

International Journal of Sustainable Energy Planning and Management

Allocation of Annual Electricity Consumption and Power Generation Capacities Across Multiple Voltage Levels in a High Spatial Resolution

Ludwig Hülk^{*}, Lukas Wienholt^b, Ilka Cußmann^c, Ulf Philipp Müller^c, Carsten Matke^b, Editha Kötter^a

^a Reiner Lemoine Institut

^b DLR Institute of Networked Energy Systems

^c Flensburg University of Applied Sciences, Centre for Sustainable Energy Systems

ABSTRACT

The electrical energy system is in transition. There is continuing expansion of renewable generation, which causes a spatial and temporal shift of energy flows at different voltage levels of the power grid. On the consumption side, demand patterns are expected to change. The limited availability of data with high spatial resolution hinders independent and transparent assessments. In view of this, the research project *open_eGo* is developing methods focusing on electricity networks and open-science principles. In this work, electricity demand and power generation are allocated to their corresponding voltage levels and network nodes. By combining data on municipal boundaries with Voronoi cells we created catchment areas for each substation. On the demand side, OpenStreetMap data is used for mapping different demand sectors. We show that a consistent data set can be produced in a high spatial resolution using geographical data processing. Our results apply to Germany but the methodology can be adopted to other countries or regions where sufficient open data is available.

Keywords:

Energy consumption;
Geospatial analysis;
Open data, Power generation;

URL:

dx.doi.org/10.5278/ijsepm.2017.13.6

1. Introduction

Due to pollution and the increasingly visible effects of climate change, various countries are turning from conventional power generation to renewable energy [1]. In Germany, the transformation in power generation technology has already changed the geographical and temporal distribution of electricity production. The increasing share of decentralised power generation challenges the grid [2, 3]. The existing system was not designed to support a large share of power generation in distribution networks. The number of redispatch operations by transmission system operators (TSO) has been increasing in the short term [4, 5], while long-term solutions such as grid extension have been delayed [6, 7].

The research project *open_eGo* is a response to these challenges [8]. It is developing an open-source grid planning toolbox named *eGo*. This toolbox works from extra-high to low voltage levels of the power grid to determine the optimal network and storage topology. In order to properly identify and locate the need for electricity grid extension at different voltage levels, multilayer power flow simulations are performed. At the same time, such simulations could assist in the analysis of the effects of decentralisation. High resolution data sets of energy consumption, power generation and the existing grid infrastructure are compiled and made available.

In addition to status-quo analysis of the German electricity system, future scenarios up to a fully renewable electricity system are considered in the research

^{*} Corresponding author: email address: ludwig.huelk@rl-institut.de (Ludwig Hülk)

project. The scope of this study is the identification of relevant substations between voltage levels and the allocation of demand and generation data to these network nodes. The creation of grid topology models [9] and the generation of time series [10, 11, 12] within the grids are not part of this article and are addressed in [13, 14]. In the following sections, we describe our methods for identifying substations (transmission substations and transition points) and allocate annual electricity consumption and power generation capacities through areas we generate (high voltage grid districts and medium voltage grid districts). The low voltage infrastructure located inside the identified load areas is not considered.

Several publications [15, 16, 17, 18, 19] implemented an allocation of generation and demand. The spatial resolutions differ due to differences in the guiding research questions. The method presented here achieves a higher spatial resolution, as the results must be suitable for calculations in the distribution grid comparable to [20]. A high spatial resolution is necessary when simulating medium and low voltage grid levels. Further, there are more than 880 distribution system operators in Germany [5]; as a result, a uniform digital data set on the medium- and low-voltage grid levels is not available.

Although developing a data set based on publicly available sources covering all voltage levels is challenging, an open-source approach is possible thanks to data from governmental authorities and the public database *OpenStreetMap* (OSM) [21]. Generally, open-source and open data increase transparency in scientific and public discussion and allow reproducible validations and comparisons [22, 23]. We develop methods for generating these data sets based on open-source software and under the strict use and compliance of open data. This makes it possible for others to use and further develop our method, the generated data or just parts thereof. The research questions that we intend to address and answer are:

1. What methods can be developed in order to allocate electricity generation and consumption across multiple voltage levels in a high spatial resolution?
2. What is the added value of the developed methods?
3. Is it possible to conduct such research using only open-source software and open data?

After a brief introduction to the general conventions of our method, we describe the processing of substations

in Section 2.2. In Section 2.3, corresponding grid areas are created for each substation. To identify and locate energy consumption, the creation of the load areas is explained in Section 2.4, followed by the method to allocate installed capacities and sector-wise annual consumption data to each load area in Section 2.5.

2. Methods

2.1. General

In this study, all methods are combined but can also be applied separately if necessary. All steps make use of the following voltage-level classification based on the German three-phase electric power grid, see Figure 1.

The grid is divided by voltage level into an extra-high voltage (EHV), high voltage (HV), medium voltage (MV), and low voltage (LV) grid [24, 25]. The rare case of 60 kV has been assigned to HV since its principal function is regional transmission, similar to 110 kV. In addition, there is a single-phase AC power grid for railway operations (HVrail with 110 kV) which we disregard. The direct transformation from EHV to MV is uncommon in the German electricity grid and is not taken into account.

In *open_eGo*, substations connecting EHV and HV grids are called transmission substations and substations between transmission and distribution level (HV-MV) are defined as (power) transition points of the electricity grid while distribution substations connect the LV grids. The transition points are of central

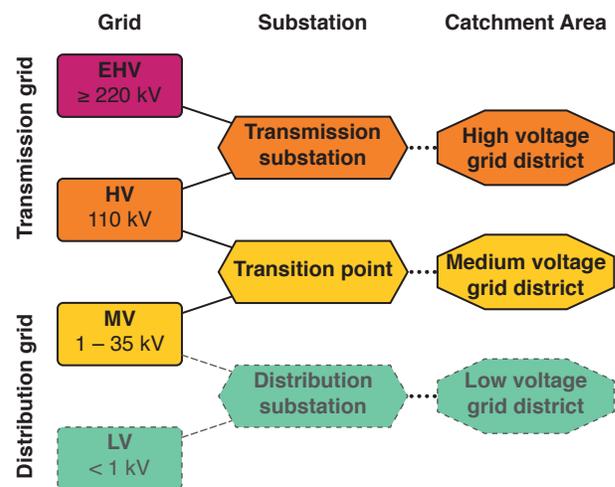


Figure 1: Simplified scheme of voltage levels

Figure by Ludwig Hülk / CC BY SA 4.0

importance to the *eGo*-tool since they are the smallest unit for power flows carried out at the transmission grid levels and at the same time considered as slack nodes from the distribution grid perspective. An area associated to only one transition point and its corresponding medium voltage grid will further be called *grid district*. These catchment areas are used for collecting electrical energy consumption and electrical generation associated to each substation at the different voltage level.

In the *open_eGo* project and in this study in particular, the data from OSM is of great importance because it influences the implementation and the outcome of the power data allocation [9]. *OpenStreetMap* is an open database project which collects and publishes data mapped by its numerous users. Data in OSM is crowd-sourced by conventions instead of strict rules [26]. There are three general data types in OSM: *nodes* (points), *ways* (lines or polygons), and *relations* (structures). Besides geographic information, OSM items can contain information (*tags*) that is stored in the form of *key* and *value* pairs (e.g. voltage = 110,000). The coverage and quality of OSM is inhomogeneous, but constantly improving due to a growing community of commercial and scientific users. Quality assurance tools are applied [27], but datasets derived from OSM can still be incomplete, outdated or might even contain false information [28, 29]. We accounted for this while minimizing the number of assumptions we made based on these data sets.

2.2. Substations

To identify relevant substations, similar to [30], we apply methods of the SciGRID model [31] to derive an OSM-based power dataset for EHV and HV. Apart from the transition points we identify transmission substations in order to allocate industrial consumers in Section 2.5. The complete process starting from the raw data set and ending with valid substation data is visualised in Table 1.

Due to the inhomogeneous data quality in OSM (see Section 2.1), we cannot only consider substations mapped as a surrounding area by data type *way*, but also have to consider substations mapped by data type *node* and in the case of EHV substations also by data type *relation*. In addition, we assume that substations have missing or incomplete voltage information indicating a connection to the respective voltage level.

In order to find valid transmission substations and transition points, substations are identified that contain one of the most frequently used conventions for substations as tag: *power* = {*substation*, *sub_station*, *station*}. Additionally, substations must have voltage information indicating a connection to the HV or EHV grid. Substations that have been associated to the railway grid (HV_{Rail}) via the frequency or operator tag are not considered in *open_eGo*. Moreover, we investigate rejected substations that have missing voltage information or that indicate no direct reference to a grid. These substations are accepted in this second investigation if a power line is found that starts within the area of the substation.

Table 1: Process of abstraction and filtering from raw OSM data to coherent substations

Step	Assumption	Result
1. Raw OSM data processing	power = *	extracted open power dataset of all voltage levels
2. Filter substations	type <i>node</i> or <i>way</i> or <i>relation</i>	
3. Filter voltage levels	power = substation or sub_station or station for EHV substations: voltage ≥ 220,000 for HV substations: voltage = 110,000 or 60,000 or line starts/ends at substation	substations of all voltage levels EHV substations HV substations HV _{Rail} substations
4. Filter relevant substations	situated within administrative boundary frequency ≠ 16.7 or 16.67 operator ≠ DB_Energie or DB Energie GmbH or DB Netz or DB Netz AG substation ≠ transition or traction	EHV substations HV substations
5. Aggregate substations	aggregate substations that are situated within a distance of 75 m from their boundary	Transmission substations Transition points

In cases where accepted substations are very close to each other ($d < 150$ m), the one with the smaller area is rejected since we assume it belongs to its larger neighbour e.g. they are cut by a street, but belong to the same electrical entity. Aggregating neighbouring substations is, moreover, helpful when assigning loads and generators to the respective grid districts as described in Section 2.5. Finally, all valid substations are abstracted to their geometrical centre, forming the initial set of points for the following analyses.

2.3. Grid districts

Currently, comprehensive and consistent data about the grid infrastructure in Germany is not publicly available. Power lines are contained in the OSM data set but are not complete for medium voltage. Network operators are only obliged to publish aggregated statistics [32], which are not suitable for modelling and representing individual georeferenced grids. Therefore, the approach is to allocate load and generation to the electrical grid and each substation using geographic catchment areas. An existing approach is the assignment to the closest point via a Voronoi partition [19, 33]. For the grid districts, we expand this mathematical approach based on the substation locations from Section 2.2 with existing administrative boundaries. We assume that in the past, concessions for the construction of MV electricity networks have often been awarded on the municipal or district level. Boundary lines appear to be more natural than straight lines from the Voronoi method. In reality, villages and settlements are predominantly located entirely inside one municipality and thus intersections between load areas will be few. The outcomes are three polygon layers: Two with only Voronoi cells for transmission substations (HV grid districts) and transition points as well as one improved layer with a combination of the Voronoi cells and municipal boundaries (medium voltage grid districts).

Municipalities are the lowest level of administrative division and therefore the smallest official territorial unit in Germany [34] for which information is publicly available. Another benefit is the availability of matching statistical data at the municipal level. In the official data, municipalities are given either as single areas (polygon) or as areas consisting of multiple parts (multipolygon) with additional exclaves or islands. The total number of the municipalities and the course of the boundaries have changed regularly in the past decades, resulting in growth of shapes with many bulges and fragments [35].

First, all multipolygons are split in individual parts in which three different types are distinguished: a) islands located in the sea or in a different country b) fragments bordering other municipal boundaries c) fragments completely within other polygons (enclaves), see Figure 2. Fragments from a) and b) are treated independently while all enclaves are removed and assigned to the surrounding municipality.

The municipality polygons resulting from this first step are related to the transition points from Section 2.2. By counting the number of substations in each polygon they are assorted in three types: 1) single point 2) multiple points and 3) no point. Each group is processed differently, as described in the following passage, and merged afterwards.

The first group comprises all polygons with exactly one substation per municipality. In this particular case, the entire area is set as grid district and no additional calculations are required.

The second group contains all municipalities with more than one substation. The area is supplied by multiple stations and therefore the polygon has to be divided. For this purpose, a Voronoi diagram is created from all transition points. This common mathematical operation with Euclidean distance is used to generate a set of Voronoi cells corresponding to each point containing the nearest area [36, 37, 38]. Van Leeuwen et al. applied this method in a similar context [33]. All processed municipality areas are cut with this Voronoi layer, resulting in divided polygons for each transition point and additional fragments without points. Accordingly, the marginal areas are joined with the resulting grid districts inside the same original municipality using a next neighbour algorithm.

The third group consists of all polygons from the processed municipality layer which do not contain any substations. Again, a next neighbour algorithm is used to find the shortest distance to the nearest transition point.

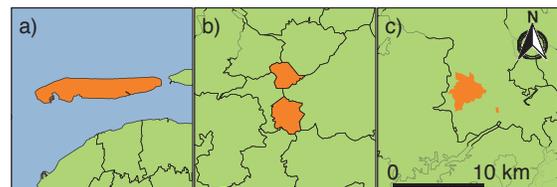


Figure 2: Types of municipality exclaves (orange):

a) Islands b) Exclaves c) Enclaves

© GeoBasis-DE I BKG 2016 (Data changed)

Figure by Ludwig Hülk / CC BY SA 4.0

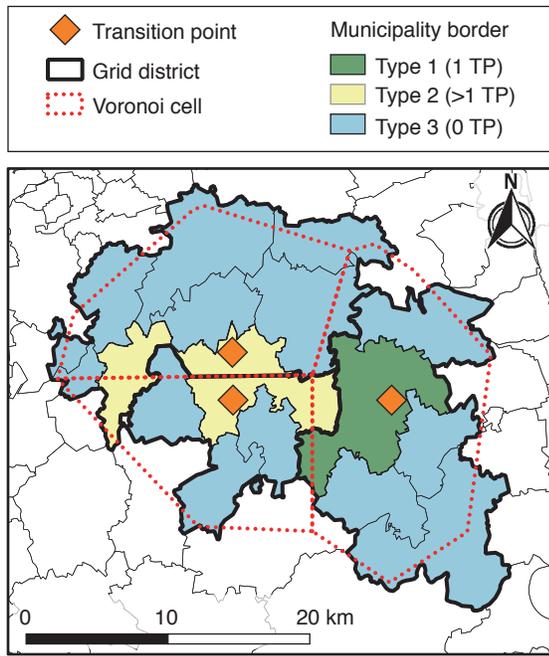


Figure 3: Example of the grid district processing
 © GeoBasis-DE I BKG 2016 (Data changed)
 © OpenStreetMap contributors.
 Figure by Ludwig Hülka / CC BY SA 4.0

After the three categories are processed separately, the resulting polygons are combined. Polygons that are assigned to the same substation are joined, resulting in one grid district for each transition point, see Figure 3. The grid districts consist of all possible combinations from polygons from the three categories. Some grid districts are defined by the boundaries of one or multiple municipalities while others are defined partly by municipal boundaries and partly by straight cuts from the Voronoi polygons. In addition to this processed layer for the medium voltage level, the original Voronoi is made for each of them. We apply this approach on the EHV level due to its simplicity, its use in similar studies (cf. [33]) and the assumption that on the EHV level, administrative boundaries are less relevant to the network connection of load and generation than the distance to the nearest transmission substation.

2.4. Load areas

Due to the high spatial resolution used in *open_eGo*, the aim is to locate all areas with an electrical energy consumption connected to the electrical grid. While generating capacities are represented by a manageable number of single points, the energy consumption

consists of far more points and is therefore abstracted as an area. We define these resulting polygons as *load areas*. This is based on the assumption that the data indicates areas with specific structures or buildings which can be regarded as electrified and therefore connected to the electrical grid [39]. All points inside the load areas are connected via LV grids.

The main input data for the generation of load areas are the OSM land use data [40], OSM data on industrial infrastructure and high resolution population data based on the 2011 census for Germany [41]. As described in Table 2, the generation of load areas is a concatenation and coalescence of different geometries. Specific OSM tags are selected and assigned to one of the four electricity demand sectors: residential, retail, industrial and agricultural [42]. Published electricity consumption statistics also use this classification and by matching the sectors they can be assigned more precisely. The composition of the load areas becomes more diverse and leads to different load profiles and peak-loads.

First, the extensive OSM polygon data set for Germany is filtered for the previously mentioned tags. Because some objects close to the borders are outside the country, crossing polygons are blended with the official national border. The resulting OSM polygons are then processed in two ways: They are combined to form a load area and separated with the respective electricity demand sector to allocate the consumption for each sector more accurately. The geometrical procedure will be described for the final load areas because the approach is the same for the four sectors.

Industrial areas whose allocated demand exceeds a certain limit are identified as large industrial consumers. We assume they are directly connected to the transmission substation. The methodology to allocate the electricity demand is described in detail in Section 2.5.

As observed in the data, different land-use areas overlap or have minor gaps. To combine and unite these polygons, a buffer of 100 m and a merger function is

Table 2: OSM tags assigned to demand sectors [40]

Sector	Tag (key = value)
residential	landuse = residential
retail	landuse = commercial or retail
industrial	landuse = industrial or port aeroway = terminal or gate man_made = works or wastewater_plant
agricultural	landuse = farmyard or greenhouse_horticulture

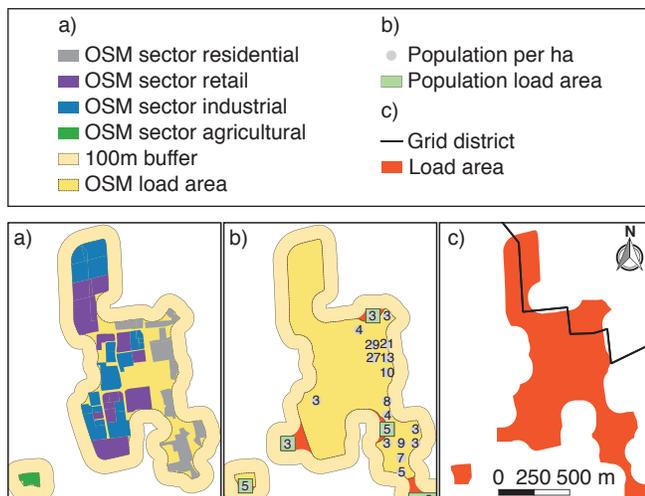


Figure 4: Load area processing: a) OSM landuse b) Census population c) Grid district combination

© GeoBasis-DE / BKG 2016 (Data changed) © Statistisches Bundesamt, Wiesbaden; Datenlizenz by-2-0; (Data changed)

© OpenStreetMap contributors. Figure by Ludwig Hülk / CC BY SA 4.0

applied to all areas. By reverting this subsequently with a negative 100 m buffer, the output geometries can consist of multiple OSM polygons that are jointed and smoothed. All overlaps and gaps of up to 200 m in size are eliminated resulting in OSM load area, see Figure 4a).

A similar method is applied to the population dataset, see Figure 4b). The population is taken from the population grid based on the German population census of 2011 with a high spatial resolution of 1 ha. To reduce calculation time, only lattice points that have a population value and are not inside the previously generated OSM loads are considered. This is based on the assumption that people are living in areas which are not yet documented in OSM. These are called census loads and are melted with the OSM loads to obtain melted loads, see Figure 4c). The results are polygons that cover the selected OSM demand sectors and all registered 80.3 million inhabitants in 2011 [43]. There is a growth of 1.65 million inhabitants from the 2011 census to 2015 which is not taken into account in this work.

In order to precisely assign the loads to the transition points, the melted loads are cut with the processed medium voltage grid districts to receive distinct load areas. In the case of an intersection, the area is divided into multiple load areas. Each load area is assigned to its substation and is entirely within the corresponding grid

district. This step is also performed on the four aforementioned OSM sector layers, see Figure 4a) to match the data. In a last step, additional parameters are determined for each load area: area and perimeter, respective shares per sector and the number of inhabitants per load area. This data is then used to assign specific consumption and can even be used to generate low voltage grids.

2.5. Allocation of generation and demand

To perform power flow simulations, comprehensive knowledge of the spatial distribution of generation and demand is essential. Therefore, the grid topology as well as data on generation and demand need to be linked. In *open_eGo*, the annual electricity consumption is assigned to the identified load areas by using several data sources and methodologies. We derive data on annual German electricity demand from publications by the German Federal Statistical Office providing information on the sectoral electricity demand per federal state [44]. The methodology for a spatial allocation of demand varies according to the specific sector. The annual electricity consumption of households is allocated based on the distribution of population assuming a direct correlation between the number of inhabitants and the electricity consumption in the reference area. This assumption is supported by several publications [45, 18, 46, 15, 19].

For the industrial and retail sectors, a correlation between electricity consumption and gross value added (GVA) is supposed, which follows similar assumptions made in [16, 33, 47]. GVA data is provided for different sectors on the level of administrative districts (German: *Landkreise*) [48]. Consistent with [49, 33], the annual consumption of the industrial and retail sectors is broken down to the level of administrative districts using the GVA. In contrast to other approaches, we further increase the spatial resolution by allocating the annual consumption not only to administrative districts, but also to every load area within the grid districts. The basis for this allocation is the respective share of industry or retail as defined in Section 2.4.

Active power generators are taken from official and publicly available registers such as the power plant registry published by the German Federal Network Agency (*Bundesnetzagentur*) [50] and a renewable energy system (RES) registry published by a solar industry trade group [51]. These registries provide information on power plants with different degrees of

accuracy. Data gaps in the voltage level are filled on the basis of the installed capacity applying Table 3.

With information provided on the voltage level of generators, we are able not only to allocate generators spatially, but also to a specific grid level. Generators with a voltage level of 110 kV or lower are assigned to transition points with voltage tags of 110 kV by performing a spatial comparison between the generator location and the medium voltage grid districts. Generators with voltage levels over 110 kV are allocated to the transmission substations using the high voltage grid districts.

For electricity demand, we assume that every residential, retail or agricultural load is allocated to transition points, while some parts of the industrial sector are assumed to have a direct connection to a transmission substation. Large industrial consumers are identified using a threshold of 130 GWh/a annual electricity consumption and allocated to the high voltage level, see Figure 5. This value is based on a synthetic time series for industry consumption with a peak load of approximately 0.132 kW per 1000 kWh/a annual consumption based on [54] and assumptions taken from [52, 53]. The high voltage grid districts from Section 2.3 are used for allocation.

Table 3: Installed capacity assigned to voltage level based on [52, 53]

Voltage level	Inst. Capacity	Allocation
1 EHV	> 120 MW	Transmission substation
3 HV	17.5 – 120 MW	HV Grid
4 HV-MV	4.5 – 17.5 MW	Transition point
5 MV	0.3 – 4.5 MW	MV Grid
6 MV-LV	0.1 – 0.3 MW	Distribution substation
7 LV	< 0.1 MW	Load area

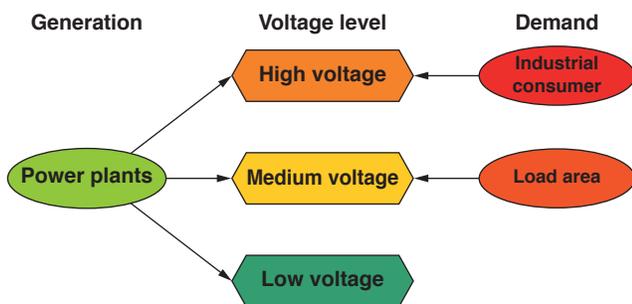


Figure 5: Simplified scheme of data allocation
Figure by Ludwig Hülka / CC BY SA 4.0

This results in a data set where every load and electrical generator is assigned to one specific substation, which provides a basis for other calculations such as power flow simulations. The creation of time series is not treated in this paper.

3. Results and Validation

Results obtained applying the methods described in Section 2 derive from the OSM data set dated 1 October 2016. Furthermore, the results are validated where other data sets are available.

3.1. Substations

After the first step, 4,078 HV substations are identified, which consist of 23 nodes, 2,845 ways and 1,210 substations that are identified via 110 kV lines starting or ending within a substation area. Starting from these 4,078 substations, 361 substations were eliminated during the subsequent process of filtering irrelevant items or those belonging to the railway grid. At the final aggregation stage, another 111 substations are discarded in favour of neighbouring transition points with a larger area. Thus, the final data set consists of 3,606 transition points.

In order to validate the resulting transition points, we compare the total number of transition points with a similar study by van Leeuwen [33] and carry out regional comparisons with data provided by DSOs. Generally, the number of transition points at the HV level is consistent with the findings of others. For instance, van Leeuwen [33] states that there are around 4,000 substations of HV and EHV in Germany. Since we can disregard 119 sole EHV substations and aggregate several transition points, the numbers can be considered to be in the same range.

Next, we compare our number of transition points with real data for particular regions in Germany provided by the respective DSOs. We focus on three DSO-districts, one of them in Southern Germany [55], the others in Northwestern [56] and Northeastern [57] Germany. Within these areas, we compare the number of substations given by the DSO on their maps with the number of transition points found by the method described in Section 2.2. The results in Table 4 indicate that the method is applicable. In the Northern German regions, the number of transition points exceeds the number of substations given by the DSO due to the fact that substations not connected to the MV grid – such as

Table 4: Comparison of transition points with DSO-data

Region	Substations by DSO	Transition Points	%
S	118	116	- 1.7
NW	122	136	+ 11.5
NE	31	36	+ 16

EHV power-plant connections – are not part of the DSO-data.

Apart from the transition points, we also abstract transmission substations relevant for the allocation of industrial consumers. The abstraction and filtering is applied to the HV transition points analogously. The final data set consists of 435 EHV transmission substations, which are however not disjunct to the HV transition points in cases where substations contain EHV and HV voltage tags.

3.2. Grid districts

The number of grid districts is 3,606 and thus identical to the number of transition points, see Figure 6. The entire grid district area (357,188 km²) is consistent with the official land area of Germany given by the Federal Statistical Office [58]. This comparison shows that the entire area of Germany is considered and can be assigned to transition points. In order to understand the composition and results of the grid districts, the municipal boundaries are examined first. There are 12,353 individual polygons in the 11,431 municipalities, showing a total number of 922 exclaves. In Section 2.3 and Figure 2, three different categories of exclaves are distinguished. While islands and normal exclaves (a and b) with 342 polygons are treated as independent areas, the 580 enclaves (c) are assigned to the surrounding area resulting in 12,011 polygons.

The combination of this data set with the transition points results in 1,717 areas with exactly one transition point (type 1), 9,746 areas without a transition point (type 3) and 548 areas with more than one transition point (type 2). Type 3 contains 1,889 transition points. After combining the area types, there are grid districts with four different possible compositions of these three types. Counts and mean areas of each composition are shown in Table 5. A combination of the types 1 and 2 is not possible because they are both assigned to one specific transition point."

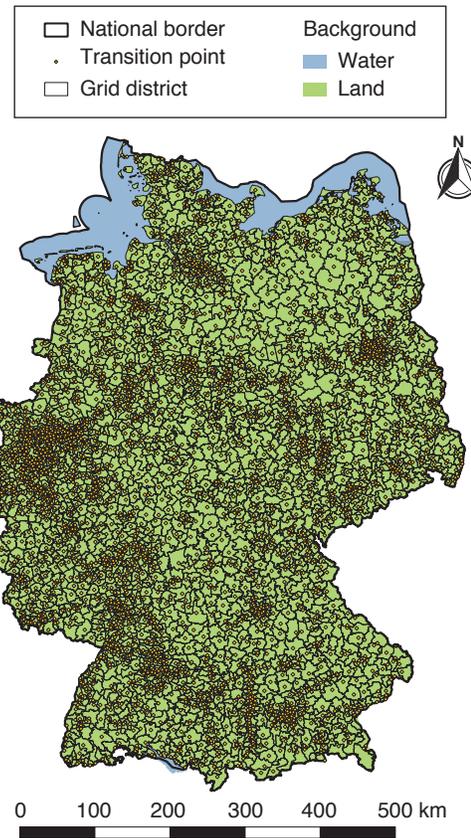


Figure 6: Transition points and grid districts in Germany
 © GeoBasis-DE / BKG 2016 (Data changed)
 © OpenStreetMap contributors.
 Figure by Ludwig Hülk / CC BY SA 4.0

Table 5: Number and average size of grid district types[1 (1 transition point (TP)); 2 (>1 TP); 3 (0 TP)]

Types	Count	Mean area
1 + 3	1,374	163 km ²
2 + 3	693	104 km ²
1	343	78 km ²
2	1,196	28 km ²
total	3,606	99 km ²

In rural areas, medium voltage lines can be up to 30 km in length [59] [60]. For the computed grid districts to have a size similar to real MV grids, the maximum distance from the substation to the consumer should not be greater than 30 km. In our results, only one grid district could be found where the distance between transition point and grid district outline exceeded 30 km. The largest distance between a transition point and its corresponding load area is 24 km, so the geographical

extents of the computed grid districts are consistent with typical supply areas of real MV grids.

3.3. Load areas

There are 208,486 load areas with a total area of 43,503 km². The size and appearance of the load areas vary. The majority (67%) is smaller than 5 ha while 0.5% is larger than 500 ha comprising 33% of the total area. The computed grid districts and load areas can be used as inputs to model distribution grids as applied in [14]. Fragmentation of load areas should be avoided, particularly in rural areas, because a connection of a village or town to multiple substations is rarely made in practice.

We quantify to which extent the grid districts and the Voronoi polygons separate load areas by evaluating the length of their cutting edges. By applying the grid districts, the cutting length is reduced from 9,350 km (Voronoi) to 6,797 km (grid district) (a reduction of 27%). The total number of cut load areas is reduced by 1,804 (from 19,530 to 17,726). In relation to the total number, in every other grid district one load area is spared.

3.4. Allocation of consumption and generation

In 2011, electricity consumption in Germany was 506 TWh, of which 45.4% was for industrial use [44]. Households and retail accounted for 26.6% and 28%, respectively. We identified 295 industrial large scale consumers directly connected to the EHV grid and consuming 26.6 TWh of electricity annually, or 16% of the total industrial consumption.

Figure 9a) shows the area-specific electricity consumption by grid district. This data does not include the annual electricity demand of large scale industrial consumers as these are directly connected to the transmission substations. There is higher specific electricity consumption per hectare in the urban centres of Germany due to higher population density and more industrial and retail consumers.

Renewable and conventional power plants obtained from the registries [50] and [51] were assigned to the appropriate grid level and the corresponding substation. The total installed capacity obtained from the registries was 210 GW, 94 GW of which are from renewable power plants. Figure 9b) visualizes the installed capacity of renewable and conventional power plants connected at ≤ 110 kV. Of all renewable power plants, over 90% are connected to the lower grid

levels, compared with just 30% of the conventional power plants.

The correctness of the grid districts affects an accurate connection of generation and demand to the grid model. The lack of official and accessible grid data makes an overall evaluation difficult. Our validation focuses on the Northwestern DSO-region (see Table 4) representing 122 DSO substations with their corresponding supply areas. By vectorisation of the DSO's grid map [56] 119 substations can be spatially linked to corresponding transition points and are used for comparison. We compare three parameters: grid district area, installed generation capacity and annual electricity consumption. We focus on the spatial distribution of generation and demand, as these are crucial to the quality of the grid model. The distribution within grid districts and Voronoi areas is compared to that found in real DSO supply areas.

Almost 40% of the modelled catchment areas estimate the distribution of generation of the benchmark with an accuracy of $\pm 10\%$ normalised by the average installed power of the DSO's grids (see histogram in Figure 8). The modelled distributions tend to underestimate the real distributions. The Voronoi estimate is less skewed towards underestimation. The interquartile range of both estimates is approximately 48% ranging from the lower quartile of -39% to a upper quartile of 9% (in case of the grid districts) (see boxplot in Figure 8). The sample shows outliers, such as 16% of the grid districts under- or overestimate the benchmark by more than $\pm 110\%$. The Voronoi cell estimation generates fewer outliers (11%).

In contrast to the estimation of generation, the estimation of the consumption tends to overestimate the distribution based on the real DSO grid areas (see boxplot in Figure 7). Once again, the grid districts have a more skewed distribution than the Voronoi cells. The interquartile range is broader at 62% (grid districts) respectively 56% (Voronoi), while 36% of the grid districts meet the benchmark with an accuracy of $\pm 10\%$ normalised by the average annual electricity consumption of the DSO's grids. Fewer grids of the Voronoi model (4% less than grid district distribution) meet this prime category (see histogram in Figure 7). The occurrence of outliers of the two models is similar. An overestimation of more than 110% apply for 13% of the modelled grids.

These validation results apply to a sample of 119 transition points representing 3.3% of the total

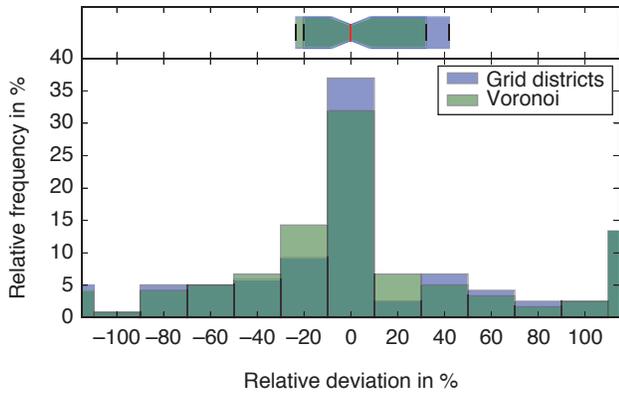


Figure 7: Deviation of electricity consumption for the grid district and Voronoi distribution compared to the DSO distribution normalised by its mean (49 GWh/a). *Bottom*: Histogram (the very left and very right bins have been trimmed for better visualisation and contain all remaining values). *Top*: Boxplot displaying median and quartiles
Figure by Ulf Philipp Müller / CC BY SA 4.0

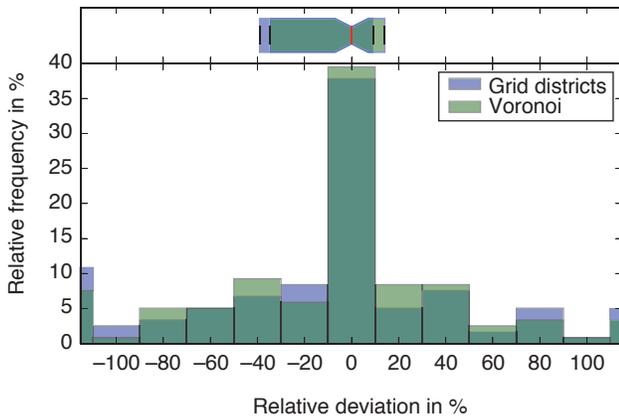


Figure 8: Deviation of installed power of generation facilities for the grid district and Voronoi distribution compared to the DSO distribution normalised by its mean (39 MW). *Bottom*: Histogram (the very left and very right bins have been trimmed for better visualisation and contain all remaining values). *Top*: Boxplot displaying median and quartiles
Figure by Ulf Philipp Müller / CC BY SA 4.0

amount of transition points. It is expected that the analysis would evoke different results in other areas due to regional variation of structural grid characteristics. Thus, the validation results may not necessarily be representative.

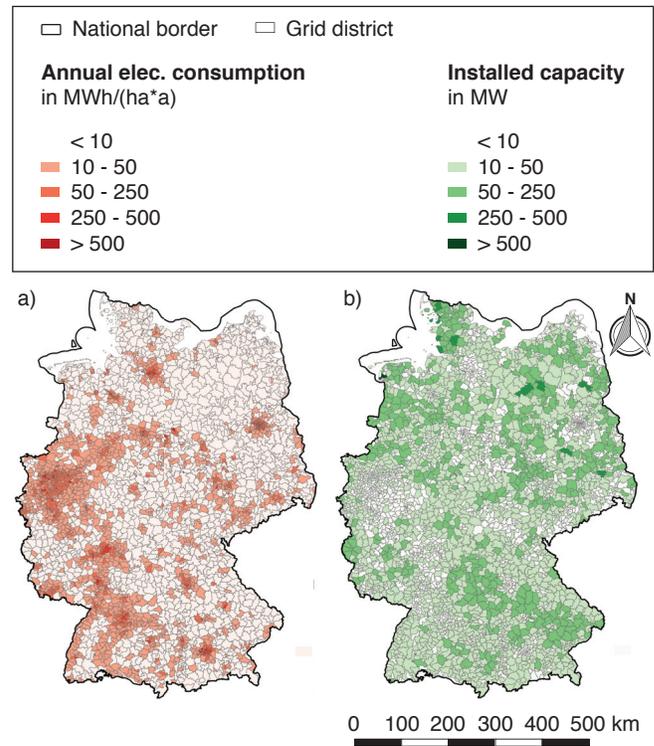


Figure 9: Electricity consumption and installed capacity per grid district
© GeoBasis-DE / BKG 2016 (Data changed)
© OpenStreetMap contributors.
Figure by Ilka Cußmann / CC BY SA 4.0

4. Discussion and Conclusion

Our approach generates a characteristic and consistent data set of the German energy system through allocation of electricity demand and power generation. The primary outputs are substations, their catchment areas (grid districts) as well as the respective annual electricity demand and generation capacity. The main goal of this work is to find a coherent combination of methods to locate and assign data to the different voltage levels.

We identified 3,606 electrical substations between HV and MV levels (transition points) and 435 substations connecting EHV and HV levels (transmission substations). The results we generated are consistent with data published by DSOs, which supports our applied methodology. Single deviations which occurred can be explained and do not undermine the applicability of the approach.

For each substation we define one corresponding catchment area. The high voltage grid districts are

created using the well-known Voronoi algorithm while the method to generate the downstream medium voltage grid districts includes and merges municipal boundaries and Voronoi polygons.

We identified 208,486 areas with electrical demand connected to the MV grid. Their total area corresponds to the land use for housing and transport in Germany. The annual electricity consumption in Germany of 479.4 TWh/a is distributed to these load areas. In the EHV grid, the electricity consumption of 26.6 TWh/a is divided into 295 large industrial consumers.

The total installed capacity of the allocated power plants is 210 GW including 94 GW of renewable power plants and is assigned to the respective voltage levels and substations. This assignment was conducted for 879 conventional and around 1.6 million renewable power plants. The accuracy of this allocation is highly influenced by the data quality provided in the power plant registries. Incorrect or implausible records should be identified and rectified in the future to ensure the consistency of the created data set.

As addressed in the first research question, we developed and combined methods to allocate demand and generation for multiple voltage levels. This work creates the base data in a high geographical resolution to analyse the German distribution and transmission grid. The investigation shows that an allocation to certain voltage levels using Voronoi cells or the presented grid districts is applicable. As a validation, the quantities assigned to the transition points are analysed using sample values from a DSO. The comparison shows a small deviation in favour of the Voronoi cells.

With subsequent work, the resulting data can be used for power flow simulations. Firstly, the annual electricity consumption and installed generation capacities have to be transformed into hourly time series. Secondly, grid topologies connecting the transition points are necessary to carry out power flow simulations. These subsequent works are not considered at this point, but directly use the results presented here.

The grid district method aims at improving the data collection with respect to the requirements of the MV grid modelling. One of them is to avoid multiple connections of load areas to the MV grid. By applying the grid districts, the number of affected load areas decreased and the overall cutting length is reduced. Another requirement is the geographical extent and the maximal length of MV lines. A verification confirms that all grid districts are compliant. As stated in the

second research question, our methodology generates added value in terms of the spatial resolution and a more accurate allocation to the individual voltage levels and specific substations.

The collection of methods used in this paper depends heavily on data provided by OSM, so the accuracy of the results is necessarily limited by the integrity of the OSM data. In order to compensate this weakness, other land use data can be included in future work. This also applies to the localisation and assignment of the consumption sectors. A further increase in accuracy might be achieved by taking additional methods into account, such as regional input-output models. Nevertheless this would require adaptations in the applied methods and increase their complexity. Such adaptations are not expected with the focus given in this manuscript.

Due to the chosen open licenses, the methods and results can be used for other applications and further developments.

As suggested in the third research question, this work exclusively processes open data and uses open-source software [61, 62]. The source code (mainly SQL scripts with *PostGIS* [63] functions) is published under an open license (*AGPL*) [64] and can be accessed at a public repository [65]. The processed and resulted data is stored in the *OpenEnergy Database (oedb)* [66], mainly under *ODbL* [67]. The oedb is a *PostgreSQL* database, developed within the research project *open_eGo*.

Acknowledgement

The authors thank the *German Federal Ministry for Economic Affairs and Energy* for funding the research project *open_eGo*. Funding code: 0325881A-E

References

- [1] International Energy Agency, World Energy Outlook 2016 (2016). URL <http://www.worldenergyoutlook.org/publications/weo-2016/>
- [2] K. Schaber, F. Steinke, T. Hamacher, Transmission grid extensions for the integration of variable renewable energies in Europe: Who benefits where?, *Energy Policy* 43 (2012) 123–135. doi:10.1016/j.enpol.2011.12.040.
- [3] M. Bartels, C. Gatzert, M. Peek, W. Schulz, R. Wissen, A. Jansen, J. P. Molly, B. Neddermann, H. P. Gerch, E. Grebe, Y. Saßnick, W. Winter, Planning of the grid integration of wind energy in Germany onshore and offshore up to the year 2020,

- International Journal of Global Energy Issues 25 (3/4) (2006) 257. doi:10.1504/ijgei.2006.008995.
- [4] F. Hinz, D. Möst, The effects of reactive power provision from the distribution grid on redispatch cost, in: 2016 13th International Conference on the European Energy Market (EEM), IEEE, 2016, pp. 1–6. doi:10.1109/EEM.2016.7521274.
- [5] Bundesnetzagentur, Monitoringbericht 2016. URL https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/DatenaustauschUndMonitoring/Monitoring/Monitoringbericht2016.pdf?jsessionid=63005C9469CA815143063966F11BCA67?blob=publicationFile&v=2
- [6] A. Steinbach, Barriers and solutions for expansion of electricity grids—the german experience, *Energy Policy* 63 (2013) 224–229. doi:10.1016/j.enpol.2013.08.073.
- [7] Bundesnetzagentur, Netzausbau – Vorhaben aus dem Energieleitungsausbaugesetz (2016). URL https://www.netzausbau.de/leitungsvorhaben/de.html?cms_map=3
- [8] open eGo Project, accessed 23 August 2017.
- [9] W. Medjroubi, U. P. Müller, M. Scharf, C. Matke, D. Kleinhans, Open data in power grid modelling: New approaches towards transparent grid models, *Energy Reports* 3 (2017) 14–21. doi:10.1016/j.egy.2016.12.001.
- [10] S. Pfenninger, I. Staffell, Long-term patterns of european pv output using 30 years of validated hourly reanalysis and satellite data, *Energy* 114 (2016) 1251 – 1265. doi:10.1016/j.energy.2016.08.060.
- [11] I. Staffell, S. Pfenninger, Using bias-corrected reanalysis to simulate current and future wind power output, *Energy* 114 (2016) 1224 – 1239. doi:10.1016/j.energy.2016.08.068.
- [12] T. Boßmann, I. Staffell, The shape of future electricity demand: Exploring load curves in 2050s germany and britain, *Energy* 90 (2015) 1317 – 1333. doi:http://dx.doi.org/10.1016/j.energy.2015.06.082.
- [13] U. P. Müller, L. Wienholt, D. Kleinhans, I. Cußmann, W.-D. Bunke, G. Pleßmann, J. Wendiggensen, The eGo grid model: An open source approach towards a model of German high and extra-high voltage power grids, in: SciGRID International Conference on Power Grid Modelling, 2017, submitted.
- [14] J. Amme, G. Pleßmann, J. Bühler, L. Hülk, E. Kötter, P. Schwaegerl, The eGo grid model: An open-source and open-data based synthetic medium-voltage grid model for distribution power supply systems, in: SciGRID International Conference on Power Grid Modelling, 2017, submitted.
- [15] C. Rathke, Entwicklung eines Modells für die integrierte Simulation der europäischen Übertragungsnetze und Strommärkte, Dissertation, Gottfried Wilhelm Leibniz Universität Hannover (2013).
- [16] T. Rendel, Erweiterung und Plausibilisierung eines Modells für die integrierte Simulation des europäischen Verbundnetzes und Strommarktes, Ph.D. thesis, Gottfried Wilhelm Leibniz Universität Hannover (2015).
- [17] A. Singh, P. Eser, N. Chokani, R. Abhari, High resolution modeling of the impacts of exogenous factors on power systems—case study of germany, *Energies* 8 (12) (2015) 14168–14181. doi:10.3390/en81212424.
- [18] R. Mackensen, Herausforderungen und Lösungen für eine regenerative Elektrizitätsversorgung Deutschlands, Dissertation, Universität Kassel (2011).
- [19] U. P. Müller, I. Cusmann, C. Wingenbach, J. Wendiggensen, AC Power Flow Simulations within an Open Data Model of a High Voltage Grid, in: V. Wohlgemuth, F. Fuchs-Kittowski, J. Wittmann (Eds.), *Advances and New Trends in Environmental Informatics*, Springer International Publishing Switzerland, 2016, pp. 181–191.
- [20] S. Köppl, F. Böing, C. Pellingner, Modeling of the transmission grid using geo allocation and generalized processes, in: *Trends in Mathematics*, Springer International Publishing, 2017, pp. 189–204. doi:10.1007/978-3-319-51795-7_12.
- [21] OpenStreetMap contributors, OpenStreetMap, accessed 23 August 2017.
- [22] F. Wiese, G. Bökenkamp, C. Wingenbach, O. Hohmeyer, An open source energy system simulation model as an instrument for public participation in the development of strategies for a sustainable future, *Wiley Interdisciplinary Reviews: Energy and Environment* 3 (5) (2014) 490–504. doi:10.1002/wene.109.
- [23] G. Schlömer, C. Blaufuß, L. Hofmann, Modelling of low-voltage grids with the help of open data, in: D. Schulz (Ed.), *NEIS Conference 2016: Nachhaltige Energieversorgung und Integration von Speichern*, Springer Fachmedien Wiesbaden, Wiesbaden, 2017, pp. 21–25. doi:10.1007/978-3-658-15029-7_4.
- [24] Bundesnetzagentur, Datendefinitionen zu Anlage 1 der Entscheidung über die Datenerhebung für das Vergleichsverfahren und die Anreizregulierung bei den Netzbetreibern der Elektrizitätswirtschaft. URL <http://docplayer.org/30756485-Datendefinitionen-zu-anlage-1.html>
- [25] X. Tao, Automatisierte Grundsatzplanung von Mittelspannungsnetzen, Ph.D. thesis, RWTH Aachen (2007).
- [26] O. W. contributors, Editing Standards and Conventions, accessed 23 August 2017.
- [27] OpenStreetMap Wiki contributors, Quality assurance, accessed 23 August 2017.
- [28] P. Neis, D. Zielstra, A. Zipf, Comparison of volunteered geographic information data contributions and community development for selected world regions, *Future Internet* 5 (2) (2013) 282–300. doi:10.3390/fi5020282.

- [29] H. Senaratne, A. Mobasher, A. L. Ali, C. Capineri, M. M. Haklay, A review of volunteered geographic information quality assessment methods, *International Journal of Geographical Information Science* 31 (1) (2016) 139–167. doi: 10.1080/13658816.2016.1189556.
- [30] C. Matke, W. Medjroubi, D. Kleinhans, S. Sager, Structure analysis of the German transmission network using the open source model SciGRID, in: *Advances in Energy System Optimization, Trends in Mathematics*, Birkhäuser Basel, 2017, pp. 177–188. doi:10.1007/978-3-319-51795-7_11.
- [31] C. Matke, W. Medjroubi, D. Kleinhans, SciGRID – An Open Source Reference Model for the European Transmission Network (v0.2) (Jul. 2016). URL <http://www.scigrd.de>
- [32] Bundesministerium der Justiz und für Verbraucherschutz, Verordnung über die Entgelte für den Zugang zu Elektrizitätsversorgungsnetzen (Stromnetzentgeltverordnung – §27) (2005). URL <http://www.gesetze-im-internet.de/bundesrecht/stromnev/gesamt.pdf>
- [33] T. van Leeuwen, S. Dierkes, L. Verheggen, H. Schuster, F. Köhne, A. Moser, Ermittlung von Transitflüssen im Hochspannungsnetz durch mehrere Verknüpfungspunkte mit dem Übertragungsnetz, in: 13. Symposium Energieinnovation, 12.-14.2.2014, Graz/Austria, 2014. URL https://www.tugraz.at/fileadmin/user_upload/Events/Eninnov2014/files/kf/KF_vanLeeuwen.pdf
- [34] Bundesamt für Kartographie und Geodäsie, Administrative areas (Verwaltungsgebiete) VG250 (2016). URL http://www.geodatenzentrum.de/geodaten/gdz_rahmen.gdz_div?gdz_spr=eng&gdz_akt_zeile=5&gdz_anz_zeile=1&gdz_unt_zeile=13&gdz_user_id=0
- [35] Statistische Jahrbücher für die Bundesrepublik Deutschland, The consolidation of municipalities in Germany (2017). URL [https://de.wikipedia.org/wiki/Gemeinde_\(Deutschland\)#Zusammenlegung_von_Gemeinden](https://de.wikipedia.org/wiki/Gemeinde_(Deutschland)#Zusammenlegung_von_Gemeinden)
- [36] A. Okabe, B. Boots, K. Sugihara, S. N. Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, John Wiley & Sons, 2000.
- [37] C. M. Gold, P. R. Remmele, T. Roos, Fully Dynamic and Kinematic Voronoi Diagrams in GIS, 1999.
- [38] P. Dong, Generating and updating multiplicatively weighted voronoi diagrams for point, line and polygon features in GIS, *Computers & Geosciences* 34 (4) (2008) 411–421. doi:10.1016/j.cageo.2007.04.005.
- [39] J. D. Melo, A. Padilha-Feltrin, E. M. Carreno, Data issues in spatial electric load forecasting, in: 2014 IEEE PES General Meeting | Conference & Exposition, IEEE, 2014. doi:10.1109/pesgm.2014.6939848.
- [40] OpenStreetMap, OSM Landuse (2017). URL <http://wiki.openstreetmap.org/wiki/Key:landuse>
- [41] Statistisches Bundesamt (Destatis), Ergebnisse des Zensus 2011 - Bevölkerung im 100 Meter-Gitter, Tech. rep., Statistisches Bundesamt (11.07.2016). URL <https://www.zensus2011.de/SharedDocs/Aktuelles/Ergebnisse/DemografischeGrunddaten.html>
- [42] J. Mei, Y. Goude, G. Hebrail, N. Kong, Spatial estimation of electricity consumption using socio-demographic information, in: 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), IEEE, 2016. doi:10.1109/appeec.2016.7779596.
- [43] Statistisches Bundesamt, Population (2015). URL https://www.destatis.de/EN/FactsFigures/SocietyState/Population/CurrentPopulation/Tables/_Irbbev03.html
- [44] Länderarbeitskreis Energiebilanzen, Energiebilanzen der Länder (2015). URL <http://www.lak-energiebilanzen.de/laenderbilanzen/>
- [45] F. U. Leuthold, Economic Engineering Modeling of Liberalized Electricity Markets: Approaches, Algorithms, and Applications in a European Context, Ph.D. thesis, Technische Universität Dresden (2009).
- [46] M. Robinius, F. ter Stein, A. Schwane, D. Stolten, A top-down spatially resolved electrical load model, *Energies* 10 (3) (2017) 361. doi:10.3390/en10030361.
- [47] F. Andersen, H. Larsen, N. Juul, R. Gaardestrup, Differentiated long term projections of the hourly electricity consumption in local areas. the case of Denmark west, *Applied Energy* 135 (2014) 523–538. doi:10.1016/j.apenergy.2014.08.075.
- [48] Statistisches Bundesamt, VGR der Länder – Bruttoinlandsprodukt / Bruttowertschöpfung in den Kreisen (2016). URL https://www.destatis.de/DE/Publikationen/Thematisch/VolkswirtschaftlicheGesamtrechnungen/AlteAusgaben/VGR_KreisergebnisseBand1Alt.html
- [49] M. Robinius, F. ter Stein, S. Schiebahn, D. Stolten, Lastmodellierung und -visualisierung mittels Geoinformationssystemen, in: 13. Symposium Energieinnovation in Graz/Austria, 2014.
- [50] Bundesnetzagentur, List of Power Plants (11 2015). URL http://www.bundesnetzagentur.de/cIn_1911/DE/Sachgebiete/ElektrizitaetundGas/Unternehmen_Institutionen/Versorgungssicherheit/Erzeugungskapazitaeten/Kraftwerksliste/kraftwerksliste-node.html
- [51] Deutsche Gesellschaft für Sonnenenergie e.V., Die EEG-Anlagen der Region "Bundesrepublik Deutschland" (2014). URL <http://www.energymap.info/energieregionen/DE/105.html>
- [52] P. Konstantin, *Praxisbuch Energiewirtschaft: Energieumwandlung, -transport und -beschaffung im liberalisierten Markt*, Springer Science & Business Media, 2009.

- [53] A.-C. Agricola, B. Höflich, P. Richard, J. Völker, C. Rehtanz, M. Greve, B. Gwisdorf, J. Kays, T. Noll, J. Schwippe, A. Seack, J. Teuwsen, G. Brunekreeft, R. Meyer, V. Liebert, Ausbau- und Innovationsbedarf der Stromverteilnetze in Deutschland bis 2030 (kurz: dena-Verteilnetzstudie): Endbericht. URL https://shop.dena.de/fileadmin/denashop/media/Downloads_Dateien/esd/9100_dena-Verteilnetzstudie_Abschlussbericht.pdf
- [54] B. Schachler, Bewertung des Einsatzes von Kraft-Wärme-Kopplungsanlagen hinsichtlich einem CO₂-Emissionen bei wachsenden Anteil Erneuerbarer Energien, Masterarbeit, Technische Universität Berlin (2014).
- [55] LEW Verteilnetz GmbH, LVN Netzplan (2014). URL <https://www.lew-verteilnetz.de/media/5655/netzplan.pdf>
- [56] EWE Netz GmbH, Stromnetzgebiet der EWE Netz GmbH: Trafobezirke (2016). URL <https://www.ewe-netz.de/einspeiser/mehr/portraet-stromnetz>
- [57] WEMAG Netz GmbH, HS-Netzkarte (2015). URL http://www.wemag-netz.de/export/sites/netz/zentrale_dokumente/HS-Netzkarte_WNG_2015.jpg
- [58] Statistisches Bundesamt, Area and population (2014). URL http://www.statistikportal.de/Statistik-Portal/en/en_jb01_jahrtab1.asp
- [59] Forschungsgemeinschaft für Elektrische Anlagen und Stromwirtschaft e. V., Technischer Bericht 302, Ein Werkzeug zur Optimierung der Störungsbeseitigung für Planung und Betrieb von Mittelspannungsnetzen, Tech. rep., Forschungsgemeinschaft für Elektrische Anlagen und Stromwirtschaft e. V. (2008).
- [60] E. Sissimatos, Technik und Einsatz von hochtemperatursupraleitenden Leistungstransformatoren, Ph.D. thesis, Universität Hannover (2005).
- [61] Quantum GIS Development Team, Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project. (v2.14) (2017). URL <http://qgis.osgeo.org>
- [62] PostgreSQL Global Development Group, PostgreSQL (v9.3), <http://www.postgresql.org> (2017).
- [63] Postgis (v2.1.2) - geographic information systems extensions to postgresql, <http://postgis.net/> (2017).
- [64] Free Software Foundation, GNU Affero General Public License Version 3 (AGPL-3.0), accessed 23 August 2017 (18 November 2016).
- [65] eGo dataprocessing © Flensburg University of Applied Sciences, Centre for Sustainable Energy Systems © Europa-Universität Flensburg, Centre for Sustainable Energy Systems © Reiner Lemoine Institut © DLR Institute for Networked Energy Systems, open eGo DataProcessing (eGoDP), accessed 23 August 2017. [66] OpenEnergy Platform, OpenEnergy DataBase (oedb), accessed 23 August 2017.
- [67] Open Knowledge International, Open Database License v1.0 (ODbL-1.0), accessed 23 August 2017 (December 2007).