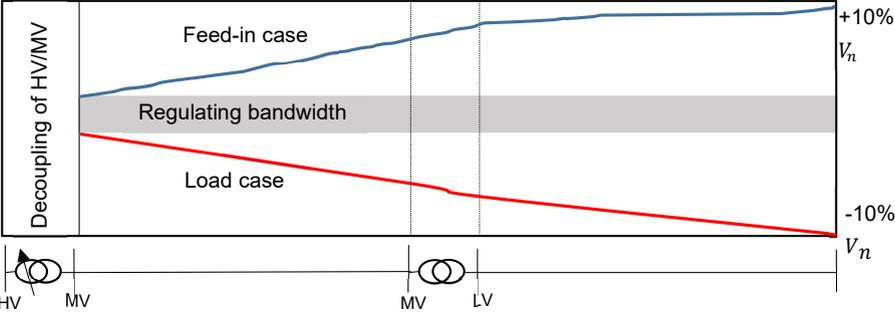


Figure 2-4 Distribution of voltage band between medium-voltage and low-voltage networks with a standard distribution transformer, Adapted from [24]

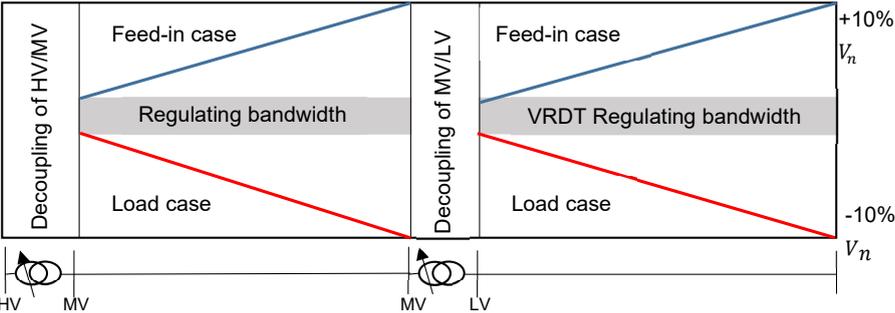


Figure 2-5 Distribution of voltage band between medium-voltage and low-voltage with a voltage regulated distribution transformer (VRDT), Adapted from [24]

Apart from the overvoltage and undervoltage violations, an unsymmetric system exhibits voltage unbalance. In a system, voltage unbalances lead to power losses on the network, degradation, and malfunctioning of the protection equipment. The voltage unbalance is quantified using the voltage unbalance factor, which is the measure of deviation from balanced conditions of voltage magnitudes and phase angles [25]. More precisely, it is the measure of the ratio between negative and positive sequence voltages [7]. Equation 2-2 defines the voltage unbalance factor expressed as a percentage [7]. Similar to the overvoltage and undervoltage limits, according to EN 50160, this constraint's violation limit is $+2\%$ for 95% of data measured over a week (see, Table 2-2).

$$VUF = \frac{V_2}{V_1} \times 100 \quad 2-2$$

V_2 is the negative sequence component shown in Equation 2-4, and V_1 is the positive sequence voltage component shown in Equation 2-3.

$$V_1 = \frac{V_{ab} + a V_{bc} + a^2 V_{ca}}{3} \quad 2-3$$

$$V_2 = \frac{V_{ab} + a^2 V_{bc} + a V_{ca}}{3} \quad 2-4$$

Where $a = \angle 120^\circ$, $a^2 = \angle 240^\circ$ and V_{ab} , V_{bc} , V_{ca} (line voltages) [7].

The next important operating limiting factor to study is grid component loading. Transformers, overhead lines, underground cables, and switchgear are all examples of components in a network. However, transformers, power cables, and power lines are the main components of interest in this thesis, because switch gears are not modeled as part of synthetic networks. Each power network component includes specifications that regulate how much load it can support. Moreover, each component also varies in terms of its current-carrying capacity, with each designed to handle specific current ratings. If these components are utilized/operated beyond their operational limits, violations are triggered, and the components will be overloaded. In theory, overloading of the components results in their failure, eventually leading to operational failures [26]. According to this thesis, if a component's loading/utilization exceeds 100% of its rated capacity, the component is subject to grid component loading violation [27]. The limits for the grid components considered in this thesis are listed in Table 2-3.

Table 2-3 Operational limits for considered grid component

Grid component	Limits
Transformer	> 100% of rated capacity
Overhead line	> 100% of rated capacity
Underground cable	> 100% of rated capacity

Another significant operating limiting factor on the power networks is power flow constraints, which include reverse power flows and power losses. As previously stated, distributed generation can reach a high penetration rate in distribution networks. Thus, the introduction of distributed generators influences the voltage and power flows. By increasing generation at the secondary/low-voltage side of the transformer, the power flow is reversed by facilitating the flow of power from the generator to the distribution transformer. For example, whenever the overall network generation exceeds local demand, power flows from lower voltage level networks to higher voltage level networks, resulting in Reverse Power Flows (RPFs) [28, 29]. Reverse power flow increases the voltage at the generator node, which is associated with the overvoltage violations that occur when the voltage surpasses the upper limit. The reverse

power flow can be controlled or regulated by lowering the voltage on the transformer's secondary side using an on-load tap changing transformer, which eventually decreases the voltage below the upper limit. During high reverse power flows, the minimum tap changing position can be reached, which may lead to a loss of control over the voltage [30]. When the network undergoes such repercussions, protective equipment will sustain damage and negatively impact power quality [11]. Therefore, reverse power flow reduction strategies, such as distributed renewable energy generation curtailment, should be implemented to reduce the amount of reverse power flowing backward to the higher voltage side network [31]. However, before implementing such strategies, it is necessary to comprehend and analyze the implications of the future integration of distributed generation into the distribution networks. In this context, it is unquestionably crucial to investigate the instances in which the power flow is reversed.

Increased distributed demand and generation penetration into distribution networks may also jeopardize network resilience and lead to increased power losses. However, the majority of network power losses are attributed to transformer and line/cable I^2R losses [32], where I is the current and R is the resistance of the line which is defined as the measure of the opposition of the flow of current through the power line. Here, the current quantity and the length of the lines/cables are directly proportional to these network losses. Therefore, It may also be claimed that more the number of feeders on a network, the greater the power losses [33]. Additionally, it may be asserted that an increase in load demand will increase net power network losses. The total power losses in the network are calculated as depicted in Equation 2-5.

$$P_l = \sum_{i=1}^N I^2 R_i \quad 2-5$$

Where, I = current, R_i = Resistance of feeder i , N = number of distribution feeders.

In brief, the operational limiting factors include voltage limiting factors, grid component loading, and power flow constraints, which are the most significant operational constraints associated with the integration of distributed demand and generation into the distribution networks. Therefore, several researchers conducted experiments with various combinations of demand and generation to investigate how the integration of distributed demand and generation in distribution networks impacts the limiting factors. In this regard, a comprehensive review of the literature from the last decade is conducted. The details of the publications are listed in Table 2-4 along with the technologies, including distributed generation such as PV, distributed demand such as battery electric vehicles and heat pumps, and Energy Storage Systems (ESS). Table 2-4 also includes factors, such as voltage limiting factors, grid component loading, and power flow constraints, which are evaluated in the literature.

Table 2-4 Literature study on the impact of considered distributed demand and generation over operational limiting factors (OLFs) in the distribution networks

Technologies /OLFs	Overvoltage	Undervoltage	Voltage unbalance	Voltage stability	Harmonics	Transformer loading	Line loading	Reverse power	Power losses
PV	[21, 33, 35, 36, 61, 62, 64, 65, 83, 98, 111-114]	[36, 64, 65]	[33, 64, 98, 99]	[21, 34, 85-87]	[21, 73, 83]	[64, 65, 73]	[64, 65]	[35, 60-62]	[21, 33-37]
EV/PHEV/BEV	X	[18-20, 27, 32, 38, 41, 42, 66, 74, 75, 78, 84, 102-105]	[25, 27, 43, 70, 75, 77, 79, 100, 101]	[25, 32, 39, 40, 46, 68, 70, 71, 74, 79, 88-90]	[67, 75, 84]	[20, 27, 40, 66, 69, 70, 74-79]	[27, 66-71]	X	[19, 20, 32, 38-47]
EV & PV	[48, 106]	[48, 106]	X	[49]	X	[49]	X	[49]	[48, 49]
EV, PV, & ESS	X	X	X	[5, 26, 80]	[5]	[10, 26, 80]	[26]	X	[5, 12, 50]
PV & ESS	[15]	[15]	X	[52, 91-94]	X	X	X	[31]	[51, 52]
RES	[71]	X	X	[53, 95, 96]	X	X	X	X	[53]
RES & ESS	X	X	X	[97]	X	X	X	X	X
RES & EV	[54-56, 72, 107, 108]	[54-56, 63, 72, 107, 108]	X	[58]	X	X	[72]	[28, 55, 63]	[28, 54-59]
RES & HP	[82, 109, 110]	[81, 82, 109, 110]	X	X	X	[81, 82]	X	X	X

It is evident in Table 2-4 that researchers attempted to integrate individual technologies into distribution networks to ascertain their effects. For example, 38% of the literature reviewed in this study relates to integrating electrical vehicles in distribution networks. Additionally, 20% and 8% of the literature evaluated the impacts of PV and renewables in the distribution networks, respectively. The integration of electric vehicles, however, leads to significant challenges in the distribution networks. To overcome the problems associated with integrating electric vehicles into the distribution networks, renewable energy sources and PV are also being integrated into the networks. In this regard, 15% of the literature reviewed examines the renewables & electric vehicle combination. However, there are further challenges that emerge with this combination. Therefore, to address the impacts of demand and generation that includes electric vehicles and renewables in the distribution networks, the use of energy storage is also introduced. Accordingly, 8%, 3%, and 1% of the evaluated literature studied PV & energy storage systems, electric vehicles & energy storage systems, and renewables & energy storage system combinations, respectively. As mentioned previously, heat pump technology is also gaining traction in modern power systems. But there are relatively modest publications concerning the integration of heat pumps in distribution networks. According to the reviewed literature, only 4% of the articles considered heat pumps with renewables in the distribution networks. Figure 2-6. displays the percentage of publications considered for individual and combined technologies in the distribution networks.

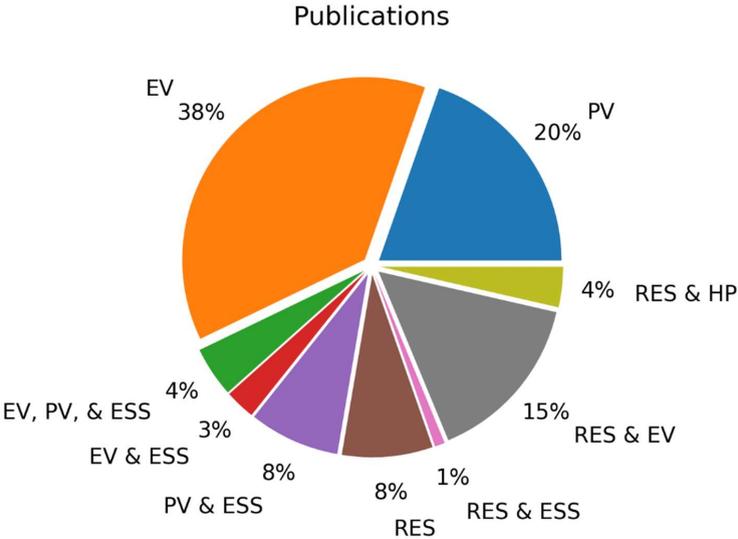


Figure 2-6 Percentage publications for a different combination of technologies integrated into the distribution networks; the combination of technologies include Electric Vehicles (EV), Photovoltaics (PV), Renewable Energy Sources (RES), Energy Storage Systems (ESS), and Heat Pumps (HP)

These studies combinedly indicate that the two technologies that greatly influence distribution networks are electric vehicles and solar photovoltaics. However, given the recent surge in interest in the usage of heat pumps in the building sector, this technology should be evaluated in conjunction with electric vehicles and PVs in the distribution networks. Yet, according to the investigation undertaken for this thesis, there is relatively less literature available pertaining to the combination of these three technologies, namely PV, electric vehicles, and heat pumps in the distribution networks. Consequently, to fill the gap in research, this thesis will analyze the integration of PVs, battery electric vehicles, and heat pumps in the distribution networks and calculate the operational limits that arise.

When analyzing the operational limiting factors, the literature reviewed also entailed an assessment of several limiting factors for individuals and different combinations of technologies pertaining to distributed demand and generation in the distribution networks. The heatmap in Figure 2-7 illustrates the number of publications associated with the limiting factors analyzed in conjunction with distributed demand and generation in the distribution networks.

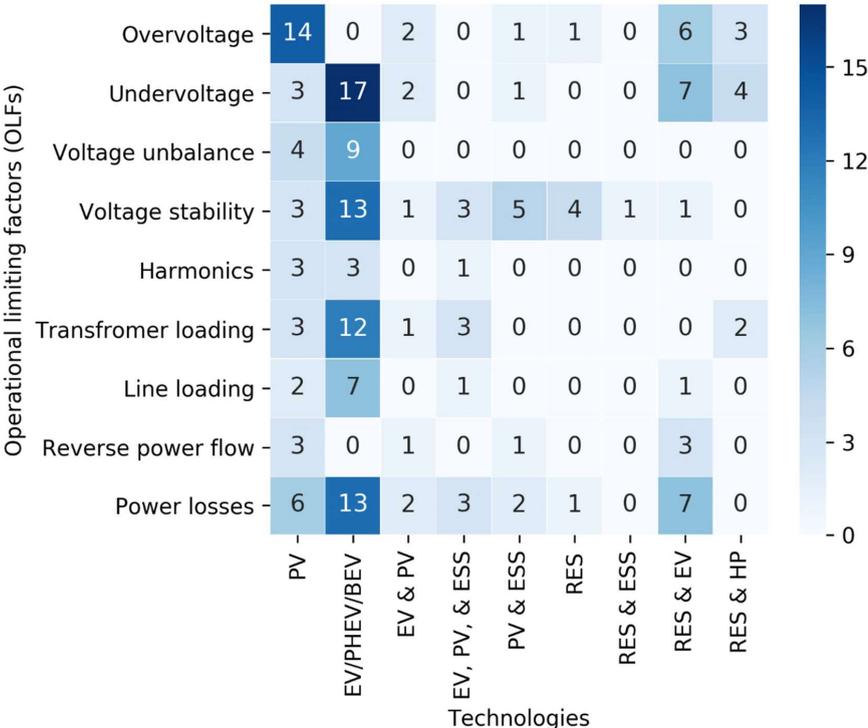


Figure 2-7 Total number of publications referring to Operational Limiting Factors (OLFs) with Distributed Demand and Generation (DDG) which include Electric Vehicles (EV), Photovoltaics (PV), Renewable Energy Sources (RES), Energy Storage Systems (ESS), and Heat Pumps (HP)

Through the literature review in this thesis, one pertinent conclusion is the pertaining role of overvoltage when introducing distributed generation into the distribution networks. Unsurprisingly, when generation increases with low demand, the voltage on the node/bus increases, resulting in overvoltage violations. Together, these studies further indicate that when integrating distributed demand in the distribution networks, the undervoltage limiting factor was given prominence. It is understandable that when load demand is high, the likelihood of a voltage drop is also high. Furthermore, significance is also attributed to transformer loading, line/cable loadings, and power loss occurrences when distributed demand is introduced into the distribution networks. Considering the factors discussed above, overvoltage, undervoltage, transformer loading, line/cable loading, reverse power flows, and power losses are the most relevant evaluation criteria when incorporating combined demand and generation in the distribution networks. Thus, to address the research question associated with the future integration of PV, electric vehicles, and heat pumps into the distribution networks, this thesis will focus on operational limiting factors related to voltage limiting factors, grid component loadings, and power flow constraints, such as overvoltage, undervoltage, transformer loading, line loading, and reverse power flows.

2.1.2 Methods and Tools for Assessing Limiting Factors

Since the adequate Operational Limiting Factors (OLFs), which will be the primary emphasis of this thesis have been defined, this subsection will explore the methodologies for assessing operational limiting factors in the distribution networks. Limiting factors can be evaluated from both the consumer's and the network operator's perspective [115]. From the consumer perspective, network operators have minimal control over consumer-side demand and generation [4]. Thus, users are free to install generation units and consume electricity at their discretion. The owners of distributed generators, however, should adhere to a set of guidelines. When customers have a disproportionate influence over the integration of distributed demand and generation units in the distribution networks, no predictable patterns can be discerned in the location of distributed demand and generation units and the time of demand in the networks. Thus, to address limiting factors violations in the distribution networks, it is optimal to employ deterministic and probabilistic methods. In contrast, from the network operator's perspective, an optimization method can be employed for the integration of demand and generation to minimize x , where x could be the investment cost, limiting factor violations [4].

However, numerous studies have investigated the limiting factors by attempting to utilize different methodologies, including deterministic, probabilistic, and optimization methods. However, all these methods have one thing in common: they all use power flow calculations to determine the voltage at all nodes, the current in all lines, and the power at the transformer. Table 2-5 showcases the literature that utilized deterministic, probabilistic, and optimization approaches.

Table 2-5 Overview of methods used for assessing Operational Limiting Factors (OLFs) in the literature

Method	References
Deterministic method	[21, 31, 32, 37, 41, 42, 46, 61-63, 66, 67, 69, 73, 83, 84, 86, 91, 106, 108, 116, 117]
Probabilistic method / Stochastic method	[26, 33, 35, 36, 48, 49, 51, 64, 71, 74, 77, 78, 82, 88, 94, 96, 97, 101, 102, 105-113, 116-122]
Optimization methods	[5, 10, 12, 19, 25, 28, 35, 38-40, 42-45, 47, 49, 50, 53, 55-59, 63, 68, 70, 72, 79, 80, 89, 93, 97, 102, 114, 118, 123-126]

The deterministic method does not focus on randomness when evaluating the limiting factors. Instead, all variables are considered beforehand and assigned to distinct network nodes prior to performing power flow calculations [4]. These variables may be any aspect associated with the network and distributed demand and generation parameters, which include the placement of DDGs, their size, the number of residents in dwellings, and the overall number of dwellings in a building, among other things. Deterministic methods can be of two types. On the one hand, when performing power flow simulations, constant input is used without considering variations in the input profiles. In other words, the value of distributed demand and generation output does not vary. This type of method is referred to as the constant generation method. In this instance, the maximum/peak value or the average value is typically considered when assessing the limiting factors violations. On the other hand, power flow simulations can be performed with time series input wherein the distributed demand and generations value varies with each timestep of the simulation. This is referred to as the time series method. However, limiting factor violations in these methodologies are detected using scenario-based or sensitivity-based analysis. Figure 2-8 illustrates the main flow diagram of the deterministic approach for assessing operational limiting factor violations.

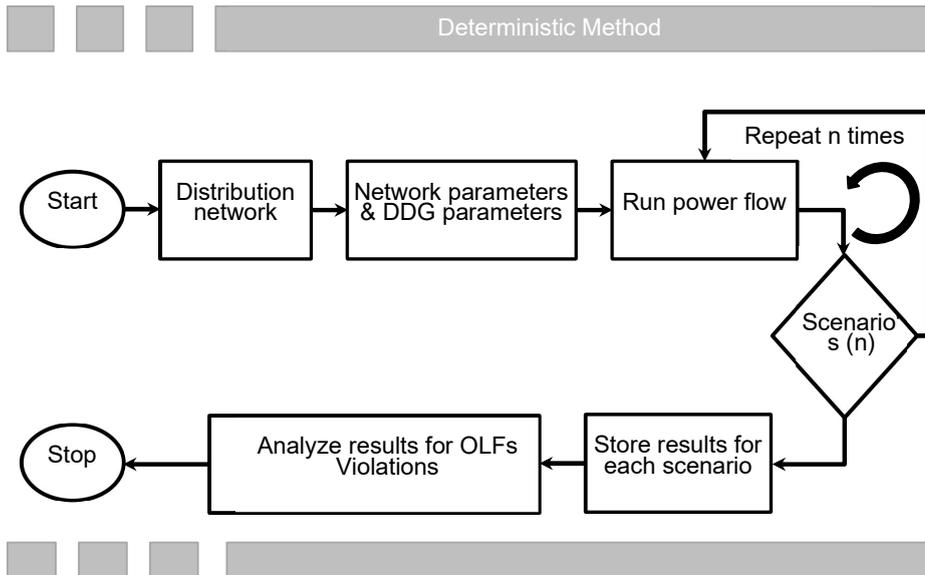


Figure 2-8 Flow diagram for assessing Operational Limiting Factors (OLFs) violations using a deterministic method

The deterministic method for analyzing limiting factors and hosting capabilities has been applied by several researchers in literature and is detailed in Table 2-5. All these studies focused on various limiting factors, depending on the technology used in the distribution networks. When distributed demands, which include electric vehicles and heat pumps, are integrated into the distribution networks, undervoltage, transformer loading, and line loading are typically examined [66, 67, 73]. Similarly, when distributed generation, such as solar PV, is introduced into the distribution networks, overvoltage and reverse power flows are investigated [21, 83, 106, 108].

As mentioned previously, there is substantial uncertainty surrounding the integration of distributed demand and generation in the distribution networks [17]. The number of distributed demand and generation users, the location of distributed demand and generation units, and the capacity of distributed demand and generation units are only a few of the uncertainties. These uncertainties substantially impact the number of violations associated with operational limiting factors. All these variables are fixed in the deterministic method before performing power flow simulations. Uncertainties, however, are not deemed to be a part of the deterministic method. Therefore, a probabilistic method is an optimal choice in terms of expressing the uncertainties associated with the integration of distributed demand and generation in the distribution networks. In the probabilistic method, numerous power flow simulations are performed with the aforementioned probability distribution of the parameters such as distributed demand and generation units installed, the capacity of distributed demand and generation, and location of distributed demand and generation [26, 110, 116]. The probability distributions of these parameters help gauge the various possible combinations

when using distributed demand and generation units in the distribution networks. Monte Carlo simulations are typically employed to generate this randomness in the simulations [33]. The flow diagram of the probabilistic approach for assessing operational limiting factor violations is depicted in Figure 2-9. The literature that used the probabilistic method for analyzing operational limiting factors and hosting capabilities is detailed in Table 2-5.

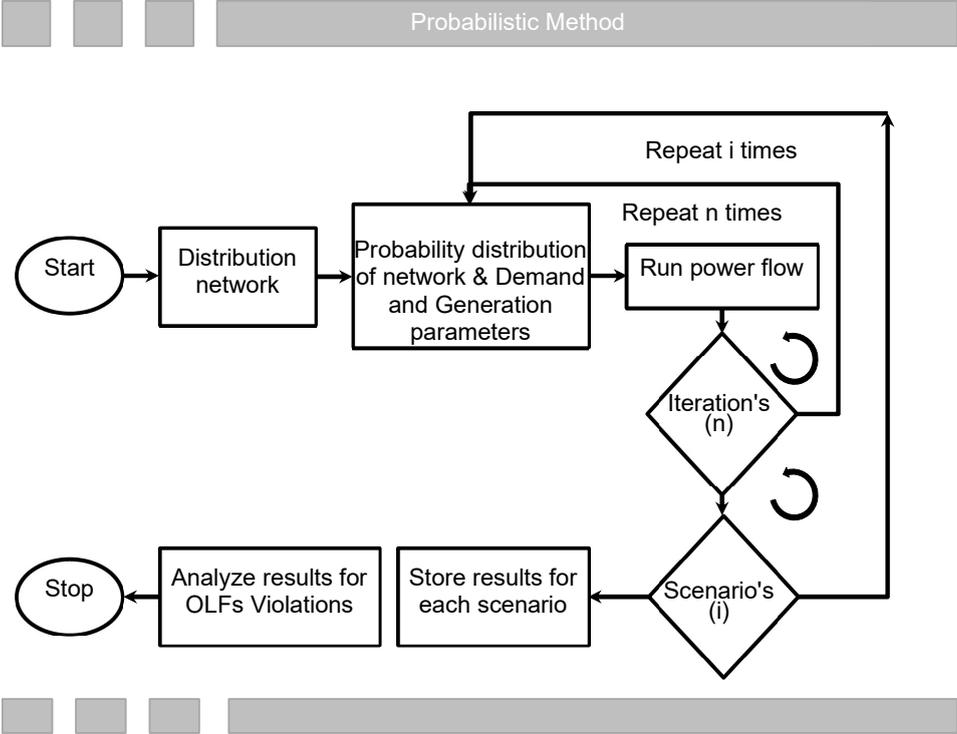


Figure 2-9 Flow diagram for assessing Operational Limiting Factors (OLFs) violations using a probabilistic method

In addition to the deterministic and probabilistic methods, an optimization approach method is utilized if a system integration is planned from the distribution network operator perspective. This method identifies operational limiting factors violations and attempts to mitigate those problems [4]. An optimization problem is introduced in the optimization method to minimize operational limiting factors violations [19, 42, 93, 114], investment costs [19], or to maximize distributed demand and generation utilization [59] in the distribution networks. However, the objective function is subject to several different network constraints [19], such as positioning distributed demand and generation units only in specific places, charging electric vehicles only at night, voltage band limit, and feed-in electricity only while electric vehicles in the network are charging. The standard form of optimization is depicted in Equation 2-6 [127].

$$\begin{aligned} & \text{minimize } f(x) \\ & \text{subject to } g_i(x) \leq 0, i = 1, \dots, m, \\ & \quad h_j(x) = 0, j = 1, \dots, n \end{aligned}$$

Where f is the objective function to be minimized, $g_i(x) \leq 0$ are called inequality constraints, $h_j(x) = 0$ are called equality constraints

In the context of the use of an optimization method to optimally integrate distributed demand and generation, most of the literature reviewed in this thesis used a single-objective function. In most of the studies, when integrating distributed demand and generation units in the distribution networks, the single-objective function is frequently constructed to minimize operating costs [19, 104], investment cost [50], CO₂ emissions [25], energy losses [27, 42, 45, 47], and load fluctuations [19]. Moreover, optimization problems are formulated for optimal power flows, with particular emphasis on the optimal siting of solar PV generation units [12, 53, 93, 97], optimal siting of electric vehicle charging stations [38, 89, 94], and battery placements of large PV generators [49]. A single-objective function is also utilized to address issues such as smart charging for electric vehicles [10, 124] and battery management [42]. A multi-objective function was also developed in certain literature to optimize the position and the size of electric vehicle charging stations [37, 39, 68, 118], optimal positioning of PV units in addition to battery management [35], and combined optimal placement and sizing of electric vehicles and Distributed Generators (DGs) [59]. To solve these optimization problems and formulate an optimal solution, several techniques were employed in the literature. Genetic Algorithms (GA) [44, 93], Particle Swarm Optimization (PSO) [43, 89, 94], Evolution Algorithm [123], Artificial Bee Colony [56], Grey Wolf Optimization [56], global improved binomial mutations for Butterfly Optimization (BO) [19] are the most common strategies employed in the studies that were reviewed.

In all the literature reviewed here, 26% of the research took the deterministic method into consideration, 36% of the studies considered the probabilistic method, and 41% considered the optimization method, as illustrated in Figure 2-10. The evidence presented here indicates that the optimization method was employed in most of the published studies. However, each of these methods presents distinct advantages and disadvantages. The deterministic approach, for instance, does not include uncertainties regarding PV generation and its location, electric vehicle charging locations, and usage. Nevertheless, it is an adequate method if a known network with predefined distributed demand and generation sites, consumption, and generation patterns exists. However, a deterministic approach is inadequate when the consumption pattern and distribution of distributed demand and generation units' placements in distribution networks are unknown. Consequently, any uncertainty in generation and consumption is addressed by employing the probabilistic method through multiple iterations of power flow simulations by modifying parameters relating to distributed demand and generation in the distribution networks. In this case, several random inputs are provided that are dependent on the chosen probability distribution. However, this method is highly sophisticated due to the size of the input parameters, the number of iterations, and the number of scenarios used in the calculations. This is because the time required to run these simulations is

significantly higher than the time required when using the deterministic method. This method, however, is appropriate when there are fewer networks to evaluate. Conversely, this method is infeasible when dealing with a high number of networks. The optimization approach requires numerous iterations to obtain an optimal solution and is unsuitable for a high number of networks. In addition, when it comes to integrating distributed demand and generation in the distribution network from the consumer perspective, this strategy is inadequate.

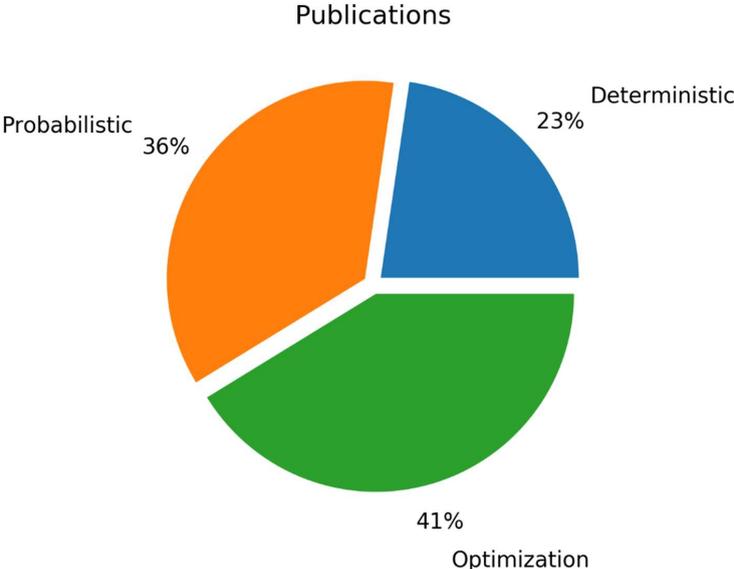


Figure 2-10 Assessment methods addressed in the literature

Considering the information provided until now, this thesis aims to delve into the integration of future distributed demand and generation into the distribution networks with an emphasis on the entire country. Therefore, the analysis of thousands of networks is involved. A deterministic method is, hence, an optimal choice for identifying the most vulnerable networks by studying all the operational limiting factors discussed.

So far, we have covered the operational limiting factors as well as the methods of analysis. Now, the tools required for the analysis must be investigated. Upon deciding to implement the deterministic method to study the operational limiting factors, specific tools are required to conduct analysis and execute power flow simulations. As demonstrated in [128], various commercial and open-source tools are available for energy and power system simulations. Each of these tools possesses distinct technical and economical capabilities. Each tool focuses on a single or a combination of capabilities, such as power flows, dynamic analysis, fault current analysis, energy dispatch optimization, investment cost optimization, and long-term scenarios.

This thesis, however, is concerned with the analysis of integrating future distributed demand and generation in the distribution networks in the context of operational limiting factors violations. Hence, tools that are capable of simulating power flows are sufficient. Power flow simulations can be performed with a wide variety of tools, such as CYME [69, 129], NEPLAN [130, 131], pandapower [132], PyPSA [133], DlgSILENT/PowerFactory [31, 86, 134], Synergi Electric [135], DER-CAM [136], GridLAB-D [137], HYPERSIM [138], IPSA 2 [139], MATPOWER [140], OpenDSS [141], RAPSIm [142], ReEDS [143], and SIMPOW [144]. These tools are distributed under a combination of commercial and open-source licenses. As indicated above, the thesis is focused on analyzing distribution networks across the country, which necessitates modeling several thousand networks; thus, programming-based tools can simplify their operations. The Scripting language modules pandapower [132] and PyPSA [133], for example, could be among those that are capable of performing power flow simulations and are purely based on Python programming language. These are open-source scripting languages and are free to use. Among these tools, PyPSA [133] is used by Europower [145], which is an in-house transmission grid model. This information coupled with the capabilities of each tool has led to the adoption of PyPSA for power flow simulations in this thesis. PyPSA [133] or Python for Power System Analysis tool was developed by Tom Brown et al., [133] which is entirely based on python and is an open source tool. In addition to performing power flow simulations, PyPSA also provides investment and decision support.

Until now, this section discussed the operational limiting factors, their effects, and violation limits in the distribution networks. Additionally, methods and tools are explored, and the deterministic method was chosen to analyze the impacts of distributed demand and generation in the distribution networks with PyPSA as a tool for power flow simulations. However, to accomplish the goal of this thesis, network topologies – the most critical and central component of the analysis – are necessary. Therefore, the following section will discuss the various network topologies available and their associated limitations.

2.1.3 Network Topologies for Operational Limiting Factors Assessment

To evaluate the possible consequences of future distributed demand and generation in the distribution networks, the network topologies are essential as they play an important role in accommodating anticipated demand and generation. Therefore, this section will examine the proposed network topologies in the literature and their limitations. The literature studied as a part of operational limiting factors violations in the distribution networks due to distributed demand and generation utilizes real, test, IEEE, and typical/benchmark network topologies to perform the analysis, detect the violations associated with the operational limiting factors, and mitigate the consequences. Table 2-6 highlights the literature that employed various network topologies.

Table 2-6 Literature overview with different network topologies

Network Type	References
Real	[10, 15, 21, 27, 28, 31, 35-37, 43, 46, 48, 51, 60, 61, 64-67, 69, 70, 74, 76, 77, 80-84, 88, 89, 91, 92, 100, 102, 108-110, 112-114, 116, 120-123, 125, 146, 147]
Test	[25, 26, 50, 58, 103, 118, 119, 148-150]
IEEE	[5, 12, 18-20, 32, 34, 38-42, 44, 45, 53-57, 68, 72, 75, 85, 86, 93-95, 97, 99, 106, 107, 111, 124]
Typical / Benchmark	[33, 47, 52, 71, 96, 98, 117, 151]

The literature listed in Table 2-6 under the category of “real” employ real-world network topologies, which are basically retrieved from distribution system operators to conduct analysis. For instance, Barbosa et al. [27] investigated the hosting capacity of electric vehicles on 75,550 real-world low-voltage networks, which were obtained from Brazilian utility operators. Similarly, Unahalekhaka et al. [31] leveraged the network of the provincial electricity authority in Thailand to develop and test an algorithm that can be employed to reduce the reverse power flow when installing photovoltaic power plants by situating and identifying the right size of batteries. In this manner, the authors in the reviewed literature utilized different real-world networks to determine the hosting capacities of distributed demand and generation, violations underlying the operational limiting factors, and the development of algorithms to alleviate these violations. However, real-world network topologies are helpful in examining the effects of distributed demand and generation on a particular network with known demand and generation. It is far-fetched to utilize a real-world network from one specific area to investigate another. Additionally, only a small percentage of real-world networks are accessible online and security-based concerns make it particularly challenging to access real-world networks from distribution system operators. Therefore, several authors have employed test networks to determine the wide-ranging issues of network vulnerabilities and implications of distributed demand and generation.

The test networks are essentially replicas of real-world networks and are derived by analyzing and adapting various real-world networks. However, they are generated using a variety of approaches, such as eliminating outliers from the real-world networks with regard to sensitivity data, clustering network parameters, and designing manually [152]. For instance, Yiju et al. [26] constructed a test network by converting a UK-based low-voltage network to an Australian-based low-voltage network with the objective of studying the impacts of PV and battery storage systems in the distribution networks. Similarly, various other test networks have been established to address the challenges posed by the unavailability of real-world networks. In this context, networks such as a semi-urban low-voltage network [25], a 21-node sample distribution network [118], a residential test grid [148], Roy Billinton test system [149], a 33 [103], and 69 bus microgrid [46] are developed. Thus, the test networks are created to generalize the results from these networks to other networks by replicating the real-world networks.

In this context, the Institute of Electrical and Electronics Engineers (IEEE) created IEEE networks ranging in size from 13 to approximately 342 nodes for each network. Although there is a wide range of networks, the researchers in the majority of the literature utilized IEEE 13, 33, and 34 bus networks to conduct various experiments. Similar to the test networks, these IEEE networks are based on real-world network topologies. For research purposes, power networks designed by IEEE are publicly available and accessible online. Table 2-7 summarizes some of the IEEE networks that are publicly accessible online. Since these networks are freely available, they can be modified to meet specific requirements. For example, the transformer's capacity, the lines/cables, or the network's total length may be modified. Adjusting the component parameters such as transformer capacity or line specifications creates a minimal shift in the network structure. However, adding a new node or removing a node can completely alter the network's topology. Therefore, there is less flexibility in modifying the specified network topologies to meet the predefined requirements.

Table 2-7 Institute of Electrical and Electronics Engineers (IEEE) test networks, Adapted from [152]

Network	Nodes	Length (km)	Loads	Reference
IEEE13 Bus	13	2.5	9	[18, 20, 95, 152]
IEEE 33 Bus	33		27	[12, 32, 34, 39, 41, 42, 53-55, 68, 85, 97, 106, 107, 124]
IEEE 34 Bus	34	94	24	[75, 86, 94, 111, 152]
IEEE 37 Bus	37	5.5	25	[99, 152]
IEEE 38 Bus	38			[19]
IEEE 69 Bus	69			[34, 39, 53, 56]
IEEE 123 Bus	123	12	114	[152]
IEEE 342 Bus	342	15.2	624	[152]

It is important to note that these networks are less flexible. Moreover, network characteristics such as frequency, voltage, network circuit length, and the number of nodes in a network vary by location. For instance, the network from Europe differs from that of Asia and North America [153]. This can be attributed to the fact that the total number of consumers connected to a network in Europe is higher than the total number of consumers connected to North America because of the spatial distribution. Therefore, the network's length in Europe is significantly greater than the network's length in North America. Additionally, the cabling level in Europe is higher than in North America [152].

For these reasons, to differentiate networks between regions, researchers created benchmarks, representatives, and typical networks that combine typical features for a specific country. Specifically, Kerber in [151] developed Germany's typical low-voltage networks. These network topologies were developed with the aim of estimating the capacity of photovoltaic systems to be hosted in low-voltage grids. These typical grids can be applied to a variety of situations. Similarly, Meinecke et al. [154] created benchmark datasets for

Germany using publicly available data under Simbench [154], and Dickert et al. [155] used principal component analysis and clustering to construct Dickert's LVDNs benchmark grids expressly for Germany. In the same way, Kiaee et al. [156] created UKGDS, which mimics typical United Kingdom (UK) distribution grids. In North America, there are systems such as Cinvalar's System [157], Baran's System [158], Salama's System [159], and CIGRE Systems [160] designed to reflect the specific features of the region. The CIGRE system, however, combines the characteristics of both Europe and North America.

In all the studies reviewed concerning the impacts of distributed demand and generation on distribution grids, the authors utilized various network types that were recently explored. Together, these studies indicate that IEEE, Real, Test, and Typical networks account for 33%, 49%, 10%, and 8% of the total literature analyzed, respectively. Surprisingly, real networks provided by the distribution system operators are employed in half of the cases. The rest of the studies, however, seem to be distributed across test, IEEE, and typical networks. Figure 2-11 illustrates the distribution of various network topologies utilized in the literature reviewed.

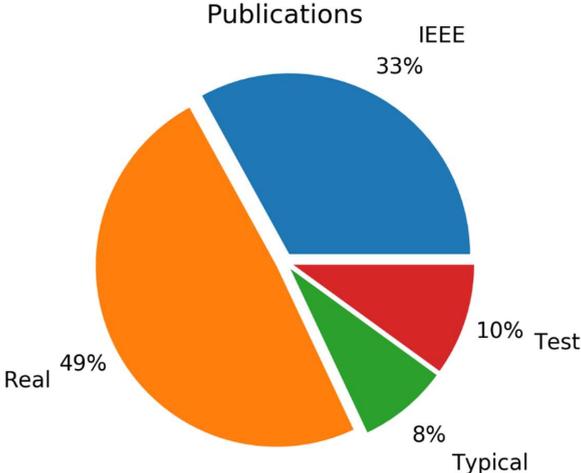


Figure 2-11 Network topologies considered in the publications

Based on the information provided up to this point, a clear picture of network types and their limitations becomes evident. As illustrated in Figure 2-11, most of the literature employed real-world networks to conduct research. However, real-world networks from distribution system operators are challenging to obtain. Even if real networks are acquired, they are typically protected by non-disclosure agreements for security reasons. This may prove to be disadvantageous for researchers looking to develop new algorithms and methodologies. To circumvent this, researchers explored openly available network topologies, such as Test and IEEE networks. However, their limited number precludes the incorporation of distinct structural properties. Additionally, the features vary by location, rendering the test and IEEE networks ineffective when focusing on a particular region. Certain researchers are concerned with typical networks that are developed in response to a region's unique characteristics. These networks,

however, are insufficient for large-scale investigations in which geographic location is essential. In most cases, typical networks/benchmark networks can only provide general conclusions at the country level, and they are not effective in conducting in-depth analyses for specific locations.

These limitations imposed by real, test, IEEE, and typical networks led to the emergence of the concept of synthetic network topologies with an emphasis on geographical coordinates. In the following section, a literature review will be presented, which focuses on existing approaches for constructing synthetic network topologies.

2.2 Synthetic Electrical Power Network Models

Various tools capable of generating synthetic network topologies are being developed due to the limitations of current network topologies. This section discusses models that can generate synthetic electrical grids irrespective of geographical location. Electrical power networks are primarily represented mathematically as graph networks and analyzed using complex network analysis [161]. Nodes and edges are the standard components of graph networks. When the graph networks are defined by physical constraints and principles, they evolve into power networks, with nodes representing transformers, substations, generators, and loads, and edges representing power lines and cables.

Research activities on synthetic networks were initiated with graph networks and complex network analysis and progressed from there. Erdos and Rényi [162] were the first to discuss random graphs, which are constructed by selecting pairs of nodes at random and linking them with the edges. The only requirements for this method are the network's order and size. Later, Watts and Steven [163] compared power networks to small-world networks. In small-world networks, the majority of the nodes are not neighbors and instead, are only reachable via a few edges from other nodes [164]. In this work, random graphs are generated using small-world characteristics. Barabási and Albert [165] created a method for generating networks with a power-law degree distribution. This study employs a preferential attachment paradigm, wherein the node with the highest degree has a greater probability of acquiring new edges forming a network. The requirement, in this case, is the number of connections per node, which is referred to as the node degree. However, graph networks built on random graphs and small-world qualities may not adhere to the topological aspects of real-world power networks. This is corroborated by Pagani and Aiello [166], who reviewed the literature and observed variations between real-world and small-world networks. Nevertheless, synthetic networks based solely on topological considerations may not be suitable for power networks due to the limits imposed by the power flow. Therefore, electrical characteristics, in addition to topological properties, are required to depict power networks.

Wang et al.[167] presented an upgraded version of small-world networks called RT-nested-Small world algorithms for generating synthetic power grids and subsequently representing these limitations. This model takes into consideration both statistical information about network topologies as well as electrical characteristics. Similar considerations were also made in [168, 169] to generate synthetic power grids by modifying small-world algorithms. A different

methodology was presented in [170] that introduces cluster and connect algorithms, generating synthetic networks. The models, however, construct a partial power grid. Nevertheless, these models lack the ability to depict the power grid through the spatial use of geographical coordinates. Due to the models' limitations, they are only partially suitable for conducting in-depth research on power grid vulnerabilities. To circumvent the limitations surrounding geospatial considerations, location-based comprehensive modeling tools emerged as a promising area of research.

In the field of location-based synthetic power grid research, Birchfield et al. [171, 172] described a greenfield strategy, which helps construct synthetic power grids using a combination of complex network approaches, electrical characteristics, and publicly available data. In this model, the grid topology was determined by picking edges from the Delaunay triangulation of geographically distributed loads and generators. However, the data obtained from publicly available sources are postal zip codes and population, which were used to assign loads to post-office zip codes, and the population helped determine the load size. Nonetheless, real generator sites were utilized, which were obtained from the energy infrastructure administration. In [173], a similar approach was employed to create synthetic power flow models with geographic locations. Furthermore, [174] developed a random growth model that takes into account geographical nodes and links. In this case, the algorithm produces a minimum spanning tree and adds additional links to the neighboring nodes. This strategy, however, bears some resemblance to the historical development of electrical systems. In this context, a complex network technique was employed in [175], which utilizes the historical evolution of power grids to construct synthetic power grids by replicating the influence of various economic and technical elements on their growth.

All these models reviewed here produce synthetic power grids, although not every model resembles an actual power grid. Some models employ complex network analysis to generate topologies, while others improve them based on electrical characteristics. Additional features, such as geographic coordinates for loads and locations of existing generators taken from system operators were also implemented. In any case, these models were derived from an entirely automated process to transform them into a semi-automated one with additional information. However, power lines are automatically synthesized within the network according to various methodologies. To construct realistically similar network topologies, [176] developed the network imitation method based on learning (NIMBLE) by examining the structural features of the western interconnection grid of the United States (WI). NIMBLE develops synthetically spatial power networks with features that are akin to a given network without disclosing the precise locations of power lines and substations. This methodology, however, requires the locations of power lines and substations from system operators. Additionally, [164, 175] coupled complex network analysis with electrical parameters to create a synthetic transmission network and compare the results to those of Spain, Portugal, and France.

The models presented thus far utilize complex network analysis and require additional information from system operators to synthesize power networks. Therefore, to address the shortcomings of these models, new models using publicly accessible map data are under

development. OpenStreetMap (OSM) [177] data is one of the most widely used map data that can be accessed freely and is a collaborative initiative that began in 2004, intending to create an editable global geographic database. Using the OSM data, models such as SciGRID [178], osmTGmod [179], GridKit [180], and the Common Information Model (CIM) [181] are capable of constructing extra-high voltage transmission grids. These models use data retrieved from OSM regarding extra-high voltage power lines, substations, and other power-related properties to build the network topology. Although the data is identical, the approach employed to build the transmission grid varies between models. Heitkoetter et al. [182] conducted a mathematical, visual, and electrical study of these free-source models. In this context, AutoGridComp [182] was developed to compare models based on a variety of variables, including graphs, Geographical Information Systems (GIS), and power. The eGo grid model [183] also developed a synthetic extra-high voltage transmission grid and a high-voltage grid for Germany by supplementing the osmTGmod [179] model with additional information. However, all these models are intrinsically focused on the generation of synthetic transmission grids; only a limited number of models have been developed to generate synthetic distribution networks. Figure 2-12 depicts the map illustrating synthetic transmission models that have been built for various countries.

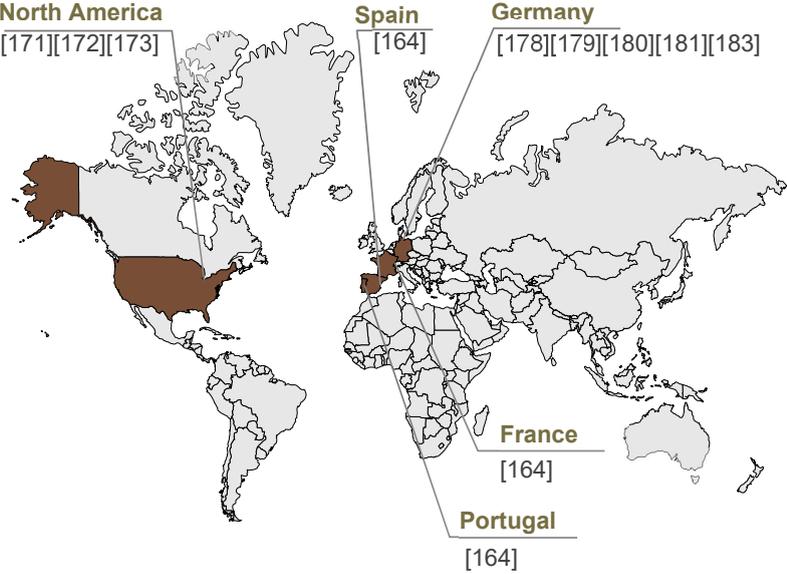


Figure 2-12 Countries with synthetic transmission grid models

In the context of georeferenced synthetic distribution grid models, Domingo et al. [184] developed a Reference Network Model (RNM). This work produced two distinct forms of reference network models. The first is a greenfield model, which can construct distribution networks from scratch while considering substations and distributed generators, among other things. Under the brownfield model, on the other hand, the existing distribution networks are

reinforced. Moreover, the reference network model has the capability to generate simultaneous high-, medium-, and low-voltage networks. With it as a foundation, Mirna et al. [185] built the Distribution Network Model Platform called DiNeMo in order to construct synthetic European representative networks. This platform, however, collects additional data from a wide range of distribution system operators. Similarly, Meinecke et al. [154] developed a model named SimBench to generate representative datasets for high-, medium-, and low-voltage networks. In this model, several data points pertaining to distribution networks are gathered from distribution system operators. Nevertheless, only benchmark and representative networks are made available through the DiNeMo and SimBench platforms. Besides these models, Pisano et al. [186] presented a synthetic distribution network model based on open data and geographical information. However, substation locations, high-voltage nodes, population, land use statistics, and administrative boundaries were all considered from external sources. Based on the topological distinctions between European and North American electrical distribution networks, Carlos et al. [187] developed the RNM-US model for building synthetic distribution grids in the United States of America, which is also capable of modeling high-, medium-, and low-voltage networks. A study that is not included in the cluster of studies listed above is [188] as it uses OSM data to model high-, medium-, and low-voltage networks. Overall, these models are location-specific and require considerable external data.

To utilize more publicly available data instead of less accessible data, the eGo grid model was developed, which generated medium-voltage synthetic distribution networks [189]. This model identifies load areas and medium-voltage grid districts before generating network topology. However, the network topology was generated using a Capacitated Vehicle Routing Problem (CVRP) and a local search metaheuristic. This model developed 3608 synthetic medium-voltage networks for Germany. However, low-voltage networks remained undeveloped. Additionally, [190, 191] created a cluster-based low-voltage synthetic network in which a large planning zone was divided into mini-zones while implementing a greenfield methodology for network and transformer optimization. However, countries where synthetic distribution networks have been established, are depicted in Figure 2-13.

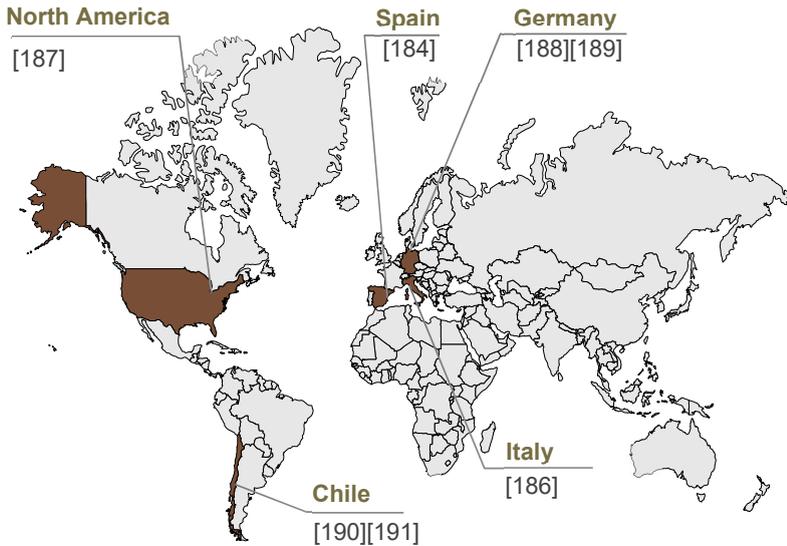


Figure 2-13 Countries with synthetic distribution grid models

To summarize, there are numerous approaches that are currently being employed for the development of synthetic transmission grids utilizing publicly available OSM data. However, only a few models can generate geo-referenced synthetic distribution networks, and they have various shortcomings, including the fact that they generate synthetic networks that are restricted to specific locations, require a high degree of heuristics and a lot of external input in addition to open data. Therefore, this thesis focuses on the development of geo-referenced synthetic distribution network topologies for impact analysis of future distributed demand and generation on distribution networks. The following section discusses the supporting literature that contributes to model development.

2.3 Supplementary Literature for Future Demand and Generation Deployments in Synthetic Networks

To estimate geo-referenced synthetic distribution network topologies, validate them, and analyze them, it is essential to analyze the literature for certain data, modeling tools, and previously developed algorithms. In this context, several graph-related properties such as betweenness centrality [192] and clustering coefficient [193] were utilized. This section assesses graphical features as well as real-world electrical properties from the distribution system operator's observatory [1]. While choosing a nation to develop algorithms that can produce synthetic network topologies, the Pearson correlation coefficient [194, 195] is used to examine correlations between the indicators and European Union nations (in, Chapter 3).

As discussed above, OpenStreetMap data [196] are considered for generating geo-referenced synthetic networks in the thesis. However, extremely high computational performance and storage are required to handle data from OSM for all countries at once. Therefore, data

pertaining to a specific country were selected and downloaded from the Geofabrik server [197]. The osmosis [198] and osm2pgsql [199] tools were used for data processing. OSM data alone is insufficient for the development of synthetic networks. Open power system data [200] is used to obtain data on the location of power generators. After extracting the essential components from OSM data, data related to electrical components were obtained from [201, 202] and compared to the extracted data to validate the retrieved components. To understand the OSM data deficiency and the incompleteness of OSM data, the retrieved buildings must be labeled to assign specific load profiles to perform power flow simulations. Therefore, a detailed literature review was conducted to achieve building-type classification [203-211]. But this thesis classified the building types extracted from OSM data after identifying the limitations of these investigations. Nevertheless, to increase the data quality of the OSM data prior to categorization, data from additional sources such as CORINE [212], building heights data [213], and census data [214] were also used. Moreover, several labels extracted from OSM data were streamlined into a few necessary labels considering [215] during the preprocessing stage. After preprocessing the data, various models were evaluated, and the best-performing model was chosen for classification [216-219]. However, to validate the classification model's output, the data set from the federal statistical office of Germany [220] was considered (Implementation and Validation, Chapter 4).

In addition to the data for developing geo-referenced synthetic distribution network topologies, synthetic data describing distributed demand and generation were considered to include in the power flow simulations [221-223].

Apart from the open and synthetic data, additional algorithms [224-231] were used while estimating synthetic network topologies. Overall, synthetic low-, medium-, high-, and extra-high-voltage networks are estimated and validated [232] using these datasets and algorithms. Finally, to determine the influence of future distributed demand and generation in the distribution networks, a national energy systems scenario [233] was used as a baseline, and additional scenarios are generated. Using all this information, the following chapters demonstrate the evolution of geo-referenced synthetic network topologies and the assessment of future distributed demand and generation in the distribution networks.

2.4 Summary

This chapter began with the following question: “*What are the challenges involved in distribution networks?*”. First, the role of future distributed demand and generation to minimize CO₂ emissions was discussed. Secondly, a literature study was conducted concerning the effects of distributed demand and generation in the distribution networks. This study includes operational limiting factors and their associated violation limits, methods of analysis, tools, and network topologies. Finally, it was concluded that, on the one hand, a significant penetration of distributed demand and generation will disrupt network operation by bringing voltage violations, thermal loadings, and power flow violations. On the other hand, there is insufficient infrastructure available to analyze and plan the effects of future demand and generation in the distribution networks. Based on this information, it also becomes possible to ask the following

question “*Why is it necessary to estimate geo-referenced synthetic distribution network topologies?*”. The network infrastructure modeling required for analyzing the effects of future demand and generation is limited, as evidenced by the fact that most of the literature reviewed uses real, test/IEEE, and typical/benchmark network topologies. These network topologies are specific to a particular location and type and are typical to a nation. To find an alternative to these topologies, a literature study was carried out pertaining to synthetic power grid models. Based on the findings of this study, geo-referenced synthetic network models are published for transmission networks, while only a few models are offered for distribution networks. In addition, the available models are limited to specific locations, have a high degree of heuristics, and require significant external input data. Therefore, it is necessary to estimate geo-referenced synthetic distribution network topologies to answer the overarching question of this thesis.

Key messages:

- ✚ Voltage limiting factors, grid component loadings, and power flow constraints are key operational limiting factors that must be considered when making decisions to integrate future distributed demand and generation in the distribution networks.
- ✚ Only 4% of the literature reviewed considered the combination of heat pumps and distributed generation combination.
- ✚ Only limited literature has explored the combination of PV, heat pumps, and battery electric vehicles.
- ✚ When integrating distributed demand and generation into the distribution networks, overvoltage, undervoltage, transformer loading, line loading, and reverse power flows are all significant factors.
- ✚ According to consumer perspective, a deterministic and probabilistic method is preferred when estimating the operational limiting factors. While optimization is desired from the operators' standpoint.
- ✚ If numerous input parameters and networks are considered for simulation, a probabilistic technique is more computationally intensive.
- ✚ Because the purpose of this thesis is to analyze the integration of future distributed demand and generation for an entire nation, only a deterministic method is feasible.
- ✚ For power flow simulations, the open source PyPSA is chosen.
- ✚ Synthetic networks are essential because of the challenges associated with obtaining real-world networks and the inaccuracy associated with the performance on the test, IEEE, and typical networks and their reflection of the state of the entire country.
- ✚ Numerous synthetic transmission grid models have been developed using OpenStreetMap data; however, only limited models are available for generating synthetic distribution networks using open data. Moreover, these models use external data and heuristics.
- ✚ Considering these challenges and opportunities, this thesis will develop geo-referenced synthetic network topologies to evaluate future distributed demand and generation in the distribution networks.

3 Distribution Networks Characteristics

Which graphical and electrical parameters are necessary for the generation and validation of geo-referenced synthetic distribution networks?

Is it possible to estimate geo-referenced synthetic distribution networks for any country in Europe using algorithms developed for a specific country?

Topics covered:

-  Graphical parameters necessary for the development of synthetic distribution network topologies
-  Electrical parameters essential for the development of synthetic distribution network topologies
-  Examining real-world network parameters and the reliability of supply indicators for European countries
-  Analysing the structural and performance-based characteristics of real-world networks to determine whether a generic algorithm is capable of generating synthetic network topologies for European countries

In the previous chapter, the estimation of geo-referenced synthetic distribution network topologies is suggested as a solution to address the thesis's main research question. With this background, this chapter will provide more contextual information surrounding the estimation of the geo-referenced synthetic distribution network topologies. First, the graphical and electrical characteristics of the distribution networks are discussed. Then, real-world distribution network indicators are analyzed to answer whether developing generic algorithms utilizing data from a specific country can be used to estimate synthetic distribution networks for other European countries.

The information provided in this chapter will be used in the subsequent chapters, either in estimating the synthetic network topologies or in validating developed networks. For instance, the total length of low-voltage networks in Germany is employed to validate geo-referenced synthetic low-voltage networks by comparing their total lengths.

3.1 Graphical Characteristics

Graphical metrics are particularly important when evaluating the network robustness of electrical network topologies. In the case of synthetic network topologies, graphical properties may be incorporated into the development process or serve as validating indicators for estimated synthetic network topologies. The graphical features characterize the network's interconnectivity, linkages, and integration of various components. These characteristics,

however, help estimate synthetic distribution network topologies through an additional dimension of precision and accuracy.

In principle, all networks – not just electric networks – are fundamentally depicted as graphs. A graph, (G) , is better represented as a collection of nodes $(n_i \in N)$ and edges $(e_{i,j} = (n_i, n_j) \in E)$ and the graph networks' nodes, when specifically referred to as electrical networks, can be loads, generators, transformers, and substations. Furthermore, the edges (E) are typically reserved for overhead power lines and underground power cables and are created by connecting nodes across the network.

At this point, it is necessary to point out that electrical network graphs are undirected and weighted. At any edge, the weight represents the maximum power flow or current that a line can carry, depending on the line type or length. Additionally, weights such as those on the edges are also included in the power network graph nodes. In this context, the weights assigned to the nodes represent the quantity of injected or withdrawn active and reactive power. Moreover, as mentioned earlier, power network graphs are undirected, because power can flow in any direction. In general, power flows from a higher voltage to a lower voltage level. Occasionally, however, due to increased generation on the low-voltage side, power flows in the opposite direction. This phenomenon is attributed to the increase in voltage at the generation.

The power network graph, on the other hand, is mathematically expressed as an adjacent matrix (A) . This is a two-dimensional symmetric matrix with n rows and n columns $(n \times n)$, where n is the total number of nodes in a graph. Figure 3-1 illustrates an adjacent matrix representation for a simple graph. If there is an edge between the nodes n_i and n_j in this graph, the entries in this adjacent matrix $A_{i,j}$ is equal to one; otherwise, $A_{i,j}$ is equal to zero. For example, in the given sample graph, there are five nodes and four edges. Therefore, it creates an adjacent matrix of (5×5) . The entities in the matrix are, however, determined by the edges. In this case, there is an edge between the nodes n_1 and n_2 . So, $A_{1,2}$ and $A_{2,1}$ is equal to 1. In contrast, there is no edge between n_3 and n_4 . So, $A_{3,4}$ and $A_{4,3}$ is equal to 0. This adjacent matrix is used in power flow analysis.

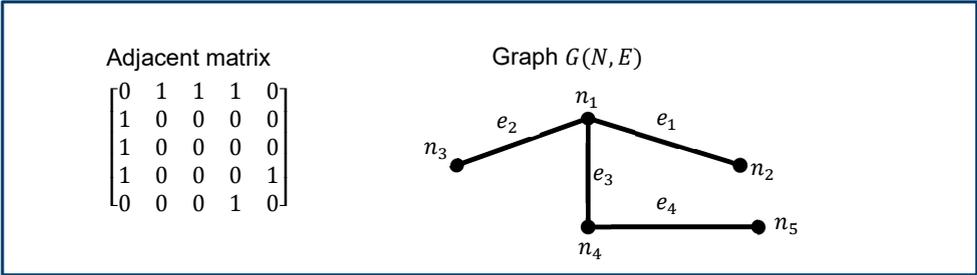


Figure 3-1 Sample graph with its adjacent matrix

With the graph of the power network defined, it is essential to consider its properties associated to estimate and validate geo-referenced synthetic distribution networks. From the sample graph in Figure 3-1, it is evident that the graph structure, nodes, and edges are indeed the three primary components of the graph. The characteristics of these individual components are, however, listed in Table 3-1. Several of these characteristics, including graph structure, number of nodes, number of edges, clustering coefficient, total length are crucial when generating synthetic distribution networks and perhaps validating estimated networks. Therefore, this section details each of these characteristics.

Table 3-1 Graphical characteristics corresponding to nodes, edges, and graph structure

Graph $G(N, E)$	Nodes ($n_i \in N$)	Edges ($e_{i,j} \in E$)
Radial structure	Network size	Number of edges
Ring structure	Node degree	Mean edge length
Mesh structure	Node degree distribution	Average shortest path length
	Betweenness centrality	Total network length
	Clustering coefficient	

3.1.1 Graph Structure

The graph structures of power networks vary according to the voltage level. The three most common distribution network structures are radial, ring, and mesh. In a radial network, there does not exist a closed path that connects the nodes. Nevertheless, several feeders connect the nodes with the transformer. For instance, in Figure 3-1, consider n_1 as a transformer node. Then, there will be three feeders with an edge $e_1, e_2,$ and (e_3, e_4) . In this type of configuration, there exists only one way. The power will travel from the transformer node to reach the feeder's end node. For instance, in Figure 3-1, the power to the node n_5 is only accessible via n_4 from n_1 . When these graph structures (i.e., radial) are used when designing power networks, it leads to low operating costs and simplified operations. It is one of the primary reasons why radial networks are frequently encountered in low-voltage networks. However, these network types come with certain disadvantages such as supply failure due to line breakdown. Figure 3-2 depicts radial, ring, and mesh structures in the power grid.

The ring configuration is another prevalent network form usually seen in low- and medium-voltage networks. Figure 3-2 illustrates the ring topology of the power grid and it is evident that this network setup always contains closed paths with a circuit breaker. In this setup, the feeder begins at the transformer node and terminates at the transformer node, thereby linking all the other nodes along the course of the feeder. In this type of network topology, there are some instances where they are connected with two or more transformers to continue maintaining supply in the event of a transformer failure. Additionally, during line overloading and breaking, a circuit breaker disconnects the fault line, providing an uninterrupted power supply. This type of network structure is common in medium-voltage networks and some low-voltage networks,

where a frequent failure of a line or transformer occurs. When ring networks are operated as an open ring by opening the circuit breaker, it results in the formation of a radial network.

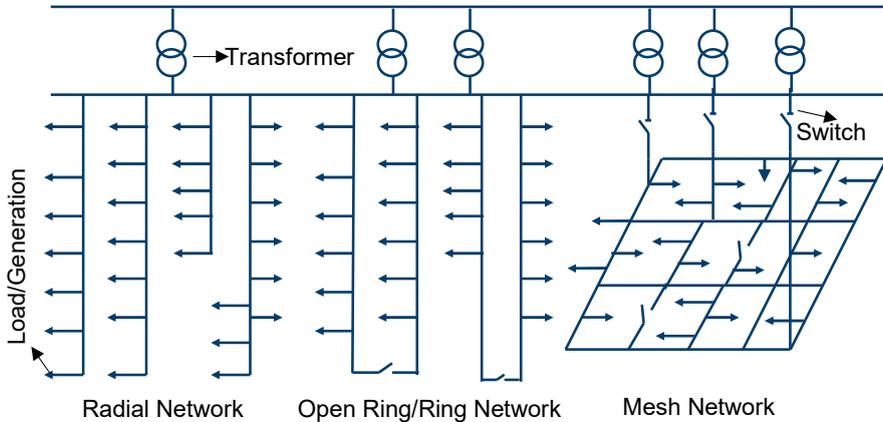


Figure 3-2 Power network graph structures

Finally, the mesh structure in the power grid is frequently utilized in high-voltage distribution networks and extra-high-voltage transmission networks. This arrangement, as represented in Figure 3-2, contains multiple transformers and substations. This configuration enables the transfer of power from one location to another through several routes. In addition, this network configuration always adheres to $(n - 1)$ security. In other words, in the case of a line or transformer failure, the network will ensure the uninterrupted flow of supply.

Apart from network structures, various other graph-based properties must be studied to estimate and validate geo-referenced synthetic distribution networks. Power networks are not usually simple structures; they can be quite complex. Occasionally, the connections between nodes and edges exhibit complex patterns. As such, ignoring patterns and focusing on nodes and edges using various indicators provides insight into the power network topologies. There are currently several statistical measures available to understand the topology through nodes and edges in a network. The most significant ones are already reported in Table 3-1. Nonetheless, the following subsections discuss the properties of nodes and edges in greater detail.

3.1.2 Node Specific Characteristics

Nodes are a fundamental component of a graph network, and the properties associated with nodes are vital when estimating and validating geo-referenced synthetic networks. Therefore, it is extremely crucial to gain a rudimentary understanding of these properties.

Network size (or) Total number of nodes:

The total number of nodes in a graph is the fundamental attribute that can be used to describe a graph network. The total number of nodes (T_N) determines the network size in terms of power networks. In a group, this metric equals the sum of all nodes. For instance, the basic graph seen in Figure 3-1 comprises five nodes. However, to describe the graph as a power network, all nodes are represented as demand or generation nodes, along with transformers. However, the total number of nodes helps determine the similarity of the networks.

Node degree:

It is possible to determine the connection to each node based on the total number of nodes by exclusively focusing on a single node and ignoring all other nodes. The connection to the nodes is formed by the edges in a graph. By observing the connections to each node, the node degree of each node can be determined in the graph. The average node degree of a network is defined as the average number of edges that connect each node in a graph. For example, if a graph has a total of N nodes and N_k is the degree of each node, and the average node degree is calculated by taking the mean of N_k . The average node degree is shown in Equation 3-1. This is a crucial indicator that provides information about the network topology. In other words, if the value of the indicator is more than two, the network is deemed to be mostly mesh structured. For instance, in Figure 3-1, the average node degree is 1.6, which is less than two. This indicates that the structure is not mesh. Hence, this indicator will be employed to validate the developed geo-referenced synthetic distribution networks. For instance, the distribution networks are almost always radial or ring structures. If the average node degree of a network exceeds two, then the developed networks will fail to meet the requirements of the considered network structure.

$$N_{Adeg} = \frac{1}{n} \sum_{i=1}^n n_{k,i} \quad \forall i \in N \tag{3-1}$$

Node degree probability distribution:

The node degree distribution reveals how many nodes have a similar degree. For instance, imagine a network with (N) number of nodes, wherein two have a degree (X), three have a degree (Y), and so forth. Using this information, one could determine that most nodes in the network have the same degree. While this may gather a limited amount of information, it is still advantageous for the network structure. Therefore, we consider node degree probability distribution while generating geo-referenced synthetic networks. The degree distribution of nodes $p_d(k)$ is defined as the fraction of nodes in a network with degree (K). If a network has (n_k) nodes of degree (k), the node degree probability distribution is represented mathematically in Equation 3-2. The node degree distribution specifies how many nodes have a similar node degree. Typically, the majority of nodes in an electrical network with a radial structure have a small degree. Nevertheless, if a node with a large node degree is observed within a radial structured electrical network, it is most likely the substation/transformer node.

Therefore, when validating the geo-referenced synthetic networks that are produced, this indicator is used to ensure that the transformer is properly positioned within the network.

$$p_d(k) = \frac{n_k}{N} \quad 3-2$$

Betweenness centrality:

To locate a node that is positioned on the shortest path between any other nodes, the betweenness centrality indicator is an optimal measure. The term "betweenness centrality" refers to the frequency at which nodes are on the shortest path between other nodes, normalized to $0 \leq B_c \leq 1$. Thus, the greater the value of a node's betweenness centrality, the more frequently that node is located on the shortest path. The betweenness centrality is calculated as indicated in Equation 3-3 [192] and is significant when identifying the node that can hold transformers or substations in a geo-referenced synthetic distribution network.

$$B_{c(n)} = \sum_{s,p \in N} \frac{\sigma(s,p|n)}{\sigma(s,p)} \quad 3-3$$

Here, N is the set of nodes, $\sigma(s,p)$ is the total number of shortest paths, and $\sigma(s,p|n)$ is the total number of those routes passing through some node (n).

Clustering coefficient:

The clustering coefficient is a metric that is used to identify nodes that may cluster in a graph. However, the clustering coefficient of a node is defined as the ratio of possible edges between nodes' neighbors. Accordingly, the clustering coefficient [193] of a graph is equal to the average clustering coefficient of its nodes. Equation 3-4 indicates the mathematical equation for the same. The indicator of the clustering coefficient is used to validate the estimated geo-referenced synthetic distribution networks. In fact, to validate the low- and medium-voltage networks, the clustering coefficient of the networks should be equal to zero.

$$C = \frac{1}{N} \sum_{i=1}^N \frac{e_G(i)}{\lambda_G(i)} \quad 3-4$$

Here, $e_G(i)$ is the number of edges between the neighbors of node i , $\lambda_G(i)$ is the total number of edges that can exist among neighbors of i .

In conclusion, this section ascertained that the above-mentioned indicators, such as network size, average node degree, node degree probability distribution, betweenness centrality, and clustering coefficient are helpful in generating and validating geo-referenced synthetic distribution networks.

It is also important to identify the edge-specific indicators which are crucial in the process of developing geo-referenced synthetic distribution networks. In the following section, an overview of the parameters pertaining to the edges is provided.

3.1.3 Edge Specific Characteristics

Edges, along with Nodes, form the core components of graph networks, and the edge-specific indicators are crucial to estimate, analyze, and validate the geo-referenced synthetic distribution networks. Therefore, it is essential to hold a firm grasp on the edge-specific indicators.

Number of edges:

The behavior of a network is influenced by the total number of edges in a graph. The sum of all edges in a graph, wherein at least one node with a degree one exists, provides the total number of edges. In the context of power system networks, the edges indicate overhead and underground power lines and cables. The total number of edges helps identify the networks' resemblance with real-world networks when performing validation with other networks.

Total network length:

Total network length, on the other hand, is primarily used to validate the generated geo-referenced synthetic networks by comparing the total network length of real-world networks to the total network length of the predicted networks. The total length of a network is defined as the sum of the length of all its edges. In Equation 3-5, the formula for computing the network's total length is provided. The total length of the network varies amongst networks. For example, the network length of a low-voltage network located in a suburban area is greater than that of a network located in a rural area.

$$T_{len} = \sum_{j=1}^m L_m \quad \forall j \in E \quad 3-5$$

Mean edge length:

Mean edge length takes all edges in a graph network into consideration and is extremely important when examining nodes that are spread apart. The mean edge length is an effective comparative measure when considering other graphs and is determined by dividing the network's total length by the total number of edges. Additionally, the mean edge length, like the total length, aids in determining the network's type. For instance, the network type can be distinguished as rural, semi-urban, or urban networks by comparing the mean edge length of the network to that of real-world networks.

Average shortest path length:

The average shortest path length is another significant edge-related property. For a network, it is defined as the mean length along the shortest path for all possible path of network nodes. Average shortest path length can be calculated by Equation 3-6 and is employed to validate networks and locate the placement of the transformer in the development of geo-referenced synthetic low-voltage networks. This parameter aids in the generation of radial graph networks in the medium-voltage network and its application is detailed in Chapter 5.

$$aspl = \frac{1}{2n(n-1)} \sum_{i \neq j} l(n_i, n_j) \quad 3-6$$

Here, $l(n_i, n_j)$ is the shortest path between the nodes n_i and n_j and n is the number of nodes in the network.

In conclusion, the above-mentioned indicators, such as the number of edges, total network length, mean edge length, and average shortest path length are seen to be useful in generating and validating geo-referenced synthetic distribution networks, especially for validating the developed networks.

In this section, several indicators related to graphical network are explored in detail, including graph structures, node-specific properties, and edge-specific properties. Among these, certain indicators can be used directly to estimate geo-referenced synthetic distribution networks, while others can be used to validate the estimated synthetic networks. However, to validate the generated geo-referenced synthetic networks, these parameters that are related to nodes and edges were retrieved from the developed networks data and compared to real-world networks. Chapter 5 will examine various approaches for validating generated geo-referenced synthetic networks.

3.2 Electrical Characteristics

Till this point, complex network parameters were analyzed in terms of graph features. However, several requisite electrical properties should be studied and analyzed before the graph network can be transformed into an electrical network. Unlike the graphical properties, some electrical characteristics are time-dependent and might vary with time. Therefore, all the electrical characteristics necessary for estimating and validating synthetic networks, whether time-dependent or not, will be introduced. In this section, two distinct types of characteristics are explored in detail namely, power-specific, and electric network component characteristics.

3.2.1 Power Specific Characteristics

In a network, power demand and supply are essential parameters that must be carefully considered while estimating geo-referenced synthetic distribution networks. It is necessary to ensure that supply always meets demand, failing which, the network will exhibit certain violations. Therefore, these critical power characteristics should be balanced in a network. Additionally, to calculate power demand and supply in the network, the graph nodes should be configured as electrical nodes with substations, consumers, and power plants. In this context,

power is imported when a node is configured as load while if the node is configured as a generator, power will be exported to that node.

Power demand:

In any network, the total active power demand ($P_{D,total}$) is defined as the sum of the individual power demands at each node specified as load. However, if a node in the graphical network is identified as a load node, the time series active power associated with that load is integrated at that node. To calculate the total power demand in the network, the summation of all the active power at load nodes must be computed. However, Equation 3-7 corresponds to the total active power demand in the network.

$$P_{D,total} = \sum_{i=1}^n P_{D,ni} \forall ni \in N, i \in L \quad 3-7$$

In this equation, L is the set that contains load nodes.

Power supply:

In addition to power demand, there exists power supply. In any network, the total active power supply is defined as the sum of the active power supplied by each generator. However, if a node in the graphical network is identified as a generator node, the time series active power associated with that generation is integrated at that node. Equating the sum of all the active power at generation nodes provides the total power generation in the network. Equation 3-8 corresponds to the total active power supply in the network.

$$P_{S,total} = \sum_{j=1}^n P_{S,nj} \forall nj \in N, j \in G \quad 3-8$$

Where, G is the set that contains generator nodes.

Additionally, if a node has both demand and generator units, the difference between demand and supply will be demanded or supplied from that node.

In this section, it is clear that the power supply and demand are two parameters that must be assigned to each node based on its state to convert a graphical network to an electrical network. The following section will explain the characteristics of the components of the electrical network.

3.2.2 Electrical Network Component Characteristics

Substations, power lines, and electrical nodes are the three primary components of a network. Thus, they must be properly utilized to ensure network stability. The utilization metrics of these components will determine the network's stability, in both synthetic and real-world networks.

In fact, the network component metrics are introduced in this section because they are essential during the validation phase and when evaluating the network's performance.

Substation/transformer utilization:

The transformer is the first component in the electrical network that is of a high level of significance. The transformer's utilization factor is the feature that characterizes the state of the transformer. Transformer or substation utilization is defined as the average ratio of each substation's or transformer's apparent power supply to its maximum supply. Equation 3-9 illustrates the formula corresponding to transformer utilization. This parameter is significant for monitoring network operating limit violations, which were covered in Chapter 0. It is necessary to ensure that transformer or substation utilization is always at a minimum to ensure continuous functioning and the maintenance of the component's life. In reality, numerous types of transformers with varying capacities are available today. Chapter 5 discusses various types of transformers that are available for each network level. For instance, in low-voltage networks, the capacities of the available transformer types range from 100 kVA to 630 kVA [232]. In rare cases, a transformer type with 1000 kVA capacity will be included in the low-voltage networks. For the geo-referenced synthetic network, a transformer type will be selected based on the network's load.

$$U_t = \frac{1}{n_i} \sum_{i=1}^n \frac{|S_{s,i}|}{|S_{s,max}|} \quad 3-9$$

Here, n_i is the number of substations, $S_{s,i}$ is apparent power at the i^{th} substation, and $S_{s,max}$ maximum capacity of the substation.

Power line utilization:

Like transformer utilization, power line utilization is an essential factor of the electrical network that must be addressed. In real-world power networks, the two types of power lines include overhead lines and underground cables. Each type of power line has a unique size and current carrying capacity. The most popular overhead lines and underground cables utilized in the real-world networks will be discussed in Chapter 5. As mentioned previously, it is necessary to learn about power lines utilization to study the stability of the network. Equation 3-10 illustrates network's power line utilization. It is the value of the average apparent power flow through the lines at their maximum capacity. When the utilization of power lines is known, the line that violates the limits can be reinforced before a breakdown occurs. However, the graph network edges are configured with electrical line parameters to represent as an electrical network.

$$U_{pl} = \frac{1}{m} \sum_{i=1}^m \frac{|S_{m,i}|}{|S_{s,max}|} \quad \forall m_j \in E \quad 3-10$$

Here, $S_{m,i}$ is the apparent power flow in i^{th} line, m is the total number of electrical lines in the network, and $S_{s,max}$ is the maximum apparent power that a line can carry.

To summarize, electrical loads, generators, transformers, and power lines are configured on graphical nodes and edges to depict a graphical network as an electrical network. For the network to remain stable, the power-specific and network component characteristics should be satisfied and within acceptable limits. These factors will be considered in chapter 5 when estimating geo-referenced synthetic network topologies. The following section will provide the real-world networks graphical and electrical characteristics.

3.3 Real-world Graphical and Electrical Network Characteristics

Apart from the graphical and electrical parameters that an electrical network is supposed to hold, real-world parameters are necessary to validate and estimate geo-referenced synthetic distribution network topologies. As stated above, it is challenging to obtain data from the distribution system operator. Similarly, estimating geo-referenced synthetic networks without any data for validation is naive. However, the distribution system operator observatory [1] contains various parameters from several distribution system operators across Europe, which are categorized for different countries. Therefore, all these metrics or indicators are evaluated to gather knowledge necessary for estimating geo-referenced synthetic networks and validating developed networks.

In this context, the characteristics collected from the distribution system operator’s observatory [1] must be analyzed comprehensively to determine whether a general algorithm for producing geo-referenced synthetic distribution network topologies can scale to any European country. To do this, network indicators and system reliability indicators are examined using data from the distribution system operator observatory (2020). Table 3-2 provides a list of the indicators that are available for analysis.

Table 3-2 List of real-work networks indicators and system reliability indicators taken from distribution system operators observatory [1]

Network indicators	Reliability indicators
Total network length	SAIDI
Low-voltage network length	SAIFI
Low-voltage underground cable length	
Low-voltage overhead line length	
Medium-voltage network length	
Medium-voltage underground cable length	
Medium-voltage overhead line length	
Low-voltage transformers	

The observatory of the distribution system operators gathered network statistics from different distribution system operators in 27 European nations. According to the study [1], the

distribution system operators in each country range from 1 to 883 , for a total of 2400 Distribution System Operators (DSOs). Among them, Germany has more distribution system operators of any country studied (i.e., 883 as per [1] for year 2020). Among these 883 distribution system operators, 80 serve a consumer base of more than 100,000.

The data for real-world networks, which were collected from the distribution system operators, were classified into two categories: network indicators and reliability of supply indicators (see Table 3-2). In this section, the statistics for each indicator in each category will be discussed. These indicators combinedly help determine if generic algorithms built on a specific country can be used to generate geo-referenced synthetic network topologies for all other European countries.

3.3.1 Network Indicators

This section elaborately discusses the network indicators for the electrical networks. The indicators relating to lines, cables, and substations considered in network structural indicators [1] will play a significant role in both estimating and validating the geo-referenced synthetic distribution network topologies.

As previously stated, an electrical network's primary components are edges, nodes, and transformers. However, collecting information related to edges, such as the length of each network from distribution system operators, is extremely difficult and unrealistic. Therefore, the study [1] only examines the overall circuit length for each voltage level in the distribution level and the proportion of overhead and underground power lines. Secondly, nodes essentially correspond to the end-consumers, and, for a variety of reasons, they are not documented in the study and therefore not considered in this section's analysis. Finally, from this study [1], the total number of transformers in each European nation is considered.

Since all the required indicators are collected, this section first analyzes the overall circuit length, including low-, medium-, and high-voltage networks for all European nations that are considered in this study.

Total circuit length:

The total network length at all voltage levels at the distribution level – low-voltage, medium-voltage, and high-voltage – is presented in Figure 3-3 for each of the 27 European nations reviewed in the study [1]. The total network length expressed here is in kilometers (km). Figure 3-3 demonstrates that when all voltage levels in the distribution level are considered together, Lithuania (LU) has the shortest total circuit length of 8,477 km. Germany (DE), on the other hand, has the longest route at 1.7 million kilometers. It is interesting to note that in Figure 3-3, only 30% of the nations have a total electrical power network length exceeding 250,000 km. Interestingly, except France (FR), the total network length of all other nations is less than the low-voltage network length of Germany (DE). The number of distribution system operators (i.e., 883) operating distribution networks in Germany clearly indicates the existence of several networks within the distribution system. Thus, the longer overall circuit length when all

networks are included can then be justified. The total circuit length for each nation that is specified in this study will be used to validate the geo-referenced synthetic distribution networks.

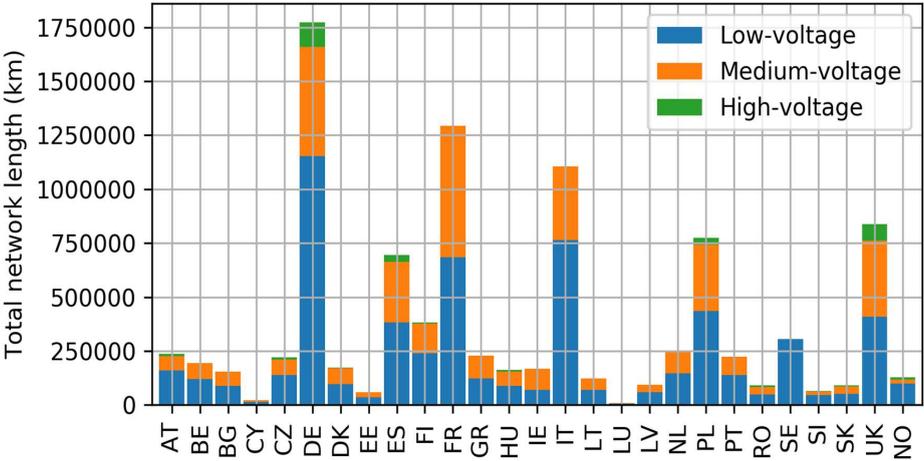


Figure 3-3 Total network length (km) considering low-, medium-, and high-voltage levels for European nations [1]

However, to validate geo-referenced synthetic networks at each voltage level precisely, the total lengths associated with low-, medium-, and high- voltage levels are evaluated separately. The length of high- voltage network for a few nations is illustrated here because low- and medium-voltage networks are operated at the distribution level in all the countries. In some nations, high-voltage is also operated at distribution level in addition to low- and medium-voltage. On the other hand, some nations refer to high-voltage networks as a transmission network. For example, EU member states such as Cyprus (CY), Estonia (EE), France (FR), Italy (IT), Lithuania (LT), and Latvia (LV) run low- and medium-voltage as distribution grids and high-voltage as transmission grids (See Appendix A, Figure 9-1). This may explain why Lithuania (LT) has the shortest distribution network length, as previously discussed. Due to the fact that different nations operate distribution networks at varying voltage levels, low-, medium-, and high-voltage networks are examined independently.

With this background, the length of the low-voltage network is analyzed first in each of the countries examined in this thesis.

Low-voltage network length:

Figure 3-4 illustrates the total length of low-voltage networks in all European nations considered in this study. As mentioned earlier, the power lines can either be overhead lines or underground cables depending on the location of the operation. Moreover, the parameters associated with each line type vary due to the operational characteristics. Thus, in generating

geo-referenced synthetic networks, the proportion of overvoltage power lines and underground cables for each voltage is crucial. In Figure 3-4, separate underground and overhead line is also mentioned in addition to the total circuit length. From the data in Figure 3-4, it is apparent that the proportion of overhead lines and underground cables for the Netherlands (NL), Slovakia (SK), and Norway (NO) could not be accessed due to a lack of data. Nevertheless, for the other nations, there appears to be a distinct ratio of different line types. Comprehensively observing this data reveals that Germany has the longest low-voltage network length of approximately 1.2 million kilometers. Unsurprisingly, Germany is an outlier with regard to the operation of low-voltage networks. Another interesting fact is that there are three countries - Germany (DE), Denmark (DK), and the United Kingdom (UK) – with an underground cabling percentage of greater than 80%. Out of these three countries, Denmark controls a whopping 95% of underground cable, surpassing Germany by 8%. From these observations, it is apparent that Denmark is moving away from traditional distribution networks. Nonetheless, Germany's low-voltage network is the longest in comparison to other nations, and 87% of it is situated underground, indicating that Germany is ahead of the curve in terms of network modernization. In contrast, Greece (GR) and Hungary (HU) continue to have more than 90% of their network covered by overhead lines. Despite of these differences, the data pertaining to the share of underground cable and overhead line will still be used in generating geo-referenced synthetic distribution networks and data regarding the total low-voltage length while validating the synthetic networks in chapter 5.

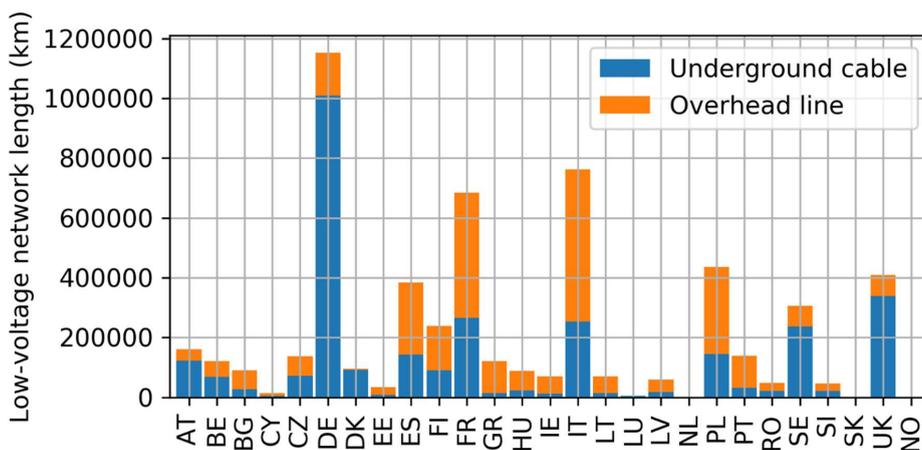


Figure 3-4 Total low-voltage network length (km) representing underground cable and overhead line length [1]

Medium-voltage network length:

In this study, the medium-voltage circuit length, and the proportion of underground cable and overhead line length were considered in this regard. The statistics corresponding to them are presented in Figure 3-5 for the European countries considered in this study. Surprisingly,

France (FR) possesses the longest medium-voltage network, surpassing Germany by nearly 0.1 million kilometers. Nonetheless, Germany's medium-voltage circuit length is around 0.5 million kilometers. In terms of the underground cable length and overhead line length, the underground length of Belgium (BE) and Denmark (DK) exceeds 80%. On the other hand, Germany ranks third with a slightly lower underground cable percentage of 75%. This dataset will be used in the study to generate synthetic networks. Moreover, these data will also help choose the nation when developing generic algorithms.

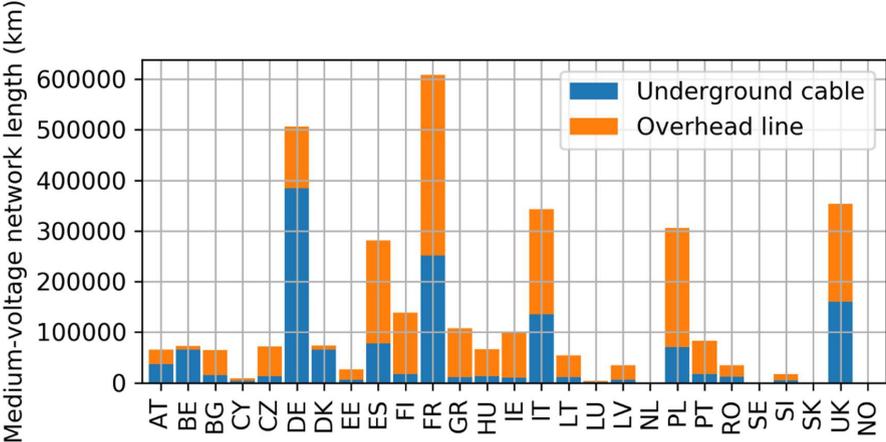


Figure 3-5 Total medium-voltage circuit length (km) representing underground cable and overhead line length [1]

In addition to low- and medium-voltage, high-voltage voltage network lengths are examined for some countries. Only 80% of the nations examined in the study [1] operate high-voltage networks at distribution level. Among these countries, only 60% or 75% of reviewed EU nations operating high-voltage network at the distribution level provided data pertaining to the high-voltage network length. Unsurprisingly, similar to the low-voltage network, the longest high-voltage electrical network is installed in Germany, approximately spreading across 0.1 million km. However, the overall high-voltage network level length for the considered EU nations in the study [1] is depicted in Figure 3-6. It is evident in the chart that Germany (DE) is still an outlier in terms of the total high-voltage network length installed. To conclude, by reviewing the data for all the voltage networks at distribution level, it is evident that Germany (DE) has substantial data in terms of network lengths.

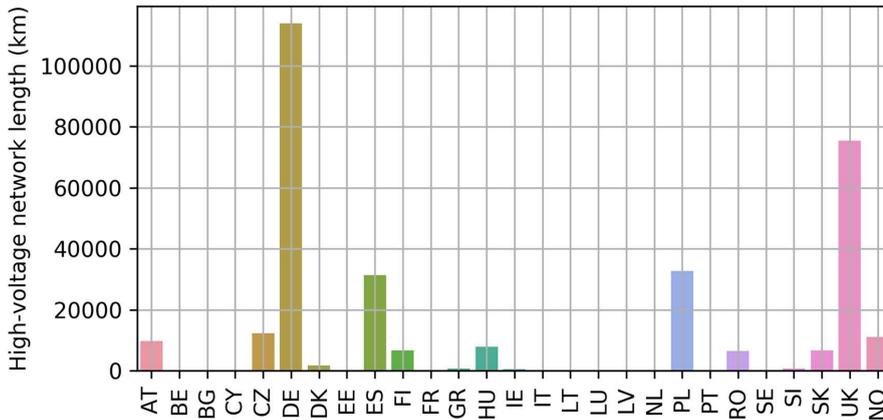


Figure 3-6 Total high-voltage network length represented in kilometers (km) [1]

The final network indicator required to develop geo-referenced synthetic distribution networks is the total number of low-voltage transformers. Figure 3-7 illustrates the total number of low-voltage transformers installed in the EU nations considered in this study. There seems to be a directly proportional relationship between the total length of the low-voltage networks combined and the number of transformers. Conversely, France (FR) and the United Kingdom (UK) have a shorter circuit length than Germany but have more transformers. For instance, the number of transformers in France (FR) exceed those in Germany by 289,100. However, the total low-voltage circuit length is 40% shorter than Germany. Thus, external sources will be able to determine the network structure independently of internal components and characteristics. Due to this limited information, it is not possible to present a precise explanation about more transformers with a shorter network length. Nevertheless, in the case of the shortest length, there are instances where more transformers are required owing to the loading condition and spatial distribution. However, the primary objective here is to collect data that will aid in the development of geo-referenced synthetic distribution network topologies. The data are utilized in chapter 5 to develop synthetic networks. Furthermore, this information is also used to choose the nation for the purpose of developing algorithms by determining the correlation between the nations.

In summary, the total network length indicator is utilized to validate the geo-referenced synthetic network topologies that have been developed. The share of underground and overhead lengths in each voltage level will be used to construct the final electrical network. Finally, the total number of low-voltage networks serves as the foundation for the entire developmental process of synthetic networks.

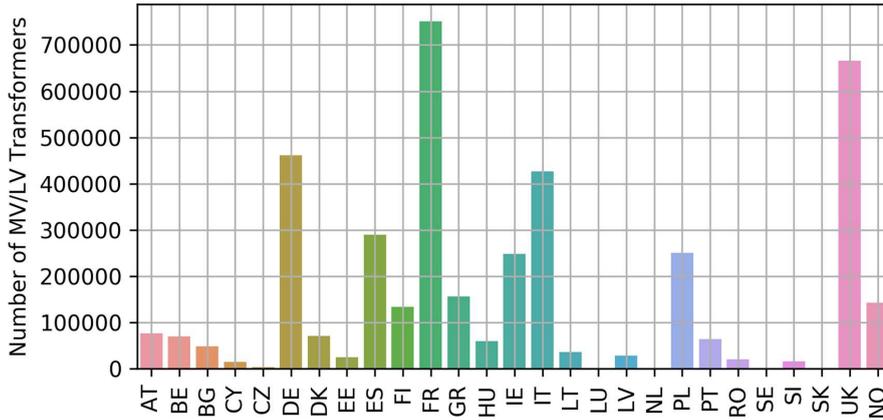


Figure 3-7 Total number of low-voltage transformers [1]

In addition to network indicators, this thesis considers reliability indicators for EU nations to investigate the quality of service provided by the distribution system operators. It is also crucial because a supply loss in the network indicates that the network is configured ineffectively in terms of integrated demand and supply conditions. Therefore, it is necessary to investigate the reliability of supply indicators prior to building synthetic distribution networks.

3.3.2 Reliability of Supply Indicators

In this analysis, two forms of reliability indicators are considered: SAIDI and SAIFI, which are also referred to as reliability of supply indicators. These indicators vary for each unsupplied load according to the network's topology and location within the network.

SAIDI:

SAIDI (System Average Interruption Duration Index) is defined as the average interruption duration for each consumer served [32]. In other words, SAIDI indicates the system condition in terms of the duration of interruptions wherein the interruptions may be planned or unplanned. Unplanned interruption in a system highlights the system's ineffective configuration and uncertainties in the system. Therefore, it is crucial to investigate such interruptions. Equation 3-11 corresponds to SAIDI. SAIDI is measured in units of time, mostly in minutes or in hours.

$$SAIDI = \frac{\sum U_i N_i}{N_i} \quad 3-11$$

Here, U_i is the annual interruption time for i^{th} location and N_i is the number of consumers at i^{th} location.

Data from the distribution system operators' observatory [1] was used to study SAIDI in minutes, ranging from 2002 to 2014 for the EU nations that were studied. For this analysis, the

most recent statistics from the year 2014 are further analyzed. Figure 3-8 illustrates the amount of time lost in 2014. Note that both Cyprus (CY) and Slovak Republic (SR) did not contribute any data pertaining to the SAIDI for the year 2014. Nevertheless, as illustrated in Figure 3-8, Slovenia (SI) has observed the longest time of system disruption. However, 80% of the countries considered have recorded less than three hours of unplanned interruption per year (i.e., 2014). This means the existing networks are out of order for a substantial amount of time and require further maintenance to function properly. However, it is also necessary to determine the frequency with which the interruptions occurred in the networks to ascertain the status of the network in each country. Therefore, SAIFI is considered further.

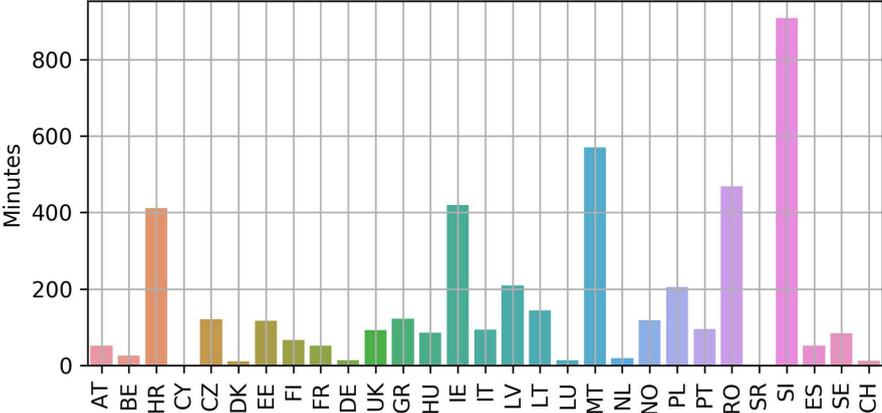


Figure 3-8 SAIDI (Minutes lost in 2014) [1]

SAIFI:

SAIFI (System Average Interruption Frequency Index) is defined as the number of times a system customer experiences interruption during a particular period of time [32]. This indicates the system's condition. SAIFI calculations can be performed using the Equation 3-12 and it is measured in units of interruptions per consumer.

$$SAIFI = \frac{\sum \gamma_i N_i}{N_i} \tag{3-12}$$

Here, γ_i is the failure rate for i^{th} location, N_i is the number of consumers at i^{th} location.

Figure 3-9 shows the SAIFI recorded for the year 2014. Cyprus (CY), Latvia (LV), and Lithuania (LU) have not provided any information regarding SAIFI for the year 2014. Among countries that provided this information, Slovak Republic (SR) and Romania (RO) have recorded the highest number of interruptions per consumers. On the other hand, Denmark (DK), Germany (DE), the Netherlands (NL), and Sweden (SE) have reported the lowest unit of interruptions.

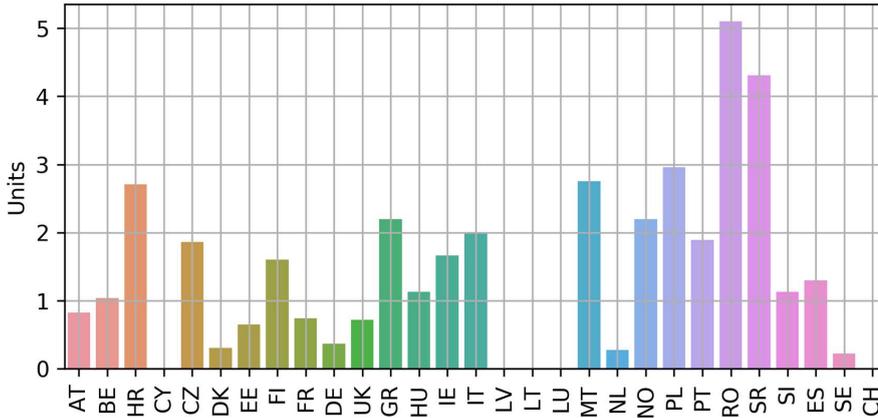


Figure 3-9 SAIFI (Average interruptions per consumer in 2014) [1]

Since the important network and reliability of supply indicators have been analyzed in detail, the following section examines the conditions for selecting a nation and determining if the generic algorithms developed for a nation is sufficient to construct geo-referenced synthetic network topologies for other EU nations. This study is conducted by analyzing network indicators and reliability indicators together.

3.3.3 Analysis on Network and Reliability Indicators

Data pertaining to real network indicators and reliability indicators that are received from the distribution system operator observatory [1] and described in the previous section were further evaluated to answer the following question:

"Is it possible to estimate geo-referenced synthetic distribution networks for any country in Europe using algorithms developed for a specific country?"

This question is addressed by studying the correlation between the characteristics and the considered EU nations by using a correlation coefficient measure. A correlation measure helps determine the relationship between specific features of data that has been provided. Indeed, correlation is a well-known measure of similarity between two features. For the given data, the correlation coefficient of the features lies between -1 and 1 . For instance, if two features are linearly dependent on one another, the correlation coefficient between them is $+1$ or -1 . If no relationship exists between the features, the correlation coefficient is 0 . A value closer to one indicates a high positive relationship, whereas a value closer to -1 indicates a strong negative correlation. On the other hand, a value closer to zero indicates a lower association, whereas a value equal to zero indicates no correlation between the features.

In this study, the first step entailed the evaluation of correlations between network and reliability indicators for all evaluated EU nations using the Pearson correlation coefficient [194, 195].

Pearson correlation coefficient is the ratio of a variable's covariance to the product of its standard deviations. Equation 3-13 denotes the formula for calculating the Pearson correlation coefficient. Applying Equation 3-13 on network and reliability indicators provides the correlation between the features. Figure 3-10 illustrates the correlation coefficient between network indicators and reliability indicators.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X\sigma_Y} \tag{3-13}$$

Where, $cov(X,Y)$ is the covariance of two variables and $\sigma_X\sigma_Y$ is the product of variables' standard deviations.

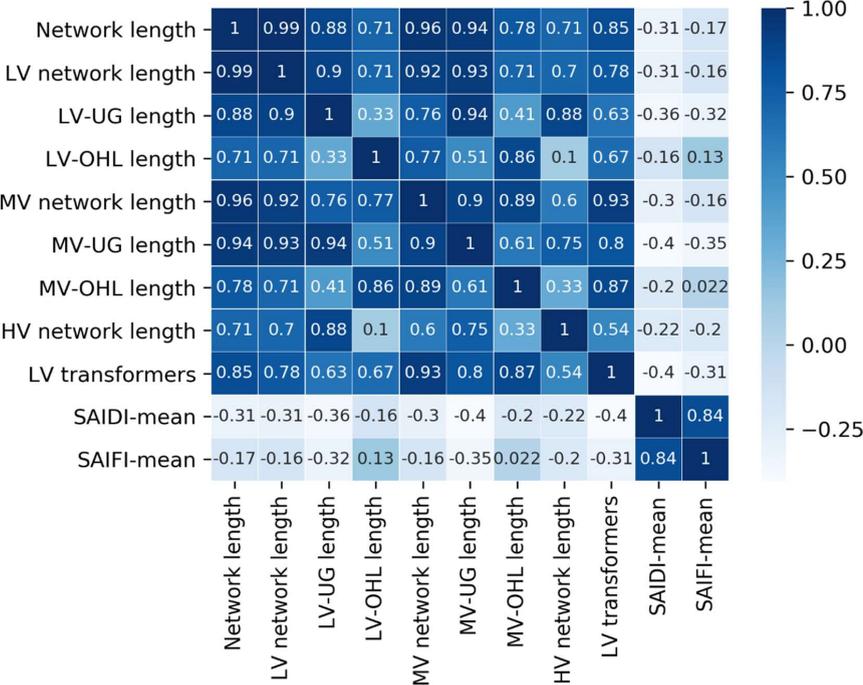


Figure 3-10 Correlation between network indicators and reliability indicators¹

Considering the total network length as a target variable, all other network indicators are closely related to each other with a correlation coefficient that exceeds 0.7. Since the total network length is the primary parameter associated with the network, it is deemed as the target variable. The correlation analysis further reveals that the reliability indicators are weakly related to the network indicators. Because reliability indicators are highly dependent on the load conditions in each country and engage in a lower level of interaction with network indicators, it can also be concluded that reliability indicators have minimal influence while generating geo-

¹ LV: Low-voltage; MV: Medium-voltage; HV: High-voltage; UG: Underground cable; OHL: Overhead line

referenced synthetic network topologies. Nevertheless, since the network metrics seem to be highly correlated across all considered EU nations, this study further examines the correlations between the EU nations on these indicators.

Similar to the correlation of indicators, the correlation between considered EU nations were calculated using the Pearson correlation coefficient without excluding any of the network indicators. Here, Germany was chosen as the target variable because it has the longest network covering all voltage levels. Additionally, its total network length for an individual voltage level is the longest when compared to other EU nations. Furthermore, the reliability indicator values are significantly lower than those in other nations considered. Therefore, Germany was considered as the target variable, while Figure 3-11 illustrates the correlation coefficients for all other nations.

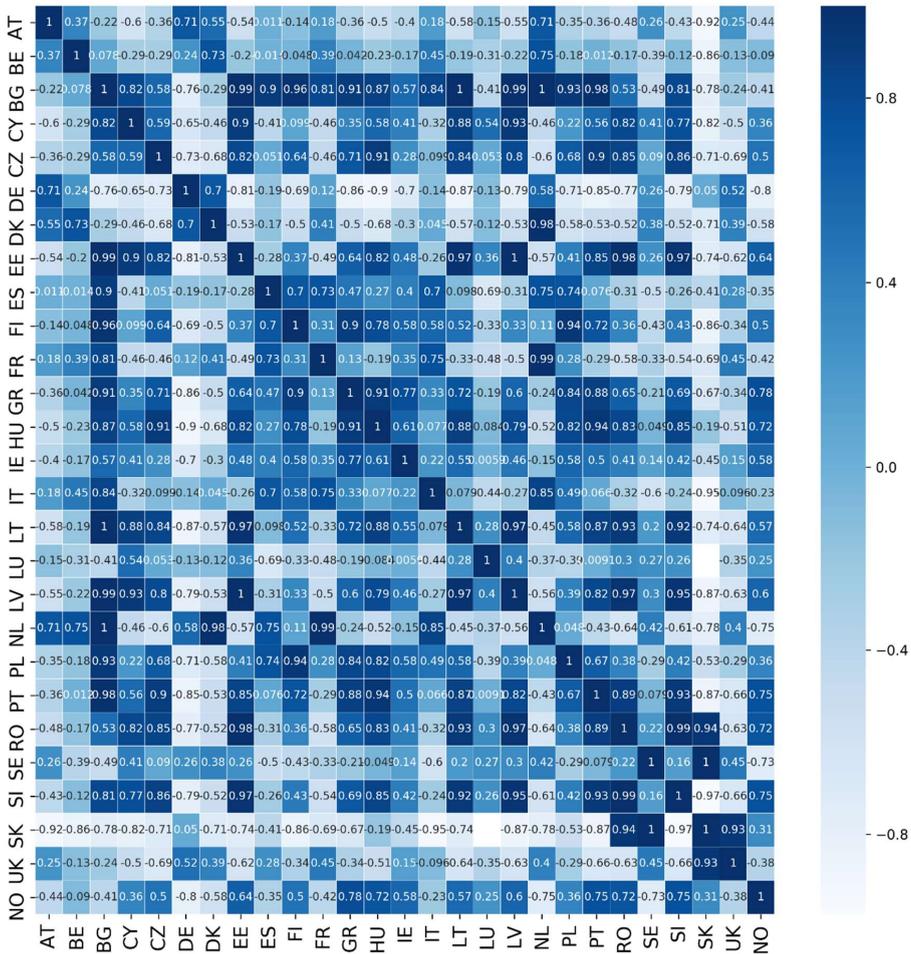


Figure 3-11 Correlation between considered EU nations

It is evident from the correlation results that 75% of the nations considered are correlated with Germany (DE). At the same time, Germany is also in correlation with 75% of the considered EU nations. Data from the study reveals that Germany has the longest total distribution network length and the lowest rate of SAIDI and SAIFI. Interestingly, if another country data is chosen to develop geo-referenced synthetic distribution networks, the challenges associated with higher node counts, longest power line lengths, and, most importantly, hundreds of thousands synthetic networks, will remain unresolved. However, the obstacles associated with these variables are described in Chapter 5, when discussing the geo-referenced synthetic network development process.

Therefore, the data retrieved from Germany would be the optimal choice for developing and training the algorithms. However, it can be argued that the algorithm developed using the data retrieved from Germany can be utilized to generate geo-referenced synthetic networks for other nations as well, since the algorithms will be trained on the most extreme data available when compared to data from other EU nations.

In this chapter, data from Germany was chosen to develop the algorithms to estimate geo-referenced synthetic network topologies. In the following chapter, the open and synthetic data required for developing algorithms and distribution network topologies for Germany will be discussed.

3.4 Summary

In Chapter 0, the development of geo-referenced synthetic distribution network topologies was emphasized as the next step. This chapter begins to answer the question “*Which graphical and electrical parameters are necessary for the generation and validation of geo-referenced synthetic distribution networks?*” To answer this question, several electrical power networks' graphical characteristics, such as graph structures, node specific, and edge specific properties as well as electrical characteristics, such as power specific, and component specific characteristics, are described. By examining these characteristics, several parameters required for the generation and validation of the geo-referenced synthetic networks are uncovered. To generate geo-referenced synthetic networks, first graph structures are necessary. In this case, radial networks are identified as a suitable choice for low-voltage networks, whereas ring structure operated as open-ring is suitable for medium-voltage networks. Second, in order to locate the placement of the transformers, the betweenness centrality and the average shortest path length characteristics will be used. Finally, the process of selecting a transformer type must not overlook the transformer utilization factor. Furthermore, to validate the developed geo-referenced synthetic distribution networks, several indicators that were studied will be utilized. First, clustering coefficient is used to identify the radial networks. Second, the number of nodes, number of edges, and the total network length will be compared with the real-world network characteristics. Third, node degree probability distribution is employed to validate the transformer node. Finally, transformers and line utilization are used to monitor network violations. Following the outcome of the preceding question, several real-work network and reliability indicators are analyzed for 27 EU nations to

answer the question “*Is it possible to estimate geo-referenced synthetic distribution networks for any country in Europe using algorithms developed for a specific country?*” The analysis concludes that using data from Germany facilitates the development of generic algorithms capable of generating geo-referenced synthetic distribution networks for other nation. The conclusion was attributed to the fact that 75% of the nations considered in this study are in correlation with Germany, and the extreme data qualities may effectively incorporate the uncertainties.

Key messages:

- ✚ The average node degree of a network reveals its structure. If the average degree of a network is greater than two, the structure is most likely mesh.
- ✚ Radial networks have zero clustering coefficient.
- ✚ Germany has a total circuit length of 1,772,696 km, which includes low-, medium-, and high-voltage networks [1].
- ✚ Germany's total low-voltage circuit length is 1,152,138 km [1].
- ✚ In Germany, 87% of low-voltage power lines are underground cables.
- ✚ 75% of Germany's medium-voltage power lines are underground cables.
- ✚ Germany has 461,000 low-voltage transformers [1].
- ✚ Considering the network and the reliability of supply indicators for all European countries, data from Germany were chosen to develop algorithms to create geo-referenced synthetic networks.

4 Open and Synthetic Data for Modeling

What data are required for estimating geo-referenced synthetic network topologies?

What are their limitations?

How can these issues be addressed?

Topics covered:

-  OpenStreetMap data acquisition and analysis
-  Development of classification model to classify buildings extracted from OpenStreetMap data
-  Statistical validation of the predicted buildings
-  Synthetic load and generation profiles

The foundation of this thesis entails furnishing the demand for the geo-referenced synthetic distribution network topologies and studying the graphical and electrical characteristics of distribution networks that are useful in generating synthetic networks, which were identified in the preceding chapters. However, this chapter provides an in-depth background of various open and synthetic data, with an emphasis on estimating geo-referenced synthetic distribution network topologies and a distribution grid model. It is worth noting that developing a model using open data will impose minimal restrictions on its use and keep it open for public research. All of the data required for model development are publicly available. For instance, consider OpenStreetMap data, which is populated with nearly all the components required for the development of geo-referenced synthetic networks, such as buildings, power lines, transformers, etc. These data are openly available. Nevertheless, the quality of the data should be analyzed before it is used for model development. Therefore, this chapter focuses on the assessment of OpenStreetMap (OSM) [196] dataset, as it is the most essential dataset for this thesis. In this context, this chapter describes the acquisition of OSM data, its analysis, as well as its limitations. Furthermore, buildings retrieved from OSM data are classified using state-of-the-art machine learning techniques in order to improve data quality and circumvent the limitations. Herein, several explicit and implicit algorithms are studied, and performance indices are used to analyze each model's performance in order to choose the optimal performance model. Furthermore, with the help of public statistics, the validation of the classification model is demonstrated. Following the assessment of building types, this chapter will provide synthetic load profiles for each building type, residential rooftop Solar Photovoltaic (PV) profiles, Battery Electric Vehicle (BEV) profiles, and Heat Pump (HP) profiles. In order to address the overarching question of this thesis, it is necessary to curate adequate load and generation profiles for the buildings.

4.1 OpenStreetMap Data Acquisition

In this thesis, the objective is to develop geo-referenced synthetic distribution network topologies using publicly available data. Thus, data pertaining to the location and type of buildings, the position and type of electrical transformers, the location and voltage level of electrical substations, power lines, and roads are all required. Therefore, this thesis relies heavily on OpenStreetMap [196] to obtain the necessary data.

The OSM is an open-source initiative that aims to develop a freely modifiable and publicly available world map. The OSM platform is established through crowdsourcing activities and enables users to freely contribute to its data by incorporating any missing information. A registered user can access its interface online to edit any of the information on the OSM platform. Simultaneously, any user may freely access and use this geographical data. Due to its growing volunteers as well as the high usage, the OSM's Geographical Information System (GIS) data is undergoing constant improvements. Therefore, the latest studies that rely on the GIS feature use OSM data due to the free availability of these geographic features. The OSM platform provides information about objects with a variety of properties, including buildings, roads, railroads, waterways, natural, aerial ways, boundaries, highways, military, public transportation, route, and power and all of this data are categorized according to their tags.

As previously stated, the OSM raw data are freely available and can be downloaded in either Extensible Markup Language (XML) or Protocol Buffer Binary File (PBF) format. To extract relevant information, understanding the structure of this data is essential. As illustrated in Figure 4-1, data in OSM are organized by three objects: node, way, and relation. For each of these objects, two distinct elements are employed to define them: *<attribute>* and *<tag>*. The data attributes, on the other hand, are represented as key-value pairs. Each object is assigned to an identification number (id), which is unique to each object and is one of the most significant attributes for distinguishing one object from another in a collection.

```

<osm version="0.6">
<node id="260734600" uid="13281" user="Francois" lat="23.81" lon="90.43" >
  <tag k=" amenity" v="school"/>
  <tag k="name" v="International School"/>
</node>
<way id="2492201729">
  <nd id="260734600" />
  <nd id=" 260734603" />
  <tag k="Building" v="yes" />
  <tag k="Building" v="school" />
</way>

```

Figure 4-1 OSM data example for representation in XML format

As previously stated, these objects are modified by individual users; each of these objects is further characterized by security-related attributes such as timestamp, user, and user ID. In addition to these essential attributes, each node is also defined by geographical details such as latitudes and longitudes. As a result of these geometrical attributes, the node object is represented as point geometry. Furthermore, each node encompasses a set of tags, which are expressed as key-value pairs. Similarly, the way object has a collection of nodes that represent an open or closed object. Thus, for the way object, the geometry would include a line for the open object and a multi-polygon for the closed object. Finally, the relation object contains data pertaining to the node and way relationships. The relation object stores a logical or graphical relationship between the nodes and ways. Figure 4-2 illustrates the general attributes and tags associated with these three objects.

Node	Way	Relation
<ul style="list-style-type: none"> . ID . Time stamp . User . User id . Latitude . Longitude . Free form tags 	<ul style="list-style-type: none"> . ID . Time stamp . User . User id . Nodes . Free form tags 	<ul style="list-style-type: none"> . ID . Time stamp . User . User id . Nodes . Ways . Free form tags

Figure 4-2 Fundamental attributes for node, way, and relation objects in the OSM data

The OSM dataset is of high variety and contains a high volume of data because this dataset contains all-encompassing geographical information about roads, buildings, etc., which fall under the high-volume category, and offers detailed information about various variables such as address, type, etc., which are high variety. For the planet, the raw dataset from the OSM is available as planet OSM data. However, the OSM server offers the planet's OSM data that provides information on the entire planet. But it will take up a large amount of space because the OSM file for planet is of high volume and variety. Additionally, the dataset is stored in XML/PBF format, making the process of retrieving the information highly complex. Consequently, Geofabrik server [197] was utilized to retrieve regional data. Geofabrik server stores data extracts from the OSM project, which is updated regularly. Therefore, the most recent data for Germany at the time of model development was downloaded from this server for modeling purposes [197].

However, the data include redundant information that is futile for estimating geo-referenced synthetic distribution networks. Therefore, an additional tool is required to extract the essential information from the regional file. Osmosis [198], a Java command-line tool was used to process the OSM data. This tool was used to filter nodes, ways, and relations objects. Power, buildings, and road objects are required to generate geo-referenced synthetic distribution networks. Thus, the key-value pairs in Table 4-1 are utilized to filter the tags in the three aforementioned objects. A command line shown below is inputted in the Osmosis tool to extract all objects related to power, buildings, and roads. That is why for each power, building, and highway key, a value of * was assigned.

Table 4-1 Key-value pairs for filtering the OSM data

Data	Key	Value	Output
Power	power	*	Substations, Transformers, Power lines, Power pylons, and Generators
Buildings	building	*	All type of buildings
Road	highway	*	Streets, highways, and footpaths

```
osmosis --read-pbf file=/germany-latest.osm.pbf --tf accept-nodes power=* highway=* building=*
--tf reject-ways --tf reject-relations --read-pbf file=/germany-latest.osm.pbf --tf accept-ways
building=* highway=* power=* --tf reject-relations --used-node --merge --write-pbf
file=/germany_bph.osm.pbf
```

In this case, the process should accept all nodes that correspond to the keys for power, building, and highway, as well as all associated values. Additionally, similar key-value pairs are accepted for way objects, ensuring that all components are gathered during the filtering process. However, despite filtering, the filtered file is still in PBF file format, which is unsuitable for data analysis and modeling purposes, particularly in this case, when the data is inputted into machine learning algorithms. As a result, the data were sent to the PostgreSQL server via the osm2pgsql tool [199].

```
osm2pgsql -r pbf -c -d DE_Building -U postgres -H localhost -S /bph.style --hstore -s -C 1200  
germany_bph.osm.pbf
```

Executing the preceding query creates three tables in the PostgreSQL server, each of which contains points, lines, and polygons. The point table is composed of several point elements with varying characteristics. Here, the most prominent feature is 'geometry,' which contains geometrical details. The points table geometry is comprised of 'POINT' geometries. In addition, the database also contains all the data essential to denote the point object type. The second table is a line table, which contains the geometry 'LINE'. The lines mainly symbolize all streets, highways, and roads, as well as power lines. In addition, each element is represented by its appropriate tag. Lastly, the polygon table contains the geometry 'MULTIPOLYGON'. Most of the items held here are residential buildings, non-residential buildings, and other enclosed structures, with each object including multiple tags to indicate the building's type and certain additional attributes. However, Figure 4-3 depicts a geographical representation of points, lines, and polygons in the form of a map for Germany. The World Geodetic System 1984 (WGS 84) is the Coordinate Reference System (CRS) (EPSG:3857) used for these datasets.

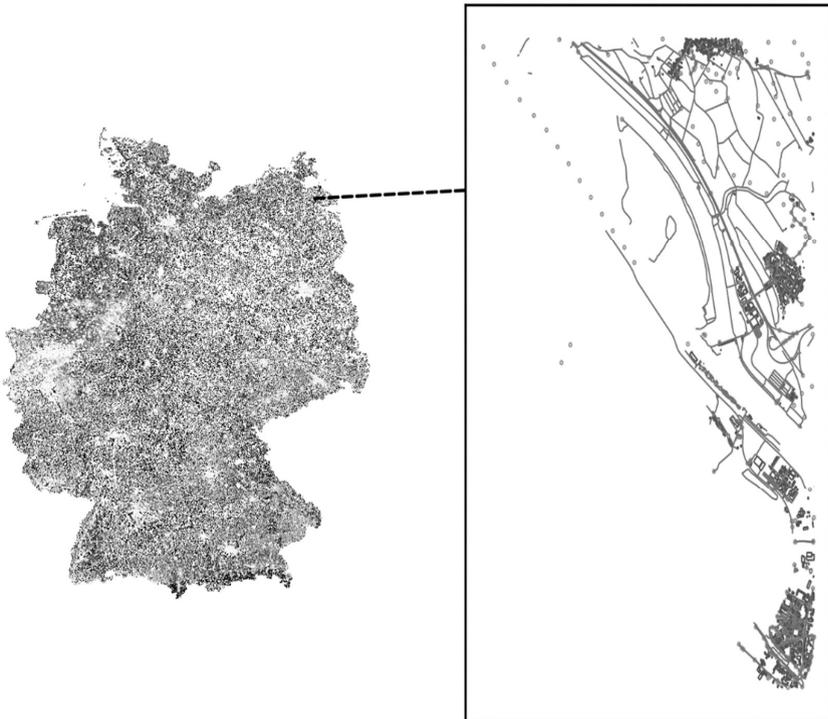


Figure 4-3 Geographical representation of points, lines, and polygons for Germany

Figure 4-3 illustrates the high volume of data represented in the form of points, lines, and polygons. The zoomed region in Figure 4-3 shows all the roads, footpaths, power pylons,

railway stations, railroads, etc., This information is still irrelevant for generating geo-referenced synthetic distribution network topologies. Therefore, further analysis of the extracted data is followed in the next section.

4.2 OpenStreetMap Data Analysis

After acquiring data from OpenStreetMap using a series of methods, the data were further evaluated to address data-related challenges. As previously stated, the data necessary to estimate geo-referenced synthetic distribution networks include buildings, roads, and power related components such as transformers, substations, and power lines. So, these data should be retrieved from the point, line, and polygon datasets and analyzed for the purpose of model estimation. The first step is to collect the power-related components from the given datasets.

To do so, the point dataset is evaluated first. For Germany, the point dataset has 4,484,462 data points in total. Additionally, each point object is composed of 71 attributes, which aid in distinguishing the various point types. Figure 4-4 depicts the major attributes of the points dataset.

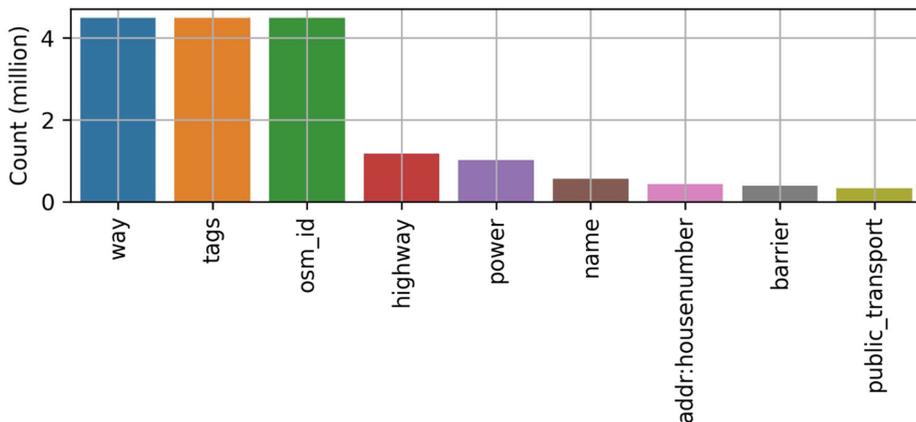


Figure 4-4 Count of attributes in point dataset for Germany

As illustrated in Figure 4-4, 1,015,806 points are associated with power. All other attributes, however, are not mandatory to extract transformers and substations. Hence, data points with only the power attribute are filtered out before being used for further analysis. Additionally, each point associated with the power attribute contains a tag denoting the power components. Pole, tower, generator, substation, transformer, portal, and terminal are the primary tags that contain the most amount of data compared to other labels (see, Figure 4-5).

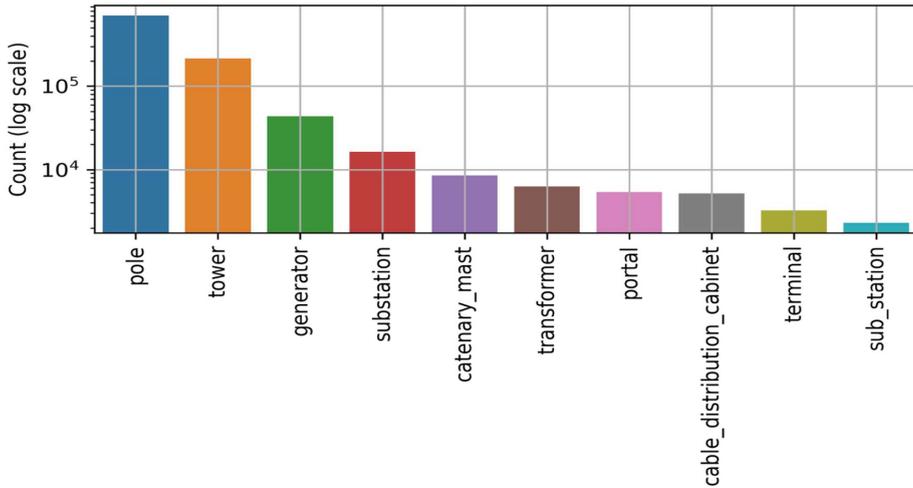


Figure 4-5 Distribution of power attribute tags

It is important to note that pole and tower tags are used to identify pylons, particularly those operating at high and extra-high voltage levels. The generator indicates the location of the power plant. Open Power System Data [200] contains publicly available data on the specific sites of power generators. During the data filtering process, generator stations accessible via OSM data are omitted. Regardless, the tags representing substation, transformer, and sub_station are used to filter out transformer locations, since a substation resembles a closed facility and frequently contains a common wall or fence which is not a point object. Additionally, extra tags pertaining to the voltage are evaluated to differentiate the transformers at various voltage levels, particularly using voltage-specific labels. However, because of user inactivity, not all points associated with transformers have these tags. Voltage-specific tags are present on 13.3% of the points with substation, transformer, and sub_station tags. Figure 4-6 illustrates the voltage-specific labels affixed to the points filtered using substation, transformer, and sub_station labels. For instance, 20000 label is assigned to 1500 points.

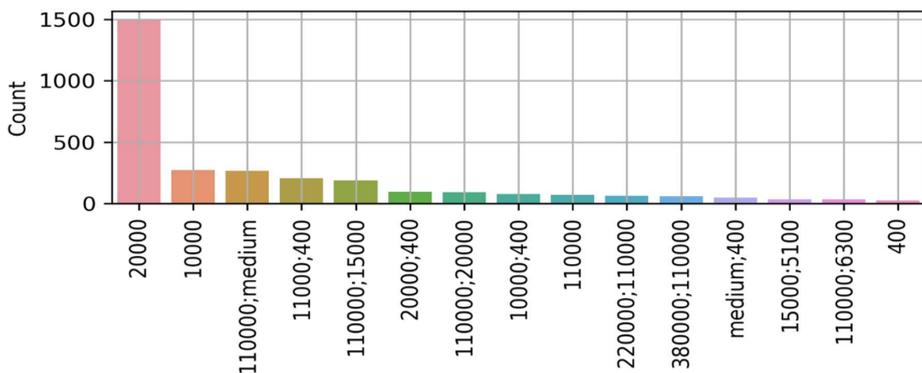


Figure 4-6 Voltage tags for point objects

However, a few points from each voltage-specific label are visualized spatially to review the labels and verify their consistency across voltage levels. This analysis indicates that the tags 20000, 10000, 10000;400, medium;400, 15000;5100, and 400 denote MV/LV transformers. Likewise, 110000;medium, 110000;15000, 110000;20000, 110000, 110000;6300, and 110000;10000 denote HV/MV transformers. Finally, 220000;110000 and 380000;110000 are EHV/HV transformers, respectively (see, Table 4-2). Using these tags assigned to each point, a total of 2,317 MV/LV transformer locations, 653 HV/MV transformer locations, and 129 EHV/HV transformer locations are recognized.

Table 4-2 Transformer point labels in OSM data

Point label	Assumption
20000, 10000, 10000;400, medium;400, 15000;5100, and 400	MV/LV transformers ²
110000;medium, 110000;15000, 110000;20000, 110000, 110000;6300, and 110000;10000	HV/MV transformers
220000;110000 and 380000;110000	EHV/HV transformers

However, as previously stated, medium and high-voltage substations are closed facilities and are commonly portrayed as polygons. The polygon dataset is further processed to identify the locations of low-voltage transformer and medium- and high-voltage substation. The polygon dataset contains 29,702,788 polygon objects with a total of 71 attributes, which are used to identify the type of the object. Figure 4-7 illustrates the attributes with a higher proportion of polygons.

² LV: Low-voltage; MV: Medium-voltage; HV: High-voltage

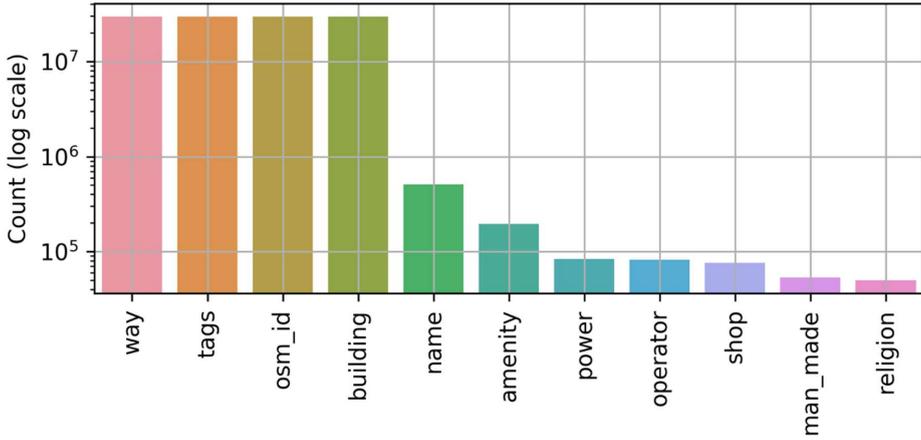


Figure 4-7 Count of polygons according to various attributes

As illustrated in Figure 2, there are 82,913 polygons associated with power and 29,619,992 polygons related to buildings in Germany. The buildings will be processed subsequently, followed by the extraction of power-related components. The retrieved polygons using power attribute are further divided into substations, generators, sub_stations, plants, transformers, and switchgear. As in the previous step, the generator and plants objects are omitted due to the availability of open data from open power system data [200]. Polygons, including substations, transformers, and sub_stations are of particular interest, and their voltage-specific tags are evaluated to determine power component locations at specific voltage levels. Approximately, 13.5% of power components are labeled with voltage-specific information. Figure 4-8 illustrates the distribution of several voltage-specific tags on power-related polygon objects.

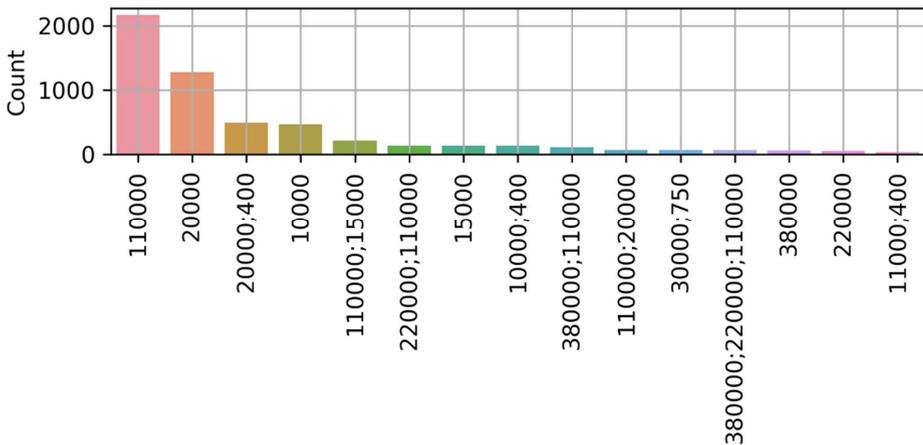


Figure 4-8 Distribution of voltage tags on polygon objects of power components

These tags are further analyzed and interpreted to extract the locations of substations at various voltage levels. After analyzing the tags via spatial plotting of the polygons and based on technical understanding, it was determined that tags 20000, 20000;400, 10000, 15000, 10000;400, and 30000;750 represent MV/LV transformer³, 110000, 110000;15000, 110000;20000, 110000;10000, 110000;30000, and 110000;medium represent HV/MV substation, and 220000, 220000;110000, 380000;110000, 380000, and 220000 represent EHV/HV substation (see Table 4-3). This information enables the identification of a total of 2,728 MV/LV transformers, 2,595 HV/MV substations, and 473 EHV/HV substations with their locations.

Table 4-3 Transformer polygon labels in OSM data

Polygon label	Assumption
20000, 20000;400, 10000, 15000, 10000;400, and 30000;750	MV/LV transformers
110000, 110000;15000, 110000;20000, 110000;10000, 110000;30000, and 110000;medium	HV/MV transformers
220000, 220000;110000, 380000;110000, 380000, and 220000	EHV/HV transformers

The geographical validation of extracted points and polygons revealed that two HV/MV transformers are contained within each medium voltage substation location. To determine the final HV/MV and EHV/HV substation locations, the points and polygons are merged and the HV/MV transformer points intersecting with the HV/MV substation polygons and the EHV/HV transformer points overlapping the EHV/HV substation polygons are removed. The point dataset supplied locations for several MV/LV transformers. Figure 4-9 illustrates the spatial distribution of MV/LV transformers, HV/MV substations, and EHV/HV substations and demonstrates that the number of MV/LV transformer locations that were identified were far fewer than expected. This may be attributed to the fact that the majority of MV/LV transformers are housed indoors within residential and commercial buildings. Although MV and HV substations are comparatively lower than existing LV transformers, the number of HV/MV and EHV/HV substation locations identified are more than the number of MV/LV transformers. The reason is that since these substations are large structures, they are obviously visible in satellite images. The validation of the substations and transformers identified follows.

³ LV: Low-voltage; MV: Medium-voltage; HV: High-voltage; EHV: Extra-high-voltage

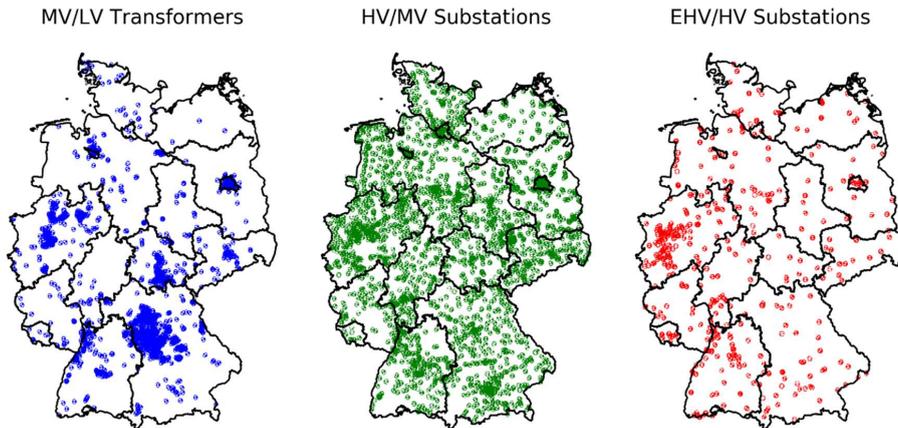


Figure 4-9 Spatial representation of extracted transformer and substation locations for low-, medium-, and high-voltage levels

To evaluate the OSM-based data on transformers and substations, the data are compared to the existing information about low-voltage transformers, medium-voltage substations, and high-voltage substations as per the official records. There are 500,000 MV/LV⁴ transformers, 4,500 HV/MV substations, and 800 EHV/HV transformers installed and currently operational in Germany [201]. Figure 4-10 compares the data extracted from OSM to existing data. It is evident from the comparison that there is a percent error of -98.99% , -42.33% , and -40.87% for MV/LV transformers, HV/MV substations, and EHV/HV substations, respectively. Due to the significant inaccuracy percentage for MV/LV transformers coupled with the small number of detected transformers, these extracted locations are ineffective in estimating geo-referenced synthetic low-voltage network topologies. In chapter 5, the process of synthetically estimating the locations of MV/LV transformers will be discussed. Interestingly, the percentage inaccuracy for HV/MV and EHV/HV substation locations is less than 50%, indicating that the extracted data helps develop geo-referenced synthetic medium- and high-voltage networks. Various approaches were developed to estimate geo-referenced synthetic distribution network topologies while preserving the information extracted from OSM data to include these data in the model.

Apart from transformers and substations, another critical component of the model is power lines. Germany's power lines can be recognized using the previously extracted line dataset. The line dataset includes power lines, streets, highways, other roads, trains, and airways. As a result, it is necessary to filter the dataset to detect power lines and differentiate between different voltage levels. The line dataset is also examined as a part of this process. There are 11,130,492 line-objects in the line dataset, of which 89,717 are associated with power

⁴ LV: Low-voltage; MV: Medium-voltage; HV: High-voltage; EHV: Extra-high-voltage

components. Each of these line objects has tags that correspond to voltage, frequency, wire, and circuits, among other things. Voltage-specific tags are essential in order to distinguish between power lines operating at different voltage levels.

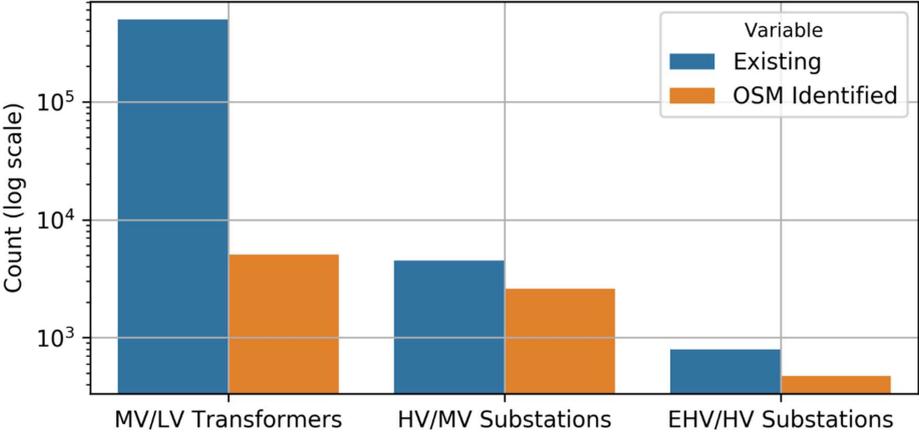


Figure 4-10 OSM identified data comparison with existing one for transformers and substations count⁵ [201]

However, only 48% of retrieved power line objects include voltage-specific tags and the frequency attribute aids in differentiating between power lines used for electrical transmission and distribution and power lines used for railroad transportation. This is because the frequency for power transmission and distribution is 50 Hz, whereas the frequency for railroad transport is 16.6 Hz. A further point to note is that only 30% of the line objects have frequency tags, with 7% corresponding to railway-based power lines. To identify transmission and distribution lines, power lines are filtered with a 50 Hz frequency and voltage specification. Figure 4-11 illustrates the total number of power lines according to the voltage specifications.

⁵ LV: Low-voltage; MV: Medium-voltage; HV: High-voltage; EHV: Extra-high-voltage

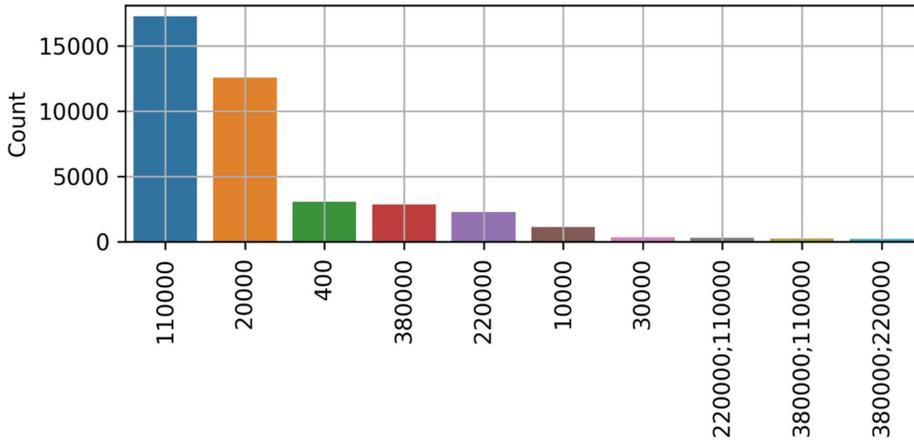


Figure 4-11 Voltage-specific tags for power line objects

By examining the voltage tags mentioned for line objects, we can ascertain that 380000, 220000 indicates EHV⁶ lines, 110000, 220000;110000, 380000;110000 depicts HV lines, 20000, 10000, 30000 represents MV lines, and 400 denotes LV lines (see Table 4-4). Additionally, 41.7%, 14.4%, 34.3%, and 9.5% of the total lines extracted were HV, EHV, MV, and LV power lines, respectively. For obvious reasons, they are referred to as power lines, as underground cables are not visible in satellite imagery and are not apparent in OSM data. Figure 4-12 illustrates power lines extracted from OSM data for HV and EHV, where green lines denote alternating current lines, and additional red lines denote direct current lines connecting borders.

Table 4-4 Power line labels in OSM data

Line label	Assumption
380000, 220000	EHV lines
110000, 220000;110000, 380000;110000	HV lines
20000, 10000, 30000	MV lines
400	LV lines

In order to assess the quality and effectiveness of the extracted power lines, they are compared to existing power lines in Germany. However, statistical comparisons of identified power lines were adapted due to the scarcity of accessible data and the difficulty of conducting spatial comparisons due to their vast number. The statistical parameter compared was the total circuit length for all voltage levels and each specific voltage level. In Germany, the total circuit length for all electrical power lines that include underground cables, irrespective of voltage level, is

⁶ LV: Low-voltage; MV: Medium-voltage; HV: High-voltage; EHV: Extra-high-voltage

1,816,857 km [202]. These power lines and cables are distributed as 1,173,065 km, 511,164 km, 96,658 km, and 35,970 km for LV, MV, HV, and EHV levels⁷, respectively. Table 4-5 illustrates the total circuit length for various voltage levels along with the nominal voltage and voltage regulation for each substation.

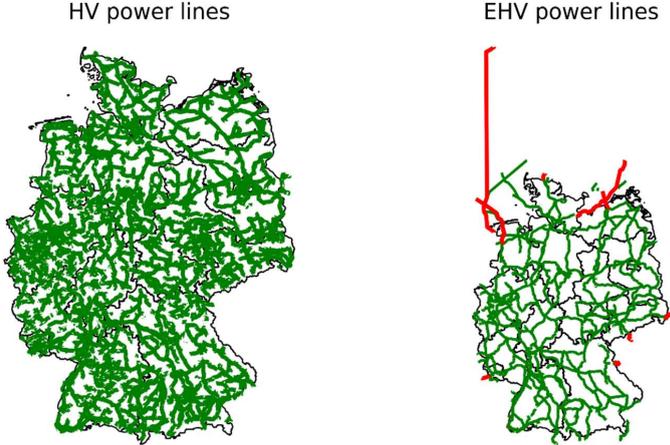


Figure 4-12 OSM extracted HV and EHV power lines

Table 4-5 Power grid characteristics in Germany that include, nominal voltage, total length, and voltage control according to voltage levels [Adapted from [202]]

Voltage level	Nominal voltage	Total circuit length	Voltage regulation
Low voltage	0.4 kV	1,173,065 km	Fixed voltage ratio
Medium voltage	10, 20, and 30 kV	511,164 km	Fixed/Automated control
High voltage	110 kV	96,658 km	Automated control
Extra-high voltage	220 kV, 380 kV	35,970 km	Automated control

In the illustration, it is evident that the overall circuit lengths for low- and medium-voltage levels are significantly longer than those for other voltage levels. However, despite their extraordinary lengths, the OSM identified power lines at the LV and MV levels are modest and insignificant. The outcome is unsurprising because the majority of electrical power supply lines are underground cables. According to data [1], 85.9% and 75.2% of total circuit length for LV and MV, respectively, are underground cables. Due to the increased percentage of underground cables, it is evident that the power lines transporting electricity remain unidentified when using OSM data. On the other hand, HV and EHV power lines are overhead wires that are generally identifiable in the OSM data. Therefore, due to their identifiable nature, the extracted HV and EHV power lines are compared to Table 4-5 and depicted in Figure 4-13.

⁷ LV: Low-voltage; MV: Medium-voltage; HV: High-voltage; EHV: Extra-high-voltage

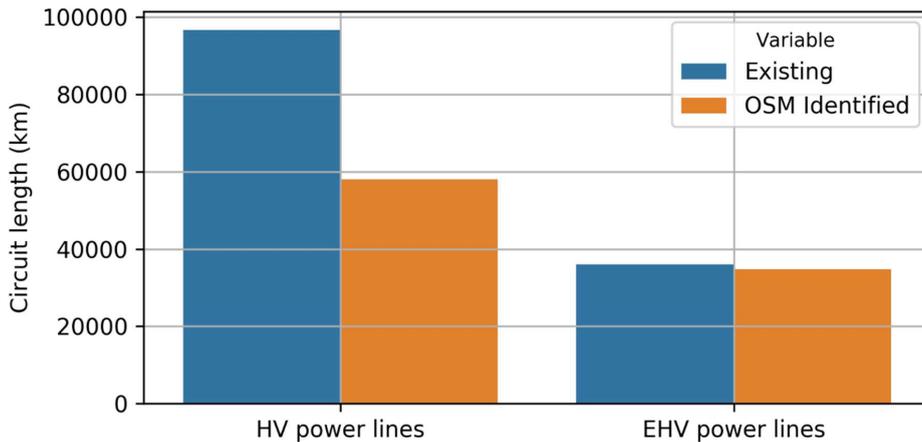


Figure 4-13 OSM identified power line length comparison with existing data for HV and EHV levels

According to Figure 4-13, the percentage inaccuracy for HV and EHV power lines⁸ is -39.9% and -3.2%, respectively. The negative percentage inaccuracy implies that the identified power lines are fewer in number than those already in place. According to the findings, all of the EHV power lines are visible in OpenStreetMap and are extracted without data loss. As a result, creating geo-referenced synthetic network topologies for EHV networks is a straightforward process. Consequently, SciGRID [178], osmTGmod [179], GridKit [180], and the Common Information Model (CIM) [181] used OSM data to develop the synthetic EHV network or the transmission network for Germany. Additionally, the percentage error is higher for extracted HV power lines than for EHV power lines. This necessitates the development of concrete approaches for generating HV networks from OSM data using extracted power lines. Compared to the HV and EHV power lines extracted from OSM data, the LV and MV power lines discovered are insignificant compared to the already installed LV and MV power lines. Consequently, advanced approaches were established to extract power lines from road networks and develop geo-referenced synthetic LV and MV networks, which will be explored in greater detail in Chapter 5.

Following the extraction of the most significant power components, such as transformers, substations, and power lines, the data corresponding to network nodes are the final data necessary for developing geo-referenced synthetic network topologies. The nodes could be either load or generation nodes, depending on their configuration. The former is generally located in enclosed structures, except in use cases such as street lighting and some other commercial locations such as mines. Therefore, buildings are essential for determining network nodes, particularly load nodes and data extracted from the OSM aids in identifying the locations and attributes of buildings. As already mentioned, closed structures in the OSM data

⁸ LV: Low-voltage; MV: Medium-voltage; HV: High-voltage; EHV: Extra-high-voltage

are represented by polygons with specific tags representing a type of polygon. The polygons with the tag *<building>* were filtered in order to extract their locations and their attributes.

The filtered polygon dataset contains 29,619,992 buildings and 71 associated attributes. All of these features, however, are redundant and contain multiple missing values. As a result, only a few critical features among the available options are examined. After a few redundant features are eliminated, the total comes down to twelve. Additionally, each building is labeled; the essential labels and the labels associated with the majority of buildings are illustrated in Figure 4-14.

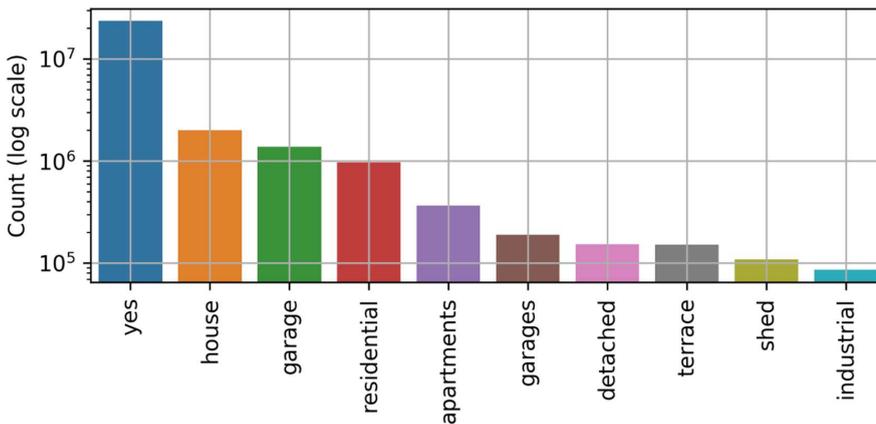


Figure 4-14 Building labels in filtered OSM data

Figure 4-14 illustrates the significant number of buildings, which are denoted by the label 'yes' totaling to 23,570,792. That is, 79.5% of the structures with the 'yes' tag are of unknown type. Some of the other essential labels include house, garage, residential, apartment, detached, and industrial. Along with these labels, there are 'no' and 'power' labels as well. However, buildings labeled 'power' are already used to designate power-related components such as transformers and substations. As such, they are omitted from the list along with buildings labeled 'no,' so the rest can serve as load nodes. The resulting dataset contains approx. 29 million buildings with twelve distinct characteristics. However, it is vital to define the type of a building before integrating it into electrical networks to ensure that appropriate load profiles are assigned for the power flow analysis. However, only 20.5% of buildings have specific tags. Therefore, it is vital to predict the building types prior to designing geo-referenced synthetic network topologies. Therefore, the following section discusses building type classification using state-of-the-art machine learning methods.

4.3 Building Type Classification

Buildings retrieved from OSM data merely provide their spatial positions along with certain limited information about their type. Several studies have been conducted regarding the classification of buildings using a variety of different methodologies. Classification of building

types can be accomplished manually or automatically. However, manual classification of buildings is a highly time-consuming process. As a result, automated classification is the most advanced solution available. According to the literature reviewed in this study, automated classification leveraged until now relies on several building attributes. Steiniger et al. [203] classified urban structures based on their perceptual qualities, which include built-up area, density, size, shape, and orientation. Similarly, Hecht et al. [204] pioneered the use of extensive data sources, such as 3D models to automate the classification of buildings. Additionally, Henn et al. [205] employed 3D building information to train Support Vector Machines (SVMs), a machine learning technique used in classification applications. In addition to three-dimensional information, Wurm et al. [206] considered one-dimensional variables such as length and two-dimensional variables such as the area. This method, however, is constrained by its emphasis on unidentified features. In the field of artificial intelligence, the use of data from Light Detection and Ranging (LIDAR) and image recognition is also highly popular. In this context, Lu et al. [207] and Zheng et al. [208] classified building types using spatial and landscape features extracted from LIDAR data. The characteristics of these image data function as the primary basis for classifying building types. However, the utilization of this data necessitates the use of a preset tag for each image during the training phase of any machine learning model. It is also extremely challenging to obtain images/data with predetermined tags. This is why many researchers are utilizing the OSM dataset, which contains essential information such as geometry, type, area, length, and address, all of which are essential for training a model. Forget et al. [209] conducted several tests using the OSM dataset to classify built-up areas using supervised learning. Fan et al. [210], on the other hand, analyzed rural structures using deep learning based on Convolutional Neural Networks (CNNs). A machine learning method was used by Bast et al. [211] to tag building types to improve the quality of building-level information in OSM.

There is a gap in the literature in classifying the building types extracted from the OSM data with various missing values and inconsistencies in the predefined labels in the OSM data, which will be discussed in this section. Therefore, a methodology was developed to predict building types from buildings-based data extracted from OSM. This was accomplished by augmenting the model with additional features from external sources, which correspond to the buildings and accounting for missing values for the features to improve the prediction results. The approach used to generate building tags is depicted schematically in Figure 4-15.

The methodology is initiated by collecting and combining features of the input data, which includes Coordination of Information on the Environment (CORINE) [212] dataset, building heights in Berlin [213] dataset, and the 2011 census data for Germany [214], in addition to the core OSM buildings dataset. The merged dataset is used to train a machine learning model that addresses the dataset's inherent problems. Finally, data post-processing as well as active validation of predicted data against publicly available statistics were performed.

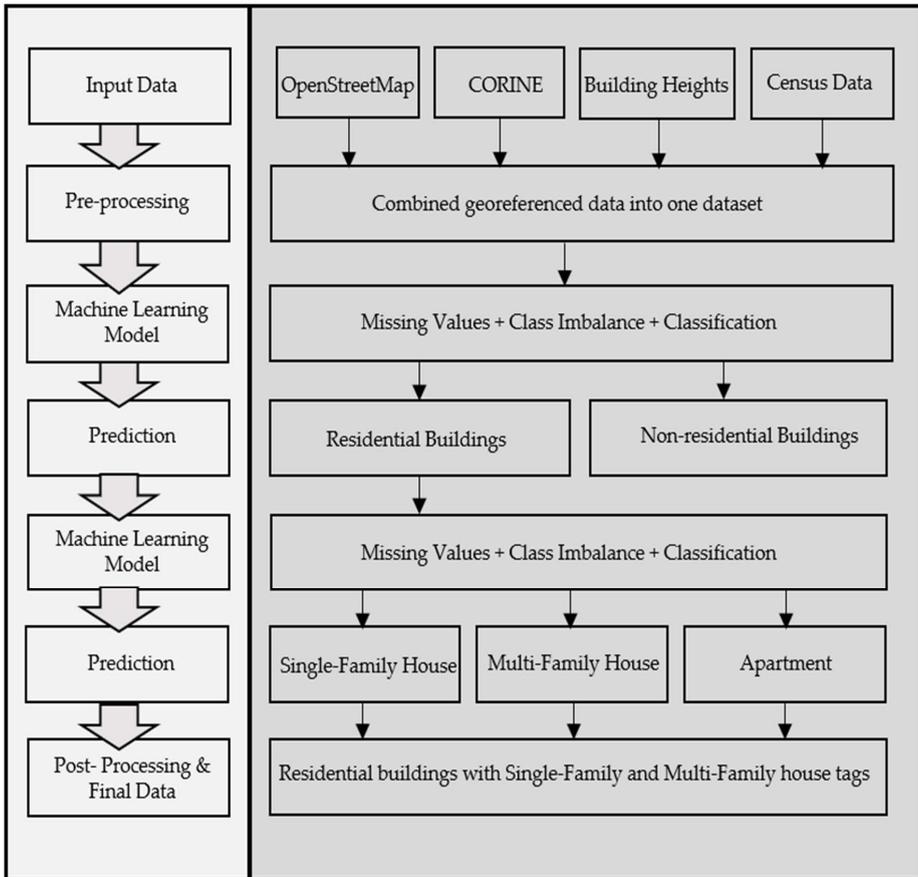


Figure 4-15 Methodology for predicting building types on building footprints of Germany

When predicting building types, the OSM data was appended with additional information retrieved from the datasets mentioned above. The following section will discuss the process of appending and preparing the data for training the machine learning model.

4.3.1 Data Pre-processing

As previously stated, the extracted buildings add up to approx. 29 million buildings with 12 distinct features such as way, way_area, building, osm_id, amenity, shop, leisure, office, construction, highway, foot, and place. However, these twelve features were reduced to five that include foot, height, area, road network, and building because the data for the rest of the features contain more than 90% missing values, which would negatively impact the model. However, these six features are insufficient to predict building tags accurately and as such, they are supplemented with data from other external data sources.

First, buildings' heights are a highly feature in direct relation to the buildings. However, retrieving the heights of individual buildings is impossible, as no such dataset exists so far. However, the Copernicus project's urban atlas [213] defines building heights for several large

cities, including Berlin's, which was retrieved from the urban atlas database [213]. This dataset includes a raster layer with a resolution of ten meters that provides information about building heights. In order to integrate the dataset to OSM data, geographical coordinate reference systems are extremely important. Berlin's buildings height dataset supports the ETR89 Coordinate Reference System (CRS) (i.e., EPSG:3035). However, the coordinate reference system should be uniform for each dataset while appending the data. Figure 4-16 depicts the raster layers containing information about the height of buildings in Berlin, Germany [213]. As per the box plot in Figure 4-16, building heights range from 3 m to 30 m with a few outliers. However, the values of 50% of building heights are between 5 and 15 meters.

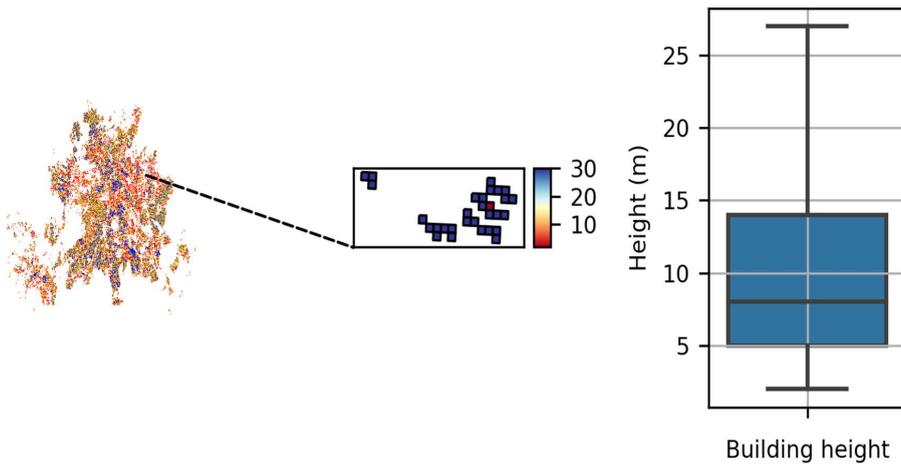


Figure 4-16 10 m raster layer containing building height information on the left and box plot showing range of building heights on the right

Apart from conventional building features such as area and height, predicting building tags becomes easier with the land-use property. The CORINE land cover databases, which were created as part of the Copernicus project [212], contain information on land use. However, the dataset is based on the classification of satellite images and holds a feature with 44 labels. The labels in the dataset correspond to the continuous urban fabric, discontinuous urban fabric, industrial or commercial units, and airports labeled as 111, 112, 121, and 124, respectively. Appendix B has extensive definitions of each label and its representation. The dataset was retrieved from the Copernicus land monitoring service due to the high-resolution land use offering and contains 167,119 geometries with a variety of labels, which are explained elaborately in Appendix B. However, the majority of polygon geometry were classified as pastures or non-irrigated arable land (20% and 14%, respectively). Additionally, 0.6%, 12.7%, and 4% of data reflect continuous urban fabric, discontinuous urban fabric, and industrial or commercial units, respectively. As stated above, the reference system of the geometry is vital for appending the data to OSM buildings. Therefore, this dataset was projected from ETR80 (EPSG:3035) to the World Geodetic System 1984 (WGS 84). (EPSG:3857). This projection

was necessary because the primary dataset, which is the OSM buildings, is in the WGS 84 (EPSG:3857) coordinate system. Figure 4-17 illustrates the geometric representation of the CORINE land cover data for Germany and demonstrates that most of the continuous urban fabric and discontinuous urban fabric areas are concentrated in major cities such as Berlin, Bremen, Hamburg, and others. Furthermore, we can see that North Rhine-Westphalia has a high concentration of urban regions as well.

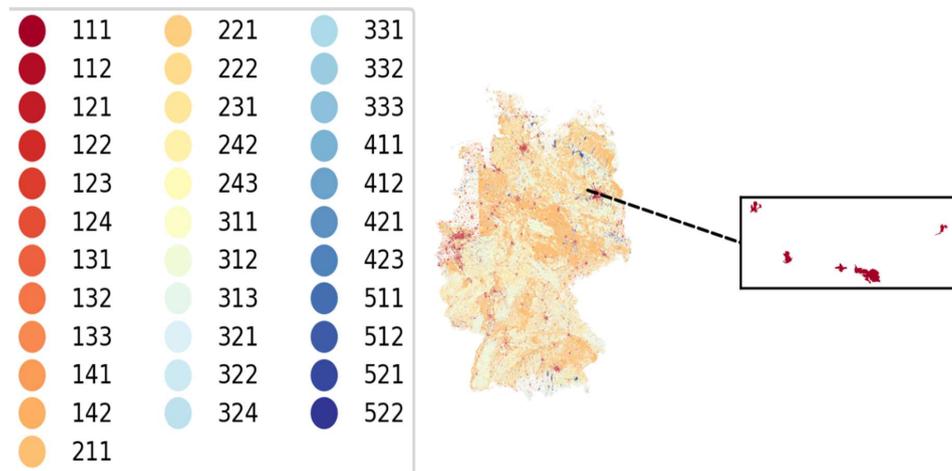


Figure 4-17 CORINE land cover data for Germany

Until now, only parameters pertaining to building types and their geographic locations have been examined. Further, the census 2011 dataset was used to incorporate certain statistical data regarding the total number of buildings in Germany and the total number of buildings by municipality. In 2011, the survey collected data on the overall number of buildings in Germany, including the number of apartments, single-family houses, multi-family houses, and residential buildings. This dataset contains the total number of buildings per 100 m x 100 m grid cell. The data is further sub-divided to provide information on single-family houses, two-family houses, multi-family houses, and apartment buildings. Grid cell locations for various types of buildings are depicted in Figure 4-18. Moreover, each grid cell provides the buildings count for those with living spaces. That is, quasi commercial, industrial, and public spaces are disregarded.

Table 4-6 summarizes the various types of buildings, their abbreviations, and the total number of grid cells for each type. However, each grid cell has a varied number of buildings. Based on the data in each grid cell, the total number of buildings with living space is approx. 19 million. The buildings with living spaces are classified as single-, two-, and multi-family houses (detached and semi-detached), as well as apartment buildings. In this study, a detached house is described as a free-standing building regardless of its type, a semi-detached house is identified as a building constructed against another building, a terraced house is defined as a

structure built against two other buildings, and other building types are categorized as buildings that are not detached houses, semi-detached houses, or terraced houses. However, statistical data related to the aforementioned building types should be integrated into each building retrieved from OSM data as a feature.

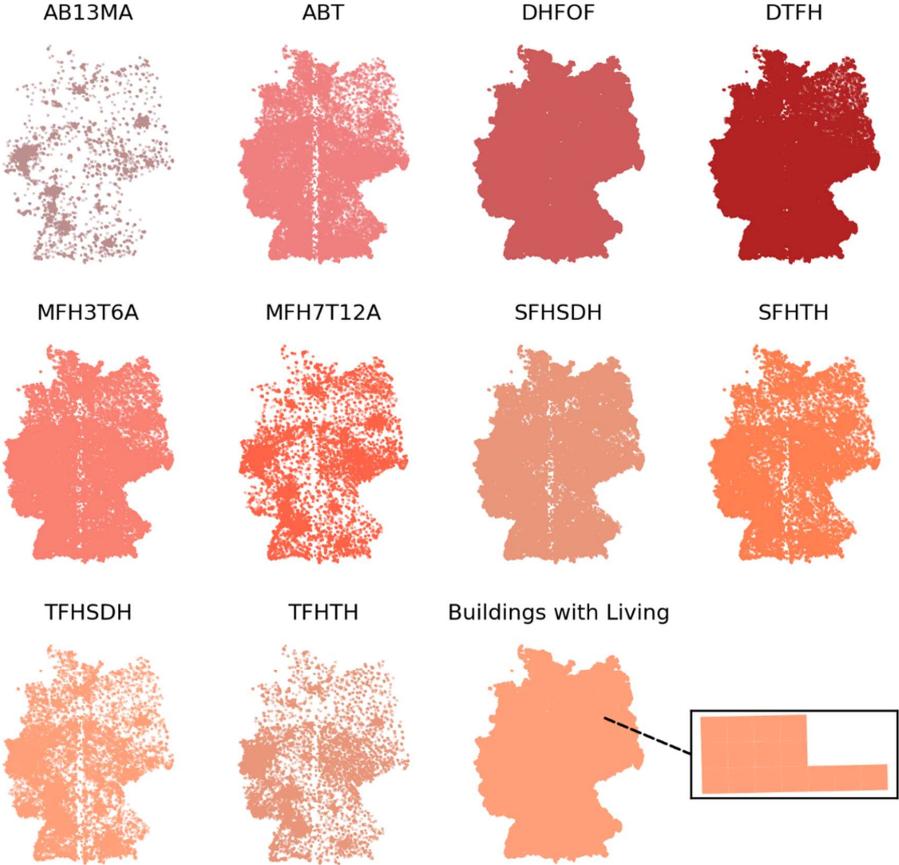


Figure 4-18 Number of buildings with living spaces for different building types per 100m X 100m grid cell

Table 4-6 Number of grid cells per building type in Germany

Building type	Abbreviation	Number of grid cells
DHFOF	Detached house for one family	1,637,974
SFHSDH	Single-family house: semi-detached	411,851
SFHTH	Single-family house: terraced	292,062
DTFH	Detached two-family houses	552,705

TFHSDH	Two-family house: semi-detached	65,266
TFHTH	Two-family house: terraced	50,873
MFH3T6A	Multi-family house: 3–6 dwellings	437 990
MFH7T12A	Multi-family house: 7–12 dwellings	153 802
AB13MA	Apartment building: 13 or more units	36,214
ABT	Another building type	104,205

Here, the procedure for extracting features from statistical data is described in detail. Before deriving this, features from building heights and CORINE land cover datasets are appended to the OSM buildings. In order to combine the data from these datasets, it is necessary for their coordinate reference systems to match. Therefore, EPSG:3857 was selected since it corresponds to OSM buildings. If the dataset is formatted in line with a different reference system, it must first be converted to the predefined system before proceeding.

The CORINE land cover data is appended to the building dataset as a feature. This is accomplished by matching the CORINE data geometries (polygons) with the buildings that overlap these polygons. Thus, the intersection of building polygons with CORINE data geometries enables the use of CORINE data as a feature for intersected buildings. Similarly, the heights of berlin buildings were added to OSM buildings by intersecting them with the raster data from the building heights dataset. Finally, census data with the 10 characteristics listed in Table 4-6 and buildings with living spaces are assigned to the OSM data. Each OSM building will be assigned a value, and an ID based on the grid cell with which it intersects. These statistics indicate the number of buildings per building type in a grid cell. These features are used to provide eleven more features with a percentage probability. These additional features were developed by calculating the fraction of total buildings for different building types in a grid cell, which are derived from the census data, by the total number of OSM buildings in that grid cell. Equation 4-1 contains the mathematical expression used to calculate this.

$$Z_{i,j} = \frac{TB_{i,j}}{TB_{OSM}} \times 100 \quad 4-1$$

$$\forall i = \{DHFOF, SFHSDH, SFHTH, DTFH, TFHSDH, TFHTH, MFH3T6A, AB13MA, ABT, \text{buildings with living}\}$$

$$\forall j = \text{grid cells}$$

The left-hand side of Figure 4-19 depicts the total number of buildings with living space contained within a grid cell (i.e., three). However, when the grid cell is mapped using OSM buildings, a total of six buildings are found. Thus, based on Equation 4-1, 50% of the buildings in that grid are likely to be buildings with living space, which are also referred to as residential buildings, and all buildings in that grid receive 50% as an additional feature. Similarly, on the right side of Figure 4-19, an extra building type count is considered (i.e., DHFOF). In this case, the total number of DHFOF in the same grid is two. As a result, two of the six buildings are DHFOF. Thus, each building in that grid has a 33.3% probability of becoming a DHFOF building. When this technique is applied to other building types, each building receives a percentage probability for the building types, resulting in eleven additional features.

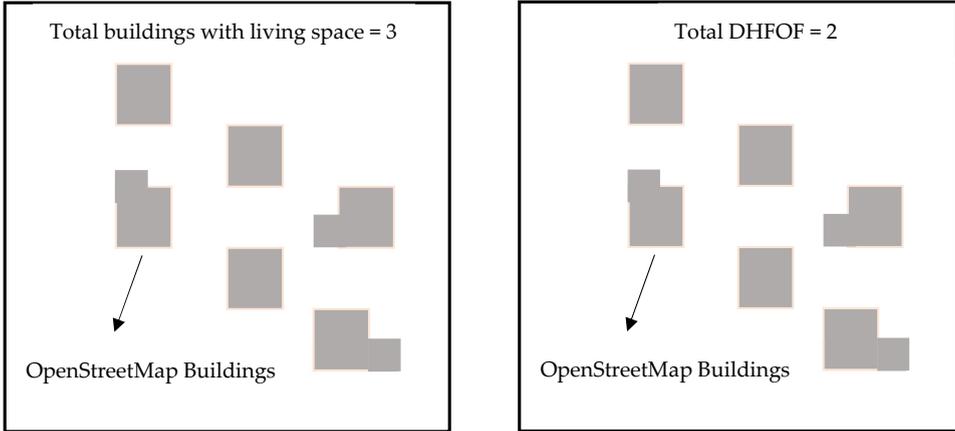


Figure 4-19 OSM buildings mapped with census 100 X 100 m² grid cells

Based on these data, the building with a greater than 100% likelihood of having living space is classified as a residential building and labeled as *<residential>* in the building type feature. In fact, this building type feature is considered the target class and stores essential building tags. Before any of the buildings are labelled in line with the aforementioned assumption, it is necessary to address the significant uncertainties in the labels indicated in Figure 4-14. Buildings retrieved from OSM data have over 1575 distinct labels. The reason for this ambiguity includes unclear representation of buildings, such as spelling errors and the use of different languages. Nevertheless, some of these ambiguities are summarized in Table 4-7.

Table 4-7 Uncertainties in OSM building labels

Building label	Differing representation
garage	garages
farm	farm_auxiliary
Terrace	Terrasse
House	Haus Hause house
Youth Centre	Jugendzentrum
Nursing home	Pflegeheim
deconstructed	deconstructed

Through manual refinement, the number of unique labels was reduced to 895. However, there are still a huge number of unique values that are not required for the training model. Therefore, the target class labels were reduced to 25 using a Wiki model, where different building types are categorized [215] and renamed *<building_class>*, which now serves as the target class for classification and the 25 designations of the target class also include residential, garages, industrial, commercial, etc. Indeed, the label *<yes>* is a member of the target class, which must be predicted. There are also several buildings that are unfit for living and are labeled as *<yes>*.

For instance, garages are depicted on the map as distinct polygons, with some of them labeled <yes>. Hence, pre-labeling these buildings aids in the model's effective training. It is important to note that the individual's need determines the size of the garages, although some adhere to standard garage dimensions. Figure 4-20 illustrates the area of OSM buildings that are already designated as garages/attachments. It is clear from this figure that the areas of 75% of buildings are less than 35 m² that are represented as garages/attachments in OSM data. Therefore, in addition to the typical garage sizes of 20 m² to 40 m², buildings with an area of less than or equal to 35 m² are considered garages. However, not all buildings with a floor size of less than 35 m² are garages. However, they are most likely not residential buildings.

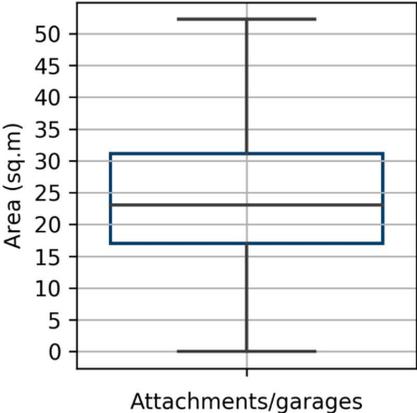


Figure 4-20 Box plot representing area of buildings labeled as garages

Following all these stages, the dataset contains 29,497,772 buildings with 30 features and 10,359,617 buildings with 25 unique labels. However, before preprocessing, the overall number of buildings with labels was 6,047,266, which is 20.4% of total OSM buildings. This number rises by 15% after preprocessing. The enhancement in labeled data enables large-sample training of the model. Figure 4-21 illustrates the percentage improvement in the labeled data before and after preprocessing processes. Finally, the labeled data are used to train a machine learning model using the specified features to predict the unlabeled dataset.

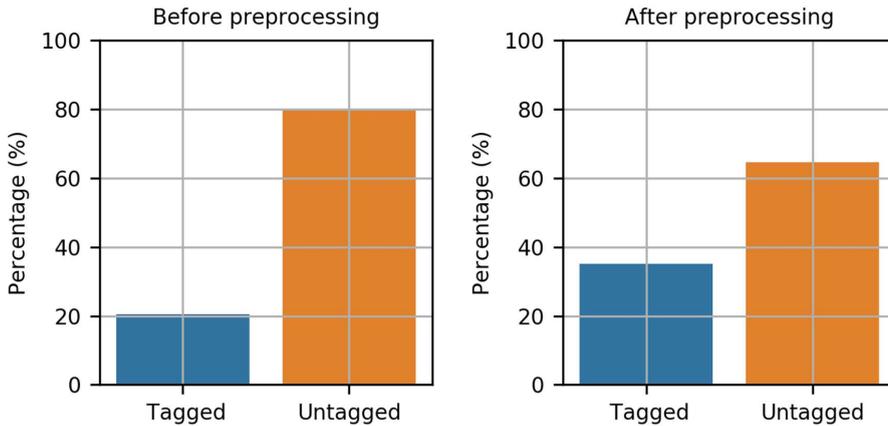


Figure 4-21 Percentage labeled and unlabeled data before and after preprocessing

However, before creating a classification model, a secondary examination of the dataset was conducted, revealing that the features in the dataset contain many missing values. Upon examination of all the features, it became apparent that there are at least 77% missing values. As such, the created model may have several inaccuracies. As a result, the developing model must account for these missing values. Thus, the missing values in all features are further examined along with the label distribution in the target class. The distribution of labels in the target class is depicted in Figure 4-22 and most of them are attachments accounting for 52.8% of total labeled data. Residential and commercial buildings are associated with 35.9% and 9.02% of labelled data, respectively. The remaining classifications, such as agricultural, school, and hospital, are fairly insignificant. As a result, if a model is trained on these data, the predicted value will be biased toward the widely shared labels. Thus, the developing model should consider not only missing variables but also class imbalance issues.

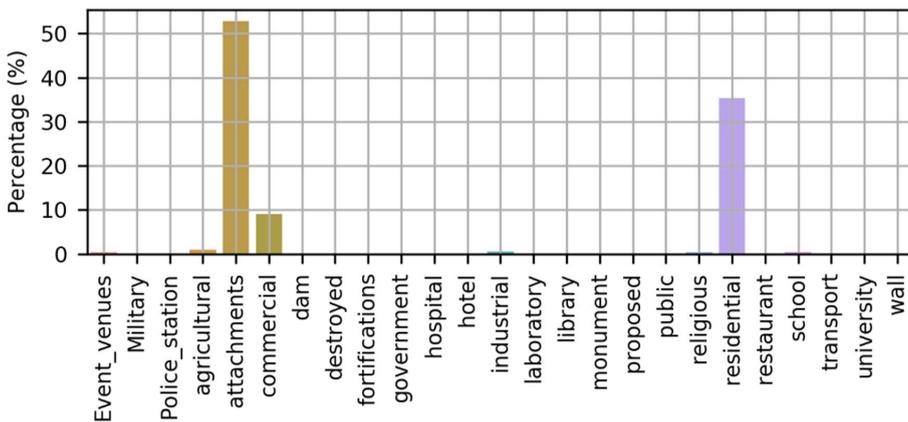


Figure 4-22 Distribution of labels in the target class

The next section illustrates a classification model for predicting building types in response to the preprocessing of OSM data and identifying issues associated with class imbalance and missing values in the dataset.

4.3.2 Classification Model

After defining the datasets and identifying the underlying challenges, different machine learning classification models were employed to address the issues of missing values and class imbalance. In this context, experiments were conducted on the datasets to determine the model that fits the data optimally. However, a detailed discussion of the classification models that were employed follows. The classification of building types was performed undertaking a two-fold approach. The first stage is to classify buildings as residential or non-residential. The second stage is to subclassify all predicted residential buildings into single- and multi-family houses. The approach employed is represented in Figure 4-23.

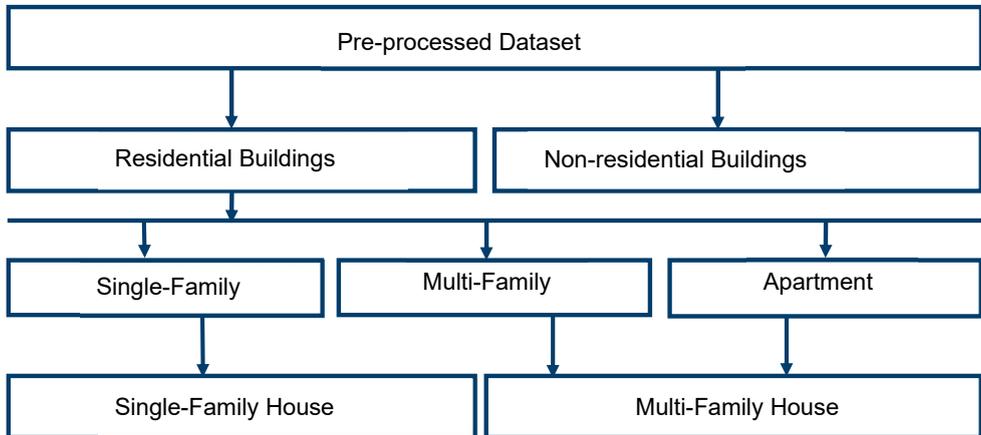


Figure 4-23 Methodology employed for building type classification

However, the machine learning algorithm used to categorize buildings should consider missing values and class imbalance issues while training on OSM buildings with tags. The majority of the effort is directed towards processing the data; roughly 20% of the task entails training the model and making predictions on the unlabeled dataset. Herein, two distinct types of models were investigated: those that account for missing values and class imbalance. The first is the implicit method, which combines the solutions for missing values, class imbalance, and classification into a single architecture. The second method is the explicit method, in which each task is completed independently.

The implicit approach employed HexaGAN [216] and a modified Artificial Neural Network (ANN) [217]. Since this thesis does not pertain significantly to algorithms. An explicit method was developed in tandem with the implicit method to address missing values and class imbalance problems. This method utilized Multiple Imputation by Chained Equations (MICE)

[218] for imputing missing values, Synthetic Minority Oversample Techniques (SMOTE) [219], and cost-sensitive learning for imbalance classification (Class-Weighting) (CS) that helps address class imbalances in the target class labels. Finally, a random forest classifier was used to classify the different types of buildings.

The implicit techniques were evaluated on both baseline and preprocessed datasets containing labeled data since they are addressed and offered the precision score with baseline dataset. On the other hand, the explicit method was trained exclusively on preprocessed OSM dataset because this method contains separate algorithms in each step. Nevertheless, model comparison was carried out using the F1 score for all three models, wherein the scores range from 0 to 1. The experiments were repeated ten times with five-fold cross-validation using labeled data that was split 80:20 between training and testing data. The decision to limit the experiment to ten times is ascribed to the fact that increasing the number of experiments resulted in increased time due to the large volume of data involved in training the model. Nevertheless, classification performance demonstrates that when implicit approaches are deployed and trained on baseline data, they achieve results that are nearly identical to model specifications. However, when trained on the OSM dataset, a score of 0.81 was produced, which is 17% lower than the baseline dataset values. Furthermore, the F1 score obtained when the explicit method was used in conjunction with the OSM dataset was 0.995. The classification performance of all the implemented models is summarized in Table 4-8.

Table 4-8 F1 scores for implementation models

Model	Model Results	Implementation + Results	OSM Data Results
HexaGAN	0.97±0.02	0.97	0.81
ANN	0.81	0.79	0.63
MICE+CS+RF	-	-	0.99

The F1 scores clearly demonstrate that the explicit method produced acceptable results. Therefore, model MICE combined with class weighting and a random forest classifier fit the OSM data better. So, this model was chosen to predict the types of buildings in the unlabeled OSM dataset. When the trained model was applied to predict the building types for the unlabeled buildings, it resulted in a label for each unlabeled building. The predicted building types yielded the results listed in Table 4-9. Residential, attachments, commercial, and industrial were the primary labels identified. However, the share of residential labels was significantly higher than the share of all other labels. Here, any buildings other than residential structures are classified as non-residential buildings.

Table 4-9 Total buildings per building type predicted

Building Type	Predicted Count	Percentage
Residential	19,747,802	66.947
Attachments	5,583,658	18.929
Commercial	3,127,442	10.602

Industrial	460,698	1.562
Hospital	217,103	0.736
Hotel	114,013	0.387
Agricultural	105,888	0.359
Government	15,198	0.052
Event venues	35,935	0.122
School	46,452	0.157
Religious	33,645	0.114
Transport	3,117	0.011
University	2,927	0.010
Military	1,048	0.004
Others	2,866	0.010

The residential buildings categorized in the first step of the two-fold approach provided 19,747,802 as the result, which are further classified into different house types. The house type should be classified into single-family house, multi-family house, and apartment. To accurately predict the house type, the training and testing data sets should contain appropriate house type tags. However, merely 2% of the entire labeled dataset comprises labels of the appropriate type (i.e., single-family house, multi-family house, and apartment). The majority of the labeled house types are categorized as detached, semi-detached, and terrace. It is important to note that detached houses are buildings that do not share a wall with another building, semi-detached houses share at least one wall with another building, and a terraced house consists of several connected houses. But these labels are different from those expected. Therefore, the above labels are adapted to align with the expected labels as specified in Table 4-10.

Table 4-10 Adaptation of OSM building labels to proposed labels

OSM Building Type	Proposed House Type
Detached	Single-Family House
Semi-Detached	Multi-Family House
Terrace	Multi-Family House
Apartment	Apartment
House	To be predicted
Residential	To be predicted

Thus, detached houses are classified as single-family houses, whereas semi-detached and terrace houses are classified as multi-family houses. Furthermore, the terms house and residential continue to be used interchangeably. Thus, it was determined that buildings that were labelled as houses and residential should be predicted for single-, multi-, and apartment buildings. These assumptions, however, were made in order to improve the dataset's quality. Additionally, to enhance the quantity of labeled data, the approach utilized in the preprocessing stage to pre-label residential buildings using percentage probability characteristics was employed here. If the probability of a building being of a particular kind is greater than or equal

to 100%, the appropriate house type label is attached to the building. Following these stages, single-family houses, multi-family houses, and apartments make up 60.0%, 21.8%, and 18.0%, respectively. However, a closer examination of the percentage distribution reveals that the dataset contains a class imbalance, with a disproportionate share of single-family houses.

As stated above, there are 1,769,997 samples in the full training and test dataset. Using these samples, we constructed a model that accounts for missing values (i.e., MICE), an oversampled SMOTE with weighted data to account for class imbalance, and a random forest for classification. The fitted model was then used to predict house types of 17,977,805 residential buildings, which included both those labeled as residential and houses. Finally, applying the model to predict the residential building types led to three categories: single-family homes, multi-family homes, and apartment buildings. Table 4-11 summarizes the number and percentage share of the various housing types predicted. Nevertheless, it is significant to verify the outcomes of the model prediction. Geographical validation of the predicted buildings is out of scope for this study due to the data volume. However, statistical validation on the predicted building types and house types is performed in the following subsection.

Table 4-11 Predicted house types count and percentage

House Type	Count	Percentage
Single-family House	14,378,635	72.81
Multi-Family House	4,350,183	22.03
Apartment	1,018,984	5.16

4.3.3 Statistical Validation

The labeled OSM dataset following prediction comprises of 19,747,802 residential buildings, including 14,378,635 single-family houses, 4,350,183 multi-family houses, and 1,018,984 apartment buildings. However, validating predicted results at the building level is rather difficult and detailed statistics must be used to evaluate the results. Therefore, the results were statistically validated using publicly available data. Germany has a total residential building count of 19,053,216 [220]. However, the machine learning model outlined previously, which considers not just classification but also missing values and class imbalance was employed to anticipate the total number of residential buildings as 19,747,802. That is, the total number of predicted residential buildings deviates by 3.4%. The comparison of the residential building count in Germany with official figures and predicted results is shown in Figure 4-24. The variation in the result is seemingly minor.

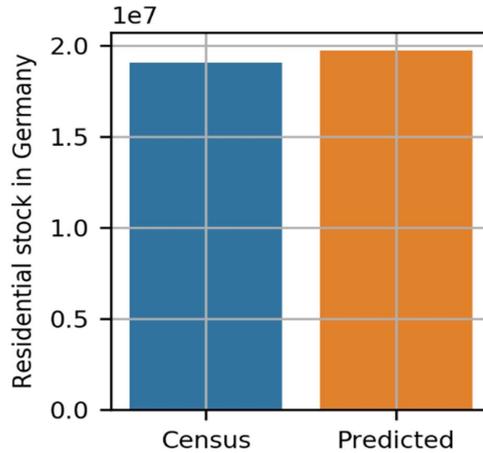


Figure 4-24 Comparison of actual residential buildings count with predicted count

Furthermore, it is hardly accurate that barring 3.4% of non-residential buildings predicted to be residential, rest are actually residential. As a result, additional validation spreading across the whole of Germany is required, which extends beyond the validation of the count. Therefore, additional validation was conducted for each German federal state to accomplish this task. Figure 4-25 illustrates the predicted residential building count by federal state and the associated data from the state's official data.

The comparison demonstrates that the percentage inaccuracy spans between -18.6% and 22.7% . That is, -18.6% indicates that approximately 19.0% of residential buildings are anticipated to be non-residential, while 22.7% indicates that some of the non-residential buildings are predicted to be residential. The data from Figure 4-25 clearly demonstrates that the two states of Baden-Württemberg and North Rhine-Westphalia is where it is over predicted. However, when all federal states are considered, the root mean square error is 12.7% . The relative error could be explained by the fact that certain states have a higher density of buildings than others. In addition, buildings used for training in some of the states are fewer, which could account for the percentage error. To further reduce error and predict appropriate labels, the training dataset could be enhanced. In other words, the training data should be free of missing values, and the distribution of the labels in the training data should be nearly equal.

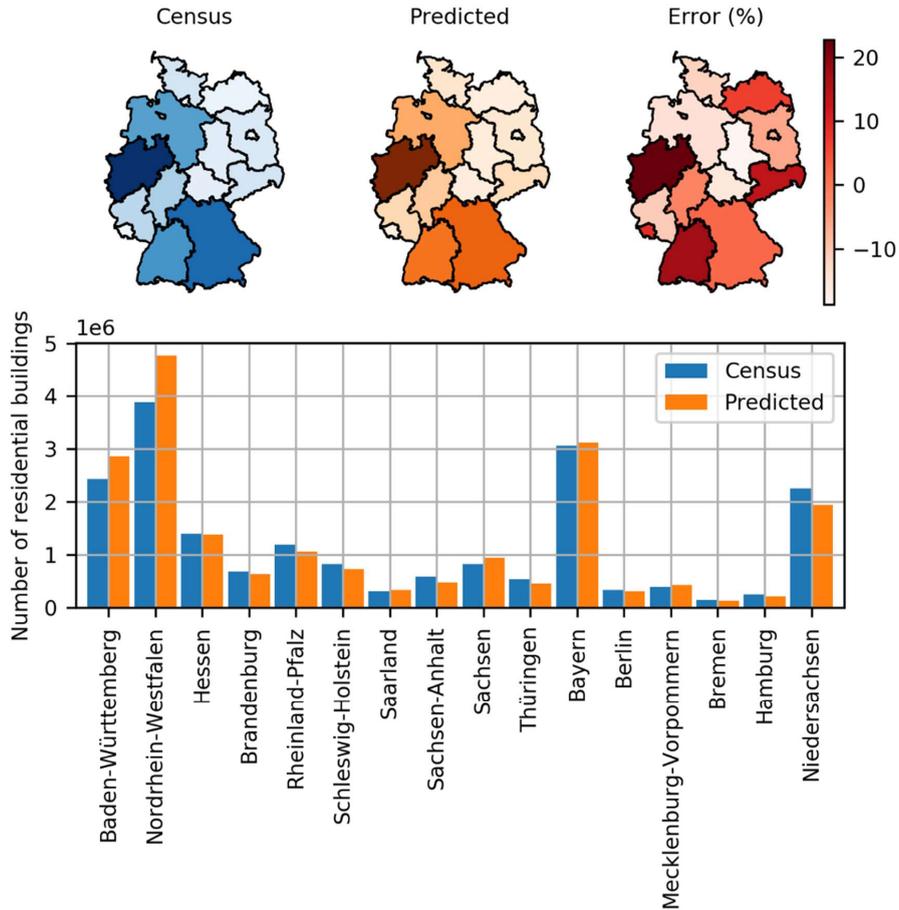


Figure 4-25 Residential buildings count comparison for each federal state of Germany

As previously indicated, residential buildings are further classified into single-family houses, multi-family houses, and apartment buildings, which are proven through comparison to official statistics. Germany has a total of 12,707,978 single-family houses [220]. However, the total number of single-family homes predicted is 14,378,638. When these predicted buildings were compared to Germany's single-family housing count, a relative error of 13.1% is observed. That is, around 13.5% of buildings are classified as single-family houses, despite their actual type being different. Additionally, data pertaining to the predicted multi-family houses and apartment buildings are combined and compared to multi-family houses in Germany. In Germany, there are a total of 6,345,238 of this type [220]. However, 5,369,167 multi-family houses and apartment buildings are predicted. The comparison of predicted results to the official figures presents an inaccuracy of -15.3% . This is because official data include the total number of multi-family houses, which comprise two-family houses, multi-family houses, and residential establishments or apartments. Figure 4-26 compares the values predicted by

machine learning models for single- and multi-family houses in Germany compared with official statistics.

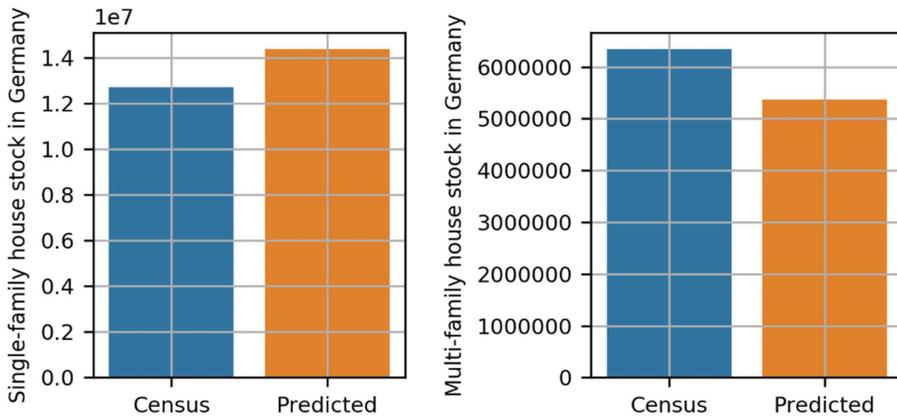


Figure 4-26 Predicted single- and multi-family houses comparison with official statistics

Since there was a substantial deviation of single-family house prediction over official statistical data, a detailed geographical validation on German federal states was conducted. Figure 4-27 depicts the predicted single-family houses in each federal state in comparison to official recorded data. As seen in Figure 4-27, the percentage inaccuracy for predicted single-family houses ranged from -16 to 50% . That is, a negative percentage suggests that the number of predicted buildings are less than actual numbers. On the other hand, the positive value shows that the number of buildings anticipated as single-family houses are higher than real single-family houses. However, there is a significant divergence between predicted and actual values for three states: Baden-Württemberg, North Rhine-Westphalia, and Sachsen, which have deviations of 37.6% , 38.3% , and 50.8% , respectively. The substantial difference between the single-family house prediction and the residential building prediction can be attributed to the several additional assumptions undertaken by the model for the single-family house prediction. That is, neither the training nor the testing datasets contain any labels that must be predicted. To close this gap, detached, semi-detached, and terrace residences are classified as single- and multi-family homes. However, not every detached house is a single-family house, and not every semi-detached house is a multi-family house. They can occasionally be two distinct houses in reality. Additionally, this task is entirely predicated on the assumption and specified labeling of the target class using census data. In the future, these issues can be overcome by improving the actual needed labeling in OSM data.

Summing up, this is the first time that OSM data-extracted buildings have been tagged according to building and house types by using state-of-the-art machine learning methods.

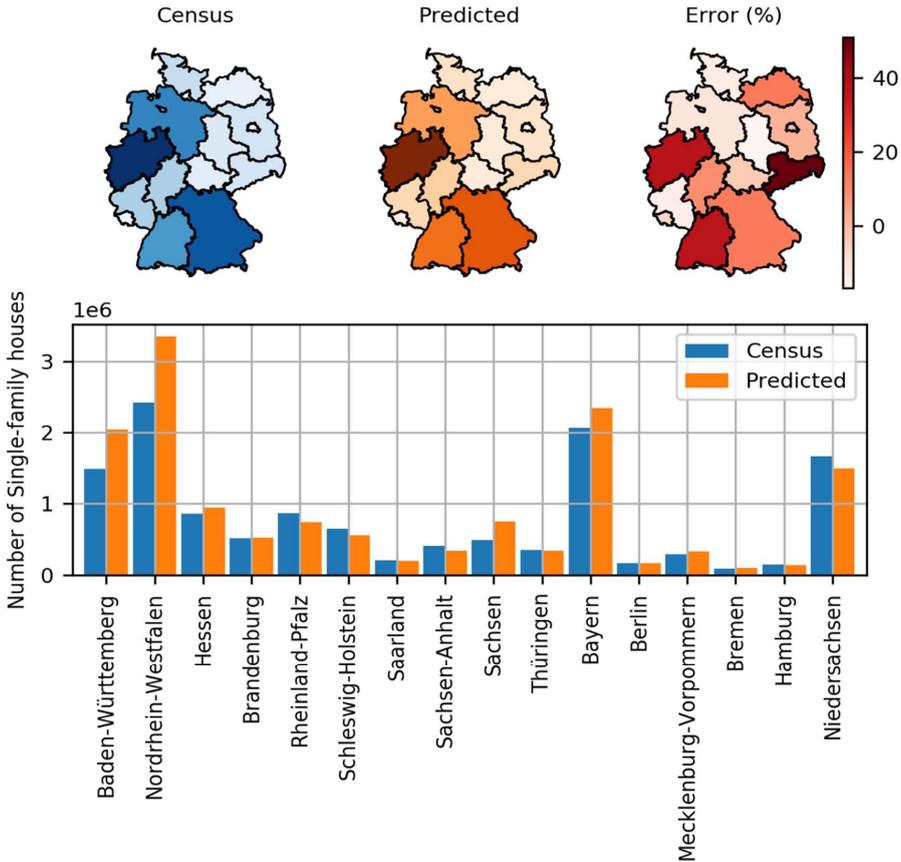


Figure 4-27 Single-family houses count comparison for each federal state of Germany

These buildings with spatial attributes and types are a prerequisite for developing geo-referenced synthetic distribution network topologies (Chapter 5) and for the distribution grid model in Chapter 6. For the electrical network, the building types are valuable for allocating load profiles. Therefore, the following section will address the generation of synthetic load profiles. In addition, the section will include various other demand and generation profiles.

4.4 Synthetic Load Profiles for the Classified Buildings

After predicting the building types, the main research question can only be answered using load and generation profiles to examine the vulnerabilities of the developed geo-referenced synthetic distribution networks and the implications of their future integration into the distribution networks. Load profiles such as residential load profiles without heat pumps, individual heat pump profiles, and battery electric vehicle charging profiles are derived in this section. In addition, generation profiles such as residential solar rooftop photovoltaic are also presented.

First and foremost, residential load profiles are generated using the framework Fine.Building [221]. The load profiles for a one-person household, a two-person household, a three-person household, a four-person household, and a five-person household are synthesized in this section. However, it is difficult to distribute these load profiles in residential buildings. Nevertheless, as noted previously, the number of predicted residential buildings are 19,747,802, including 14,378,635 single-family houses, 4,350,183 multi-family houses, and 1,018,984 apartment building types. Additionally, Germany has approximately 40,828,717 available dwellings [220]. As a matter of fact, it is evident that a single-family residence contains a single dwelling, but an apartment complex contains an average of 13 dwellings. Using this information, multi-family houses are assigned an average of five dwellings.

Each dwelling will have a load profile. According to official data [214], 42.3% of households are occupied by single people. Similarly, 33.2%, 11.9%, 9.1%, and 3.5% of households are occupied by two people, three people, four people, or five or more people, respectively. As a result, one-person, two-person, three-person, four-person, and five-person household profiles are generated using FINE.Building framework [221], as illustrated in Figure 4-28.

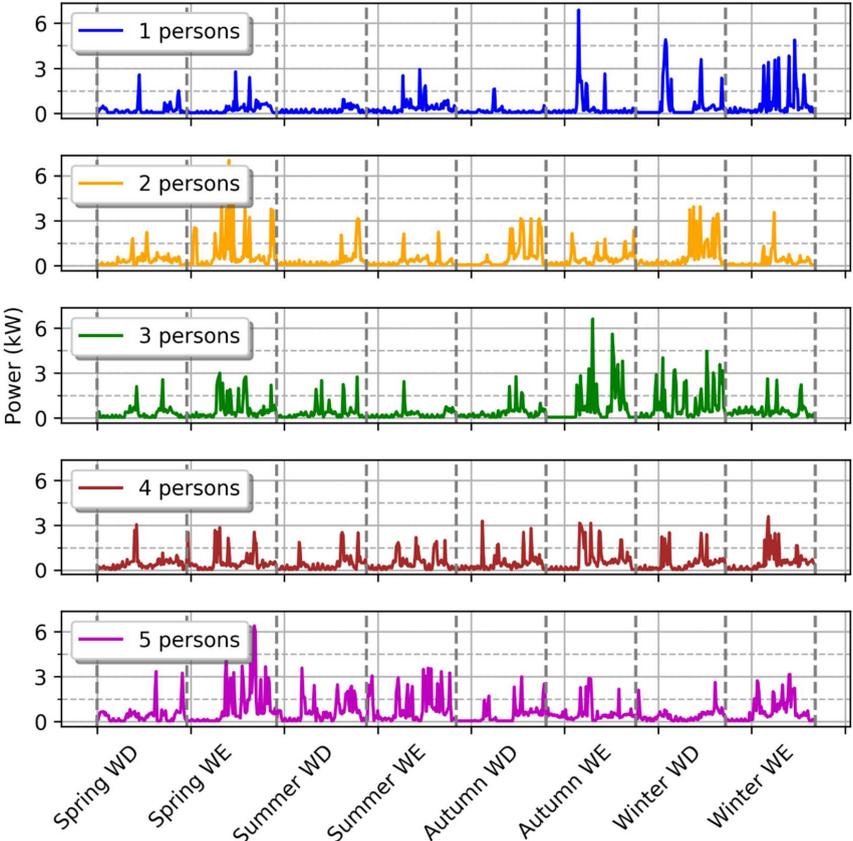


Figure 4-28 Household load profiles for one, two, three, four, and five persons

These profiles have a 15-minute temporal resolution. Fine.Building [221] can generate synthetic profiles for any fiscal year. Based upon these data, load profiles for eight days (weekday (WD) and weekend (WE) for four seasons) are developed to reduce the complexity of the distribution grid model (see Chapter 6). These profiles are distributed randomly with a probability of 0.42, 0.33, 0.11, 0.1, and 0.04 for all dwellings in Germany (see, Figure 4-29).

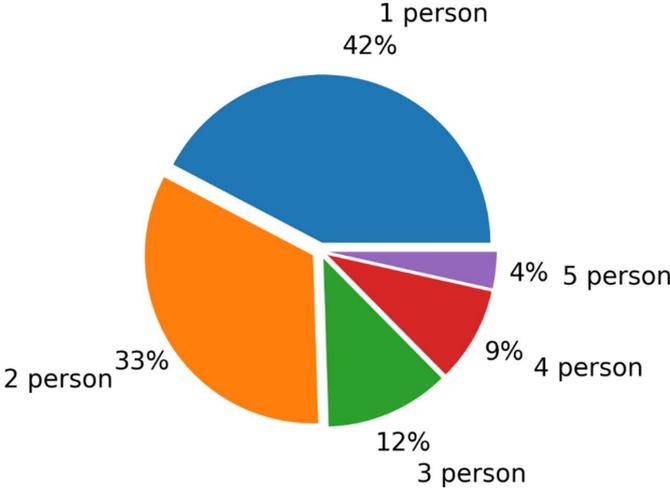


Figure 4-29 Households as a percentage of total dwellings in Germany.

The residential load profiles presented above are generated when there are no heat pumps in the dwellings. In Chapter 1, the future implication of the increased integration of heat pumps into the residential sector is already discussed. Therefore, this section synthesizes additional heat-pump profiles for residential dwellings.

In this context, using FINE.Building [221], individual heat pump profiles for one-person, two-person, three-person, four-person, and five-person houses are gathered. These profiles, like the household profiles, are collected on weekdays and weekends for four seasons with a 15-minute time resolution. Furthermore, these profiles are associated with each building in accordance with the residential load profile allocation. That is, if a home is a single-family house, a profile with the probability depicted in Figure 4-29 will be chosen at random. If the randomly picked profile is for a two-person home, the heat pump profile for that household will serve as the selection for that dwelling. When a building is designated as a single-family dwelling, just one profile is selected. Additionally, if a building is a multi-family house or an apartment, five and thirteen profiles are selected, respectively. However, the share of dwellings equipped with a heat pump will be decided based on future penetration scenarios in the residential sector. Figure 4-30 illustrates the synthetic heat pump profiles which will be assigned to the residential dwellings. Heat pumps are primarily used in the winter and only partially in the spring and autumn.

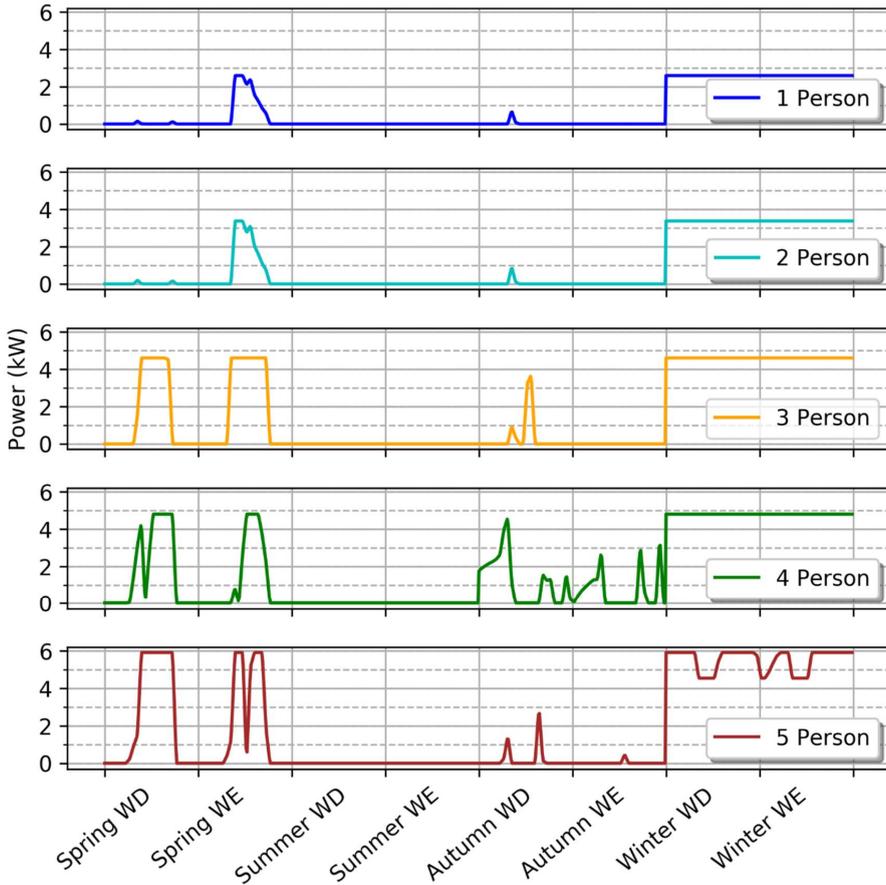


Figure 4-30 Heat pump load profiles for one, two, three, four, and five-persons in a household

We will now discuss the new distributed demand associated with Battery Electric Vehicles (BEVs). As mentioned in Chapter 1, a high penetration of battery electric vehicles will eventually integrate into distribution networks. To assess their impact in this context, it is necessary to understand the integration and electricity consumption pattern. These patterns are elaborately established by Linssen et al [222]. This study considers three distinct groups of battery electric vehicle users. Full-time employees were considered in the first group, pensioners, housewives/men were considered in the second group, and part-time employees are considered in the third group. The departure and arrival times and the distance traveled by each group of users are analyzed and the average value for each group is showcased in Table 4-12.

Table 4-12 Average time and distance travelled by the battery electric vehicle users in different clusters

	Parameter	Weekday (Time)	Weekend (Time)
Cluster 1	Departure from home (HH:MM)	07:27	09:47
	Arrival at home (HH:MM)	18:47	11:41
	Driving distance (km)	46.81	17.04
Cluster 2	Departure from home (HH:MM)	09:20	13:12
	Arrival at home (HH:MM)	12:02	18:20
	Driving distance (km)	18.88	38.84
Cluster 3	Departure from home (HH:MM)	16:24	04:24
	Arrival at home (HH:MM)	19:02	18:17
	Driving distance (km)	20.89	42.06

However, these values are used to generate load profiles for the weekday and weekend for three groups. These developed profiles are intended for cases of uncontrolled or unregulated charging. In other words, once the vehicle returns home from work or other routine activity, it connects to the electricity grid and charges until it is completely charged. These profiles are repeated for each season since the arrival and departure times, as well as the distance traveled, are not dependent on the season. Moreover, since household profiles have a 15-minute time resolution, these battery electrical vehicle profiles are also scaled to a 15-minute time resolution. The profiles generated are illustrated in Figure 4-31. Here, the profiles evaluated are for vehicles connected to a 3.7 kW power source and the average energy usage of vehicles is assumed to be 17.5 kWh/100 km.

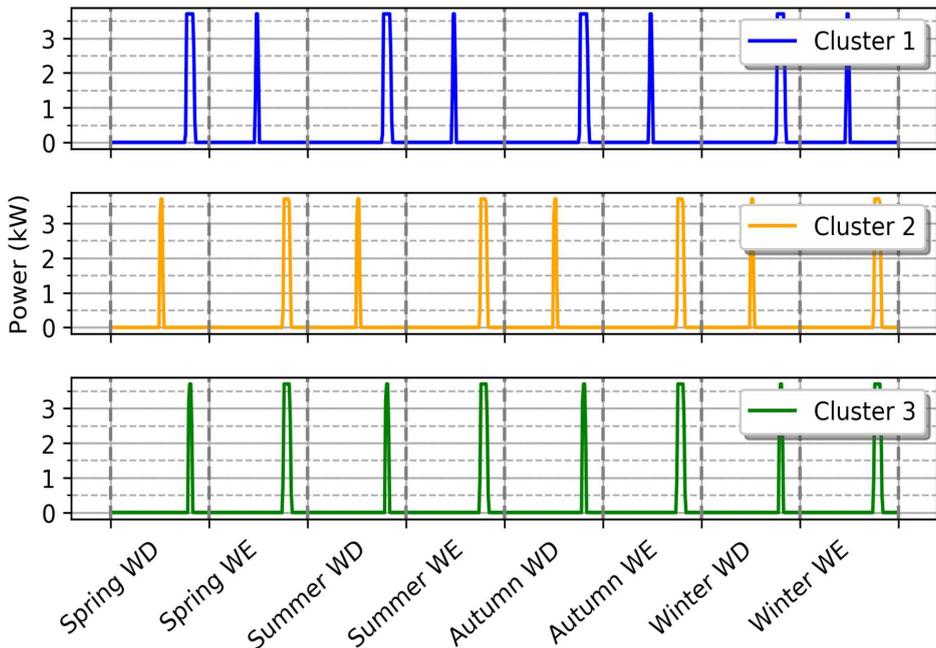


Figure 4-31 Battery electrical vehicle load profiles for three different clusters

As discussed in Chapter 1, there is a high probability of the heavy integration of rooftop PV into the distribution networks. In Chapter 6, the profiles corresponding to solar rooftop PV generation will be introduced to assess the influence of solar rooftop PV generation. However, it is worth mentioning here that residential rooftop PV profiles are generated for different locations in Germany.

Residential rooftop PV profiles are generated to analyze future decentralized generation in the distribution networks. The consideration of solar rooftop PV on each building takes into account distinct solar rooftop PV profiles for each building and the variation of solar irradiation based on location and time. To reduce model complexity only four profiles are assumed in the thesis: east, west, north, and south locations of Germany. In addition, four seasons of Spring, Summer, Autumn, and Winter with 15-minute time resolution are addressed to account for the variation in time. The rooftop PV generation profiles are modeled using the Pvlb library [223]. However, the distribution grid model in Chapter 6 consider these generated profiles (see Figure 4-32).

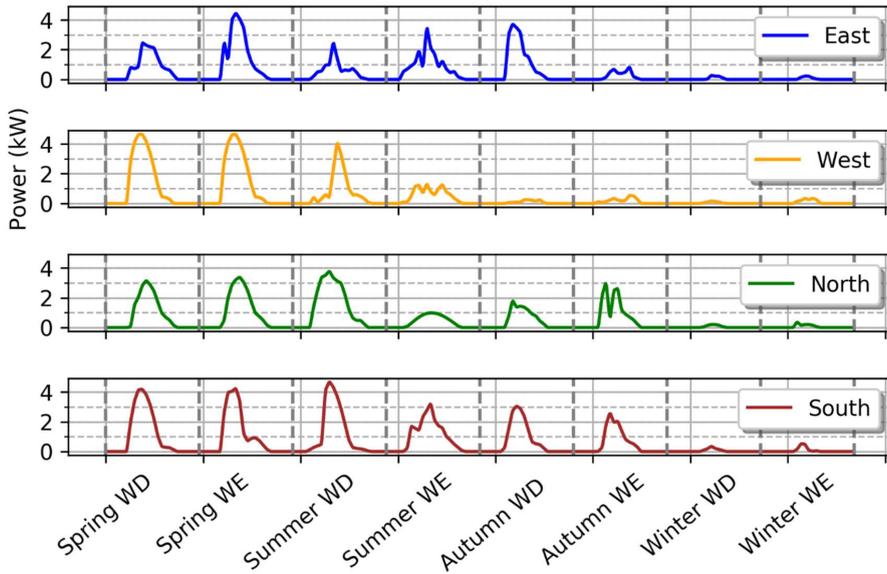


Figure 4-32 Residential PV profiles at a location of East, West, North, and South of Germany

To conclude, this chapter discussed open and synthetic data that were necessary for generating geo-referenced synthetic network topologies and distribution grid modeling to answer the overarching question of the thesis. In the next chapter, the modeling of geo-referenced synthetic distribution networks using open data is discussed.

4.5 Summary

To address the thesis's overarching question, the study conducted in Chapter 0 encourages the development of geo-referenced synthetic distribution networks. However, to assist in the development of geo-referenced synthetic networks, Chapter 3 covered the graphical and electrical characteristics, as well as numerous other real-world network and reliability indicators. As a result of the analysis in Chapter 3, it was determined that the data available from Germany should be used to develop algorithms for estimating geo-referenced synthetic distribution network topologies. Based on this context, this chapter addresses the question "*What data are required for estimating geo-referenced synthetic network topologies?*" Since the networks are being developed on a national scale, the model requires geographical data such as buildings, power lines, transformers, and substations. To retrieve this information, OpenStreetMap was selected due to the availability of required data and the comprehensive geographical data contained within this dataset. However, an elaborate analysis of the dataset reveals that the data required for estimating geo-referenced synthetic network topologies includes buildings to locate end-consumers, road infrastructure for power lines, medium-voltage transformer locations, high-voltage substation locations, and high-voltage power lines. The dataset, however, has some limitations. The limitations include the absence of adequate tags for buildings, which are essential to assign end-consumer load profiles. Identifying this

limitation introduces another question: "*How can these issues be addressed?*" This chapter addressed this issue by developing a classification model to categorize buildings into residential and non-residential buildings using state-of-the-art machine learning methods by supplementing the OSM data with additional data. Additionally, the model was used to further categorize residential buildings as single-family houses, multi-family houses, and apartment buildings. In this manner, this chapter successfully handled the limitations of the data. Additionally, synthetic load and generation profiles are established to aid in the assignment of load and generation profiles to the residential buildings.

Key messages:

- ✚ Open data for estimating geo-referenced synthetic distribution network topologies are extracted from OSM data using a series of methods and tools include osmosis, osm2pgsql, and Python.
- ✚ For high-voltage and extra-high voltage power lines extracted from OSM data, percentage errors of -39.9% , and -3.2% were observed.
- ✚ 29,494,772 buildings were extracted from OSM data, yet only 20.5% of these buildings holds tags.
- ✚ To improve the quality of data, additional data from CORINE, building heights, and census data were incorporated into the data extracted from OSM.
- ✚ The classification model categorizes 19,747,802 buildings as residential buildings.
- ✚ Single-family houses, multi-family houses, and apartment buildings are predicted to account for 14,378,633, 4,350,183, and 1,018,984 of the residential buildings, respectively.
- ✚ Residential building validation results in a percentage error of 3.4%, whereas the combined validation of single-family, and multi-family house and apartments results in a percentage error of 13.14% and -15.3% , respectively.
- ✚ Residential load profiles and heat pump profiles will be assigned to residential dwellings with a probability of 0.42, 0.33, 0.11, 0.1, and 0.04 for one-, two-, three-, four-, and five-person households, respectively.

5 Estimation of Synthetic Distribution Network Topologies

How are geo-referenced synthetic distribution network topologies estimated, what are the algorithms utilized, and how are they developed?

Topics covered:

-  Estimation of geo-referenced synthetic low-voltage networks
-  Estimation of geo-referenced synthetic medium-voltage networks
-  Estimation of geo-referenced synthetic high- and extra-high voltage network as single network

As was concluded in the preceding chapters, the overarching question of this thesis “*Are the distribution networks adequate for the anticipated increase in demand and generation in the future?*” can only be answered through the estimation of the geo-referenced synthetic distribution network topologies. Therefore, this chapter will describe the general modeling approach for estimating geo-referenced synthetic distribution network topologies, explicitly emphasizing the German distribution networks. As noted in Chapter 3, the methods established here can also be utilized to construct synthetic networks for various other European nations. It is important to note that the distribution grid networks vary by country. For instance, low-voltage, medium-voltage, and high-voltage are all considered distribution networks in Germany. In contrast, only low-voltage and medium-voltage are considered distribution networks in the majority of countries, while high-voltage is regarded as a transmission grid. As a result, a novel methodology is introduced and discussed in Section 5.1 and Section 5.2 that can assist with generating geo-referenced synthetic low-voltage and medium-voltage network topologies. Furthermore, using an open model for estimating high-voltage and extra-high voltage will be briefly discussed in Section 5.3. When generating network topologies, each voltage level of the distribution network that is being constructed will begin with an introduction to its networks. Following the collection of data necessary for estimating specific voltage level networks, the relevant graphical and electrical networks will be built using these data. To be precise, the modeling procedure will be divided into two steps: graphical network estimation and electrical network estimation. Additionally, the process of validating the generated networks is discussed in Section 5.4. The validation of the generated synthetic low-voltage networks will be provided to determine the model efficiency and gain insights that can help enhance the developed model in the future. Figure 1-3 illustrates the overall scheme for building this network structure while Figure 5-1 illustrates the general structural overview of modeling each voltage level.

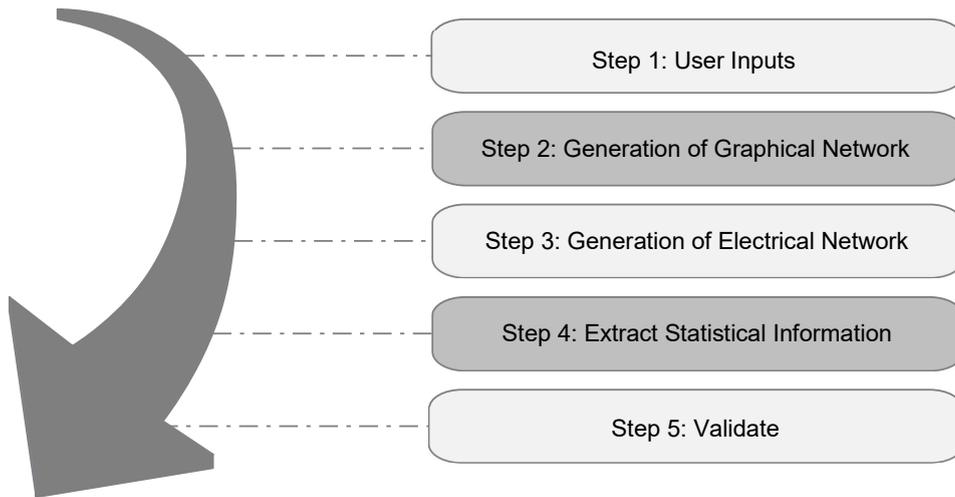


Figure 5-1 A block diagram representation of the modeling approach used to estimate geo-referenced synthetic network topologies.

5.1 Geo-referenced Synthetic Low-voltage Networks

Since public research does not have access to proprietary information, this study has placed a strong emphasis on utilizing only publicly available data to develop geo-referenced synthetic distribution networks. To do so, a bottom-up approach will be utilized. Low-voltage networks are the primary networks that supply electricity to end-users and thus, they are first in line within the process. The low-voltage networks are fundamentally comprised of a secondary medium-voltage to low-voltage (MV/LV) transformer and end-consumers. These two aspects are connected by underground cables or overhead lines, as appropriate. Several protection components are also connected to low-voltage networks, which include switches, fuses, circuit breakers, etc. However, these protection components are not considered in the modeling since this thesis aims to perform steady-state analysis. As such, the suggested methodology utilizes available data to determine end-user locations, power lines, and low-voltage transformers. The following section will, however, discuss the user input data required for estimating geo-referenced synthetic low-voltage networks. In addition, if the data for the model development are insufficient, heuristics will be assumed, which will be discussed in greater detail in the subsequent sections.

5.1.1 User Input

The first significant data input necessary for model development is information pertaining to the end-user location. Therefore, data regarding the locations of buildings are extracted from OpenStreetMap. However, in this case, only residential buildings are considered because most of the non-residential buildings, such as industrial, and commercial buildings, are connected to medium- and high-voltage networks. As discussed in the previous chapter, residential buildings are further categorized into single-family houses, multi-family houses, and apartment

buildings. As an initial step, residential building footprints along with the house type are considered for end-user locations. In this step, approx. 19.7 million residential buildings have been identified to establish synthetic low-voltage networks.

To develop appropriate synthetic low-voltage networks, the low-voltage transformer locations are also necessary in addition to the end-user locations. In Chapter 3, it was determined that open data about the locations of low-voltage transformers are not properly covered in the OSM data. There is, however, information available regarding the total number of low-voltage transformers currently installed in each EU nation. In Germany, there are approx. 500,000 low-voltage transformers in operation [201]. This information should be used when anticipating the identification of the transformer locations. In addition to the transformer location, knowledge about transformer type is also essential to build comprehensive low-voltage networks. Accordingly, four major transformer types with capacities ranging from 160 kVA to 630 kVA are utilized. Table 5-1 summarizes the most frequently used transformer capacities.

Table 5-1 Frequently used MV/LV transformer capacities in Germany [Adapted from [154]]

Low-voltage transformer types	Capacity	Voltage
MV/LV transformer type I	160 kVA	20/04 kV and 10/0.4 kV
MV/LV transformer type II	250 kVA	20/04 kV and 10/0.4 kV
MV/LV transformer type III	400 kVA	20/04 kV and 10/0.4 kV
MV/LV transformer type IV	630 KVA	20/04 kV and 10/0.4 kV

In addition to transformers and end-user locations, power lines are another essential component of low-voltage networks that must be considered. Examining the location data discussed in Chapter 4 corresponding to power lines in the low-voltage network, minimal power lines were identified in the OSM dataset. Furthermore, according to the findings in Chapter 3, 87% of the total power lines in low-voltage networks are underground cables. Typically, low-voltage power lines run along the sides of most roads. Therefore, in this method, road infrastructure abstracted from OpenStreetMap data is utilized to model the low-voltage power lines. However, the modeling of power lines will be discussed in the following section. Each of the modeled power lines should include a line type that can be assigned. Therefore, due to the high percentage share of underground cables, conventional cables from the literature are employed, such as those listed in Table 5-2. Each of these cable types has a specific current rating in the range of 123 A to 364 A. The data in Table 5-2 reveals that increasing the diameter of the cable increases the line's current carrying capacity.

Table 5-2 Standard low-voltage cable types in Germany [Adapted from [155]]

Cable	I_r (A)	R ($\frac{\Omega}{km}$)	X ($\frac{\Omega}{km}$)
NAYY 4 X 35 mm ²	123	0.868	0.08
NAYY 4 X 120 mm ²	245	0.253	0.08

NAYY 4 X 150 mm^2	275	0.206	0.08
NAYY 4 X 240 mm^2	364	0.125	0.08

In addition to the basic network components discussed above, another vital piece of data pertains to the distributed generators installed in low-voltage networks. The information regarding the renewable energy systems installed in the low-voltage networks is available in Open Power System Data (OPSD) [200]. Open power system data contains detailed information about the locations as well as the capacities of various distributor generators installed in low-voltage networks (see, Appendix D). Amongst the generators that are available, solar PV accounts for most of the installed capacity in low-voltage distribution networks. Since the necessary information has been discussed in this section, the final user input used to create the model is listed in Table 5-3.

Table 5-3 Final user input data for estimating synthetic low-voltage networks

End-user	Geographically represented residential building footprints
Transformers	Total low-voltage transformer count (Location by the developed model)
Power lines	Road infrastructure including streets, highways, footpaths, and bike paths
Power plants	Open power system data extracted low-voltage distributed generator locations and their capacities
Structure	Radial structure

With the input data, the following section discusses the model for estimating synthetic low-voltage networks.

5.1.2 Modeling Methodology

The previous section discussed the primary user input. In this section, the methods used for generating geo-referenced synthetic low-voltage networks will be delivered along with their implementation. Data pertaining to Germany is used by the developed algorithms to generate geo-referenced synthetic low-voltage networks for the country. However, as depicted in Figure 5-2, the methodology used to generate geo-referenced synthetic low-voltage networks has been divided into two layers: the development of the graphical network and the development of the electrical network.

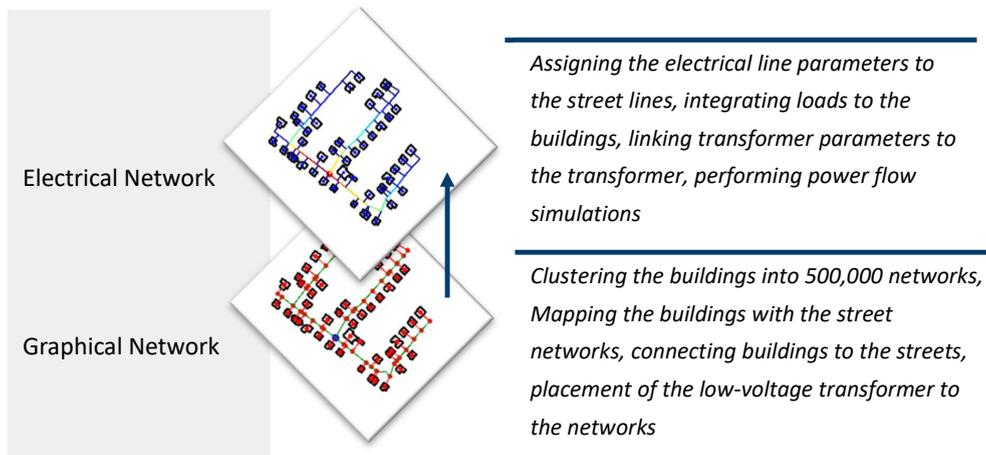


Figure 5-2 Schema for low-voltage network modeling

The following subsections demonstrate low-voltage graphical and electrical networks.

5.1.2.1 Graphical Network

When developing geo-referenced synthetic low-voltage networks, the first layer is the graphical network. In the graphic layer, the buildings must first be clustered into 500,000 clusters. This step was necessary because there are 500,000 low-voltage networks in Germany and each low-voltage network contains different buildings. Additionally, low-voltage networks' primary users are located in residential buildings. As a result, the 19,747,802 residential buildings are clustered into 500,000 clusters. Numerous methods, such as unsupervised machine learning algorithms, are investigated to execute the clustering task, including K-means clustering [224], hierarchical clustering [225], and a density-based spatial clustering algorithm with noise (DBSCAN) [226]. However, neither of the two algorithms described here, hierarchical or DBSCAN, is suitable in this scenario because DBSCAN is based on the number of points with a given radius. This algorithm considers two hyperparameters. Therefore, when the building's dataset is clustered using the DBSCAN algorithm, some points are excluded to create outliers, as illustrated in Figure 5-3. Figure 5-3 also illustrates how, given some data points, the requirement is to cluster the data points into three clusters. However, the additional information this algorithm requires is a radius. Based on the radius, a centroid is generated to form a circle. Here, all the points inside the circle will be grouped, while the data points outside the circles that are represented as outliers are neglected and not considered in any cluster. This is the primary disadvantage of this algorithm.

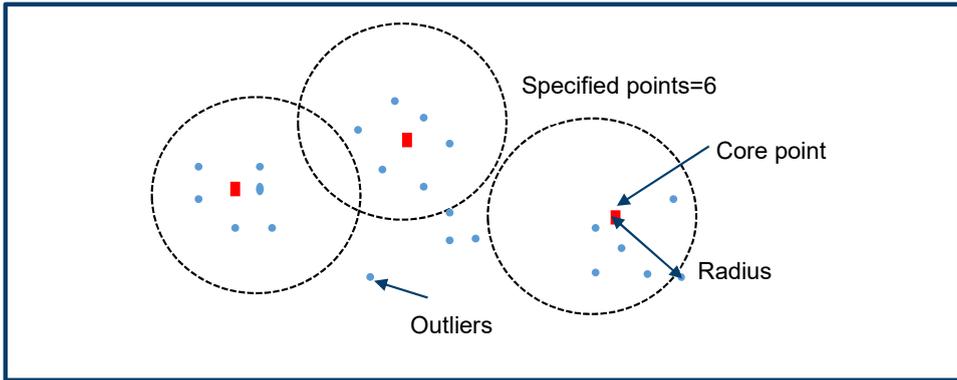


Figure 5-3 An example showing the application of DBSCAN

On the other hand, hierarchical clustering is unsuitable for this application due to the time complexity. $O(n^2)$, where n is the input size, is quadratic and requires a significant amount of time. Subsequently, this technique is incapable of handling large amounts of data. Therefore, K-Means clustering was chosen as the final method for clustering the set of 19,747,802 residential buildings. However, during implementation, this algorithm's shortcoming became apparent. Since the data is of high volume and variety, it causes memory constraints. Basically, K-Means works with distances between the data points, and holding the Euclidean distance between each pair of nodes requires a significant amount of memory space. This issue was overcome by implementing a new approach using Joblib and DASK [228] on K-Means (Scikit-learn) [229], which clusters data points using several CPUs in parallel. Yet, even with the parallel use of several computers, a memory outage occurs. This is because the K-Means algorithm has a space complexity of $O(N(D + K))$ [227], where N is number of points, D is number of dimensions, and K is number of centers. Additionally, K-Means also has time complexity of $O(NKI)$ [227], where I is the number of iterations. However, since there is still a potential to address the memory issue, scalable machine learning implementation using DASK-ML [228] is utilized further. When these two techniques were implemented on the residential building dataset to cluster up to 500,000 clusters, it was observed that only 100 clusters can be achieved for K-Means (Scikit-learn + Joblib) [229]. In contrast, using DASK machine learning algorithms [228], a maximum of 10,000 clusters can be achieved. Figure 5-4 illustrates how K-Means (Scikit-learn + Joblib) [229] and K-Means are distinct by using scalable DASK machine learning algorithms [228]. However, a memory outage has occurred after the mentioned clusters that are 100 and 10,000 for K-Means (Scikit-learn + Joblib) [229] and DASK-ML [228].

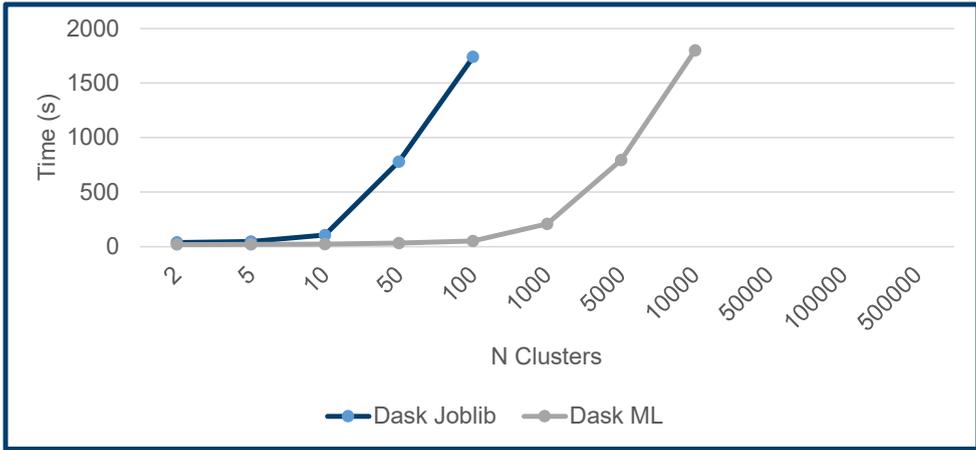


Figure 5-4 Time required to cluster data points using K-Means (Scikit-learn + Joblib [229]) and K-Means (DASK-ML [228])

Therefore, residential building data is separated by NUTS-3 level, and the total number of low-voltage networks is distributed based on the density of the residential buildings. For example, Figure 5-5 depicts the low-voltage network distributed to the NUTS-3 level in the context of Germany. The figure also demonstrates that Berlin has the highest number of residential buildings and low-voltage networks are assigned accordingly. However, the total number of low-voltage networks in each administrative district varies between a few hundred and approximately 9000. Nevertheless, the majority of the districts have low-voltage networks ranging from 500 to 1500 networks. Therefore, when using DASK-ML, K-Means clustering can be achieved when each administrative district is clustered according to the low-voltage networks count.

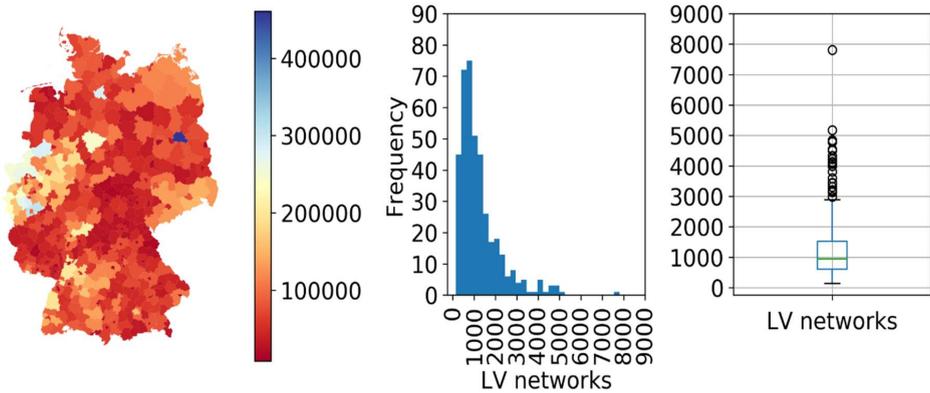


Figure 5-5 Distribution of low-voltage networks to administrative districts in Germany

K-Means clustering was also accomplished using DASK-ML on residential buildings scattered across 401 NUTS-3 regions. Each administrative district's residential buildings were clustered

according to the district's low-voltage networks (refer, Figure 5-5). Thus, this process helped resolve the memory bottleneck, as the number of clusters required in each district is smaller than the number of clusters resulting in memory problems (refer to Figure 5-4). Overall, combining the individual clusters at each administrative district (i.e., 401) resulted in the outcome of 500,000 clusters, wherein each cluster will have a certain number of residential buildings. The approach that was developed to complete this task is depicted in Algorithm 1.

Algorithm 1: Clustering the residential buildings to match low-voltage networks

Algorithm 1: Clustering the residential buildings / Datapoints with latitude and longitudes	
Input	Number of clusters (K) Geo-referenced residential building footprints ($B = \{B_1, B_2, \dots, B_n\}$) Administrative district shapefiles (NUTS-3)
Steps	1: Extract centroid coordinates for each building ($B_c = \{B_{c1}, B_{c2}, \dots, B_{cn}\}$). 2: Utilize NUTS-3 shapefiles to map residential building centroids to extract buildings from each administrative district $B_N = \{B_{N1}, B_{N2}, B_{N3}, \dots, B_{N401}\}$. 3: Distribute clusters required/low-voltage networks required in proportion to the number of residential buildings within administrative districts $K_N = \{K_{N1}, K_{N2}, \dots, K_{N401}\}$. 4: for $i = 1, \dots, 401$ do Cluster K_{Ni} using scalable K-Means on B_{Ni} Each residential building is assigned a number between 0 and K_{Ni} that corresponds to the cluster. 5: Collect K clusters by combining the residential buildings in each administrative district $B_K = \{B_{K1}, B_{K2}, \dots, B_{KK}\}$.
Return	Residential buildings with an additional characteristic indicating the cluster number $B_K = \{B_{K1}, B_{K2}, \dots, B_{KK}\}$

However, when Algorithm 1 is applied with the known total number of low-voltage networks in Germany (i.e., 500,000), the residential buildings are clustered into 500,000 clusters. The building density was used to determine the identified clusters. However, this would not be the case in real-world networks, as low-voltage networks are generally designed based on the load carried by each building. Therefore, an algorithm is developed by modifying the K-Means clustering technique to help limit the network's or cluster's peak load. Algorithm 2 illustrates the method for limiting the network's peak load so that the network remains feasible.

Algorithm 2 Clustering the residential buildings by limiting the cumulative peak load

Algorithm 2: Residential buildings /Data points clustered by weights	
Input	Number of clusters (K) Geo-referenced residential building footprints ($B = \{B_1, B_2, \dots, B_n\}$) Administrative district shapefiles (NUTS-3) Maximum low-voltage transformer capacity (P)
Steps	1: Create centroid coordinates for each building ($B_c = \{B_{c1}, B_{c2}, \dots, B_{cn}\}$).

	<p>2: Utilize NUTS-3 shapefiles to map residential building centroids to extract buildings from each administrative district $B_N = \{B_{N1}, B_{N2}, B_{N3}, \dots, B_{N401}\}$.</p> <p>3: Distribute clusters / low-voltage networks in proportion to the number of buildings within administrative districts $K_N = \{K_{N1}, K_{N2}, \dots, K_{N401}\}$.</p> <p>4: <i>for</i> $i = 1, \dots, 401$ <i>do</i></p> <p style="padding-left: 2em;">Choose the number of clusters K_{Ni}</p> <p style="padding-left: 2em;">Randomly select K_{Ni} data points to start as a cluster centroid</p> <p style="padding-left: 2em;">Calculate the Euclidean distance ($d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}$), between the data points to the centroids. Where, p, q contains cartesian coordinates</p> <p style="padding-left: 2em;">Find $argmin$ ($\arg \min_{x \in S} f(x) = \{x \in S : f(s) \geq f(x) \forall s \in S\}$) gives the closest centroid to each data point</p> <p style="padding-left: 2em;">Calculate the sum of weights of each data point that belongs to each centroid (i.e., the sum of peak loads corresponds to the building)</p> <p style="padding-left: 2em;"><i>If</i> cumulative weight exceeds the given constant(P) (i.e, transformer capacity)</p> <p style="padding-left: 2em;">The farthest data point in the current cluster will be assigned to the next nearest centroid until the cumulative weight $\leq P$.</p> <p style="padding-left: 2em;">Repeat for all the clusters</p> <p style="padding-left: 2em;">The remaining residential buildings, after all the clusters satisfy will be grouped to form different clusters (i.e., exceeding $K_{Ni} \rightarrow K_{Ni} + x_i$).</p> <p style="padding-left: 2em;">Each residential building is then assigned a number between 0 and $K_{Ni} + x_i$ that corresponds to the cluster.</p> <p>5: Collect $K + X$ clusters by combining the residential buildings in each administrative district $B_{KX} = \{B_{K1}, B_{K2}, \dots, B_{K_{K+X}}\}$.</p>
Return	Residential buildings with an additional characteristic indicating the cluster number $B_{KX} = \{B_{K1}, B_{K2}, \dots, B_{K_{K+X}}\}$

When Algorithm 2 was implemented on the OSM dataset, it worked efficiently for a small set of data points. However, since this study employs a large volume of buildings data, this solution takes an inordinate amount of time. Therefore, clustering was performed without considering Algorithm 2, and weights were applied to each residential building using the load profiles mentioned in Chapter 4. Following the load profile assignment, the peak load for each building was selected, and the cumulative of each cluster was then calculated. If the cluster peak loads exceeded the highest transformer capacity by 1.5 times or more, the cluster was further clustered until the network's peak load fell under the limit.

By implementing these methods, the residential buildings are clustered into 500,000 clusters ($K_j \forall j = \{0, 1, \dots, 499999\}$), as illustrated in Figure 5-6, which showcases the 500,000 groups. Each group has a specific number of residential buildings. For example, for the cluster (i.e., $j = 23988$) there are 48 residential buildings grouped together. Likewise, each cluster is assigned to a certain number of residential buildings.

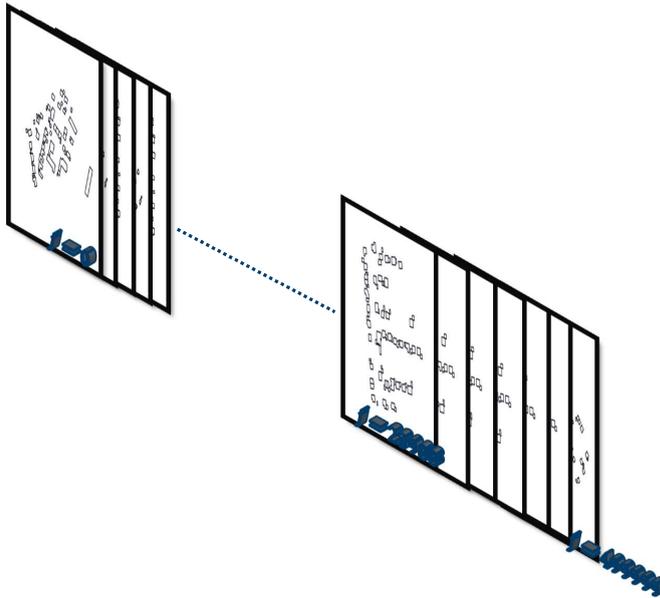


Figure 5-6 Residential buildings in the clusters

Once the residential buildings were assigned under each group, the next step involves producing a graphical low-voltage network for each group with a certain number of residential buildings. To illustrate this process, the cluster (K_{23988}) is selected as an example from the clusters shown in Figure 5-6. The methods for developing graphical and electrical low-voltage networks are explained herein.

The next phase in the modeling process entails collecting information regarding the roads in each cluster. In this context, Algorithm 3 implements a process to extract the road infrastructure (line geometry) from each cluster. In fact, the road infrastructure is considered for modeling power lines due to limited low-voltage power line visibility in the OSM, since 87% of low-voltage power lines in Germany are underground cables. Since all underground cables align along roads, the roads from the OSM dataset are retrieved for each cluster. As we are developing geo-referenced networks, all the objects considered are geometrical entities. For instance, buildings have polygon geometries, while their centroids have point geometries, and roads have line geometries. Henceforth, buildings, building centroids, and roads will be referred to as polygons, nodes, and edges, respectively.

Algorithm 3 Collecting road-infrastructure from OpenStreetMap spatially (line geometries)

Algorithm 3: Collecting road-infrastructure from OpenStreetMap spatially	
Input	Nodes (NB) Extracted line geometries from OpenStreetMap for Germany (EG) Buffer (R)
Steps	1: Collect the nodes 2: Select the center node from the set of nodes.

	3: Produce a bounding box (BB) (i.e., a polygon geometry) with a buffer (R) from the selected center node 4: Map the geographical bounding box with the line geometries (EG) 5: Extract the line geometries that intersect the bounding box (BB). The extracted lines (E) contain some nodes (NE)
Return	Nodes ($N = NB + NE$), Edges (E)

Given the residential buildings in the cluster (i.e., K_{23988}) and a buffer of ten kilometers, Algorithm 3 extracts road infrastructure proximal to the residential buildings in the cluster. The buffer distance of ten kilometers was chosen to ensure that there is no possibility of missing lines near the buildings. Now that the nodes (i.e., residential building centroids, and road infrastructure end points) (N) and edges (i.e., roads) (E) have been retrieved in the given cluster, the nodes corresponding to the residential buildings are connected to the nearby edges. The nodes in the supplied cluster include the centroids of the residential building and the nodes on the edges. Therefore, nodes pertaining to residential building centroids are referred to as building nodes (NB), whereas nodes pertaining to edges are referred to as edge nodes (NE). The following stage entails the use of Algorithm 4 for connecting the building nodes to the nearest edge.

Algorithm 4 Connecting nodes and edges

Algorithm 4: Connecting nodes and edges	
Input	Nodes (N) Edges (E)
Steps	1: Collect the edges 2: Create a R-Tree from the edges (E) 3: <i>for</i> $i = 1, \dots, count(NB)$ <i>do</i> Project the building node (NB_i) perpendicular to the nearest line in the R-Tree formed by the edges (E) Generated new node (NEE_i) on the nearest line at the point where building node is projected Generate a new line (EN_i) joining building node (NB_i) and new node formed (NEE_i) Bisect the edge at the new node formed (NEE_i), so that the new edges (EE_i) are formed from the old edge
Return	Nodes ($N = NB + NE + NEE$), Edges ($E = E + EN + EE$)

Because there is always a service drop between the power line and the building, additional nodes are generated on the major edges by projecting the building node perpendicular to the nearest edge. The service drop results from the connecting of the residential building centroid with the newly formed node. This is achieved by using the Algorithm 4.

When the discussed methods are implemented on the cluster (i.e., K_{23988}), a graph structure is produced as represented in Figure 5-7. However, these methods do not lead to the

establishment of a proper graphical network, since they form a graph consisting of primary edge nodes, newly produced nodes, building nodes, edges, and newly formed edges. But in most cases, there will be missing edges and closed edges. Furthermore, as stated above, a ten kilometers buffer was considered, which gives rise to several edges that are not adjacent to the given residential building. This fact is clearly evident in Figure 5-7.

In order to properly generate a radial graph network, these uncertainties should be eliminated. Undesirable edges were eliminated by undertaking the following procedure. Each edge in the graph comprises of at least two nodes. However, if edges exist without a newly created node, those edges are removed from the graph. This enables the retention of just those edges containing building nodes, newly generated nodes, and some edge nodes, by eliminating undesirable edges from the network.

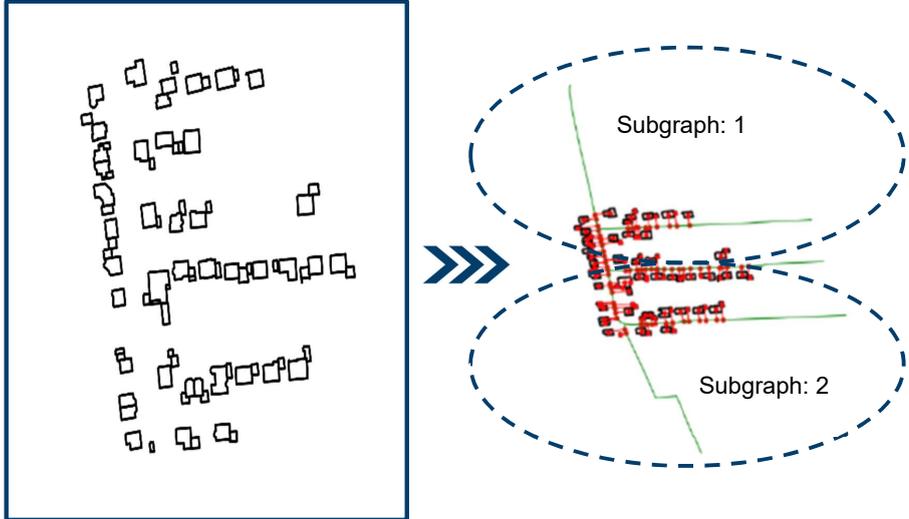


Figure 5-7 Graph generated by implementing Algorithm 4 on the considered cluster

Nevertheless, this process was not used to create a comprehensive graphical network due to the chances of producing subgraphs, as illustrated in Figure 5-7. There may be n number of subgraphs depending on the size of the network. However, this example contains only two subgraphs. Hence, these subgraphs must be joined to form a complete graphical network. If there are more than two, combine them one at a time to reduce the number of subgraphs to one. Algorithm 5 helps achieve this goal by connecting each subgraph by constructing a new edge between the two shortest node pair one from each subgraph. This procedure is repeated until there is only one subgraph, resulting in a complete graphical network without any uncertainty in the graph.

Algorithm 5 Constructing a single graph from subgraphs

Algorithm 5: Constructing a single graph from subgraphs	
Input	Nodes (N) Edges (E)
Steps	1: Collect the edges (E) 2: Create a graph $G(E, N)$ 3: Check for the subgraphs (SG) 4: <i>While</i> $SG > 1$ <i>do</i> Calculate the distance between edge nodes ($EN + EE$) of one subgraph (SG_i) to another subgraph (SG_{i+1}) Collect the node pair with minimum distance Generate a line geometry joining these node pairs (EG_i) Update the edges with the new edge ($E = E + EG_i$)
Return	Nodes ($N = NB + NE + NEE$), Edges ($E = E + EN + EE + EG$)

Implementing the Algorithm 5 on the graph structure shown in Figure 5-7 resulted in the graphical network shown in Figure 5-8. This graphical network contains edges, building nodes and edge nodes.

However, before the generation of electrical networks, the final phase in the graphical network generation process is to identify a node suitable for transformer placement.

Algorithm 6 is used to determine the transformer position in each graphical network with nodes and edges. Herein, a node is chosen for optimal transformer placement, which is enabled by positioning the transformers in the network's center, considering the weights on each edge. In fact, this approximation was considered to minimize the voltage drop on the farthest node. The identification of such a node is primarily determined by closeness centrality and the shortest path between the nodes. After obtaining the node suitable for transformer placement, a node must be named as *<transformer>*. This way, during the generation of an electrical network, transformer parameters can be associated with that node.

Algorithm 6 Identifying transformer location

Algorithm 6: Identifying transformer location	
Input	Nodes (N) Edges (E)
Steps	1: Collect the nodes (N) 2: <i>for</i> $i = 1, \dots, N$ <i>do</i> Calculate the closeness centrality $C(x)$ for each node (N_i) Where, $C(x) = \frac{N}{\sum_{i=1}^N d(x,y)}$, where $d(x, y)$ shortest path calculated using Dijkstra algorithm [230] between two nodes x and y Store the closeness centrality for each node in the network

	3: Higher the closeness centrality closer the node to all other nodes
	4: Rename the node with higher closeness centrality as <i><transformer></i>
Return	Nodes ($N = NB + NE + NEE$), Edges ($E = E + EN + EE + EG$)

When Algorithm 6 is implemented on the graph generated above for the considered cluster (i.e., K_{23988}), the location of placement of the transformer is provided, as depicted in Figure 5-8. The blue color node in the network denotes the node that contains the transformer. Additionally, the red nodes within the polygons represent building nodes (NB). The red nodes along the edges represent edge nodes ($NE + NEE$). The red color lines represent service drops (EN), and the green color edges represent power lines ($E + EE_i + EG_i$).

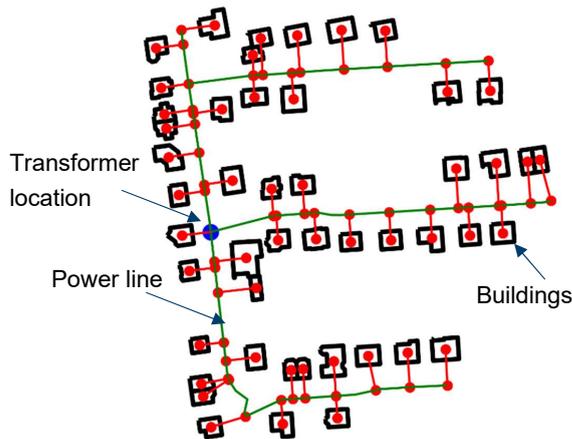


Figure 5-8 A graphical network generated for a cluster (K_{23988})

Additionally, an extension algorithm, labelled as Algorithm 7, was developed to facilitate the future integration of additional nodes or edges in the graphical network for network reinforcement. For instance, if a new building is constructed adjacent to an existing building, the load from the new building should be added to the existing network. Therefore, this reinforcement algorithm will be used to incorporate that building into the existing network.

However, Algorithm 7 was also developed to integrate distributed generators that already exist at the location of the low-voltage networks. There is publicly available data on distributed generators at various voltage levels in open power system data [200]. Distributed low-voltage generators such as solar, wind, hydro, biomass, geothermal, and others are extracted from there and analyzed. Appendix D illustrates the location of renewable energy sources in low-voltage networks, along with their capacities. Here, it is evident that around 23 GW of solar photovoltaic (PV) has been placed in low-voltage networks and is dominant over other technologies. The other technologies include bioenergy, wind onshore, and hydro accounts for 0.54 GW, 0.22 GW, and 0.03 GW, respectively.

Algorithm 7 Reinforcement of graphical network

Algorithm 7: Reinforcement of graphical network	
Input	Nodes (N) Edges (E) New nodes (NG)
Steps	1: Generate the graph network $G(N, E)$ 2: <i>for</i> $i = 1, \dots, NG$ <i>do</i> Calculate the distance from the new node (NG_i) to all other nodes (N) Collect the node from graph nodes (N) which has the least distance to the new node (NG_i) Generate new edge (EGR_i) 3. Update the nodes ($N = N + NG$) 4: Update the edges ($E = E + EGR$)
Return	Nodes ($N = NB + NE + NEE + NG$), Edges ($E = E + EN + EE + EG + EGR$)

When the data corresponding to the existing distributed generators is further analyzed, it reveals that the locations corresponding to them are not identical to the pinned locations of the power plants. Therefore, when the given location is integrated inside the low-voltage network graphical structure, additional challenges emerge. For instance, when the location inside a low-voltage network is given, the attributes corresponding to the location would include the total capacity of the generator units and the number of generators. In reality, the given number of generators are spatially distributed. Therefore, integrating the given capacity of all generators results in additional technical violation.

However, as mentioned previously, Algorithm 7 can handle network reinforcement, by connecting the distributed generators to the existing network in the future when the exact generator locations are specified. Nonetheless, when the graphical network's nodes and edges and the locations of the new distributed generators are provided, Algorithm 7 constructs the graphical network. To demonstrate this, previously considered cluster (i.e., K_{23988}) from 500,000 clusters was chosen. A graphical network with nodes and edges is generated by applying the series of previously developed algorithms. However, when given the new generator location in addition to the nodes and edges of the graph generated, Algorithm 7 constructs the new graphical network that is depicted in Figure 5-9. The black dot in Figure 5-9 represents the location of distributed power plants, and the yellow edge represents the new edge connecting the power plants to the existing network. On the other hand, the blue color node in the network denotes the node containing the transformer. Furthermore, the red nodes within the polygons represent building nodes, the red nodes along the edges represent edge nodes, the red color lines represent service drops, and the green color edges represent power lines.

Like the integration of power plants into the network, if a new node or building is constructed at the location of the given network, Algorithm 7 aids in constructing a new graphical network without disturbing the previous configuration.

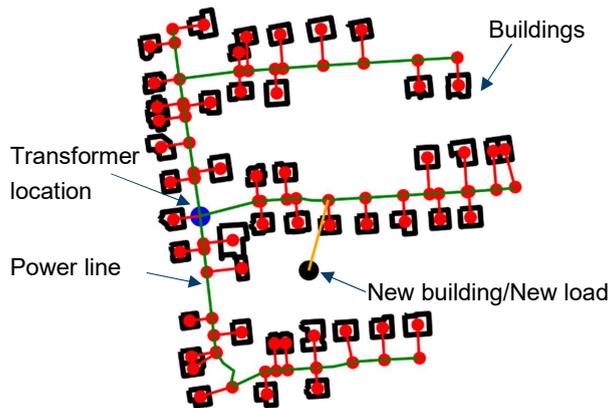


Figure 5-9 An illustration of the reinforcement on the graphical network (K_{23988})

However, the scenario calculations of the thesis do not include the extension of the graphical network with renewable energy sources that are already installed. This is because the future penetration of photovoltaic is significantly combined with the installation on buildings' rooftops. Therefore, each building will receive the rooftop PV depending on the scenario that is being considered. In a few cases and circumstances, the rooftop PV potential will be 100% utilized by 2050. In this case, a rooftop PV unit will be assumed for each building. Accordingly, rooftop PV generation will be integrated into the network. Chapter 6 will cover the integration of rooftop PV and other scenarios in the low-voltage networks.

Summing up, by executing all of these algorithms (i.e., Algorithm 1 to Algorithm 6) sequentially on the 500,000 clusters, each containing a few buildings, 500,000 low-voltage graphical networks are generated. Additionally, each graphical network has a node location that acts as a low-voltage transformer. Appendix E showcases a few of the 500,000 geo-referenced synthetic low-voltage networks. Finally, with the aid of these 500,000 graphical networks, the final stage of geo-referenced synthetic low-voltage networks, which entails generating electrical network, is accomplished in the following section.

5.1.2.2 Electrical Network

To facilitate the complete conversion of the graphical network into an electrical low-voltage network, certain electrical parameters must be provided as features to the graphical network. Transformers, loads, generators, and power lines are all essential parameters of an electrical network. Firstly, the transformer type for each established network should be identified. To perform this, loads should be assigned to each building. Therefore, the nodes corresponding to the building nodes (NB) in the graphical network are assigned with load profiles. As discussed in Chapter 4, each building node associated with the building type will receive its respective load profiles. For example, take a single-family house. In that case, a single load

profile is randomly chosen from one-person, two-person, three-person, four-person, and five-person households with probabilities of 0.42, 0.33, 0.11, 0.1, and 0.04, respectively. Similarly, multi-family houses receive five load profiles, and apartments get 13 load profiles, as per the assumption discussed in Chapter 4.

After assigning load profiles to the network, each network's transformer type at the transformer location is determined by calculating the network's peak load. The box plot in Figure 5-10 illustrates the peak load for all 500,000 networks generated. Here, 50% of the network's peak load is between 200 and 600 kW. However, the maximum peak load considered across all networks is approximately 1300 kW, and the minimum is 3 kW. Consequently, since the network's peak load is known, the transformer types are allocated from those available and demonstrated Table 5-1.

In this context, four basic types of transformers are frequently used in low-voltage networks, which are listed in Table 5-1. There are certain assumptions that are made when allocating a specific type of transformer to a network. In fact, the assumptions are made since data corresponding to the real-world networks are inaccessible. The assignment of transformer types is accomplished by considering the network's peak loads. For instance, if the network's peak load is less than 200 kW, a transformer with a capacity of 160 kVA is more appropriate and economical. Similarly, when the network's peak demand is greater than 200 but less than 300, a 250 kVA capacity transformer is assigned to the network. Additional assumptions that are made are illustrated in Table 5-4.

Furthermore, the transformers considered here have a maximum capacity of 630 kVA. Unfortunately, some networks are unable to meet the peak demand due to a lack of transformer capacity. As a result, it is evident that some networks will fail to converge for power flow analysis. There will be an opportunity to assess the networks' vulnerability to demand and the possibility of inserting transformers with a capacity of 1000 kVA or more. However, to conduct a preliminary study and estimate the electrical network, these four types of transformers are chosen.

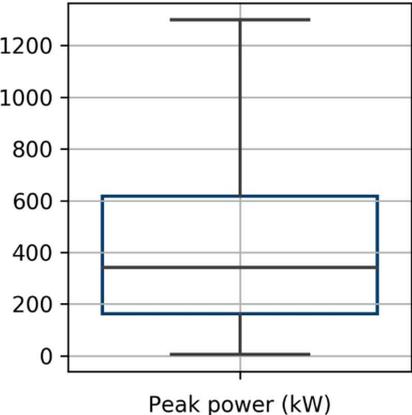


Figure 5-10 Peak power of the estimated 500,000 networks after assigning load profiles for each dwelling

Table 5-4 Assumptions considered for allocation of transformer type to low-voltage network

Peak load	Transformer type	Capacity
≤200 kW	MV/LV transformer type I	160 kVA
>200≤300 kW	MV/LV transformer type II	250 kVA
>300≤550 kW	MV/LV transformer type III	400 kVA
>550 kW	MV/LV transformer type IV	630 kVA

Figure 5-11 illustrates the distribution of low-voltage transformers over the 500,000 low-voltage networks, under the considered assumptions. It is also clear from the figure that the majority of networks are equipped with 630 kVA and 160 kVA transformers. Moreover, 33%, 30%, 23%, and 14% of the 500,000 networks are allotted with 630 kVA, 160 kVA, 400 kVA, and 250 kVA transformers, respectively. Furthermore, it is obvious from this figure that 33% of networks with 630 kVA transformations will have a high probability of demonstrating failure under power flow analysis. This is assumed because it is evident in Figure 5-10 that 25% of the networks have more than 600 kW peak.

In addition to the transformer and loads assignment to the graphical nodes, the power lines are an essential component that must be incorporated in the graphical network. Therefore, all network edges, service drops, and main power lines are assigned with cable types. Cables are chosen because 87% of powerlines in Germany are underground cables; consequently, for the convenience of computation, 100% of lines are deemed as cables. Therefore, from the list of cables indicated in Table 5-2, the cable type NAYY 4x150 mm² with a current carrying capacity of 275 A was chosen.

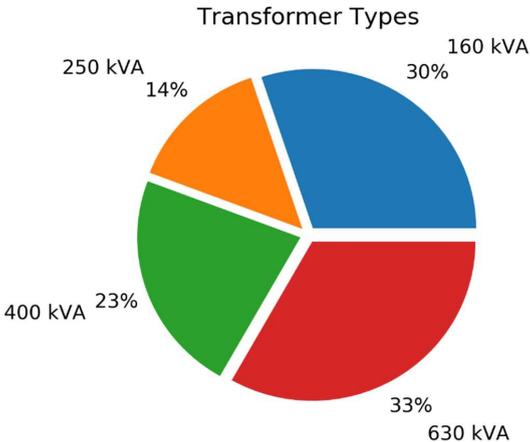


Figure 5-11 Low-voltage transformer type distributed to the estimated 500,000 low-voltage networks

Finally, based on the proposed scenarios, all future distributed demand and generation will be assigned to buildings on low-voltage networks. Here, the low-voltage network's nominal voltage is assumed to be 0.4 kV. When these electrical components are integrated into the graphical networks, they are ultimately converted into an electrical network, which enables the network for power flow analysis. The example cluster (K_{23988}) used while generating the graphical network that is assigned all of these components is illustrated in Figure 5-12, along with line loading at a specific time instant as per the power flow analysis.

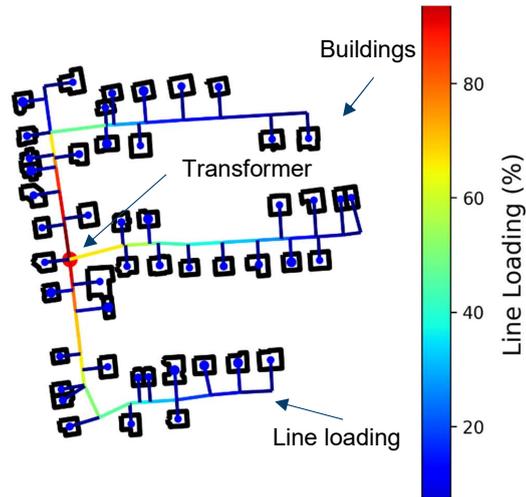


Figure 5-12 A synthetic low-voltage network among 500,000 geo-referenced synthetic networks showing line-loading at a specific time

Until now, the methods and implementation of the geo-referenced synthetic low-voltage networks were discussed. However, to synthesize entire distribution networks, medium-voltage networks should be synthesized. Therefore, the following section looks at the development of geo-referenced synthetic medium-voltage networks.

5.2 Geo-referenced Synthetic Medium-voltage Networks

This section presents the development of synthetic medium-voltage networks for Germany. As discussed in Chapter 1, distribution networks are comprised of two or three voltage levels, depending on the nation. In Germany, it is necessary to develop low-, medium-, and high-voltage levels, to generate synthetic distribution networks. Thus, this section models the medium-voltage level, which is immediately above the low-voltage level.

In principle, medium-voltage networks, like low-voltage networks, consist of nodes and edges. However, the properties of each of these nodes and edges are distinct from low-voltage networks. Typically, medium-voltage networks operate between a range of 1 kV and 30 kV. In these networks, the primary connecting points or nodes are low-voltage network transformers, non-residential, and commercial buildings. In addition, in this voltage level, a high-voltage to medium-voltage transformer (HV/MV) (i.e., medium-voltage transformer) is also a node. In

general, medium-voltage networks have a ring structure and are protected by circuit breakers that automatically disconnect in the event of a fault. In Germany, 85% of the medium-voltage networks operate as open rings [201]. The open ring topology is formed by opening a switch in the ring network. This operation represents the ring network as a radial network. In this thesis, algorithms are developed for generating geo-referenced synthetic medium-voltage networks as a radial topology since the medium-voltage networks operate in an open ring configuration. The following section discusses the data pertaining to the necessary components, such as MV-transformers, LV-networks, non-residential buildings, power lines, etc., in building these networks.

5.2.1 User Input

The connection points or nodes are necessary to estimate geo-referenced synthetic medium-voltage network topologies. First, it is important to collect location information for connection points, such as low-voltage transformers and non-residential buildings. The locations of low-voltage transformers are determined using the geo-referenced synthetic low-voltage network topologies generated in the preceding section. Each of the 500,000 geo-referenced synthetic low-voltage networks has a low-voltage transformer with location and power-specific attributes. Figure 5-13 depicts the locations of low-voltage transformers estimated synthetically in the process of generating low-voltage networks in Germany. It is evident from Figure 5-13 that major cities have a high number of low-voltage transformers installed. This can be attributed to the high density of consumers, which necessitates a higher number of transformers. However, due to their greater number, the exact locations of these transformers cannot be easily identified on the map.

In addition to the low-voltage transformers' locations, the other end-consumers connection points associated with medium-voltage networks are non-residential building locations, excluding garages. Although not all non-residential buildings are connected to medium-voltage networks, this assumption is made in the thesis to include all non-residential buildings in the medium-voltage networks since there is scarce information regarding the type of non-residential building. There are 4,166,532 non-residential buildings (excluding garages) that are being investigated for their inclusion in the process of generating geo-referenced synthetic medium-voltage networks. Since the footprints of non-residential buildings are polygonal geometries, the centroids corresponding to each polygon are used as the connecting points in the algorithms developed.

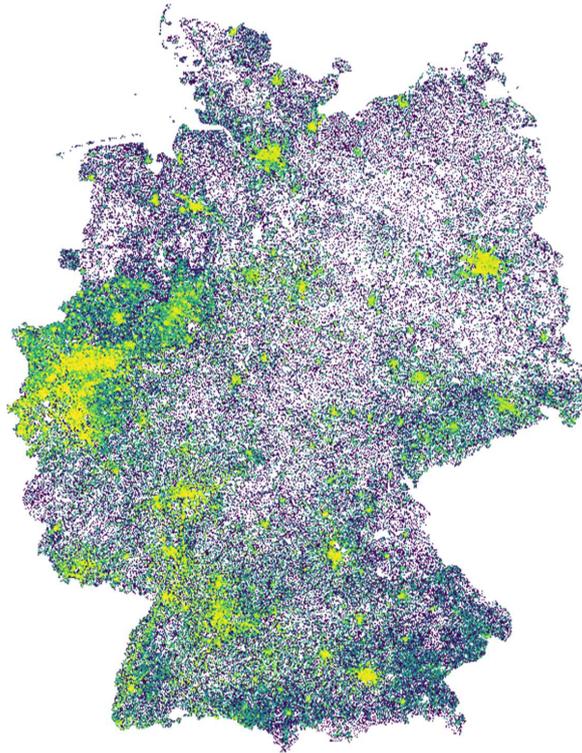


Figure 5-13 Low-voltage transformer locations extracted from geo-referenced synthetic low-voltage networks estimated in the previous section

The location of medium-voltage transformer is the final node required for the estimation of geo-referenced synthetic medium-voltage network topology. The medium-voltage transformers and their locations are significant components in medium-voltage networks as the primary goal of constructing geo-referenced synthetic network topologies is to utilize publicly available data. Therefore, in Chapters 3 and 4 many publicly available datasets are extracted and described. Based on the information from the preceding chapters, it was determined that approximately 4,500 medium-voltage transformers are deployed in Germany. But the OpenStreetMap data analysis identified only 2,595 locations of medium-voltage transformers in OpenStreetMap data (see, Figure 5-14). Figure 5-14 also illustrates that medium-voltage transformers are primarily concentrated in large cities. Nevertheless, 1,905 medium-voltage transformers locations could not be identified in the OpenStreetMap data. These transformer locations are further estimated using a method described in the next section, which utilizes the identified medium-voltage transformer locations.

After determining the locations of all medium-voltage transformers, it is necessary to assign a transformer type to each medium-voltage network. As a result, medium-voltage transformers are derived from the standard types available in Germany. Table 5-5 contains a list of available medium-voltage transformers. The standard medium-voltage transformer types have an

apparent power of ranging from 25 MVA to 63 MVA and an operating voltage between 10 kV and 20 kV [133].

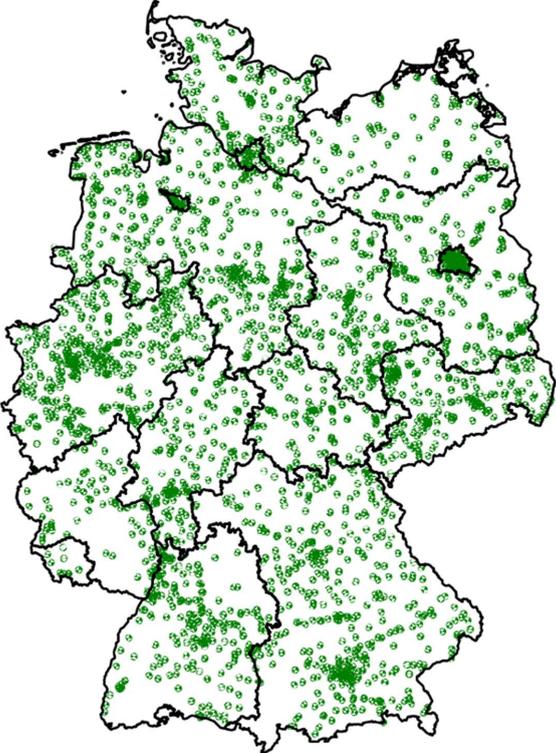


Figure 5-14 Medium-voltage transformer locations retrieved from OpenStreetMap

Table 5-5 Standard medium-voltage transformers available in Germany [Adapted from [133]]

Transformer type	Apparent power	Nominal voltage (High voltage)	Nominal voltage (Low voltage)	Frequency
25 MVA 110/10 kV	25 MVA	110 kV	10 kV	50 Hz
25 MVA 110/20 kV	25 MVA	110 kV	20 kV	50 Hz
40 MVA 110/10 kV	40 MVA	110 kV	10 kV	50 Hz
40 MVA 110/20 kV	40 MVA	110 kV	20 kV	50 Hz
63 MVA 110/10 kV	63 MVA	110 kV	10 kV	50 Hz
63 MVA 110/20 kV	63 MVA	110 kV	20 kV	50 Hz

The final essential component required in the modeling process is the power lines that connect end-consumers and medium-voltage transformers. As mentioned in Chapter 3, 75% of medium-voltage power lines are underground cables that are not visible in OpenStreetMap data. Since underground cables run parallel to the roads, road infrastructure is utilized to model

power lines in medium-voltage networks. However, to represent electrical networks, power line parameters must be assigned to each modeled line from the road infrastructure. During this process, we assumed that all power lines in medium-voltage networks are underground cables to reduce the model's complexity. As a result, commonly available cable types in Germany are included, as indicated in Appendix F.

In addition to network components, power plants within the medium-voltage network are also recognized as connection points (refer to Appendix D). As per the available data, the highest renewable energy feed-in power is supplied at the medium-voltage level. In this context, wind power dominates other technologies at this voltage level, with a capacity of 22.2 GW already installed by end of 2019. Next, 15.1 GW of open filed PV power generation is installed at this voltage level. Therefore, due to the significant feed-in from renewable energy sources and the availability of suitable locations for these installations [200], power plants are also considered as connection points and are integrated into geo-referenced synthetic medium-voltage networks.

In summary, the final user input is demonstrated in Table 5-6 to generate geo-referenced synthetic medium-voltage networks.

Table 5-6 Final user input data for estimating geo-referenced synthetic medium-voltage networks

Connection points	Low-voltage transformer locations extracted from geo-referenced synthetic low-voltage networks (Section 5.1), non-residential buildings predicted in Chapter 4.
Transformers	2595 medium-voltage transformer locations identified from the OpenStreetMap data (Chapter 4), Total medium-voltage substations count [201]
Power lines	Road infrastructure including streets, highways, footpaths, and bike paths
Power plants	Open power system data extracted medium-voltage distributed generator locations and their capacities [200]

Considering input data, the following section discusses the model for estimating geo-referenced synthetic medium-voltage networks.

5.2.2 Modeling Methodology

As with low-voltage network estimation, geo-referenced synthetic medium-voltage network estimation includes two major stages and employs user input. First, the graphical layer is developed by connecting various connection points and network components via road infrastructure. Second, in the electrical layer, several electrical properties are integrated within the graphical network, including transformer types, transformer loads, and line types. By presenting these steps sequentially, the network becomes suitable for performing power flow analysis. Figure 5-15 illustrates the schema relating to the modeling approach.

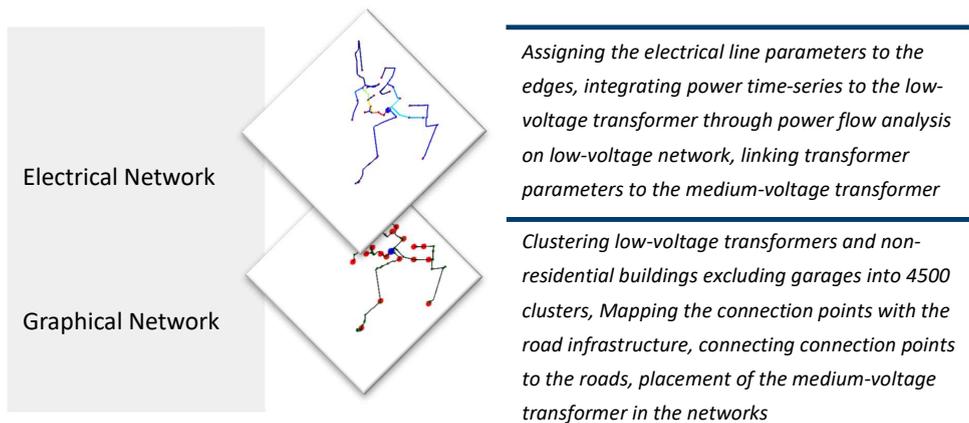


Figure 5-15 Schema for medium-voltage network modeling

5.2.2.1 Graphical Network

Similar to low-voltage network estimation, the first step in establishing a graphical layer is to cluster the connection points by combining low-voltage transformers and non-residential buildings. The previous chapter has discussed numerous unsupervised machine learning techniques such as hierarchical clustering [225], DBSCAN [226], and K-Means[224]. However, K-Means was employed in the development of low-voltage networks due to its advantages over other algorithms. Therefore, the same technique was used to cluster the connection points into 4500 clusters to develop 4500 medium-voltage networks. However, clustering 4500 clusters on connection points is not as complex as clustering approx. 19.7 million residential buildings into 500,000 clusters. Therefore, direct clustering is possible in this case without dividing the networks into administrative districts. As a result, the merged dataset of low-voltage transformers and non-residential buildings is subjected to direct K-Means clustering. On each cluster, applying the series of developed methods produced medium-voltage networks and corresponding medium-voltage transformer.

However, as stated in Chapter 4, 2595 medium-voltage transformer/substation locations have been already geo-referenced in OSM. Incidentally, this represents 58% of all transformers in existence. Thus, this information must be integrated in the clustering process and these 2595 substation locations should help cluster the connection points to 2595 clusters, while the remaining 1905 clusters should be independent of those 2595 clusters. To accomplish this, a clustering technique that uses the identified medium-voltage substation locations as known centroids and the medium-voltage transformer locations that should be forecasted as unknown centroids is devised. However, the final 1905 medium-voltage transformer locations will be determined after creating the graphical network. Algorithm 8 helps cluster the data points with some known and unknown centroids.

Algorithm 8 Clustering with known centroids and predicting unknown centroids

Algorithm 8: Clustering with known centroids and predicting unknown centroids

Input	Number of clusters (K) Data points ($D = \{D_1, D_2, \dots, D_n\}$) Known centroids ($KK = \{KK_1, KK_2, \dots, KK_j\}$) Unknown centroids ($KUK = K - KK$)
Steps	1: Choose number of clusters (K) 2: Initialize random centroids (i.e., unknown centroids ($KUK = \{KUK_1, KUK_2, \dots, KUK_u\}$)) 3: Combine known centroids (KK) and unknown centroids (KUK) 4: Calculate the Euclidean distance ($d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}$) between the data points to the centroids. Where, p, q contains cartesian coordinates 5: Find $argmin$ ($\arg \min_{x \in S} f(x) = \{x \in S : f(s) \geq f(x) \forall s \in S\}$) gives the closest centroid to each data point 6: Update the unknown centroids (KUK) by taking the average of the data points corresponding to the unknown centroids 7: Known centroids (KK) remain same 8: Repeat from step 3, until optimize 9: Finally, each data point will be assigned with the cluster number and corresponding cluster centroid
Return	Data points with an additional characteristic indicating the cluster number $K = \{K_1, K_2, \dots, K_k\}$, and centroids (CK), where, $CK = \{CK_1, CK_2, \dots, CK_k\}$

As stated before, Algorithm 8 was developed to cluster the data points around the existing medium-voltage substations and cluster the remaining data points to rest of the clusters. For a more streamlined understanding of the process see Figure 5-16, where the blue dots represent the data points that need to be clustered in this case. For our use case, some medium-voltage substation locations have been identified. Therefore, in this example, the known centroids, which are the identified medium-voltage substations, are depicted in yellow (X). Now, the data points must be clustered into four clusters with two known centroids (i.e., identified medium-voltage substation locations) and two unknown centroids (i.e., unknown medium-voltage substation locations). First, consider two data points from the existing data as centroids and use K-Means++ [231] strategy to finalize the initial centroids. This strategy ensures that the initial centroids are far away from each other. As a next step, merge the unknown centroids with the known centroids. Here, the initial centroids generated will be from both known and unknown centroids. It is important to note that the clusters are formed by applying pairwise distances and selecting the points with the shortest distance to the cluster centroids. By taking the average of the data points surrounding the unknown centroids, new centroids are yielded. The algorithm's objective is not to modify existing centroids; instead, it updates unknown centroids. Repeat this method until the optimal solution is established. Finally, combine two known centroids and two unknown centroids to produce four clusters from a given datapoint. In the Figure 5-16, four clusters are formed using the given data points without disturbing the original existing centroids.

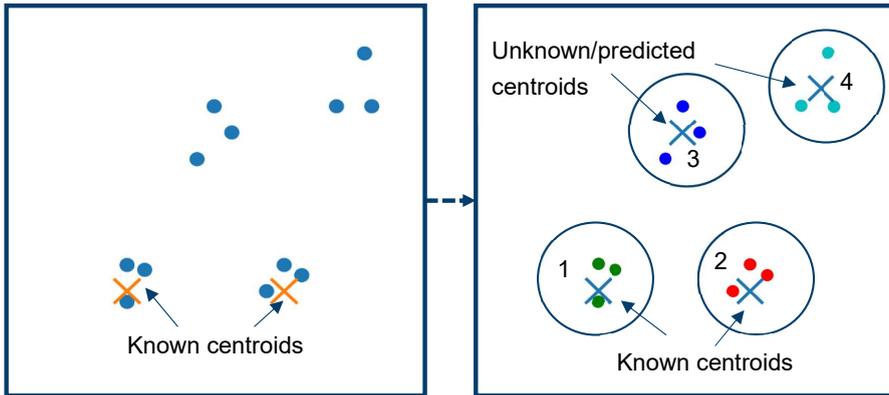


Figure 5-16 Sample implementation of the Algorithm 8

By applying Algorithm 8 on low-voltage transformers and non-residential structures, which are connection points with 2596 known medium-voltage substation locations and 1905 unknown medium-voltage substation locations, 4500 clusters can be generated. Since there is a significant number of data points to cluster, this method is computationally intensive and requires more than two weeks. Furthermore, due to the limitations of the developed method, scalable coding is required. Nevertheless, there is always the option of using scalable K-Means clustering, which helps cluster data from buildings with up to 500,000 networks in the process of low-voltage development. As a result, 4500 clusters (i.e., $K_j \forall j = \{0, 1, \dots, 4499\}$) emerged from the implementation of this clustering algorithm on the non-residential and low-voltage transformer locations, without utilizing existing medium-voltage substation locations.

Henceforth, cluster (K_{4490}) is being explored for future model development to describe the process of generating geo-referenced synthetic medium-voltage networks. After collecting the datapoints in a cluster, several other methods are developed sequentially. However, the next phase in the modeling process is integrating the road infrastructure with the cluster's connection points. The method developed for low-voltage networks is utilized to collect data on the road infrastructure directly adjacent to the connection point. Therefore, Algorithm 3 is used for this purpose, with inputs such as cluster connection points, road infrastructure, and a buffer of 20 kilometers. The buffer of 20 km is considered because the distance between connection points is substantially greater than in low-voltage networks. However, connecting these road infrastructures to the connection points is not straightforward and cannot be accomplished using the algorithm devised for low-voltage networks. This is because the points are separated by kilometers with many lines going through the pair of data points. For instance, in Figure 5-17, multiple lines flow between the nodes once this road infrastructure is gathered near selected connection nodes or connection points. In Figure 5-17, the red dots represent connection points, while the lines represent the road infrastructure retrieved using Algorithm 3.



Figure 5-17 Road infrastructure in close proximity to the connection points, retrieved using Algorithm 3

As a next step, the data points must first be connected to the road infrastructure by using Algorithm 4 that was developed for low-voltage networks. First, this algorithm is fed with an input of connection points and road infrastructure. Then, the algorithm generates additional nodes on the adjacent road infrastructure, establishing new edges between the connection points and the newly produced nodes.

The next step is to tidy up the lines between each pair of nodes to create a radial network. While it is true that medium-voltage networks should be connected in a ring topology, the literature indicates that 85% of medium-voltage networks operate in an open ring configuration with the circuit breaker detached. As a result, our primary objective is to examine a network's functioning topology. Subsequently, in this thesis, medium-voltage networks are generated as radial networks. Thus, connect the end nodes of each two feeders if necessary and create a ring topology as illustrated in Figure 5-18.

In the Figure 5-18, the open-ring network consists of two switches. Each of these two switches are opened to represent four feeders. On the other hand, if these switches are closed, there will be only two feeders with a closed path. Therefore, if there is a requirement to model a ring network, the end connection points of two adjacent feeders will be connected with the new line. This will eventually result in the formation of a ring network topology.

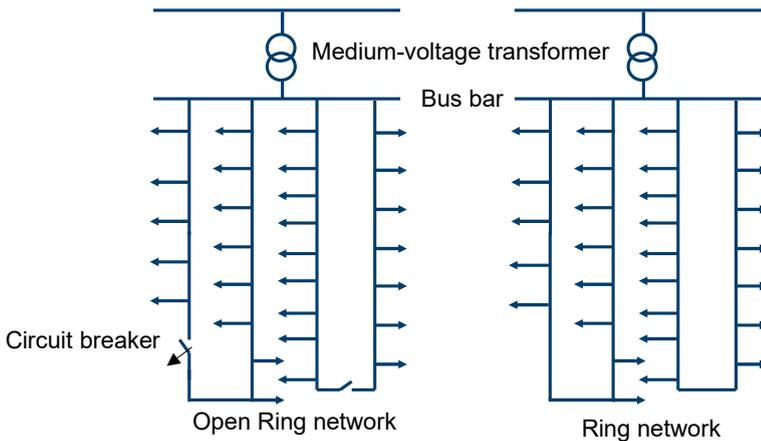


Figure 5-18 Ring topology operating as radial network (open ring) by disconnecting a line using circuit breaker

The next step is to eliminate the multiple unsuitable edges between the nodes, which are illustrated in Figure 5-17. A method was developed to remove the unnecessary edges and connect the nodes with the least possible weight. In this scenario, the length of the roads will be used to calculate the weights. By applying this method to all connection points and road infrastructure, a radial network is formed. Algorithm 9 demonstrates the creation of a radial medium-voltage network.

Algorithm 9 Estimating radial medium-voltage network

Algorithm 9: Estimating radial medium-voltage network	
Input	Nodes (N) Edges (E)
Steps	<ol style="list-style-type: none"> 1: Create a graph with edges ($G(N, K)$) 2: Check for the graph properties 3: Number of subgraphs (SG) 4: <i>while</i> $SG > 1$ <i>do</i> <ul style="list-style-type: none"> Calculate the distance between nodes of one subgraph (SG_i) to another subgraph (SG_{i+1}) nodes Collect the node pair with minimum distance Generate a line joining these node pairs Update the edges with the new edge 5: Merge the edges with same end points into single edge 6: Perform shortest path calculations from one node to all other nodes 7: Apply minimum spanning tree on the shortest path graph to secure shortest path connecting all the nodes 8: This minimum spanning tree application removes unwanted nodes and edges 9: Update the graph with remaining nodes and edges

When Algorithm 9 is applied on the chosen cluster (K_{4490}), it yields the graphical network as depicted in Figure 5-19. The red dots in Figure 5-19 denote the connection points, while the black lines denote the edges connecting the nodes following the roads. Furthermore, Algorithm 7 also incorporates power plants if they are located near the generated medium-voltage graphical network.

The graphical network is missing the node that serves as the transformer, which is undeniably a major element of any electrical network. Therefore, two approaches are designed to integrate the transformer into the graphical network. First, if the network is constructed without considering the locations of existing medium-voltage substations, the algorithm designed for developing synthetic low-voltage networks is used. Thus, Algorithm 6 is employed, which takes the nodes and edges as input and provides additional information about the transformer's location as an output. However, in the second approach, if the networks are clustered based on the position of a known medium-voltage substation, the cluster with the known substation location will receive the exact location of the substation as the transformer location. Moreover, Algorithm 6 will be used for other networks clustered around an unknown transformer location to determine the position of medium-voltage transformers. Finally, Figure 5-19 illustrates the graphical network with the transformer position shown in blue.

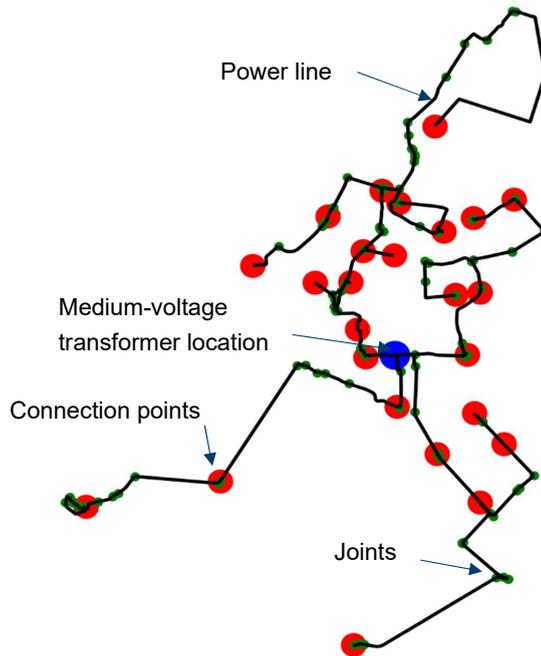


Figure 5-19 Final medium-voltage graphical network for considered cluster (K_{4490})

In conclusion, by executing all of the aforementioned methods sequentially on the 4500 clusters, each containing a few non-residential buildings and low-voltage network transformers, 4500 geo-referenced synthetic medium-voltage graphical networks are generated. Additionally, each graphical network has a node location that acts as a medium-voltage transformer. With these 4500 graphical networks, the following section covers the final stage of geo-referenced synthetic medium-voltage networks, which entails generating electrical network.

5.2.2.2 Electrical Network

The graphical network cannot function as a medium-voltage electrical network without the addition of electrical features. Therefore, electrical characteristics for the graphical nodes and edges should be introduced to convert the graphical network into an electrical network. For the nodes, the electrical demand and generation nodes are the first components. Once the power flow analysis on low-voltage networks is completed, all low-voltage transformers nodes will receive the power profiles straight from the power output at the low-voltage transformer. The non-residential nodes are the next in line to receive the load profiles, wherein the load profiles for non-residential buildings must be assigned. Currently, this thesis does not perform any scenario simulations on medium-voltage networks. Moreover, an accurate non-residential profile for each non-residential building is also absent. As a result, common business load profiles are applied to non-residential buildings to understand and implement the model.

Additionally, generator nodes such as solar, wind, biomass, and others are incorporated into the graphical network. Each of these nodes has information about the type of generation and its capacity. Therefore, based on the generation type and capacity, generation nodes can be allocated normalized profiles for each technology. However, in this example, the methodology is illustrated by solar PV profiles irrespective of the technology because the thesis's primary objective is to construct synthetic network topologies rather than analyze distributed demand's consequences in medium-voltage networks. Nonetheless, Chapters 6 and 7 will go into detail into future distributed generation and demand in low-voltage networks.

A medium-voltage transformer node, which is the final node, must be equipped with the appropriate transformer type for the network. The transformer types can be selected from the standard transformers provided in the previous section (refer to Table 5-5). However, the transformer is selected based on the load and generation in the network. The overall load on a network, on the other hand, is determined by the network's demand and supply. With the peak load in consideration, a transformer will be selected from the list in Table 5-5 at a factor of 1.5 times the peak load. For instance, if the overall network load is x kW, a transformer with a capacity of $x * 1.5$ is selected for the medium-voltage network.

The final step is to assign electrical parameters to the graphical network's edges. As previously stated, 79% of medium-voltage power lines are underground cables. As a result, cable type from the standard cables that are available must be assigned. Appendix F lists the standard cable types that are available. For model representation purposes, each power line is assigned with a power cable of type "NA2XS2Y 3x1x185". After configuring the graphical network with

all the electrical characteristics, the network will be capable of performing power flow simulations. Figure 5-20 illustrates the line loading associated with each projected line segment using power flow analysis.

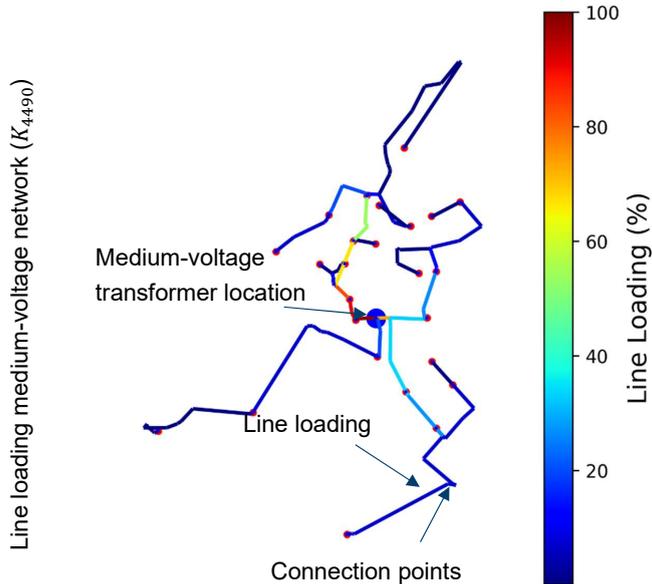


Figure 5-20 A medium-voltage network among 4500 networks showing line-loading at a specific time

Since the development of geo-referenced synthetic medium-voltage networks was discussed, the high-voltage network is the next step in developing electrical distribution grid. The following section will address the development of geo-referenced synthetic high- and extra-high-voltage networks as a single grid.

5.3 Geo-referenced Synthetic High and Extra-high Voltage Networks

In Germany, high-voltage networks form the highest level of a distribution grid. However, they are considered transmission grids in some European countries. In principle, high-voltage networks operate at a nominal voltage of 110 kV and require significant space for substation locations. Interestingly, overhead power lines are used to transport and distribute high-voltage power. As such, the substations and the power lines that maintain and transfer the power are visible in satellite images and are essentially stored in OpenStreetMap data.

This also holds true for the extra-high-voltage level, where the nominal voltage value ranges between 220 kV and 380 kV. Thus, all high-voltage and extra-high-voltage power lines and substations are visible in satellite data. Since high-voltage is classified as distribution and transmission grids in different places, the tagging that corresponds to high-voltage and extra-high-voltage in OpenStreetMap data is highly ambiguous. Therefore, this thesis models both

high-voltage and extra-high-voltage networks as a single grid, although the objective is to develop geo-referenced synthetic distribution networks.

As previously stated, high-voltage and extra-high-voltage level components can be accessed in OpenStreetMap data due to their size and voltage level. This thesis employs models based on available geo-referenced data from OpenStreetMap to build synthetic high-voltage and extra-high-voltage grids. However, there are specific concerns surrounding the OSM data with regards to power tags and network routing information. The power tags are essential for extracting network information. These tags, however, are less accurate since contributors rarely adhere to the OSM provider's recommendations. Additionally, there is no adequate mapping of the electrical properties of the transmission and distribution line. Besides, a substantial amount of data is either missing or incorrectly tagged. For instance, lines or substations corresponding to high voltage are occasionally represented with extra-high-voltage tags. Chapter 0 delved into the various available approaches for generating extra-high voltage grids that are devised to address these concerns.

However, each model has several advantages and disadvantages. In [182], the existing transmission models were compared and the results were presented. After comparing numerous existing models, the GridKit model [180] was chosen for generating extra-high-voltage levels due to its significant topological complexity. Additionally, GridKit [180] was proposed for integration of renewable energy sources into the European power system via EuroPower, which is an inhouse transmission grid model. This model can also generate both high voltage and extra-high voltage grids. Therefore, GridKit is utilized in this thesis to create synthetic high- and extra-high-voltage network topologies. This thesis generates synthetic high-voltage and extra-high-voltage levels, representing the entire power system, from low-voltage distribution grid to extra-high-voltage transmission grid. The following section provides user inputs and the methodology for generating high- and extra-high-voltage networks.

5.3.1 User Input and Methodology

GridKit facilitates the entire process of generating extra-high-voltage and high-voltage in a consecutive manner. The procedure employed when implementing the GridKit model is depicted in Figure 5-21.

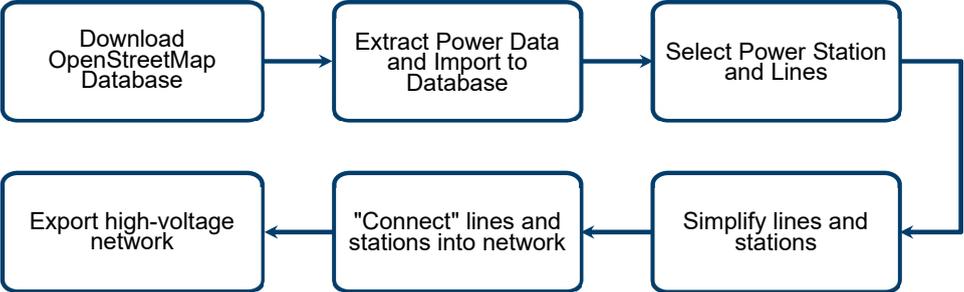


Figure 5-21 Overview of the GridKit process [Adapted from [180]]

As illustrated in Figure 5-21, the procedure begins with the extraction of OpenStreetMap data into databases. This process has already been completed in order to create geo-referenced synthetic low- and medium-voltage networks. As a result, the extracted and filtered data were inputted directly into the GridKit tool with some adjustments in order to extract high-voltage and extra-high voltage power lines and nodes. This adjustment includes the supply of nodes with voltage tags more than or equal to 110 kV. Additionally, all power lines with a voltage tag greater than or equal to 110 kV are considered. Only electricity lines with a frequency of 50 Hz are extracted, leaving 16.7 Hz that corresponds to railway power lines. Feeding these data into the GridKit tool generates the high-voltage and extra-high-voltage network for Germany, as illustrated in Figure 5-22.

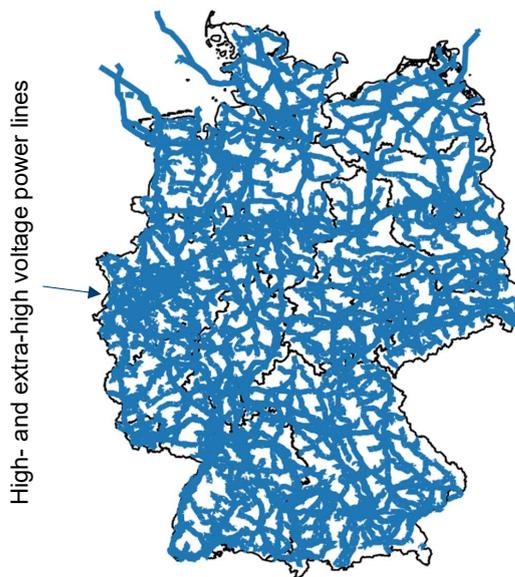


Figure 5-22 Synthetic high- and extra-high voltage grid extracted using GridKit tool [180]

As evident from the data, the complete high- and extra-high-voltage network topology has 9042 node entries. These nodes include the internal nodes that connect two power lines. Additionally, the generated network also contains 12,330 lines that connect all the nodes. A grid topology is nothing more than a graphical network. To perform power flow analysis, more electrical parameters must be incorporated in the graphical network. Therefore, electrical attributes such as transformers and line types were provided with varying resistance, inductance, and capacitance and they are assigned to nodes and edges depending on their tags. The catalog of Germany's various transformer and line types are provided in Table 5-7 and Table 5-8.

Table 5-7 Standard transformer types at high- and extra-high voltage level [183] [201]

Type	Apparent power	Nominal voltage (High voltage)	Nominal voltage (Low voltage)
HV Type-I	300 MVA	380 kV	110 kV
HV Type -II	200 MVA	220 kV	110 kV
EHV Type-I	1000 MVA	380 kV	220 kV

Table 5-8 Standard power lines at high- and extra-high-voltage level [183] [201]

Voltage	R ($\frac{\Omega}{km}$)	X ($\frac{\Omega}{km}$)	L ($\frac{nF}{km}$)	I (kA)
110 kV	0.109	1.2	9.5	
220 kV	0.080	0.32	11.5	1.3
330 kV	0.025	0.25	13.7	2.6

Interestingly, the high- and extra-high-voltage networks provide multiple circuits between nodes. The number of circuits between nodes is calculated by dividing the tags corresponding to the lines (*cables*) traversing between them by three. This is because Germany uses a three-phase ac system [138]. Equation 5-1 illustrates the formula for calculating the number of circuits responsible for connecting two grid locations.

$$N_{circuit} = \frac{Cables_{osm}}{3} \quad 5-1$$

Finally, the development of grid topologies and the assignment of electrical parameters to high-voltage and extra-high-voltage networks helps model a variety of topics, including the integration of future renewable energy sources into power systems, market simulations, and dispatch challenges. Developing high- and extra-high-voltage grid networks completes the synthetic generation of the entire power network.

In conclusion, the synthesizing of the power system began from the low-voltage level. Gradually, the study moved upwards towards the extra-high-voltage level utilizing a bottom-up method as illustrated in Figure 1-3. Following the establishment of the synthetic power networks, the following section will cover the validation of the approach, which is pertinent.

5.4 Validation of Geo-referenced Synthetic Low-voltage Networks

The previous section discussed the process of generating synthetic distribution grids in conjunction with the extra-high-voltage transmission grid. In this section, a methodology and a set of validation criteria are provided for geo-referenced synthetic low-voltage networks. The decision to specifically consider geo-referenced synthetic low-voltage networks was influenced by the fact that the data required to validate them is relatively abundant, and this thesis analyzes future distributed demand and generation in low-voltage distribution grids.

It is important to point out here that geo-referenced synthetic networks are explicitly designed for research applications and are not actually related to any real grids. As such, they are not subject to any confidentiality obligations. However, synthetic networks that resemble and function like real networks are essential. It is necessary to validate geo-referenced networks before they can be used for various applications, they must be validated. Thus, this section focuses on the indices and results from the generated geo-referenced synthetic low-voltage networks for quantitative, operational, and visual validation. This is accomplished through an elaborate comparison against real-world datasets. Hence, three types of validation are conducted for this purpose: statistical validation, operational validation, and real-world geographical visualization. Figure 5-23 provides an overview of the proposed validation measures.

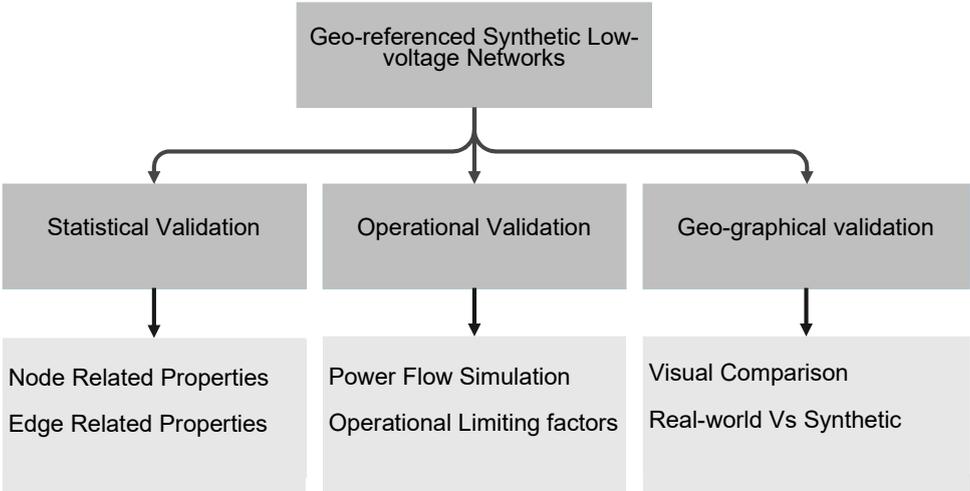


Figure 5-23 An overview of proposed validation measures

5.4.1 Statistical Validation

The statistical validation method compares the data gathered from the generated geo-referenced synthetic networks to the data collected from real-world networks. However, confidentiality obligations and limitations render real network data unavailable. Therefore, statistical validation was conducted using some representative data as well as real-world networks’ data acquired from the literature. Chapter 3 provides the validation metrics, which are included as part of the topological properties. In Section 5.1, various methods were applied to generate 500,000 geo-referenced synthetic low-voltage networks and retrieve statistical information about their topological and electrical properties.

In this context, the first metric examined in statistical validation is the total number of nodes in a network. Figure 5-24 depicts a box plot identifying the number of nodes per network for 500,000 synthetically generated networks. In these established networks, there are anywhere between two and approximately 200 nodes, with 50% of the networks including 40 to 110

nodes. As previously stated, this thesis is entirely dependent on data from the literature due to the lack of access to real data. Džanan et al. [232] analyzed 180 real low-voltage networks in Germany and published the range of total loads for each network type - rural I, rural II, rural III, semi-urban I, semi-urban II, and urban. In Figure 5-25, the range of loads in each network is illustrated in relation to the various locations where the networks are situated.

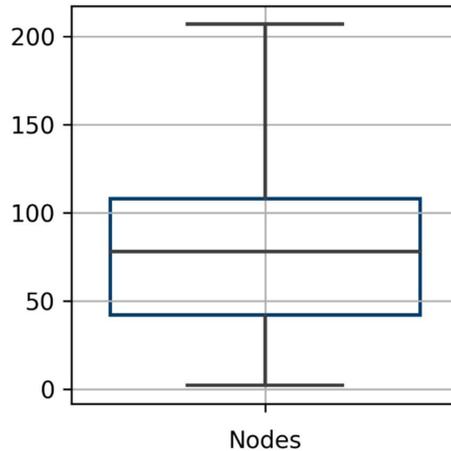


Figure 5-24 A box plot representing the total number of nodes in each network for 500,000 geo-referenced synthetic low voltage networks

Figure 5-25 illustrates that in real-world networks, the number of loads per network ranges from 2 to 200. They are, however, concentrated in specific geographic areas such as rural, semi-urban, and urban areas. Interestingly, the data collected from generated synthetic networks are relatively equivalent to published data [232]. For instance, the range of nodes determined in the geo-referenced synthetic low-voltage distribution networks ranges approximately from 2 to 220. Here, depending on the location, the range is not segregated for each type. Understandably, it is also challenging to segregate 500,000 geo-referenced synthetic networks to rural, semi-urban, and urban types. As illustrated in Figure 5-25, when all real-world network types are considered, the minimum and the maximum of total nodes is approximately 1 and 225, respectively. Therefore, it can be asserted that generated geo-referenced synthetic networks nodes are in the range observed for real-world networks.

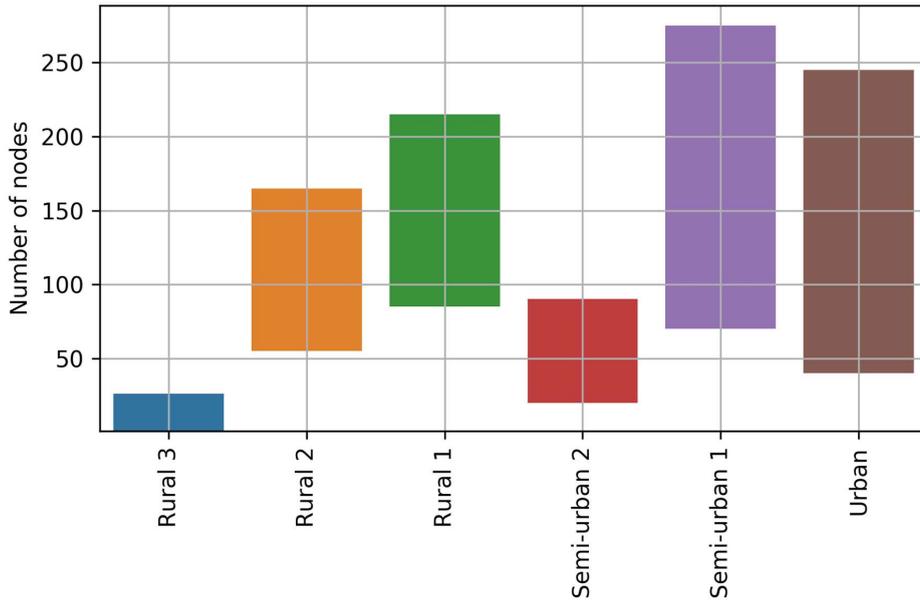


Figure 5-25 Range of total number of nodes of real low-voltage networks for various rural, semi-urban, and urban areas in Germany

Next, the mean node degree is a significant topological parameter that corresponds to the nodes. As indicated in Figure 5-26, the mean node degree is for generated 500,000 geo-referenced synthetic low-voltage networks. As stated in Chapter 3, if the average degree of the nodes exceeds two, the network is assumed to be a mesh network. Additionally, the literature indicates that low-voltage networks are radial in structure and thus, they are modeled accordingly. Thus, in a network, the mean degree of each node should be lower than two. However, as illustrated in Figure 5-26, almost all networks have a mean node degree of less than two. This could be attributed to the fact that some networks have only one node on each feeder, leading the transformer's node degree to dominate the average node degree of the network. Nevertheless, it is clear from Figure 5-26 that all networks are radial in structure, satisfying the average node degree less than two.

Therefore, by comparing these two characteristics - number of nodes and mean node degree - with real networks and statistical details, respectively, it can be concluded that the estimated synthetic networks have node-specific graphical properties comparable to those of real-world networks.

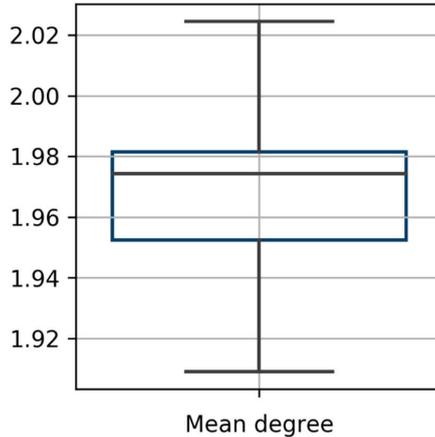


Figure 5-26 A box plot illustrating the mean node degree of 500,000 geo-referenced synthetic low voltage networks

Furthermore, edges/lines are the later component characteristic that must be compared. The total number of edges in the 500,000 low-voltage networks generator is between one and approximately 200 (refer to Figure 5-27). This number is proportional to but one fewer than the number of nodes, which is accurate as per Kerber's typical low-voltage networks [151]. Figure 5-27 depicts the length-related properties of the edges, such as the circuit length of each network, the average shortest path length, and the average length of the edges. Since there is a shortage of data in this context, the average lengths of the networks are compared to those of typical networks. The average edge length of synthetically built networks varies between 2 and 60 meters with 50% of the networks reporting edge lengths ranging from 25 to 42 meters. However, in a typical network, the average line lengths range between 6 and 50 meters. Therefore, the average edge length of the estimated geo-referenced synthetic low-voltage networks is close to the average edge length of typical networks developed using real network parameters. However, due to a lack of topological data for the real-world networks, the average shortest path lengths were not compared. By considering the circuit lengths of each developed network, the total circuit length of geo-referenced synthetic low-voltage networks is compared to statistical data.

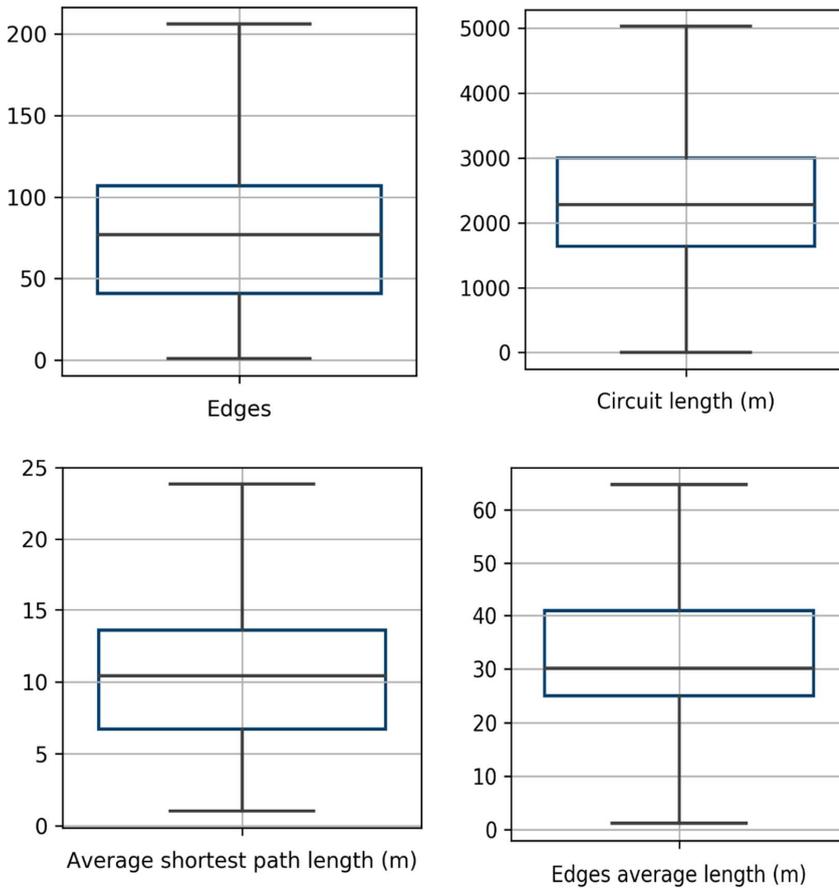


Figure 5-27 Edge related characteristics extracted from 500,000 geo-referenced synthetic low voltage networks

As stated above, when all low-voltage networks are considered, the most critical parameter associated with edge characteristics is the total network length. Benjamin et al. [202] provide the total length of Germany's low-voltage network, which is 1,173,065 km. However, when all 500,000 geo-referenced synthetic low-voltage networks are considered, the final length of the low-voltage network is 1,183,367 km. Consequently, there is a small mismatch of 13,002 km between the officially recorded and estimated lengths. Since the deviation from the actual length is only 0.89%, the predicted networks from a statistical perspective are close to real-world networks. The total length of existing low-voltage networks and estimated networks is depicted in Figure 5-28.

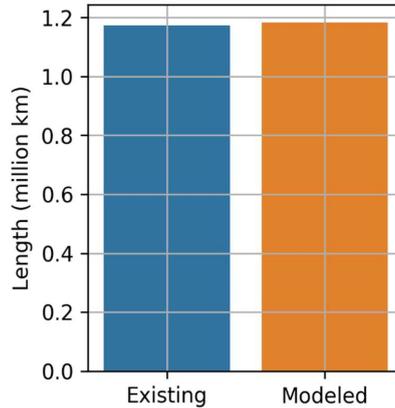


Figure 5-28 Comparison of existing total low-voltage network length for Germany with the total network length of geo-referenced synthetic networks estimated

The next sub-section carries out the operational validation of the networks.

5.4.2 Operational Validation

Operational validation is also considered in this thesis. In this form of validation, an open-source power flow analysis tool is used to determine the power flow convergence and identify the operational limiting factors satisfying grid code constraints. To conduct operational validation, 500,000 geo-referenced synthetic low-voltage networks are simulated using solely residential load profiles assigned to the residential buildings. Future distributed demand and generation, on the other hand, are not considered in this situation. To perform operational validation, networks were simulated using PyPSA (Python for power system analysis [133], an open-source power flow analysis tool).

The simulation of synthetic networks is evaluated for the parameters that should be exhibited by the network. Table 5-9 details the parameters and their associated targets for each network. Once power flow analysis is conducted and the parameters indicated in Table 5-9 are assessed for all 500,000 geo-referenced synthetic networks, 499,608 synthetic networks are converged for power flow analysis. The networks converged under power flow simulation demonstrate satisfactory results for the target set specified in Table 5-9. In fact, only 0.08% of networks do not converge when power flow simulations are performed. A comprehensive analysis of non-converged networks reveals that these networks hold significantly more apartments and that the accumulated peak demand exceeds the capacity of the highest transformer capacity available in the catalog. Nonetheless, non-converged networks can be further clustered to ensure that peak loads remain below the transformer capacity limit as specified in the catalog. Thus, this resolves the non-converged network problem. Additionally, the transformer that is not frequently used in low-voltage networks (i.e., > 630 kVA) can also be incorporated to reduce non-converged networks.

Table 5-9 Power flow parameters and target set for operational validation

Power flow parameter	Target
Convergence	Yes
Minimum voltage magnitude	> 0.95 with out voltage regulated distributed transformer > 0.90 with voltage regulated distributed transformer
Maximum voltage magnitude	< 1.05 with out voltage regulated distributed transformer < 1.10 with voltage regulated distributed transformer
Under voltage count	0 (Count at each time snapshot for total considered snapshots)
Over voltage count	0 (Count at each time snapshot for total considered snapshots)
Line loading	< 100%
Transformer loading	< 100%
Component loading time snapshots	< 0

Following the statistical and operational validation, comparing the real-world network geographically is the next essential validation. The following section dives into this aspect.

5.4.3 Geographical Validation

The final validation metric is a graphical comparison of the real-world network, and the geo-referenced synthetic network generated using the model. In this case, an internal validation was performed on an existing network that was obtained from a distribution system operator (not shown due to the non-disclosure agreement). From the 500,000 geo-referenced low-voltage network, the network synthetically generated at the same position as the existing network was selected when comparing the geo-referenced network to the real network visualizing geographically. Comparing the two networks geographically reveals that 91% of the buildings covered in the real networks correspond to the developed geo-referenced synthetic network. Additionally, the geo-referenced synthetic network’s power lines are almost identical to those seen in the real-world network. However, further comparison reveals that the transformer placement is not identical.

Nevertheless, with the algorithms proposed in this thesis, high-accurate low-voltage networks can be produced, given the buildings in the low-voltage networks by the distribution system operator. Based on the validation measures conducted, it can be concluded that the developed geo-referenced synthetic low-voltage networks can imitate real-world low-voltage networks.

5.5 Summary

After analyzing the data required for estimating geo-referenced synthetic distribution networks, this chapter addresses the following question: *“How are geo-referenced synthetic distribution network topologies estimated, what are the algorithms utilized, and how are they developed?”* The focus of this chapter was to develop methods for estimating geo-referenced synthetic network topologies. The entire process of creating geo-referenced synthetic distribution networks followed a bottom-up approach, beginning with low-voltage networks and progressing to high- and extra-high-voltage transmission grid. Low-voltage networks are developed first by clustering classified residential buildings into the specified number of low-voltage networks (i.e., 500,000). Here, scalable K-Means clustering algorithm was used due to the high volume and variety of the dataset. After clustering the buildings, the road infrastructure was extracted near the clustered residential buildings. All the edges (i.e., road infrastructure) and nodes (i.e., residential buildings) were connected to form radial networks using several developed methods. Additionally, the transformer location that was identified through the developed method validated that the transformer is proximal to all nodes in the network. Finally, electrical parameters such as loads and generators are assigned to each node, and low-voltage transformer type to transformer node. In addition, power line parameters were assigned to all the edges in the networks. This way, graphical networks were transformed to geo-referenced synthetic low-voltage electrical networks. Following that, 4,500 geo-referenced synthetic medium-voltage networks were estimated. To accomplish this, low-voltage network transformers and non-residential buildings are clustered to 4,500 medium-voltage networks. Then, using a series of developed methods, graphical networks were generated with a node indicating the location of the medium-voltage transformer. When all the electrical parameters corresponding to the nodes and edges were applied, the developed graphical networks were transformed into geo-referenced synthetic medium-voltage electrical networks. Finally, high- and extra-high-voltage were generated as single grid utilizing GridKit tool. This helped establish geo-referenced synthetic distribution networks and extra-high-voltage transmission network representing the entire power grid. Finally, in order to validate the established geo-referenced synthetic low-voltage networks, several validation measures such as statistical, operational, and geo-graphical visualization validations were undertaken in this chapter.

Key messages:

- ✚ A bottom-up approach is used to construct geo-referenced synthetic distribution networks.
- ✚ Geo-referenced synthetic low-voltage networks were generated by developing several algorithms such as those clustering the buildings to match low-voltage networks, clustering the buildings by limiting the cumulative peak load, collecting road infrastructure from OSM data, connecting nodes and edges, constructing a single graph from subgraphs, and identifying transformer locations.
- ✚ The development of geo-referenced synthetic medium-voltage networks employs algorithms developed for low-voltage networks, as well as clustering with known centroids and predicting unknown centroids, and estimating radial medium-voltage networks.
- ✚ GridKit is used with additional assumptions to create a geo-referenced synthetic high- and extra-high-voltage network.
- ✚ According to statistical validation, data generated from 500,000 geo-referenced synthetic low-voltage networks are highly comparable to published data. Additionally, comparing the overall circuit length of developed synthetic networks to the length of existing networks reveals a percent inaccuracy of only 0.89%.
- ✚ 499,608 geo-referenced synthetic low-voltage networks are converged during operational validation, while only 0.08% of networks do not converge when subjected to power flow simulations.
- ✚ Validation using geographical visualization demonstrates that the networks generated are nearly identical.

6 Distribution Grid Model and Scenario Simulations

What are the scenarios defined for future distributed demand and generation integration into the distribution networks?

How can the effects of distributed demand and generation in the distribution networks be quantified?

Topics covered:

- ✚ Distribution grid model development for integrating future distributed demand and generation in the distribution networks
- ✚ Development of base case scenario in line with the objective of 95% CO₂ emission reduction by 2050
- ✚ Development of additional scenarios for sensitivity analysis
- ✚ Demonstration of developed power flow model on an exemplary network from 500,000 geo-referenced synthetic networks developed

The primary objective of the thesis is to analyze the effects of future distributed demand and generation in the distribution networks. In line with this goal, Chapter 0 suggests the development of geo-referenced synthetic distribution networks. Accordingly, Chapter 5 estimated geo-referenced synthetic distribution networks using information derived from Chapters 3 and 4. In this chapter, a distribution grid model is introduced to holistically capture the effects of future distributed demand and generation in the distribution networks. In this context, several scenarios are proposed and developed with varying levels of rooftop solar PV, battery electrical vehicles, and heat pumps penetration into the low-voltage distribution networks. Here, low-voltage distribution networks are considered due to their ability to accommodate the majority of distributed demand and generation units. Subsequently, for each of these scenarios, the distribution grid model will be used to identify Operational Limiting Factors (OLFs) violations in the low-voltage networks. The block diagram representation of the distribution grid model is illustrated in Figure 6-1.

The distribution grid model leverages data that were defined and retrieved in previous chapters. For instance, Chapter 4 provides the demand, which include residential, battery electric vehicles, and heat pumps, and generation, which is rooftop solar PV, time-series profiles with a 15-minute resolution. These time-series profiles will be provided as an input to the geo-referenced synthetic networks. In Chapter 5, the geo-referenced synthetic low-voltage networks considered are generated, which includes 500,000 geo-referenced synthetic low-voltage networks. As discussed in Chapter 0, the deterministic method was chosen, to perform power flow analysis due to the large number of low-voltage networks, since performing power flow simulations on these networks is time consuming. In this chapter, significant information pertaining to the scenarios that define the future integration of distributed demand and

generation in low-voltage distribution grids will be explored. The scenarios illustrate the various forecasts in installed capacity for distributed demand and generation.

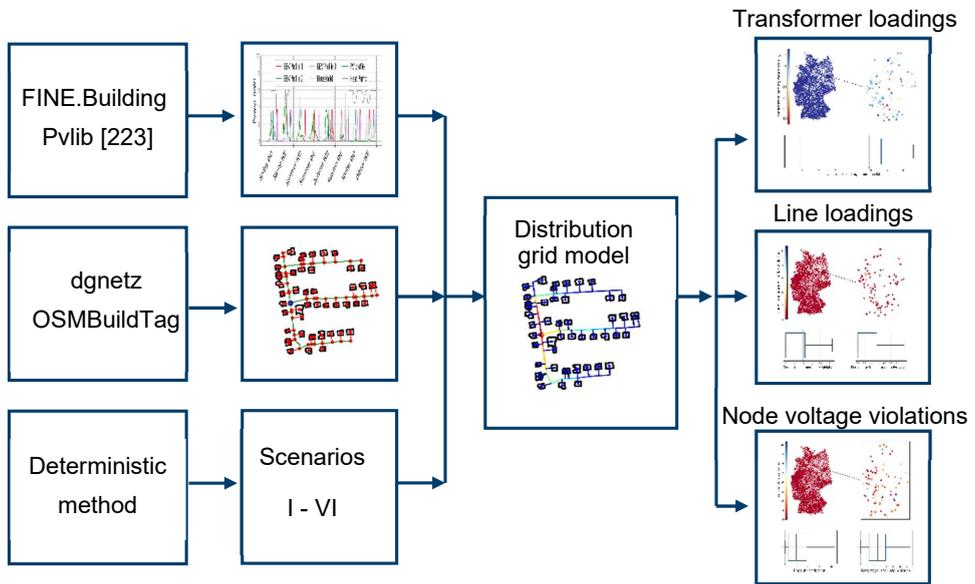


Figure 6-1 Block diagram representation of the distribution grid model

When the given information is applied, the distribution grid model provides information regarding operational limiting factors such as transformer violations, line violations, and voltage violations. Additionally, it also aids in the decision-making process for improved demand and generation integration. Table 2-3 provides the limitations for transformer and line loading violations whereas Table 2-2 defines the limitations for voltage alleged violations.

Having discussed the key parameters of the distribution grid model, the next section explores the suggested scenarios for future distributed demand and generation integration in distribution grids.

6.1 Scenarios

In this study, scenarios are developed to ascertain the trends of the influencing effects on network vulnerability through a period of time. This includes, in particular, the growth of installed renewable energy generation capacity and distributed demand over time. This thesis places an emphasis on distribution networks. The most relevant distributed generation technology is rooftop Photovoltaic (PV), while distributed demand technologies include Heat Pumps (HPs) and Battery Electric Vehicles (BEVs), which are all included in the development of scenarios.

As mentioned in Chapter 1, by 2050, governments aim to reduce GHG emissions by 80% to 95% relative to 1990 levels [3]. These targets presented here are as of 2019. In this context,

the energy system model named NESTOR [233] analyzes transition pathways for the German energy system to reduce CO₂ emissions by 95% by 2050. When the model was subjected to an optimization process, it resulted in a diverse energy mix with renewable energy sources and demand. However, since this thesis primarily deals with the distributed demand and generation in the distribution networks, the installed capacities of rooftop PV, battery electric vehicles, and heat pumps will be analyzed in terms of their impact on the low-voltage distribution networks.

According to Robinius et al., [233], by 2050, the energy system should include 62.7 GW of rooftop solar PV, 26% Battery Electric Vehicles, and 11% Plug-In Hybrid Electric Vehicles, with heat pumps heating 82.9% of total building area. While this mix corresponds to distributed demand and generation, other technologies are also included in addition to these technologies to facilitate a 95% reduction in carbon dioxide emissions. However, it is worth noting that operational challenges for the networks will emerge when this additional demand and generation are introduced into the distribution networks.

To quantify the potential for operational limiting factor violations in low-voltage networks as per the national scenario of Robinius et al., [233] is considered as the base case scenario for integrating distributed demand and generation in distribution grids. Initially, the base case scenario is constructed using data regarding the distributed demand and generation technology that is acquired from [233] (refer, Table 6-1).

Table 6-1 Distributed demand and generation mix in the energy systems by 2050 in Germany

Technology	2050
Solar rooftop PV	62.7 GW Installed capacity
Battery electric vehicles	26% of total vehicles
Plug-In Hybrid electric vehicles	11% of total vehicles
Total vehicles	42.7 million
Heat pump	82.9% of building living space

Note that all the information pertaining to different technologies is in different units. As discussed in the preceding chapters, the geo-referenced synthetic low-voltage networks contain information covering the buildings and their tags. Therefore, metadata on the distributed demand and generation mix in the energy system in 2050 should be elucidated to buildings. This must then be used as a base case scenario for conducting simulations to ascertain the effects of distributed demand and generation in the low-voltage networks.

In this context, Germany should have 62.7 GW of solar rooftop PV by 2050, according to Robinius et al., [233]. As per our research, Germany currently has a total of 19,747,802 residential buildings. Thus, when residential buildings are considered, 62.7 GW of solar rooftop PV should be converted into the percentage penetration of solar PV on all residential buildings. By 2050, with 62.7 GW of solar rooftop photovoltaic (PV), 19,747,802 residential buildings, and a typical rooftop installation of 6 kW, solar rooftop PV should be deployed on 53% of buildings.

Similarly, in terms of electrical vehicle types, there are varying levels of penetration that includes battery electric vehicles and plug-in hybrid electric vehicles. To simplify calculations, this thesis treats all electric vehicles as battery electric vehicles. Based on this assumption, battery electric vehicle penetration was estimated to make up 37% of total vehicles by 2050. However, low-voltage networks incorporate buildings, and the given percentage penetration can be converted to pertain to buildings by equating one vehicle to one building. Based on this assumption and the number of buildings, it is recommended that 80% of buildings be equipped with battery electric vehicles by 2050.

As a next step, heat pump (HP) percentage penetration was converted to a percentage of dwellings. This calculation was made based on the assumption that the area of residential buildings is proportional to the number of households. However, based on the hypothesis, 83% of dwellings should be equipped with a heat pump by 2050. Therefore, by transforming the various combinations of distributed generation and demand in the energy system by 2050 to the percentage of buildings, a base case scenario is established for future distributed demand and generation. Figure 6-2 illustrates the percentage penetration of distributed demand and generation, which include battery electric vehicles and heat pumps, and PV in low-voltage networks as a percentage of total buildings.

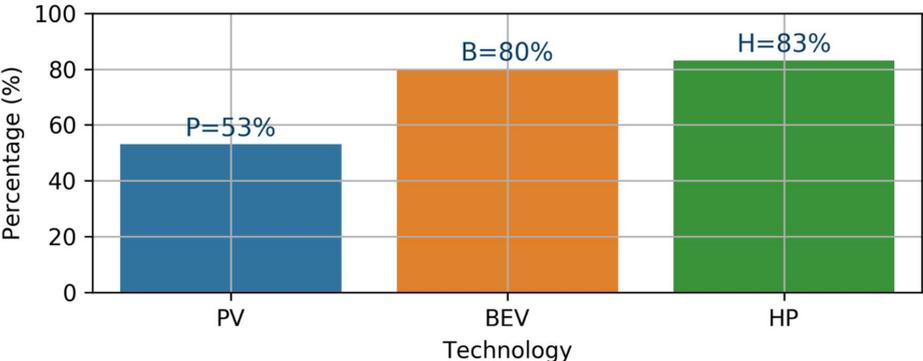


Figure 6-2 Percentage of future distributed demand and generation in distribution grids by 2050; A base case scenario is represented as PBH; where P, B, H corresponds to PV, BEV, and HP respectively⁹

Once the base case scenario was determined, six additional scenarios were established to account for any possible variation from the expected percentage penetration by 2050. In order to develop an additional six scenarios, initial considerations should first be made. Henceforth, rooftop PV base case penetration, battery electric vehicle base case penetration, and heat pump base case penetration are referred as P (53%), B (80%), and H (83%), respectively. Similarly, the Up arrow (↑) represents a 10% increase, and the down arrow (↓) represents a 10% decrease. Therefore, if there is a representation of (P↑), it indicates a 10% increase to

⁹ P: Photovoltaics (PV); B: Battery Electric Vehicles (BEV); H: Heat Pumps (HP)

the base case PV penetration. For instance, if base case PV penetration is 53%, then (P↑) indicates a 10% increase in PV penetration, resulting in 58.3% PV.

Based on this consideration, six scenarios have been developed. In the first three scenarios, one technology's percentage penetration is fixed as the base scenario, while the percentage penetration of the other two technologies is adjusted by ±10% interchangeably. For instance, in Scenario 1 (S1), solar PV is fixed at the base scenario percentage penetration (i.e., P= 53%), while the other two technologies are adjusted according to Table 6-2 to generate four unique cases. The four scenario cases for scenario 1 are named as PB↑H↑¹⁰, PB↑H↓¹¹, PB↓H↑, and PB↓H↓. Here, the percentage penetration for each technology is indicated directly in the name. If the scenario case name is represented as PB↑H↑, it represents the base case PV penetration of 53%, battery electric vehicles penetration increased by 10% (i.e., 88%), and heat pump penetration increased by 10% (91.3%). Similarly, for scenario case with name as PB↑H↓ represent PV penetration of 53%, battery electric vehicles penetration increased by 10% (i.e., 88%), and heat pump penetration decreased by 10% (74.7%).

Similar to Scenario 1, Scenario 2 (S2) and Scenario 3 (S3) are considered, each with four scenario cases and distinct names for the scenario cases. Thus, S2 scenario cases are P↑BH↑¹², P↑BH↓, P↓BH↑, and P↓BH↓. Likewise, S3 scenario cases are P↑B↑H¹³, P↑B↓H, P↓B↑H, and P↓B↓H. The scenario cases for the first three scenarios result in twelve scenario cases with varying percentage penetration for PV, battery electric vehicles, and heat pump technologies. Moreover, each of the scenario names represent the percentage penetration of each technology as mentioned previously. Table 6-2 illustrates the combinations for the first three scenarios.

Table 6-2 Consideration of Scenario 1 (S1), 2 (S2), and 3 (S3) with four cases

Scenario	Scenario 1 (S1)			Scenario	Scenario 2 (S2)			Scenario	Scenario 3 (S3)		
Name	PV	BEV	HP	Name	PV	BEV	HP	Name	PV	BEV	HP
PB↑H↑	P	B↑	H↑	P↑BH↑	P↑	B	H↑	P↑B↑H	P↑	B↑	H
PB↑H↓	P	B↑	H↓	P↑BH↓	P↑	B	H↓	P↑B↓H	P↑	B↓	H
PB↓H↑	P	B↓	H↑	P↓BH↑	P↓	B	H↑	P↓B↑H	P↓	B↑	H
PB↓H↓	P	B↓	H↓	P↓BH↓	P↓	B	H↓	P↓B↓H	P↓	B↓	H

Figure 6-3 depicts each scenario for the percentage penetration of each technology by applying the combinations listed in Table 6-2. As represented in Figure 6-3, each scenario

¹⁰ PB↑H↑: PV = 53%, BEV = 88%, HP = 91.3%
¹¹ PB↑H↓: PV = 53%, BEV = 88%, HP = 74.7%
¹² P↑BH↑: PV = 58.3%, BEV = 80%, HP = 91.3%
¹³ P↑B↑H: PV = 58.3%, BEV = 88%, HP = 83%

case has a unique name that corresponds to a different percentage penetration of each technology into the low-voltage distribution networks.

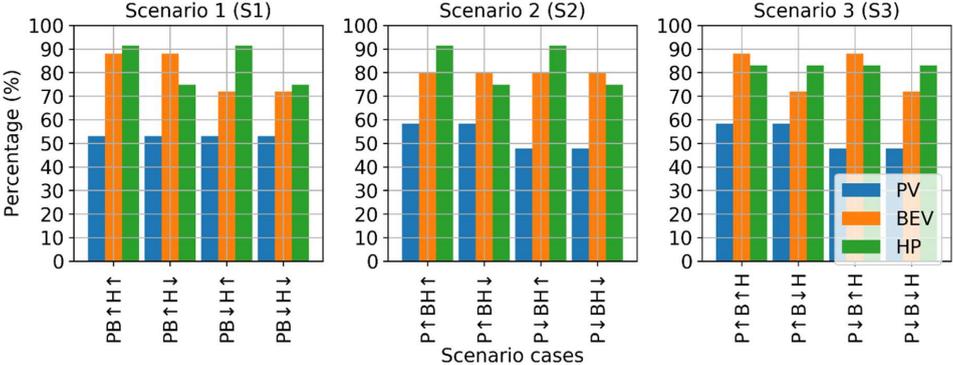


Figure 6-3 Percentage penetration of PV, battery electric vehicles, and heat pumps for Scenarios 1 (S1), 2 (S2) and 3 (S3)

In the first three scenarios, only one technology is fixed to the base case scenario, two technologies are fixed to the basic case scenario and only one technology changes $\pm 10\%$ in the other three scenarios. These variations are depicted in Table 6-3. That is, for Scenarios 4 (S4), 5 (S5), and 6 (S6), two technologies are fixed to base case percentage penetration and one technology from PV, battery electric vehicles, and heat pumps are varied sequentially for three scenarios. For example, in Scenario 4 (S4), battery electric vehicles and heat pumps are fixed to base case percentage, and PV is varied by $+10\%$ in one scenario case and -10% in other scenario case creating two unique scenario names: P↑BH¹⁴, and P↓BH¹⁵, respectively. Similarly, for Scenario 5 (S5) and Scenario 6 (S6), two scenario cases are considered for each, which create PB↑H, PB↓H, PBH↑, and PBH↓, respectively. As mentioned previously, the scenario cases have a unique name and represent the percentage penetration for each technology.

Table 6-3 illustrates the varying penetration of technologies derived for Scenarios 4 (S4), 5 (S5), and 6 (S6).

Table 6-3 Consideration of Scenario 4 (S4), 5 (S5), and 6 (S6) with two cases

Scenario	Scenario 4 (S4)			Scenario	Scenario 5 (S5)			Scenario	Scenario 6 (S6)		
Name	PV	BEV	HP	Name	PV	BEV	HP	Name	PV	BEV	HP
P↑BH	P↑	B	H	PB↑H	P	B↑	H	PBH↑	P	B	H↑
P↓BH	P↓	B	H	PB↓H	P	B↓	H	PBH↓	P	B	H↓

¹⁴ P↑BH: PV = 58.3%, BEV = 80%, HP = 83%
¹⁵ P↓BH: PV = 47.7%, BEV = 80%, HP = 83%

Figure 6-4 illustrates the percentage penetration of PV, battery electric vehicles, and heat pumps for scenarios S4, S5, and S6, with two scenario cases.

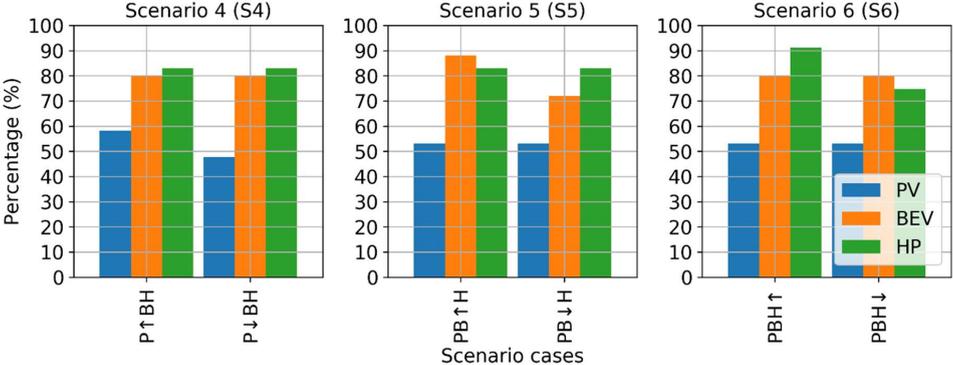


Figure 6-4 Percentage penetration of PV, battery electric vehicles, and heat pumps for Scenarios 4 (S4), 5 (S5) and 6 (S6)

In conclusion, 19 scenario cases, including the base case, were developed to evaluate the future distributed demand and generation in the geo-referenced synthetic low-voltage distribution networks.

After six scenarios were defined with four cases for scenarios S1, S2, and S3, and two cases were defined for scenarios S4, S5, and S6 along with the base scenario, a power flow analysis was performed on each geo-referenced synthetic low-voltage network. For each of the 500,000 synthetic low-voltage networks, a total of 19 simulations each were executed with a scenario case to obtain information about operational limiting factor violations, including transformer loadings, line loadings, and node voltage limit violations. The following section will look at the simulation procedure and results for one of the estimated 500,000 geo-referenced synthetic low-voltage networks.

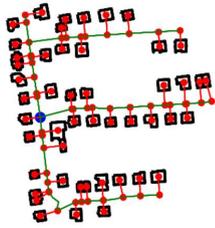
6.2 Methodology Demonstration on an Exemplary Geo-referenced Synthetic Network

As mentioned, each of the 500,000 geo-referenced synthetic low-voltage networks are simulated with 19 scenario cases with different percentage penetration for PV, battery electric vehicles, and heat pumps in terms of their integration into low-voltage networks. However, to better understand the simulation procedure and the results generated by each network, an elaborate case study is conducted on one of the 500,000 geo-referenced network topologies generated in the preceding chapter.

The geo-referenced synthetic network topology of the network (N_{23988}) was examined. As previously stated, the distribution grid model is demonstrated here to guide the reader through the results. The distribution grid model is provided with certain inputs including load and

generation profiles, synthetic networks, analysis method, and a tool for conducting power flow analysis. Table 6-4 contains the data used in the distribution grid model for an exemplary network.

Table 6-4 Parameters used to apply the distribution grid model to an exemplary network

Input parameter		Value
Network (N_{23988}) 	Total nodes	96
	Total edges	95
	Load nodes	48
	Transformer	630 kVA, 20/0.4 kV
	Total circuit length	2.4 km
Residential load	Chapter 4	Single family houses (1 Household), multi-family houses (five households), and apartments (13 households) Each household receive a load profile from 1-person, 2-person, 3-person, 4-person, and 5-person households as per the probability of 0.42, 0.33, 0.11, 0.1, and 0.04, respectively
Rooftop solar PV	Chapter 4	Generation profiles assigned to the randomly selected buildings according to the location and percentage penetration
Battery electric vehicle	Chapter 4	Battery electric vehicle randomly selected from given clusters and assigned to buildings according to the percentage penetration
Heat pump	Chapter 4	Heat pump profiles are assigned to the dwellings according to the percentage penetration
Power factors	$\cos(\varphi)$	0.93 for individual loads and 1.0 for distributed generation [154]
Power flow analysis	PyPSA	Version 0.13.2

Algorithm 10 was developed to extract information from all nodes and transformers and study the effects of distributed demand and generation in the geo-referenced synthetic low-voltage networks. The inputs used to employ Algorithm 10 are demonstrated in Table 6-4.

Algorithm 10 Flow algorithm for extracting information by performing power flow analysis

Algorithm 10: Flow algorithm for extracting information by performing power flow analysis	
Input	Table 6-4
Steps	<p>1: Select the network</p> <p>2: <i>for</i> $i = 1, \dots, 19$ <i>do</i> (19 scenarios developed in previous section)</p> <p style="padding-left: 40px;">Randomly assign the household profiles of one-person, two-person, three-person, four-person, and five-person households to each building according to the building type and with a probability of 0.42, 0.33, 0.11, 0.1, and 0.04, respectively.</p> <p style="padding-left: 40px;">According to the scenario, take the percentage of buildings that holds PV, BEV, and HP.</p> <p style="padding-left: 40px;">Assign the load and generation profiles as per the recommendation of the previous step</p> <p style="padding-left: 40px;">Perform sequential non-linear power flow analysis</p> <p style="padding-left: 40px;">Record the voltage magnitude at all the nodes, time-series power at the transformer, and lines.</p> <p style="padding-left: 40px;">Calculate the operational limiting factors that include transformer minimum power, transformer maximum power, transformer loading, reverse power flow, line loading, over voltage violations, and under voltage violations.</p>
Return	Operational limiting factors for the network for all scenarios considered.

Once the distribution grid model received inputs, non-linear power flow analysis was performed for each time step for all 768-time snapshots corresponding to weekdays (WD) and weekends (WE) during the Spring, Summer, Autumn, and Winter seasons. In the model, power flow analysis was performed using the Newton-Raphson method and after its convergence, power flow simulation results, time series of voltage magnitude at all nodes, active and reactive power at the low-voltage transformer and at each line were provided.

However, using these time-series results obtained for the base case scenario (PBH), several operational limiting factors violations are studied for each network, such as transformer loading violations, line loadings, reverse power flow violations, under voltage violations, and over voltage violations. As discussed in Chapter 0, if the transformer's load exceeds predefined limit, it is referred to as overloading. For the network (N_{23988}) considered in this exemplary simulation, Figure 6-5 illustrates the transformer loading. It is evident from the figure that the overloading occurred for a maximum of four-time instances during the operation of 768-time snapshots, wherein each time snapshot is of 15-minute. Furthermore, transformer overloading was observed during the winter, when heat pumps operate at high rate. Based on the number of times the transformer is overloaded, a decision will be made. Chapter 7 will derive conclusions based on the transformer loading by studying the 500,000 geo-referenced synthetic low-voltage networks.

Figure 6-5 illustrates that certain time periods exhibit negative loading on the transformer, which is reverse power flows. Here, for at least 25 time instances the power is fed into the

medium-voltage network, primarily in summer and autumn, when rooftop solar PV generates the most energy. However, the frequency with which reverse power flows occur will be considered in Chapter 7, and essential conclusions will be drawn.

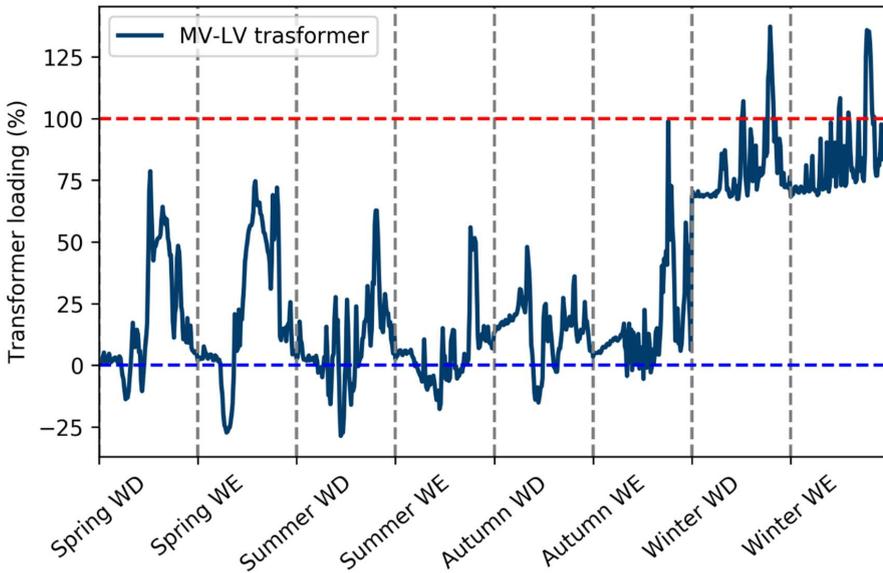


Figure 6-5 Transformer loading of the geo-referenced synthetic low-voltage network (N_{23988})

Furthermore, on each line, the line loading will be considered using active power at the line and its nominal power. Similar to transformer loading, the line is considered to be overloaded in case the line's capacity limit exceeds. Figure 6-6 illustrates the line loading on each line for a specific time instance on the network that is considered. As illustrated in Figure 6-6, lines located near transformers are subjected to greater loading than lines located further away from the transformer. Chapter 7 will examine the time instances and number of lines that are overloaded for all 500,000 geo-referenced synthetic networks.

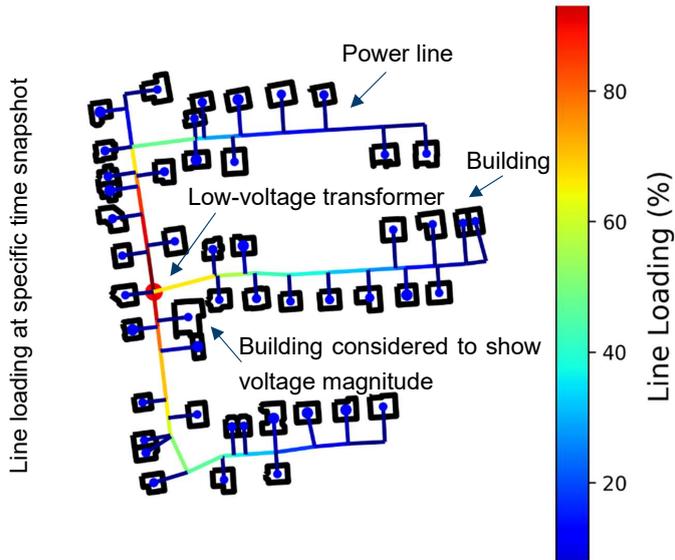


Figure 6-6 Line loading on each line for a specific time instance of simulated geo-referenced synthetic low-voltage network (N_{23988})

Finally, when the voltage magnitude at each bus in the network is analyzed, overvoltage and undervoltage violations will be identified on each node. For instance, Figure 6-7 illustrates the voltage magnitude of a random bus in the simulated network. As discussed in Chapter 0, if there is no decoupling between medium- and low-voltage networks, the node voltage should be between 1.05 and 0.95 per-unit. However, if low-voltage network contains a voltage regulating distribution transformer, the voltage at any node should fall between 1.1 and 0.9 per-unit. As demonstrated in Figure 6-7, the node is under voltage violation in more than two-time steps. However, these two instances are not considered as voltage violations if a voltage regulating distribution transformer is installed in the low-voltage network. Nevertheless, Chapter 7 will draw conclusions by identifying the nodes and the total time during which each node is under violations in a network for all 500,000 networks.

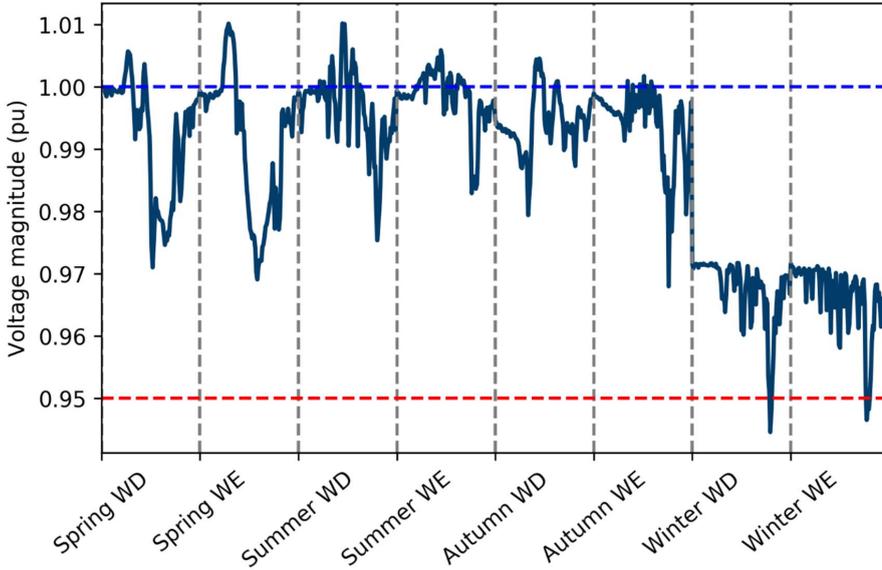


Figure 6-7 Voltage magnitude at a random bus in the simulated geo-referenced synthetic low-voltage network (N_{23988})

To summarize, the results presented here are intended to demonstrate the model's performance; thus, the simulations are not detailed. Nonetheless, the distribution grid model is applied to all 500,000 geo-referenced synthetic low-voltage distribution networks, simulating 19 different scenarios. Based on the outcomes of the simulations that are conducted, the following chapter will discuss the results, and the implications associated with future distributed demand and generation in low-voltage distribution networks.

6.3 Summary

Chapter 5 developed geo-referenced synthetic distribution network topologies in order to study the vulnerabilities of distribution networks by the integration of distributed demand and generation. In this process, several methods were employed along with data introduced in previous chapters. This chapter addresses the question "*What are the scenarios defined for future distributed demand and generation integration into the distribution networks?*" in order to address the thesis's primary research question. This question was answered by considering a base case scenario using installed capacities from a national scenario that optimized the energy mix to achieve 95% reduction in CO₂ emissions by 2050 (Disclaimer: Targets as of 2019). The optimization process for this study considered a variety of energy technologies. For simplification, solar rooftop PV, battery electric vehicles, and heat pumps are considered in this thesis due to their substantial presence in distribution networks. Along with the base case scenario, six additional scenarios with 18 scenario cases were developed to determine the possibility of capacity deviations by 2050. After establishing the scenarios, this chapter attempted to address the following question: "*How can the effects of distributed demand and*

generation in the distribution networks be quantified?" This was addressed by developing and applying a distribution grid model to the defined scenarios and running power flow simulations on 500,000 geo-referenced synthetic distribution network topologies. The effects of the considered distribution demand and generation are quantified by the frequency of occurrences of operational limiting factor violations in the networks, such as transformer loading, line loading, reverse power flows, and voltage limiting violations.

Key messages:

- ✚ The distribution grid model developed in this chapter uses inputs from previous chapters and performs power flow simulations to ascertain the operational constraints of transformers, lines, and nodes.
- ✚ Base case scenario is developed for 95% CO₂ emission reduction in Germany, 2050.
- ✚ Under the base case scenario, the percentage penetration of PV, battery electric vehicles, and heat pumps are 53%, 80%, and 83%, respectively.
- ✚ Six additional scenarios were developed to assess the sensitivity of the distributed demand and generation in the future.
- ✚ Each scenario has a unique name, with an upward arrow (↑) indicating a 10% increase over the base case scenario and a downward arrow (↓) indicating a 10% decrease over the base case scenario.

7 Results and Discussion

Are the distribution networks adequate for significant increase in distributed demand and generation in the future?

Which technical violations exist with regard to future distributed demand and generation, and how could they be addressed?

Topics covered:

- ✚ Application of distribution grid model on 500,000 geo-referenced synthetic low-voltage networks
- ✚ Analyzing the Operational Limiting Factors (OLFs) corresponding to transformers, lines, and voltage at each node
- ✚ Simple demand side management application

The preceding chapters looked at the development of geo-referenced synthetic distribution networks and a distribution grid model to help answer this thesis's overarching research question. In this chapter, the main findings are leveraged to answer the research question. To accomplish this, the distribution grid model developed in Chapter 6 is applied to the 500,000 geo-referenced synthetic low-voltage networks that were developed in Chapter 5. The objective is to examine the impacts of future distributed demand and generation under the scenarios presented in Chapter 6. Within this framework, all geo-referenced synthetic low-voltage networks are initially assigned with electrical parameters pertaining to transformer types, line types, and household loads. Moreover, as stated in Chapter 4, residential load profiles are distributed in accordance with the probability of various households in Germany. Additionally, along with the active power required by residential loads, reactive power is evaluated using a power factor ($\cos(\varphi)$) of 0.93. It is important to note that all the time-series corresponding to the load and generation are season-specific for weekdays (WD) and weekends (WD) across four seasons (i.e., Spring, Summer, Autumn, and Winter), with a 15-minute time resolution totaling to 768-time-snapshots.

The developed scenarios allocate distributed demand, which include battery electric vehicles and heat pumps, and distributed generation or residential rooftop PV to each network to conduct scenario simulations. For instance, under the base case scenario, 53% of buildings obtain rooftop PV profiles, 80% of buildings receive battery electric vehicle profiles, and 83% of dwellings in the network receive heat pump profiles. Reactive power as time-series was also considered along with active power, with a power factor ($\cos(\varphi)$) of 0.93 for distributed demand and 1.0 for distributed generation. With this complete setup, the scenario simulations are performed using the distribution grid model.

However, simulating 19 scenario cases for 786-time snapshots on 500,000 geo-referenced synthetic networks requires a significant amount of computational power. Therefore, they are performed on an in-house high-performance cluster computer (CAESAR) that consists of 25 compute nodes with a combined total of 1200 logical CPUs. To solve large-scale optimization problems, each compute node includes 1TB of RAM and the simulations are performed using 15 compute nodes. It took approximately 15 days to perform the simulations. Therefore, it can be concluded that even when approximately 800 CPUs are used, the computation time is significantly high. Therefore, the study ruled out the use of a probabilistic method to study the consequences of future distributed demand and generation in low-voltage networks.

In the distribution grid model, a deterministic method is used to extract information about the time-series power at transformers, lines, and the voltage magnitudes at each bus depicted in Chapter 6. Additionally, in order to record the aforementioned information for all networks, a significant amount of storage space is required. Therefore, only the results necessary for analyzing the impacts of demand and generation on the networks are stored and analyzed, including minimum and maximum transformer utilization, time and percentage time snapshots of transformer loading, reverse power flow, line loading, under-voltage, and over-voltage limiting factor violations. The results pertaining to the operational limiting factors for each of 500,000 synthetic networks will be discussed sequentially in the following section. This chapter will also address alternative strategies for mitigating operational limitation factor violations.

7.1 Transformer Violations

The transformer is the first critical component that is analyzed under various scenarios of distributed demand and generation in the low-voltage distribution networks. Here, the violations examined that involve transformers include thermal overloading and reverse power flow violations. In low-voltage networks, transformers must operate within certain specified limits. If the load on the transformer exceeds the limits, large currents flow, increasing copper losses and potentially causing irreversible damage to the transformer. However, the transformer load should not exceed the limits specified in Table 2-3. In other words, to ensure optimal operation, the transformer's rated capacity that is 100% of utilization should not be exceeded. In general, this limit is applicable for a specified period of time; for example, if the transformer operates for one hour at greater than 100% capacity, it is considered a violation of the limit. Nonetheless, this thesis considers it a violation when the transformer is loaded above 100% for each time snapshot simulated (15 minutes).

In this setting, this section will investigate the time period during which the transformer violates the limitations and causes violations. As previously stated, there are two major violations that a transformer could cause: thermal overloading and reverse power flows. First, thermal overloading of the 500,000 low-voltage transformers will be investigated. Thermal overloading occurs when the transformer is utilized at a capacity greater than 100% of its rated capacity. Thus, in the context of a variety of scenario cases, the transformer loadings determined from the simulations performed on 500,000 geo-referenced synthetic low-voltage distribution networks will be investigated. For each transformer, the maximum transformer loading

recorded during an eight-day simulation period will be evaluated for all scenarios. Figure 7-1 illustrates the maximum loading of the transformers observed in the base case scenario.

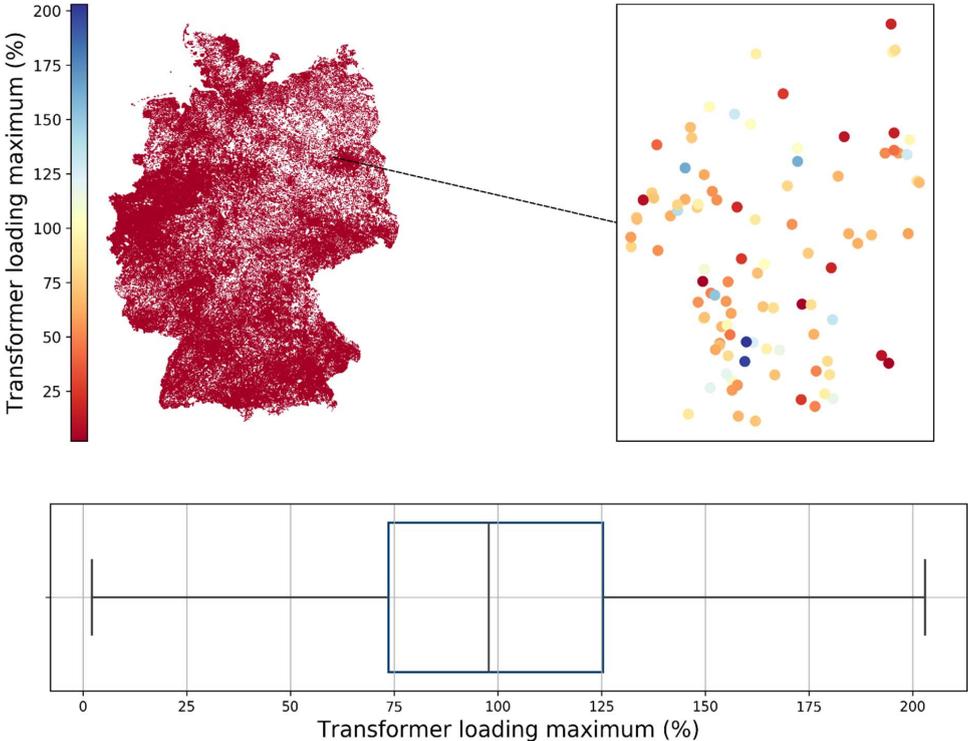


Figure 7-1 Maximum transformer loading observed for 500,000 networks under base case scenario (PBH)

Here, the analysis pertains to the maximum transformer loading that each transformer encountered during the simulated time snapshots. The loading at each time step is simulated using the power output time-series of the transformer in conjunction with its transformer rating. The maximum of the time-series will be taken into consideration to determine the maximum transformer loading for each transformer. Consequently, the maximum transformer loading value indicates the peak utilization of the transformer during the simulated time steps. Thus, it can be assumed that no additional loading occurred for each transformer beyond this utilization.

In Figure 7-1, the map on the top left represents the maximum loading of all low-voltage transformers located geographically. Since there is a large number of transformers, it is impossible to properly view each transformer. The zoomed in level at a specific area is displayed on the right-hand side at the top, along with the maximum loading on each transformer at that position. It is evident from this that each transformer has a distinct maximum loading, and due to the large number of transformers, the discrepancies are not visible on the

map. Nonetheless, the box plot depicts the maximum loading that occurred in each transformer.

Figure 7-1 illustrates that the maximum loading of transformers ranges from 2% to nearly 200%. The 2% value indicates that during the time snapshots studied, the transformer's maximum loading was 2% of its capacity. That is, less than 2% of the transformer capacity was utilized in all the time snapshots simulated. Similarly, 200% is the maximum loading that the transformer experienced for all time-snapshots that were simulated. However, from Figure 7-1, it can be seen that the median is at 100%. This means that 50% of the transformers in the 500,000 geo-referenced synthetic low-voltage networks demonstrated a maximum loading of less than 100%. Therefore, for these 50% of the networks, no other time instances lead to a transformer loading value that is greater than 100%. These 50% of the transformers will never be subject to thermal overloading as per the limit for overloading violations (i.e., >100%). Nonetheless, during 786-time snapshots with a 15-minute time resolution, the remaining 50% of transformers utilized more than 100% of their capacity at least once. Additionally, the third quartile is at 125%, which means that 25% of transformers between the median and third quartile use an additional 25% of their capacity. However, 25% of the transformers utilized only 25% of additional transformer capacity, which is significantly less than the capacity used by the transformers between the third quartile and the maximum value. Moreover, 25% of transformers with 25% excess transformer capacity usage can easily be compensated for in a variety of ways. However, there is a need to pay attention to the remaining **25%** of transformers with a maximum transformer utilization exceeding **125%**. However, it is important to remember that these are the results of the base case scenario (PBH) with 53% rooftop PV, 80% of battery electric vehicle, and 83% of heat pump penetration in the low-voltage distribution networks.

Therefore, 18 additional scenario cases will be examined to determine the possibility of a deviation in installed capacities by 2050. The results of the simulation of 18 additional scenario cases are depicted in Figure 7-2. The 18 other scenario cases include PB↑H↑, PB↑H↓, PB↓H↑, PB↓H↓, P↑BH↑, P↑BH↓, P↓BH↑, P↓BH↓, P↑B↑H, P↑B↓H, P↓B↑H, P↓B↓H, P↑BH, P↓BH, PB↑H, PB↓H, PBH↑, PBH↓. Here, P, B, H represent the base case percentage penetration of 53%, 80%, and 83%, respectively. The upward arrow (↑) represents the increase in percentage penetration by 10% over the base case scenario. Likewise, downward arrow (↓) represents the decrease in percentage penetration by 10% over the base case scenario. Considering all these scenario cases, the box plots in Figure 7-2 represent the maximum loading observed at each transformer for the simulated time-period for all 500,000 geo-referenced synthetic low-voltage networks.

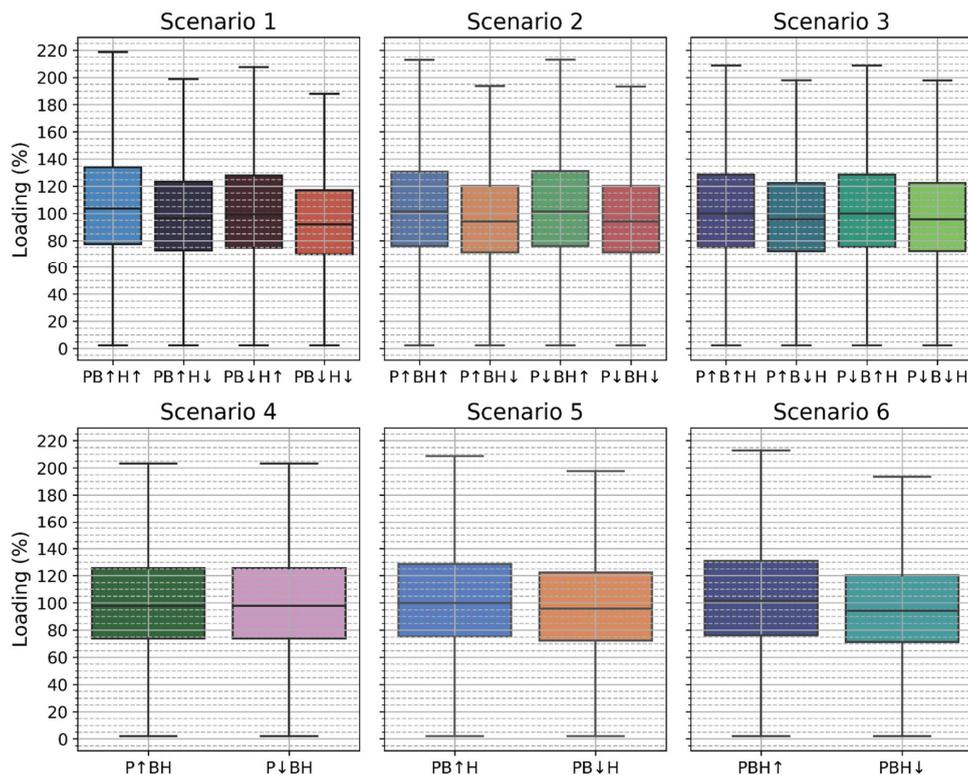


Figure 7-2 Maximum transformers loading (%) observed for all the geo-referenced synthetic low-voltage networks for 18 various scenario cases

As illustrated in Figure 7-2, for most of the scenario cases, the consequences of distributed demand and generation on transformer overloading are nearly identical to those of the base case scenario (i.e., PBH, Figure 7-1). However, the results for $PB\downarrow H\downarrow$, $P\uparrow BH\downarrow$, $P\downarrow BH\downarrow$, and $PBH\downarrow$ indicate that around 40% of the low-voltage transformers exhibit transformer overloading violations. Furthermore, this is around 10% less than the base case scenario (PBH). In the base case scenario, 50% of the transformers experience violations at least once. From $PB\downarrow H\downarrow$, $P\uparrow BH\downarrow$, $P\downarrow BH\downarrow$, and $PBH\downarrow$ scenario cases, it was observed that $H\downarrow$ is the common denominator and denotes a 10% reduction in the percentage of base scenarios (i.e., 74.7%). Thus, heat pump utilization seems to be primarily responsible for the thermal loading of the transformers. It is also evident from the time-series of the transformer loading reported in Chapter 6 for an exemplary network (i.e., N_{23988}). From Figure 7-3, it is clear that transformer overloading occurs only during the winter. Therefore, heat pumps which are primarily operated during winter season (refer, Figure 4-30) are primarily responsible for overloading of the transformers.

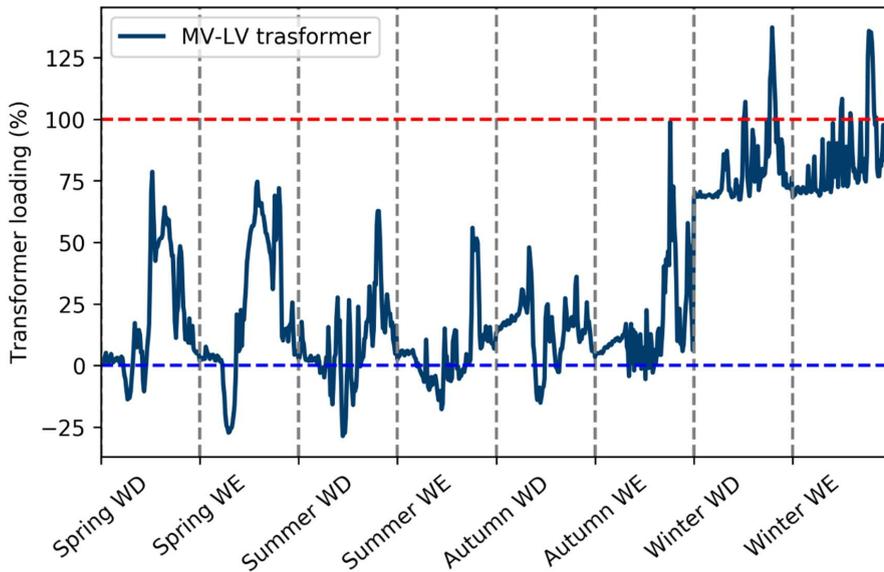


Figure 7-3 Transformer loading seen in Winter when more heat pumps are operating for an exemplary network (N_{23988})

Based on the two results that were derived above - heat pumps with 10% less penetration than base scenario and the transformer overloading observed in winter - heat pumps are identified as the primary driving technology of thermal overloading violations in transformers. Additionally, one could argue that during the winter, rooftop photovoltaic generation is at a minimum, resulting in transformer overloading. However, rooftop PV generation (see Figure 4-32) and demand for heat pumps (see Figure 4-30) are completely opposite. In other words, it entails minimal PV generation and maximum heat pump utilization. It may be for only a brief period during the day that rooftop PV generation reaches its peak. In comparison, heat pump demand is constantly at its peak during the winter. Thus, even when the rooftop PV is operating at maximum capacity during winter, it is the heat pump demand that overloads the transformer. Nonetheless, when compared to rooftop PV generation at its minimum, the instances of overloading will be reduced when rooftop PV generates in the winter.

A comparison of these four scenario cases (viz., $PB_{\downarrow}H_{\downarrow}$, $P_{\uparrow}BH_{\downarrow}$, $P_{\downarrow}BH_{\downarrow}$, and PBH_{\downarrow}) demonstrates that $PB_{\downarrow}H_{\downarrow}$ results in slightly fewer violations with a maximum violation of less than 200%. This scenario case (i.e., $PB_{\downarrow}H_{\downarrow}$) has a 10% lower penetration rate for battery electric vehicles and heat pumps than the base case scenario. In this scenario case, 65% of low-voltage network transformers operate without thermal overloading. However, 35% of transformers experience at least one violation within the total time period of the simulation. Additionally, 10% of transformers have utilized more than 20% of their capacity at least once in the total time period of the simulation. This indicates that 10% of transformers consume around 120% of their capacity. However, the additional 20% can readily be offset by auxiliary

services. However, it is necessary to address the remaining 25% of transformers with a maximum load exceeding 120%.

In conclusion, under the base case scenario (PBH), 50% of the low-voltage transformers are within the limits. The additional 50% of the transformers encounter overloading at least once in total time snapshots simulated. Furthermore, 25% of these 50% transformers operate at less than 125% of their rated capacities and thus, several steps can be undertaken to regulate them. It is crucial to address the remaining 25% of the transformers that are operating at higher than 125% of their rated capacities. During these simulations, it was concluded that heat pumps are the leading source of overvoltage violations irrespective of PV generation. Moreover, a decrease in the penetration of battery electric vehicles and heat pumps can reduce the number of transformers that need additional consideration in the future.

Until now, we have observed the maximum loading on each transformer over all time instances that were simulated. This indicates the number of transformers that are expected to exhibit transformer overloading violations when demand and generation are integrated into the low-voltage distribution networks in the future. However, experiencing violations for a short period of time may be insignificant, while persistent overloading causes critical damage to the transformer. Therefore, additional analysis is performed on the transformer loading time series to determine the total time period during which the transformers are overloaded. This is performed for all 19 scenario cases, including the base case scenario, on all the geo-referenced synthetic network transformers.

Figure 7-4 illustrates the time span during which each transformer is overloaded, as well as the percentage of total time the transformers are overloaded. Here, the time and percentages are provided for a year and derived from 768-time instances with 15-minute time snapshots each. This study will consider the total time during transformer overloading. For instance, if a transformer in a low-voltage network operates at more than 100% of its capacity once during a 15-minute time snapshot in the morning, twice in the afternoon, and once in the evening, this pattern will continue for 365 days. This means the transformer exceeds its capacity 1460 times (i.e., 365 hours) or 4% of the time over the year. These values were collected and assessed for all 500,000 geo-referenced synthetic low-voltage network transformers.

The previous analysis revealed that only 50% of the 500,000 geo-referenced synthetic low-voltage network transformers are subjected to transformer overloading violations. Figure 7-4, showcases that the minimum value, first quartile, and median are all equal to zero. This indicates that 50% of transformers are not subject to any violations, as seen in previous analyses. Therefore, 50% of the transformers from the 500,000 geo-referenced synthetic networks (i.e., 250,000 transformers) are experiencing transformer overloading violations.

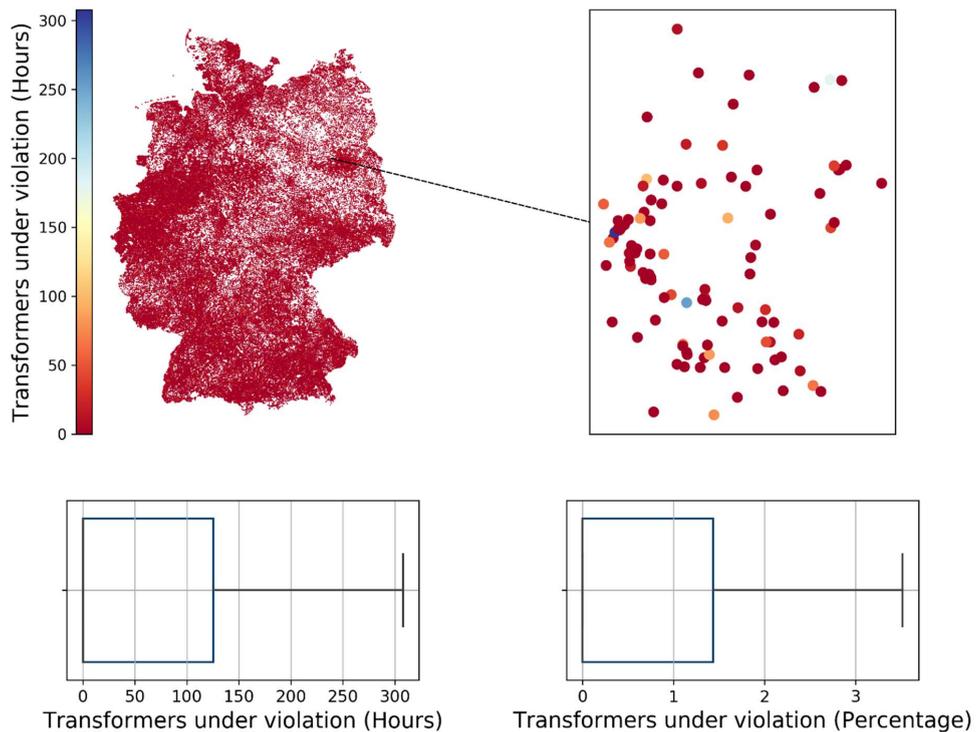


Figure 7-4 Transformers under violations (Hours) and percentage of total time (8760 hours) for 500,000 geo-referenced synthetic low-voltage networks under base scenario (PBH)

Furthermore, 25% of the transformers positioned between the median and third quartile exhibit thermal loading violations ranging from approx. 11 to 125 hours per year. The remaining 25% of transformers above the third quartile have more than 125 hours and fewer than 310 hours under overloading (i.e., 1.5 to 4% of time of a year). Thus, it is apparent that only 25% of the transformers in the 50% of transformers with violations require critical attention.

As illustrated in Figure 7-3, the transformer for the considered network (N_{23988}) is overloaded approximately six times during the total simulated time instances, which is approx. 68 hours a year. As a result, this network falls between the median and third quartile of data. From this, it is clear that the transformer is frequently overloaded during the winter season. It can also be claimed that networks with similar characteristics will fall under this 25% of the networks. However, since the percentage of total time is less than 2%, these instances can be controlled through alternative means.

However, as stated above, several additional scenarios are considered in addition to the base scenario. The results for these scenarios are depicted in Figure 7-5.

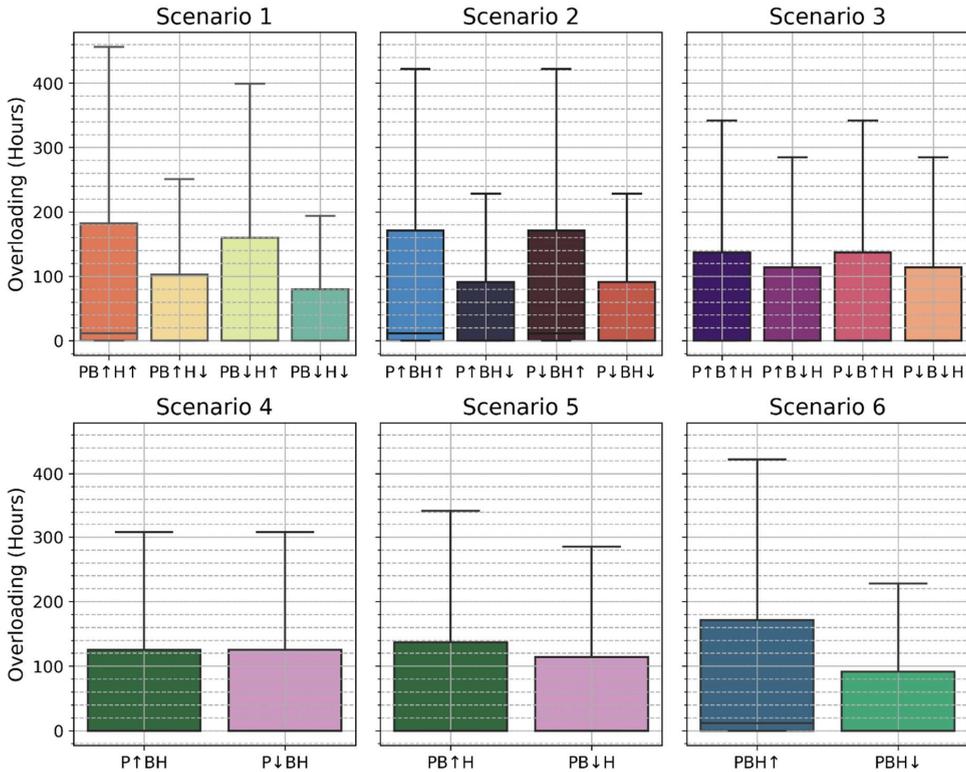


Figure 7-5 Transformers under overloading violations (Hours) observed for all the 500,000 geo-referenced synthetic low-voltage network transformers for 18 various scenario cases

According to Figure 7-5, 50% of the transformers do not exhibit any violations in all scenarios except P↑BH↑, P↓BH↑, and PBH↑. However, for these scenario cases, there are additional networks exhibiting overloading violations. In all of these scenarios, it is observed that there is an additional 10% of heat pump penetration (H↑) (i.e., 91.3%) over the base case scenario. On the other hand, PB↓H↓, P↑BH↓, P↓BH↓, and PBH↓ have transformers with a lower violation count. These scenarios feature a 10% reduction in heat pump penetration (H↓) compared to the base case scenario. Therefore, the thermal loading of the transformer is primarily affected by the heat pump penetration.

This study also considers the average time in hours for which transformers are overloaded over a period of one year for all scenarios, including the base case scenario. Figure 7-6 illustrates the average values for all scenarios. The results indicate that all scenario scenarios, except for **PB↓H↓**, have comparably similar average values. This scenario has a similar percentage penetration of residential solar PV as the base scenario, but the percentage penetration of distributed demand - battery electric vehicles and heat pumps - is 10% less than the base scenario (**B↓H↓**).

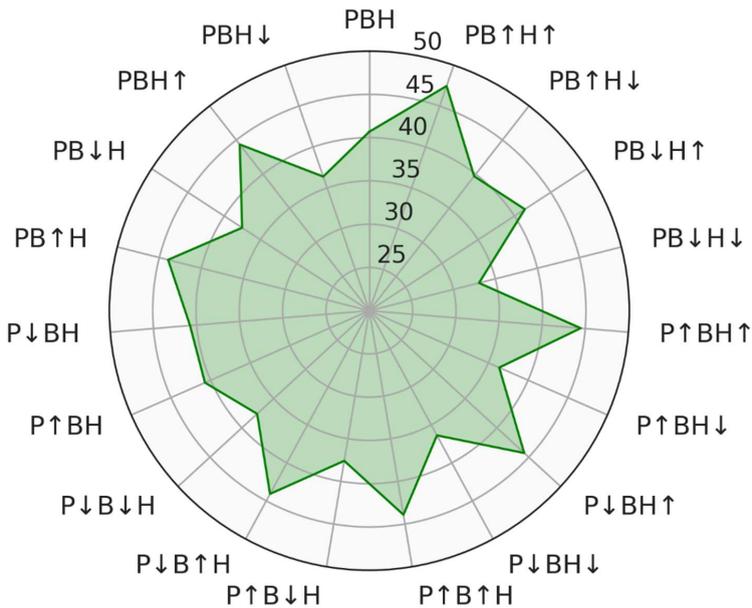


Figure 7-6 Average time transformers overloaded (Hours) for 500,000 networks for 19 scenario cases

Therefore, the **PB↓H↓** scenario case is analyzed carefully to identify the proportion of transformers that require special attention and identify the technologies that are sensitive to transformer overloading. PB↓H↓ scenario case, similar to the base scenario, shows that 50% of transformers are in violation. However, 25% of these 50% of transformers experience violations ranging from 1 to 7 times over the 768-time snapshots simulated. This adds up to approx. 11 hours to 80 hours a year. However, these 25% of transformers are not critical and may be regulated by various techniques. Additionally, based on the previous analysis, the majority of violations were recorded in winter, which can be attributed to heat pumps. In such instances, reducing the use of heat pumps can help improve the situation. Furthermore, another 25% of transformers are in violation for more than 200 hours but less than approx. 300 hours. Consequently, 25% of all transformers assessed require special attention.

In conclusion, most of the overloading time instances occur during the winter season, as evidenced by the exemplary network (N_{23988}) time series. Furthermore, the networks that experiences overvoltage violations increase with an additional 10% heat pump penetration (H↑). In contrast, a decreased penetration (H↓) to base case scenario reduces the overloading time. Thus, fewer additional services will be required to mitigate the overloading violations by reducing the use of heat pumps during winter. Nevertheless, 25% of the transformers which are critical needs additional attention.

However, transformer overloading is not the only violation that can affect the transformer. The reverse power flows at the transformer also negatively impacts them and the protection

equipment, if not properly maintained. Therefore, the effects of future distributed demand and generation on transformer's reverse power flows are analyzed. First, the minimum transformer loading on each transformer should be observed because the power flow is in the opposite direction and minimum loading can provide the maximum power flow in reverse direction that is from low-voltage to medium-voltage.

By simulating 500,000 geo-referenced synthetic low-voltage distribution networks under the base case scenario and 18 scenario cases, the minimum transformer loading was determined. Figure 7-7 illustrates the minimum transformer utilization observed under the base scenario for all geo-referenced synthetic low-voltage networks.

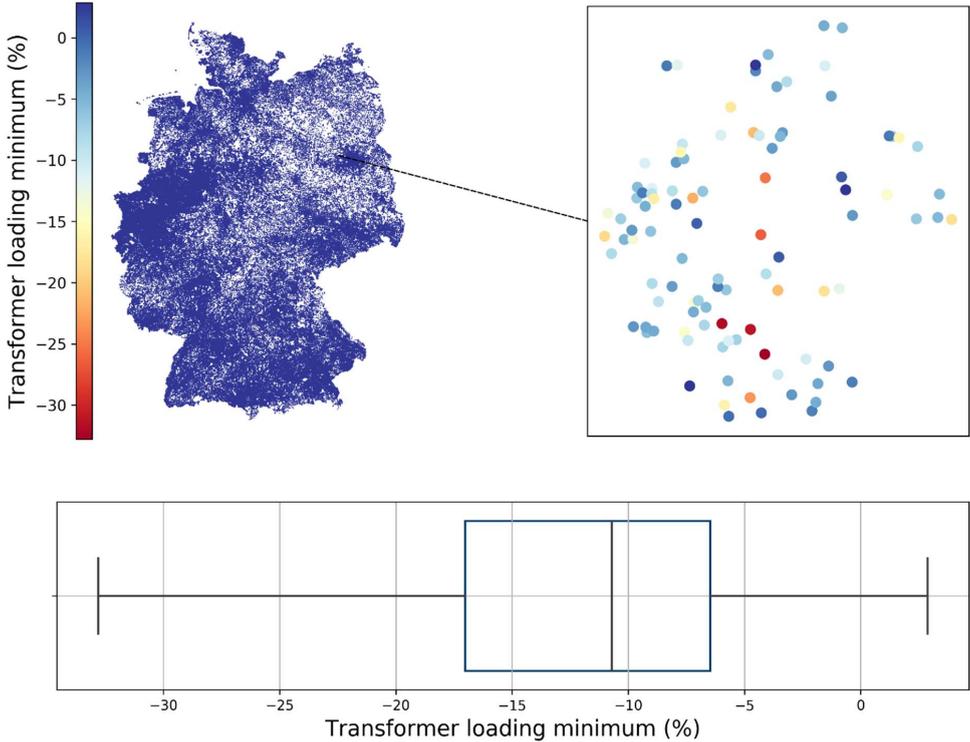


Figure 7-7 Minimum transformer loading observed for 500,000 geo-referenced synthetic networks under base case scenario (PBH)

The minimum transformer loading was determined here by applying minimum on the transformer loading time series. The minimum of all the time snapshots simulated for eight days is identified based on the analysis in previous sections. The minimum transformer loading determines the transformer's minimal utilization. This could be either positive or negative, depending on the network's circumstances. Negative implies reverse power flow. In other words, the transformer's capacity is being used to transmit power from the lower voltage side to the higher voltage side. This phenomenon occurs when there is a surplus of power generated by distributed generation and a deficiency of demand on the low-voltage side. When

a transformer's capacity is being used to transfer power from the higher voltage to the lower voltage side, it is known as forward power flow and is indicated as positive shows. The phenomenon occurs when there is a high demand requirement and a low level of generation on the low-voltage side (see, Figure 7-3).

In Figure 7-7, the map on the top left represents the minimum loading of all geo-referenced synthetic low-voltage transformers located geographically. Due to the large number of transformers, it is impossible to clearly view each transformer. However, the zoomed in level at a specific area is displayed on the right-hand side at the top, along with the minimum loading on each transformer at that position. It is evident from the figure that each transformer has a distinct minimum loading, and the differences are not visible on the map due to the large number of transformers. Nonetheless, the box plot depicts the minimum loading experienced by each transformer.

The box plot in Figure 7-7 represents the minimum loading observed for each transformer in the base case scenario (PBH). The results demonstrate that the minimum value for each transformer ranges from -37% to $+5\%$. In a few 15-minute time snapshots, the minimum transformer utilization observed for a few/one transformers of 500,000 transformers is around -37% . This value signifies that 37% of the transformer's capacity is being used to transmit power toward the higher voltage. However, as illustrated in Figure 7-7, more than 75% of transformers have experienced negative transformer utilization at least once, indicating that there is a significant probability of them exhibiting reverse power flow violation. Reverse power flow violations are related to the transformer's negative utilization. According to the analysis, 50% of transformers exhibit negative transformer violations, which are lesser than -11% . The remaining 50% of transformers exhibit a value greater than -11% for a specific time snapshot. However, the map demonstrates that the overall geographical distribution is blue, not mixed, because transformers a greater negative loading than 11% are essentially scattered and isolated from all other transformers with loading higher than -11% . That means there are diverse values ranging from -17% to -11% . Additionally, the proper distribution is obscured by the high density of transformers and the tiny size of the map.

Nevertheless, it is significantly concerning that 50% of the 500,000 transformers experience reverse power flow by utilizing more than 11% of the transformer's capacity are of significant concern. The risk of permanent damage to the protection equipment is greater if the percentage of transformer utilization for reverse power flow use is higher. For instance, if an onload tap change transformer is installed to change the transformer taps and maintain the voltage limits at the low voltage side due to the reverse power flow, abrupt tap shifts occur resulting in an abrupt increase of the current. This causes the overheating of transformer windings.

However, 25% of these 50% of transformers demonstrate less than 17% and greater than 11% reverse power flow utilization. Therefore, in addition to the 50% of transformers exhibiting reverse power flow by consuming less than 11% of transformer capacity, 25% of the transformers are in the median and third quartile. The violations exhibited by the latter can be

easily compensated for by employing ancillary services. Finally, low-voltage transformers that are inside the third quartile and exhibit the maximum value, which is **25%** of 500,000 transformers, are crucial for reverse power flows.

However, this conclusion is drawn in the context of the base case scenario (PBH) with 53% PV, 80% battery electric vehicles, and 83% heat pumps. Subsequently, to determine the probability of deviations in the installed capacities, the previously developed scenario cases are examined. Figure 7-8 illustrates several additional scenario cases that were investigated for minimum transformer utilization. As depicted in Figure 7-8 the additional scenarios do not significantly differ from the base case scenario. Nonetheless, $P_{\downarrow}BH_{\uparrow}$, $P_{\downarrow}BH_{\downarrow}$, $P_{\downarrow}B_{\uparrow}H$, $P_{\downarrow}B_{\downarrow}H$, and $P_{\downarrow}BH$ exhibit some variation and in all these scenario cases, P_{\downarrow} is common. This represents a 10% reduction in PV penetration (i.e., 47.7%) relative to the base case (i.e., 53% of PV).

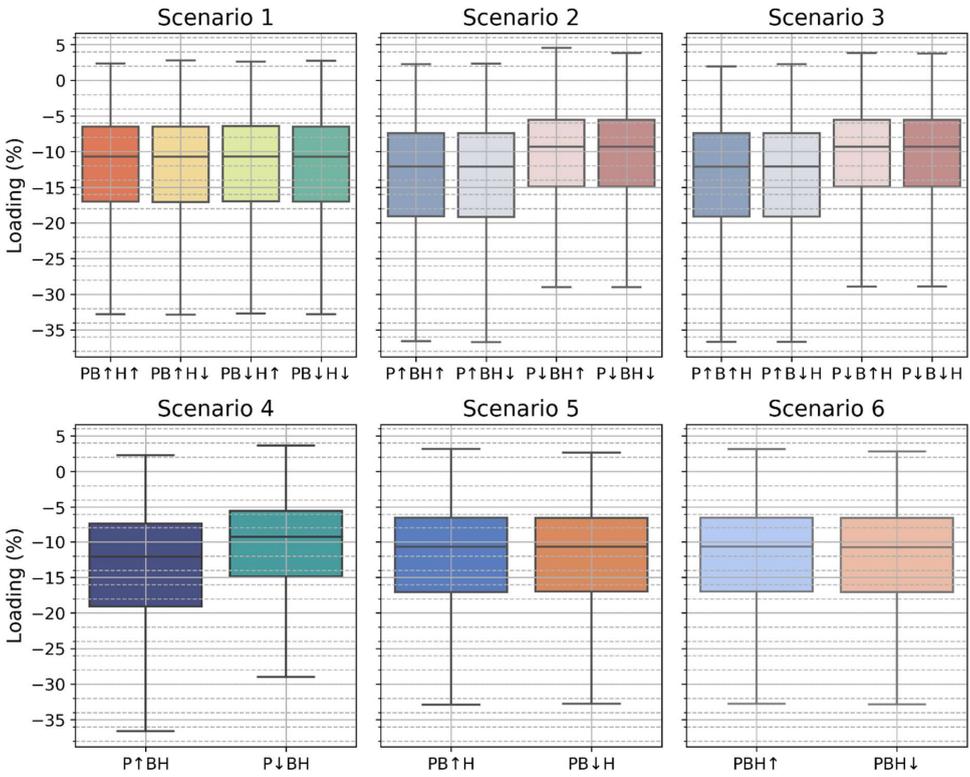


Figure 7-8 Minimum transformer loading (%) observed for all the considered low-voltage networks for 18 various scenario cases

In the base case scenario, 75% of the transformers exhibit reverse power flows utilizing less than 17% of the transformer capacity. However, in the aforementioned scenario cases, where P_{\downarrow} is common, that denotes a 10% reduction in the percentage of base scenario (i.e., 47.7%). Here, 75% of the transformers have reverse power flow utilization that is less than 15% of the

transformer capacity, which is 2% less than the base case scenario and other scenarios. Additionally, the minimum value observed for these scenario cases is significantly lower than the base case scenario and other scenario cases, where PV is fixed or $P \uparrow$ (i.e., PV penetration increased by 10% over base scenario). Thus, it can be stated that PV generation accounts for the majority of reverse power flows at the transformers. Additionally, as demonstrated by the time series of transformer loading reported in Chapter 6 for an exemplary network (N_{23988}) and depicted in Figure 7-3, negative transformer loading occurs during the summer and spring. Moreover, as evidenced by the PV generation profiles in Chapter 4, rooftop PV generates more energy in the spring and summer than in the autumn and winter.

According to these observations, a 10% reduction in PV penetration from the base case scenario and the minimum loading of the transformer during the summer and spring indicates that PV generation causes transformer reverse power flows. However, to determine whether changes in the distributed demand - battery electric vehicles and heat pumps - impact reverse power flow, average values for all scenario cases were evaluated. When 500,000 geo-referenced synthetic networks transformers were observed for 19 scenario cases, including the base scenario (PBH), the average value of the minimum transformer loading is depicted in Figure 7-9.

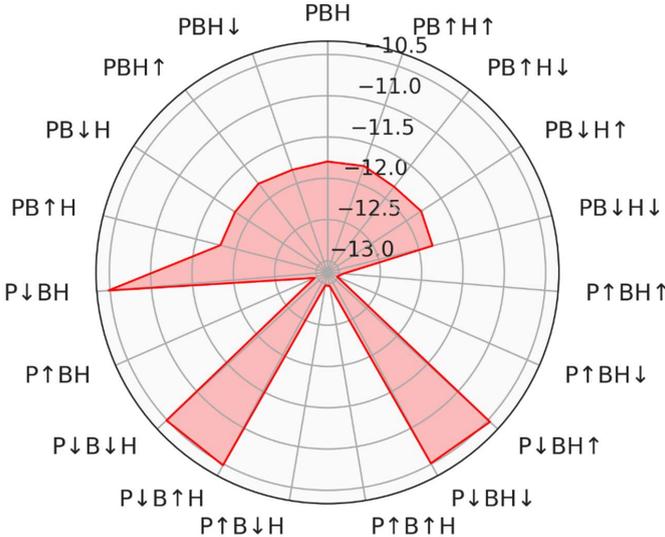


Figure 7-9 Average minimum transformer loading for 500,000 networks for 19 scenario cases

Considering the average results of all the transformers examined in all scenario cases, $P \downarrow BH \uparrow$, $P \downarrow BH \downarrow$, $P \downarrow B \uparrow H$, $P \downarrow B \downarrow H$, and $P \downarrow BH$ produce better results than the base scenario. The extreme points in Figure 7-9 correspond to the lowest average minimum loading observed and denotes $P \downarrow BH \uparrow$, $P \downarrow BH \downarrow$, $P \downarrow B \uparrow H$, $P \downarrow B \downarrow H$, and $P \downarrow BH$. Here, the deviations are negligible, and the only pattern observed is that $P \downarrow$ is common in all scenario cases. In other words, when PV penetration is decreased by 10% over the base case, the deviations are negligible.

Furthermore, it is also apparent that higher battery electric vehicles penetration with the same heat pump number can also help reduce reverse power flows.

In conclusion, under the base case scenario, more than 75% of the transformers exhibited reverse power flow at the transformer at least once. This was determined by examining the minimum loading of the transformers. However, only 25% of the transformers were identified as utilizing more than 17% of their capacity to transfer power toward the high-voltage side and demonstrating greater influence on the protection equipment in that network. Additionally, it was determined that solar rooftop PV is the primary cause of the reverse power flow at the transformer. Furthermore, it may also be concluded that distributed demand has minimal influence on the reverse power flow. In any scenario case, regardless of the increase or decrease in the penetration, they are not reducing reverse power flows at the transformers. That is, there is a significant mismatch between generation and demand, resulting in these effects.

The violation involving reverse power flow must be investigated further to determine whether the given penetration levels of distributed demand and generation are suitable for future integration. This is important because previous analysis demonstrates only the maximum utilization of the transformers as the reason for reverse power flow. In this analysis, the technology causing the violations is identified along with the number of networks that are impacted at least once as a result of distributed demand and generation. Therefore, it is crucial to determine how frequently these incidents occur during the simulated time in order to draw conclusive findings. Thus, a further evaluation is conducted to determine the time period over which the transformer experiences reverse power flow violations at each transformer.

The analysis above identified the maximum transformer loading observed for reverse power flow for each network and concluded that the maximum observed is -37% in the base case and -29% percent in other scenarios where PV is 10% less than the base case scenario (P↓), regardless of any other demand technologies. However, within the simulated time, the loading was observed just once. To establish that the transformer is under violation, it is critical to examine the amount of time over which the transformer exceeds its limits. Thus, the number of time snapshots depicting reverse power flow for each transformer experiences is considered regardless of its capacity. However, Figure 7-10 illustrates the reverse power flow time for a time period of a year for 500,000 geo-referenced low-voltage network transformers in the base case scenario.

Figure 7-10 illustrates the time indicated for reverse power flow instances for a year. It can be observed from the figure that the minimum value is zero and the maximum value is approximately 2000 hours. Here, the displayed reverse power flow time is estimated for a year from an 8-day simulation period with 15-minute time snapshots. Furthermore, the minimum time zero indicates that some of the transformers have been operating without reverse power flow violation, while 2000 hours indicates that some transformers are exhibiting reverse power violation for 2000 hours. In other words, for 24% of the total time, the transformers are feeding power back into the medium-voltage network, regardless of their capacity. As previously

stated, the maximum percentage usage is -37% . Therefore, the rate at which it feeds back does not exceed the -37% limit. However, for some of the transformers, the maximum amount of time spent feeding is 24% of the overall time considered, which is a significant concern.

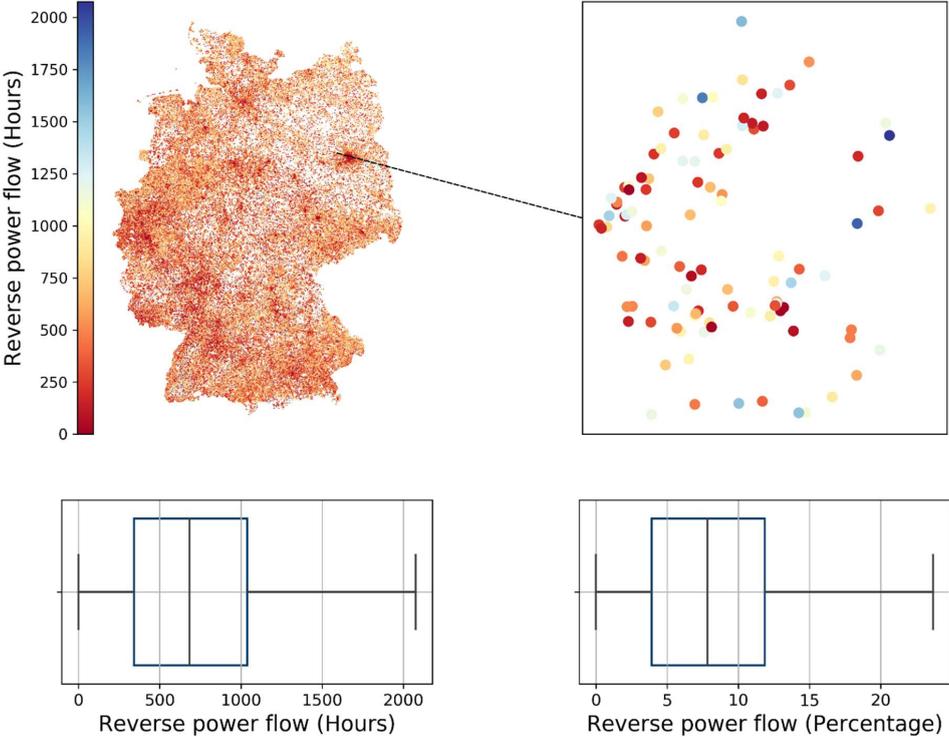


Figure 7-10 Transformers under reverse power flow (Hours) for a year for 500,000 low-voltage transformers for base case scenario (PBH)

However, from Figure 7-10, it can be seen that the interquartile range indicates that 50% of the transformers experience reverse power flows between 340 and 1000 hours. This equates to 4% to 12% of the total time considered. Additionally, the interquartile range is close to the minimum value, indicating that 75% of transformers exhibit violations lasting fewer than 1000 hours. However, the remaining 25% of the transformers exhibit reverse power flows between 1000 to 2000 hours, or about 12% to 24% of the total time for a year. Therefore, it can be concluded that 75% of the transformers are operating under reverse power flows for periods ranging from 340 to 2000 hours, and 50% for periods ranging from 340 to 1000 hours. These data demonstrate unequivocally that residential rooftop PV generation is not being properly utilized at the time of generation, leading to feeding such generation back to the medium-voltage side, resulting in violations.

However, Figure 7-11 illustrates the other sensitivity calculation scenarios that are considered. It is clear from the Figure 7-11 that for the majority of scenarios, the violations are the same as in the base scenario (PBH). However, the violations are less severe in $P\uparrow BH\downarrow$, $P\downarrow BH\downarrow$, $P\downarrow B\uparrow H$,

$P\downarrow B\downarrow H$, and $P\downarrow BH$ when compared to the base case (PBH). Based on the preceding analysis, the likelihood of violation decreases in these cases as the solar rooftop PV penetration decreases in low-voltage networks.

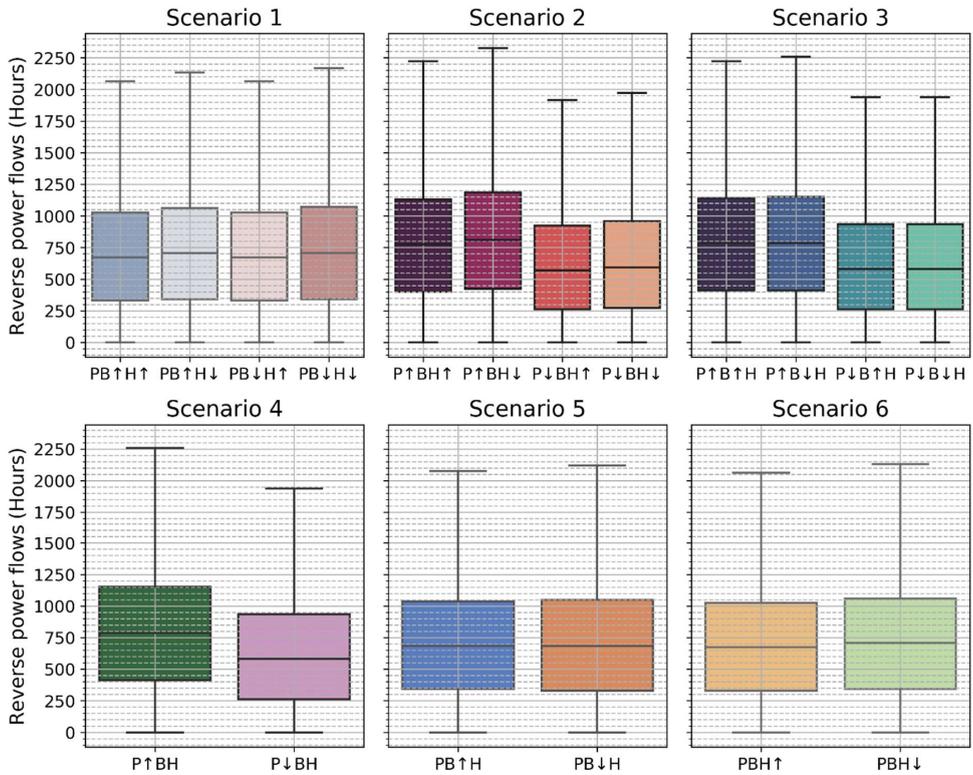


Figure 7-11 Transformer reverse power flow (Hours) observed for all the considered low-voltage network transformers for 18 various scenario cases

By observing Figure 7-11, it may be argued that either an increase in the demand ($H\uparrow$, $B\uparrow$) or a decrease in the demand ($B\downarrow$, $H\downarrow$) has a minimal effect on the reverse power flows. But, when PV is decreased ($P\downarrow$), a significant change can be observed in the reverse power flow.

Furthermore, this study also considers the average time of reverse power flow for the transformers for all scenarios. Accordingly, Figure 7-12 depicts the average time of reverse power flows at all transformers in each scenario, which includes the base scenario and the other 18 scenario cases. The average number of times the transformer is subjected to reverse power violations is between 621 and 810 hours for a year.

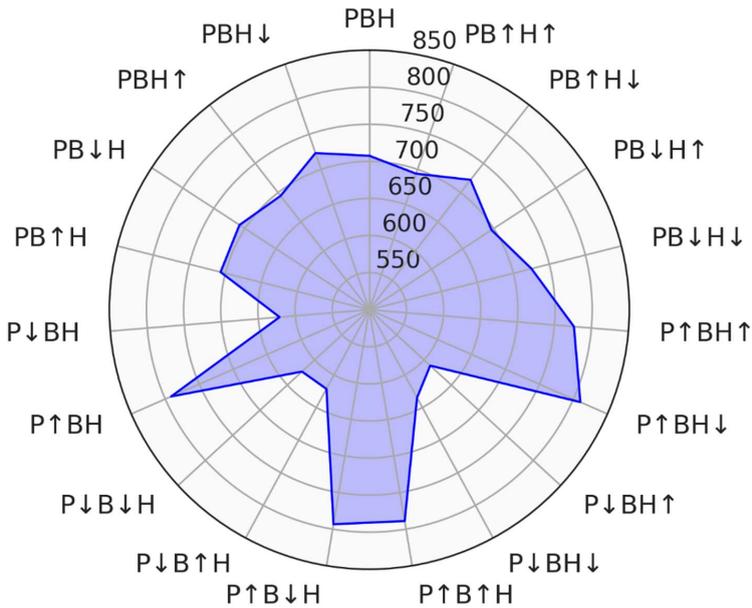


Figure 7-12 Average time reverse power flows at 500,000 transformers for 19 scenario cases

As illustrated in Figure 7-12, a lower count indicates fewer violations, while a higher value indicates that the transformers are exhibiting reverse power flows more frequently. The figure clearly showcases that the scenario cases such as $P\uparrow BH\downarrow$, $P\uparrow B\uparrow H$, $P\uparrow B\downarrow H$, and $P\uparrow BH$ experience around an average of approx. 800 hours of reverse power flow over the course of a year. In these scenarios $P\uparrow$ is the common denominator, and irrespective of battery electric vehicles and heat pump demand, the violations are higher when compared to the base case scenario. In contrast, $P\uparrow BH\downarrow$, $P\downarrow BH\downarrow$, $P\downarrow B\uparrow H$, $P\downarrow B\downarrow H$, and $P\downarrow BH$ are the scenario cases where the reverse power flows are less than the base case scenario (PBH). In this case, $P\downarrow$ is the common denominator and the violations decreased regardless of battery electric vehicles and heat pump penetration.

In conclusion, most of the reverse power flows occurred during spring and summer, as evident from the minimum transformer loading at the transformer (see Figure 7-3). Furthermore, by analyzing various several scenarios, it is evident that the scenarios with $P\downarrow$ decreases the violations and $P\uparrow$ increase the violations, regardless of distributed demand in the network. It can be further concluded that the high reverse power flow is a result of a mismatch between the generation and demand in the network. Therefore, the generation should be effectively utilized to minimize violations involving reverse power flows. Finally, it can be concluded that more than 75% of the 500,000 transformers are exhibiting reverse power flows, but only 25% of the 500,000 networks are in critical need of attention.

Since transformer violations such as transformer loading and reverse power flows have been discussed, it is now time to compare the reverse power flow violations with the transformer

thermal loadings to determine whether reverse power flows, when appropriately employed, can compensate for the transformer's thermal loading.

In this context, Figure 7-13 compares the average reverse power flow duration over the course of a year for all of the examined geo-referenced synthetic low-voltage networks to the average thermal overloading during a simulated year for all scenarios.

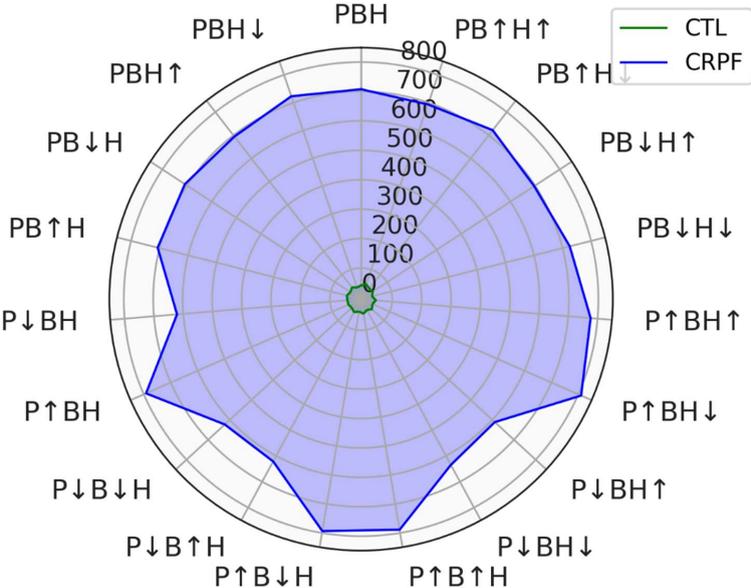


Figure 7-13 Comparison of transformer thermal overloading duration with reverse power flow duration

According to the observed results, reverse power flow violations occur significantly more often due to the widespread use of distributed generating (i.e., residential solar rooftop PV). Additionally, this occurs when the network has a high generation rate and a low demand rate. On the other hand, the transformer is thermally loaded because of the significant penetration of distributed demand - battery electric vehicles and heat pumps – particularly the latter, from the scenarios analyzed. In contrast to reverse power flows, the thermal loading of the transformer occurs due to the network's high demand and low generation. Therefore, by utilizing distributed generation (i.e., rooftop PV) effectively, reverse power flows and thermal loading on the transformer is reduced. Meanwhile, it is also important to point out that the transformer loading duration is relatively minimal in proportion to the duration of reverse power flow. Thus, when the solar PV power is stored or used concurrently, most reverse power flow situations can efficiently minimize transformer loadings. But, according to the analysis conducted, the majority of the transformer thermal loading occurs during the winter, where heat pumps are utilized more and there is less generation by solar PV. In contrast, reverse power flows are observed in the spring and summer, where there is minimal operation of heat

pumps. Therefore, transformer loading violations that occur during the winter cannot be easily compensated with the solar PV. Nevertheless, if the loading occurs during the spring or summer, additional battery integration can compensate for the mismatch between generation and demand, thereby reducing thermal loading and reverse power flows.

After analyzing the results pertaining to transformer violations under various scenarios of distributed demand and generation integration in low-voltage distribution networks, the following section looks at the violations concerning power lines.

7.2 Power Line Violations

After examining transformer violations, the next significant component that might be impacted are power lines. When the current carried by a power line exceeds its current carrying capacity, the line is said to be in violation. In general, the power lines can operate beyond their limits for short duration of time without any interruption. However, this thesis considers a line to be in violation if more than 100% of its rated capacity is drawn through it. The term "power line" refers to the line that connects two nodes without connecting to the feeder. However, the feeder is made up of multiple line segments.

When 500,000 geo-referenced synthetic low-voltage networks were simulated for the future integration of distributed demand and generation under 19 scenario cases, lines in the networks occasionally exhibited thermal loadings. As seen in Figure 5-27, each network has many lines, ranging from one to around 200 lines. Thus, Figure 7-14 illustrates the total number of lines that experienced a violation during at least one simulation over the complete duration of simulations conducted for all networks. The total count indicates that if a line is violated once or numerous times over the duration of the simulation, the line is deemed to be violated. Furthermore, just because a network's total number of lines experiencing violations is a constant value (for example, 3 lines), it does not necessarily mean that all of those lines are experiencing violations simultaneously.

As shown in the Figure 7-14, each network has between zero and 44 lines that are in violation at least once throughout the time snapshots studied. Additionally, it is shown that 50% of networks contain more than 10 lines subjected to thermal loading. However, 50% of total networks in the interquartile range had between five and 20 lines. Among these 50% of networks, between 6% and 19% of total lines are subjected to thermal loading. Nonetheless, practically all networks include at least some lines that are subject to thermal loading violations. As a result, a particular emphasis should be placed on decreasing the load on the network's lines while implementing future distributed demand and generation.

However, various other scenario scenarios were examined to see whether the total number of lines in the network may be lowered. If the total number of lines subjected to thermal loading in each network is decreased, the reason is investigated further. For each network across all 500,000 geo-referenced synthetic low-voltage networks, the lines experiencing thermal loading count are displayed in Figure 7-15 as a box plot for each scenario case.

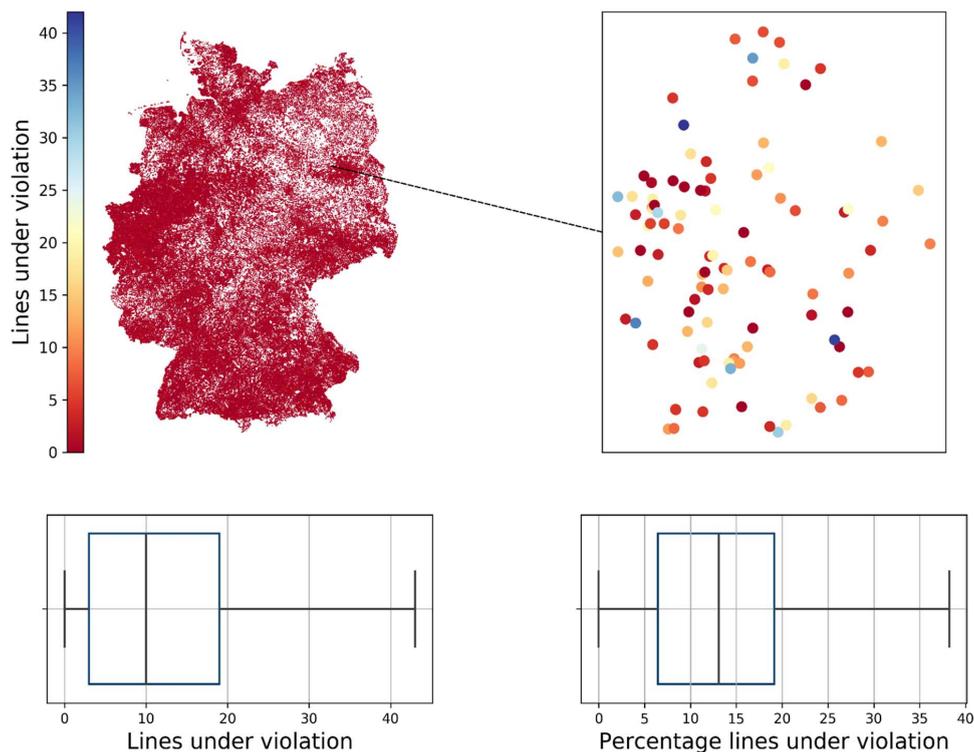


Figure 7-14 Lines under thermal loading for each network for 500,000 geo-referenced synthetic low-voltage networks for base case scenario (PBH)

In Figure 7-15, it is evident that nearly all the scenario cases produce comparable results, and the variances are negligible. Except in the cases of $PB\uparrow H\downarrow$, $PB\downarrow H\downarrow$, $P\uparrow BH\downarrow$, $P\downarrow BH\downarrow$, $P\downarrow B\downarrow H$, $P\downarrow BH$, and $PBH\downarrow$, the minimum value is zero, and the maximum value is 44. That is, the number of lines subject to violations varies between zero and 45, which is consistent with the base case. However, in other cases, the minimum value remains constant, and the maximum value is reduced to 40, with the third quartile lying at 18. That is, 75% of low-voltage networks contain less than 18 lines that are thermally loaded. In these scenarios, there is at least one distributed demand penetration that is less than 10% of the base scenario or distributed generation that is 10% greater than the base scenario. Furthermore, from these scenario cases, it was observed that $H\downarrow$ is common for most scenario cases. $H\downarrow$ denotes a 10% reduction over the percentage of base scenario (i.e., 74.7%). Therefore, it could also be claimed that heat pump has influence on the line loading. Furthermore, this is also evident from the time-series of line loading for a line taken from the exemplary network (N_{23988}) simulated in Chapter 6. In Figure 7-16, it is clear that line loading on a line occurs almost constantly in winter. However, heat pumps are the primary technology that consume power during the winter.

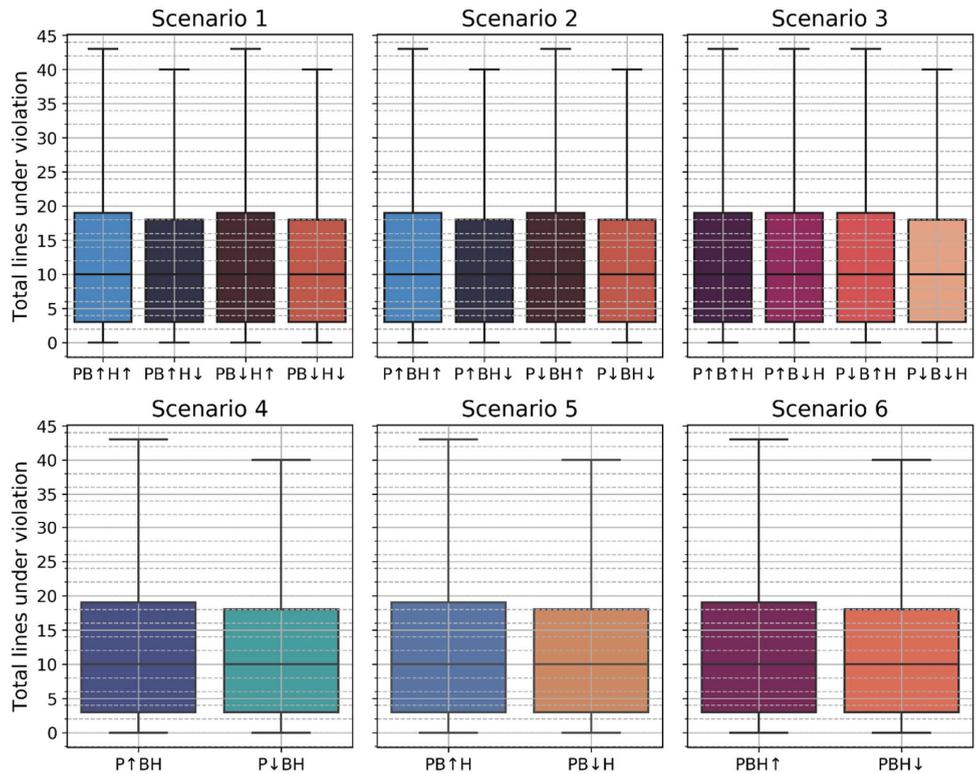


Figure 7-15 Total lines under thermal loading observed for all the considered low-voltage networks for 18 various scenario cases

In conclusion, under the base case scenario (PBH), almost all the networks have at least one line exhibiting thermal loading violation. Moreover, 50% of the networks have at least ten lines that are subjected to thermal loading violations. The remaining 50% of the networks have more than ten lines with thermal violations and fewer than 45 lines demonstrating thermal violations. Interestingly, in scenarios with less lines under violations than base case scenario, heat pumps seem to be the common technology in place. In those scenarios, a 10% reduction in its penetration level relative to the base case is observed. Furthermore, the majority of line overloading is seen in the winter season. Therefore, by using this information, line loading on the lines can be reduced by decreasing demand on the network during winter season or reinforcing the network with an additional line that is parallel to the line experiencing thermal loading violations.

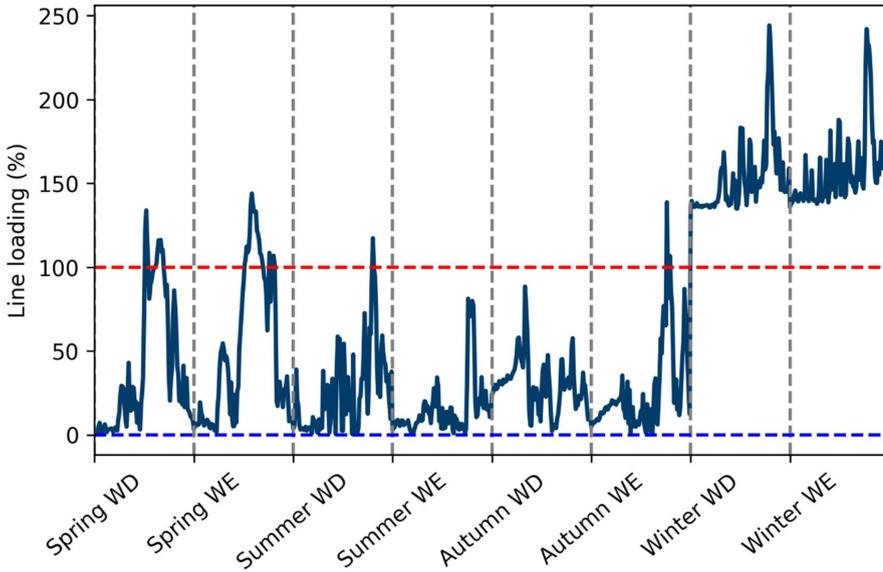


Figure 7-16 Line loading seen on a line from exemplary network (N_{23988}) simulated in Chapter 6

However, this analysis does not provide a clear outline regarding the number of lines a network needs for reinforcement. Thus, the average number of lines under loading for each network is examined and depicted in Figure 7-17.

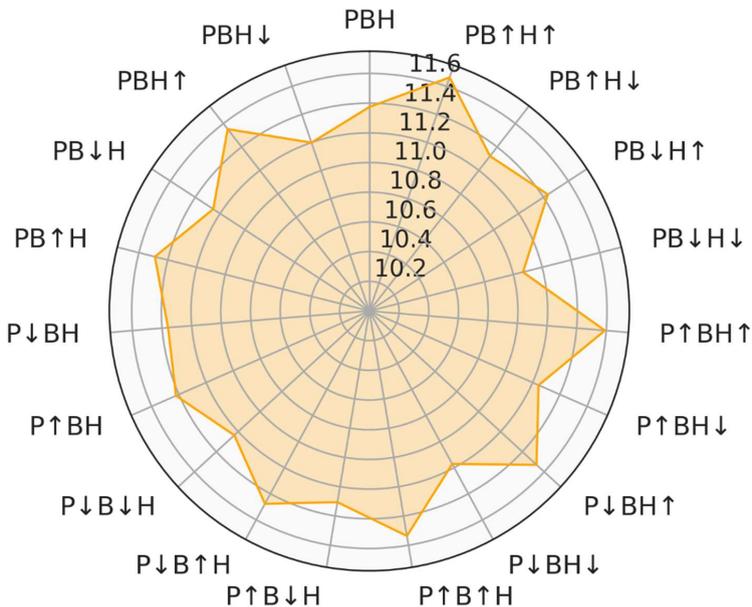


Figure 7-17 Average lines under thermal loading for 500,000 networks for 19 scenario cases

The average number of lines under thermal violations in 500,000 geo-referenced synthetic networks is between ten and eleven. This demonstrates unequivocally that all scenario cases, including the base scenario, showcase similar results. Therefore, an average of approximately 11 lines must be reinforced to accommodate the future distributed demand and generation in the low-voltage networks without experiencing line loading on any low-voltage network. On average, this adds up to approx. 11% of lines in 500,000 geo-referenced synthetic low-voltage networks.

It is important to note that the thermal stress on lines in a network is primarily concentrated on those proximal to the transformers. As illustrated in Figure 7-18, the lines highlighted in red are more stressed than the lines highlighted in blue. Clearly, these lines are located near the transformers. Furthermore, it is apparent that the lines adjacent to the transformers experience higher stress when the feeder has a larger number of loads. As illustrated in Figure 7-18, the lines adjacent to the transformer for feeder 1 are subjected to more stress than the lines adjacent to the transformer for feeders 2 and 3. This is because feeder 1 has 19 load connections, while feeders 2 and 3 have 16 and 13 load connections, respectively.

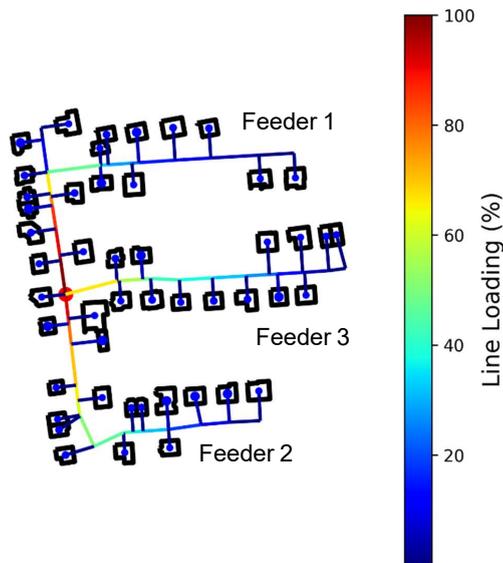


Figure 7-18 Exemplary network (N_{23988}) from the 500,000 networks showing the line loading of each at a specific time snapshot

To alleviate line loading, the load on the feeders should be minimized by compensating with distributed generators. The preceding analysis demonstrated that reverse power flows exist because of distributed generation (i.e., solar rooftop PV). Additionally, thermal loading occurs on the transformers as a result of a mismatch between supply and demand. Finally, it may be

stated that in order to minimize thermal loading on distribution lines and transformers, effective utilization of generation is necessary.

In conclusion, almost all the networks have at least one line under thermal loading violation. The thermal loading on the lines can be alleviated by reinforcing the network. In this case, reinforcement is necessary for 11% of the average number of lines for 500,000 networks. On the other hand, additional methods like demand side management, controlled battery electric vehicles charging, etc., should be integrated into the network to reduce line loading.

The following section focuses on the node voltage violations due to the integration of distributed demand and generation in the low-voltage distribution networks.

7.3 Voltage Violations

Following the transformer and line violations, the next critical parameter to evaluate are the node voltages. Voltages at the nodes should remain within a particular range in any network. In case of a voltage drop or rise at the node, undervoltage or overvoltage violations occur, respectively. As a result, two voltage violations will be addressed in this section. Table 2-2 discusses and lists the limitations for overvoltage and undervoltage violations.

Simulations performed on the 500,000 geo-referenced synthetic low-voltage networks under various future distribution and demand scenarios resulted in node undervoltage and overvoltage violations. Each network has several load nodes, depending on the predicted synthetic network configuration and the number ranges between two and 200, as seen in Figure 5-24. The voltage magnitude at each node fluctuates according to the circumstances as well as the node's placement relative to the transformer. If there is no distributed generation at the point of load, the voltage magnitude declines down the feeder. Therefore, the load farther away from the transformer encounters less voltage magnitude than the load proximal to the transformer. On the other hand, a voltage rise is experienced by a node with a generator if more power is generated than is required.

The voltage magnitude should be kept within a specified range to avoid any violations at the nodes. According to grid codes, when a typical transformer is installed in the network, the voltage at each node shall vary by no more than $\pm 5\%$. However, when a voltage regulating distribution transformer is included in the network, the allowable voltage range is between $\pm 10\%$.

However, this section will first investigate undervoltage violations concerning future distributed demand and generation in each network. Thus, the nodes that are experiencing undervoltage violations are recorded for each network. Figure 7-19 illustrates the total number of nodes that experience undervoltage violations in networks with both typical and voltage regulating distributed transformers.

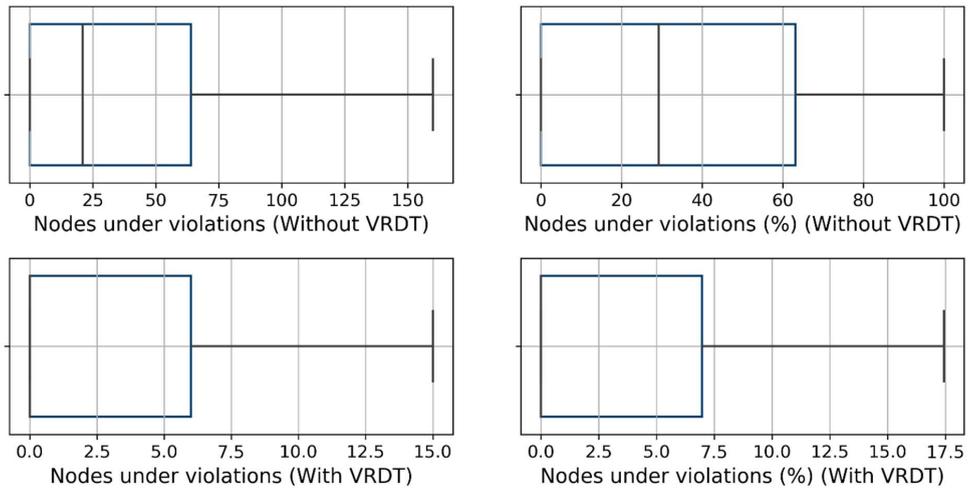


Figure 7-19 Nodes under undervoltage violations for each network for all 500,000 networks with and without voltage regulating distributed transformer (VRDT)

As illustrated in Figure 7-19, when networks are constructed using typical transformers, 25% of networks have zero nodes due to undervoltage violations. However, 50% of networks with nodes ranging from one to sixty-five that exhibited undervoltage violations. That is, 25% of networks have less than 20 nodes that experienced undervoltage violations, whereas the remaining 25% of networks have a range of 20 to 65 nodes that exhibited undervoltage violations. Additionally, the remaining 25% of networks have a wide range of total nodes ranging from 65 to 160, or around 65% to 100% of the network's nodes under undervoltage violations.

These, however, are the outcomes when networks are configured using a typical transformer. Conversely, if the network contains a voltage regulating distribution transformer, the available bandwidth for voltage deviation increases, and the number of nodes experiencing undervoltage violations reduces. This phenomenon is depicted in Figure 7-19, which clearly shows that the number of nodes with undervoltage violations has decreased significantly. For 75% of networks, the number of nodes experiencing undervoltage violations is less than 6. This number is further decreased by 90% when a voltage-regulated distributed transformer is used instead of a typical transformer. Additionally, 50% of networks implementing VRDT do not report any undervoltage violations for the nodes in those networks. That is, only 25% of networks have nodes with violations ranging from one to six or less than 7% of total nodes. Additionally, networks above the third quartile feature nodes with a count greater than six and a count less than 15, which is extremely diversified and contains more than 6% of the network's nodes but less than 17.5%.

However, a variety of alternate scenarios are examined to see if a certain deviation in the distribution demand and generation results in lesser violations. For various scenarios

employing a typical transformer, the total number of nodes subjected to undervoltage violations is illustrated in Figure 7-20.

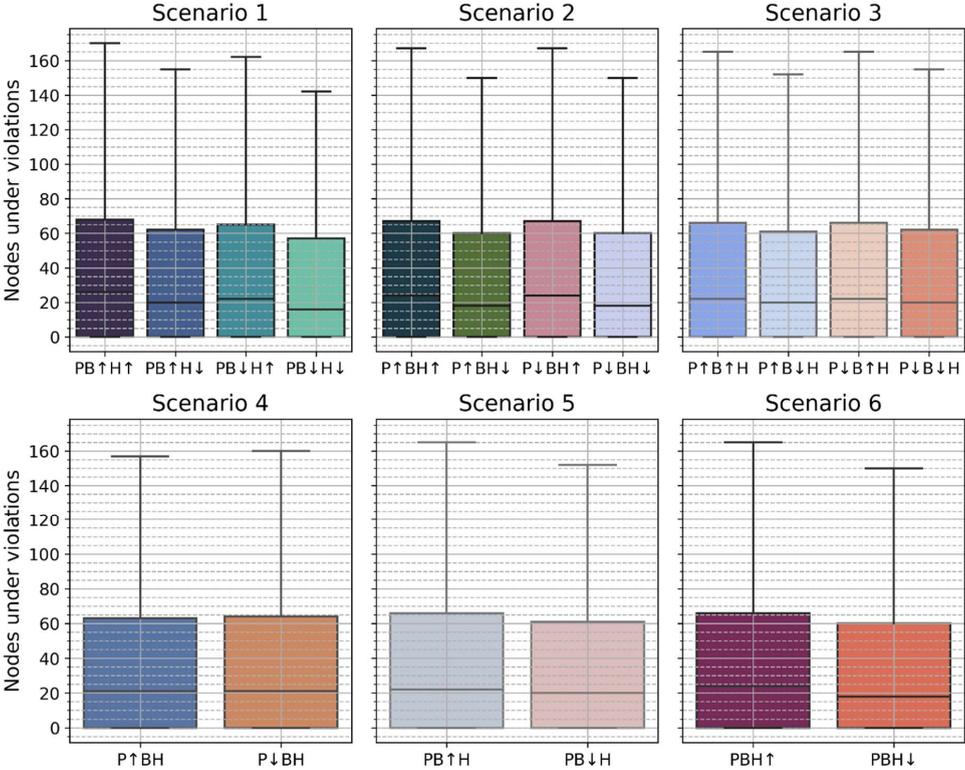


Figure 7-20 Nodes under undervoltage violations in each network for 500,000 networks for various scenarios without voltage regulating distributed transformer

Figure 7-20 showcases that practically all of the scenarios provide the same results, except for PB↓H↓, which uses fixed PV value of 53% in line with the base scenario, and battery electric vehicles and heat pumps with percentage penetration less than 10% of the base scenario, which is 72% battery electric vehicles and 74.7% heat pumps. However, the difference is negligible in comparison to other scenarios. The highest difference that was observed is 25% when the median value is used.

Therefore, it can be concluded that the network with a voltage regulating distribution transformer produces promising results. Furthermore, in several other scenarios, simulations are carried out on 500,000 networks with voltage regulating distributed transformers installed in the networks. Figure 7-21 illustrates the total number of nodes experiencing undervoltage violations for each network in 500,000 networks under 18 additional scenario cases with a voltage regulating distribution transformer.

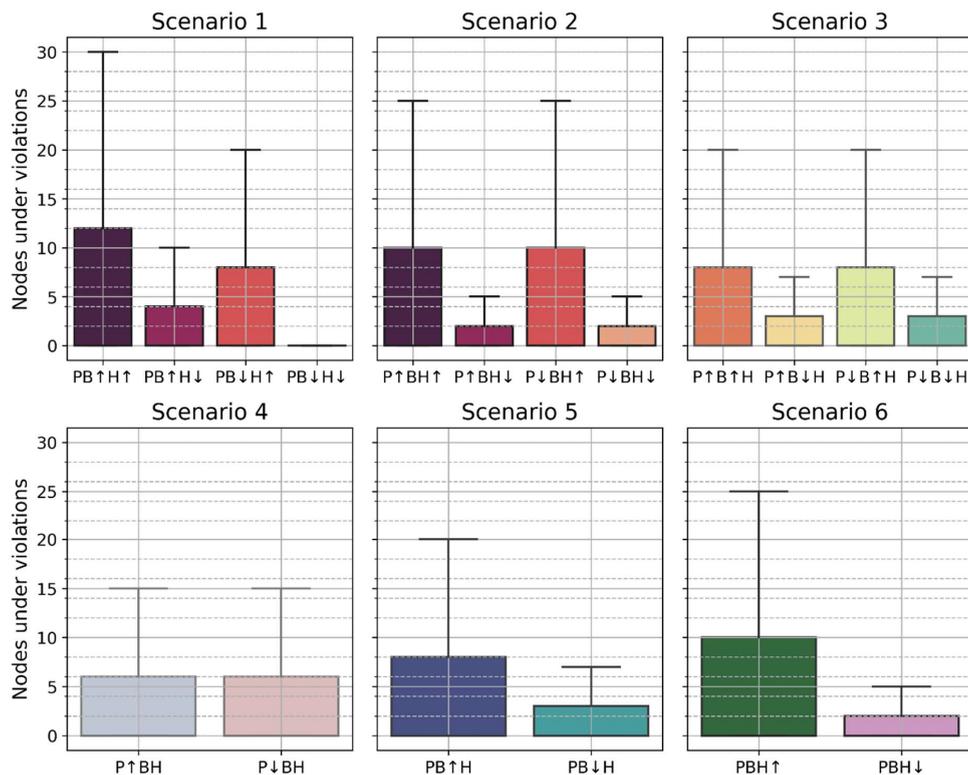


Figure 7-21 Nodes under undervoltage violations in each network for 500,000 networks for various scenarios with voltage regulating distributed transformer

As illustrated in Figure 7-21, PB↓H↓ is the most promising combination (i.e., Fixed PV (53%), –10% battery electric vehicles and heat pumps over base scenario values of 72% and 74.7%, respectively). Figure 7-21 is self-explanatory as scenario PB↓H↓ results in zero nodes violations across all 500,000 networks equipped with voltage regulating distribution transformers. This analysis helps draw a conclusion that that undervoltage violations are caused by distributed demand integration in the network. Reducing the base scenario by a small percentage results in unrestricted network operation, when battery electric vehicles and heat pumps were decreased by 10%.

Not all existing networks, however, include voltage regulating distribution transformers. As a result, additional expenditure is required to install these transformers in the networks.

Finally, the overvoltage violation is the other type of voltage violation that a node can experience in a network. Overvoltage violation occurs more frequently when there is a large amount of generation and little demand. Typically, the voltage magnitude on the feeder increases as it flows from the node proximal to the transformer to the node at the feeder's end. That is, the node at the feeder's end receives a higher voltage than the node near the transformer. This phenomenon is fundamentally opposite to the voltage drop. Overvoltage

violations are recorded by simulating 500,000 geo-referenced synthetic low-voltage networks for future distributed demand and generation. Surprisingly, none of the simulated networks designed with and without a voltage regulated distribution transformer exhibited any violations at any node. This is because the amount of generation exceeding the amount of demand in the network is insufficient to create a voltage rise of more than 5%.

In conclusion, voltage violations in relation to undervoltage violations are mostly caused by distributed demand. Future distributed demand and generation are comparable under the base scenario and the other scenario instances. While 25% of networks contain zero nodes due to undervoltage violations, 75% of networks have an average of 31 nodes that are experiencing undervoltage violations. Nonetheless, these results apply to a network configuration with a typical transformer. However, when a voltage-regulated distributed transformer is used, the total number of nodes subject to violations is reduced by 90% in the base scenario. Additionally, when different scenarios are explored, the scenario with 10% less distributed demand led to 72% battery electric vehicles and heat pumps and 74.7% heat pump penetration, causing zero undervoltage violations. Finally, overvoltage violations are not exhibited in the network nodes for all networks in the base scenario and several other scenarios involving future demand and generation irrespective of whether a typical transformer or a voltage regulated distribution transformer is employed.

After examining the violations associated with transformers, lines, and nodes in low-voltage networks, the following section discusses a sample demand side management strategy involving controlled battery electric vehicles charging at home. This attempt, when coupled with controlled battery electric vehicles charging, significantly reduced violations when compared to the absence of demand and generation management.

7.4 Sample Demand Side Management using Controlled BEV Home Charging

Using demand to balance generation is often suggested as an approach to minimize violations of reverse power flow and transformer loading. This can be accomplished through demand-side management. To determine whether demand-side management is a viable solution, a simple demand-side management strategy is used to shift the load curve to the location of PV generation. Figure 7-22 illustrates the demand and generation profiles that were evaluated.

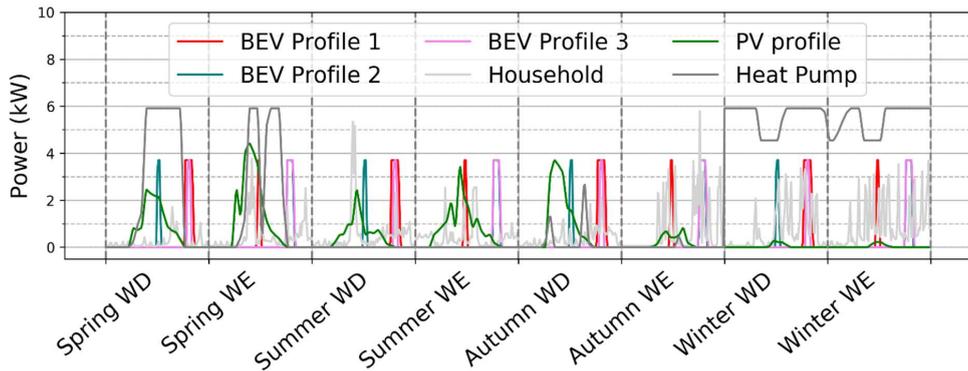


Figure 7-22 Comparison of demand and generation profiles for mismatch identification in the absence of demand side management

As illustrated in the Figure 7-22, the generation is not always utilized by demand. For instance, on an autumn weekday (WD) when PV is operating at full power, there is no usage recorded. During such periods, reverse power flow violations occur. Similarly, on the same day, battery electric vehicles 1 and 3 are charging at different times. However, if all users charged simultaneously, thermal loading of the transformers might occur. This phenomenon results in line loading. Therefore, to facilitate simple demand-side management, battery electric vehicles that are typically charged after work in the autumn should be charged before leaving for work. This implementation shifts the demand curve and matches that of the generation curve. Figure 7-23 illustrates the profiles of a moved battery electric vehicle profile 1 and battery electric vehicle profile 3 in the autumn weekday (WD).

Based on these implementations, a simulation is carried out on one of the geo-referenced synthetic networks (N_{23988}) illustrated in Section 6.2. When a power flow simulation is run under a base case scenario, it yields the results displayed in Figure 7-24. All violations associated with transformer loading, such as maximum transformer loading (TLMAX), minimum transformer loading (TLMIN), count of transformer loading instances (CTL), reverse power flow, such as count of reverse power flow instances (CRPF), and undervoltage, which includes count of under voltage violations with and without Voltage Regulating Distribution Transformer (VRDT) (CUVAN and CUVAN_VRT) are presented. Figure 7-24 illustrates that the outcomes with simplified controlled charging are preferable to those without demand-side management. The use of the generation with the controlled charging of battery electric vehicles reduces the reverse power flow due to rooftop PV. At the same time, battery electric vehicles demand is satisfied by the rooftop PV, thereby reducing thermal loading of transformers and lines. Furthermore, since voltage regulating distribution transformer increased voltage bandwidth, no undervoltage violations were observed in case of the use of voltage regulating distributed transformer.

However, if a complex demand-side management strategy is used, the results will be further improved. Moreover, combining $PB\downarrow H\downarrow$, which is essentially 53% PV, 72% battery electric

vehicles, and 74.7% heat pump, with simplified controlled battery electric vehicles charging also results in fewer violations than the base case scenario.

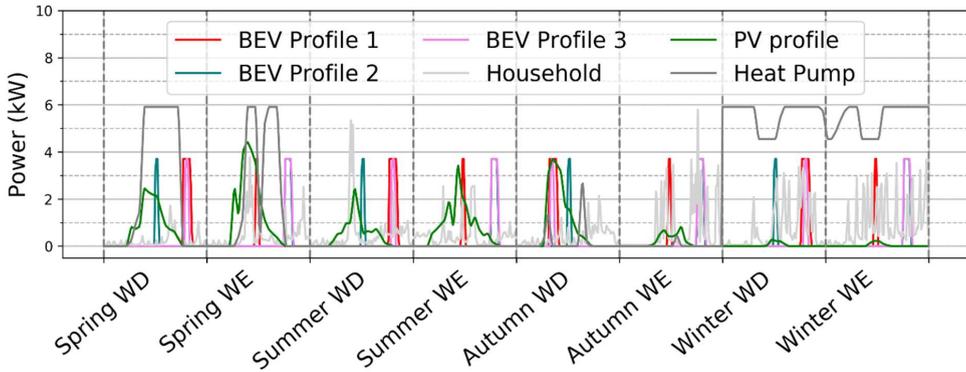


Figure 7-23 Comparison of demand and generation profiles for mismatch identification with simple demand side management

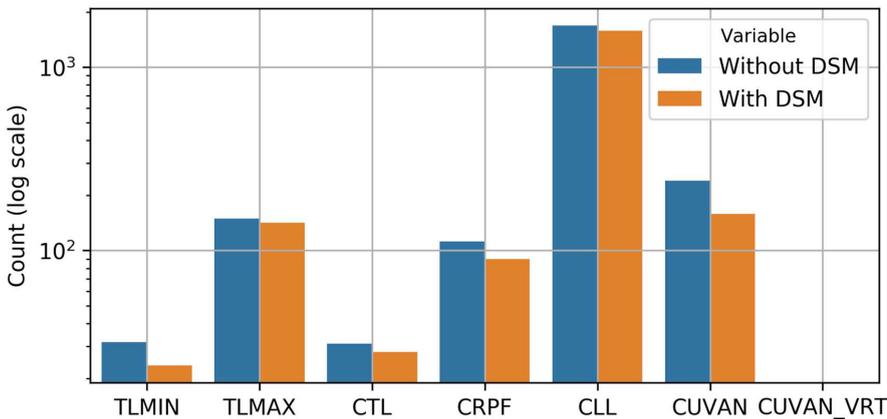


Figure 7-24 Technical violations with and without simplified controlled battery electric vehicles home charging

To summarize, the base case scenario is achievable for most networks, although for a few networks, certain technical violations may emerge. There are, however, various options for reducing violations, including demand-side management using simplified battery electric vehicles charging, and grid reinforcements.

7.5 Summary

With the objective of addressing the thesis's primary research question, this chapter examined the effects of future distributed demand and generation on low-voltage distribution networks using geo-referenced synthetic low-voltage networks developed in Chapter 5. To address the

question: “Are the distribution networks adequate for significant increase in distributed demand and generation in the future?” simulations were performed on the 500,000 geo-referenced synthetic low-voltage networks under a base case scenario with 53% PV, 80% battery electric vehicles, and 83% heat pump penetration that reduces 95% CO₂ emissions by 2050. However, the additional 18 scenarios are considered to determine the sensitivity. Under this study, various parameters corresponding to transformers, lines, and nodes are analyzed. According to this analysis, 25% of the low-voltage networks are critical due to transformer overloading. Additionally, more than 75% of low-voltage networks experience reverse power flows, but only 25% of the networks are critical and require additional attention. In addition, 11% of the lines should be reinforced on average when all 500,000 low-voltage networks are considered. Finally, on average 34% of nodes in networks should be handled for undervoltage violations. Nevertheless, these are the findings for a base case scenario in which CO₂ emissions are reduced by 95% by 2050. However, it can be concluded that by addressing the aforementioned violations, the low-voltage distribution networks will be capable of handling significant future increases in distributed demand and generation. This conclusion, addresses the following question: “Which technical violations exist with regard to future distributed demand and generation, and how could they be addressed?” Based on the analysis of the 18 additional scenarios, it can be concluded that technical violations associated with transformer overloading are a result of the high heat pump penetration. Heat pump operation, particularly in the winter, results in transformer overloading. Similarly, reverse power flows are primarily a result of high rooftop PV penetration, regardless of distributed demand. Most of the reverse power flows occur during the spring and summer. Therefore, it can be concluded that reducing or managing consumers' heat pump utilization reduces transformer thermal overloading. Additionally, reverse power flows can be reduced by curtailing solar PV or by adding battery storage and managing energy consumption in the spring and summer. This implementation has the potential to reduce the number of lines and nodes that experience thermal overloading and undervoltage violations, respectively. Finally, by incorporating voltage regulating distributed transformers into the aforementioned implementations, undervoltage violations at low-voltage nodes are minimized. This way, technical violations associated with future distributed demand and generation can be addressed, thereby preparing distribution networks for future demand and generation integration.

Key messages:

- ✚ The distribution grid model simulations are performed on 500,000 geo-referenced synthetic low-voltage networks under the base case scenario and 18 other scenario cases.
- ✚ Under reverse power flow violations, more than 75% of the networks exhibit violations, but only 25% of 500,000 transformers are considered critical.
- ✚ The primary source responsible for reverse power flows at the transformers is distributed generation or solar rooftop PV, particularly in the spring and summer.
- ✚ Under transformer thermal loading, the same 25% of 500,000 transformers are in critical condition.
- ✚ Distributed demand, particularly heat pumps, is the primary source of transformer thermal loading, particularly, in the winter.
- ✚ Under power line violations, 11% of lines need reinforcement.
- ✚ Mismatch in the distributed demand and generation profiles is the major cause of power line violations.
- ✚ Undervoltage violations affect an average of 34% and 31% of nodes for base case scenario and best-case scenario (PB↓H↓), respectively. Those, however, become null and void when a voltage regulating distributed transformer is included in the network for best-case scenario (PB↓H↓).
- ✚ However, PB↓H↓ or more precisely, 53% PV, 72% battery electric vehicles, and 74.7 percent heat pump is a better fit than the base scenario, as it can reduce some technical violations.
- ✚ Grid reinforcement measures for transformers and power lines are mandatory. Demand-side management options include controlled BEV

8 Conclusions

Residential rooftop solar photovoltaic (PV) integration into distribution networks can significantly reduce residential reliance on fossil-fuel-generated electricity. Additionally, its integration into the distribution networks lowers overall dependency on fossil fuel. Similarly, battery electric vehicles and heat pumps have the potential to minimize carbon dioxide emissions by reducing the need for fossil fuels utilizing residential rooftop PV. However, the power system implications of large-scale integration of these distributed demands and generation in the distribution networks are unclear on a national scale. Therefore, this thesis is driven by the following research questions:

Are the distribution networks equipped to handle the significant increase in distributed demand and generation in the future?

Generally, distribution networks have been constructed with the current state of the system in mind. Therefore, it is not economically viable to upgrade every network considering future integrations without proper study and analysis. Furthermore, if the distribution networks are viewed without regard for the future introduces several issues when it comes to integrating future distributed demand and generation. Therefore, this thesis addressed the following problem statement:

Which technical violations exist with regard to future distributed demand and generation, and how could they be addressed?

It is possible to evaluate the effects of future distributed demand and generation at a basic level, using test network topologies, real network topologies from any distribution system operator, or typical network topologies for the considered nation. However, evaluating future demand and generation on a broad scale is complicated, despite receiving some networks from system operators. For example, due to confidentiality and security considerations, distribution system operators are prohibited from providing geo-referenced real network topologies. As each nation has hundreds of distribution system operators, data from a single distribution system operator would prove to be insufficient. As a result, the problem statement for this thesis is as follows:

How can the effects of future distributed demand and generation on distribution grids be studied when there are no real-world networks to investigate?

Extrapolating the effects of future distributed demand and generation to the national level produces implausible findings when applied to a single network or typical networks. To address the primary impediment in studying the research question, a methodology using a data-driven approach is devised to generate geo-referenced synthetic distribution network topologies. The geo-graphical networks that have been built can be used to study the primary research question.

A package is designed to generate geo-referenced synthetic networks ranging from low-voltage distribution grids to extra-high voltage transmission grids using open data.

Several observations are made throughout the process of developing geo-referenced synthetic network topologies and determining the impact of future distributed demand and generation in the distribution networks. The derived conclusions are segmented into those related to model development and impact assessment.

The following are the major conclusions drawn from the model developed to study future distributed demand and generation in the distribution networks:

- The performed investigation revealed that most publications on distribution networks rely on real networks, test networks, and typical networks. Additionally, due to the limitations of these networks and the synthetic network models published in the literature, there is a need to study geo-referenced synthetic network topologies at the national level.
- By gathering and analyzing the graphical and electrical features of real-world networks, a clear understanding regarding the construction of geo-referenced synthetic network topologies has emerged.
- After examining real-world electrical network topologies and electrical parameters for European countries, including network indicators and reliability of supply indicators, it was determined that when data from a selected EU nation is used to develop methods, they be employed to generate geo-referenced synthetic network topologies for all European nations.
- The geo-referenced networks generator (dgnetz) generates synthetic networks using open data. Statistical validation of the geo-referenced synthetic low-voltage networks reveals a +33% inaccuracy in the total network length when compared to the actual network length. This conclusion implies that it is implausible to consider all buildings retrieved from OpenStreetMap and only residential structures should be considered.
- To address the reliability problem noted above, osmbuiltag model categorized the extracted buildings from OpenStreetMap data into residential and non-residential buildings using a state-of-the-art machine learning algorithm. Using actual data, it was validated, and the study reveals that the predicted residential buildings are 3.4% off from the actual residential buildings detected in Germany. The osmbuiltag model further classified residential buildings into single-family and multi-family dwellings, with a percentage error of +13.14% and -15.3%, respectively. Examining the results, it is concluded that with increased data quality, which includes fewer missing values and a large amount of training data, results in more accurate outcomes.

-
- dgnetz generates geo-referenced synthetic network topologies ranging from low-voltage distribution networks to extra-high-voltage transmission grids using anticipated residential building data. Due to the lack of appropriate tags in the OpenStreetMap data, high-voltage and extra-high voltage are modeled as a single grid in this case. This modelling led to the conclusion that new algorithms with additional data for classifying power-related components in high-voltage and extra-high-voltage networks enable the generation of high-voltage geo-referenced distribution networks independent of extra-high voltage networks.
 - The 500,000 geo-referenced synthetic low-voltage networks of Germany are statistically validated by comparing many topological metrics to real-world parameters from the literature. When residential buildings predicted using the osmbuildtag model are used, the overall circuit length obtained for synthetic low-voltage networks only deviates by 0.89% from the actual total low-voltage network length of Germany. Additionally, the validation of operational and geographical visualization demonstrates that dgnetz can produce nearly accurate synthetic networks.

The most important conclusions that can be drawn about the future distributed demand and generation in distribution networks are listed below:

- The analysis on 500,000 geo-referenced synthetic low-voltage networks demonstrates that future distributed demand generation with a base scenario of 53% solar rooftop PV, 80% battery electric vehicles, and 83% heat pump integration in distribution networks exhibit some violations related to transformers, lines, and electrical nodes. However, PB↓H↓ with constant PV and 10% less battery electric vehicles and heat pump, or more precisely, 53% PV, 72% battery electric vehicles, and 74.6% heat pumps, reduce the violations associated with transformer loadings, lines, and under-voltage violations compared to the base scenario.
- Additionally, the results indicate that increasing distributed generation in the distribution network results in over-voltage and reverse power flow violations. Rooftop PV, which operates at its peak in spring and summer, causes reverse power flows at the transformers. Additionally, increased distributed demand results in the thermal loading of transformers and power lines, in addition to under-voltage violations. On the other hand, heat pumps operate at their peak in Winter and cause thermal overloading.
- Over 75% of networks are in violation of reverse power flow, although only 25% of 500,000 transformers are considered critical. The same 25% of 500,000 transformers are in critical condition when subjected to thermal overloading of the transformers.
- Additionally, 11% of power lines are impacted by the integration of demand and generation in the networks. As a result, these 11% of power lines must be reinforced to ensure continued operation in the considered low-voltage networks. The integration

of demand and generation affects 34% of nodes. These, however, are reduced to 31% in the best-case scenario (PB↓H↓). Additionally, when a voltage-regulating distributed transformer is added to the network in the best-case scenario (PB↓H↓), these affected nodes become null and void.

- Additionally, it can be inferred that efficient demand-side management, and the installation of battery units can help reduce reverse power flow violations and thermal loading without requiring subjecting existing grids to restructuring. This method is only viable if end-users apply demand-side management in conjunction with battery storage to store generated energy when it is not in use. Otherwise, then reinforcement of specific networks will be required, which would incur additional costs.

Summing up, the models, geo-referenced synthetic networks, and results of thesis indicate that distribution system operators should prioritize enhancing the integration of distributed demand and generation. A feasible option would be to reinforce grids that are experiencing severe violations and foster demand side management options in distribution grids, including battery storages, controlled battery electric vehicles charging, and managing heat pump utilization. Additionally, the distribution grid model associated with the dgnetz and osmbuildtag tools can be used in the future to conduct further research on a variety of related topics, including the development of smart grids, examining the impact of new demand, such as battery-powered heavy-duty vehicles, on distribution grids, integrating high penetration of renewable energy sources in medium-voltage networks, and grid reinforcement studies.

9 Summary

To determine the impacts of future distributed demand and generation in the distribution networks under the scenario of reducing 95% of CO₂ emissions by 2050, this research develops geo-referenced synthetic distribution network for the entire nation. However, this thesis is divided into the following chapters to draw conclusions from the research that was conducted:

In Chapter 1, distributed demand and generation are introduced as alternatives to fossil fuel usage and as a means of reducing greenhouse gas emissions. In addition, the research question of whether existing distribution networks are adequate for the integration of future demand and generation is introduced along with problem statements. Finally, the overall thesis structure is introduced by describing the primary contribution of each chapter.

Chapter 0 discussed the challenges and opportunities involved in distribution grids. Distributed demand, which includes battery electric vehicles and heat pumps, and distributed generation or the residential rooftop solar PV are introduced in detail. Adopting such technologies typically alters the operation of power systems, resulting in network violations. In this context, various operational limiting factors and associated violation limits, methods of analysis, tools, and network topologies are studied. Based on the literature reviewed, it is concluded that a significant penetration of distributed demand and generation will interrupt network operation by causing thermal loadings, voltage violations, and power flow violations. Furthermore, it is identified that there is insufficient infrastructure available to analyze and plan the effects of future demand and generation in the distribution networks. Therefore, synthetic models are studied. This chapter identified the potential to develop geo-referenced synthetic distribution networks based on the identified limitations to answer the research question of this thesis.

Following the decision in Chapter 0 to develop geo-referenced synthetic distribution network topologies, Chapter 3 explored several graphical and electrical parameters. By examining these parameters, several attributes required for the generation and validation of the geo-referenced synthetic networks are uncovered. Firstly, radial networks and open-ring are identified as suitable for low-voltage and medium voltage networks, respectively. Secondly, betweenness centrality, and average shortest path length are utilized in generating the networks. Finally, clustering coefficient, number of nodes, number of edges, total network length, and node degree probability distribution are used to validate the developed geo-referenced synthetic networks. An analysis on the real-world network indicators for EU nations revealed that open data of Germany allows the development of algorithms capable of generating geo-referenced synthetic networks that can be applied in any other EU nations. This chapter ends by choosing Germany as a test to develop algorithms for geo-referenced synthetic networks based on these findings.

Since it was decided that algorithms must be developed to generate geo-referenced synthetic networks by utilizing data from Germany, Chapter 4 focuses on the open data of Germany for developing geo-referenced synthetic network topologies. Data regarding the geo-graphical

locations such as buildings, power lines, transformers, substations, and installed renewable energy sources are extracted and analyzed. Analyzing the open data reveals that buildings which are necessary for the estimation of geo-referenced synthetic networks lack labels corresponding to the building types, such as residential, non-residential, single-family house, multi-family house, etc. Therefore, a classification model is developed to address this limitation, which classified approx. 19 million buildings as residential buildings out of approx. 29 million buildings. These 19 million residential buildings are classified into 14 million single-family houses, 4 million multi-family houses, and one million apartment buildings. However, statistical validation reveals a 3.4% variance of residential buildings from actual buildings in Germany. Additionally, a percentage error of 10.14% is observed for single-family houses and –15.3% percent for multi-family houses and apartments combined. In addition to these open data, synthetic data such as load profiles for buildings, battery electric vehicle profiles, heat pump profiles, and residential rooftop solar PV profiles are established.

Analyzing the data required for estimating geo-referenced synthetic networks, Chapter 5 demonstrates the methodology for implementing geo-referenced synthetic distribution networks. This chapter discussed strategies for progressively generating synthetic networks ranging from low-voltage distribution grids to extra-high-voltage transmission grids. Firstly, low-voltage networks are modeled by developing algorithms that can assist with various tasks, such as clustering residential buildings to match low-voltage networks, clustering buildings by limiting cumulative peak load, collecting road infrastructure to generate power lines, combining nodes and edges, constructing multiple sub-graphs into a single graph, identifying transformer locations, and reinforcing graphical networks. These graphical networks are then transformed to electrical networks by allocating transformer types, and line types to nodes and edges in the networks. Secondly, similar to low-voltage network development, medium-voltage network development was driven by the introduction of graphical network and electrical network development. These steps made extensive use of the algorithm designed for low-voltage networks. Several new methods were also developed, including clustering with known centroids plus predicting unknown centroids, which helped with identifying medium-voltage transformer locations that are not already identified in OpenStreetMap data, and estimating radial medium-voltage network as some of the developed algorithms. Just like low-voltage networks, transformer and line types are assigned to the graphical network to convert it to medium-voltage electrical networks. Finally, high-voltage distribution networks and extra-high-voltage transmission networks are constructed by combining them into a single network because some countries regard high-voltage to be a distribution network while others consider it to be a transmission network. In addition to the development of synthetic networks, three types of validation methodologies for geo-referenced low-voltage synthetic networks are presented, which are statistical, operational, and geographical network comparison validation. These validations demonstrated that the networks formed are extremely accurate. The comparisons of the overall network for the generated synthetic networks and the actual network length revealed a difference of only 0.89%. Finally, the chapter provides 500,000 geo-referenced synthetic low-voltage networks throughout Germany, as well as a working

approach for synthesizing geo-referenced synthetic medium-voltage, high- and extra-high-voltage networks as a single network.

In Chapter 6, a distribution grid model is developed using geo-referenced synthetic network topologies to estimate the impact of future distributed demand and generation in the low-voltage distribution networks. Here, scenarios for future integration of distributed demand and generation are included with the aim of reducing CO₂ emissions by 95% by 2050. Additionally, technologies that play a significant role in distribution networks are identified, including residential rooftop solar PV, battery electric vehicles, and heat pumps and the percentage penetration of these technologies that can reduce CO₂ emissions by 95% is used as a base case scenario. In this base case scenario, 53% PV, 80% battery electric vehicles, and 83% heat pumps are considered. However, to conduct sensitivity analysis, six additional scenarios are included, with four cases in the first three scenarios and two cases in the next three scenarios, totaling 18 scenarios. Finally, to demonstrate the distribution grid model's application, one network is chosen from generated 500,000 geo-referenced synthetic networks.

Finally, Chapter 7 simulated the deployment of the distribution grid model to 500,000 geo-referenced low-voltage networks using 19 scenario cases including base case scenario. Here, various operational limiting factors are analyzed for each network and violations are logged. Under the analysis, several conclusions are made. First, 75% of the networks are under violations, but 25% of 500,000 transformers are considered critical in case of reverse power flows. Second, the same 25% of 500,000 transformers are in critical condition for transformer loading as well. Thirdly, 11% of power lines and 34% of nodes are under thermal overloading, and under voltage violations, respectively. With these results, it can be concluded that the 500,000 low-voltage networks undergo certain violations with respect to future distributed demand and generation. However, grid reinforcement measures for transformers and power lines are a potential solution. Furthermore, demand-side management options like controlled battery electric vehicles charging, controlled heat pump utilization, and battery storage and consumption can reduce grid reinforcement measures to some level.

In this manner, a comprehensive answer in response to the research question was established by implementing state of the art artificial intelligence techniques to classify high volume and variety of data into residential and non-residential buildings, modeling geo-referenced synthetic networks by developing new algorithms, and analyzing the effects of future distributed demand and generation across the country.

Appendix

A. Distribution Grid Operating Voltage Levels for EU Nations

Distribution system operators operate at different voltage levels for distributing power to end consumers. Most of the EU nations operate at low-, medium-, and high-voltage. However, some countries include, Cyprus, Estonia, France, Italy, Lithuania, and Latvia operate only low- and medium-voltage lines. In those countries, high-voltage level is considered as the transmission grid. Figure 9-1 illustrates the distribution grid voltage level used in EU nations.

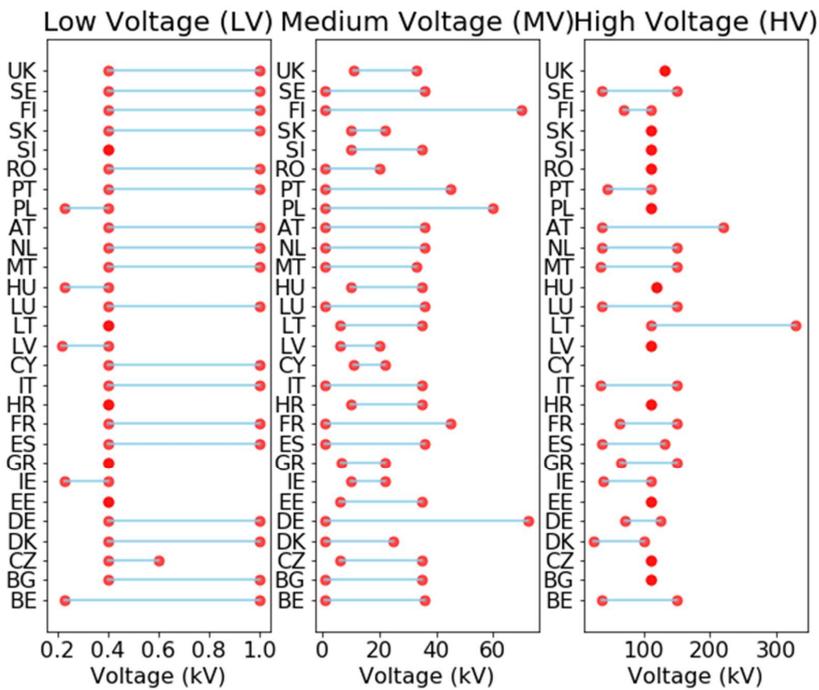


Figure 9-1 Voltage levels used in EU nations

B. Corine Land Cover (CLC) Features

Classes representing the land cover type in the CLC dataset is shown in Table 9-1.

Table 9-1 Corine land cover classes

Class	Meaning
111	Continuous urban fabric
112	Discontinuous urban fabric
121	Industrial or commercial units
122	Road and rail networks and associated land
123	Port areas
124	Airports
131	Mineral extraction sites
132	Dump sites
133	Construction sites
141	Green urban areas
142	Sport and leisure facilities
211	Non-irrigated arable land
212	Permanently irrigated land
213	Rice fields
221	Vineyards
222	Fruit trees and berry plantations
223	Olive groves
231	Pastures
241	Annual crops associated with permanent crops
242	Complex cultivation patterns
243	Land principally occupied by agriculture
311	Broad-leaved forest
312	Coniferous forest
313	Mixed forest
321	Natural grasslands
322	Moors and heathland
323	Sclerophyllous vegetation
324	Transitional woodland-shrub
331	Beaches, dunes, sands
332	Bare rocks
333	Sparsely vegetated areas
334	Burnt areas
335	Glaciers and perpetual snow
411	Inland marshes
412	Peat bogs

C. Uncertainties in OSM Building Labels

Due to the fact that OSM data is contributed by a large number of users worldwide, there are several uncertainties in the data's labeling. This is due to language differences and, in some cases, spelling errors. Table 2 summarizes some of the ambiguities found in OSM.

Table 9-2 Uncertainties in OSM building labels

Building label	Differing representation
garage	garages
farm	farm_auxiliary
Terrace	Terrasse
House	Haus Hause house
Youth Centre	Jugendzentrum
Nursing home	Pflegeheim
deconstructed	deconstructed

D. Renewable Energy Sources (Open Data)

The renewable energy sources (RES) locations and their installed capacities in Germany are taken from OPSD [200]. Firstly, RES locations with their capacities installed in the low-voltage network for Germany is illustrated in Figure 9-2.

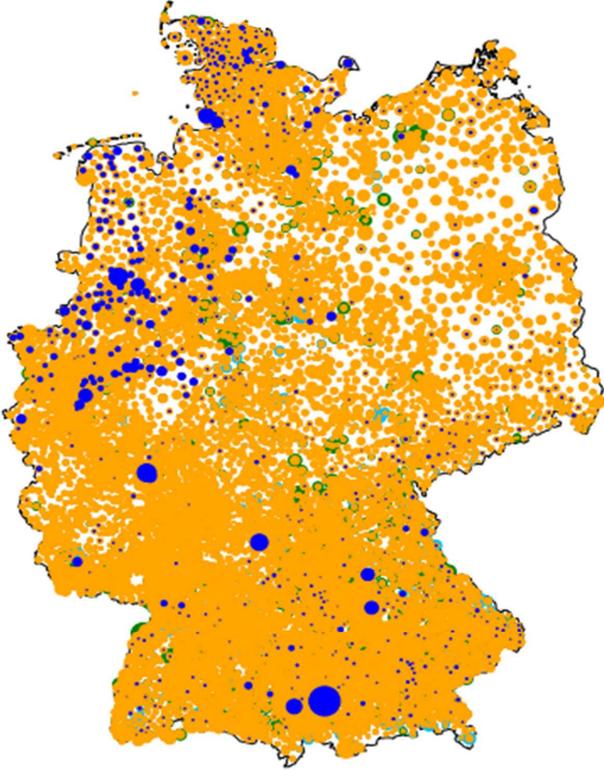


Figure 9-2 Renewable energy sources in low-voltage networks

Moreover, the installed capacities for each technology with maximum peak capacity for each technology is shown in Table 9-3. From the Figure 9-2, it is clear that solar PV is dominating in the low-voltage networks.

Table 9-3 Total RES installed in low-voltage networks

Technology	MWp	Total Installed capacity (GW)
------------	-----	-------------------------------

Solar	6.90	23.07
Bioenergy	6.02	0.54
Wind	3.30	0.03
Hydro	3.20	0.22
Storage	$0.07 * 10^{-1}$	$7.50 * 10^{-6}$

RES locations with their capacities installed in medium-voltage networks for Germany is illustrated in Figure 9-3.

● Solar
 ● Wind
 ● Hydro
 ● Bioenergy
 ● Geothermal
 ● Storage

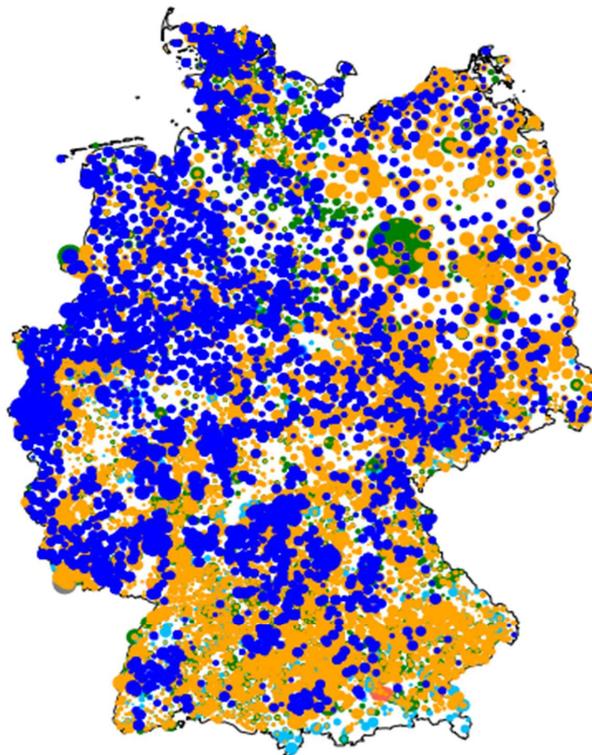


Figure 9-3 Renewable energy sources in medium-voltage networks

However, the installed capacities for each technology with maximum peak capacity for each technology in medium-voltage network is shown in Table 9-4. From the Figure 9-3 Figure 9-2, it is clear that wind is dominating in the medium-voltage networks

Table 9-4 Total RES installed in medium-voltage networks

Technology	MWp	Total Installed capacity (GW)
Solar	24.71	15.11
Bioenergy	135.00	5.96
Wind	24.00	22.22
Hydro	6.23	0.93
Geothermal	7.00	0.04

E. Geo-referenced Synthetic Low-voltage Networks

A few of the 500,000 geo-referenced synthetic low-voltage networks generated are shown in Figure 9-4, Figure 9-5, and Figure 9-6.

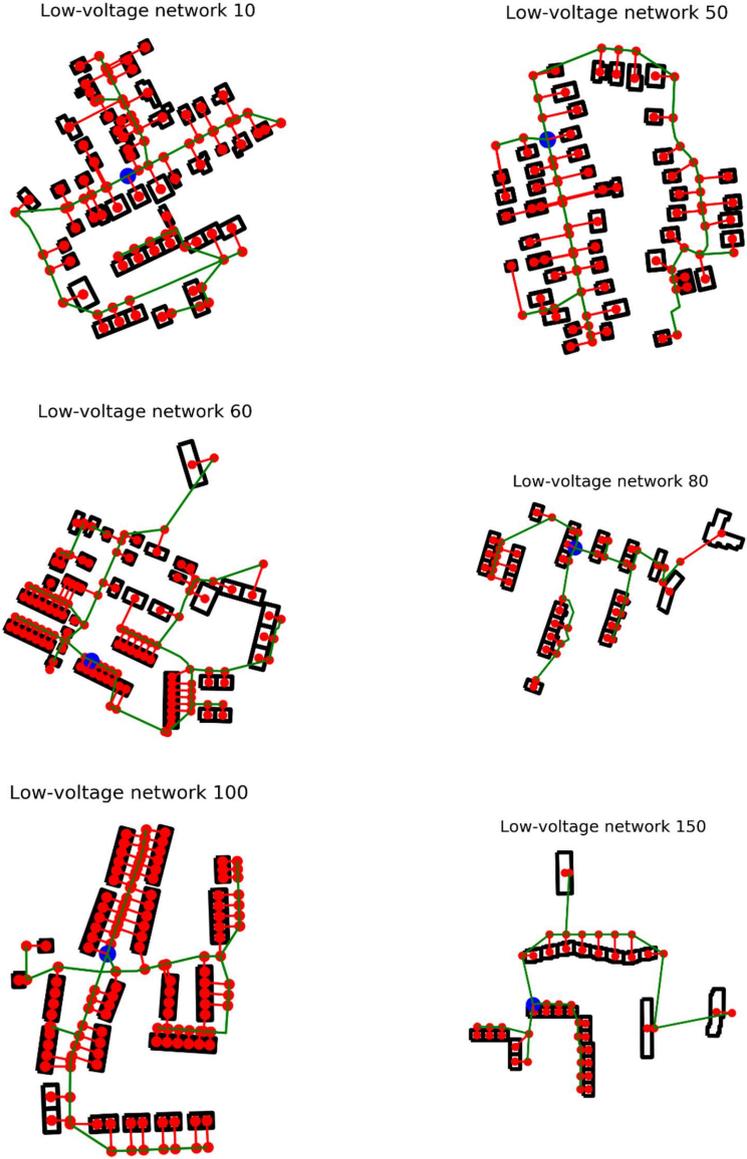


Figure 9-4 Few geo-referenced synthetic low-voltage distribution networks -I

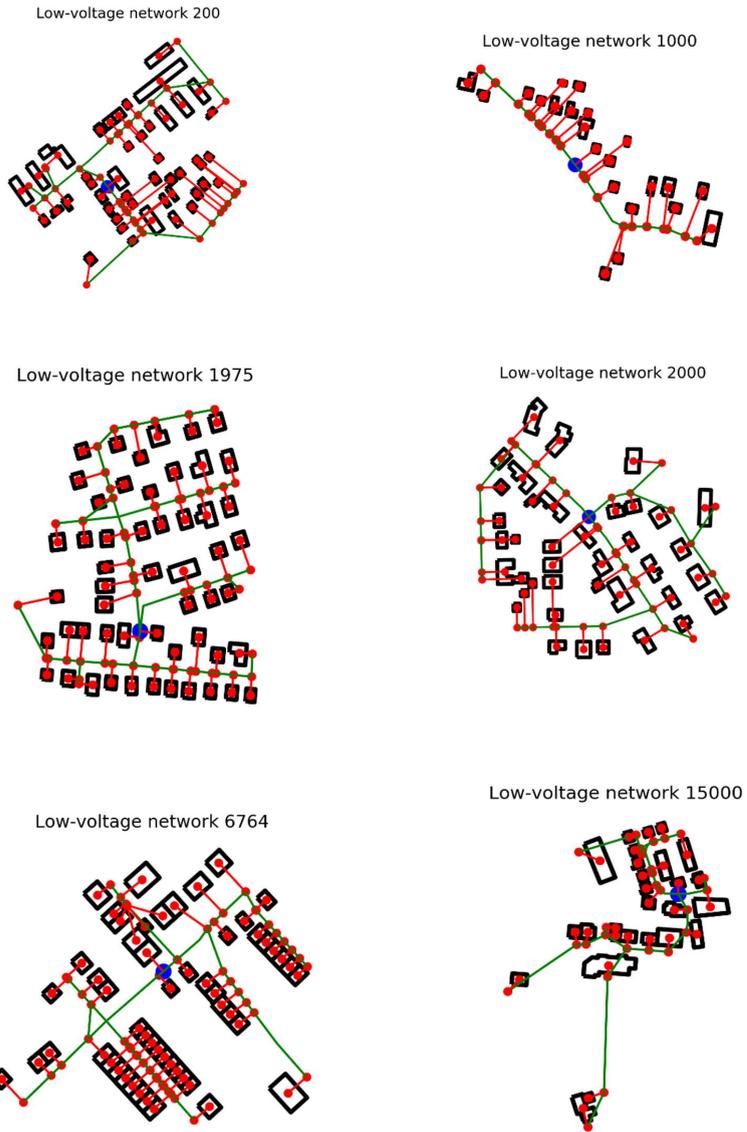


Figure 9-5 Few geo-referenced synthetic low-voltage distribution networks -II

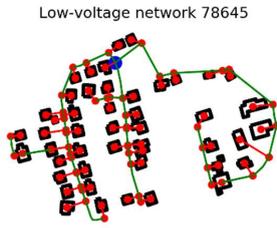
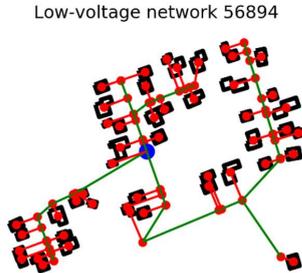
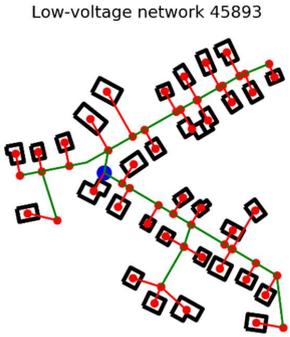
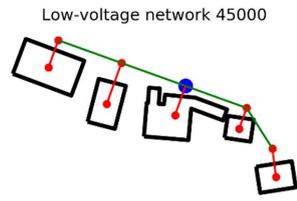
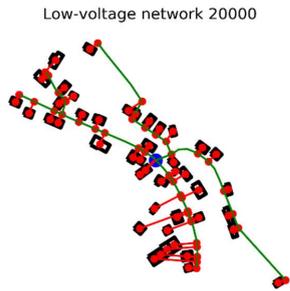


Figure 9-6 Few geo-referenced synthetic low-voltage distribution networks -III

F. Standard Medium-voltage Cable Types

Standard medium-voltage cable and overhead power line types available in Germany are shown in

Type	I (A)	R (Ohm/km)	L (mH/km)	C (μ F/km)
NA2XS2Y 3x1x185	357	0.1664	0.38	0.41
NA2XS2Y 3x1x240	417	0.125	0.36	0.47
NA2XS2Y 3x1x300	466	0.1	0.35	0.495
NA2XS2Y 3x1x400	535	0.078	0.34	0.57
NA2XS2Y 3x1x500	609	0.061	0.32	0.63
NA2XS2Y 3x1x150	319	0.206	0.4011	0.24
NA2XS2Y 3x1x240	417	0.13	0.3597	0.304
NA2XS2S(FL)2Y 3x1x300	476	0.1	0.37	0.25
NA2XS2S(FL)2Y 3x1x400	525	0.078	0.36	0.27
NA2XS2S(FL)2Y 3x1x500	598	0.06	0.34	0.3

G. Figures

Figure 1-1 Standard power grid structure from generation to demand	2
Figure 1-2 Overview of thesis	4
Figure 1-3 Schema for estimating geo-referenced synthetic distribution network topologies..	6
Figure 2-1 Overview of distributed demand and generation in distribution networks.....	11
Figure 2-2 Classification of operational limiting factors in the distribution networks	13
Figure 2-3 Voltage magnitude along the feeder in case of generation and load (left), and only load (right). Adapted from [15].....	14
Figure 2-4 Distribution of voltage band between medium-voltage and low-voltage networks with a standard distribution transformer, Adapted from [24].....	16
Figure 2-5 Distribution of voltage band between medium-voltage and low-voltage with a voltage regulated distribution transformer (VRDT), Adapted from [24].....	16
Figure 2-6 Percentage publications for a different combination of technologies integrated into the distribution networks; the combination of technologies include Electric Vehicles (EV), Photovoltaics (PV), Renewable Energy Sources (RES), Energy Storage Systems (ESS), and Heat Pumps (HP)	20
Figure 2-7 Total number of publications referring to Operational Limiting Factors (OLFs) with Distributed Demand and Generation (DDG) which include Electric Vehicles (EV), Photovoltaics (PV), Renewable Energy Sources (RES), Energy Storage Systems (ESS), and Heat Pumps (HP).....	21
Figure 2-8 Flow diagram for assessing Operational Limiting Factors (OLFs) violations using a deterministic method	24
Figure 2-9 Flow diagram for assessing Operational Limiting Factors (OLFs) violations using a probabilistic method	25
Figure 2-10 Assessment methods addressed in the literature	27
Figure 2-11 Network topologies considered in the publications	31
Figure 2-12 Countries with synthetic transmission grid models.....	34
Figure 2-13 Countries with synthetic distribution grid models	36
Figure 3-1 Sample graph with its adjacent matrix	42
Figure 3-2 Power network graph structures	44
Figure 3-3 Total network length (km) considering low-, medium-, and high-voltage levels for European nations [1]	53
Figure 3-4 Total low-voltage network length (km) representing underground cable and overhead line length [1]	54

Figure 3-5 Total medium-voltage circuit length (km) representing underground cable and overhead line length [1]	55
Figure 3-6 Total high-voltage network length represented in kilometers (km) [1]	56
Figure 3-7 Total number of low-voltage transformers [1].....	57
Figure 3-8 SAIDI (Minutes lost in 2014) [1].....	58
Figure 3-9 SAIFI (Average interruptions per consumer in 2014) [1]	59
Figure 3-10 Correlation between network indicators and reliability indicators	60
Figure 3-11 Correlation between considered EU nations.....	61
Figure 4-1 OSM data example for representation in XML format.....	67
Figure 4-2 Fundamental attributes for node, way, and relation objects in the OSM data.....	67
Figure 4-3 Geographical representation of points, lines, and polygons for Germany	69
Figure 4-4 Count of attributes in point dataset for Germany.....	70
Figure 4-5 Distribution of power attribute tags	71
Figure 4-6 Voltage tags for point objects	72
Figure 4-7 Count of polygons according to various attributes	73
Figure 4-8 Distribution of voltage tags on polygon objects of power components	74
Figure 4-9 Spatial representation of extracted transformer and substation locations for low-, medium-, and high-voltage levels	75
Figure 4-10 OSM identified data comparison with existing one for transformers and substations count [201]	76
Figure 4-11 Voltage-specific tags for power line objects.....	77
Figure 4-12 OSM extracted HV and EHV power lines	78
Figure 4-13 OSM identified power line length comparison with existing data for HV and EHV levels.....	79
Figure 4-14 Building labels in filtered OSM data.....	80
Figure 4-15 Methodology for predicting building types on building footprints of Germany.....	82
Figure 4-16 10 m raster layer containing building height information on the left and box plot showing range of building heights on the right.....	83
Figure 4-17 CORINE land cover data for Germany	84
Figure 4-18 Number of buildings with living spaces for different building types per 100m X 100m grid cell.....	85
Figure 4-19 OSM buildings mapped with census 100 X 100 m ² grid cells.....	87
Figure 4-20 Box plot representing area of buildings labeled as garages.....	88

Figure 4-21 Percentage labeled and unlabeled data before and after preprocessing	89
Figure 4-22 Distribution of labels in the target class	90
Figure 4-23 Methodology employed for building type classification.....	90
Figure 4-24 Comparison of actual residential buildings count with predicted count.....	94
Figure 4-25 Residential buildings count comparison for each federal state of Germany	95
Figure 4-26 Predicted single- and multi-family houses comparison with official statistics.....	96
Figure 4-27 Single-family houses count comparison for each federal state of Germany	97
Figure 4-28 Household load profiles for one, two, three, four, and five persons.....	98
Figure 4-29 Households as a percentage of total dwellings in Germany.....	99
Figure 4-30 Heat pump load profiles for one, two, three, four, and five-persons in a household	100
Figure 4-31 Battery electrical vehicle load profiles for three different clusters.....	102
Figure 4-32 Residential PV profiles at a location of East, West, North, and South of Germany	103
Figure 5-1 A block diagram representation of the modeling approach used to estimate geo-referenced synthetic network topologies.....	106
Figure 5-2 Schema for low-voltage network modeling	109
Figure 5-3 An example showing the application of DBSCAN	110
Figure 5-4 Time required to cluster data points using K-Means (Scikit-learn + Joblib [229]) and K-Means (DASK-ML [228]).....	111
Figure 5-5 Distribution of low-voltage networks to administrative districts in Germany.....	111
Figure 5-6 Residential buildings in the clusters.....	114
Figure 5-7 Graph generated by implementing Algorithm 4 on the considered cluster	116
Figure 5-8 A graphical network generated for a cluster ($K23988$).....	118
Figure 5-9 An illustration of the reinforcement on the graphical network ($K23988$).....	120
Figure 5-10 Peak power of the estimated 500,000 networks after assigning load profiles for each dwelling	122
Figure 5-11 Low-voltage transformer type distributed to the estimated 500,000 low-voltage networks.....	122
Figure 5-12 A synthetic low-voltage network among 500,000 geo-referenced synthetic networks showing line-loading at a specific time.....	123
Figure 5-13 Low-voltage transformer locations extracted from geo-referenced synthetic low-voltage networks estimated in the previous section	125
Figure 5-14 Medium-voltage transformer locations retrieved from OpenStreetMap	126

Figure 5-15 Schema for medium-voltage network modeling	128
Figure 5-16 Sample implementation of the Algorithm 8	130
Figure 5-17 Road infrastructure in close proximity to the connection points, retrieved using Algorithm 3	131
Figure 5-18 Ring topology operating as radial network (open ring) by disconnecting a line using circuit breaker.....	132
Figure 5-19 Final medium-voltage graphical network for considered cluster (<i>K4490</i>)	133
Figure 5-20 A medium-voltage network among 4500 networks showing line-loading at a specific time	135
Figure 5-21 Overview of the GridKit process [Adapted from [180]]	136
Figure 5-22 Synthetic high- and extra-high voltage grid extracted using GridKit tool [180]..	137
Figure 5-23 An overview of proposed validation measures.....	139
Figure 5-24 A box plot representing the total number of nodes in each network for 500,000 geo-referenced synthetic low voltage networks	140
Figure 5-25 Range of total number of nodes of real low-voltage networks for various rural, semi-urban, and urban areas in Germany	141
Figure 5-26 A box plot illustrating the mean node degree of 500,000 geo-referenced synthetic low voltage networks	142
Figure 5-27 Edge related characteristics extracted from 500,000 geo-referenced synthetic low voltage networks	143
Figure 5-28 Comparison of existing total low-voltage network length for Germany with the total network length of geo-referenced synthetic networks estimated.....	144
Figure 6-1 Block diagram representation of the distribution grid model.....	150
Figure 6-2 Percentage of future distributed demand and generation in distribution grids by 2050; A base case scenario is represented as PBH; where P, B, H corresponds to PV, BEV, and HP respectively	152
Figure 6-3 Percentage penetration of PV, battery electric vehicles, and heat pumps for Scenarios 1 (S1), 2 (S2) and 3 (S3).....	154
Figure 6-4 Percentage penetration of PV, battery electric vehicles, and heat pumps for Scenarios 4 (S4), 5 (S5) and 6 (S6).....	155
Figure 6-5 Transformer loading of the geo-referenced synthetic low-voltage network (<i>N23988</i>)	158
Figure 6-6 Line loading on each line for a specific time instance of simulated geo-referenced synthetic low-voltage network (<i>N23988</i>)	159
Figure 6-7 Voltage magnitude at a random bus in the simulated geo-referenced synthetic low-voltage network (<i>N23988</i>).....	160

Figure 7-1 Maximum transformer loading observed for 500,000 networks under base case scenario (PBH).....	165
Figure 7-2 Maximum transformers loading (%) observed for all the geo-referenced synthetic low-voltage networks for 18 various scenario cases	167
Figure 7-3 Transformer loading seen in Winter when more heat pumps are operating for an exemplary network (N23988).....	168
Figure 7-4 Transformers under violations (Hours) and percentage of total time (8760 hours) for 500,000 geo-referenced synthetic low-voltage networks under base scenario (PBH).....	170
Figure 7-5 Transformers under overloading violations (Hours) observed for all the 500,000 geo-referenced synthetic low-voltage network transformers for 18 various scenario cases	171
Figure 7-6 Average time transformers overloaded (Hours) for 500,000 networks for 19 scenario cases.....	172
Figure 7-7 Minimum transformer loading observed for 500,000 geo-referenced synthetic networks under base case scenario (PBH).....	173
Figure 7-8 Minimum transformer loading (%) observed for all the considered low-voltage networks for 18 various scenario cases	175
Figure 7-9 Average minimum transformer loading for 500,000 networks for 19 scenario cases	176
Figure 7-10 Transformers under reverse power flow (Hours) for a year for 500,000 low-voltage transformers for base case scenario (PBH).....	178
Figure 7-11 Transformer reverse power flow (Hours) observed for all the considered low-voltage network transformers for 18 various scenario cases.....	179
Figure 7-12 Average time reverse power flows at 500,000 transformers for 19 scenario cases	180
Figure 7-13 Comparison of transformer thermal overloading duration with reverse power flow duration	181
Figure 7-14 Lines under thermal loading for each network for 500,000 geo-referenced synthetic low-voltage networks for base case scenario (PBH).....	183
Figure 7-15 Total lines under thermal loading observed for all the considered low-voltage networks for 18 various scenario cases.....	184
Figure 7-16 Line loading seen on a line from exemplary network (N23988) simulated in Chapter 6	185
Figure 7-17 Average lines under thermal loading for 500,000 networks for 19 scenario cases	186
Figure 7-18 Exemplary network (N23988) from the 500,000 networks showing the line loading of each at a specific time snapshot.....	186

Figure 7-19 Nodes under undervoltage violations for each network for all 500,000 networks with and without voltage regulating distributed transformer (VRDT).....	188
Figure 7-20 Nodes under undervoltage violations in each network for 500,000 networks for various scenarios without voltage regulating distributed transformer	189
Figure 7-21 Nodes under undervoltage violations in each network for 500,000 networks for various scenarios with voltage regulating distributed transformer	190
Figure 7-22 Comparison of demand and generation profiles for mismatch identification in the absence of demand side management.....	192
Figure 7-23 Comparison of demand and generation profiles for mismatch identification with simple demand side management.....	193
Figure 7-24 Technical violations with and without simplified controlled battery electric vehicles home charging.....	193
Figure 9-1 Voltage levels used in EU nations	205
Figure 9-2 Renewable energy sources in low-voltage networks	208
Figure 9-3 Renewable energy sources in medium-voltage networks	209
Figure 9-4 Few geo-referenced synthetic low-voltage distribution networks -I	211
Figure 9-5 Few geo-referenced synthetic low-voltage distribution networks -II	212
Figure 9-6 Few geo-referenced synthetic low-voltage distribution networks -III	213

H. Tables

Table 2-1 Summary of operational limiting factors accompanied by performance indicators.	12
Table 2-2 Voltage limits according to European standards EN 50160 [22, 23].....	15
Table 2-3 Operational limits for considered grid component	17
Table 2-4 Literature study on the impact of considered distributed demand and generation over operational limiting factors (OLFs) in the distribution networks	19
Table 2-5 Overview of methods used for assessing Operational Limiting Factors (OLFs) in the literature	23
Table 2-6 Literature overview with different network topologies	29
Table 2-7 Institute of Electrical and Electronics Engineers (IEEE) test networks, Adapted from [152].....	30
Table 3-1 Graphical characteristics corresponding to nodes, edges, and graph structure	43
Table 3-2 List of real-work networks indicators and system reliability indicators taken from distribution system operators observatory [1]	51
Table 4-1 Key-value pairs for filtering the OSM data	68
Table 4-2 Transformer point labels in OSM data	72
Table 4-3 Transformer polygon labels in OSM data.....	74
Table 4-4 Power line labels in OSM data.....	77
Table 4-5 Power grid characteristics in Germany that include, nominal voltage, total length, and voltage control according to voltage levels [Adapted from [202]].....	78
Table 4-6 Number of grid cells per building type in Germany.....	85
Table 4-7 Uncertainties in OSM building labels	87
Table 4-8 F1 scores for implementation models	91
Table 4-9 Total buildings per building type predicted.....	91
Table 4-10 Adaptation of OSM building labels to proposed labels	92
Table 4-11 Predicted house types count and percentage	93
Table 4-12 Average time and distance travelled by the battery electric vehicle users in different clusters.....	101
Table 5-1 Frequently used MV/LV transformer capacities in Germany [Adapted from [154]]	107
Table 5-2 Standard low-voltage cable types in Germany [Adapted from [155]].....	107
Table 5-3 Final user input data for estimating synthetic low-voltage networks	108

Table 5-4 Assumptions considered for allocation of transformer type to low-voltage network	122
Table 5-5 Standard medium-voltage transformers available in Germany [Adapted from [133]]	126
Table 5-6 Final user input data for estimating geo-referenced synthetic medium-voltage networks.....	127
Table 5-7 Standard transformer types at high- and extra-high voltage level [183] [201].....	137
Table 5-8 Standard power lines at high- and extra-high-voltage level [183] [201]	138
Table 5-9 Power flow parameters and target set for operational validation	145
Table 6-1 Distributed demand and generation mix in the energy systems by 2050 in Germany	151
Table 6-2 Consideration of Scenario 1 (S1), 2 (S2), and 3 (S3) with four cases	153
Table 6-3 Consideration of Scenario 4 (S4), 5 (S5), and 6 (S6) with two cases	154
Table 6-4 Parameters used to apply the distribution grid model to an exemplary network..	156
Table 9-1 Corine land cover classes.....	206
Table 9-2 Uncertainties in OSM building labels	207
Table 9-3 Total RES installed in low-voltage networks	208
Table 9-4 Total RES installed in medium-voltage networks	209

I. Abbreviations

PV	Photovoltaics
OLF	Operational Limiting Factor
BEV	Battery Electric vehicle
CO ₂	Carbon dioxide
HP	Heat Pump
DDG	Distributed Demand and Generation
GHG	Greenhouse Gas Emission
DG	Distributed generator
DER	Distributed Energy Resources
PV	Photovoltaic
DSO	Distribution System Operator
LV	Low-voltage
MV	Medium-voltage
HV	High-voltage
EHV	Extra-high-voltage
OSM	OpenStreetMap
OLF	Operating Limiting Factor
RPF	Reverse Power Flow
VRDT	Voltage Regulating Distribution Transformer
VSI	Voltage Stability Index
VDI	Voltage Deviation Index
OPSD	Open Power System Data
XML	Extensible markup Language
PBF	Protocolbuffer Binary Format

J. References

- [1] G. Pretico, M. Flammini, N. Andreadou, S. Vitiello, G. Fulli, and M. Masera, "Distribution system operators observatory 2020," *Publications Office of the European Union*, 2021.
- [2] "Report Of The Conference Of The Parties On Its Twenty-first Session, Held In Paris From 30 November To 13 December 2015," United Nations Framework Convention on Climate Change, Paris, Tech. Rep., 2015. [Online]. Available: <https://unfccc.int/resource/docs/2015/cop21/eng/10a01.pdf>
- [3] "Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions," Brussels, Belgium, 2011. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52011DC0112>
- [4] 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.
- [5] S. Deilami and S. Muyeen, "An Insight into Practical Solutions for Electric Vehicle Charging in Smart Grid," *Energies*, vol. 13, no. 7, p. 1545, 2020.
- [6] L. Li, L. Wang, C. Sheng, W. Sun, and Y. Li, "Analysis on voltage deviation inactive distribution network and active voltage management," in *2014 China International Conference on Electricity Distribution (CICED)*, 2014: IEEE, pp. 1610-1614.
- [7] P. Pillay and M. Manyage, "Definitions of voltage unbalance," *IEEE Power Engineering Review*, vol. 21, no. 5, pp. 50-51, 2001.
- [8] O. Ipinnimo, S. Chowdhury, S. Chowdhury, and J. Mitra, "A review of voltage dip mitigation techniques with distributed generation in electricity networks," *Electric Power Systems Research*, vol. 103, pp. 28-36, 2013.
- [9] P. Kang, W. Guo, W. Huang, J. Le, and T. Mao, "Power quality problem and key improvement technology for regional power grids," *International Journal of Emerging Electric Power Systems*, vol. 21, no. 3, 2020.
- [10] C. de Mattos Affonso and M. Kezunovic, "Technical and economic impact of PV-BESS charging station on transformer life: A case study," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4683-4692, 2018.
- [11] J. P. Holguin, D. C. Rodriguez, and G. Ramos, "Reverse power flow (RPF) detection and impact on protection coordination of distribution systems," *IEEE Transactions on Industry Applications*, vol. 56, no. 3, pp. 2393-2401, 2020.
- [12] A. Alzahrani, H. Alharthi, and M. Khalid, "Minimization of Power Losses through Optimal Battery Placement in a Distributed Network with High Penetration of Photovoltaics," *Energies*, vol. 13, no. 1, p. 140, 2020.
- [13] Y. H. Bhatt, D. V. Bhatt, and V. H. Pakka, "Voltage Stability Index and Voltage Deviation Improvements using Intelligent Algorithms for Capacitor Sizing and Placement," in *2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2018: IEEE, pp. 422-427.
- [14] C. Masters, "Voltage rise: the big issue when connecting embedded generation to long 11 kV overhead lines," *Power engineering journal*, vol. 16, no. 1, pp. 5-12, 2002.
- [15] Y. Wang, K. Tan, X. Y. Peng, and P. L. So, "Coordinated control of distributed energy-storage systems for voltage regulation in distribution networks," *IEEE transactions on power delivery*, vol. 31, no. 3, pp. 1132-1141, 2015.

-
- [16] J. Modarresi, E. Gholipour, and A. Khodabakhshian, "A comprehensive review of the voltage stability indices," *Renewable and Sustainable Energy Reviews*, vol. 63, pp. 1-12, 2016.
- [17] 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. 105445, 2020.
- [18] C. Z. El-Bayeh, K. Alzaareer, B. Brahmi, and M. Zellagui, "A Novel Algorithm for Controlling Active and Reactive Power Flows of Electric Vehicles in Buildings and Its Impact on the Distribution Network," *World Electric Vehicle Journal*, vol. 11, no. 2, p. 43, 2020.
- [19] B. K. Jha, A. Singh, A. Kumar, D. K. Dheer, D. Singh, and R. K. Misra, "Day ahead scheduling of PHEVs and D-BESSs in the presence of DGs in the distribution system," *IET Electrical Systems in Transportation*, vol. 10, no. 2, pp. 170-184, 2019.
- [20] S. Habib *et al.*, "A framework for stochastic estimation of electric vehicle charging behavior for risk assessment of distribution networks," *Frontiers in Energy*, pp. 1-20, 2019.
- [21] A. Anzalchi, A. Sundararajan, A. Moghadasi, and A. Sarwat, "High-penetration grid-tied photovoltaics: Analysis of power quality and feeder voltage profile," *IEEE Industry Applications Magazine*, vol. 25, no. 5, pp. 83-94, 2019.
- [22] P. Q. A. Guide, "Voltage Disturbances," *Standard EN*, vol. 50160, 2004.
- [23] B. Standard, "Voltage characteristics of electricity supplied by public distribution networks," *BS EN*, 2007.
- [24] FNN, "Voltage Regulating Distribution Transformer (VRDT) – Use in Grid Planning and Operation ", 2016.
- [25] 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.
- [26] Y. Ma, D. Azuatalam, T. Power, A. C. Chapman, and G. Verbič, "A novel probabilistic framework to study the impact of photovoltaic-battery systems on low-voltage distribution networks," *Applied Energy*, vol. 254, p. 113669, 2019.
- [27] T. Barbosa, J. Andrade, R. Torquato, W. Freitas, and F. C. Trindade, "Use of EV hosting capacity for management of low-voltage distribution systems," *IET Generation, Transmission & Distribution*, vol. 14, no. 13, pp. 2620-2629, 2020.
- [28] G. Caneponi, F. Cazzato, S. Cochi, M. Di Clerico, M. C. Falvo, and M. Manganelli, "EV charging stations and RES-based DG: A centralized approach for smart integration in active distribution grids," *International Journal of Renewable Energy Research (IJRER)*, vol. 9, no. 2, pp. 605-612, 2019.
- [29] U. P. Müller *et al.*, "Integrated techno-economic power system planning of transmission and distribution grids," *Energies*, vol. 12, no. 11, p. 2091, 2019.
- [30] M. Thomson, "Automatic voltage control relays and embedded generation," *Power Engineering Journal*, vol. 14, no. 2, pp. 71-76, 2000.
- [31] P. Unahalekhaka and P. Sripakarach, "Reduction of Reverse Power Flow using the Appropriate Size and Installation Position of a BESS for a PV Power Plant," *IEEE Access*, 2020.
- [32] S. Deb, K. Tammi, K. Kalita, and P. Mahanta, "Impact of electric vehicle charging station load on distribution network," *Energies*, vol. 11, no. 1, p. 178, 2018.
- [33] C. H. Tie, C. K. Gan, K. A. Ibrahim, and M. Shamshiri, "Probabilistic evaluation of the impact of residential photovoltaic system on Malaysia low-voltage network using Monte

-
- Carlo approach," *Journal of Renewable and Sustainable Energy*, vol. 7, no. 6, p. 063110, 2015.
- [34] J. Shukla, B. K. Panigrahi, and P. K. Ray, "Stochastic reconfiguration of distribution system considering stability, correlated loads and renewable energy based DGs with varying penetration," *Sustainable Energy, Grids and Networks*, vol. 23, p. 100366, 2020.
- [35] H. Sadeghian and Z. Wang, "A novel impact-assessment framework for distributed PV installations in low-voltage secondary networks," *Renewable Energy*, vol. 147, pp. 2179-2194, 2020.
- [36] T. P. Abud, R. S. Maciel, and B. S. Borba, "Influence of local market economic analysis on PV generation stochastic approach in LV distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 112, pp. 178-190, 2019.
- [37] G. Tévar, A. Gómez-Expósito, A. Arcos-Vargas, and M. Rodríguez-Montañés, "Influence of rooftop PV generation on net demand, losses and network congestions: A case study," *International Journal of Electrical Power & Energy Systems*, vol. 106, pp. 68-86, 2019.
- [38] M. S. K. Reddy and K. Selvajothi, "Optimal placement of electric vehicle charging station for unbalanced radial distribution systems," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, pp. 1-15, 2020.
- [39] V. K. B. Ponnamp and K. Swarnasri, "Multi-Objective Optimal Allocation of Electric Vehicle Charging Stations in Radial Distribution System Using Teaching Learning Based Optimization," *International Journal of Renewable Energy Research (IJRER)*, vol. 10, no. 1, pp. 366-377, 2020.
- [40] H. Suyono, M. T. Rahman, H. Mokhlis, M. Othman, H. A. Illias, and H. Mohamad, "Optimal scheduling of plug-in electric vehicle charging including time-of-use tariff to minimize cost and system stress," *Energies*, vol. 12, no. 8, p. 1500, 2019.
- [41] Y. Kongjeen and K. Bhummikittipich, "Impact of plug-in electric vehicles integrated into power distribution system based on voltage-dependent power flow analysis," *Energies*, vol. 11, no. 6, p. 1571, 2018.
- [42] L. Gong, W. Cao, K. Liu, J. Zhao, and X. Li, "Spatial and temporal optimization strategy for plug-in electric vehicle charging to mitigate impacts on distribution network," *Energies*, vol. 11, no. 6, p. 1373, 2018.
- [43] A. Awasthi, K. Venkitesamy, S. Padmanaban, R. Selvamuthukumar, F. Blaabjerg, and A. K. Singh, "Optimal planning of electric vehicle charging station at the distribution system using hybrid optimization algorithm," *Energy*, vol. 133, pp. 70-78, 2017.
- [44] M. Manbachi, H. Farhangi, A. Palizban, and S. Arzanpour, "A novel Volt-VAR Optimization engine for smart distribution networks utilizing Vehicle to Grid dispatch," *International Journal of Electrical Power & Energy Systems*, vol. 74, pp. 238-251, 2016.
- [45] Z.-j. Pan and Y. Zhang, "A novel centralized charging station planning strategy considering urban power network structure strength," *Electric Power Systems Research*, vol. 136, pp. 100-109, 2016.
- [46] N. Leemput, F. Geth, J. Van Roy, J. Büscher, and J. Driesen, "Reactive power support in residential LV distribution grids through electric vehicle charging," *Sustainable Energy, Grids and Networks*, vol. 3, pp. 24-35, 2015.
- [47] E. Veldman and R. A. Verzijlbergh, "Distribution grid impacts of smart electric vehicle charging from different perspectives," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 333-342, 2014.

-
- [48] J. Hernández, F. Ruiz-Rodriguez, and F. Jurado, "Modelling and assessment of the combined technical impact of electric vehicles and photovoltaic generation in radial distribution systems," *Energy*, vol. 141, pp. 316-332, 2017.
- [49] M. ElNozahy, T. K. Abdel-Galil, and M. Salama, "Probabilistic ESS sizing and scheduling for improved integration of PHEVs and PV systems in residential distribution systems," *Electric Power Systems Research*, vol. 125, pp. 55-66, 2015.
- [50] P. M. de Quevedo, G. Munoz-Delgado, and J. Contreras, "Impact of electric vehicles on the expansion planning of distribution systems considering renewable energy, storage, and charging stations," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 794-804, 2017.
- [51] D. Shaw-Williams, C. Susilawati, G. Walker, and J. Varendorff, "Valuing the impact of residential photovoltaics and batteries on network electricity losses: An Australian case study," *Utilities Policy*, vol. 60, p. 100955, 2019.
- [52] A. Cortés, J. Mazón, and J. Merino, "Strategy of management of storage systems integrated with photovoltaic systems for mitigating the impact on LV distribution network," *International Journal of Electrical Power & Energy Systems*, vol. 103, pp. 470-482, 2018.
- [53] M. A. Tolba, H. Rezk, V. Tulsy, A. A. Z. Diab, A. Y. Abdelaziz, and A. Vanin, "Impact of optimum allocation of renewable distributed generations on distribution networks based on different optimization algorithms," *Energies*, vol. 11, no. 1, p. 245, 2018.
- [54] S. K. Injeti and V. K. Thunuguntla, "Optimal integration of DGs into radial distribution network in the presence of plug-in electric vehicles to minimize daily active power losses and to improve the voltage profile of the system using bio-inspired optimization algorithms," *Protection and Control of Modern Power Systems*, vol. 5, no. 1, pp. 1-15, 2020.
- [55] Z. Guo, Z. Zhou, and Y. Zhou, "Impacts of Integrating Topology Reconfiguration and Vehicle-to-Grid Technologies on Distribution System Operation," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 1023-1032, 2019.
- [56] J. Salehi and A. Abdolahi, "Optimal scheduling of active distribution networks with penetration of PHEV considering congestion and air pollution using DR program," *Sustainable Cities and Society*, vol. 51, p. 101709, 2019.
- [57] G. Battapothula, C. Yammani, and S. Maheswarapu, "Multi-objective simultaneous optimal planning of electrical vehicle fast charging stations and DGs in distribution system," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 4, pp. 923-934, 2019.
- [58] M. H. Moradi, M. Abedini, S. R. Tousi, and S. M. Hosseini, "Optimal siting and sizing of renewable energy sources and charging stations simultaneously based on Differential Evolution algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 73, pp. 1015-1024, 2015.
- [59] B. Liu, X. Huang, J. Li, X. Qian, and J. Cheng, "Multi-objective planning of distribution network containing distributed generation and electric vehicle charging stations," *Power system technology*, vol. 39, no. 2, pp. 450-456, 2015.
- [60] S. Freitas, T. Santos, and M. C. Brito, "Impact of large scale PV deployment in the sizing of urban distribution transformers," *Renewable Energy*, vol. 119, pp. 767-776, 2018.
- [61] J. D. Watson, N. R. Watson, D. Santos-Martin, A. R. Wood, S. Lemon, and A. J. Miller, "Impact of solar photovoltaics on the low-voltage distribution network in New Zealand," *IET Generation, Transmission & Distribution*, vol. 10, no. 1, pp. 1-9, 2016.

-
- [62] M. Hasheminamin, V. G. Agelidis, V. Salehi, R. Teodorescu, and B. Hredzak, "Index-based assessment of voltage rise and reverse power flow phenomena in a distribution feeder under high PV penetration," *IEEE Journal of Photovoltaics*, vol. 5, no. 4, pp. 1158-1168, 2015.
- [63] F. Cazzato, M. Di Clerico, M. C. Falvo, S. Ferrero, and M. Vivian, "New Dispatching Paradigm in Power Systems Including EV Charging Stations and Dispersed Generation: A Real Test Case," *Energies*, vol. 13, no. 4, p. 944, 2020.
- [64] 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.
- [65] T. Aziz and N. Ketjoy, "Enhancing PV penetration in LV networks using reactive power control and on load tap changer with existing transformers," *IEEE Access*, vol. 6, pp. 2683-2691, 2017.
- [66] S. Johansson, J. Persson, S. Lazarou, and A. Theocharis, "Investigation of the Impact of Large-Scale Integration of Electric Vehicles for a Swedish Distribution Network," *Energies*, vol. 12, no. 24, p. 4717, 2019.
- [67] H. Wang, W. Zhou, K. Qian, and S. Meng, "Modelling of ampacity and temperature of MV cables in presence of harmonic currents due to EVs charging in electrical distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 112, pp. 127-136, 2019.
- [68] K. Kasturi, C. K. Nayak, and M. R. Nayak, "Electric vehicles management enabling G2V and V2G in smart distribution system for maximizing profits using MOMVO," *International Transactions on Electrical Energy Systems*, vol. 29, no. 6, p. e12013, 2019.
- [69] M. A. Awadallah, B. N. Singh, and B. Venkatesh, "Impact of EV charger load on distribution network capacity: A case study in Toronto," *Canadian Journal of Electrical and Computer Engineering*, vol. 39, no. 4, pp. 268-273, 2016.
- [70] J. De Hoog, T. Alpcan, M. Brazil, D. A. Thomas, and I. Mareels, "Optimal charging of electric vehicles taking distribution network constraints into account," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 365-375, 2014.
- [71] Y. Mu, J. Wu, N. Jenkins, H. Jia, and C. Wang, "A spatial-temporal model for grid impact analysis of plug-in electric vehicles," *Applied Energy*, vol. 114, pp. 456-465, 2014.
- [72] J. Zhao, Z. Xu, J. Wang, C. Wang, and J. Li, "Robust distributed generation investment accommodating electric vehicle charging in a distribution network," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4654-4666, 2018.
- [73] A. Chidurala, T. K. Saha, and N. Mithulananthan, "Harmonic impact of high penetration photovoltaic system on unbalanced distribution networks—learning from an urban photovoltaic network," *IET Renewable Power Generation*, vol. 10, no. 4, pp. 485-494, 2016.
- [74] Y. Zou, J. Zhao, X. Gao, Y. Chen, and A. Tohidi, "Experimental results of electric vehicles effects on low voltage grids," *Journal of Cleaner Production*, vol. 255, p. 120270, 2020.
- [75] R. Kumar and D. Saxena, "Impact of Plug-In Electric Vehicles on Faulted Distribution System," *Arabian Journal for Science and Engineering*, vol. 45, no. 3, pp. 1599-1614, 2020.
- [76] R. Godina, E. M. Rodrigues, N. Paterakis, O. Erdinc, and J. P. Catalao, "Innovative impact assessment of electric vehicles charging loads on distribution transformers using real data," *Energy Conversion and Management*, vol. 120, pp. 206-216, 2016.

-
- [77] M. Neaimeh *et al.*, "A probabilistic approach to combining smart meter and electric vehicle charging data to investigate distribution network impacts," *Applied Energy*, vol. 157, pp. 688-698, 2015.
- [78] M. K. Gray and W. G. Morsi, "Power quality assessment in distribution systems embedded with plug-in hybrid and battery electric vehicles," *IEEE Transactions on Power Systems*, vol. 30, no. 2, pp. 663-671, 2014.
- [79] L. Hua, J. Wang, and C. Zhou, "Adaptive electric vehicle charging coordination on distribution network," *IEEE Transactions on Smart Grid*, vol. 5, no. 6, pp. 2666-2675, 2014.
- [80] S. M. Arif, T. T. Lie, B. C. Seet, S. M. Ahsan, and H. A. Khan, "Plug-In Electric Bus Depot Charging with PV and ESS and Their Impact on LV Feeder," *Energies*, vol. 13, no. 9, p. 2139, 2020.
- [81] A. Marszal-Pomianowska, J. Widén, J. Le Dréau, P. Heiselberg, B. Bak-Jensen, and I. D. de Cerio Mendaza, "Operation of power distribution networks with new and flexible loads: A case of existing residential low voltage network," *Energy*, p. 117715, 2020.
- [82] A. Navarro-Espinosa and P. Mancarella, "Probabilistic modeling and assessment of the impact of electric heat pumps on low voltage distribution networks," *Applied Energy*, vol. 127, pp. 249-266, 2014.
- [83] T. E. Castelo de Oliveira, M. Bollen, P. F. Ribeiro, P. de Carvalho, A. C. Zambroni, and B. D. Bonatto, "The concept of dynamic hosting capacity for distributed energy resources: Analytics and practical considerations," *Energies*, vol. 12, no. 13, p. 2576, 2019.
- [84] L. González, E. Siavichay, and J. Espinoza, "Impact of EV fast charging stations on the power distribution network of a Latin American intermediate city," *Renewable and Sustainable Energy Reviews*, vol. 107, pp. 309-318, 2019.
- [85] P. A. Aghdam and H. Khoshkhou, "Voltage stability assessment algorithm to predict power system loadability margin," *IET Generation, Transmission & Distribution*, vol. 14, no. 10, pp. 1816-1828, 2020.
- [86] M. Ghaffarianfar and A. Hajizadeh, "Voltage stability of low-voltage distribution grid with high penetration of photovoltaic power units," *Energies*, vol. 11, no. 8, p. 1960, 2018.
- [87] M. M. Aly, M. Abdel-Akher, Z. Ziadi, and T. Senjyu, "Assessment of reactive power contribution of photovoltaic energy systems on voltage profile and stability of distribution systems," *International Journal of Electrical Power & Energy Systems*, vol. 61, pp. 665-672, 2014.
- [88] J. Su, T. Lie, and R. Zamora, "Modelling of large-scale electric vehicles charging demand: A New Zealand case study," *Electric Power Systems Research*, vol. 167, pp. 171-182, 2019.
- [89] L. Luo *et al.*, "Optimal planning of electric vehicle charging stations comprising multi-types of charging facilities," *Applied energy*, vol. 226, pp. 1087-1099, 2018.
- [90] S. Shafiee, M. Fotuhi-Firuzabad, and M. Rastegar, "Investigating the impacts of plug-in hybrid electric vehicles on power distribution systems," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1351-1360, 2013.
- [91] P. Arboleya, A. Koirala, L. Suárez, B. Mohamed, and C. González-Morán, "Impact Evaluation of the New Self-Consumption Spanish Scenario on the Low-Voltage Terminal Distribution Network," *IEEE Transactions on Industry Applications*, vol. 55, no. 6, pp. 7230-7239, 2019.
- [92] F. Lamberti, V. Calderaro, V. Galdi, and G. Graditi, "Massive data analysis to assess PV/ESS integration in residential unbalanced LV networks to support voltage profiles," *Electric Power Systems Research*, vol. 143, pp. 206-214, 2017.

-
- [93] O. Babacan, W. Torre, and J. Kleissl, "Siting and sizing of distributed energy storage to mitigate voltage impact by solar PV in distribution systems," *Solar Energy*, vol. 146, pp. 199-208, 2017.
- [94] H. Farzin, M. Moeini-Aghtaie, and M. Fotuhi-Firuzabad, "Reliability studies of distribution systems integrated with electric vehicles under battery-exchange mode," *IEEE Transactions on Power Delivery*, vol. 31, no. 6, pp. 2473-2482, 2015.
- [95] M. Islam, M. Nadarajah, and M. J. Hossain, "Short-term voltage stability enhancement in residential grid with high penetration of rooftop PV units," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 4, pp. 2211-2222, 2018.
- [96] X. Ran and S. Miao, "Three-phase probabilistic load flow for power system with correlated wind, photovoltaic and load," *IET Generation, Transmission & Distribution*, vol. 10, no. 12, pp. 3093-3101, 2016.
- [97] E. Hooshmand and A. Rabiee, "Energy management in distribution systems, considering the impact of reconfiguration, RESSs, ESSs and DR: A trade-off between cost and reliability," *Renewable energy*, vol. 139, pp. 346-358, 2019.
- [98] M. Wajahat, H. A. Khalid, G. M. Bhutto, and C. Leth Bak, "A comparative study into enhancing the PV penetration limit of a LV CIGRE residential network with distributed grid-tied single-phase PV systems," *Energies*, vol. 12, no. 15, p. 2964, 2019.
- [99] J. Yaghoobi, M. Islam, and N. Mithulananthan, "Analytical approach to assess the loadability of unbalanced distribution grid with rooftop PV units," *Applied energy*, vol. 211, pp. 358-367, 2018.
- [100] K. Knezović and M. Marinelli, "Phase-wise enhanced voltage support from electric vehicles in a Danish low-voltage distribution grid," *Electric Power Systems Research*, vol. 140, pp. 274-283, 2016.
- [101] F. Shahnian, A. Ghosh, G. Ledwich, and F. Zare, "Predicting voltage unbalance impacts of plug-in electric vehicles penetration in residential low-voltage distribution networks," *Electric Power Components and Systems*, vol. 41, no. 16, pp. 1594-1616, 2013.
- [102] Q. Deng *et al.*, "Evaluation of Accommodation Capability for Electric Vehicles of a Distribution System Considering Coordinated Charging Strategies," *Energies*, vol. 12, no. 16, p. 3056, 2019.
- [103] M. M. Aly, E. Abdelkarim, and M. Abdel-Akher, "Mitigation of photovoltaic power generation fluctuations using plug-in hybrid electric vehicles storage batteries," *International Transactions on Electrical Energy Systems*, vol. 25, no. 12, pp. 3720-3737, 2015.
- [104] J. Y. Yong, V. K. Ramachandaramurthy, K. M. Tan, and N. Mithulananthan, "Bi-directional electric vehicle fast charging station with novel reactive power compensation for voltage regulation," *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 300-310, 2015.
- [105] R.-C. Leou, C.-L. Su, and C.-N. Lu, "Stochastic analyses of electric vehicle charging impacts on distribution network," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1055-1063, 2013.
- [106] F. J. Ruiz-Rodriguez, J. C. Hernández, and F. Jurado, "Voltage behaviour in radial distribution systems under the uncertainties of photovoltaic systems and electric vehicle charging loads," *International Transactions on Electrical Energy Systems*, vol. 28, no. 2, p. e2490, 2018.
- [107] X. Liu, "Research on Flexibility Evaluation Method of Distribution System Based on Renewable Energy and Electric Vehicles," *IEEE Access*, vol. 8, pp. 109249-109265, 2020.

-
- [108] F. J. Ruiz-Rodríguez, J. C. Hernández, and F. Jurado, "Probabilistic load-flow analysis of biomass-fuelled gas engines with electrical vehicles in distribution systems," *Energies*, vol. 10, no. 10, p. 1536, 2017.
- [109] C. Protopapadaki and D. Saelens, "Heat pump and PV impact on residential low-voltage distribution grids as a function of building and district properties," *Applied Energy*, vol. 192, pp. 268-281, 2017.
- [110] 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.
- [111] L. Wang, M. Yuan, F. Zhang, X. Wang, L. Dai, and F. Zhao, "Risk assessment of distribution networks integrating large-scale distributed photovoltaics," *IEEE Access*, vol. 7, pp. 59653-59664, 2019.
- [112] A. Argüello, J. D. Lara, J. D. Rojas, and G. Valverde, "Impact of rooftop PV integration in distribution systems considering socioeconomic factors," *IEEE Systems Journal*, vol. 12, no. 4, pp. 3531-3542, 2017.
- [113] A. F. Crossland, D. Jones, N. S. Wade, and S. L. Walker, "Comparison of the location and rating of energy storage for renewables integration in residential low voltage networks with overvoltage constraints," *Energies*, vol. 11, no. 8, p. 2041, 2018.
- [114] C. Li, V. R. Disfani, Z. K. Pecenak, S. Mohajeryami, and J. Kleissl, "Optimal OLTC voltage control scheme to enable high solar penetrations," *Electric Power Systems Research*, vol. 160, pp. 318-326, 2018.
- [115] K. Duwadi, A. Ingalalli, and T. M. Hansen, "Monte Carlo Analysis of High Penetration Residential Solar Voltage Impacts using High Performance Computing," in *2019 IEEE International Conference on Electro Information Technology (EIT)*, 2019: IEEE, pp. 1-6.
- [116] M. Li *et al.*, "GIS-based probabilistic modeling of BEV charging load for Australia," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3525-3534, 2018.
- [117] A. Ahmadian, M. Sedghi, B. Mohammadi-ivatloo, A. Elkamel, M. A. Golkar, and M. Fowler, "Cost-benefit analysis of V2G implementation in distribution networks considering PEVs battery degradation," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 961-970, 2017.
- [118] H. Tayarani, H. Jahangir, R. Nadafianshahamabadi, M. Aliakbar Golkar, A. Ahmadian, and A. Elkamel, "Optimal charging of plug-in electric vehicle: Considering travel behavior uncertainties and battery degradation," *Applied Sciences*, vol. 9, no. 16, p. 3420, 2019.
- [119] M. B. Ndawula, S. Z. Djokic, and I. Hernando-Gil, "Reliability Enhancement in Power Networks under Uncertainty from Distributed Energy Resources," *Energies*, vol. 12, no. 3, p. 531, 2019.
- [120] H. R. Galiveeti, A. K. Goswami, and N. B. D. Choudhury, "Impact of plug-in electric vehicles and distributed generation on reliability of distribution systems," *Engineering science and technology, an international journal*, vol. 21, no. 1, pp. 50-59, 2018.
- [121] X. Fu, H. Sun, Q. Guo, Z. Pan, X. Zhang, and S. Zeng, "Probabilistic power flow analysis considering the dependence between power and heat," *Applied energy*, vol. 191, pp. 582-592, 2017.
- [122] K. Qian, C. Zhou, and Y. Yuan, "Impacts of high penetration level of fully electric vehicles charging loads on the thermal ageing of power transformers," *International Journal of Electrical Power & Energy Systems*, vol. 65, pp. 102-112, 2015.

-
- [123] S. Wang, L. Yu, L. Wu, Y. Dong, and H. Wang, "An improved differential evolution algorithm for optimal location of battery swapping stations considering multi-type electric vehicle scale evolution," *IEEE Access*, vol. 7, pp. 73020-73035, 2019.
- [124] R. Mehta, D. Srinivasan, A. M. Khambadkone, J. Yang, and A. Trivedi, "Smart charging strategies for optimal integration of plug-in electric vehicles within existing distribution system infrastructure," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 299-312, 2016.
- [125] T. Yunusov, D. Frame, W. Holderbaum, and B. Potter, "The impact of location and type on the performance of low-voltage network connected battery energy storage systems," *Applied Energy*, vol. 165, pp. 202-213, 2016.
- [126] Y. Zheng, Z. Y. Dong, Y. Xu, K. Meng, J. H. Zhao, and J. Qiu, "Electric vehicle battery charging/swap stations in distribution systems: comparison study and optimal planning," *IEEE transactions on Power Systems*, vol. 29, no. 1, pp. 221-229, 2013.
- [127] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [128] H.-K. Ringkjøb, P. M. Haugan, and I. M. Solbrekke, "A review of modelling tools for energy and electricity systems with large shares of variable renewables," *Renewable and Sustainable Energy Reviews*, vol. 96, pp. 440-459, 2018.
- [129] EATON. "CYME Power Engineering Software." <http://www.cyme.com/> (accessed 10 March, 2020).
- [130] NEPLAN. "Target Grid Planning." <https://www.neplan.ch/description/target-grid-planning/> (accessed 15 April, 2020).
- [131] S. Kadam, B. Bletterie, and W. Gawlik, "A Large Scale Grid Data Analysis Platform for DSOs," *Energies*, vol. 10, no. 8, p. 1099, 2017.
- [132] L. Thurner *et al.*, "pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6510-6521, 2018.
- [133] T. Brown, J. Hörsch, and D. Schlachtberger, "PyPSA: Python for power system analysis," *arXiv preprint arXiv:1707.09913*, 2017.
- [134] DlgSILENT. "DlG SILENT PowerFactory." <http://www.digsilent.de/index.php/products-powerfactory.html> (accessed 25 March, 2020).
- [135] S. G. S. (SGS). "Enhanced Hosting Capacity Analysis." http://mnsolarpathways.org/wp-content/uploads/2018/10/mn-solar-pathways_pv-hosting-capacity-report.pdf (accessed 25 March, 2020).
- [136] "Distributed Energy Resources Customer Adoption Model (DER-CAM)." <https://building-microgrid.lbl.gov/projects/der-cam> (accessed 25 March, 2020).
- [137] "GridLAB-D Simulation Software." <http://www.gridlabd.org/> (accessed 15 April, 2020).
- [138] "Power system simulation Power system Analysis Hypersim - Opal-RT." <http://www.opal-rt.com/systems-hypersim/> (accessed 15 April, 2020).
- [139] "IPSA 2 - IPSA POWER." http://www.ipsa-power.com/?Page_id=766 (accessed 15 April, 2020).
- [140] R. D. Zimmerman, C. E. Murillo-Sánchez, and D. Gan, "Matpower," *PSEERC.[Online]. Software Available at: http://www.pserc.cornell.edu/matpower*, 1997.
- [141] "EPRI | Smart Grid Resource Center: Simulation Tool—OpenDSS." <http://smartgrid.epri.com/SimulationTool.aspx> (accessed 25 April, 2020).

-
- [142] M. Pöchacker and W. Elmenreich, "Model implementation for the extendable open source power system simulator RAPSIm," in *2015 12th International Workshop on Intelligent Solutions in Embedded Systems (WISES)*, 2015: IEEE, pp. 103-108.
- [143] "Regional Energy Deployment System (ReEDS) Model - NREL." <http://www.nrel.gov/analysis/reeds/> (accessed 25 April, 2020).
- [144] "Simpow | Solvina:." <http://www.solvina.se/simpow> (accessed 25 April, 2020).
- [145] C. Syranidou, J. Linssen, D. Stolten, and M. Robinus, "Integration of large-scale variable renewable energy sources into the future European power system: On the curtailment challenge," *Energies*, vol. 13, no. 20, p. 5490, 2020.
- [146] S. Küfeoğlu and M. G. Pollitt, "The impact of PVs and EVs on domestic electricity network charges: A case study from Great Britain," *Energy policy*, vol. 127, pp. 412-424, 2019.
- [147] K. R. Reddy and S. Meikandasivam, "Smart Distribution Network with Integration of Stochastic Renewable Energy Sources and Plug-in Electric Vehicles: Challenges and Issues," *Journal of Green Engineering*, vol. 8, no. 4, pp. 431-474, 2018.
- [148] J. Hu, Z. Li, J. Zhu, and J. M. Guerrero, "Voltage stabilization: A critical step toward high photovoltaic penetration," *IEEE Industrial Electronics Magazine*, vol. 13, no. 2, pp. 17-30, 2019.
- [149] H. Kheradmand-Khanekehdani and M. Gitizadeh, "Well-being analysis of distribution network in the presence of electric vehicles," *Energy*, vol. 155, pp. 610-619, 2018.
- [150] V. Aravinthan and W. Jewell, "Controlled electric vehicle charging for mitigating impacts on distribution assets," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 999-1009, 2015.
- [151] G. Kerber, "Aufnahmefähigkeit von Niederspannungsverteilnetzen für die Einspeisung aus Photovoltaikkleinanlagen," Technische Universität München, 2011.
- [152] F. E. Postigo Marcos *et al.*, "A review of power distribution test feeders in the United States and the need for synthetic representative networks," *Energies*, vol. 10, no. 11, p. 1896, 2017.
- [153] S. Meinecke, L. Thurner, and M. Braun, "Review of Steady-State Electric Power Distribution System Datasets," *Energies*, vol. 13, no. 18, p. 4826, 2020.
- [154] S. Meinecke *et al.*, "Simbench—a benchmark dataset of electric power systems to compare innovative solutions based on power flow analysis," *Energies*, vol. 13, no. 12, p. 3290, 2020.
- [155] J. Dickert, M. Domagk, and P. Schegner, "Benchmark low voltage distribution networks based on cluster analysis of actual grid properties," in *2013 IEEE Grenoble Conference*, 2013: IEEE, pp. 1-6.
- [156] M. Kiaee, A. Cruden, and D. Infield, "Demand side management using alkaline electrolyzers within the UKGDS simulation network," 2011.
- [157] S. Civanlar, J. Grainger, H. Yin, and S. Lee, "Distribution feeder reconfiguration for loss reduction," *IEEE Transactions on Power Delivery*, vol. 3, no. 3, pp. 1217-1223, 1988.
- [158] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Power Engineering Review*, vol. 9, no. 4, pp. 101-102, 1989.
- [159] C.-T. Su, C.-F. Chang, and J.-P. Chiou, "Distribution network reconfiguration for loss reduction by ant colony search algorithm," *Electric power systems research*, vol. 75, no. 2-3, pp. 190-199, 2005.

-
- [160] S. Barsali, *Benchmark systems for network integration of renewable and distributed energy resources*. 2014.
- [161] G. A. Pagani and M. Aiello, "A complex network approach for identifying vulnerabilities of the medium and low voltage grid," *International Journal of Critical Infrastructures* 7, vol. 11, no. 1, pp. 36-61, 2015.
- [162] P. Erdos and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci.*, vol. 5, no. 1, pp. 17-60, 1960.
- [163] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *nature*, vol. 393, no. 6684, pp. 440-442, 1998.
- [164] R. E. González, "A complex-network approach to support transmission expansion planning," Universidad Pontificia Comillas, 2019.
- [165] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509-512, 1999.
- [166] G. A. Pagani and M. Aiello, "The power grid as a complex network: a survey," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 11, pp. 2688-2700, 2013.
- [167] Z. Wang, A. Scaglione, and R. J. Thomas, "Generating statistically correct random topologies for testing smart grid communication and control networks," *IEEE transactions on Smart Grid*, vol. 1, no. 1, pp. 28-39, 2010.
- [168] Z. Wang, R. J. Thomas, and A. Scaglione, "Generating random topology power grids," in *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*, 2008: IEEE, pp. 183-183.
- [169] Z. Wang and R. J. Thomas, "On bus type assignments in random topology power grid models," in *2015 48th Hawaii International Conference on System Sciences*, 2015: IEEE, pp. 2671-2679.
- [170] J. Hu, L. Sankar, and D. J. Mir, "Cluster-and-Connect: An algorithmic approach to generating synthetic electric power network graphs," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2015: IEEE, pp. 223-230.
- [171] A. B. Birchfield, K. M. Gegner, T. Xu, K. S. Shetye, and T. J. Overbye, "Statistical considerations in the creation of realistic synthetic power grids for geomagnetic disturbance studies," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1502-1510, 2016.
- [172] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Transactions on power systems*, vol. 32, no. 4, pp. 3258-3265, 2016.
- [173] K. M. Gegner, A. B. Birchfield, T. Xu, K. S. Shetye, and T. J. Overbye, "A methodology for the creation of geographically realistic synthetic power flow models," in *2016 IEEE Power and Energy Conference at Illinois (PECI)*, 2016: IEEE, pp. 1-6.
- [174] P. Schultz, J. Heitzig, and J. Kurths, "A random growth model for power grids and other spatially embedded infrastructure networks," *The European Physical Journal Special Topics*, vol. 223, no. 12, pp. 2593-2610, 2014.
- [175] R. Espejo, S. Lumbreras, and A. Ramos, "A complex-network approach to the generation of synthetic power transmission networks," *IEEE Systems Journal*, vol. 13, no. 3, pp. 3050-3058, 2018.
- [176] S. Soltan, A. Loh, and G. Zussman, "A learning-based method for generating synthetic power grids," *IEEE Systems Journal*, vol. 13, no. 1, pp. 625-634, 2018.
- [177] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *IEEE Pervasive computing*, vol. 7, no. 4, pp. 12-18, 2008.

-
- [178] W. Medjroubi, C. Matke, and D. Kleinhans, "SciGRID—An Open Source Reference Model for the European Transmission Network," 2015.
- [179] M. Scharf and A. Nebel, "osmTGmod: Open Source German Transmission Grid Model Based on OpenStreetMap v0. 1.3," 2017.
- [180] B. Wiegmanns, "Improving the topology of an electric network model based on open data," Faculty of Science and Engineering, 2016.
- [181] J. Rivera, P. Nasirifard, J. Leimhofer, and H.-A. Jacobsen, "Automatic generation of real power transmission grid models from crowdsourced data," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5436-5448, 2018.
- [182] W. Heitkoetter, W. Medjroubi, T. Vogt, and C. Agert, "Comparison of Open Source Power Grid Models—Combining a Mathematical, Visual and Electrical Analysis in an Open Source Tool," *Energies*, vol. 12, no. 24, p. 4728, 2019.
- [183] U. P. Mueller *et al.*, "The eGo grid model: An open source approach towards a model of German high and extra-high voltage power grids," in *Journal of Physics: Conference Series*, 2018, vol. 977, no. 1: IOP Publishing, p. 012003.
- [184] C. M. Domingo, T. G. San Roman, A. Sánchez-Miralles, J. P. P. Gonzalez, and A. C. Martinez, "A reference network model for large-scale distribution planning with automatic street map generation," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 190-197, 2010.
- [185] M. Grzanic, M. G. Flammini, and G. Pretico, "Distribution network model platform: A first case study," *Energies*, vol. 12, no. 21, p. 4079, 2019.
- [186] G. Pisano *et al.*, "Synthetic models of distribution networks based on open data and georeferenced information," *Energies*, vol. 12, no. 23, p. 4500, 2019.
- [187] C. Mateo *et al.*, "Building large-scale us synthetic electric distribution system models," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5301-5313, 2020.
- [188] J. Kays, A. Seack, T. Smirek, F. Westkamp, and C. Rehtanz, "The generation of distribution grid models on the basis of public available data," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2346-2353, 2016.
- [189] J. Amme, G. Pleßmann, J. Bühler, L. Hülk, E. Kötter, and P. Schwaegerl, "The eGo grid model: An open-source and open-data based synthetic medium-voltage grid model for distribution power supply systems," in *Journal of Physics: Conference Series*, 2018, vol. 977, no. 1: IOP Publishing, p. 012007.
- [190] A. Navarro and H. Rudnick, "Large-scale distribution planning—Part I: Simultaneous network and transformer optimization," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 744-751, 2009.
- [191] A. Navarro and H. Rudnick, "Large-scale distribution planning—Part II: Macro-optimization with Voronoi's diagram and tabu search," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 752-758, 2009.
- [192] U. Brandes, "A faster algorithm for betweenness centrality," *Journal of mathematical sociology*, vol. 25, no. 2, pp. 163-177, 2001.
- [193] J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, and J. Kertesz, "Generalizations of the clustering coefficient to weighted complex networks," *Physical Review E*, vol. 75, no. 2, p. 027105, 2007.
- [194] M. E. Newman, "Mixing patterns in networks," *Physical review E*, vol. 67, no. 2, p. 026126, 2003.
- [195] J. G. Foster, D. V. Foster, P. Grassberger, and M. Paczuski, "Edge direction and the structure of networks," *Proceedings of the National Academy of Sciences*, vol. 107, no. 24, pp. 10815-10820, 2010.

-
- [196] O. Contributors, "Planet Dump Retrieved from <https://planet.osm.org>," URL <https://www.openstreetmap.org>, 2017.
- [197] Geofabrik. "OpenStreetMap data download." <https://download.geofabrik.de/europe/germany.html> (accessed 20 January, 2019).
- [198] OSMOSIS. "OSMOSIS—A command line Java application for processing OSM data." <http://wiki.openstreetmap.org/wiki/Osmosis> (accessed 10 January, 2019).
- [199] osm2pgsql. "Osm2pgsql—An OSM data importer for Postgis databases." <https://osm2pgsql.org/> (accessed 10 January, 2019).
- [200] "Open Power System Data." <https://open-power-system-data.org> (accessed 10 January, 2019).
- [201] B. Verteilernetzstudie, "Moderne Netze für Deutschland (Verteilernetzstudie)," *Abschlussbericht der E-Bridge, IAEW und OFFIS im Auftrag des BMWi*, 2014.
- [202] B. Bayer, P. Matschoss, H. Thomas, and A. Marian, "The German experience with integrating photovoltaic systems into the low-voltage grids," *Renewable energy*, vol. 119, pp. 129-141, 2018.
- [203] S. Steiniger, T. Lange, D. Burghardt, and R. Weibel, "An approach for the classification of urban building structures based on discriminant analysis techniques," *Transactions in GIS*, vol. 12, no. 1, pp. 31-59, 2008.
- [204] R. Hecht, G. Meinel, and M. Buchroithner, "Automatic identification of building types based on topographic databases—a comparison of different data sources," *International Journal of Cartography*, vol. 1, no. 1, pp. 18-31, 2015.
- [205] A. Henn, C. Römer, G. Gröger, and L. Plümer, "Automatic classification of building types in 3D city models," *Geoinformatica*, vol. 16, no. 2, pp. 281-306, 2012.
- [206] M. Wurm, A. Schmitt, and H. Taubenböck, "Building types' classification using shape-based features and linear discriminant functions," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 1901-1912, 2015.
- [207] Z. Lu, J. Im, J. Rhee, and M. Hodgson, "Building type classification using spatial and landscape attributes derived from LiDAR remote sensing data," *Landscape and Urban Planning*, vol. 130, pp. 134-148, 2014.
- [208] Y. Zheng and Q. Weng, "Model-driven reconstruction of 3-D buildings using LiDAR data," *IEEE geoscience and remote sensing letters*, vol. 12, no. 7, pp. 1541-1545, 2015.
- [209] Y. Forget, C. Linard, and M. Gilbert, "Supervised classification of built-up areas in sub-Saharan African cities using Landsat imagery and OpenStreetMap," *Remote Sensing*, vol. 10, no. 7, p. 1145, 2018.
- [210] H. Fan, A. Zipf, and Q. Fu, "Estimation of building types on OpenStreetMap based on urban morphology analysis," in *Connecting a Digital Europe Through Location and Place*: Springer, 2014, pp. 19-35.
- [211] H. Bast, S. Storandt, and S. Weidner, "Fine-grained population estimation," in *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2015, pp. 1-10.
- [212] "Corine land cover." <https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=download> (accessed 31 August, 2020).
- [213] "Building height 2012." <https://land.copernicus.eu/local/urban-atlas/building-height-2012> (accessed 31 August, 2020).

-
- [214] "Zensus 2011." <https://www.zensus2011.de/DE/Home/Aktuelles/DemografischeGrunddaten.html?nn=3065474> (accessed 8 August, 2020).
- [215] "Category:buildings and structures in germany by type." https://en.wikipedia.org/wiki/Category:Buildings_and_structures_in_Germany_by_type (accessed 10 September, 2020).
- [216] U. Hwang, D. Jung, and S. Yoon, "Hexagan: Generative adversarial nets for real world classification," in *International Conference on Machine Learning*, 2019: PMLR, pp. 2921-2930.
- [217] M. Smieja, Ł. Struski, J. Tabor, B. Zieliński, and P. Spurek, "Processing of missing data by neural networks," *arXiv preprint arXiv:1805.07405*, 2018.
- [218] P. Royston and I. R. White, "Multiple imputation by chained equations (MICE): implementation in Stata," *J Stat Softw*, vol. 45, no. 4, pp. 1-20, 2011.
- [219] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321-357, 2002.
- [220] "Database of the Federal Statistical Office of Germany." <https://www-genesis.destatis.de/genesis/online> (accessed 5 November, 2020).
- [221] L. Kotzur, D. Stolten, and H.-J. Wagner, *Future grid load of the residential building sector* (no. RWTH-2018-231872). Lehrstuhl für Brennstoffzellen (FZ Jülich), 2019.
- [222] J. F. Linßen, *Modellierung zeitlich aufgelöster Ladeenergienachfragen von batterieelektrischen Fahrzeugen und deren Abbildung in einem Energiesystemmodell*. Forschungszentrum Jülich GmbH, Zentralbibliothek, 2019.
- [223] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, "pvlib python: A python package for modeling solar energy systems," *Journal of Open Source Software*, vol. 3, no. 29, p. 884, 2018.
- [224] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern recognition*, vol. 36, no. 2, pp. 451-461, 2003.
- [225] F. Nielsen, "Hierarchical clustering," in *Introduction to HPC with MPI for Data Science*: Springer, 2016, pp. 195-211.
- [226] K. Khan, S. U. Rehman, K. Aziz, S. Fong, and S. Sarasvady, "DBSCAN: Past, present and future," in *The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)*, 2014: IEEE, pp. 232-238.
- [227] X. Jin and J. Han, "K-Means Clustering," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb Eds. Boston, MA: Springer US, 2010, pp. 563-564.
- [228] M. Rocklin, "Dask: Parallel computation with blocked algorithms and task scheduling," in *Proceedings of the 14th python in science conference*, 2015, vol. 130: Citeseer, p. 136.
- [229] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research*, vol. 12, pp. 2825-2830, 2011.
- [230] J.-C. Chen, "Dijkstra's shortest path algorithm," *Journal of formalized mathematics*, vol. 15, no. 9, pp. 237-247, 2003.
- [231] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," Stanford, 2006.
- [232] D. Sarajlić and C. Rehtanz, "Low voltage benchmark distribution network models based on publicly available data," in *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, 2019: IEEE, pp. 1-5.

-
- [233] M. Robinius *et al.*, *Wege für die Energiewende: kosteneffiziente und klimagerechte Transformationsstrategien für das deutsche Energiesystem bis zum Jahr 2050*. Forschungszentrum Jülich GmbH Zentralbibliothek, Verlag, 2020.

Band / Volume 634

Investigation and implementation of improved and degradation-tolerant fuel electrodes for solid oxide cells

A. Schwiers (2024), VI, 163, XIII pp

ISBN: 978-3-95806-762-2

Band / Volume 635

In Situ Time Calibration for Stationary Multichannel GPR Monitoring Systems

L. Steinbeck (2024), xvi, 98, xxxi pp

ISBN: 978-3-95806-767-7

Band / Volume 636

Erneuerbares Methanol als Ausgangsstoff für die Bereitstellung von flüssigen Kraftstoffen für den Transportsektor

F. Schorn (2024), VI, 275 pp

ISBN: 978-3-95806-769-1

Band / Volume 637

Investigation of Lower Boundary Conditions of Brominated Very Short-lived Species (VLS)

S. Zheng (2024), 2, iii, 160 pp

ISBN: 978-3-95806-770-7

Band / Volume 638

Modellgestützte Analyse zukünftigen Mobilitätsverhaltens

J. P. Reul (2024), XVI, 291 pp

ISBN: 978-3-95806-771-4

Band / Volume 639

Insights into Mechanisms of Secondary Organic Aerosol Formation: Approaching Atmospherically Relevant Conditions in an Atmospheric Reaction Chamber

Y. Baker (2024), XVII, 122 pp

ISBN: 978-3-95806-776-9

Band / Volume 640

Advancing the representation of agricultural systems in Land Surface Models: systematic model evaluations and technical model developments

T. S. Boas (2024), xxi, 145 pp

ISBN: 978-3-95806-777-6

Band / Volume 641

Imaging spatial and temporal soil water content variations of the soil-plant continuum using ground penetrating radar

L. Lärm (2024), xii, 303 pp

ISBN: 978-3-95806-778-3

Band / Volume 642

Development of Iridium-based Nanostructures for Oxygen Evolution Reaction in PEM Water Electrolysis

S. Park (2024), 135 pp

ISBN: 978-3-95806-779-0

Band / Volume 643

Multi-dimensional GPR full-waveform inversion for small-scale hydrogeophysical soil characterization

D. Hoven (2024), IX, 163 pp

ISBN: 978-3-95806-781-3

Band / Volume 644

Analyse des Gastransports in komplexen Membransystemen durch Modellierung und multiskalige Simulation

K. Wilkner (2024), VIII, 122 pp

ISBN: 978-3-95806-784-4

Band / Volume 645

Deployment of Fuel Cell Vehicles in Road Transport and the Expansion of the Hydrogen Refueling Station Network: 2024 Update

T. Grube; M. Rex (2024), iii, 26 pp

ISBN: 978-3-95806-786-8

Band / Volume 646

Modellgestützte Analyse treibhausgasneutraler Transformationsstrategien für Deutschland

T. F. Schöb (2024), XII, 228 pp

ISBN: 978-3-95806-789-9

Band / Volume 647

Future Distribution Grids Using Geo-Referenced Synthetic Network Topologies

A. Bandam (2024), ix, 237 pp

ISBN: 978-3-95806-790-5

Energie & Umwelt / Energy & Environment
Band / Volume 647
ISBN 978-3-95806-790-5