Large-Scale Inference of Geo-Referenced Power Distribution Grids Using Open Data

Working Paper

Author(s):
Oneto, Alfredo Ernesto; Gjorgiev, Blazhe; Tettamanti, Filippo; Sansavini, Giovanni

Publication date:
2023-12-02

Permanent link:
https://doi.org/10.3929/ethz-b-000646111

Originally published in:
TechRxiv, https://doi.org/10.36227/techrxiv.24607662

Large-Scale Inference of Geo-Referenced Power Distribution Grids Using Open Data

Alfredo Ernesto Oneto [1], Blazhe Gjorgiev [1], Filippo Tettamanti [1], and Giovanni Sansavini [2]

[1]Affiliation not available
[2]ETH Zurich

December 7, 2023

Abstract

Power distribution grids host an increasing amount of distributed renewable generators, electric vehicles, and heat pumps worldwide. Distribution grids, however, were not designed with the goal of incorporating large shares of these technologies. These soaring challenges demand accurate and realistic grid models to assess the need for operation strategies and reinforcements that ensure reliable and economic management. Nevertheless, real models are often unavailable due to privacy and security concerns or a lack of digitized data from distribution system operators. To address this issue, we present a framework for large-scale inference of geo-referenced low- and medium-voltage grid models using publicly accessible information on power demand and transport infrastructure. First, we develop a clustering algorithm, which detects load areas served by distribution grids. Then, we obtain the graphical grid layout, i.e., a graph with the street and pathway geometries and the load point connections inside the load area. Next, we introduce a selection method for line types that assigns cost-effective conductors to grid lines while ensuring operational constraints. We demonstrate the effectiveness of our approach by inferring all the low- and medium-voltage infrastructure in Switzerland. Remarkably, the inferred grids present overall power requirements and line lengths statistically aligned with reference grids.

Large-Scale Inference of Geo-Referenced Power Distribution Grids Using Open Data

Alfredo Oneto, Blazhe Gjorgiev, Filippo Tettamanti, Giovanni Sansavini

***Abstract*—Power distribution grids host an increasing amount of distributed renewable generators, electric vehicles, and heat pumps worldwide. Distribution grids, however, were not designed with the goal of incorporating large shares of these technologies. These soaring challenges demand accurate and realistic grid models to assess the need for operation strategies and reinforcements that ensure reliable and economic management. Nevertheless, real models are often unavailable due to privacy and security concerns or a lack of digitized data from distribution system operators. To address this issue, we present a framework for large-scale inference of geo-referenced low- and medium-voltage grid models using publicly accessible information on power demand and transport infrastructure. First, we develop a clustering algorithm, which detects load areas served by distribution grids. Then, we obtain the graphical grid layout, i.e., a graph with the street and pathway geometries and the load point connections inside the load area. Next, we introduce a selection method for line types that assigns cost-effective conductors to grid lines while ensuring operational constraints. We demonstrate the effectiveness of our approach by inferring all the low- and medium-voltage infrastructure in Switzerland. Remarkably, the inferred grids present overall power requirements and line lengths statistically aligned with reference grids.**

***Index Terms*—Geo-referenced grids, power distribution, clustering methods, data models, load modeling.**

I. Introduction

THE increasing utilization of distributed renewable generation, electric vehicles, and heat pumps poses substantial challenges to power distribution grids (PDGs) [1], and their adoption is expected to rise due to the climate strategies that aim to reduce carbon emissions worldwide [2]. Therefore, concerns such as voltage stability [3], reverse power flows [4], and overloading of transformers [5] are becoming increasingly pressing. Tackling these issues requires implementing and evaluating PDG models for computational assessments [6]. However, publicly available information about PDGs is minimal because of privacy and security issues [7], added to the lack of digitized grid data [8] of distribution system operators (DSOs). Consequently, test grids have been published to tackle the need for data [9], [10]. Still, PDGs are very diverse due to geographical and power requirement conditions; thus, test models can provide only limited insights.

Motivated by the inaccessibility of PDG data, several works propose PDG-inference frameworks. However, many contributions focus on low-voltage (LV) or medium-voltage (MV) grids, restricting their application to sub-parts of power distribution infrastructure. Additionally, many works do not leverage geographic information systems (GIS), which is crucial to perform zonal studies on PDGs [11], [12]. Even when GIS tools are employed, many methods still require expert tuning from DSOs and private sources [5], [13]. Additionally, some of these frameworks can only be adopted for small-scale areas, as they entail significant computation effort [14], [15]. Others offer a light computational burden but generate grids with insufficient detail about the electrical components adequate to perform power flow analysis [6], [16]. In summary, no existing work has covered the four characteristics necessary for a precise large-scale inference of PDGs using open data, namely, combining both LV and MV levels, geo-referencing all PDG components, being computationally efficient, and producing statistically accurate results.

In general, frameworks for inferring PDGs share three steps: 1) the identification of load areas, 2) the derivation of graphical grid layouts, i.e., the graphs assigned to each load area with the load point connections and the transport geometry[1], and 3) the selection of the conductors. Step 1) defines the geographic clusters where electric power is consumed [4]; step 2) builds geo-referenced graphs with location-specific constraints over the load areas [1]; step 3) selects the conductors to obtain the PDGs [17]. The scientific literature offers various techniques for performing these steps.

To detect load areas, [1] adopts the `k-means` algorithm [18] and [19] uses the `DBSCAN` algorithm [20]. [21] employs Voronoi partitioning, updating iteratively the number and the location of transformers, which are used as generating points of the partitions. In [22], the authors use binary image segmentation to detect contiguous demands and group them to reach a total load around a target value. Still, these works do not meet all three conditions that are imposed by technical and economic constraints of PDG's design: 1) the load areas must have a total power demand not higher than the transformer's capacity, 2) the load density in a given load area must be sufficient to avoid voltage drops higher than allowed, and 3) the number of load areas should be minimized when assigning each load to a load area while respecting 1) and 2).

The broad consensus on graphical grid layouts is that streets and public ways should be used as possible power line routes in general [14], [23]. Such information is readily available from the worldwide database OpenStreetMap (OSM) [24], which is accessible via computer tools [25], [26]. Leveraging the transport information allows us to consider local geospatial conditions and to avoid restricted-access areas [15]–[17].

The conversion of the graphical grid into a PDG must comply with topological and operational standards. As both

[1]The transport considers drivable public and service streets, and pedestrian and bike pathways.

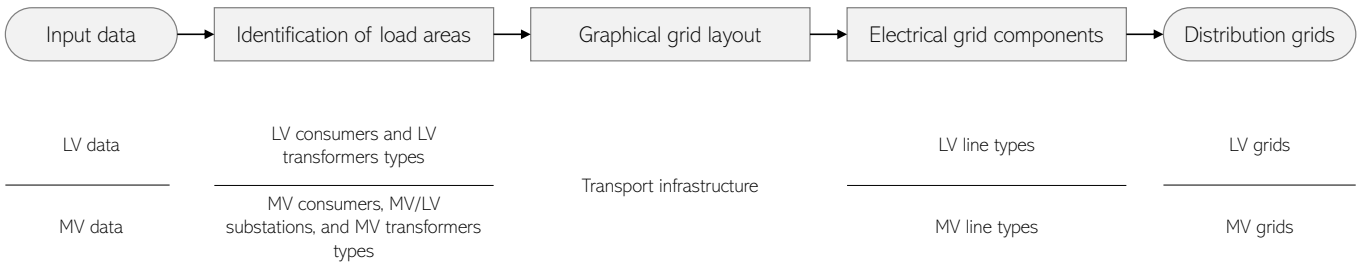

Fig. 1. Steps for inferring geo-referenced distribution grids.

LV and MV grids mainly operate with a radial topology [27], most frameworks generate topologies for PDGs based on this configuration. Moreover, they solve the conductor selection problem for each line in these grids while ensuring feasible operation conditions. For example, [17] builds tree topologies and derives cost curves for each conductor section with the technique introduced in [28]; [5], [29] use a branch-exchange algorithm. In [22], the authors apply a steepest-descent-based approach to generate radial MV grids while selecting line conductor types. Furthermore, to solve the same problem, [30], [31] adopt evolutionary algorithms, and [32] uses genetic algorithms and minimum spanning trees. However, the combinatorial nature of the line-type selection problem makes it hard to solve, and existing methods require high processing power.

Our large-scale inference of PGDs based on public data closes the identified research gaps with two main contributions. First, we propose a clustering algorithm that can detect load areas that meet the technical and economic requirements of PDGs' design. This allows us to treat the PDG inference in parallel for each load area and reduce the need to recombine or split the grids at a later step. Second, we formulate a line-type selection algorithm that leverages the topological properties of the grids to gain computation speed. We demonstrate the computational tractability of our framework by inferring both the MV and LV grid infrastructure of Switzerland. The statistical features of the obtained PDGs are aligned with those of real and synthetic reference grids, confirming the framework's accuracy.

The remainder of this article is organized as follows. Section II introduces the framework for large-scale inference of MV and LV grids. Section III details the Swiss case study. Section IV presents the grid inference and benchmark results. Section V discusses the strengths and limitations of the proposed scheme. Finally, Section VI gives concluding remarks.

II. Inference of geo-referenced distribution grids

The framework for inferring geo-referenced PDGs is summarized in Fig. 1. First, we identify load areas using the spatial distribution of loads and the transformer types. We compute them using a tailored clustering algorithm (Section II-A). Second, for each load area, we determine a graphical layout of the PDG using the transport infrastructure network (Section II-B). Third, we develop a novel technique for efficient line type selection that derives cost-effective configurations and satisfies the operational constraints (Section II-C). The framework blocks are further specialized for MV and LV grid voltage levels (Section II-D).

A. Identification of load areas

An essential problem in PDG inference is the placement of transformers that supply the spatially distributed loads within load areas. The solution aims to minimize the installation of transformers for the cost-effectiveness of DSOs investments [21]. To address this challenge, we develop a geographic clustering approach to minimize the number of load areas or, equivalently, maximize their load while adhering to technical constraints. These constraints are: 1) not exceeding a threshold of total power demand and 2) supplying proximate customers via the same transformer without incurring inadmissible voltage drops.

Our method uses two control parameters for aggregating neighboring load areas, namely, the load area's maximum power demand (τ) and the distance limit (δ). We set τ based on transformer power ratings and determine δ using standard distances between adjacent transformers. In what follows, we use 'load point' to denote a single load and 'zone' for the region under analysis, which contains the load points and areas. The procedure consists of three steps:

- *Step 1:* We initialize the load areas assignment using weighted `k-means` on the load points within a zone. To do this, we fix k as the upper approximation of the zone's total power load divided by τ, utilizing the power demands of the load points as weights.
- *Step 2:* We use weighted `k-means` to partition load areas whose total load exceeds τ until their demand is no greater than τ.
- *Step 3:* We aggregate neighboring load areas where their load centroids are at a distance no greater than δ, and the sum of their loads is less than or equal to τ.

Figure 2 summarizes the tasks performed in *Step 2*. The results of *Step 1* provide the labels $\boldsymbol{l}$, which indicate the load area where each load point is included. The load areas' indices are contained in a set C, and the subset $S \subset C$ has the ones with a total load above τ. Then, for each s in S, we execute the procedure on the right-hand side of the flowchart. The number of splits b required for the s-th load area is $\lceil \sum_{i|\boldsymbol{l}(i)=s} d_i/\tau \rceil$, where d_i is the power demand of the i-th load point and $\boldsymbol{l}(i)$ is its load area label. If b is smaller than the number of load points in the s-th load area, we run weighted `k-means` with b

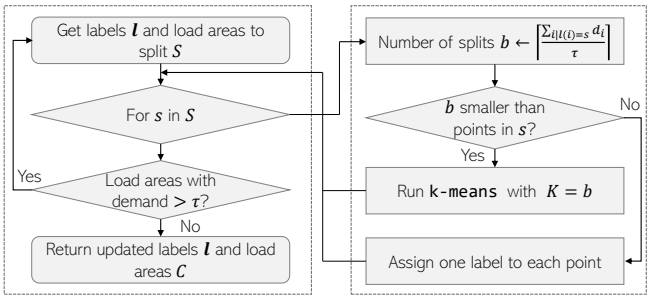

Fig. 2. Diagram of the load areas splitting algorithm.

partitions. Conversely, if b exceeds the number of load points, we create new load area indices and assign different labels to each one of the load points. After finishing the loop over S, the algorithm checks if there are load areas with power demand above τ after the updates and re-runs the splitting if it is required. Otherwise, we get the updated load area indices C and labels $\boldsymbol{l}$.

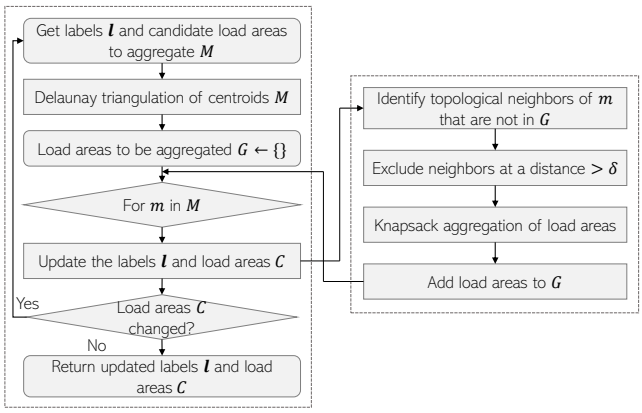

Fig. 3. Diagram of the load areas aggregation algorithm.

Fig. 3 shows the workflow of *Step 3*. The algorithm starts by getting the labels $\boldsymbol{l}$ of the load point and the load areas $M \subset C$ with a total load below τ. The Delaunay triangulation of the centroids of the loads belonging to each load area in M identifies the neighboring load areas to be aggregated. The set G stores the lists of aggregated load areas.

For each m in M, the algorithm identifies neighbors of the load centroid of m that are not already in a list within G and excludes those beyond a geographic distance greater than δ. Load area aggregation is performed using a knapsack model [33] to maximize the aggregated total load, ensuring it remains below or equal to τ:

$$\max_{\boldsymbol{w}} \quad \sum_{n \in \mathcal{H}} \alpha_n w_n \tag{1a}$$

$$\text{s.t.} \quad \sum_{n \in \mathcal{H}} \alpha_n w_n \leq \tau - \alpha_m \tag{1b}$$

$$w_n \in \{0, 1\} \quad \forall n \in \mathcal{H}, \tag{1c}$$

where $\mathcal{H}$ contains the index of load areas neighbors to m; w_n is a binary variable, which takes the value of 1 if the n-th load area is selected and 0, otherwise; and α_m is the total load of area m. The solution of (1a)-(1c) identifies the load areas to be merged with the m-th load area, and these are included in G. Once this identification is done for all m in M, the load areas are aggregated as indicated in G.

Finally, the load point labels $\boldsymbol{l}$ and the load area indices C are updated. If C has changed, the algorithm proceeds with the update of the set of candidate load areas M and executes the aggregation steps again. Otherwise, the algorithm returns $\boldsymbol{l}$ and C.

At the end of the procedure, each load area's substation is located at the weighted medoid of the power demand [34] to reduce voltage drops in the PDGs.

B. Graphical grid layout

The graphical grid layout contains the geospatial coordinates of each grid component, i.e., lines, loads, and transformers. We utilize Geographic Information System (GIS) tools to generate this layout to determine the potential line routes within the PDGs. We achieve this by leveraging publicly available transport infrastructure data, constraining the grid geometries.

The process entails several steps:

- *Data retrieval:* Data for the specific study zone are queried from OpenStreetMap (OSM) [24] with OSMnx [26], and structured in graphs, with edges representing transport paths and nodes denoting intersections or endpoints.
- *Convex envelopes:* These envelopes provide an easy way to obtain the smallest convex set that encloses all the points for the load areas within the study zone. Therefore, we can leverage them to retrieve the transport data within them. As there might be relevant streets surrounding the envelopes, we add a buffer distance to expand their boundaries and include nearby transport paths.
- *Transport graph filtering:* We query the transport paths within the respective envelope for each load area. This entails creating a graph defined by edges with at least one node inside the envelope.
- *Load-to-Grid Connection:* We determine the points where loads connect to the graphical grids by projecting the loads onto the nearest geographical point along the transport paths. Consequently, the nodes within the graphical grid encompass not only street nodes but also the projected load points, with the lines representing the connections between these nodes.

Finally, since the results of the previous steps might contain some unnecessary data, such as nodes with no demand that are connected only through one transport path. To address this, we conduct a data cleaning process for the graphical grid layout that comprises the following steps:

- *Leaf node removal:* We delete leaf nodes[2] with no demand until no more such nodes remain.
- *Redundant node elimination:* We delete redundant nodes, i.e., without demand and with multiple neighbors, all

[2]In an undirected graph like PDGs, a leaf node is a node with only one edge connecting to the rest of the graph.

of which also have no demand. We merge the edges connecting these redundant nodes.

- *Edge trimming:* We delete edges that follow transport paths that are significantly longer than the statistical population of the rest of the edges. These might be either unnecessary or replaced by straight lines built outside transport paths, as observed in real distribution grids.
- *Ensuring full connectivity:* As the edge trimming might have produced disconnected graphs, we will add straight lines to connect them. The process begins by removing graphs without loads. Then, we connect the graph with the fewest nodes by drawing straight lines from each node in this graph to its nearest node belonging to any of the other graphs. We repeat this until the graphical grid is fully connected.

C. From graphical grid layouts to power distribution grids

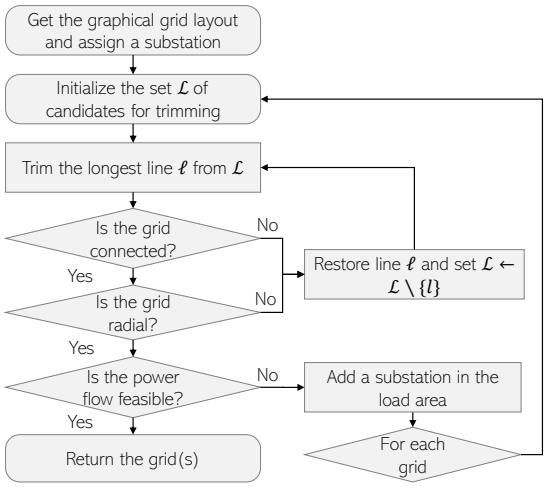

Fig. 4. Steepest-descent-based algorithm for obtaining radial grids.

The standard topology for PDG operation is radial [27]. Consequently, the graphical grids must be adapted because their initial layout is based on the meshed land transport networks. To this aim, we employ an approach grounded in steepest descent schemes for line trimming [22], [35], as illustrated in Fig. 4. The method considers running AC power flow calculations based on different line types, i.e., conductors with different ampacity and impedance.

The method encompasses the following steps:

- *Initial candidate set:* All grid lines are initial candidates for trimming, denoted as $\mathcal{L}$.
- *Line trimming:* Following [22], the longest line l in $\mathcal{L}$ is removed from the grid, and the grid is tested for connectivity and radiality. If either condition is not met, l is removed from $\mathcal{L}$ and is restored in the grid. This step repeats until $\mathcal{L}$ is empty.
- *Power flow feasibility:* AC power flow feasibility is tested for steady-state under-voltage and line congestion, considering the highest ampacity line type for all grid lines. If the power flow is infeasible, one additional substation is introduced, and the load area is split using load-weighted `k-means`. Subsequently, the previously trimmed lines are restored, and the trimming procedure is repeated. The algorithm terminates when the power flow is feasible and returns the feasible radial grid (or grids if substations are added).

After the load area splitting, grids may contain leaf nodes devoid of demand or substations. These nodes are subsequently removed.

Algorithm 1 Selection of line types

Input: 1) Set of lines E, 2) Set of nodes V, where $s \in V$ is the substation, and 3) The line types parameters, sorted by ampacity in ascending order $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N$.
Get the set $T \subset V$ of leaf nodes.
Initialize: 1) the evaluated lines $\mathcal{K} \leftarrow \{\}$, 2) the expanded lines $\mathcal{J} \leftarrow \{\}$, and the counter $i \leftarrow 0$.
repeat
  $i \leftarrow i + 1$
  Set the line type $\boldsymbol{\theta}_e \leftarrow \boldsymbol{\xi}_i$ for all line e in E
until power flow is feasible
for t in T **do**
  $\mathcal{R} \leftarrow$ line path from s to t
  $\mathcal{K} \leftarrow \mathcal{K} \cup \mathcal{R}$
  $\mathcal{R} \leftarrow \mathcal{R} \setminus \mathcal{K}$
  $\boldsymbol{\theta}_e \leftarrow \boldsymbol{\xi}_1$ for all line e in $\mathcal{R}$
  while power flow is not feasible **and** $\mathcal{R} \neq \emptyset$ **do**
    $r \leftarrow$ the first line of $\mathcal{R}$.
    if r not in $\mathcal{J}$ **then**
      $\mathcal{J} \leftarrow \mathcal{J} \cup \{r\}$
      repeat
        $\boldsymbol{\theta}_r \leftarrow$ the following line type to $\boldsymbol{\theta}_r$
      until power flow is feasible **or** $\boldsymbol{\theta}_r$ is equal to $\boldsymbol{\xi}_i$
    end if
    $\mathcal{R} \leftarrow \mathcal{R} \setminus \{r\}$
  end while
end for
for t in T **do**
  $\mathcal{D} \leftarrow$ line path from t to s
  Initialize: $\boldsymbol{\lambda} \leftarrow \boldsymbol{\xi}_1$, the highest rating observed in the path.
  while $\mathcal{D} \neq \emptyset$ **do**
    $d \leftarrow$ the first line of $\mathcal{D}$
    if $\boldsymbol{\theta}_d$ has higher rating than $\boldsymbol{\lambda}$ **then**
      $\boldsymbol{\lambda} \leftarrow \boldsymbol{\theta}_d$
    else if $\boldsymbol{\theta}_d$ has lower rating than $\boldsymbol{\lambda}$ **then**
      $\boldsymbol{\theta}_d \leftarrow \boldsymbol{\lambda}$
    end if
    $\mathcal{D} \leftarrow \mathcal{D} \setminus \{d\}$
  end while
end for
return $\boldsymbol{\theta}_e$ for all e in E

Lastly, since DSOs aim to build PDGs that must meet their customers' power requirements while efficiently using their resources, we must assign cost-efficient line types. Therefore, we evaluate assignments under steady-state conditions with peak power demand.

To select the line types, we leverage the fact that line ampacities in radial grids are non-increasing downstream from substations. Therefore, we can sequentially test line types in the paths[3] from the grid's substation to leaf nodes. This property allows the development of a computationally efficient algorithm for line-type selection in radial grids (see Algorithm 1). The inputs to the algorithm are the set of lines E and their lengths, the set of nodes V and their power demand, and the line parameter vectors $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N$, containing ampacity, resistance, and reactance parameters and sorted by increasing ampacity values.

After initialization (Lines 1-3), the algorithm identifies the line type $\boldsymbol{\xi}_i$ with the minimum ampacity, that assigned to all grid lines concurrently yields a feasible power flow (Lines 4-7). Subsequently, for each line path from the substation to a leaf node, the algorithm initially sets the line type $\boldsymbol{\xi}_1$ to all its lines (Lines 8-12). The algorithm assesses line types with increasing ampacity sequentially downstream the path until the power flow becomes feasible (Line 13-22). This process considers line types with an ampacity less than or equal to that of $\boldsymbol{\xi}_i$. This process is repeated for all paths connecting the substation to the leaf nodes (Lines 8-23). Finally, we ensure that the ampacity in the grid is non-decreasing upstream, i.e., if the ampacity decreases upstream at any line on a path, the line type is updated to a larger ampacity (Lines 24-36).

D. Differences between low- and medium-voltage grids

MV load areas are more extensive than LV load areas, necessitating more data processing concerning the land transport infrastructure for defining MV grids. Consequently, to alleviate the computational burden in the MV graphical grid data cleaning process (Section II-B), we adopt a strategy to reduce the edges set based on transport paths. Specifically, we employ the Minimum Spanning Tree approach, which selects the paths connecting all nodes with the minimum total length [36].

Furthermore, the differences between MV and LV load areas require diverse strategies for trimming long lines in the graphical grid data cleaning process. Given the smaller size of LV load areas, their transport paths are predominantly confined to homogeneous settlement types, such as urban or rural areas. Consequently, we assume that transport path lengths within a specific LV load area originate from a relatively homogeneous statistical population. Under this condition, we can identify and eliminate outlier paths based on their lengths x_i using robust statistics techniques [37]. To achieve this, we utilize z-scores in based on robust estimates of location υ and scale η [38]:

$$\upsilon = \underset{i=1,\ldots,n}{\text{median}}(x_i) \tag{2a}$$

$$\eta = 1.483 \cdot \underset{i=1,\ldots,n}{\text{median}} |x_i - \upsilon| \tag{2b}$$

$$z_i = \frac{x_i - \upsilon}{\eta} \qquad \forall i. \tag{2c}$$

[3]A line path from a node a to a node b is a list of lines [e_1, e_2, …, e_N] such that e_1 connects a, e_2 is connected to e_1, and so on until e_N connects b.

(2a)-(2b) are the robust estimates of location and scale[4], respectively, based on the length values x_i, and (2c) are the z-scores. Scores ranging from -2.5 to 2.5 are typically considered inliers.

As MV load areas can encompass multiple settlement types, the assumption of a homogeneous population of edge lengths no longer holds. Consequently, we opt to trim transport paths exceeding a realistic line length threshold based on statistical information on line lengths. Furthermore, in MV load areas, it is crucial to account for demand points physically separated by water bodies. We tackle this challenge by integrating water body data into our clustering methodology, thereby adapting the algorithm in Section II-A. Specifically, in *Step 2*, we start by splitting load areas internally separated by lakes, and in *Step 3*, we do not aggregate clusters that are separated by lakes.

III. Case study

We demonstrate the applicability of our framework by generating MV and LV grids using public data for Switzerland, a country with a land area of $41\,290$ km^2 and a population of $8\,703\,405$ [39]. We estimate the spatial distribution of power demand at the building level by rescaling the sectorial (industrial, commercial, and residential) heating and cooling demands, which are mapped with a resolution of 100×100 m [40]. Following [22], we assume similar electricity and heat demand spatial patterns. The rescaling is based on the peak electricity demand of the Swiss cantons [41] and the consumption share of each sector [42]. The power demand is allocated to buildings using Voronoi partitions. Each one of the centroids of the 100×100 m demand map is a generating point of the Voronoi partition. Then, for each Voronoi partition, the demand of its generating point is spread among the buildings contained in the partition proportionally to their surface. Building shapes are derived from OSM [24].

Once we have assigned the loads to the buildings, we segment MV and LV consumption. First, we use typical simultaneity factors for the LV loads to get the observed peak power demand by MV/LV transformers [43]. After this, we assign loads bigger than 100 kVA to MV level and to LV otherwise [44]. To delineate LV load areas, we differentiate between urban, peri-urban, and rural zones, as defined by [45] for each Swiss municipality. Given the similar conditions between Switzerland and Germany [46], we adopt German standards as input parameters for the grouping method outlined in Section II-A. For the distance limit δ, we consider 400 meters in urban zones and 2000 meters in peri-urban and rural zones, aligning with typical distances between transformers [47]. For the power demand threshold τ, we set 630 kVA for urban and peri-urban areas and 250 kVA for rural zones, in accordance with standard values for these conditions [48]. Additionally, we assume a power factor of 0.97 for all LV loads [14].

For the MV level, we include the MV/LV transformers and the MV consumers, also using standard simultaneity

[4]The constant 1.483 is a correction factor that makes η, the median absolute deviation from the median, an unbiased estimator at the normal distribution.

factors [43]. We assume that loads greater than 4 MVA correspond to high-voltage (HV) consumers, and thus, we exclude them. Also, we keep out loads below 50 kVA, which represent less than 0.8% of the MV loads after excluding HV consumers. These loads are primarily in remote locations and can be assumed to be supplied by stand-alone microgrids. Furthermore, we consider only one distance parameter for clustering MV load areas, set to 10 km, because MV substations show no regularity in the distances separating them. In addition, we adopt one type of transformer, with a power rating of 25 MVA, previously used for a Swiss case study [22], which sets the power demand upper bound per load area. We obtain the water bodies data from [49] and consider power factors of 0.90 for all the MV loads [4].

To derive the graphical grids, we download all the land transport pathways of Switzerland from OSM [24]. We filter the paths for each segmented load area, as described in Section II-B. The boundary buffers of the load areas' convex envelopes are set to 100 m for LV and 200 m for MV. Also, for trimming LV graphical grids, we delete paths above 2.5 median absolute deviations [38] among the set longer than 20 m. In the MV graphical grids, we leave out streets longer than 1.5 km since more than 95% of reference MV lines are shorter than that [50].

TABLE I
LV AND MV LINES SPECIFICATIONS.

ID	Type	Ampacity (A)
	Low voltage	
1	NAYY4x150SE	270
2	NAYY4x240SE	364
	Medium voltage	
3	48AL1/8ST1A	210
4	94AL1/15ST1A	350
5	122AL1/20ST1A	410
6	243AL1/39ST1A	645
7	NA2XS2Y1x70	220
8	NA2XS2Y1x185	362
9	NA2XS2Y1x240	421

TABLE II
LV AND MV TRANSFORMERS SPECIFICATIONS.

Type	Voltages	MVA
LV-Rural	20/0.4 kV	0.25
LV-Urban & LV-Peri-urban	20/0.4 kV	0.63
MV	110/20 kV	25

For the characterization of electrical components, we use international standard types [51], considering the operation of MV grids at 20 kV and LV at 0.4 kV. Table I shows the line types[5] and Table II shows the transformer types. We use cables for LV grids, setting the line type 1 for rural and peri-urban and type 2 for urban zones [48]. For MV, we consider types 3, 4, 5, and 6 of overhead lines (OHL) and types 7, 8, and 9 of cables. When MV lines follow the land transport infrastructure, we select cables and place OHL if the line is straight in the graphical grid. Furthermore, more than one line can be placed in a given transport path, as in real PDGs.

The AC optimal power flows are computed using PyPSA [52]. We impose Swiss requirements [53] as steady-state voltage deviation constraints, i.e., 3% for LV grids and 2% for MV grids.

IV. RESULTS

Our approach generates a set of MV and LV PDGs designed to meet the power demands of their customers under present regulations. We analyze MV grids in Section IV-A and LV grids in Section IV-B. These PDGs can be publicly accessed in GitHub[6].

A. Medium-voltage grids

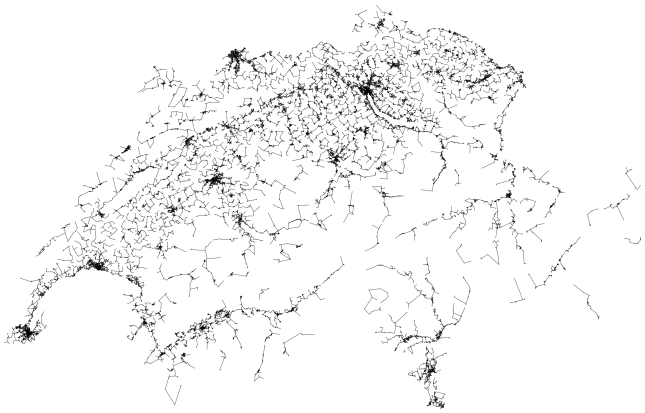

Fig. 5. The setup of 792 geo-referenced MV grids in Switzerland.

We initiate the process by locating the MV loads in the country to infer the MV grids. Our analysis reveals a total of 675 MV load areas. Subsequently, after deriving the corresponding graphical grid layouts, we characterize a total of 792 MV grids, as illustrated in Fig. 5. Remarkably, it is observed that in approximately 84% of cases, a single substation suffices to supply the MV load areas. Only 16% of the load areas require the presence of two substations, and eight load area requires three substations.

Figure 6 illustrates a single MV grid in the Canton of Vaud. This grid has 60 loads with a peak power demand of 13.32 MW and an overall line length of 29.36 km, 76.14% of which corresponds to cables. Remarkably, the derived lines follow street layouts in the zones with a high density of roads and pathways, and the derived OHLs are placed on the outskirts of the settlement.

Figure 7 presents the distribution of the relevant features of the inferred MV grids. A comparison is made with the typical cable ratio of German grids as documented in [50], where their average value is represented by a star marker in Fig. 7, panel (c). Notably, we observe a close alignment between the distributions of transformer capacity per supply point and the cable length ratio. Furthermore, we compare the overall active

[5]Note that NAYY4x240SE is not described in [51]. However, this is a common cable [48], and we obtained its data from [34].

[6]https://www.github.com/aeonetos/Swiss-PDGs

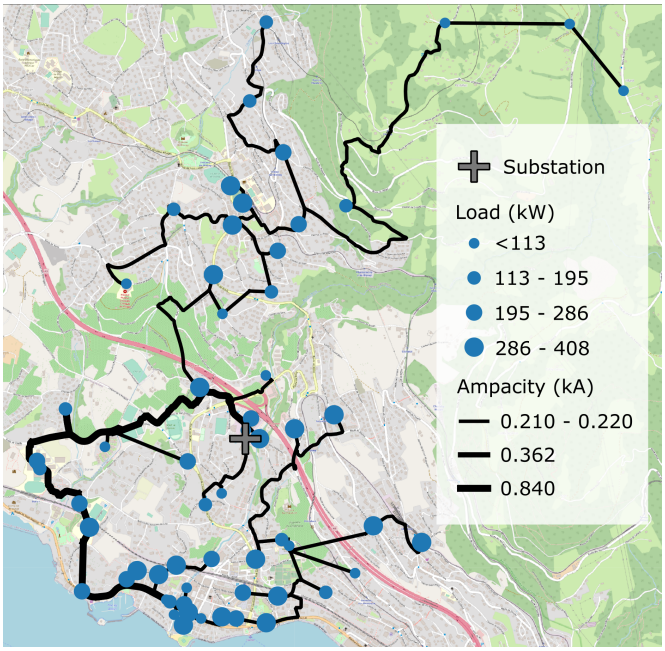

Fig. 6. Inferred MV grid located in the Canton of Vaud.

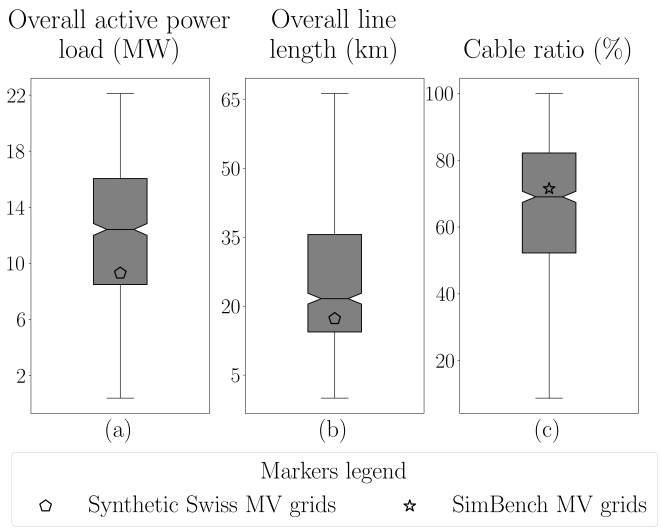

Fig. 7. Statistical description of the estimated MV grids. Pentagon and star markers represent the average values sourced from [22] and [50], respectively.

power load and line length of the inferred grids against the average values of the 776 synthetic Swiss MV grids obtained in [22], depicted using pentagonal markers in Fig.7, panel (a) and (b), and generated through an alternative approach based on different data sources. [22] infers a close number of grids with lower average power requirements than our inferred grids, which were inferred under peak simultaneous power requirements. Also, the smaller overall line lengths in [22] can be attributed to the reduced power requirements. It is worth noting that this discrepancy in overall line lengths is also influenced by our consideration of geographical constraints imposed by the layout of transportation infrastructure, which is not included in prior work.

B. Low-voltage grids

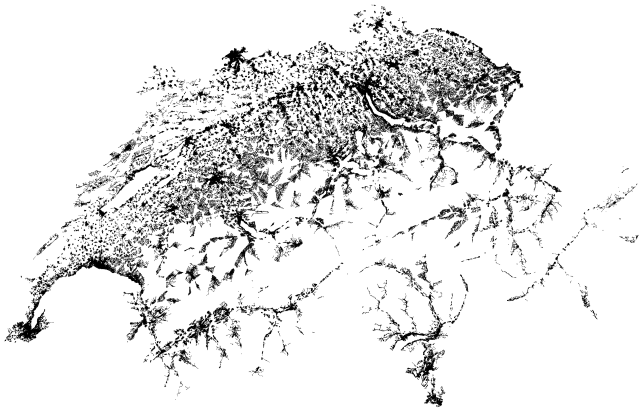

Fig. 8. The setup of 34 920 geo-referenced LV grids in Switzerland.

Our methodology identifies 31 796 LV load areas and, from this dataset, it infers 34 920 LV grids distributed across different zones, namely, 12 765 in rural zones, 6099 in peri-urban zones, and 16 056 in urban zones. The load clustering method reveals that approximately 88.83% of these grids are supplied by a single substation, while the remaining grids require two or more substations to meet the power demand. Fig. 8 shows all inferred geo-referenced LV grids in Switzerland. We observe that the density of the grids is higher in densely populated areas of the country, particularly the cities.

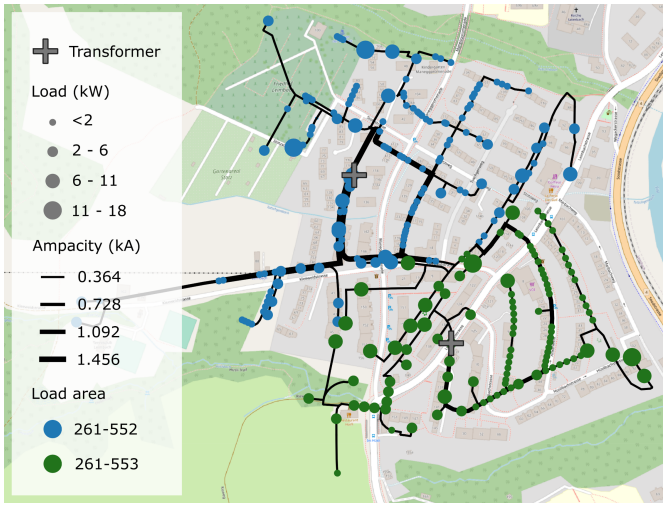

Fig. 9. Exemplary LV grids located in the Hüslibachstrasse surroundings, municipality of Zurich.

Figure 9 shows two urban LV grids in Zurich. Specifically, Grid 261-552 has 192 nodes, a peak power demand of 382 kW, and a total line length of 5.25 km; and Grid 261-553 has 149 nodes, a peak power demand of 357 kW, and a total line length of 3.01 km. Both grids are equipped with transformers rated at 630 kVA. The real distribution power system of the area is analyzed in [54], [55], revealing a substantial relation with the inferred data. In particular, the observed peak power demand of the area is 0.85 MW, in contrast to the inferred demand of 0.74 MW; the number of observed nodes is 254,

while the inferred nodes are 341. Furthermore, the real system is also equipped with two 630 kVA transformers.

Figure 10 shows the features of the LV grids for the three types of zones. We observe that the transformer capacity and the load density of the supplied area highly determine the total load and the geographic extension of the grids. Because of the lower transformer capacity, the overall active load of rural grids is significantly lower when compared to urban and peri-urban grids. Additionally, rural grids depend upon fewer feeders, consistent with their power demand. Urban areas have a higher power demand density than peri-urban areas. While these grids' overall active power load is similar, the total line length in peri-urban areas is appreciably greater. The distributions of the number of feeders of peri-urban and urban grids are almost identical despite the noticeable difference in their overall line length.

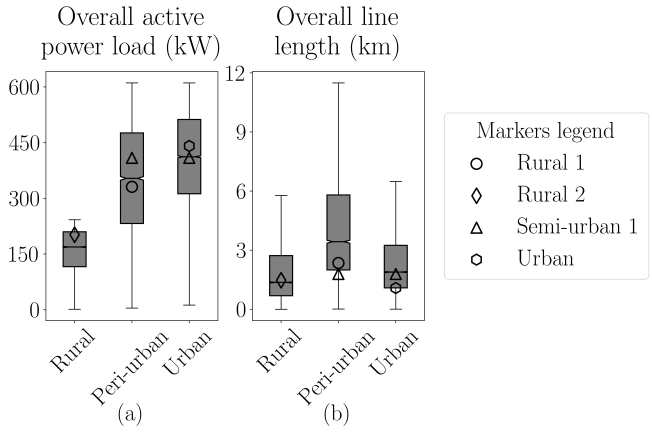

Fig. 10. Statistical description of the estimated LV grids. The markers show the values corresponding to four grid archetypes provided by [48]: Rural 1 (circle), Rural 2 (diamond), Semi-urban 1 (triangle), and Urban (hexagon).

Currently, there is a notable absence of LV benchmark grids specifically tailored for Switzerland. However, German LV grids offer a valuable alternative for evaluating the Swiss LV infrastructure, as highlighted in [46]. These German grids are readily accessible through [48] and encompass diverse LV grid archetypes, including rural, semi-urban, and urban categories, as detailed in [56].

The statistical characteristics presented in Fig. 10 closely align with the values of their respective representative grids, indicated by the markers. Note that the German grids and the inferred ones employ transformers with the same power rating in all the comparisons, except for our peri-urban grids (630 kVA) with their rural grid 1 (400 kVA).

Given potential disparities in our LV grid categorization, we conducted comparisons with the most analogous German archetype for each inferred grid. Remarkably, all the data points fall within the interquartile range, with one exception: comparing our peri-urban grids with the semi-urban grid 1 in terms of overall line length. This discrepancy may suggest that our peri-urban grids encompass regions with a lower demand density per surface area than the definition of semi-urban grid 1 used in the German dataset.

V. Discussion

Section IV demonstrates that the proposed framework, based on open data, produces accurate large-scale inference of LV and MV grids. However, it is important to acknowledge the limitations of our framework, which warrant further investigation. Firstly, we model balanced three-phase grids, while real-world PDGs are often unbalanced [57], potentially leading to voltage deviations [58]. Secondly, our modeling focuses on steady-state power flows during peak demand without accounting for transient effects on the PDGs. Additionally, we simplify the topological representation of feeders by treating lines placed on the same geographical path as parallel. Thirdly, while MV grids in practice are often meshed [4], our work infers radial topologies, reflecting their more common operational configuration. Fourthly, our framework does not identify renewable distributed generation and energy storage systems, which can significantly impact the temporal profiles of generation and demand in PDGs [59]. This should be considered in the future when significant penetration of distributed energy sources will be present in PDGs, and they have been designed or expanded accordingly. Lastly, we do not differentiate between various types of transport paths, allowing the placement of underground cables in any path, including dirt roads, which may not occur in real-world scenarios.

VI. Conclusions

In this study, we have developed a comprehensive framework for the large-scale inference of PDGs using open data sources. Our framework can infer LV and MV grids, providing precise geo-referencing for all grid components. It is not only computationally efficient but also yields statistically accurate results. To achieve this, we have developed two novel algorithms for the framework: a clustering method for detecting load areas and a selection algorithm for determining line types.

The clustering algorithm is a significant innovation as it allows for parallel processing during grid estimation, streamlining the identification of load areas that require a single transformer in most cases. Moreover, the line type selection algorithm significantly expedites the process of selecting cost-effective line types while ensuring that power flows and voltages remain within permissible limits.

To demonstrate the capabilities of our framework, we applied it to infer the power distribution infrastructure of Switzerland. Our results validate the framework's accuracy, as the inferred grids exhibit properties consistent with real-world and synthetic MV and LV reference grids.

Our framework holds great value for future research. It sets the stage for investigations, including but not limited to the assessment of novel control strategies for distributed energy sources, the analysis of distribution market designs, the characterization of hazard risk exposure, and the determination of PV hosting capacities. Furthermore, using inferred grids with high-performance computing can facilitate national-level energy system simulation and planning. These opportunities can significantly influence and shape the future of power distribution grids.

ACKNOWLEDGMENTS

This project was carried out with the support of the Swiss Federal Office of Energy SFOE as part of the SWEET EDGE project (https://www.sweet-edge.ch). The authors bear sole responsibility for the conclusions and the results. In addition, we thank Paolo Gabrielli and Nicolas Stocker for their technical advice, Katherine Lonergan for her valuable feedback on the manuscript, and Amirhassan Keshavarzzadeh for data management and processing.

REFERENCES

[1] 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)*, pp. 1–6, 2021.

[2] United Nations, "Net-zero coalition," 2022.

[3] R. Knecht, F. Carigiet, A. Schwab, P. Korba, and F. Baumgartner, "Techno-economic evaluation of voltage dependent active and reactive power control to reduce voltage violations in distribution grids," 2018.

[4] 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*, vol. 977, p. 012007, IOP Publishing, 2018.

[5] M. Grzanic, M. G. Flammini, and G. Prettico, "Distribution network model platform: A first case study," *Energies*, vol. 12, no. 21, p. 4079, 2019.

[6] H.-A. Jacobsen, P. Nasirifard, J. Rivera, and P. R. Baruah, "Inference of distribution grids based on crowdsourced grid data and drone imagery," *IEEE Transactions on Sustainable Computing*, 2019.

[7] M. Ali, C. A. Macana, K. Prakash, R. Islam, I. Colak, and H. Pota, "Generating open-source datasets for power distribution network using openstreetmaps," in *2020 9th International Conference on Renewable Energy Research and Application (ICRERA)*, pp. 301–308, 2020.

[8] C. Ma, J.-H. Menke, J. Dasenbrock, M. Braun, M. Haslbeck, and K.-H. Schmid, "Evaluation of energy losses in low voltage distribution grids with high penetration of distributed generation," *Applied Energy*, vol. 256, p. 113907, 2019.

[9] R. Allan, R. Billinton, I. Sjarief, L. Goel, and K. So, "A reliability test system for educational purposes-basic distribution system data and results," *IEEE Transactions on Power Systems*, vol. 6, no. 2, pp. 813–820, 1991.

[10] A. Koirala, L. Suárez-Ramón, B. Mohamed, and P. Arboleya, "Non-synthetic european low voltage test system," *International Journal of Electrical Power & Energy Systems*, vol. 118, p. 105712, 2020.

[11] 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.

[12] M. Liang, Y. Meng, J. Wang, D. L. Lubkeman, and N. Lu, "Feedergan: Synthetic feeder generation via deep graph adversarial nets," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1163–1173, 2021.

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

[14] 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.

[15] H. K. Çakmak, 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 Conference Europe (ISGT-Europe)*, pp. 1–6, IEEE, 2022.

[16] 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.

[17] 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.

[18] S. Lloyd, "Least squares quantization in pcm," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.

[19] S. Corigliano, T. Carnovali, D. Edeme, and M. Merlo, "Holistic geospatial data-based procedure for electric network design and least-cost energy strategy," *Energy for Sustainable Development*, vol. 58, pp. 1–15, 2020.

[20] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *kdd*, vol. 96, pp. 226–231, 1996.

[21] 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.

[22] 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.

[23] C. Mateo, F. Postigo, F. de Cuadra, T. G. San Roman, T. Elgindy, P. Dueñas, B.-M. Hodge, V. Krishnan, and B. Palmintier, "Building large-scale us synthetic electric distribution system models," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5301–5313, 2020.

[24] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org ," 2017.

[25] I. Filippidis, "OpenStreetMap Functions." https://github.com/johnyf/openstreetmap, 2023. Accessed: March 29, 2023.

[26] G. Boeing, "Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks," *Computers, Environment and Urban Systems*, vol. 65, pp. 126–139, 2017.

[27] P. Paiva, H. Khodr, J. Dominguez-Navarro, J. Yusta, and A. Urdaneta, "Integral planning of primary-secondary distribution systems using mixed integer linear programming," *IEEE Transactions on Power systems*, vol. 20, no. 2, pp. 1134–1143, 2005.

[28] S. Mandal and A. Pahwa, "Optimal selection of conductors for distribution feeders," *IEEE Transactions on Power Systems*, vol. 17, no. 1, pp. 192–197, 2002.

[29] E. Miguez, J. Cidras, E. Diaz-Dorado, and J. Garcia-Dornelas, "An improved branch-exchange algorithm for large-scale distribution network planning," *IEEE Transactions on Power Systems*, vol. 17, no. 4, pp. 931–936, 2002.

[30] J. Gomez, H. Khodr, P. De Oliveira, L. Ocque, J. Yusta, R. Villasana, and A. Urdaneta, "Ant colony system algorithm for the planning of primary distribution circuits," *IEEE Transactions on power systems*, vol. 19, no. 2, pp. 996–1004, 2004.

[31] E. Diaz-Dorado, J. Cidras, and E. Miguez, "Application of evolutionary algorithms for the planning of urban distribution networks of medium voltage," *IEEE Transactions on Power Systems*, vol. 17, no. 3, pp. 879–884, 2002.

[32] S. Najafi, S. H. Hosseinian, M. Abedi, A. Vahidnia, and S. Abachezadeh, "A framework for optimal planning in large distribution networks," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 1019–1028, 2009.

[33] J. Puchinger, G. R. Raidl, and U. Pferschy, "The multidimensional knapsack problem: Structure and algorithms," *INFORMS Journal on Computing*, vol. 22, no. 2, pp. 250–265, 2010.

[34] 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*, pp. 1–6, 2013.

[35] J. M. Nahman and D. M. Perić, "Radial distribution network planning under uncertainty by applying different reliability cost models," *International Journal of Electrical Power & Energy Systems*, vol. 117, p. 105655, 2020.

[36] R. Graham and P. Hell, "On the history of the minimum spanning tree problem," *Annals of the History of Computing*, vol. 7, no. 1, pp. 43–57, 1985.

[37] M. Hubert, P. J. Rousseeuw, and S. Van Aelst, "High-breakdown robust multivariate methods," *Statistical science*, vol. 23, no. 1, pp. 92–119, 2008.

[38] P. J. Rousseeuw and M. Hubert, "Robust statistics for outlier detection," *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, vol. 1, no. 1, pp. 73–79, 2011.

[39] The World Bank, "Switzerland data," 2020. Data retrieved from The World Bank, https://data.worldbank.org/.

[40] Swiss Confederation, "Heat/cooling demand," n.d. Accessed: 2022-10-31.

[41] Swissgrid, "Aggregated energy data of the control block Switzerland," 2022. Accessed: 2022-10-31.

[42] L. Eymann, J. Rohrer, and M. Stucki, "Energieverbrauch der schweizer kantone : Endenergieverbrauch und mittelabfluss durch den energieimport," tech. rep., ZHAW Z"urcher Hochschule f"ur Angewandte Wissenschaften, Winterthur, 2014.

[43] 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 Transactions on Power Systems*, vol. 26, no. 1, pp. 190–197, 2011.
[44] H. L. Willis, *Power distribution planning reference book*. CRC press, 2004.
[45] Swiss Energy Research for the Energy Transition, "Sweet-edge," 2021. The project's page can be accessed through https://www.sweet-edge.ch/en/home.
[46] S. Kortmann, X. Han, M. Schwarz, and G. Hug, "From a distributing to a generating network: Assessing pv hosting capacity under uncertainty in distribution grids," in *PESS 2021; Power and Energy Student Summit*, pp. 1–6, VDE, 2021.
[47] Forschungsgemeinschaft für Elektrische Anlagen und Stromwirtschaft, "Ein werkzeug zur optimierung der störungsbeseitigung für planung und betrieb von mittelspannungsnetzen," 2008.
[48] 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)*, pp. 1–5, 2019.
[49] Swiss Confederation, "Hydrography," n.d. Accessed: 2023-04-13.
[50] S. Meinecke, D. Sarajlić, S. R. Drauz, A. Klettke, L.-P. Lauven, C. Rehtanz, A. Moser, and M. Braun, "Simbench—a benchmark dataset of electric power systems to compare innovative solutions based on power flow analysis," *Energies*, vol. 13, no. 12, 2020.
[51] L. Thurner, A. Scheidler, F. Schäfer, J. Menke, J. Dollichon, F. Meier, S. Meinecke, and M. Braun, "pandapower — an open-source python tool for convenient modeling, analysis, and optimization of electric power systems," *IEEE Transactions on Power Systems*, vol. 33, pp. 6510–6521, Nov 2018.
[52] T. Brown, J. Hörsch, and D. Schlachtberger, "PyPSA: Python for Power System Analysis," *Journal of Open Research Software*, vol. 6, no. 4, 2018.
[53] J. Meyer, W. Mombauer, and M. Híckel, "The third edition of the austrian-czech-swiss-german (d-a-ch-cz) technical rule for assessment of network disturbances," in *CIRED 2021 - The 26th International Conference and Exhibition on Electricity Distribution*, vol. 2021, pp. 1026–1030, 2021.
[54] V. Poulios, E. Kaffe, F. Kienzle, H. Luternauer, and L. Küng, "Challenges with pv grid integration in urban distribution systems: A case study in the city of zurich," in *CIRED workshop*, pp. 11–12, Citeseer, 2014.
[55] V. Poulios, E. Vrettos, F. Kienzle, E. Kaffe, H. Luternauer, and G. Andersson, "Optimal placement and sizing of battery storage to increase the pv hosting capacity of low voltage grids," in *International ETG Congress 2015; Die Energiewende-Blueprints for the new energy age*, pp. 1–8, VDE, 2015.
[56] C. Kittl, D. Sarajlić, and C. Rehtanz, "k-means based identification of common supply tasks for low voltage grids," in *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pp. 1–5, 2018.
[57] P. Li, H. Ji, C. Wang, J. Zhao, G. Song, F. Ding, and J. Wu, "Optimal operation of soft open points in active distribution networks under three-phase unbalanced conditions," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 380–391, 2019.
[58] S. Weckx and J. Driesen, "Load balancing with ev chargers and pv inverters in unbalanced distribution grids," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 2, pp. 635–643, 2015.
[59] J. Jo and J. Park, "Demand-side management with shared energy storage system in smart grid," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4466–4476, 2020.