Automated Generation of Urban Medium-voltage Grids using OpenStreetMap Data

Tobias Gebhard
Institute for the Protection of Terrestrial Infrastructures
German Aerospace Center (DLR)
Darmstadt, Germany
tobias.gebhard@dlr.de

Andrea Tundis
Institute for the Protection of Terrestrial Infrastructures
German Aerospace Center (DLR)
Darmstadt, Germany
andrea.tundis@dlr.de

Florian Steinke
Energy Information Networks & Systems
Technische Universität Darmstadt
Darmstadt, Germany
florian.steinke@eins.tu-darmstadt.de

***Abstract*—Realistic geo-referenced electrical distribution grid (DG) models are of great importance for power system analysis and resilience studies. However, DG data are usually not publicly available. In this study, we develop a new process for the automated generation of medium-voltage (MV) grid topologies, specifically for urban areas, based on openly available data and open-source software. OpenStreetMap (OSM) data on power infrastructure, street layouts, and land use, are used as the only input source. In contrast to previous works on DG reconstruction, we use available OSM data on substation locations. Different existing methods are combined in a new, hybrid approach by considering the incompleteness of OSM data, taking the street network into account, and applying the Capacitated vehicle routing problem (CVRP) to find cost-optimal routes for power lines. Our method is tested with a German city as a case study. Furthermore, we verify the result using land use data and evaluate the quality of power-related OSM data. The results demonstrate that our approach can yield realistic geo-referenced MV grid topologies, even with incomplete OSM power data.**

***Index Terms*—Distribution Grid, Medium-Voltage, Automated Grid Model Generation, Open Data, OpenStreetMap**

I. Introduction

ELECTRICITY distribution grids (DGs) play a crucial role in supplying consumers with power reliably and efficiently. As the integration of renewable energy sources, such as wind and solar, along with emerging technologies like electric vehicles and heat pumps, requires adaption and transformation for future DGs, the need for realistic DG models becomes increasingly paramount. These models are indispensable for various power system analyzes in grid planning and optimization, including simulations, load flow calculations, and resilience assessments.

However, access to data on power grids, especially at the distribution level, remains a significant challenge for researchers. Unlike the transmission system, where system operators in Europe are forced to publish grid data [1], DG data are generally not publicly available due to security concerns and proprietary constraints imposed by distribution system operators (DSOs). Even when data are accessible, it is typically in static (e.g. PDF) or proprietary formats, and often lacking essential geo-referenced information [2]. This requires extensive amounts of manual labor for creating simulation models. The decentralized nature of DSOs in countries like Germany further complicates the creation of comprehensive network models. There are certain reference grids available often being used in research, which are however not representative for every region and often do not include geo-information [3].

For these reasons, the automated reconstruction of DGs for any region, based on open data and widely accepted data formats, seems promising. OpenStreetMap (OSM) can serve as a free, open, and powerful GIS data source for energy models [4], [5]. OSM provides extensive data on power infrastructure in a geo-referenced form [2]. For transmission grids, tools exist that can deliver a simulatable power system model [1], [2]. While the OSM data on extra high voltage levels are comprehensively covered, data quality and coverage decreases for the high, medium, and low voltage levels [2]. Due to the fact that power lines are mostly underground at lower voltage levels (and especially in urban areas), OSM data provide primarily overhead lines and substation locations. Thus, the reconstruction of DGs, receiving attention more recently, requires different approaches.

In this paper, a novel methodology for the automated generation of geo-referenced medium-voltage (MV) networks is presented. Leveraging OSM data on power infrastructure, streets, and land use, we combine existing techniques for low-voltage (LV) and MV grid reconstruction in a hybrid approach. Unlike previous methods, our approach utilizes OSM data on secondary substations, enhancing the accuracy and realism of the generated grid topologies. By using the street layout from OSM and assuming a certain grid topology for urban MV grids, our method is suitable for urban areas. The analogy of the Capacitated vehicle routing problem (CVRP) is used to determine optimal routes for power lines. The developed workflow for the automated process is described and evaluated using the German city of Darmstadt as an example. Moreover, we verify and evaluate the quality of power-related OSM data with land use data.

The proposed method extends previous work on MV grid reconstruction by focusing on urban areas and explicitly considering the street network. The resulting geo-referenced

This work has been conducted in the context of the "RESIKOAST" project funded from the German Aerospace Center (DLR). It has been performed in the context of the LOEWE center emergenCITY.

power grid model enables not only load flow analysis, but also more comprehensive analyzes, like disaster impact assessment and resilience studies. The use of exclusively open OSM data enables the universal application of the approach to every region in the world where OSM data are available, without relying on local data formats or restricted data access. Furthermore, our methodology provides the basis for a complete DG generation based on open data by extending the approach with LV grids in a top-down approach. Since our approach uses the known substation locations, this idea seems more promising than bottom-up approaches described in other papers.

The paper is structured as follows. Section II provides an overview of related work. Section III explains the methodology and Section IV evaluates the proposed method. Section V concludes the paper.

II. Related Work

The automated creation of power grid models representing a real world area has gained attention in research. For the different voltage levels, different approaches have been proposed.

For transmission grids, open-source tools for creating a grid model have been developed by using OSM data, namely SciGRID, GridKit, osmTGmod, and GridTool [1], [2]. These works make use of power-related OSM data, which are comprehensively available for the high and extra high voltage level, as this infrastructure is mostly overground [2]. In contrast, OSM data on DGs are often not sufficient, requiring different approaches.

Previous contributions on MV grid reconstruction have focused on using land use data [6]–[8]. The DINGO tool [6] exploits OSM land use data from [5] to create "load areas" by distinguishing between residential, retail, industrial, and agricultural. For all approaches, a clustering of LV customers is performed to obtain virtual MV-LV stations. The Voronoi decomposition is then used to assign each supply point to a given HV-MV station. A ring topology is assumed where each supply node is connected to exactly one HV-MV station. While [8] adapts the *Traveling Salesman Problem* (TSP) algorithm to obtain the cost-optimal MV ring, [6] and [7] use the more general *Capacitated Vehicle Routing Problem* (CVRP) to find optimal lines with multiple ring connections. Power lines are considered as straight point-to-point connections, neglecting local characteristics, like street structures and natural boundaries. All related works [6]–[8] focused on rural areas. Research on the generation of synthetic MV grids within urban areas, taking into account urban characteristics, is lacking.

The generation of LV grid models has also been in focus of research, especially for residential areas. Here, usually the street network from OSM is used as a base graph for possible cable routes by assuming a radial grid topology. For the load modeling, the building outlines from OSM are used [8]–[11]. The location of the substation is either assumed as given, for example obtained through on-site inspection [10] or is extracted by the algorithm using clustering [8], [9]. A stochastic approach is used in [11] to generate an ensemble of possible LV topologies in contrast to one deterministic result.

To the best of our knowledge, none of the works for MV and LV grids has considered OSM data of MV-LV substations yet. By combining the approaches from the MV and LV level in a hybrid way and using the known substation locations, we aim to improve the MV grid generation.

III. Methodology

Fig. 1 visualizes the methodology for generating a MV grid for a specified region. The locations of substations (primary and secondary) contained in OSM are the basis for the grid generation. Their respective distances are calculated based on the street network. The CVRP algorithm is then applied to obtain the optimal (minimal length) power network by considering topology constraints.

Since the substations from OSM can be incomplete, the resulting number is verified using land use data. Consequently, additional information can be included by specifying custom loads in an iterative process.

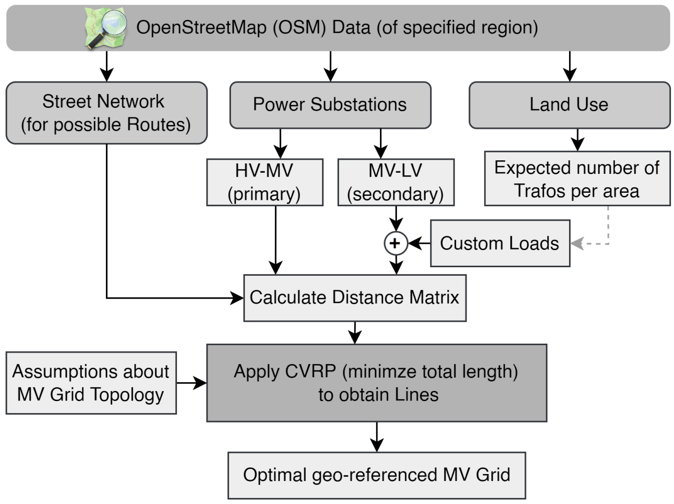

Fig. 1. Overview of the approach

The main steps for the automated process are summarized in Algorithm 1. We implement the approach in Python by building on open-source packages like OSMnx, NetworkX, and VRPy. The following subsections elaborate on the concept and different aspects in more detail.

A. Data Acquisition

All input data used are obtained from the OpenStreetMap project. We use OSM data of power infrastructure, the street network, and land use. To obtain the required data from OSM in a usable and pre-filtered form, the Overpass API [12] is used in combination with the Python package *overpy*.

1) Street Data: Representing streets and paths has been the core motivation for OSM. The quality and coverage of streets and walkable paths can be considered sufficient for most parts of the world.

In urban areas, power lines are mostly underground and follow the layout of streets [13]. Therefore, we assume the street network as the basis for potential routes of power lines. The

Algorithm 1 Main steps for automated MV grid generation

1: Specify the region of interest (e.g. a city or polygon)

2: Generate street graph for selected area using OSMnx

3: Query Overpass in selected area for all nodes and ways (representing areas) containing tag *power* with value *substation* or *transformer*

4: Perform tag filtering (e.g. voltage and frequency)

5: [Optional] Remove unwanted elements with specified list

6: [Optional] Add custom MV loads with specified list

7: Connect all power nodes and areas (corresponding centroid) to street network via closest node

8: Calculate distances between power nodes on street graph

9: Assign transformers to MV connection (between 2 substations) using the Voronoi diagram of line segments

10: Solve CVRP for each MV connection using VRPy to obtain grid topology and line courses

street network forms a geometric graph with traffic junctions as nodes and streets as edges. With OSMnx, a directed graph structure can be extracted from OSM and is directly available in Python for further processing.

2) Power-related Data: While OSM data quality and coverage for overhead lines at higher voltage levels (>100 kV) are very high [2], the situation is very different for the MV level. MV lines are underground in European countries to a large degree, e.g. about 75-95% in Germany [14]. For this reason, underground power cables are usually not represented in OSM. Comparing the worldwide number of elements with the *power* tag reveals a share of just 3.9% for the value *cable* compared to *line* and *minor_line*[1].

In contrast, substation locations are mapped more comprehensively in OSM, and are therefore used for our approach. HV-MV substations are typically mapped with the tag *power=substation* as an area (a closed way in OSM). Transformers can be mapped as a node with *power=transformer*. Due to the good coverage of primary substations in OSM and their manageable number for one city, they are assumed as given.

For MV-LV substations, both *substation* and *transformer* is used, either as node or closed way (e.g. a building). To distinguish between HV-MV and MV-LV substations, the tags *substation=minor_distribution* or *transformer=distribution*, as well as tags for the nominal voltage could be used. As this is not done consistently, we assume that all substation and transformer elements are secondary unless the voltage is specified and above 50 000. Furthermore, all elements that contain a frequency value other than 50 are dropped to avoid objects from the railway power grid.

[1]https://taginfo.openstreetmap.org/keys/power#values

3) Land Use Data: Data on land use are included as a counter approach to verify the number of secondary substations within certain areas or districts of a city. In the OSM database, land use data are contained in a high quality, especially in urban areas [4].

B. Custom Loads

Since the MV-LV substations from OSM can be incomplete or incorrect, our approach can include additional information in a flexible way. In theory, the public OSM database could be extended by everyone with new information. However, this is not always possible or practical, e.g. when partially sensitive information is available. Moreover, OSM data are only intended for accurate, verified information, but sometimes including estimated information can be useful. For a specific use-case, the additional information could be added manually after the model generation, but future OSM updates would require to repeat these steps. Therefore, we incorporate a list for locations of custom loads that are added during the automated process. In addition, elements that should not be included for various reasons, can be specified in another list via their OSM id.

C. Grid Topology

In contrast to related works that assume ring topologies, we assume that all MV lines connect two distinct substations. This configuration permits each point to be served by two separate sides and from different primary substations. A pair of substations is referred to as "connection", consisting of one or more parallel lines. All secondary substations are connected to the grid by exactly one line, not consisting of junctions.

The electrical grid can be considered as a graph, with nodes corresponding to supply points, whereas edges represent power lines. We consider all nodes representing secondary (MV-LV) substations. They are connected to the nearest node of the street network. Since we assume that power lines follow streets, the power grid can be considered as a sub-graph of the street network. On the street network, the distance of lines can be calculated, being more accurate than the euclidean distance.

D. Line Assignment

To partition the MV-LV transformers, every transformer is assigned heuristically to a connection between two primary substations. To find a reasonable assignment, often the Voronoi partition is used, which assigns any point in the plane to the nearest substation [15]. However, given our assumed topology, we want to find the closest *connection*, i.e. the closest straight line created by two substations. For this, the Voronoi diagram of line segments in the plane is used [16].

From all combinations of substation pairs, some connections might be unreasonable. Given prior knowledge, a list of substation pairs can be provided. As a heuristic approach, the Delaunay-Triangulation can be used to create sensible connections. Outer triangles with sharp angles (<45°) should be excluded, as they form unrealistic geometries.

E. Load Modeling

The load within the MV grid is mostly represented by the secondary substations, which transport the energy to the LV level. Industrial facilities might be connected directly on the MV level, which has to be considered. We do not consider distributed generation, as its effect is comparably low in urban areas.

In contrast to the location, the technical details of transformers, such as capacity, are much harder to determine and thus rarely included in OSM data. Since the exact parameters have little influence on the resulting topology, a universal transformer type can be chosen.

F. Cost Optimization with CVRP

To generate realistic MV line topologies connecting all substations, the CVRP algorithm is applied. The CVRP can find cost-optimal routes satisfying certain constraints. Thereby, the assumption is made that the DSO designed the grid in a cost-optimal way. The cost is assumed to be proportional to the grid size, i. e. the sum of all cable lengths. The CVRP algorithm is applied for every connection separately.

The CVRP is an optimization problem and can be seen as a general case of the Traveling Salesman Problem (TSP). Originally, the CVRP is designed to identify routes from a depot to another depot for a fleet of vehicles delivering goods to all specified locations with the objective of minimizing costs. We use the analogy of routes for power lines, depots for primary substations, delivery locations for secondary substations, and vehicle capacities for cable capacity limits. Following the described grid topology, we assume that routes for power lines start at a primary substation (depot), connect a number of secondary substations while respecting the limit of accumulated power capacities, and end at another primary substation.

For solving the CVRP, the costs between all depots and delivery points are determined as the shortest path distance on the street graph. This results in the distance matrix (representing a complete graph) and is passed to the CVRP solver.

The CVRP approach does not consider further constraints, like voltage limits. For such considerations, the resulting grid model needs to be evaluated with power flow simulations.

G. Verification with Land Use Data

Since power data in OSM need to be assumed incomplete, especially at the lower voltage levels [2], a verification for the MV-LV substations is advisable. To verify the number of transformers, a comparison is conducted by using land use data to estimate the power demand per area. If the number of expected transformers is larger than the number of mapped ones, additional transformers can be added at reasonable locations to the custom loads list.

1) Land Use Data from OSM: A simple estimation is to calculate the power demand per area by assuming certain values for different types of land use. The land use types and areas can be obtained from OSM or other sources. After calculating the sizes of the areas, the total demand for a certain area is divided by a representative transformer capacity to obtain the required number of transformers.

2) Statistical Energy Report Data: If official statistics about the historic energy consumption of a city or district under investigation are available, the installed capacity can be estimated from this. However, the mapping of administrative areas to power nodes can be challenging. Furthermore, such data usually refers to the average demand, which has to be transferred to an estimate for the maximum demand or capacity, as this is the relevant factor for the initial grid design.

IV. Evaluation

A. Case Study for a German City

The proposed methodology is applied to Darmstadt, a medium-sized city in Hesse (Germany) with a population of about 160 000. 6 primary substations at 110 kV are known. Given the arrangement of the substations, 7 feasible connections for the line assignment are obtained with the

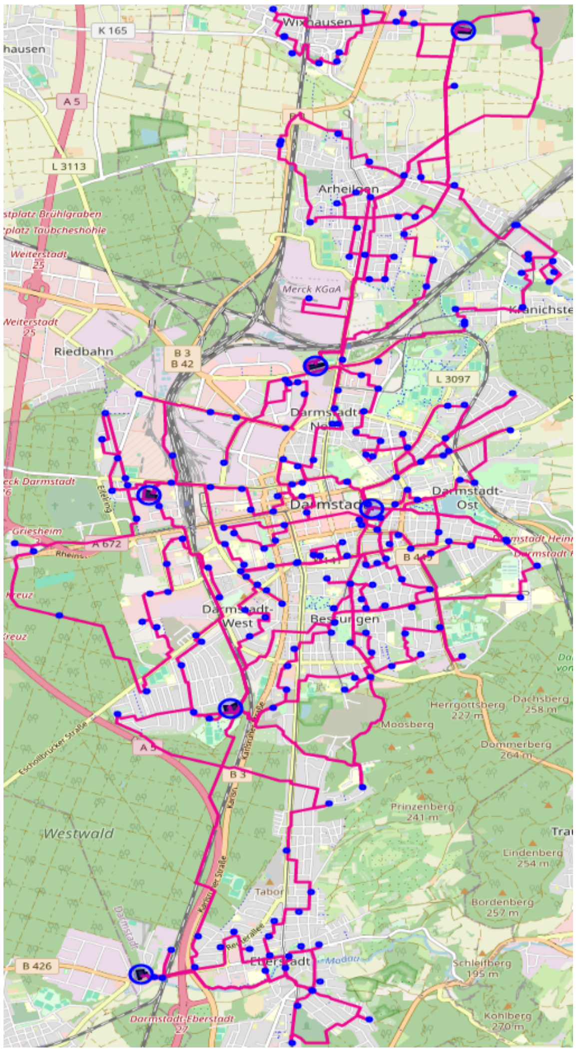

Fig. 2. Synthetic MV grid of Darmstadt (© OpenStreetMap Contributors)

Delaunay-Triangulation. A transformer capacity of 400 kVA and cable capacity of 10 MVA is chosen.

Fig. 2 shows the resulting geo-referenced MV grid, forming a non-planar graph. The number of lines per connection varies between 1 and 5. Parallel lines using the same street path appear near substations.

The coverage of secondary substations obtained from OSM appears to not be distributed equally. For example, industrial areas, having a higher consumption per area, are underrepresented.

B. OSM Data Quality Assessment

To assess the coverage of OSM substation data, the number of secondary substations is compared to the power demand estimation based on land use data. For this, we extract administrative boundaries from OSM and calculate the area, respectively for *residential*, *commercial*, *retail*, and *industrial*. We use typical power demand per area values from [17].

Table I shows the results of the comparison. In total, OSM seems to cover more than 70% of required transformers. However, the number of missing transformers varies strongly per district. In industrial areas, transformers are rarely included in OSM, likely because they are not publicly visible or the power connection is handled individually at the MV level. In contrast, there are districts where the number of OSM transformers is higher than the estimated number, indicating that the land use approach cannot sufficiently consider the height of buildings.

TABLE I
COMPARISON FOR NUMBER OF SUBSTATIONS

District	# Trafos OSM	# Trafos Land Use	Difference
Darmstadt-Mitte	22	9	+13
Darmstadt-Ost	20	15	+5
Darmstadt-Nord	45	116	-71
Darmstadt-West	32	63	-31
Arheilgen	25	24	+1
Bessungen	30	23	+7
Eberstadt	27	39	-12
Kranichstein	11	7	+4
Wixhausen	13	9	+4
Total	225	305	-80

V. CONCLUSION

This paper introduces a new approach for the automated generation of MV grid models solely based on open data and open-source software. By considering the specific characteristics of urban areas, our work extends current research in the field of power grid reconstruction.

The results show that with using OSM data on power infrastructure, a highly realistic MV grid can be constructed. The quality of the result depends mostly on the coverage of OSM-tagged MV-LV substations. Even though the coverage is often not complete, already a low share of true substation locations can create realistic topologies. Moreover, the linear increasing trend of substations tagged worldwide[2], indicating a steady increase in the OSM data coverage, leads to more realistic results in the future.

In future work, the presented approach could be combined with other approaches for areas where OSM data are rare. In addition, existing medium voltage lines in OSM (underground or overhead) could be taken into account for the CVRP.

[2]https://taginfo.openstreetmap.org/tags/power=substation#chronology

ACKNOWLEDGMENT

The authors would like to thank Rubén Chacón for his technical contribution to the implementation.

REFERENCES

[1] R. Gaugl, S. Wogrin, U. Bachhiesl, and L. Frauenlob, “GridTool: An open-source tool to convert electricity grid data,” *SoftwareX*, vol. 21, p. 101314, Feb. 2023.

[2] W. Medjroubi, U. P. Müller, M. Scharf, C. Matke, and D. Kleinhans, “Open data in power grid modelling: New approaches towards transparent grid models,” *Energy Reports*, vol. 3, pp. 14–21, Nov. 2017.

[3] 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, Jan. 2020.

[4] A. Alhamwi, W. Medjroubi, T. Vogt, and C. Agert, “OpenStreetMap data in modelling the urban energy infrastructure: a first assessment and analysis,” *Energy Procedia*, vol. 142, pp. 1968–1976, Dec. 2017.

[5] L. Hülk et al., “Allocation of annual electricity consumption and power generation capacities across multiple voltage levels in a high spatial resolution,” *Int. Journal of Sustainable Energy Planning and Management*, vol. 13, pp. 79–92, Sep. 2017.

[6] J. Amme et al., “The eGo grid model: An open-source and open-data based synthetic medium-voltage grid model for distribution power supply systems,” *Journal of Physics: Conference Series*, vol. 977, no. 1, p. 012007, 2018.

[7] J. Tran et al., “Modelling of synthetic power distribution systems in consideration of the local electricity supply task,” in *Cired*. AIM, Jun. 2019, Conf. Proc.

[8] 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 Trans. on Power Systems*, vol. 32, no. 3, p. 2346–2353, 2017.

[9] B. Abhilash, C. Syranidou, J. Linssen, and D. Stolten, “Geo-referenced synthetic low-voltage distribution networks: A data-driven approach,” in *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, Oct. 2021.

[10] H. K. Cakmak, L. Janecke, M. Weber, and V. Hagenmeyer, “An optimization-based approach for automated generation of residential low-voltage grid models using open data and open source software,” in *2022 IEEE PES Innovative Smart Grid Technologies Conf. Europe (ISGT-Europe)*, Oct. 2022.

[11] D. Tomaselli, P. Stursberg, M. Metzger, and F. Steinke, “Representing topology uncertainty for distribution grid expansion planning,” in *27th Int. Conf. on Electricity Distribution (CIRED 2023)*. IET Digital Library, Jun. 2023, Text, pp. 721–725.

[12] “Overpass api.” [Online]. Available: https://wiki.openstreetmap.org/wiki/Overpass_API/Overpass_QL

[13] C. Mateo Domingo, T. Gomez San Roman, A. Sanchez-Miralles, J. P. Peco Gonzalez, and A. Candela Martinez, “A reference network model for large-scale distribution planning with automatic street map generation,” *IEEE Trans. on Power Systems*, vol. 26, no. 1, pp. 190–197, Feb. 2011.

[14] J. Buechner et al., “Moderne Verteilernetze für Deutschland (Verteilernetzstudie),” Tech. Rep., 2014. [Online]. Available: https://www.bmwk.de/Redaktion/DE/Publikationen/Studien/verteilernetzstudie.html

[15] A. Navarro and H. Rudnick, “Large-scale distribution planning—part ii: Macro-optimization with voronoi’s diagram and tabu search,” *IEEE Trans. on Power Systems*, vol. 24, no. 2, pp. 752–758, May 2009.

[16] S. W. Bae, “An almost optimal algorithm for voronoi diagrams of non-disjoint line segments,” *Computational Geometry*, vol. 52, pp. 34–43, Feb. 2016.

[17] A. Sillaber, *Leitfaden zur Verteilnetzplanung und Systemgestaltung*, 1st ed. Wiesbaden: Springer Vieweg Wiesbaden, 2016.