Automatic Generation of Real Power Transmission Grid Models from Crowdsourced Data

Jose Rivera, *Member, IEEE,* Pezhman Nasirifard,
Johannes Leimhofer and Hans-Arno Jacobsen, *Senior Member, IEEE*

***Abstract*—Real models of electrical transmission grids are difficult to obtain. The process of generating such models from unstructured and incomplete data is tedious, and the resulting models are rarely updated. This paper proposes a novel approach for automatically extracting power-relevant data from the public and unstructured crowdsourced OpenStreetMap (OSM) and for generating topology and simulation-ready models of real transmission grids based on the relation between different grid elements such as power lines and substations. Our approach uses spatial analysis and minor assumptions to periodically generate transmission grid models based on the latest OSM data for every country on the planet. A comparison of our generated power grid models with official data from 14 countries reveals accuracy levels between 31%-94%, caused by the varying availability of OSM data for different countries. Since the crowdsourced data is continuously improving, the automated and periodical model generation approach extends the models with new power circuits as the quantity and the quality of the OSM dataset increases. We provide a platform to access our generated models at opengridmap.org. This paper describes our model generation method, presents our data access platform and evaluates the accuracy of our topological models for selected countries.**

***Index Terms*—Power grid data, transmission grid, simulation models, crowdsourcing, smart grid**

I. INTRODUCTION

THANKS to the Paris Agreement of 2015 [1], many countries around the world have committed to a significant reduction in their CO2 emissions. The development of a smart grid plays a crucial role in achieving these goals and several approaches to support the transition to renewable energy sources are currently being researched. Solutions range from significant changes in the control and structure of the current transmission and distribution grids to the introduction of new devices such as distributed generators, distributed storage units, and electric vehicles [2], [3]. However, to accurately assess the technical, financial and social impact of the proposed solutions, many complex simulation studies using realistic grid data are required. Currently, simulation research studies for power grids rely heavily on standardized test grids, such as the IEEE test feeders [4], the Power Systems Test Case Archive [5], the CIGRE test feeders [6], and the PNNL feeders [7]. Although these test cases are often used and have undoubtedly contributed to the reproducibility, and thus the comparability, of research results, it is important to point out that these models often do not take the individual properties of power grids, such as geographic features, utility strategies, and local legislation, into consideration. The assessment of different smart grid solutions would benefit from the use of real power grid data, preferably from the specific region where their implementation is planned. Moreover, open real power grid models would help with the transparency, comparability, and reproducibility of studies. The problem, however, is that the availability of open and realistic power grid models for such simulations is insufficient.

In this work, we propose a novel approach for generating real power grid models from crowdsourced data. Our approach periodically and automatically downloads crowdsourced data of OpenStreetMap (OSM) and extracts the power-relevant data and infers relations between the extracted data to create transmission grid models of the countries. The primary challenge of this approach lies in the inconsistency and the low quality of the OSM data for many areas. Since the availability of data differs significantly from country to country, the quality of the inferred models can also change significantly from one area to the other.

Several studies have made use of OSM data to generate power grid models [8]–[10]. However, most of the previous approaches rely on structured data, in other words, these approaches require the OSM relations between the crowdsourced power elements, such as power lines, substations, and power generators, to be already manually defined as OSM power circuit relations. The novelty of our approach is in inferring these power circuit relations automatically. Based on the inferred relations and the available meta-data of the power elements, we automatically generate a *Common Information Models (CIM)* [11] of the grid and subsequently produce simulation models to support power flow simulations. Depending on the availability of crowdsourced power data in OSM, we offer models for many countries with varying degrees of quality. We make our inferred models openly available in various formats at opengridmap.org. In this work, we describe our inference approach, present our platform to access and improve our models, and most importantly, evaluate our resulting models for selected countries by comparing them with the available official data.

This article complements the existing research in power grid models as follows:

1) We propose and develop a novel approach for creating

Jose Rivera, Pezhman Nasirifard, Johannes Leimhofer, and Hans-Arno Jacobsen are with the Department of Computer Science, Technical University of Munich, D-85748, Munich, Germany. (`j.rivera@tum.de`, `p.nasirifard@tum.de`)

power transmission grid models of countries by inferring OSM-like power circuit relations solely based on the public and free crowdsourced OSM data.

2) We introduce a novel algorithm for performing spatial analysis on power-relevant OSM elements to automatically infer relationships between power lines, power generators, and substations. The results of our evaluation show the high accuracy of the inferred relations in comparison to the manually defined OSM relations. Therefore, liberating OSM volunteers from creating OSM power relations manually.
3) We introduce a novel approach for converting OSM-like power relations to CIM models that can be used to create MATLAB Simulink models when required.
4) We provide a unique and novel workflow for automatically and periodically downloading the whole OSM dataset of the planet, for inferring and generating transmission grid models of the whole planet and publishing the model for public use.
5) We provide a platform at opengridmap.org that allows open and detailed access to inferred models for the whole planet in different formats.
6) We evaluate the accuracy of the inferred power grid models of 14 countries compared to the official data of the selected countries. The results of the evaluation show the high accuracy of the inferred models; however, the accuracy significantly varies across countries depending on the availability of OSM data.

The rest of the paper is structured as follows: Section II offers an overview of the related work on different inference approaches of the topological models of the power grid. Our approach for automatically generating power grid models based on crowdsourced OSM data and transferring the models to CIM and simulation models is explained in Section III. We present Transnet, our inference engine, and our model access platform, in Section IV. We evaluate our inference approach by assessing the quality of inferred topology models for several studied countries in Section V. Finally, Section VI provides our concluding remarks and future works.

II. Related Work

Our approach focuses on the use of the OpenStreetMap (OSM) data to generate power grid models. The OSM project relies on a community to crowdsource geographical data and to create a map dataset which is free to use, editable and licensed under the Open Data Commons Open Database License (ODbL) [12]. There are over 15 million mapped power elements in the OSM database[1]. This significant amount of openly available data has encouraged research that focuses on the generation of open real power grid models.

We initially introduced the idea of using OSM data to generate real power grid models in [13] and since then there have been several works on the subject. One related project is SciGRID [10], in which an inferred model of the German transmission grid is produced using OSM power relations. In another related work [9], they generate the transmission grid topology for the German state of Baden-Württemberg using OSM power relations and governmental data sources. They also generate a comprehensive simulation model and carry out several studies. Both previous works demonstrate the potential of OSM data to generate real power grid models. However, their reliance on manually defined power relations severely limits their ability to fully utilize the available OSM power elements, which for the most part, have not been organized into power circuit relations.

One of the issues with the OSM dataset is the low availability of underground grid elements because of the difficulty of mapping buried elements for the OSM community. Therefore, the majority of the OSM power-relevant elements are the overhead or visible devices. However, in this work, we focus on high voltage transmission grids with a voltage level higher than 100 kV, which are often installed above ground. For example, in Germany, 83% of low and medium voltage cables are buried underground and about 95% of high voltage cables are installed above ground [14]. We address the issues with underground power cables, especially in the distribution level, in another study [15]. There, we offer an approach with the help of crowdsourcing campaigns and applying graph theory approaches for generating distribution grid models based on the location of grid devices, consumer endpoints and road structure of an area.

Recent works have made efforts to automate the creation of OSM power circuit relations. In [16], a transmission network in Germany is obtained by extending OSM power relations. Power ways were extracted from existing OSM power relations, and missing unmapped ways were inferred using a heuristic routing algorithm. While an improvement, this approach is still reliant on previously manually defined OSM relations. Moreover, the heuristic algorithm removed many unrecognized ways making the produced model inaccurate and also ineffective for sparse networks. A significant improvement is presented in GridKit [17], which creates a model for the German electrical transmission grid. Their approach is independent of OSM power relations, and instead, it makes use of spatial analysis on available OSM power-relevant objects such as power lines and stations. The result is a graph of the German transmission grid, where the vertices present the power stations and the edges show the power lines. This result, however, is not a correct model, since the produced edges do not represent the real geographical course of power lines. Moreover, there was no evaluation of the accuracy of the inferred model. In fact, the lack of comprehensive evaluation of the inferred models is a significant problem that affects all of the previously mentioned works. Also, a method for improving the quality of the generated models is missing.

The approach presented in this work builds on our initial work in [18], on automatic OSM power circuit relation inference, and substantially extends it to provide a complete solution for creating and improving transmission grid simu-

---

[1] https://taginfo.openstreetmap.org/keys/power, accessed 2017-11-17

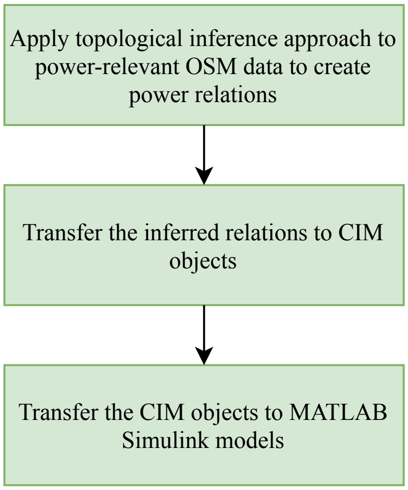


Fig. 1: Overview of the power grid models generation approach.

lation models for any geographical region. We differentiate ourselves from previous works by using spatial analysis to define OSM power relations automatically and to generate geographically accurate power grid models. Also, by using some assumptions and available OSM metadata, our approach allows the generation of simulation models for power flow calculations. Moreover, we provide a platform that continually downloads the latest OSM dataset (updated every week), automatically infers the power transmission grid of several countries and makes them openly available to the broader community. Furthermore, we flag the issues with data quality during the inference and provide an option to improve the crowdsourced data quality. Finally, we propose an approach to evaluate the accuracy of OSM inferred power grid models and apply it to evaluate our results for selected countries.

III. AUTOMATIC POWER GRID MODEL GENERATION

Our approach of automatically generating power grid simulation models is divided into three phases, as shown in Fig. 1. In the first phase, *OSM power circuit relations* are inferred from unstructured OSM data. The second phase uses the resulting OSM power circuit relations to produce CIM description files. In the third and final phase, the previously generated CIM file is used to define a parametrized *MATLAB Simulink model*, such as power flow simulations. In the following, we describe the different phases and illustrate their result with a simple example.

A. Inference of OSM Power Circuit Relations

To infer the transmission grids from OSM data, we need to extract the power-relevant objects and organize them into the circuits that represent the power grids. The OSM data is formatted in XML and structured as *Relation*, *Way*, and *Node* objects. Relations express the relationship between objects, Ways describe open or closed paths, and Nodes, which are the smallest structural entities, illustrate geographical positions containing the longitude and latitude of nodes. All these objects can carry further descriptive elements, such as key-value pairs of attribute and tag elements. Power-relevant objects contain power tags, such as *<tag k='power' v='generator' />*. As an example, Listing. 1 shows the representation of a power pole in OSM.

Listing 1: OSM node representation for power pole in XML format

```
<node id="1988705117" lat="48.4687278" lon="9.2840263">
    <tag k="power" v="pole"/>
</node>
```

The transmission grid for a particular geographical area consists of several power circuits. A power circuit represents a single power line with its end nodes. The end nodes could be any grid devices, such as a power plant or substation. In OSM, power circuits are represented as a relation, according to the OSM power routing proposal [19]. As an example, Listing. 2 shows the OSM power circuit relation with the ID of *5581802*, and its corresponding geographical representation is shown in Fig. 2. We use this power circuit relation for the rest of the paper to present our examples. The power circuit relation 5581802 is composed of two end nodes, a substation (*role* = *"substation"*) and a power plant (*role* = *"plant"*), and also several line (*role* = *"line"*) elements that represent the power line that connects the two nodes. This structure is common in all OSM power circuit relations, and we exploit it to formulate an algorithm that can automatically infer such relations from unstructured power-relevant objects.

Listing 2: OSM relation with ID 5581802 in XML format

```
<osm version="0.6" ... >
<relation id="5581802" ... >
    <member type="way" ref="23025610" role="substation"/>
    <member type="way" ref="115932301" role="line"/>
    <member type="way" ref="234156599" role="line"/>
    <member type="way" ref="234156600" role="line"/>
    <member type="way" ref="11341038" role="plant"/>
    <tag k="cables" v="3"/>
    <tag k="frequency" v="50"/>
    <tag k="operator" v="50Hertz"/>
    <tag k="route" v="power"/>
    <tag k="type" v="route"/>
    <tag k="voltage" v="380000"/>
</relation>
</osm>
```

Algorithm. 1 is our spatial approach to infer the transmission grid of a country from power-relevant OSM elements. The generated transmission grid is a list of inferred power circuit relations that we create from the power lines and grid endpoints. The algorithm receives as input the power-relevant OSM data of the country, which includes the power lines and the grid endpoints, including the power plants, generators, and substations. First, the algorithm extracts the transmission voltage levels of the country by querying the voltage tags assigned to the power lines. We only consider voltage tags that are higher than 100 kV and only if there exist at least thirty different power lines with such a voltage level. Then, for each voltage level, the algorithm filters the power lines with equivalent voltage levels, and it also filters all the grid endpoints that geographically intersect with any filtered lines. For every endpoint, the algorithm determines the lines which are intersecting with the endpoint, and the endpoint is the starting point of a new power circuit. For every intersecting line, the algorithm recursively makes attempts to find other attached

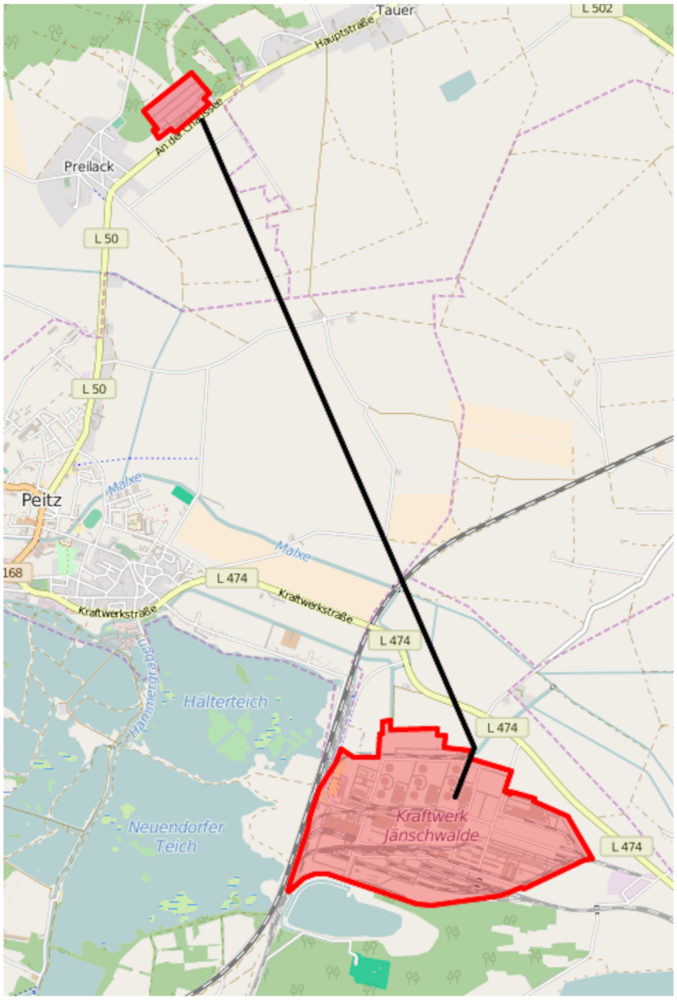


Fig. 2: OSM relation with ID 5581802 representing a power circuit of the transmission system.

power lines to the line that leads to another grid endpoint. Once the other endpoint is found, the new power circuit is complete, and the algorithm instantiates an OSM power relation object containing the two endpoints and the connecting power lines between the endpoints. Finally, the circuit is added to a list which represents the topology of the power grid. Listing. 3 shows the output of the algorithm's execution, where the already existing OSM relation with ID 5581802 is thoroughly reconstructed and recognized by our algorithm. The extracted relation contains several metadata which are not presented in Listing. 3. The metadata includes information on the location and characteristics of grid elements, such as the longitude and latitude and the voltage and frequency.

Relation 5581802 is an example of the simplest station-to-station connection. In reality, relations can have a much more complex structure. For example, Fig. 3 illustrates a T-junction connection between three stations: X, Y and Z. The inference result of this connection is two smaller relations with members *[X, a, b, Y]* and *[X, a, c, Z]*. Furthermore, to prevent the inclusion of duplicated relations, we utilize a duplication removal mechanism that ensures each circuit is represented only once in the final model and the identical circuits are removed. The inference algorithm applies duplicate removal to the inferred grid model before returning the model. As an example, neither the relations with members *[Y, b, a, X]* nor *[Y, b, c, Z]* are included in the final set. The former is equivalent to *[X, a, b, Y]* and the latter is already represented by a combination of *[X, a, b, Y]* and *[X, a, c, Z]*.

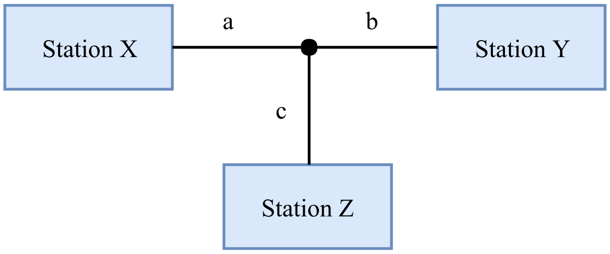


Fig. 3: Three stations connected by a T-junction.

While inferring the power circuits, we make use of a loop prevention mechanism. The loop prevention mechanism keeps a record of already covered endpoints and power lines. Whenever the inference algorithm traverses a station or power lines that are already covered, it drops the current relation inference and restarts a new relation. This mechanism avoids the inference of a relation from one endpoint to itself, and it also prevents the inference of duplicate relations. However, the loop prevention mechanism ignores the intentional loops in the grid, installed by the grid operators. During the early stages of development, we realized that including loops, by traversing every possible option from one power line to another line or endpoint, would not significantly improve the quality of the models, but it would significantly increase the complexity of the algorithm in terms of computation time and memory. Therefore, in favor of better performance, while operating on a global scale, we decided not to include loops in the inferred model.

Finally, we store the inferred topological models of each country as JSON and CSV files for further computation.

Listing 3: Partial inference output for constructing the existing OSM relation with ID 5581802

```
...
Relation 4
Station - ID: 23025610 Type: substation Name: Umspannwerk
Preilack Ref: PRL Voltage: 380000;110000
Line - ID: 234156600 Type: line Voltage: 380000
Line - ID: 234156599 Type: line Voltage: 380000
Station - ID: 11341038 Type: plant Name: Kraftwerk
J nschwalde
...
Found substation-covering relation (Id=5581802;Parts=
[115932301L, 234156599L, 234156600L, 23025610L, 11341038L]
Found substation-covering relation (Id=5581803;Parts=
[234156601L, 11341038L, ... 174126659L, 234156602L]
Most similar relation is 5581802 with 100.0%
...
```

B. Creation of CIM Models

We transform the extracted power circuit relations into CIM Models which are defined according to *ENTSO-E CIM Profile 1*[2] for European transmission networks. This profile

---

[2] https://www.entsoe.eu/digital/common-information-model/, accessed 2017-11-17

Algorithm 1: Inference of OSM power circuit relations

```
InferGrid ($GridEndpoints, PowerLines$)
    input  : $GridEndpoints$: Extracted OSM power plants and substations of country.
    input  : $PowerLines$: Extracted OSM power lines of country.
    output: $TopologicalGridModel$: Topological model of the transmission grid of country.
    gridModel = set()
    voltageLevels = extractHighVoltageLevels(PowerLines)
    foreach voltageLevel ∈ voltageLevels do
        filteredLines = filterPowerlinesWithVoltage(PowerLines, voltageLevel)
        intersectingEndpoints = filterIntersectingEndpoints(GridEndpoints, filteredLines)   // List of all endpoints that geographically intersects with any line
        foreach endpoint ∈ intersectingEndpoints do
            intersectingLines = filterIntersectingPowerLines(endpoint, filteredLines)
            foreach line ∈ intersectingLines do
                (otherEndpoint, connectingLines) = recursivelyFindOtherEndpoint(endpoint, line, filteredLines)
                if otherEndpoint != NULL AND isNotLoop(endpoint, connectingLines, otherEndpoint) then
                    gridModel.add(new Circuit(endpoint, connectingLines, otherEndpoint))
    gridModel = removeDuplicateCircuits(gridModel)
    return $gridModel$
```

describes the standard methods and parameters for modeling power circuits. To accurately convert the circuit relations into CIM models, we require converting each member of the relation, such as substations, generators and power lines, to the equivalent CIM objects according to the ENTSO-E CIM Profile 1 shown in Table. I. As an example, Fig. 4 shows the schematic representation of the resulting CIM model for our example relation with the ID 5581802. In Fig. 4, we convert the relation's substation to a CIM *Substation* which contains two instances of the *PowerTransformer* for each voltage level. Each *TransformerWinding* also provides a *Terminal* instance for connecting to other equipment. Furthermore, Fig. 4 includes a generator that is modeled as a CIM *GeneratingUnit* instance that contains a *SynchronousMachine* instance that also provides a *Terminal*. We convert the relation's power lines to *ACLineSegment* instances that provide *Terminal* instances on both sides. To construct a connection between terminals, we use *ConnectivityNode* instances. We should note that the *EnergyConsumer* object is not part of the OSM relation, but it is required to generate a valid load flow simulation model.

TABLE I: Used CIM Classes

CIM Object	Description
Substation	Container for substation equipment
PowerTransformer	Transformer in a Substation
TransformerWinding	Winding of a PowerTransformer
GeneratingUnit	Container for generator equipment
SynchronousMachine	Synchr. generator in a GeneratingUnit
EnergyConsumer	Load
ACLineSegment	Power line
Terminal	ConductingEquipment connection point
ConnectivityNode	Node, connecting two or more Terminals

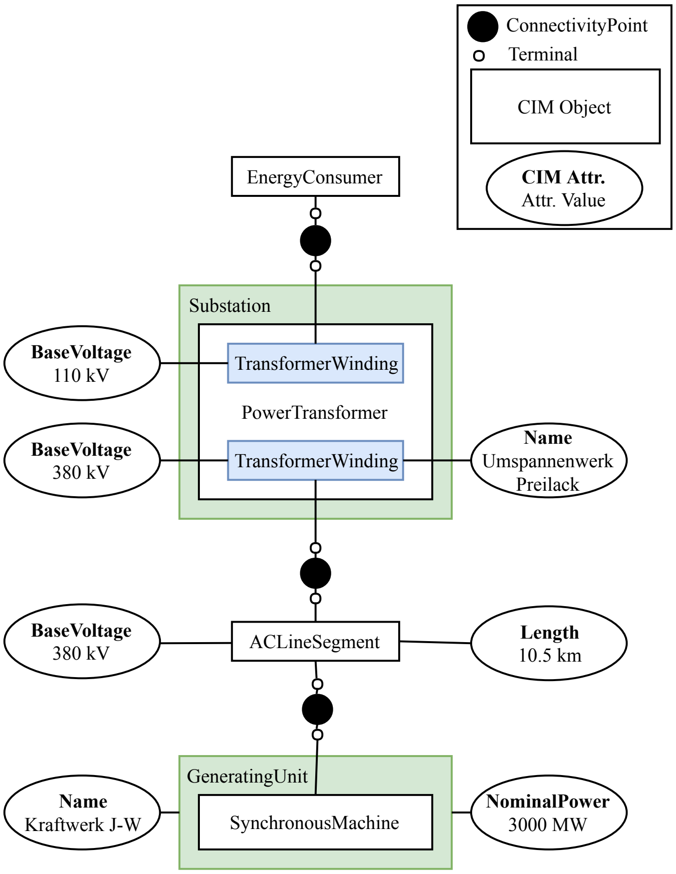


Fig. 4: CIM model for OSM relation (OSM-ID=5581802).

Although OSM data often contains some of the required information for generating correct CIM models, such as voltage level, the length of the line, and generator capacities, we cannot extract all mandatory parameters from OSM data. To compensate for the missing information, we make assumptions based on expert knowledge and data analysis techniques. For example, the load level is a parameter that the OSM data does not contain. Therefore, to estimate the loads attached to each substation, we segment the area using a Voronoi partition algorithm with the substations as seeds. Then, we multiply the average population per substation with the average per-head power consumption of inhabitants in the area, which results in an estimation of the attached load to the substation. We acquire the power consumption data from the World Bank's

Global Development Data [20] and the information on the population from the OpenGeoDB dataset [21] and also the World Bank's dataset [22]. The other missing values are the resistance and reactance of the line, which we estimate based on the characteristics of the line. We presume the material of the conductor based on the line's voltage, e.g. the high voltage lines are often made of aluminum. Given the presumed type of the conductor, the length of the line, the number of wires per conductor, which we acquire from OSM, we calculate a rough estimate for the resistance and reactance[3].

C. *Creation of MATLAB Simulation Models*

We provide an approach for parsing the CIM models and transforming the CIM objects to MATLAB Simulink models based on the model templates of the MATLAB *SimPowerSystems* library[4]. These model templates support parametrization and are available for several power grid components. We can use the resulting Simulink models for simulation and analysis studies.

As an example, Fig. 5 shows the resulting Simulink model of the generated CIM model of the OSM relation with ID 5581802. Our approach transforms the CIM objects with similar Simulink instances. The model consists of *SimPowerSystems* instances of type *Three-Phase Source* for the power plant, *Distributed Parameters Line* for the power line, *Three-Phase Transformer with Two Windings* for the substation, and *Three-Phase Series RLC Load* for the load. We fill the instances' parameters with the corresponding CIM model attributes, such as power line length and generator capacity. The *Load Flow Bus* instances (orange color) represent the measurement points of load flow simulations. We assume a balanced three-phase network where all power lines of the same voltage level have the same impedance. Each estimated load is attached to its corresponding substation. Furthermore, as an example, we include the results of a Newton-Raphson load flow simulation for the generated Simulink model in Fig. 5, where each bus shows the resulting relative voltage and the corresponding phase angle.

IV. TRANSNET INFERENCE ENGINE

Transnet is an inference engine that we designed and developed based on the discussed inference approaches. Transnet automatically generates power transmission grid models for every country on the planet based on the latest available OSM data. On a weekly schedule, Transnet downloads the latest OSM data for every continent from Geofabrik[5] and uses the Java application Osmosis[6] to filter the power-relevant OSM data. Then, the methods described in Section III are applied to every country to obtain the latest OSM power relations and models. Transnet offers the possibility for visualizing and downloading the available power transmission grid data for any particular voltage level, area or country in several formats.

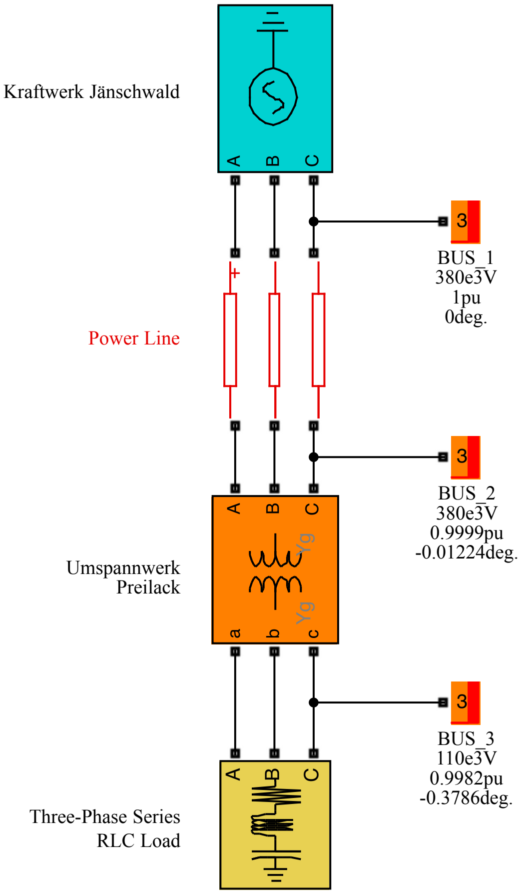


Fig. 5: Simulink model for the OSM relation with ID 5581802.

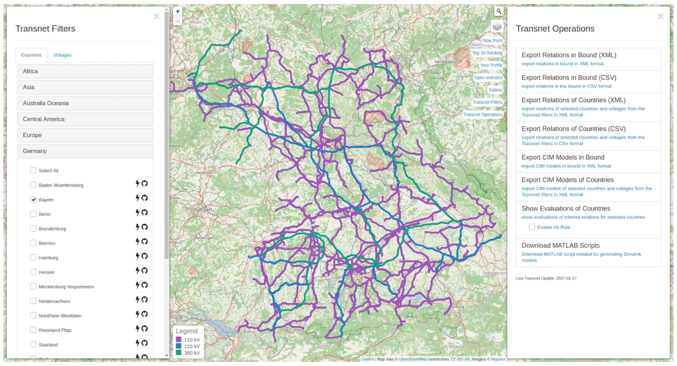


Fig. 6: Transnet layer in OpenGridMap.

A MATLAB script is also available for download, which allows the generation of MATLAB Simulink models from a downloadable CIM file. Moreover, we provide a contribution layer, where we display the discovered issues and incoherences of data integrity during the inference, and a link is provided to allow the contributors to improve and resolve these issues directly on OSM. We include a snapshot of the Transnet layer on OpenGridMap, as shown in Fig. 6.

Transnet consists of a set of Python and MATLAB scripts,

---

[3] https://www.electricalengineeringtoolbox.com/2016/01/resistance-and-reactance-per-km-of.html, accessed 2017-11-17

[4] http://de.mathworks.com/products/simpower/, accessed 2017-11-17

[5] http://download.geofabrik.de/, accessed 2017-11-17

[6] http://wiki.openstreetmap.org/wiki/Osmosis, accessed 2017-11-17

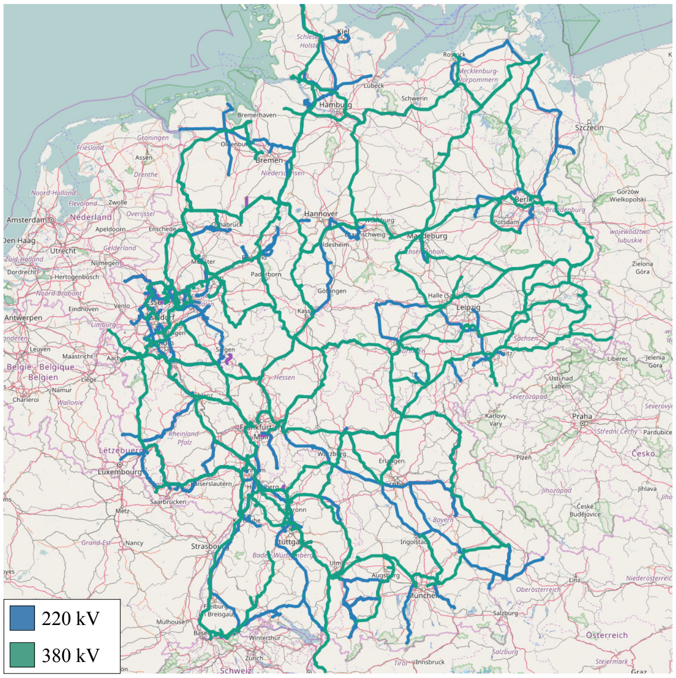


Fig. 7: OpenGridMap German transmission topology grid model.

backed by a PostgreSQL/PostGIS database. The source code of Transnet is open-sourced and freely available at github.com/OpenGridMap/transnet.

To adequately demonstrate the capabilities of Transnet, we briefly discuss the phases required to obtain a Simulink model for Germany. Fig. 7 shows the topological model of the German high voltage transmission grid, which the users download from Transnet. The CIM models of this topological model and the MATLAB script required for transforming the CIM models to Simulink models can also be downloaded from Transnet. Fig. 9 shows the resulting Simulink model for Germany. As mentioned in Section III, not all required electrical parameters for the CIM and Simulink models can be extracted from OSM data. Therefore, to parameterize our model, we estimate the loads using Germany's regional population dataset provided by *OpenGeoDB*[7] and the average per-head power consumption of a German citizen [23] on Voronoi partitions of the substations. Fig. 8 shows the Voronoi partitions for available substations. Once loads are estimated, the resulting Simulink model is simulation-ready. By using our platform, the users can follow a similar procedure to obtain power grid models for any other countries, which can be transformed into simulation models.

V. EVALUATION

In the following section, we evaluate the performance of our inference approach and the quality of our generated models. For our evaluations, we consider 14 countries that have more than five manually mapped power relations in OSM and for which official information is openly available about the length of the power lines. The selected countries are Denmark, Egypt, France, Germany, Great Britain, Ireland, Italy, Japan, Netherlands, Poland, Spain, South Africa, Turkey and the USA. These evaluations are carried out using OSM data from March 1, 2017.

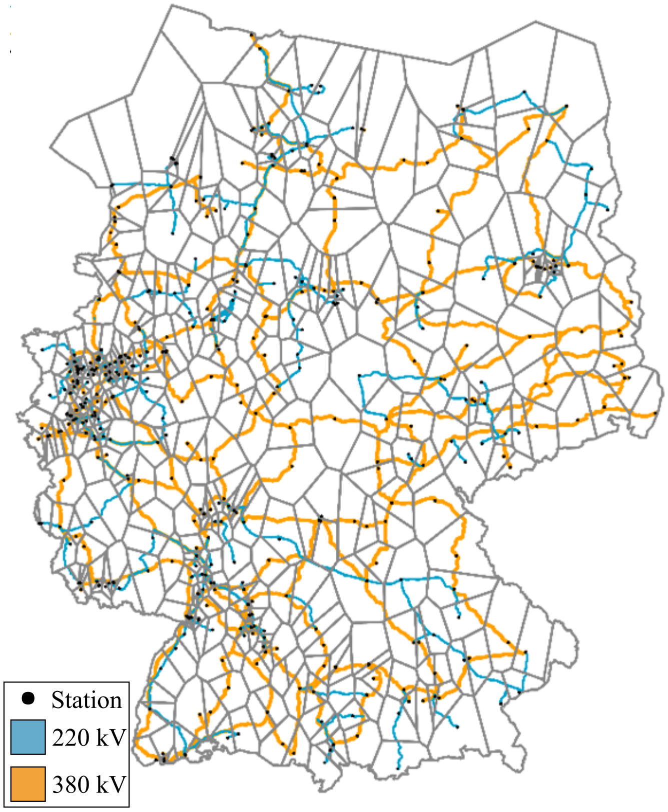


Fig. 8: Germany's substations Voronoi partition.

A. *Efficiency of Inference Algorithm*

To evaluate the accuracy of our algorithm for automatically generating OSM relations, we first compare the number of manually defined power relations in OSM with the number of inferred relations produced by our algorithm. In a second step, we quantify the ability of our algorithm for correctly identifying the manually defined relations. Using Algorithm. 2, we determine the percentage of OSM relations which are represented through our inferred relations; in other words, we measure the recall ability of our approach. An OSM relation is considered to be wholly represented when every station-to-station connection in the OSM relation is also present in either one of the inferred relations or a combination of multiple inferred relations. The evaluation algorithm requires every OSM relation to have a specific structure containing a voltage tag and precisely two stations as members. However, not every manually defined OSM relation has a valid structure. For this reason, before starting with the evaluation, we standardize OSM relations by checking the member voltage tag and member stations. When a voltage tag is missing, we estimate the voltage based on voltages of member power lines and member stations. Moreover, when

---

[7] http://opengeodb.giswiki.org/, accessed 2017-11-17

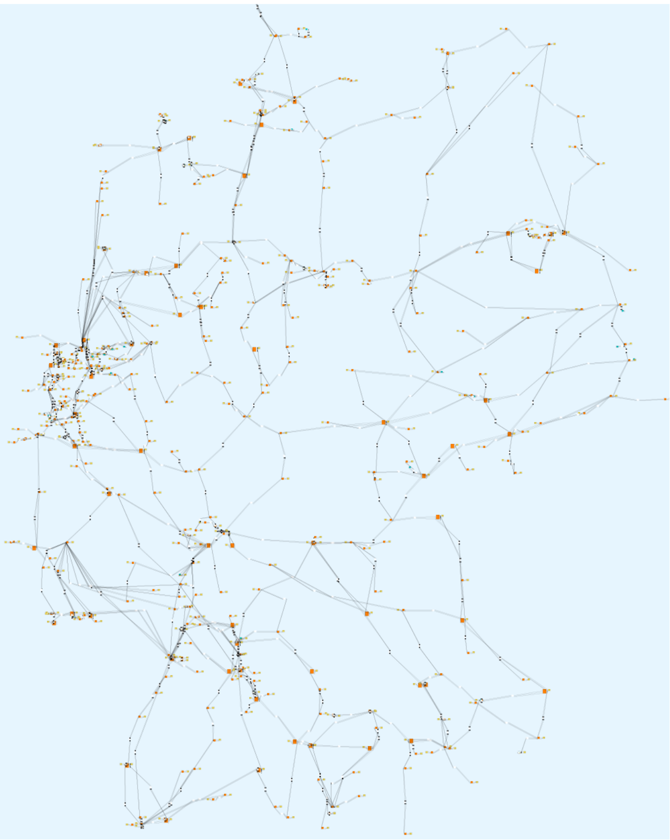


Fig. 9: Germany's generated Simulink model.

an OSM relation contains more than two member stations, we split the relation into separate relations with precisely two stations.

The result of the evaluation of the inference algorithm for the 14 selected countries is summarized in Table. II. Our approach generates a significant number of relations compared to the manually defined relations in the majority of studied countries excluding Germany. The German transmission grid is comprehensively modeled in OSM and has been manually defined into OSM power circuit relations. However, these comprehensive manually defined relations are not the case for most other countries. The results indicate that most of the studied countries lack a sufficient amount of manually defined relations.

Regarding the recall capability of our approach, as shown in Table. II, our approach achieves an 86% to 100% recall of manually defined relations. The main reason that all of the manually defined relations could not be identified is because the non-identified relations are incomplete and cannot be used. For example, the voltage levels cannot be determined, or they are missing necessary components from their structure. According to the results, we show that our automated approach exhibits performance comparable to a manual definition of power relations. Issues with incomplete and missing data are flagged and are displayed in the OpenGridMap map under *Contribute Layer* for improvement. All in all, these results reveal that our inference approach is an efficient method to automate the definition of OSM power relations and structure OSM power data for the generation of comprehensive power grid models.

Algorithm 2: Evaluation of inferred relations against OSM relations

```
ValidateRelations($(OSMRelations, InferredRelations)$)
    input : $OSMRelations$: Valid OSM relations of a country.
    input : $InferredRelations$: Inferred relations of a country.
    output: $CoveredOSMRelations$: Number of OSM relations which are covered by inferred relations.
    coveredOSMRelations = 0
    foreach OSMRelation $\in$ OSMRelations do
        stationsInRelation = getAllStationsInRelation(OSMRelation)
        numberOfHits = 0
        foreach inferredRelation $\in$ InferredRelations do
            if inferredRelation.voltage == OSMRelation.voltage then
                transitiveClosure = makeTransitiveClosureOfFirstStation(inferredRelation)
                foreach $station_i$ $\in$ stationsInRelation do
                    foreach $station_j$ $\in$ stationsInRelation do
                        if $station_i$ != $station_j$ $AND$ ($station_i$, $station_j$) is in transitiveClosure then
                            numberOfHits += 1
            if numberOfHits == getNumberOfStationPairs(stationsInRelation) then
                coveredOSMRelations += 1
    return $coveredOSMRelations$
```

B. Accuracy of Topological Models

Evaluating the accuracy of inferred real power grid models is challenging because usually the ground truth, such as the official model of the power grid, is not openly available for comparison. To evaluate the accuracy of our generated models, we compare the total length of the extracted power lines to the officially reported power line length. We use this measurement because the length of a country's power grid is one of the only parameters that is consistently reported by the transmission grid operators in most countries. We develop our measurement approach and apply it to all 14 studied countries. In our approach, the total length of an inferred power grid is defined as the sum of the power lines length of all relations. Hence, for a set of relations $R = \{r_1, ..., r_m\}$, where each relation is defined as r_i = [$station_{start}, line_1, ..., line_l, station_{end}$], the length of each relation is defined as the sum of the lengths of each power line segment $Length(r_i) = \sum_{n=1}^{l} line_n$ and

TABLE II: Evaluation Against OSM Relations

Country	**Defined Relations**	**Inferred Relations**	**Recall(%)**
Egypt	16	131	100
Netherlands	6	66	100
South Africa	18	782	100
Spain	22	558	100
Ireland	279	280	95
Italy	265	501	95
Denmark	243	271	93
France	681	909	92
Great Britain	223	849	91
Turkey	10	817	90
Japan	18	289	89
USA	18	4302	89
Germany	780	679	87
Poland	146	190	86

The recall is the percentage of manually defined OSM relations accurately identified by our approach

the length of the inferred grid as a whole is the sum of the lengths of each relation $Length(R) = \sum_{n=1}^{m} Length(r_n)$.

Table. III shows the results for the 14 studied countries, where the accuracy of the inferred models is quantified by dividing the total length of the inferred models by the officially reported length of the grid for each country. As shown, the level of accuracy is very different among countries ranging from 31%-94% accuracy. The inferred models on the map for the selected countries are shown in Fig. 10, Fig. 11 and Fig. 12. The varying accuracy among countries is directly correlated with the availability of OSM power-related data for each country. The advantage of our approach is that the models are periodically updated with the latest OSM data. Just in the period of six months from March to November 2017, the number of ways (not only power-related ways) in OSM has increased by 9%[8]. Therefore, as more data becomes available, the accuracy of our automatically generated models will improve over time.

While our results show that we do not yet offer an entirely accurate model of the power grid for the studied countries, but they confirm that our approach automatically provides models that can serve as the basis towards the definition of an entirely accurate model. Most importantly, our approach substantially reduces the effort required for manually defining an entire power grid from unstructured data.

VI. CONCLUSIONS

In this paper, we proposed and evaluated an approach for automatically generating transmission grid models from the crowdsourced and public OpenStreetMap data. We provide an inference engine, called Transnet, which periodically downloads the latest OSM data of the planet, extracts the power-relevant elements and uses spatial analysis to infer the relationship between the extracted elements to generate structured transmission grid models. Furthermore, we provide an approach to transform the generated models to CIM models as well as MATLAB Simulink simulation models. To evaluate the quality of the inferred models, we inferred topological models of transmission grids of 14 countries with high availability of OSM data, and we evaluated the quality of the inferred power relations in comparison to the manually defined OSM power relations. The result of the evaluation revealed that our algorithm inferred significantly more relations than given by the number of manually defined OSM relations. Furthermore, the inference algorithm showed a recall rate of 86%-100% of manually defined OSM relations. We also compared the length of the inferred power lines to the officially reported length of countries' power grids which showed accuracy levels of 31%-94% among countries. The difference in accuracy between countries is caused by the difference in available OSM power-relevant data, which is the primary input of our approach. However, since our models are automatically generated and periodically updated, they are expected to improve over time as more data is added to OSM. We encourage everyone to join the OpenGridMap effort and help collect relevant power grid data to build models for our planet's power systems, see opengridmap.org for information on how to join.

As part of our future work, we plan to offer an approach to track the alterations of the accuracy of our models over time as the quality and quantity of OSM data changes. Furthermore, we expect to improve the robustness of our inference engine against the OSM power-relevant elements with missing data through correlation with other data sources. Moreover, to evaluate the performance and accuracy of our CIM and Simulink model generation approach, we plan to verify the accuracy of power flow simulations of the inferred models against well-established models such as the Polish 2736. To improve the quality of simulation-ready models, we purpose extending our platform with self-validating features to ensure that the model is ready for power flow analysis. Once we established a reasonable validation framework, we aim to apply our inference algorithm to more complex low voltage grids. Finally, we want to establish a platform that provides the user with the possibility for generating simulation models for the distribution grids.

ACKNOWLEDGMENT

The authors would like to thank all the countless contributors that have helped crowdsource geographical data. This research was supported by the German Federal Ministry of Education and Research Grant (BMBF 01IS12057) and the Alexander von Humboldt Foundation.

REFERENCES

[1] United Nations, "Paris Agreement," 2015, accessed: 2017-03-10.

[2] Deutsche Energie-Agentur, "Dena-Verteilnetzstudie: Ausbau- und Innovationsbedarf der Stromverteilnetze in Deutschland bis 2030," Tech. Rep., 2012.

[3] U.S. Department of Energy, "2014 Smart Grid System Report," 2014.

[4] W. Kersting, "Radial distribution test feeders," *Power Systems, IEEE Transactions on*, vol. 6, no. 3, pp. 975–985, 1991.

---

[8] https://osmstats.neis-one.org/, accessed 2017-11-17

TABLE III: Power Grid Length Accuracy

Country	Voltage Range (kV)	Reported Length (km)	Inference Length (km)	Accuracy(%)
Ireland	110-400	7950 [24]	7459	94
Denmark	132-400	5238 [25]	4499	86
South Africa	132-765	30072 [26]	25938	86
Turkey	154-380	53216 [27]	41951	79
Great Britain	132-400	23202 [28]	17906	77
Germany	380-220	35708 [29]	26690	75
Italy	220-380	22000 [30]	14140	64
Spain	220-400	40566 [31]	25620	63
USA	230-765	248648 [32]	148751	60
Poland	220-400	13365 [33]	7818	59
France	220-400	47431 [34]	27228	57
Japan	187-500	21437 [35]	10873	51
Netherlands	220-380	2760 [36]	1393	50
Egypt	220-500	19723 [37]	6068	31

Accuracy is the power grid's inferred length divided by the official length

[5] "Power Systems Test Case Archive," http://www2.ee.washington.edu/research/pstca/, accessed: 2017-01-31.

[6] K. Strunz, R. Fletcher, R. Campbell, and F. Gao, "Developing Benchmark Models for Low-Voltage Distribution Feeders," in *Power Energy Society General Meeting, 2009. PES '09. IEEE*, 2009, pp. 1–3.

[7] Pacific Northwestern National Laborattories, "Modern Grid Initiative Distribution Taxonomy Final Report," Tech. Rep., 2008.

[8] S. Altschaffl, R. Witzmann, and T. Ahndorf, "Generating a Pss Netomac Model of the German Transmission Grid from Google Earth and Visualizing Load Flow Results," in *Proceedings of the 2014 IEEE International Energy Conference (ENERGYCON)*. IEEE, 2014, pp. 603–609.

[9] H. Çakmak, H. Maaß, F. Bach, U. G. Kühnapfel, and V. Hagenmeyer, "Ein Ansatz zur automatisierten Erstellung umfangreicher und komplexer Simulationsmodelle für elektrische Übertragungsnetze aus OpenStreetMap-Daten," *Automatisierungstechnik*, vol. 63, no. 11, pp. 911–925, 2015.

[10] W. Medjroubi, C. Matke, and D. Kleinhans, "SciGRID - An Open Source Reference Model for the European Transmission Network (v0.2)," 2015, accessed: 2017-02-20.

[11] T. Berry, "Standards for Energy Management System Application Program Interfaces," in *Proceedings of the International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (Drpt)*. IEEE, 2000, pp. 156–161.

[12] M. Haklay and P. Weber, "OpenStreetMap: User-Generated Street Maps," *IEEE Pervasive Computing*, vol. 7, pp. 12–18, 2008.

[13] J. Rivera, C. Goebel, D. Sardari, and H. Jacobsen, *OpenGridMap: An Open Platform for Inferring Power Grids with Crowdsourced Data*. Springer International Publishing, 2015, pp. 179–191.

[14] S. von Roon, M. Sutter, F. Samweber, and K. Wachinger, "Netzausbau in deutschlan," 2014, accessed: 2018-01-03. [Online]. Available: http://www.kas.de/wf/doc/kas_38837-544-1-30.pdf

[15] P. Nasirifard, J. Rivera, Q. Zhou, K. B. Schreiber, and H. Jacobsen, "A crowdsourcing approach for the inference of distribution grids," in *Proceedings of the Ninth International Conference on Future Energy Systems*, ser. e-Energy '18. ACM, 2018, pp. 187–199.

[16] W. Medjroubia, U. Müllerb, M. Scharf, C. Matkea, and D. Kleinhansa, "Open Data in Power Grid Modelling: New Approaches Towards Transparent Grid Models," *Energy Reports*, vol. 3, pp. 14–21, 2016.

[17] B. Wiegmanns, "Improving the Topology of an Electric Network Model Based on Open Data," Master's Thesis, Energy and Sustainability Research Institute, University of Groningen, 2015, accessed: 2017-03-07. [Online]. Available: http://scigrid.de/publications/16_1_BWiegmans_Master_Thesis_2015.pdf

[18] J. Rivera, J. Leimhofer, and H. Jacobsen, "OpenGridMap: Towards Automatic Power Grid Simulation Model Generation from Crowdsourced Data," 2016.

[19] OpenStreetMap, "Power Routing Proposal," 2016, accessed: 2017-02-20. [Online]. Available: http://wiki.openstreetmap.org/wiki/Proposed_features/Power_routing_proposal

[20] T. W. Bank, "Global Population by Country," https://data.worldbank.org/indicator/SP.POP.TOTL, accessed: 2018-07-12.

[21] OpenGeoDB, "OpenGeoDB Open Source Database," http://opengeodb.giswiki.org/wiki/OpenGeoDB, accessed: 2018-07-12.

[22] T. W. Bank, "Electrical Power Consumption per Capita," https://data.worldbank.org/indicator/EG.USE.ELEC.KH.PC, accessed: 2018-07-12.

[23] Statista, "BDEW. (n.d.). Pro-Kopf-Stromverbrauch in Deutschland in den Jahren 1995 bis 2015 (in Kilowattstunden)," 2015, accessed: 2016-07-11. [Online]. Available: http://de.statista.com/statistik/daten/studie/240696/umfrage/pro-kopf-stromverbrauch-in-deutschland

[24] EirGrid, "The Irish Electric Power System," accessed: 2017-03-10. [Online]. Available: http://www.cigre.org/var/cigre/storage/original/application/d265c3b09604c7fe99501f2036459645.pdf

[25] DanskEnergi, "Danish Electricity Supply 08," 2008.

[26] Eskom, "The South African Electric Power System," accessed: 2017-03-11. [Online]. Available: http://www.cigre.org/var/cigre/storage/original/application/c9bda75aeaec519077e049b9ab74789d.pdf

[27] TEIAS, "The Turkish Electric Power System," accessed: 2017-03-11. [Online]. Available: http://www.cigre.org/var/cigre/storage/original/application/affeb24295861859ceec4db5e0a28d04.pdf

[28] N. Grid, "The Great Britain Electric Power System," accessed: 2017-03-10. [Online]. Available: http://www.cigre.org/var/cigre/storage/original/application/f61d0ebd5531fcbed39ef222fc7e8948.pdf

[29] B. B. der Energie-und Wasserwirtschaft, "Deutsches Stromnetz," 2008, accessed: 2017-03-09. [Online]. Available: https://bdew.de/internet.nsf/id/DE_20100322_PM_Deutsches_Stromnetz_ist_178_Millionen_Kilometer_lang/$file/Fakten_BDEW_Stromnetze.pdf

[30] Terna, "The Italian Electric Power System," 2014, accessed: 2017-03-10. [Online]. Available: http://www.cigre.org/var/cigre/storage/original/application/81c662618e407440ca396a07dd479164.pdf

[31] REE, "The Spanish Electric Power System," 2015, accessed: 2017-03-11. [Online]. Available: http://www.ree.es/en/activities/grid-manager-and-transmission-agent

[32] U.S. Department of Energy, "National Transmission Grid Study," 2002, accessed: 2017-03-10. [Online]. Available: https://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/TransmissionGrid.pdf

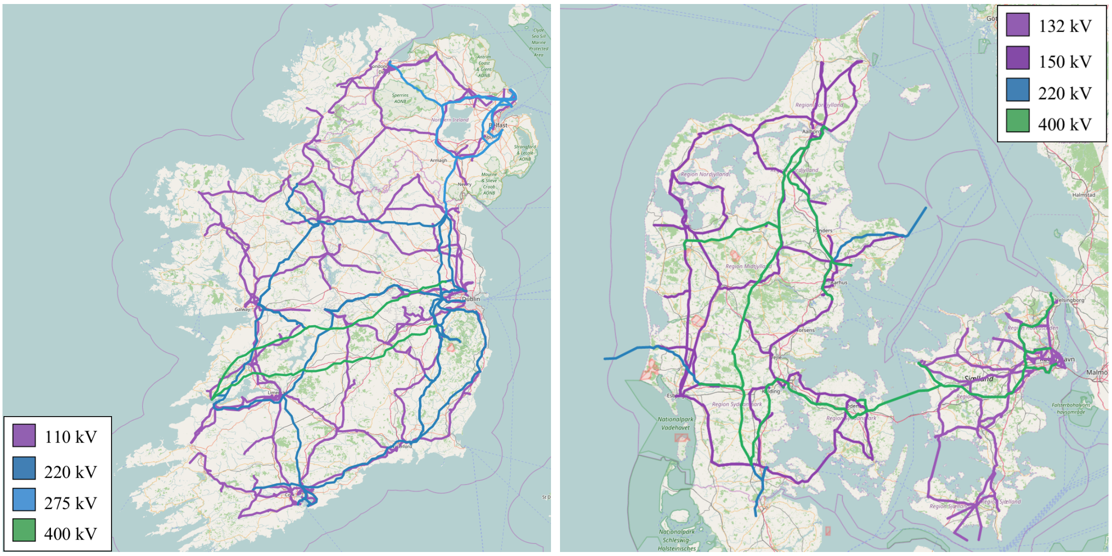
[33] PSE. S.A., "The Polish Electric Power System," accessed: 2017-03-11. [Online]. Available: http://www.cigre.org/var/cigre/storage/original/application/504bd2667347fe1a56772788e17787ea.pdf

[34] RTE, "The French Electric Power System," accessed: 2017-03-10. [Online]. Available: http://www.rte-france.com/en/screen/europe-s-biggest-transmission-system

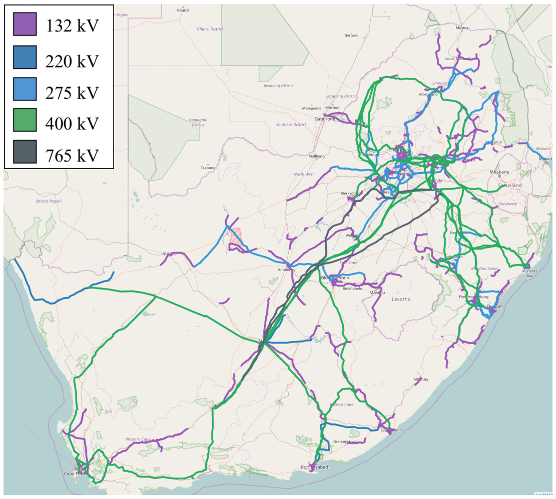
[35] FEPC, "The Japanese Electric Power System," 2013, accessed: 2017-03-10. [Online]. Available: http://www.cigre.org/var/cigre/storage/original/application/fac3623d5d889b8e3bb4821ceb21aa9f.pdf

[36] Tennet, "The Dutch Electric Power System," accessed: 2017-03-11. [Online]. Available: http://www.cigre.org/var/cigre/storage/original/application/d265c3b09604c7fe99501f2036459645.pdf

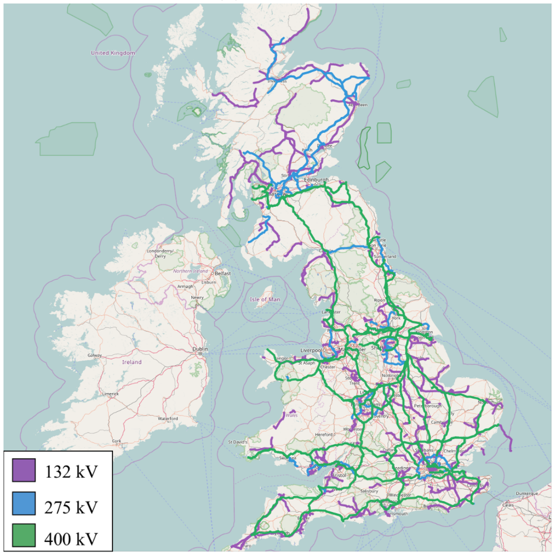
[37] MOEE, "Annual Report of Egyptian Electricity Holding Company," 2012, accessed: 2017-03-09. [Online]. Available: http://www.moee.gov.eg/english_new/EEHC_Rep/2011-2012.pdf



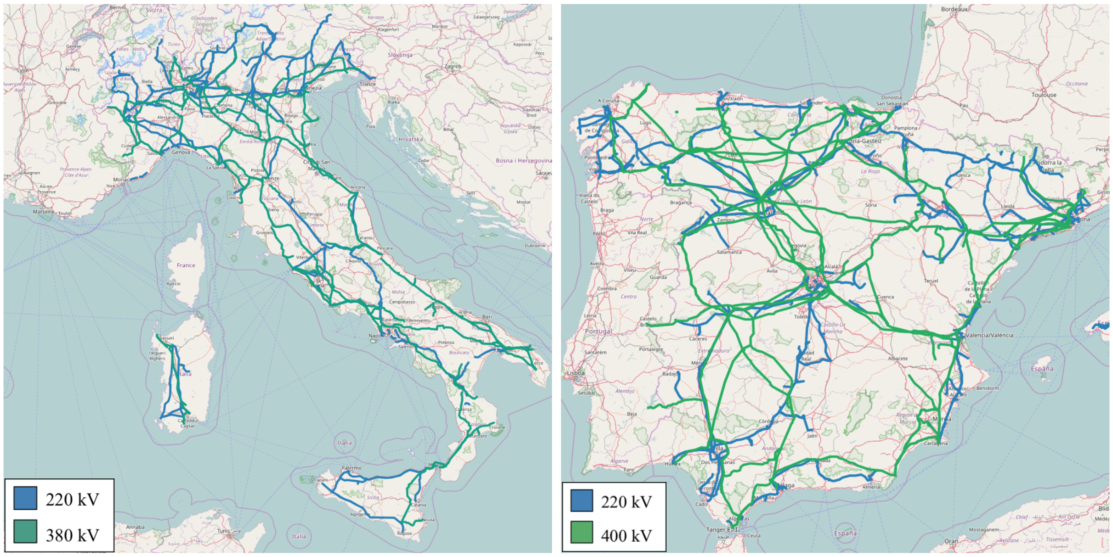
(a) Ireland (b) Denmark



(c) South Africa

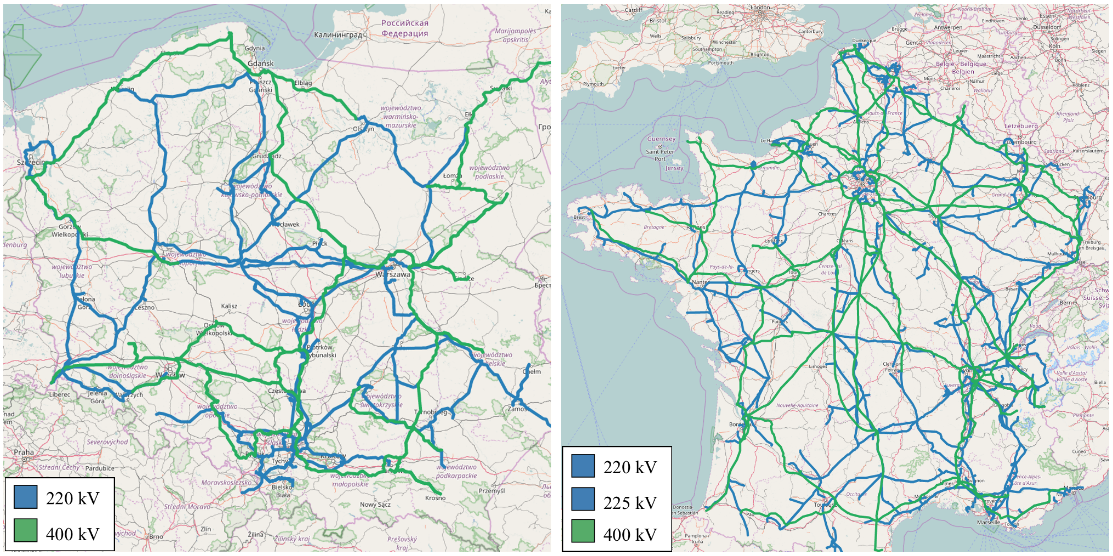


(d) Great Britain

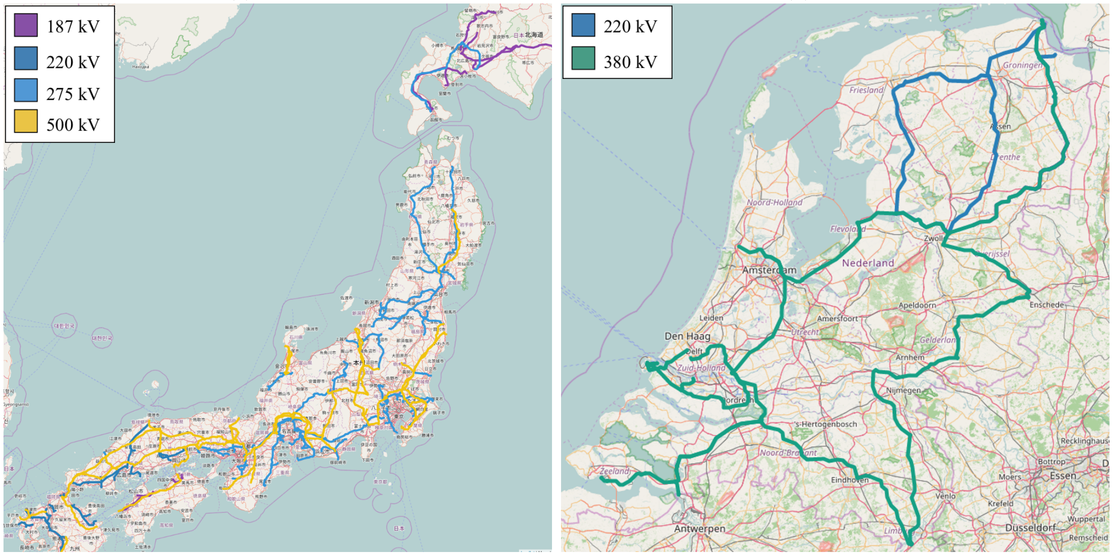


(e) Italy (f) Spain

Fig. 10: Inferred transmission grids for selected countries.



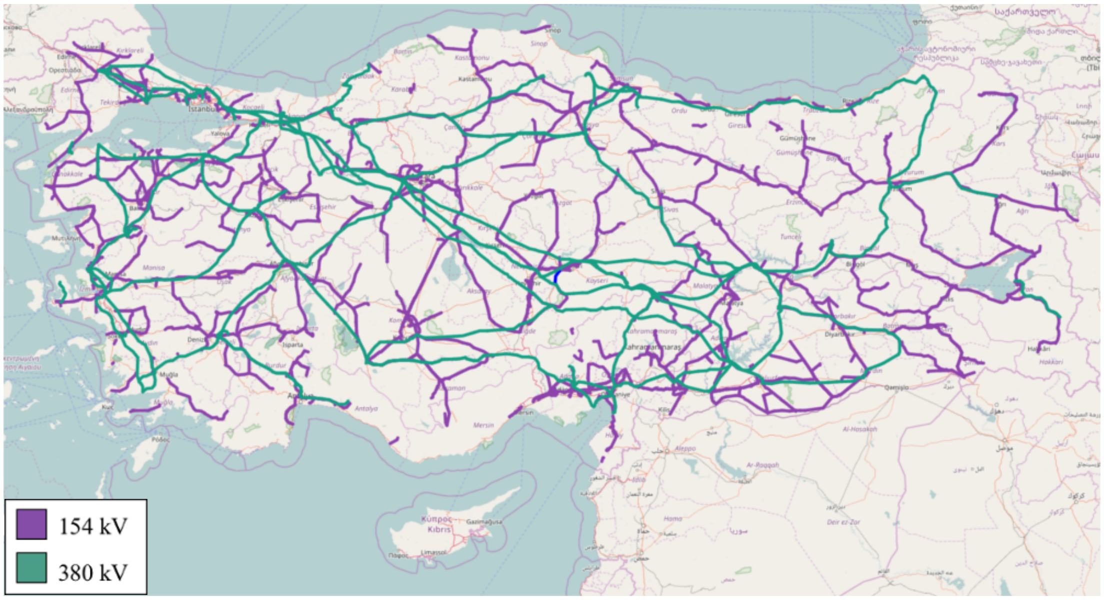
(a) Poland (b) France



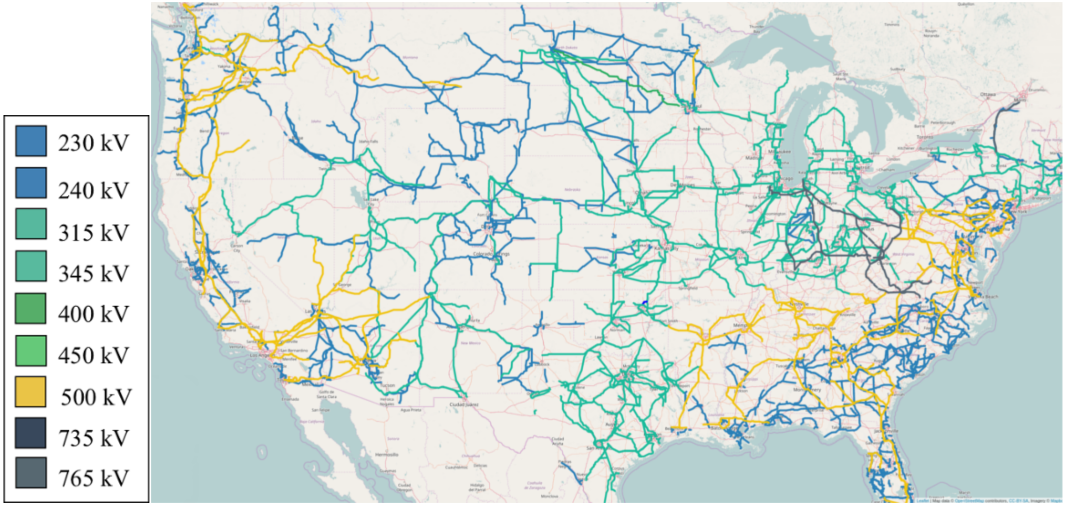
(c) Japan (d) Netherlands

(e) Egypt

Fig. 11: Inferred transmission grids for selected countries.



(a) Turkey



(b) USA

Fig. 12: Inferred transmission grids for selected countries.

Jose Rivera received a Dipl.-Ing. degree in electrical engineering and information technology and the Ph.D. degree in natural sciences from the Technical University of Munich (TUM), Germany, in 2012 and 2017, respectively. Until his graduation, he interned at Siemens Corporate Technologies. He was also a visiting student at the Massachusetts Institute of Technology (MIT) in the Laboratory for Information and Decision Systems (LIDS). His research interests include power systems, data science and distributed optimization.

Pezhman Nasirifard received his B.Sc. in Computer Software Engineering from the University of Tehran, Iran in 2013 and his M.Sc. in Computer Science from the Technical University of Munich (TUM) in 2017. He joined the Chair of Application and Middleware Systems at TUM in August 2017 as a Ph.D. candidate. His research interests include distributed systems, cloud computing and blockchain technologies.

Johannes Leimhofer received his B.Sc. in Information Systems in 2014 and his M.Sc. in Computer Science in 2016 from the Technical University of Munich (TUM). As a researcher at the Chair of Application and Middleware Systems at TUM, he collaborated with the OpenGridMap team, where he worked on inference approaches for generating high voltage transmission grid models from crowdsourced data. Currently, he works on predictive analytics solutions for processing development in the chemical industry, where he is involved in a large-scale digitalization project paving the way for chemistry 4.0.

Hans-Arno Jacobsen's pioneering research lies at the interface between computer science, computer engineering and information systems. He has held numerous patents and was involved in important industrial developments with partners like Bell Canada, Computer Associates, IBM, Yahoo! and Sun Microsystems. His principal areas of research include the design and the development of middleware systems, event processing, service computing and applications in enterprise data processing. His applied research is focused on ICT for energy management and energy efficiency. He also explores the integration of modern hardware components, such as FPGAs (Field Programmable Gate Arrays), into middleware architectures. After studying and completing his doctorate in Germany, France and the USA, Prof. Jacobsen engaged in postdoctoral research at the INRIA near Paris before moving to the University of Toronto in 2001. There, he has been supervising students as a professor in the Department of Electrical and Computer Engineering and the Department of Computer Science. After being awarded the prestigious Alexander von Humboldt Award, he started to engage in research activities at TUM.