Literature Review on Representative Distribution Network Architecture via Synthetic Network Models

Mir Toufikur Rahman, Usman Bashir Tayab, Kazi N. Hasan, and Mahdi Jalili

November 28, 2023

RMIT UNIVERSITY

Executive Summary: Literature Review

Synthetic power networks are developed to study the impacts and benefits of new energy technologies in a location-specific electrical network. While the network topologies, models, and assets are proprietary and, in many cases, are not publicly available, synthetic networks are developed using open-source data. In the pursuit of developing a synthetic network for grid integration and network flexibility studies, this report presents a comprehensive review of existing methodologies for synthetic distribution network developments. This report also provides a thorough appraisal of various validation approaches for developed synthetic networks and their demonstration in power system studies. This study aims to identify (i) the input data and software requirements for synthetic network development and (ii) the most accurate and efficient approaches for modeling and validation of existing synthetic distribution networks to be used for grid integration studies.

- This literature review identified that the GIS data-based and data-driven approaches are the most widely used methods in developing a synthetic distribution network.
- Furthermore, operational validation has been commonly utilized to measure the accuracy of the developed synthetic distribution networks.

Executive Summary: Work Package 1.4

Work Package 1.4 aims to create a comprehensive taxonomy and information database for LV and MV/HV network models across Victoria for impact assessment and power system studies considering the mass adoption of localized renewable generation and distributed energy resources, the electrification of transport, transition of domestic gas use, and uptake of new fuels. The developed networks will capture spatial characteristics and actual network topology and support different network analysis functions. These networks will be augmented by information on typical MV and HV networks made available from DNSPs.

Contents

1 Introduction

Over the past decade, power system networks, particularly distribution networks, have significantly transformed due to the integration of new energy technologies, including renewable energy resources (RES) and new types of loads (e.g., electric vehicles). According to the International Energy Agency (IEA), the developed countries are significantly investing in RESs for achieving net-zero emissions by 2050 [1]. To achieve net-zero emissions, it is essential for the energy sector to perform system-wide planning and operational studies. However, transforming desktop analysis and local trial results into regional and network topology-wide implications is a major difficulty with these technological studies. The unavailability of regional (or national) power system network models and data poses a challenge. A dispersed and incomplete set of power system networks limits the ability to test and analyze the impacts of RES integration and electric vehicles (EVs) to some extent.

The distribution network is a complex network of lines, buses, transformers, poles, junction boxes, and other equipment that delivers electricity from the high-voltage transmission grid to industrial, commercial, and domestic users. Figure 1 illustrates the typical structure of power system networks, including three major voltage levels: high (HV), medium (MV), and low voltage (LV). Distribution networks are considered the critical infrastructure of a country; hence, detailed network diagrams, asset locations, and customer data prompt privacy concerns. Consequently, very few distribution networks are publicly available that can be used as test networks. The limited availability of test networks in the literature has compelled the research community to rely only on the existing ones, often re-purposing them for objectives that extend beyond the initial aims of network creators [2]. These factors necessitate the development of synthetic networks. The concept of synthetic network modeling has been considered as an alternative to modeling complex power networks. The synthetic networks are topologically the same as the actual networks. Thus, a synthetic network can be used to assess the benefits and challenges associated with installing new energy technologies in the network.

Synthetic power distribution networks, also known as synthetic grids or synthetic power systems, are computer-generated models or simulations of real-world electrical power distribution systems. These models are created to mimic the behavior and characteristics of actual power distribution networks, but they exist in a virtual or simulated environment. Synthetic power distribution networks serve various purposes, including research, testing, training, and analysis in electrical engineering and power systems. They contribute to advancing technology, grid reliability, and resilience while minimizing risks associated with real-world experiments. Additionally, they play a crucial role in shaping the future of power distribution by supporting the integration of emerging technologies and the transition towards more sustainable and efficient power grids.

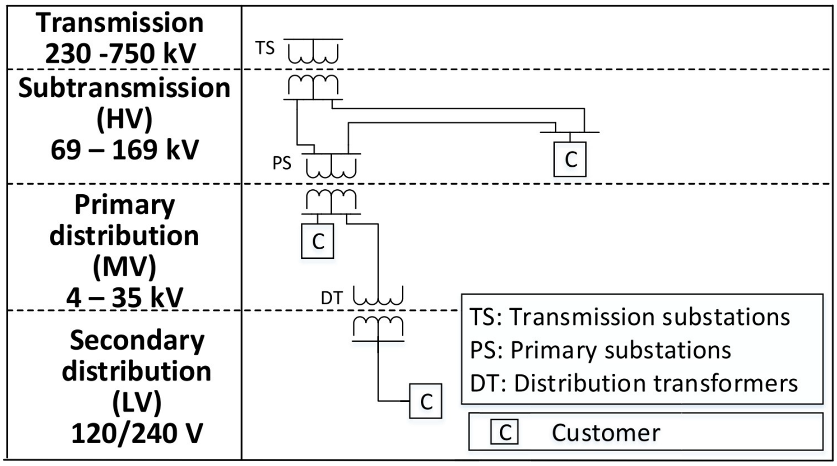

Figure 1: Structure of a typical power system network [3].

2 Synthetic Distribution Network Developments

The modeling for synthetic distribution networks has been conducted by employing various methodologies and validated through different pronged approaches. However, a comprehensive review of various methods and validation approaches for synthetic distribution networks, to the best of our knowledge, has yet to be conducted. This report provides a brief but comprehensive review of existing synthetic distribution network development, including the different approaches for modeling, validation, and the purposes of network development. In addition, the input data requirements and software used in these approaches for synthetic network modeling have been highlighted.

Figure 2 illustrates the summary of various methods, validation approaches, and demonstration examples in power system studies for developing synthetic distribution networks. Five major methods have been reported in the literature, which are (i) data-driven, (ii) random geometric graphs, (iii) geographic information systems, (iv) grid-based, and (v) hybrid approaches. Figure 2 also shows that three approaches have been utilized to validate the developed synthetic distribution network, which are (a) operational, (b) structural, and (c) statistical validation. Furthermore, the demonstration example has been categorized into three major parts based on their purpose of study, such as (A) network modeling, (B) hosting capacity, and (C) DER impact analysis.

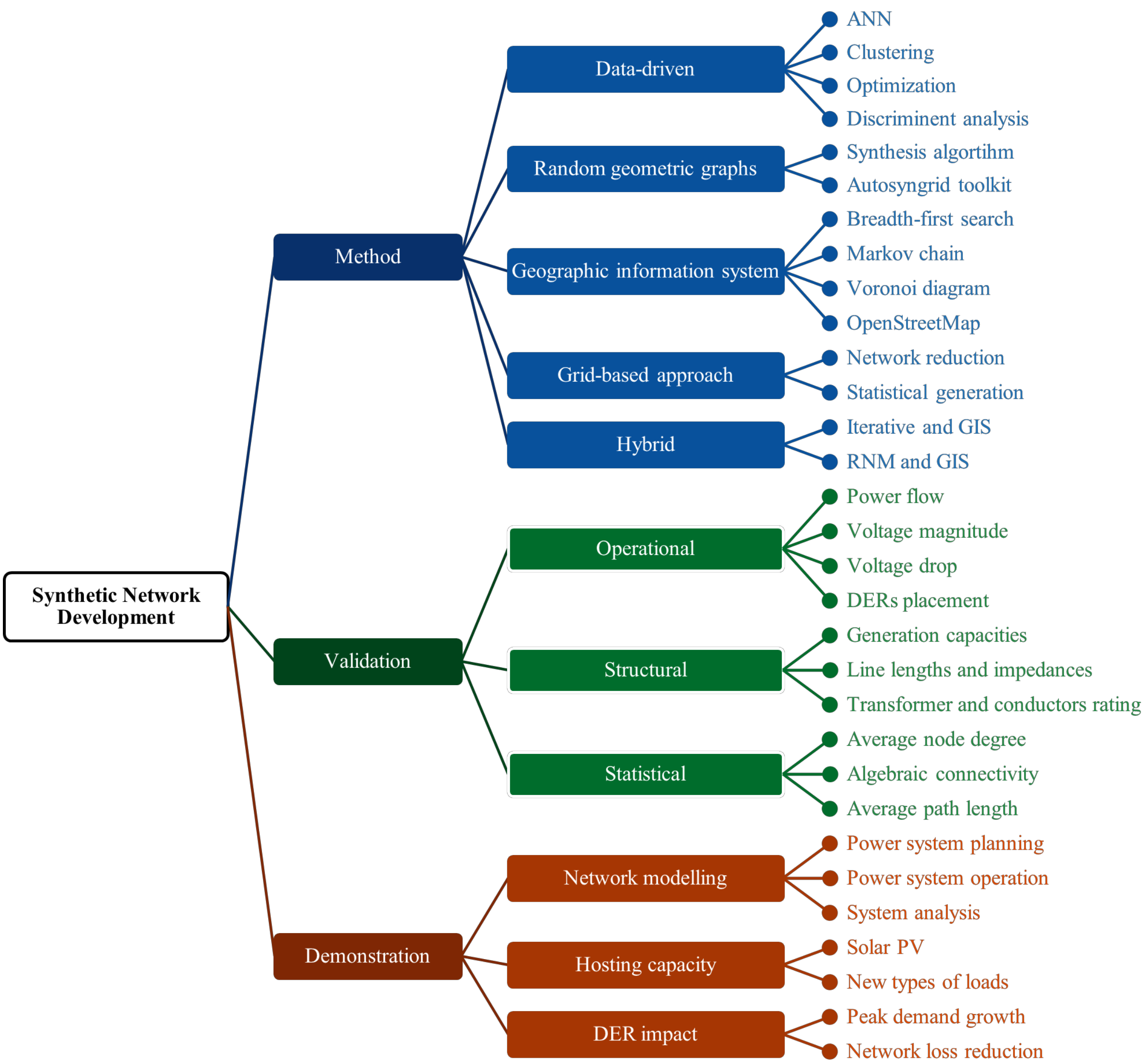

Figure 2: Summary tree showing the various methods, validation approaches, and demonstration examples of synthetic distribution network in power system studies.

3 Exiting Methodologies for Developing Synthetic Network

Developing a synthetic distribution network is essential to perform grid impact analysis of new energy technologies, including generations and loads, of a specific location if network models are publicly unavailable. Several approaches exist for creating synthetic distribution network models, as described below.

3.1 Data-driven Approaches

A data-driven approach to developing a synthetic power distribution network involves using real-world data and analytical techniques to create an electrical distribution system model or simulation. Figure 3 illustrates an application of a data-driven approach assigning load/generation profiles to a primary substation for developing a synthetic network. Data-driven synthetic network has been developed by employing a diverse range of algorithms, such as generative adversarial network [4], optimization algorithm [5], clustering algorithm [6, 7, 8, 9, 10, 11], and discriminant analysis [8]. A 60 bus, 13.8 kV synthetic distribution network has been developed by using an unbalanced graph generative adversarial network (UG-GAN) in [4], where limited actual data from the US distribution networks has been utilized to describe the topological and electrical properties of the synthetic network. In [5], a multi-objective optimization algorithm is utilized to generate a 94 bus, 22 kV synthetic distribution network. The multi-objective optimization function minimizes the investment cost of lines and substations and the operational cost of the system while considering the geographical and spatial evaluation of the Singaporean power distribution system.

In Australia, a partition-based *k-medoid* clustering algorithm has been employed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) to develop publicly available datasets and models for low-voltage [6] and medium-voltage [7] taxonomy studies. In the LV taxonomy, 23 networks have been developed varying the bus number from 2 to 152 at 0.4 kV, while in the MV taxonomy, 19 networks are found varying the bus numbers from 7 to 570 at different voltage levels of 11 kV to 33 kV. In [9], the hierarchical clustering algorithm has been employed to develop 575 synthetic distribution feeders at different voltage levels of 12.47 kV and 35 kV. This algorithm has utilized actual feeder model data from North America. Four different clustering algorithms, such as hierarchical clustering, $k-medoids^{++}$, improved $k-means^{++}$, and Gaussian Mixture Model, have been employed in [10, 11] to generate 11 synthetic distribution feeders at 0.4 kV in the North West of England. In conjunction with the clustering algorithm, a discriminant analysis has been proposed in [8] to develop 204 synthetic feeders at 22 kV and 8 synthetic feeders at 0.4 kV in the West Australian context.

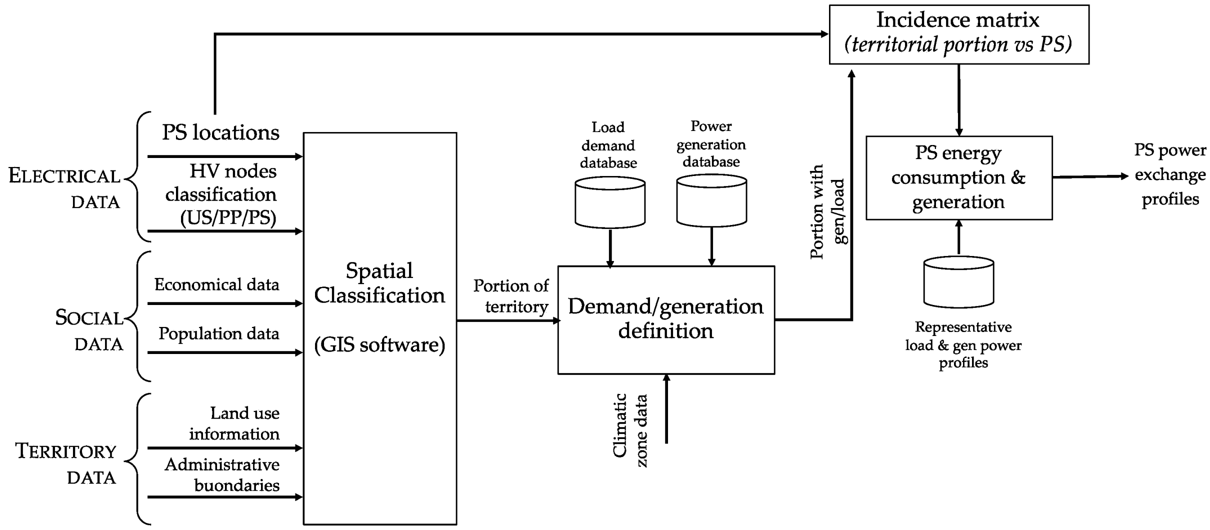

Figure 3: Illustration of data-driven approach assigning load/generation profiles to substation in a given region [12].

3.1.1 Input Data Requirements and Software

The data-driven approach for synthetic distribution network requires various data, including the number of nodes [4, 6, 7], total line length [6, 7, 10, 11, 8], geographical information [4, 5], the ratio of overhead lines [6, 7], type of customer, active and reactive power, line current and DER penetration [10, 11], the number of edges [4], spatial evaluation data [5], transformer capacity [8], voltage level and climate zone [9], and the number of customers per distribution transformer, the number of customers, and peak demand [8]. These data have been processed to develop synthetic networks using different software platforms, including OpenDSS [4, 6, 10, 11], MATLAB [4, 10, 11], MATPOWER [5], GridLab-D [9], TensorFlow [4], GAMS [5], and SINCAL and PowerFactory [7].

3.1.2 Advantages and Limitations

Data-driven approaches for synthetic distribution networks leverage real-world data that includes actual network topology, equipment specifications, and historical performance. Thus, it can provide a more realistic representation of distribution network characteristics and handle the inherent complexity of distribution networks with various interconnected components. The use of actual network data expedites the synthetic network validation process and further enables the models to be updated as new data becomes available.

However, the data-driven approaches mostly lack quality data (incomplete or outdated) and specific details required for an accurate representation of the distribution network. Also, an inappropriate selection of training data could lead to unrealistic simulations and predictions that do not accurately reflect the true characteristics of distribution networks. Furthermore, ensuring privacy while using actual data for synthetic distribution networks is a significant challenge to the scientific community.

3.2 Random Geometric Graphs (RGG)

Random geometric graphs (RGG) create synthetic distribution networks by randomly placing nodes (representing substations or load centres) in a defined area. As part of the RGGs, a synthesis algorithm is proposed in [13], and a MATLAB-based AutoSynGrid toolkit is introduced in [14] to develop synthetic MV distribution networks. The synthesis algorithm in [13] considers several properties of an actual distribution network in the Netherlands, including nodes and edges, and uses this structure to explore new statistical patterns for a 195-bus synthetic network. The developed AutoSynGrid toolkit in [14] takes the input of the size and statistical properties of the actual network from the US to generate a synthetic distribution network. The developed toolkit is capable of generating synthetic networks with a wide range of bus numbers and voltage levels. Figure 4 shows a sample synthetic case created by the AutoSynGrid toolkit indicating the generation and load buses with corresponding power flows. In addition to the generated MATPOWER format case, the tool provides evaluation results for the generated cases.

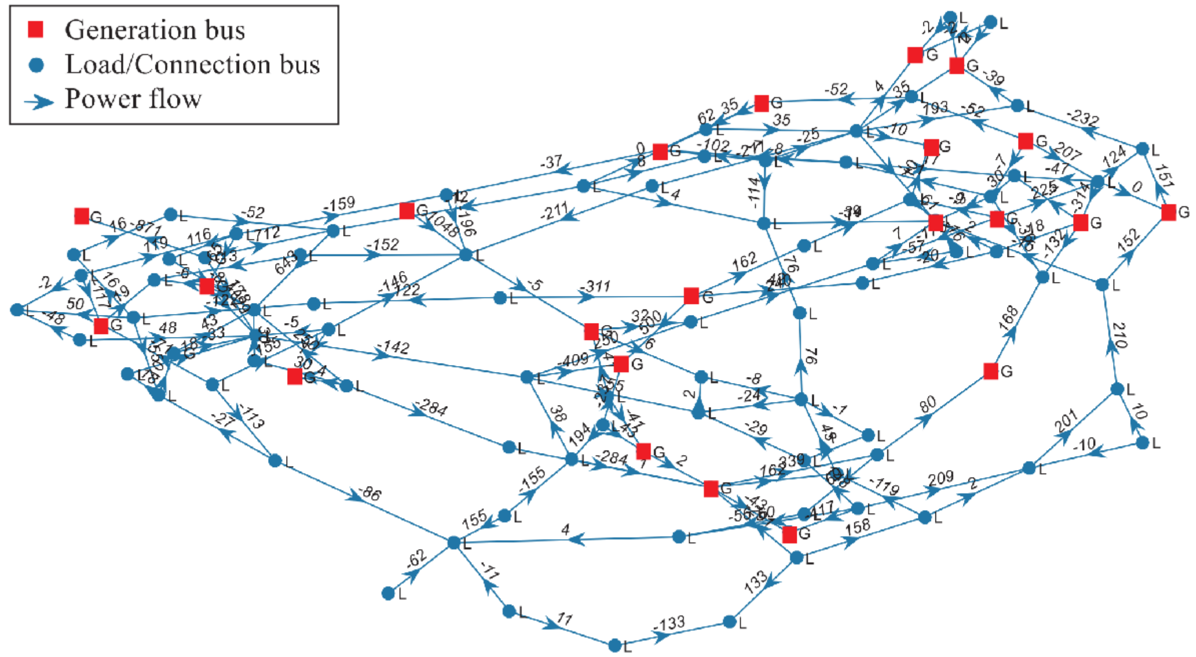

Figure 4: Illustration of RGG synthetic case created by AutoSynGrid toolkit [14].

3.2.1 Input Data Requirements and Software

The RGG approaches for the synthetic network require various data, including the node degree and hop distance from the source, load at nodes and branch currents [13], and number of buses, generation, and load settings, number of branches, bus type entropy and network loading level [14]. These data have been processed to develop synthetic networks using different software platforms, including MATLAB [14], and Phase2Phase and PostgreSQL [13].

3.2.2 Advantages and Limitations

RGGs are relatively simple models, making them easy to understand, implement, and simulate for better explorations of network behavior. They naturally capture the spatial characteristics of distribution networks and include spatial constraints, such as the minimum and maximum distances between nodes. Furthermore, random geometric graphs can be easily scaled to represent both small-scale and large-scale synthetic distribution networks.

However, RGGs for synthetic distribution networks fall short of capturing the complexities and specific features of actual networks. In RGGs, a uniform distribution of nodes is considered in the geometric space which ignores connectivity requirements, equipment capabilities, and load characteristics. Furthermore, the random nature of node placement sometimes does not accurately capture specific topological features present in real distribution networks, such as hierarchical structures, loops, or radial configurations.

3.3 Geographic Information Systems (GIS)

GIS databases provide detailed technical information, including the location of substations, feeders, transformers, capacitors, and end-user and billing information, including their location and power consumption. GIS databases have been utilized as a source of open data to generate synthetic distribution networks using breadth-first search (BFS) algorithm [15], Markov chain [16], Voronoi diagrams [17], and OpenStreetMap [18, 2, 19]. The BFS algorithm is utilized alongside the GIS data in [15] to achieve the full reconnection of a realistic model of 53 bus, 0.4 kV distribution network in the United Kingdom (UK). In [16], the Markov chain has been utilized to construct 20 number of synthetic distribution networks with different voltage levels from 0.4 kV to 6 kV based on publicly available data from the United States (US). Firstly, the nodes and edges have been connected, and then the connections of residential buildings to the substation transformer have been established based on the GIS database. In a similar fashion, a countrywide Voronoi diagram based on Switzerland is used to construct a synthetic distribution network in [17]. Figure 5 shows the result of the Voronoi partitioning in Switzerland, and it has been approximated based on their respective high-voltage (HV) service area. In the end, one synthetically generated 25 bus, 21 kV network has been demonstrated in the case of Switzerland. In [20], the geospatial and topological GIS data, including customer and asset locations, conductor types, and asset connection graphs have been utilized to develop 10 LV networks. In this effort, 80,000 LV networks across Australia have been analyzed based on common LV network topologies and features that influence PV hosting capacity.

OpenStreetMap (OSM), also a geographic database, is a popular tool used in [2] for developing a 311 bus, 0.4 kV synthetic distribution network based on the Colac region in Australia. The complete network is visualized in QGIS software, as shown in Figure 6.

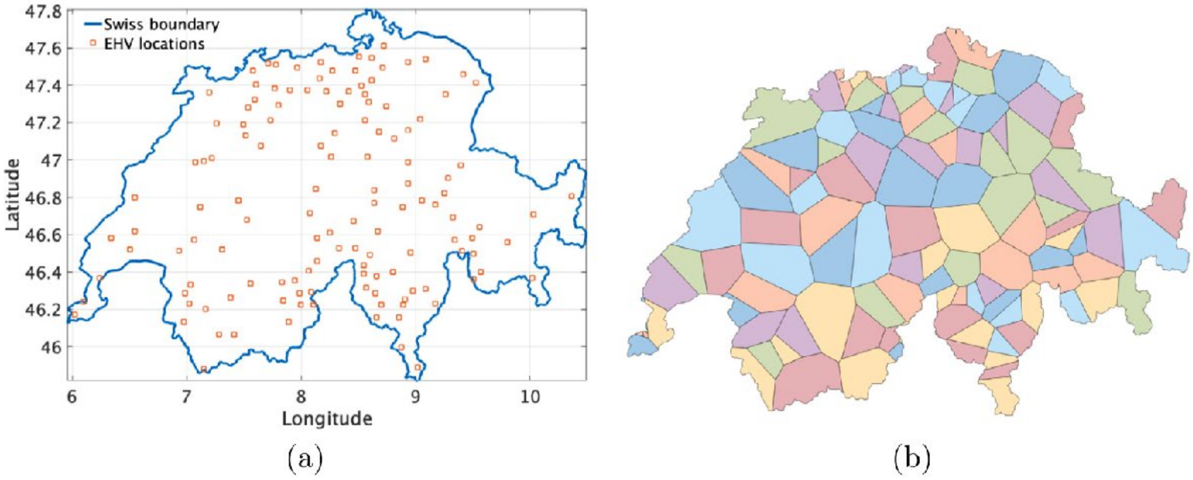

Figure 5: Demonstration of GIS: (a) Locations of the considered 148 substations in Switzerland and (b) approximated regions served by each substation after Voronoi partitioning [17].

The catalogue of network elements includes 4714 power lines, 48 distribution transformers, 4155 energy consumers, 609 electrical nodes, and one substation. The OSM has also been utilized in [18] to develop 124 numbers of 20/0.4 kV substations across Europe along with an example of a 177 bus LV distribution network. Similarly, research in [19] proposed a framework for generating synthetic networks and demonstrated a single substation model using OSM.

Figure 6: Illustration of a synthetic distribution network for Colac region in Australia [2].

3.3.1 Input Data Requirements and Software

GIS data-based approaches for the synthetic network require various data, including the number of nodes [15, 16], the number of customers [20] geographical data [17, 16, 20], substation location [17, 18], line segment [15], the number of edges and transformers, [16], transformer rating [20], length and route of the lines and voltage level [18], HV feeder type and conductor type [20], postal codes and the number of loads and generators [19]. These data have been processed to develop synthetic networks using different software platforms, including MATLAB [15, 17, 18], OpenDSS [19], Pandapower [20], and QGIS [18].

3.3.2 Advantages and Limitations

GIS data-based approaches provide high spatial accuracy, allowing for precise representation of the geographic layout of the synthetic distribution networks. These approaches also allow the integration of various spatial data sources, such as topography, land use, and infrastructure layers. Moreover, GIS data-based approaches support network analysis functions by calculating distances, optimal routes, and connectivity patterns, and various scenario modeling by manipulating spatial data.

However, the complexity of spatial data and integration challenges are some limitations of GIS data-based approaches for synthetic distribution networks. While modeling a large-scale distribution network, handling and managing spatial data in GIS software will be a major challenge. Besides that, integrating GIS software with other systems, such as simulation models or real-time data feeds, can be challenging.

3.4 Grid-Based Approaches

Grid-based layouts have been utilized to generate synthetic distribution networks in [21, 22, 23, 24], where nodes and lines are arranged in a grid pattern. As part of this grid-based approach, a statistical network generation algorithm based on fractal theory has been used in [21] to reproduce realistic multivoltage-level distribution in the UK context. Four voltage levels of 33 kV, 11 kV, 6.6 kV, and 0.4 kV have been considered in developing a 1094 feeders distribution network. Figure 7 shows the actual MV network in the UK and the synthetic network layout using a grid-based approach. It can be seen that the areas on the upper and lower right correspond to the rural type area, and the area on the lower left shows a higher load density urban area. An application of the stochastic method has been demonstrated in [22] to generate 2119 synthetic feeders with voltage levels of 15 kV and 20 kV. Similarly, the stochastic method has been used in [23, 24] to generate synthetic distribution networks based on a preliminary study of actual network models across Italy.

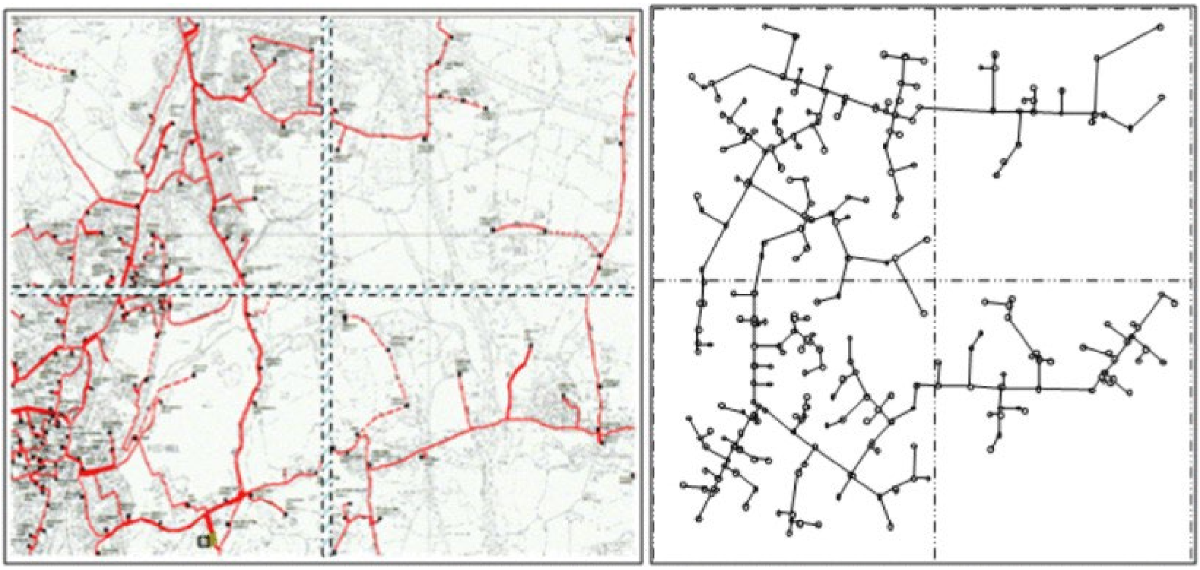

Figure 7: Illustration of a grid-based approach for synthetic network development: real MV network (left) and synthetic MV network (right) [25].

3.4.1 Input Data Requirements and Software

Grid-based approaches for the synthetic network require various data, including substation transformers [22, 23, 24, 21], the number of lines and loads, and the installed generators [22, 23, 24], the consumer settlement control parameters and load density [21].

3.4.2 Advantages and Limitations

Grid-based approaches for synthetic distribution networks provide a structured and systematic representation of the developed network. The grid layout allows for organized placement of network components, making it easier to understand and analyze. Grid-based approaches can be easily scaled to represent both small- and large-scale distribution networks.

However, grid-based approaches sometimes oversimplify spatial representation and lack the flexibility to capture the complexities of real-world distribution networks accurately. Grid-based approaches rely on spatial discretization, where continuous space is represented by discrete grid cells. This discretization may lead to a loss of accuracy in representing the intricate details of the distribution network. Moreover, grid-based approaches may oversimplify the representation of distribution networks, particularly in cases where the actual network exhibits complex spatial relationships.

3.5 Hybrid Models

Hybrid models may combine elements from various approaches, such as GIS-based models and machine learning techniques to create more realistic synthetic distribution networks. These hybrid modeling approaches have been previously applied to model the synthetic distribution network combining iterative method and GIS-based models [26], reference network model (RNM) and GIS data-based models [3, 27, 28, 29]. An iterative complex network analysis and graph theory have been combined in [26] to identify the topological attributes of an actual MV distribution network from China. Then, 30 synthetic feeders at 10 kV were modeled as graphs with nodes, then connecting lines between various nodes.

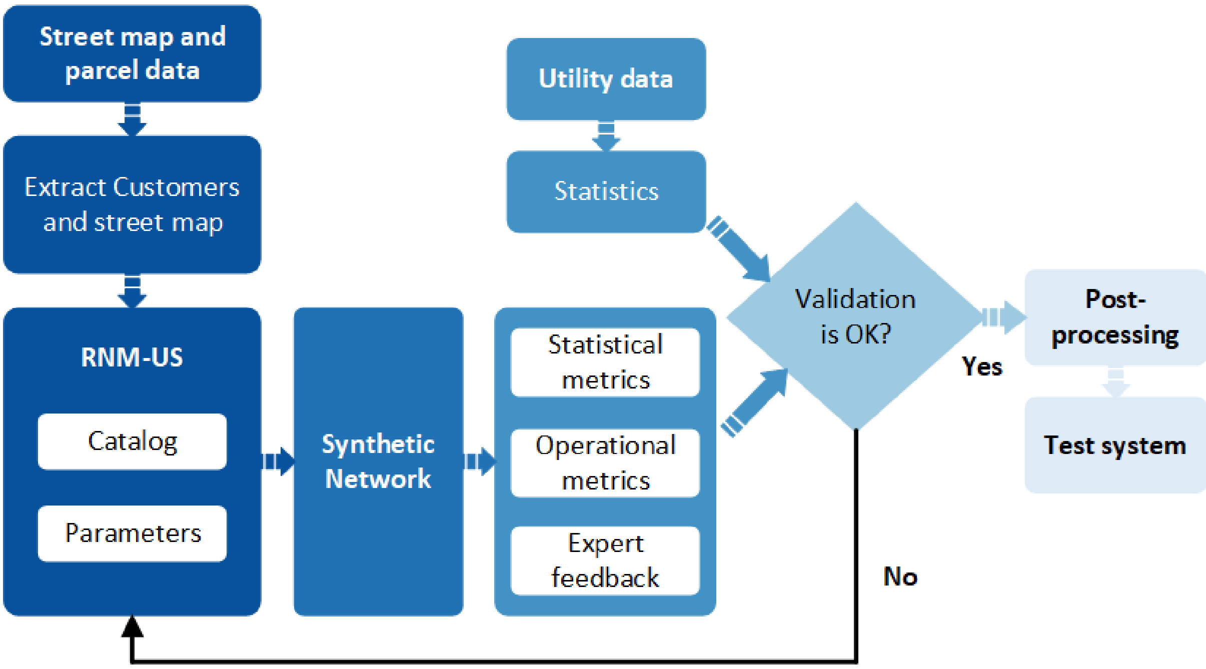

Figure 8: Demonstration of RNM method: Planning stages in RNM-US [3].

On the other hand, RNM is an established planning and analysis tool for large-scale distribution networks that need to interact with the GIS database to establish the optimal layout of power lines complying with the street map. The main features of an RNM tool have been outlined and demonstrated an automatic generation of street maps from the GIS module of the RNM in [30, 29, 3]. The methodology applied to build synthetic distribution systems is illustrated in Figure 8. The process starts with parcel or land use information and building footprint and height data using OSM. This information is used for labeling consumer load type (household, supermarket, hotel, restaurant, or school) and estimating the location of the consumers. The peak load is obtained through linear interpolation by using a set of reference building models, assuming that the energy consumption is correlated with the building volume. The system layout follows the layout of streets, which are obtained from publicly available street maps. The information is validated using a catalog of components and design criteria based on actual data from utilities. Finally, the process outputs GIS files for system visualization, OpenDSS files for power flow simulation, and CSV files describing the system components. The RNM software tool has also been used in [28] to synthetically generate the IEEE 8500 bus 12.47 kV distribution network using synthetic networks produced by RNM-US. In the RNM tool, heuristic methods have been used to determine the optimal number and sequence of phases at each feeder and allocate the phase to each final user, minimizing the unbalance [28].

3.5.1 Input Data Requirements and Software

Hybrid models for the synthetic network require various data, including the number of lines and length [29, 3, 26], distribution transformers [29, 3, 26], switchgear [26], voltage levels [29, 3, 28], loads [29, 3, 28], street maps [29, 3, 28], consumers' location, the number of phases, capacitors and the number of consumers per feeder [29, 3], and generators locations and topography [28]. These data have been processed to develop synthetic networks using different software platforms, including OpenDSS [29, 3], MATLAB [26] and RNM-US software [28].

3.5.2 Advantages and Limitations

Hybrid models basically provide better flexibility by combining the strengths of different approaches and can capture both spatial and data-driven aspects. Thus, hybrid models can provide a more accurate representation of the distribution networks. However, the integration of different modeling approaches introduces complexity in terms of model development, validation, and interpretation. For instance, combining spatial and data-driven models requires consistent and compatible datasets, which may require extensive pre-processing and demand more computational resources.

RNMs use existing real-world networks as a baseline and incorporate real data, such as infrastructure details, operational characteristics, and historical performance, providing a realistic starting point and more accurate simulations for the synthetic distribution network. However, RNMs rely on existing data, which may have limitations in terms of completeness, accuracy, or relevance [16]. Incorporating real-world data into RNMs also raises privacy concerns while handling sensitive information about infrastructure, operations, and consumers.

Table 1 summarizes the advantages and limitations of existing methods for synthetic network developments. It is seen that actual data and network topology-based methods (i.g. data-driven and grid-based approaches) capture both the spatiotemporal aspect and can analyse system vulnerability. They can also provide an accurate representation of the actual networks and are capable of handling complexity. A common limitation of these methods is that they require detailed network information, including locations, generation, loads and line parameters.

4 Validation Approaches for Synthetic Network

Validation of synthetic distribution network models is crucial to ensure that they accurately represent real-world distribution systems and produce reliable results for various analyses and simulations [3, 31]. Several validation approaches and techniques have been implemented in the literature, which is summarized in Table 2. It is seen from Table 2 that a large number of literature have considered operational validation followed by

Table 1: Advantages and Limitations of Existing Methodologies for Synthetic Network Development.

Methods	Advantages	Limitations
Data-driven	✓ Actual network topology ✓ Equipment specifications ✓ Handle complexity ✓ Easy validation ✓ Model can be updated	✗ Lack of quality data (incomplete or outdated) ✗ Ensuring privacy of actual data ✗ Selection of training data
RGG	✓ Relatively simple model ✓ Capture spatial characteristics ✓ Include spatial constraints ✓ Flexible in scalability	✗ Ignore specific features of actual network ✗ Ignore connectivity requirements (mesh or radial) ✗ Does not capture specific topological features
GIS data-based	✓ High spatial accuracy ✓ Allow various spatial data sources ✓ Allow scenario modeling ✓ Support network analysis functions	✗ Complexity of spatial data ✗ Higher challenges for large-scale ✗ Complex integration of simulation models/analysis
Grid-based	✓ Provide a structured representation ✓ Organized placement of components ✓ Flexible in scalability	✗ Lack the flexibility to capture the complexities ✗ Ignores complex spatial relationships ✗ Less accurate
Hybrid models	✓ Accurate representation ✓ Capture spatial and network topology ✓ Incorporate actual data ✓ More accurate simulations	✗ Lack of quality data ✗ Ensuring privacy of actual data ✗ Demand high computational resources ✗ Complexity in terms of model development ✗ Extensive data pre-processing

Table 2: Validation Approaches.

Approaches	Attributes
Operational Validation	Power flow analysis [4, 6, 7, 9], Voltage profiles [16, 19, 17], Line current magnitude [17], DERs placement [4, 6, 7], Downstream power [13], Voltage drop magnitude [13, 15], Minimal voltage band [18], Voltage histograms and voltage-distance plots [27], Total loss and load profiles [22, 23, 24]
Structural Validation	Generation capacities, loads, and line impedances [14, 21], Lengths of the lines [2, 21], Ratings of conductors and transformers [21], Number of nodes and the network length per kilometer [26]
Statistical Validation	Average node degree [5, 14], Sampling error [10, 11], Historical voltage readings [20], Algebraic connectivity and average path length [14]

structural and statistical validation for synthetic network development.

4.1 Operational Validation

The operational validation ensures the synthetic networks' suitability to be used by the scientific community to aid in their research. In literature, various operational data, such as voltage profile [16, 19, 17, 27], voltage drop magnitude [13, 15, 16, 19, 17, 27], minimal voltage band [18], total losses, load profiles and reverse power flow [22, 23, 24], power flow analysis [4, 9], DER placements [4, 6, 7], and current magnitude [17] have been reported while validating a synthetic network.

The resulting voltage profiles in the synthetic network are observed in [16, 19, 17, 27] and compared with the actual network. Voltage histograms for each region have been separately compared in [27]. The voltage drop magnitude along the network is one of the most indicative parameters to validate a synthetic network which has been used in [16, 19, 17, 13, 15, 27]. The voltage drop over the distance between the synthetic and actual network is particularly compared in [27] for validation purposes. In [18], the minimal voltage band throughout the synthetic network is computed through power flow simulation for an optimal substation position, which is then considered as the fitness of the generated grids.

Furthermore, the developed network is compared with the real network in [22, 23, 24] in terms of the total losses, the load profiles, and the number of transformers experiencing reverse power flow towards the transmission level. In [4], the developed synthetic networks are validated by performing power flow analysis, DERs placement, transmission, and distribution power flow co-analysis and comparing results with the actual network. Similarly, the power flow analysis [9] and DERs placement have also been studied in [6, 7]. The line current magnitudes are also important operational data of the synthetic network, which has been compared with the actual distribution network in [17].

4.2 Structural Validation

Structural validation of a synthetic network involves assessing its physical and logical design to ensure it accurately represents the real-world power network. Various structural approaches have been previously used for the synthetic network validation, such as comparing network assets/statistics between synthetic and actual network [14, 2, 26], and alternative design options mimicking real distribution networks [21]. The probability distribution that identifies common and uncommon parameters between synthetic and actual distribution systems has been used in [14]. The network generation capacities, loads, and line impedance are considered. Similar topological properties have been compared with the actual network to validate the developed synthetic network in [26]. In this case, the number of nodes and the network length per kilometer, average edge length, and average path length have been considered. In [2], the lengths of power distribution lines in synthetic and actual networks are compared to validate the synthetic network besides industry and academic experts' feedback.

4.3 Statistical Validation

Statistical validation involves performing statistical tests to assess the similarity between synthetic and real distribution network characteristics. These statistical tests include various network attributes, such as the average node degree [5, 14], sampling error [10, 11]. The average node degree has been considered for selected samples of the synthetic network in [5] to statistically compare an actual distribution network. In addition to the average node degree, algebraic connectivity and average path length have been considered in [14] to validate the synthetic network. In [10, 11], sampling error (the level of precision of a sample) of the number of feeders has been considered to statistically validate the synthetic network. In [20], the developed models were then tested against historical conditions and validated based on historical voltage readings to ensure their accuracy.

5 Demonstration Examples of Synthetic Network

Synthetic distribution network models are valuable tools in power system studies for a wide range of applications. These models offer flexibility and control and facilitate conducting experiments and analyses when the required data are not available. Key applications of synthetic distribution network models in power system studies are summarized below.

5.1 Network Modeling

Modeling a synthetic power system network involves creating a computer-based representation of a power network that closely mimics the behavior of an actual network. This type of modeling is generally conducted for various purposes, including power system planning [5, 2, 26], power system operation [6, 9, 8], system analysis [30, 15, 19, 21], and developing simulation tool [13].

While the network data are not publicly available due to privacy concerns, synthetic distribution networks can be used in power systems to ensure the efficient and reliable supply of electricity to end-users, maintaining safety and cost-efficiency. A synthetic network has been developed in power system planning process for ensuring supply-demand equilibrium, acceptable voltage limits, evaluating the need for new substations or upgrades to existing ones, forecasting the future electricity demand [6], conducting cost-benefit analysis [30], and exploring distribution automation [13] and smart grid technologies [9]. Furthermore, the distribution networks are often subject to industry-specific regulations and requirements, which have been conducted using a synthetic network. In power system operations, a synthetic distribution network has been developed for load management to avoid overloading and maintain voltage limits, implementing demand response programs, monitoring power quality, and data management and analysis for predictive maintenance.

Table 3: A Summary of Existing Synthetic Distribution Networks.

Study	Bus/Nodes	Voltage	Test system	Software platform	Year	Ref
	94 Bus	22 kV	Singapore	GAMS and MATPOWER	2018	[5]
	2 to 152 Bus [23 networks]	0.4 kV	CSIRO LV Taxonomy, Australia	OpenDSS	2021	[6]
	30 Feeders	10 kV	China	MATLAB	2018	[26]
Network Modeling	195 Bus	N/A	DSOs in the Netherlands	Phase2Phase and PostgreSQL	2017	[13]
	8 LV and 204 MV Feeders	0.4 kV and 22 kV	Western Australia, Australia	N/A	2014	[8]
	311 Bus	0.4 kV	Victoria, Australia	QGIS software	2023	[2]
	53 Bus	0.4 kV	UK	OpenDSS and MATLAB	2015	[15]
	575 Feeders	12.47 kV - 35 kV	US	GridLAB-D	2009	[9]
	1094 Feeders	33/11/6.6/0.4 kV	UK	N/A	2014	[21]
	IEEE 8500 Bus	12.47 kV	US	RNM-US software	2019	[28]
	4 to 10000 Bus	4 kV - 25 kV	Three metropolitan areas in the US	Python and OpenDSSDirect	2021	[27]
	10000 Feeders	120/240 V	US (New Mexico, North Carolina, California)	OpenDSS	2020	[31]
	49,16,869 Bus	4 kV and 35kV	North Carolina, San Francisco Bay Area, US	OpenDSS and MATPOWER	2020	[3]
	20 Networks	0.48 kV and 6 kV	Southwest Virginia, US	N/A	2022	[16]
	N/A	N/A	US (NYISO, ERCOT, WECC)	MATLAB	2022	[14]
	N/A	N/A	US	OpenDSS	2019	[19]
	N/A	0.4 kV and 20 kV	Spanish distribution networks	Reference network model tool	2011	[30]
Hosting Capacity	11 Feeders	0.4 kV	Electricity North West Limited (ENWL), UK	MATLAB and OpenDSS	2016	[10, 11]
	7 Bus to 570 Bus [19 networks]	11 kV and 33 kV [Three voltage levels]	CSIRO MV Taxonomy, Australia	SINCAL and DIgSILENT PowerFactory	2015	[7]
	55 Bus	20 kV	Switzerland	MATLAB	2021	[17]
	2119 substations	15 kV and 20 kV	Italian Grid	N/A	2020	[22]
	10 to 100 Bus [10 networks]	0.4 kV	Victoria, Australia	PandaPower	2020	[20]
DER Impact	60 Bus	13.8 kV	US	OpenDSS, MATLAB and TensorFlow	2023	[4]
	177 Bus	0.4 kV	Europe	MATLAB	2017	[18]
	3900 Feeders	N/A	Italian MV distribution system	N/A	2021	[23, 24]
	8500 Bus	1 kV and 36kV	European grid	MATLAB	2018	[29]

5.2 Hosting Capacity

Hosting capacity in a distribution network generally refers to the maximum amount of injected power by the DERs, such as solar PV, wind, and battery storage systems, or consumed power by loads, such as electric vehicles while maintaining the network's performance standards. In literature, synthetic distribution networks have been developed to simulate different scenarios and evaluate the network's ability to host additional power from solar PV [10, 11, 7, 17, 20], and new loads [22].

While identifying solar PV hosting capability using the synthetic network, various factors have been considered, such as power swings [7], voltage limits [22, 20], thermal limits [10, 11, 7, 17], and location of the network [17].

5.3 Distributed Energy Resources (DER) Impact Analysis

In today's power system network, the impact analysis of various DER technologies is crucially important for utilities and grid operators; however, such studies are not publicly available due to data privacy concerns. In this case, the synthetic distribution networks have been successfully utilized to investigate the DER impacts on peak demand growth [29], new loads and generations in different types of the networks [24], network loss reduction [4], and the local substation position [18].

6 Challenges in Synthetic Distribution Network Development

Developing a synthetic distribution network, whether for research, testing, or simulation purposes, comes with several limitations and challenges. These limitations can vary depending on the specific goals and constraints of the project, but some common limitations include data quality, generalization, uncertain conditions, scalability, security and privacy concerns, and validation and benchmarking.

Synthetic distribution networks are often created using generated data or simplified assumptions, mostly ignoring real-time data. This can result in data that may not accurately represent real-world conditions and sometimes may not fully replicate the dynamics of live distribution systems. Furthermore, synthetic networks are mostly designed for specific scenarios or applications; thus, they lack in capturing real-world distribution systems' full complexity and diversity.

In reality, distribution networks are subject to various uncertainties, such as weather conditions, equipment failures, and load fluctuations; however, replicating these uncertainties in a synthetic network can be difficult, and it may require the use of probabilistic analysis. The size of the developed network is also an important aspect of synthetic network development. A large-scale synthetic distribution network sometimes requires significant computational resources, and the resulting models may not be easily scalable to match the complexity of real-world distribution networks.

Furthermore, generating distribution networks sometimes involves various sensitive information, such as infrastructure details or customer data, which should be maintained with additional privacy and security. Finally, validating the accuracy of a synthetic distribution network model can be complex. There may be limited real-world data available for bench-marking, and validating the model's performance against real-world conditions can be challenging.

7 Proposed Approach for Developing Victorian Synthetic Distribution Network

Work package 1.4 (WP 1.4) aims to create a comprehensive taxonomy and information database for LV and MV/HV network models across Victoria and through the interfaces with transmission grid supply points. The developed network models will be used in the subsequent work packages of the "Enhanced System Planning" project to perform power system studies and techno-economic evaluation primarily.

7.1 Outline of the Proposed Methodology

The first step of this WP 1.4 is reviewing the CSIRO LV network taxonomy models [6] and mapping those into different distribution network service provider (DNSP) regions across Victoria. Powercor LV taxonomy models [20] will also be considered. Different types of networks will be chosen and developed with all relevant data based on the network characteristics. The networks will include urban, short-rural, long-rural, overhead, underground, strong, and weak networks. These networks will be augmented by information on typical MV and HV networks made available from DNSPs.

The proposed high-level architectural approach is shown in Figure 9. In order to achieve these tasks, WP 1.4 will employ a hybrid model combining grid-based and GIS data-based approaches for synthetic network development. Previously, in the literature, grid-based approaches have been utilized to generate synthetic distribution networks in [21, 22, 23, 24]. In these studies, the actual network topology, data, and network characteristics have been obtained from the UK [21] and Italy [23, 24, 22]. Furthermore, GIS data-based approaches have been successfully employed in the literature as a source of open data to generate synthetic distribution networks in [15, 16, 17, 18, 2, 19]. GIS data-based approaches require geographical and network information, such as the location of substations, feeders, transformers, capacitors, and end-user and billing information, including their location and power consumption.

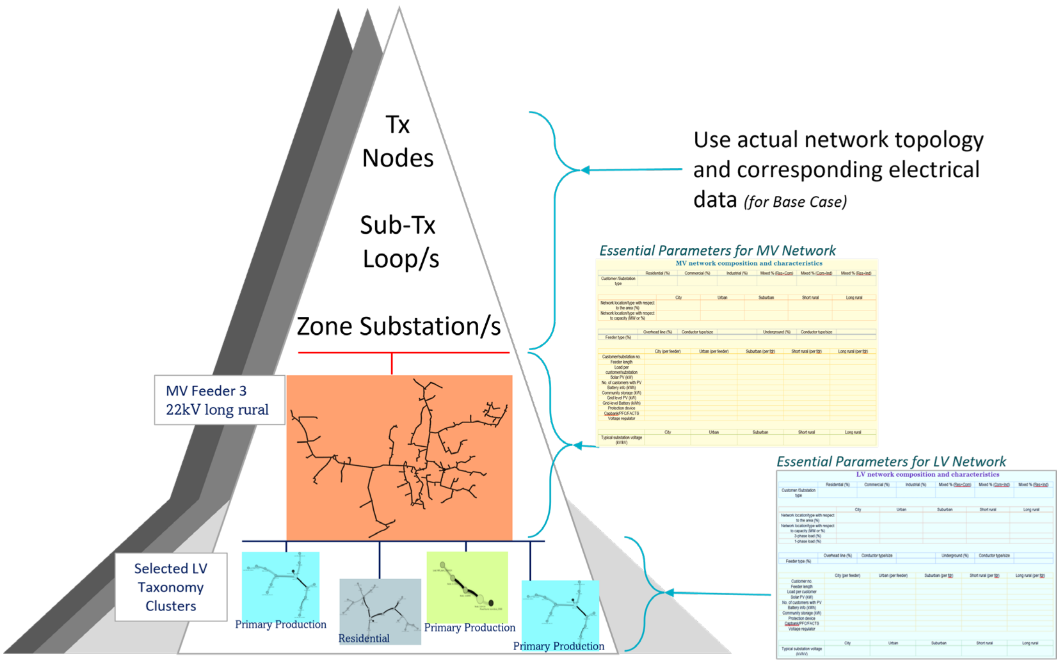

Figure 9: The proposed high-level architectural approach for Victorian synthetic distribution network.

7.2 Development of MV and HV Networks

The grid-based approach (discussed in Sub-section 3.4) includes actual transmission and sub-transmission networks up to 66 kV across Victoria. Besides the network topology, the corresponding actual network data (for the Base Case) will also be included in de-

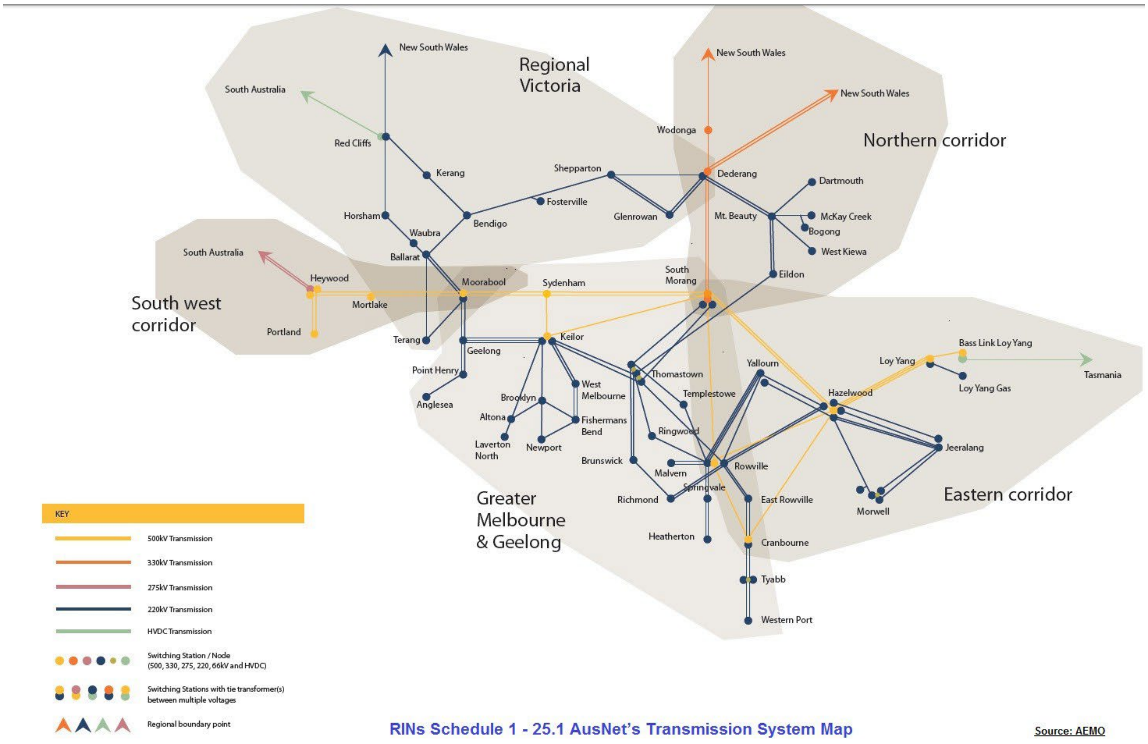

Figure 10: Ausnet's transmission system map [32].

veloping the synthetic network. Figure 10 illustrates the transmission network, including the geographical location in Victoria provided by AusNet Services [32].

Then, WP 1.4 will acquire several actual MV network models up to 22 kV from Victorian DNSPs, including their geographical information and network characteristics. In addition to the 22 kV MV network, several 11 kV and 6.6 kV networks exist in Victoria. These MV networks will include urban, short rural, long rural, overhead, underground, strong, and weak networks across Victoria. At this stage, WP 1.4 will employ the grid-based approach to select a combination of MV networks to be connected at different locations of the sub-transmission network by following the GIS database approach (as discussed in Sub-section 3.3). Figure 11 and Figure 12 illustrate the MV network map across Victoria, supplied by CitiPower and Powercor and Ausnet Services, respectively.

Generally, these networks are operated in radial mode. In some urban areas, they can often be operated in open-loop arrangement via switches installed to provide alternative points of supply and thus improve the reliability of the network. In rural areas, the average network length is 156 km (including spurs) with few alternative points of supply.

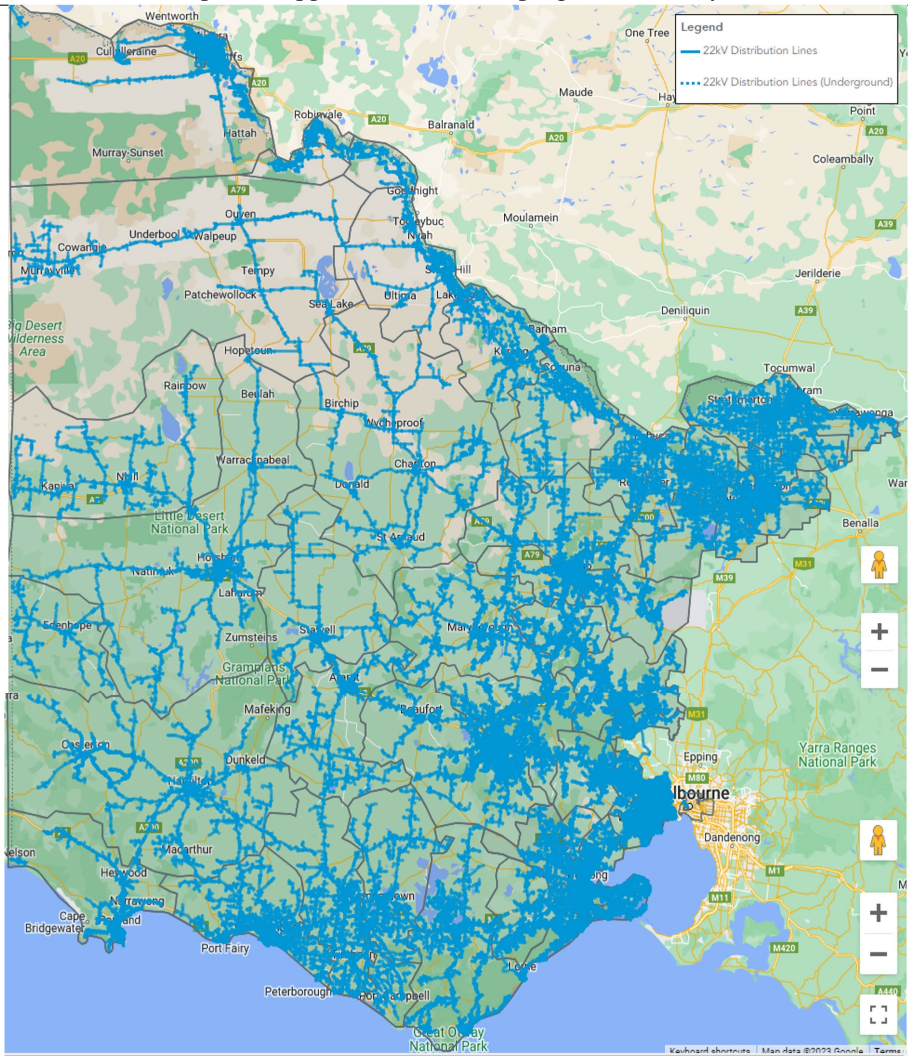

Figure 11: The MV distribution network area, Central to Western Victoria.

These MV networks are usually three-phase, but some spur lines, especially in rural areas, are single-phase supplied from two of the three available phases. Remote and low population density rural areas are often supplied by Single Wire Earth Return (SWER) MV distribution networks.

7.3 Development of LV Networks

In a similar fashion, a combination of several CSIRO LV taxonomy feeders and Powercor taxonomy networks have been picked and connected at different selected load points of MV feeders as shown in Figure 9 by following the GIS database approach. Nine feeders have been selected for WP 1.4 from the CSIRO LV taxonomy: (*i*) C (city, residential), (*ii*) E/H (city, residential), (*iii*) K (suburban, primary production), (*iv*) L

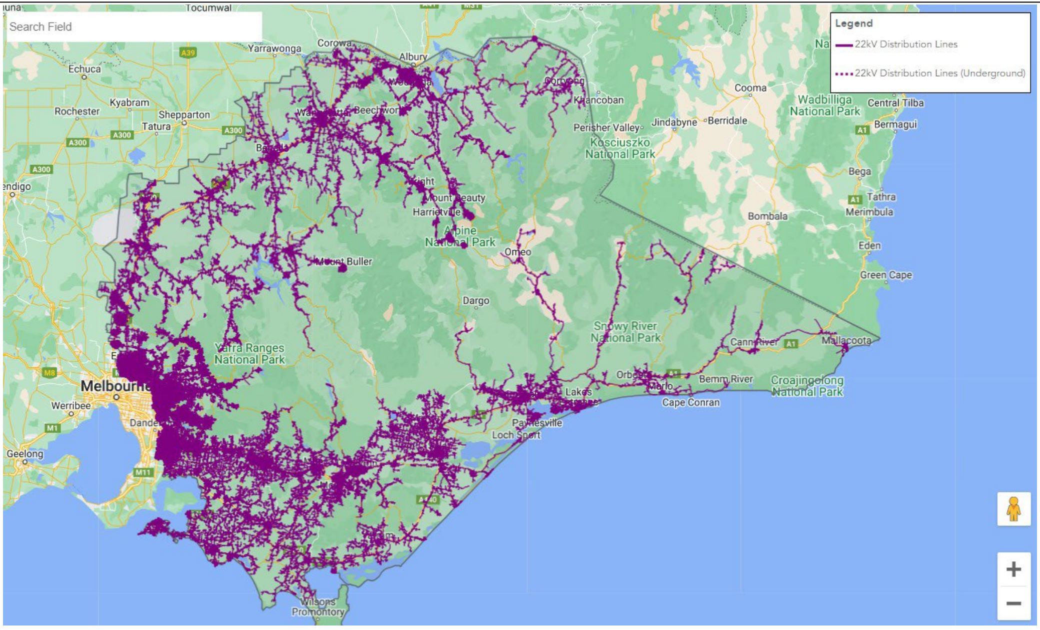

Figure 12: The MV distribution network area, central to Eastern Victoria.

(suburban, residential), (*v*) N (suburban, residential), (*vi*) Q (suburban, residential), (*vii*) S (regional, primary production), (*viii*) U (regional, primary production), and (*ix*) V (regional, residential). Figure 13 visualizes an example LV network (network C) and the basic information of the network C is presented in Table 4.

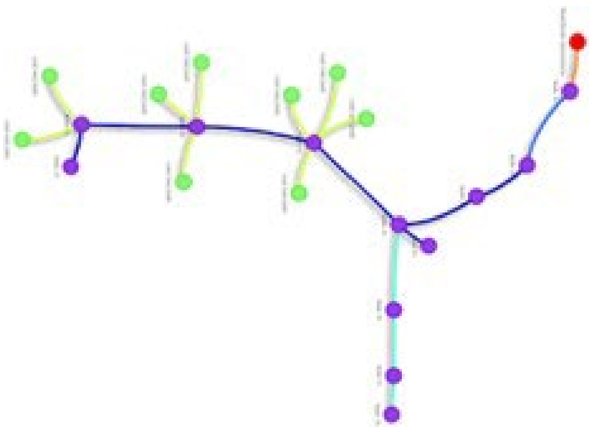

Figure 13: The LV network model for network C.

Table 4: Specifications of network C from CSIRO LV Feeder Taxonomy.

Parameters	Value
Number of loads	9
Number of lines	11
Total length	0.27 km
Number of nodes	12
The ratio of overhead lines	1

In this stage, the whole Victorian network's aggregated network view will be developed. The developed models will contain network information in a format suitable for network studies in different tools. Thereby, the range of network parameters and their sensitivities to power system studies will be reported, such as the generation portfolio – grid supply vs. DER, system loads, storage options, and their relative capacities. Also, the parameters for different power system studies need to be recorded.

While developing the network models, WP 1.4 will ensure that all further WPs can perform simulation through benchmark power system software and validate the results.

Finally, to validate the suitability of the developed Victorian representative network models, acquired datasets, and software models, a base case simulation, results, and analysis will be demonstrated for different types of power system studies, including power flow/ voltage constraints, optimal power flow and hosting capacity analysis.

8 Conclusions

8.1 Summary of Literature Review

Mass adoption of renewable DERs and the electrification of transportation are transforming today's power distribution networks. This becomes more challenging when the scientific community has very little access to actual distribution network models and data due to privacy concerns. The synthetic distribution network is an alternative to represent actual networks with critical topological and electrical characteristics.

This literature review has presented a concise but comprehensive review of existing synthetic distribution network development methodologies along with their data requirements, validation approaches, and demonstrative studies. Among the existing methods of developing a synthetic network, the GIS data-based method, an open-source database, provides better accessibility to detailed technical and geographical information for more accurate synthetic networks. Also, data-driven approaches have been demonstrated as an effective method; however, they require various actual data from the network operators. In the case of validation, operational data have been commonly applied as this can ensure their suitability for various applications in power system studies. Furthermore, in most cases, only the network modeling has been demonstrated in the literature while presenting a synthetic distribution network (rather than a demonstration of the application together).

8.2 Summary of Proposed Method for WP 1.4

WP 1.4 aims to create a comprehensive taxonomy and information database for LV and MV/HV network models across Victoria and through to the interfaces with transmission grid supply points. From the literature review, there is currently no whole-of-system modeling framework for infrastructure development downstream to the transmission system in Victoria that incorporates consistent parameters and assumptions. Furthermore, an adequate power system network model is crucial, representing Victoria's whole-distribution network for assessments of DER integration, the electrification of transport, the transition of domestic gas use, and the uptake of new fuels.

In such a scenario, WP 1.4 will utilize a hybrid model combining a grid-based and GIS database approach to develop a Victoria whole-distribution network infrastructure. In the next stage of this project, the following solid tasks will be conducted gradually.

- Acquiring actual transmission and sub-transmission networks up to 66 kV including topology and characteristics from DNSPs.
- Employing a grid-based approach to synthesize the actual transmission and sub-transmission networks removing/anonymizing sensitive information.
- Acquiring a number of actual MV networks from DNSPs and selecting a combination of MV networks (22 kV) based on spatial characteristics (GIS approach) to be connected to the sub-transmission network.
- Selecting a number of LV feeders (0.4 kV) from CSIRO LV taxonomy and DNSP Taxonomy studies and managing a combination of LV feeders based on spatial characteristics (GIS approach) to be connected to the MV networks.
- Validating the developed Victoria whole-distribution network through operational and statistical approaches and expert feedback.

- Demonstrating the base case scenario for power flow/ voltage constraints, optimal power flow, and hosting capacity analysis.

In the end, WP 1.4 will prepare the developed Victoria whole-distribution network models that will support successive studies on the impact assessment of electrification across the entire state and perform the techno-economic analysis. The power system studies would include power flow/ voltage constraints (that might be potentially performed in WPs 1.5, 1.6, 2.7, 2.10), optimal power flow (WPs 2.8, 2.9, 2.10), hosting capacity analysis (WPs 2.7, 2.8, 2.10), power system frequency stability analysis (WP 3.11), and power system reliability analysis (WP 3.12).

Acknowledgment

This work was supported by the iMOVE CRC and Centre for New Energy Technologies, ”C4NET”, through the Enhanced System Planning project (Grant number: PRJ00001911).

References

[1] S. Bouckaert, A. F. Pales, C. McGlade, U. Remme, and B. Wanner, "Net zero by 2050: A roadmap for the global energy sector," 2021.

[2] M. Ali, K. Prakash, C. Macana, M. Raza, A. Bashir, and H. Pota, "Modeling synthetic power distribution network and datasets with industrial validation," *Journal of Industrial Information Integration*, vol. 31, p. 100407, 2023.

[3] C. Mateo, F. Postigo, F. de Cuadra, T. G. S. Roman, T. Elgindy, P. Dueñas, B.-M. Hodge, V. Krishnan, and B. Palmintier, "Building large-scale U.S. synthetic electric distribution system models," *IEEE Transaction on Smart Grid*, vol. 11, pp. 5301–5313, 2020.

[4] R. Yan, Y. Yuan, Z. Wang, G. Geng, and Q. Jiang, "Active distribution system synthesis via unbalanced graph generative adversarial network," *IEEE Transaction on Power Systems*, vol. 38, pp. 4293–4307, 2023.

[5] A. Trpovski, D. Recalde, and T. Hamacher, "Synthetic distribution grid generation using power system planning: Case study of singapore," in *2018 53rd International Universities Power Engineering Conference (UPEC)*, pp. 1–6, IEEE, 2018.

[6] F. Geth, T. Brinsmead, S. West, P. Goldthorpe, G. Cross, and J. Braslavsky, "National low-voltage feeder taxonomy study," 2021.

[7] E. Oliver and C. Perfumo, "Technical report: Load and solar modelling for the NFTS feeders," *Commonwealth Sci. Ind. Res. Organisation (CSIRO), Newcastle, NSW, Australia, Tech. Rep*, 2015.

[8] Y. Li and P. J. Wolfs, "Taxonomic description for western australian distribution medium-voltage and low-voltage feeders," *IET Generation, Transmission and Distribution*, vol. 8, no. 1, pp. 104–113, 2014.

[9] K. P. Schneider, Y. Chen, D. Engle, and D. Chassin, "A taxonomy of north american radial distribution feeders," in *2009 IEEE Power and Energy Society General Meeting*, pp. 1–6, IEEE, 2009.

[10] V. Rigoni, L. F. Ochoa, G. Chicco, A. Navarro-Espinosa, and T. Gozel, “Representative residential LV feeders: A case study for the North West of England,” *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 348–360, 2016.

[11] V. Rigoni and L. N. Ochoa, “Deliverable 3.7 “characterisation of LV networks”,” 2014.

[12] G. Pisano, N. Chowdhury, M. Coppo, N. Natale, G. Petretto, G. G. Soma, R. Turri, and F. Pilo, “Synthetic models of distribution networks based on open data and georeferenced information,” *Energies*, vol. 12, p. 4500, 2019.

[13] E. Schweitzer, A. Scaglione, A. Monti, and G. A. Pagani, “Automated generation algorithm for synthetic medium voltage radial distribution systems,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 7, no. 2, pp. 271–284, 2017.

[14] H. Sadeghian and Z. Wang, “Autosyngrid: A matlab-based toolkit for automatic generation of synthetic power grids,” *International Journal of Electrical Power & Energy Systems*, vol. 118, p. 105757, 2020.

[15] L. S. R. Navarro Espinosa, Alejandro Ochoa and D. Randles, “Reconstruction of low voltage distribution networks:from gis data to power flow models,” in *23rd International Conference on Electricity Distribution*, pp. 1273–1277, 2015.

[16] R. Meyur, A. Vullikanti, S. Swarup, H. S. Mortveit, V. Centeno, A. Phadke, H. V. Poor, and M. V. Marathe, “Ensembles of realistic power distribution networks,” *Proceedings of the National Academy of Sciences*, vol. 119, no. 42, p. e2205772119, 2022.

[17] R. Gupta, F. Sossan, and M. Paolone, “Countrywide PV hosting capacity and energy storage requirements for distribution networks: The case of switzerland,” *Applied Energy*, vol. 281, p. 116010, 2021.

[18] 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, 2017.

[19] S. S. Saha, E. Schweitzer, A. Scaglione, and N. G. Johnson, "A framework for generating synthetic distribution feeders using openstreetmap," in *2019 North American Power Symposium (NAPS)*, pp. 1–6, IEEE, 2019.

[20] CitiPower and Powercor, "Future grid for distributed energy," *[Online]. Available: https://arena.gov.au/knowledge-bank/future-grid-for-distributed-energy/*, June. 2020.

[21] C. K. Gan, D. Pudjianto, P. Djapic, and G. Strbac, "Strategic assessment of alternative design options for multivoltage-level distribution networks," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1261–1269, 2014.

[22] G. Viganò, M. Rossi, C. Michelangeli, and D. Moneta, "Creation of the italian distribution system scenario by using synthetic artificial networks," in *2020 AEIT International Annual Conference (AEIT)*, pp. 1–6, IEEE, 2020.

[23] G. Viganò, M. Rossi, and D. Moneta, "Generation of artificial distribution networks for the evaluation of advanced control solutions," in *2017 AEIT International Annual Conference*, pp. 1–6, 2017.

[24] G. Viganò, M. Rossi, C. Michelangeli, and D. Moneta, "Study of the evolution of the italian mv distribution system by means of synthetic networks," in *2021 AEIT International Annual Conference (AEIT)*, pp. 1–6, IEEE, 2021.

[25] R. Espejo, S. Lumbreras, and A. Ramos, "A complex-network approach to the generation of synthetic power transmission networks," *IEEE Systems Journal*, vol. 13, pp. 3050–3058, 2019.

[26] S. Abeysinghe, J. Wu, M. Sooriyabandara, M. Abeysekera, T. Xu, and C. Wang, "Topological properties of medium voltage electricity distribution networks," *Applied Energy*, vol. 210, pp. 1101–1112, 2018.

[27] B. Palmintier, T. Elgindy, C. Mateo, F. Postigo, T. Gómez, F. de Cuadra, and P. D. Martinez, "Experiences developing large-scale synthetic us-style distribution test systems," *Electric Power Systems Research*, vol. 190, p. 106665, 2021.

[28] F. Postigoa, C. Mateoa, T. Gómeza, F. de Cuadraa, P. Dueñasb, T. Elgindyc, B.-M. Hodgec, B. Palmintierc, and V. Krishnanc, "Phase-selection algorithms to minimize cost and imbalance in U.S. synthetic distribution systems," *International Journal of Electrical Power & Energy Systems*, vol. 120, p. 106042, 2020.

[29] C. Mateoa, G. Prettico, T. Gómez, R. Cossent, P. F. Flavia Gangale, and G. Fulli, "European representative electricity distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 99, pp. 273–280, 2018.

[30] C. M. Domingo, T. G. S. Roman, Álvaro Sánchez-Miralles, J. P. P. González, and A. C. Mart´ınez, "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, 2011.

[31] R. Krishnan *et al.*, "Validation of synthetic U.S. electric power distribution system data sets," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4477–4489, 2020.

[32] AusNet Services, "Ausnet's transmission system map," *[Online]. Available: https://www.aer.gov.au/system/files/AusNet2015.*