

scientific data



OPEN
DATA DESCRIPTOR

Modelling the high-voltage grid using open data for Europe and beyond

Bobby Xiong¹✉, Davide Fioriti¹, Fabian Neumann¹, Iegor Riepin¹ & Tom Brown¹

This paper provides the background, methodology and validation for constructing a representation of the European high-voltage grid (AC lines from 220 to 750 kV and all DC lines) based on OpenStreetMap data. Grid components include commissioned substations, transmission lines and cables, transformers, and converters as well as technical parameters based on standard types. The data is provided as easy-to-access comma-separated values files which makes it suitable for model-independent, large-scale electricity and energy system modelling. For further ease-of-use, an interactive map is included to enable visual inspection. To assess the data quality, this paper compares the dataset with official statistics and representative model runs using PyPSA-Eur based on different electricity grid representations. The dataset and workflow are provided as part of PyPSA-Eur, an open-source, sector-coupled optimisation model of the European energy system. By integrating with the codebase for initiatives such as PyPSA-Earth, the benefits of this work of this work extend to the global context. The dataset is published under the Open Data Commons Open Database (ODbL 1.0) licence.

Background & Summary

Energy system models are indispensable tools in today's world in order to understand the complex interactions between energy sources, technologies, policies, and markets. They are used by researchers, industry and policy makers to enable informed decision-making in the transition to a net-zero energy system. However, conclusions drawn from such models are only as good as the underlying data and assumptions. Especially the representation of existing energy infrastructure, such as the electricity grid, can have a deciding impact on future investments derived from such models¹. While Transmission System Operators (TSOs) have their own information on the high-voltage grid, this data is often not publicly available to the level of detail needed for academic research purposes. Official institutions like the European Network of Transmission System Operators for Electricity (ENTSO-E) provide an online map² of the European high-voltage grid. There are, however, several limitations typical for these sources: (i) there is no underlying, topologically connected dataset, (ii) they are not released under an open licence, (iii) nor are they updated frequently, and (iv) their geographic detail is limited or highly stylised.

There are previous projects that have modelled the European high-voltage grid or its parts based on OSM data. Some institutions provide data for particular regions, however, all of them come with their own drawbacks: While the most trustworthy data comes from TSOs themselves, they are — with few regional exceptions³ — not georeferenced⁴ or do not cover the entirety of Europe. Datasets from previous academic projects^{5–7} are either very complex to reproduce or have not been updated for close to a decade. An overview of notable projects and datasets is listed in Table 6. Proven to be a reliable public data source, we make use of OSM to introduce a transparent workflow in order to create a representation of the European high-voltage grid. On the lack of updates of existing datasets, there are two main advantages of our work compared to previous initiatives. First, our approach uses the OSM Overpass turbo API⁸ that always allows to retrieve the latest OSM data. Second, the active OSM community, along with the large user base of PyPSA-Eur and its integration into automated workflows ensure frequent updates and validation. Debugging can be easily done with the help of the open source project OpenInfraMap⁹ and the interactive map included in our dataset. Both tools render the electricity infrastructure in a JavaScript-based map which can be accessed with any modern browser. Finally, the entire

¹Department of Digital Transformation in Energy Systems, Institute of Energy Engineering, Technische Universität Berlin, Berlin, Germany. ²Department of Energy Systems, Territory and Construction Engineering, Università di Pisa, Pisa, Italy. ✉e-mail: xiong@tu-berlin.de



Fig. 1 Process diagram depicting the creation of the European high-voltage grid from OSM data, implemented through individual Snakemake rules.

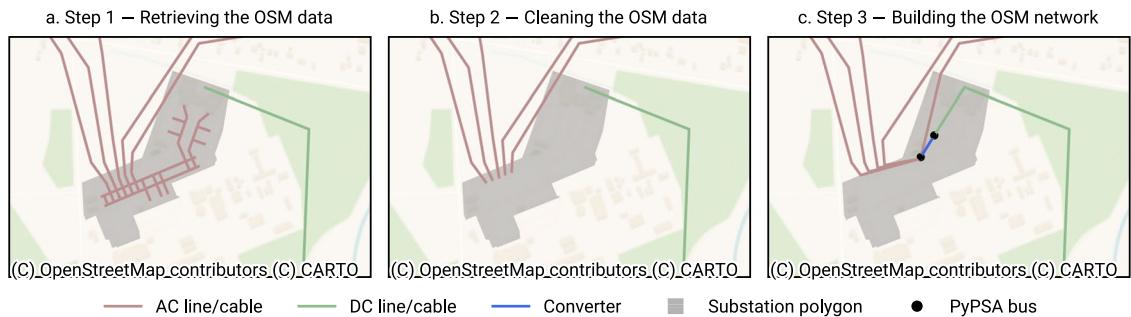


Fig. 2 Illustration of the steps to create a PyPSA-ready network from OSM data. Note that Step 4 does not make changes to the topology and is hence omitted from this illustration.

workflow is developed in Python and may hence be more accessible than other implementations which require external dependencies (e.g. SQL databases, commercial software, Java)¹⁰.

Compared to previous implementations in the global modal PyPSA-Earth¹¹, we significantly improve the work in speed and data quality by taking advantage of the topological, electrical and geographical information available for Europe in OSM. Given the generic structure of the developed workflow, it can be easily applied to other regions and fed back to the global PyPSA-Earth project. However, the output will directly depend on the OSM data quality for a particular region (e.g. whether data on the substation's geometric footprint is available). To fill in missing data, we introduce cleaning process that yields a representation of the European high-voltage grid. We benchmark the processed data against country-level statistics provided by ENTSO-E, concluding that OSM data coverage of the European high-voltage grid is high or even close to complete. These improvements will also contribute to the quality of transmission grids modelled on a global scale in PyPSA-Earth.

Methods

PyPSA-Eur is a spatially and temporally highly resolved, open-source, sector-coupled linear optimisation model that covers the European continent¹². The model is built on top of the open-source toolbox PyPSA¹³ and is suited for operational as well as expansion studies (transmission, generation, and storages). The model includes a stock of existing power plants (processed with the tool powerplantmatching¹⁴) as well as renewable potentials and availability time series (processed with atlite¹⁵). Throughout the last decade, PyPSA-Eur has gained a large user base from academia, industry, and policy makers alike and has been used in a variety of studies^{16–23}. Other open-source models exist — notable and widely-used ones include the *oemof*²⁴ and *OSeMOSYS Global*²⁵. However, both models lack the detailed geographical as well as electrical representation of the transmission grid that PyPSA-Eur provides. While the open European model *Euro-Calliope* does provide a high spatial resolution, it models the grid with net transfer capacities²⁶.

With the integration into PyPSA-Eur, we also enable compatibility with additional functions already implemented into the model, such as, but not limited to, the option to enable dynamic line rating²⁷ and the inclusion of projects under planning (e.g. European Ten-Year Network Development Plan²⁸ and the German Network Development Plan²⁹).

PyPSA-Eur is managed by a workflow management system called Snakemake³⁰. Its modular structure enables the addition of new model functionalities and data sources, these can then be toggled using a configuration file. We split the construction of the high-voltage grid into four steps and add them into the existing workflow (see Fig. 1). We also use the model for validation purposes (see section Technical Validation). While the dataset and its reconstruction is built into PyPSA-Eur, it has the potential to be used in other energy system models, too. For this purpose, we release the entire dataset, including the original and mapped technical parameters. To obtain a functioning, topologically connected representation of the European high-voltage grid based on OSM data, we take the following steps (see Fig. 2 for an application to an example region).

Step 1—Retrieving the OSM data. In OSM, geographical data is stored in ‘nodes’, ‘ways’, and ‘relations’. Being the simplest data type, nodes are defined by coordinates and associated parameters. Ways are geometric line strings that connect a set of nodes. Relations can contain nodes, ways, a combination of either or other relations⁷. For the purpose of this work, we extract ways for obtaining the outline of substations. For AC lines we use both ways and relations data to build the dataset. To harness the highest data quality, we use relational data whenever available. To avoid duplicates in the representation of an AC line, all members (ways) of a relation are removed from the dataset. This hybrid approach allows us to make use of relations whenever they are complete and exist, and fall back to ways if relations are incomplete or missing.

| Tag | Data type | Example |
|-----------|-----------|----------|
| cables | numeric | 9 |
| circuits | numeric | 3 |
| frequency | numeric | 50 Hz |
| power | string | line |
| voltage | numeric | 380000 V |

Table 1. Key tags and parameters for AC power lines and cables.

| line id | cables | circuits | frequency (Hz) | type | voltage (V) |
|------------|--------|----------|----------------|-------|----------------|
| way/1 | | 2 | 50 | cable | 380000 |
| relation/1 | 3 | | 50 | cable | 380000 |
| way/3 | 9 | 1;2 | 50 | line | 380000; 220000 |
| way/4 | 9 | 3 | 50; 50 | line | 380000; 220000 |
| way/5 | 8 | | 50 | line | 110000; 220000 |
| way/6 | | | 50 | cable | 300000 |

Table 2. Illustrative example of AC lines and cables input data.

| line id | circuits | frequency (Hz) | type | voltage (V) |
|------------|----------|----------------|-------|-------------------------|
| way/1 | 2 | 50 | cable | 380000 |
| relation/1 | 1 | 50 | cable | 380000 |
| way/3-1 | 1 | 50 | line | 380000 |
| way/3-2 | 2 | 50 | line | 220000 |
| way/4-1 | 1 | 50 | line | 380000 |
| way/4-2 | 1 | 50 | line | 220000 |
| way/5-1 | 1 | 50 | line | 110000 (removed) |
| way/5-2 | 1 | 50 | line | 220000 |
| way/6 | 1 | 50 | cable | 300000 |

Table 3. Illustrative example of AC lines and cables after cleaning. Changes highlighted in bold.

To obtain DC projects (usually cables, hereafter referred to as DC links) we use relations. There are two main reasons for this differentiation: (i) relations are the more complex data type, coverage for AC lines and cables is still scarce for Europe. For DC links, given that there are fewer of them in Europe, we have checked each project individually using the latest ENTSO-E map² and compared their technical parameters (e.g. nominal capacity) with previous datasets^{10,13}. Based on this, we consider the coverage of commissioned DC links complete (see Table 7); (ii) Some DC links contain multiple components, accessing relations allows us to efficiently aggregate and simplify the links for the purpose of static energy system modelling.

First, we retrieve the raw data using the Overpass turbo API⁸ (Fig. 2a). Specifically, we query the OSM database for electricity grid related features (see Table 5). Note that Overpass turbo has a limit on the size and number of requests for each query and provides the API under a fair use policy. To minimise unnecessary load and enable data reusability, we provide a prepared transmission grid for download via Zenodo (0.6 is the latest version at the time of publication)³¹.

Step 2 — Cleaning the OSM data. While OSM provides a rich dataset, it is not directly usable for energy system modelling. Next to geospatial coordinates, OSM includes tags that provide feature-specific additional information. Depending on the individual feature, data may however contain noise, be incomplete, or inconsistent. After importing the retrieved raw data into pandas dataframes³², we apply a series of steps including heuristics to clean and fill in the missing information (see example in Table 2). We then use the power of geopandas³³ to perform geospatial operations (including but not limited to spatial joins, intersections, buffering, etc.).

Substations and transformers. To obtain the set of substations, we filter for substations with a voltage level within the scope of interest, i.e., between AC 220 and 750 kV. Where available, we extract the polygon shape of the substations, stored in the element's geometry. This allows us to differentiate between internal and external grid components. Note that information on transformers is not extracted from OSM, as (i) we cannot adequately evaluate their coverage and (ii) this data is not sufficient to create a topologically connected network. Instead, we use a needs-based approach, i.e., adding a single transformer between buses of different voltages within the perimeter of the same substation. The nominal capacity of each transformer is determined by the maximum



Fig. 3 Example — Simplification of a DC link relation: Moyle interconnector (Scotland - Northern Ireland), OSM relation ID [6914309](#). Different colors show the four segments, i.e., ways that compose the OSM relation.

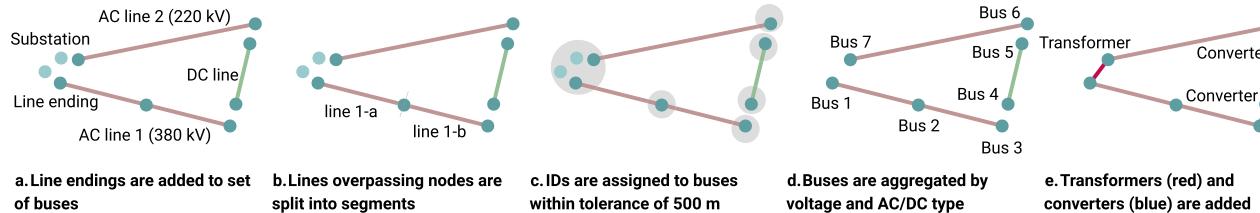


Fig. 4 Illustration of building a topologically connected grid.

aggregated capacity of connected AC lines and cables on either side. This is in line with previous approaches to obtain the ENTSO-E map based transmission grid¹².

AC lines and cables. In a first step, we clean the tag columns to only contain the correct data type and unit, as shown in Table 1 for AC power lines and cables. The minimum parameters that need to be given for a particular line or cable are ‘voltage’ (in) and ‘power’ (string: ‘line’ or ‘cable’).

We filter for the entries with a voltage level including and above AC 200 kV. While data for the mid- to low-voltage grid is also partially available in OSM, public statistics are scarce, making validation of such data difficult. As not all entries contain clean or complete data, we make heuristic assumptions to fill in the gaps, as illustrated in Tables 2 and 3.

For each line or cable we use the most specific information available that the data provides. In a three-phase AC high-voltage system, we assume three cables to form an AC circuit (e.g. way/2)³⁴. If a way contains multiple data points split by semicolons (i.e., transmission lines sharing overhead line routes), we split the entries into individual lines, accordingly. In this process, we preserve the original OSM identifier and its associated geometries. We add a numbered suffix after the split to maintain unique line ids (e.g. way/3 becomes way/3-1 and way/3-2). The given electric parameters are mapped according to the semicolon splits. In some cases however, where the number of data points across columns is not equal (e.g. way/4 and way/5), we make the following assumption: we take the floor of the number of circuits divided by the number of entries in the voltage column. In the absence of better information, this may lead to an underestimation of the real number of circuits.

If no information on cables or circuits is available, we assume a single circuit, provided that a voltage level is given (e.g. way/6). Finally, we remove all ways representing bus bars and lines fully contained within substation outline (Fig. 2b), as they are considered internal elements and do not provide additional information for static analyses.

DC links and converters. Due to their distinct electrical properties, we treat DC links differently from AC lines and cables. To avoid double counting, we remove all DC links from the original way queries. Instead, we query the OSM database for relations that contain DC links. As data on DC projects is widely available and because there are fewer of them^{2,35}, we contribute to the OSM database by adding missing parameters such as the nominal rating and voltage level. This signifies the ease of data improvements with OSM for the benefit of all. To trace future DC projects through our workflow, the following tags are required: ‘route’ = ‘power’, ‘frequency’ = 0, and ‘rating’ in ‘X MW’ format. DC components need to be correctly linked in the OSM database as member (either ‘cable’ or ‘line’) of the parent relation.

Note that some relations contain multiple DC link segments or components (e.g. converter stations or grounding), these are simplified into a single line with the sum of their nominal ratings. Figure 3 shows an example of how the Moyle interconnector from Northern Ireland to Scotland (Fig. 3a) is simplified (Fig. 3b). In this simplification, we preserve original end points of the DC link and the longest connected path. In PyPSA, converters are modelled as links connecting two buses. By analogy with how transformers are added to the network ex post, we connect the terminals of the DC link to their associated or nearest AC bus in the transmission grid. As such, we guarantee that DC links are always topologically connected.

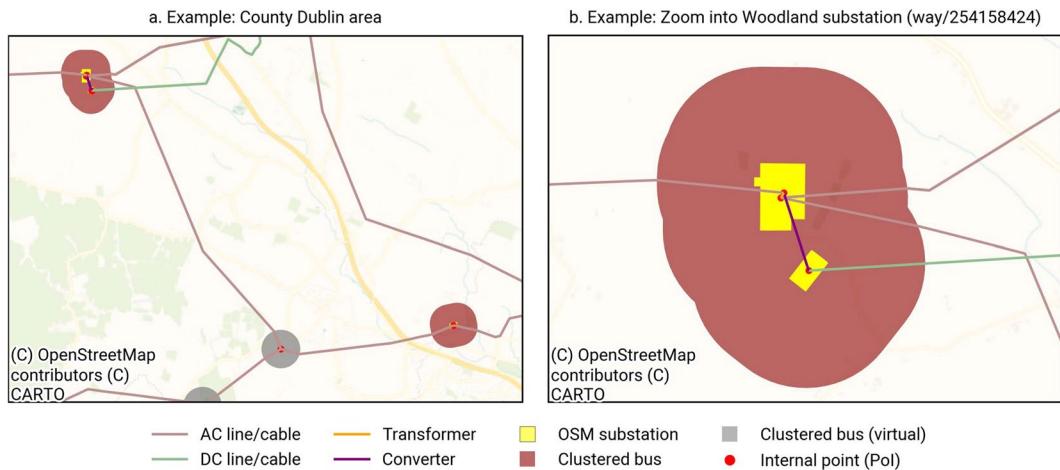


Fig. 5 Example of bus clustering. The darkred shape represents the union of the buffer around virtual buses and an original OSM substation polygon (yellow). The bright red dot represents the internal point of the union, this point sets the geographic coordinates of the obtained bus. Lines and cables are connected to the respective voltage level within the substation, transformers are added to connect buses of different voltage levels.

| | System costs (bn. €) | CAPEX (bn. €) | OPEX (bn. €) | Curtailment (TWh) | Generation (TWh) |
|-----------------|----------------------|---------------|--------------|-------------------|------------------|
| GridKit | 319.28 | 283.47 | 35.81 | 2749.41 | 3119.82 |
| OSM | 318.19 | 283.09 | 35.1 | 2742.26 | 3118.5 |
| Delta (Percent) | -0.34 | -0.13 | -1.99 | -0.26 | -0.04 |

Table 4. Comparison of key result metrics between ENTSO-E map and OSM-based transmission grid. CO₂ price: 100 €/t.

Step 3 — Building the OSM network. After having obtained a cleaned dataset, we need to ensure that the grid components are topologically and electrically connected. For this purpose, we propose a complete graph-based structure, i.e., buses/substations are connected by AC and DC lines, as well as converters and transformers. We introduce the following procedure, illustrated in Fig. 4.

First, we add the endpoints of both AC and DC lines to the set of buses (Fig. 4a). We refer to the buses created from endpoints as *virtual buses*. This step ensures that each line has two corresponding buses to which it is initially connected. Note that this may introduce potential duplicates of buses beyond those already present — which we will take care of later. Next, we specifically consider long lines that are overpassing (multiple) substations. Here we assume that in reality, the substation is electrically connected to the overpassing line. To achieve this, we split the original line into the subsegments between the intersecting nodes and add a suffix to its original identifier (Fig. 4b), to (i) preserve the uniqueness of each line and (ii) still allow for tracing back the line on OSM or OpenInfraMap⁹. We update the endpoints of the split lines, accordingly. For each virtual bus, we evaluate the number of connected lines. If a single line enters and exits the same virtual bus with identical electrical parameters — such as voltage level and number of circuits — the two lines are merged, and the connecting virtual bus is removed. This step simplifies the network by reducing the number of virtual buses, which do not correspond to real substations defined by OSM substation polygons. The remaining virtual buses have an important role in maintaining the network's topological connectivity, serving as line junctions, such as branching points that directly link multiple transmission lines (see Fig. 5, grey polygon).

Third, we cluster all buses within a radius of 500 m: From each substation polygon provided by OSM and each virtual bus, a buffer of 500 m is created (see Fig. 5). Substations and virtual buses are assigned to the same station identifier and clustered if they overlap. Then we calculate an internal point based on Pole of Inaccessibility (PoI)³⁶ which provides the geographic coordinates of the simplified substation. With this methodology, we can always find a polygon-internal point, even if the shape is non-convex. If the clustered shape contains one or multiple OSM substations, the largest substation determines the poi of the simplified bus. Connected lines are remapped, accordingly (Fig. 4d). Finally, we ensure that all given components are correctly electrically connected, i.e., AC buses of different voltage levels at the same substation are linked by transformers, DC buses which represent endpoints of DC lines are connected to the closest AC bus through converters (Fig. 4e) of equal nominal rating, respectively. Figure 5 shows an example how the clustering algorithm is applied and how lines are connected.

The proposed methodology is an enhanced version of the data processing integrated in the PyPSA-Earth model¹¹ which has previously demonstrated the potential of using OSM data for global energy system modelling. Our methodology improves the original implementation in efficiency, computational performance and quality regarding the representation of the European high-voltage grid (see Technical Validation Section).

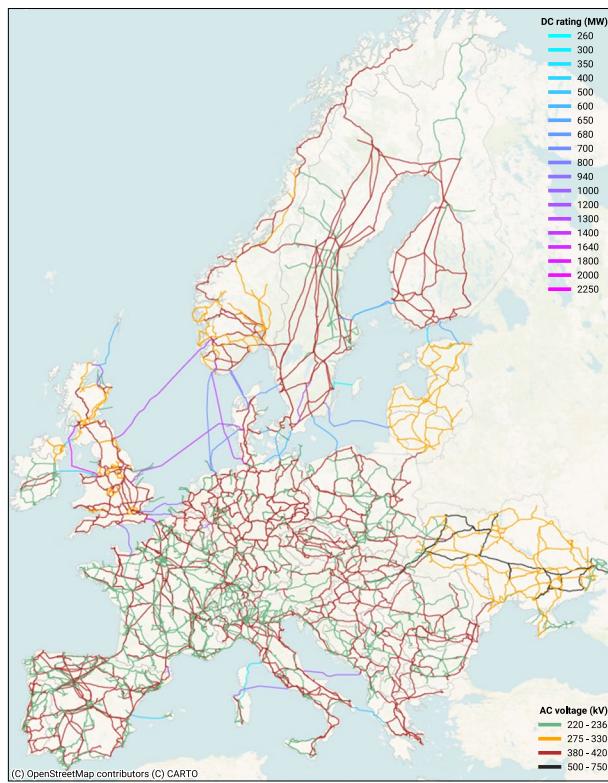


Fig. 6 Map of the OSM-based European high-voltage grid. This map was generated using the grid dataset provided with this publication³¹.

Step 4 — Creating a PyPSA-ready base network. In a final step, we create a network ready for modelling within PyPSA and PyPSA-Eur. The advantage of our dataset is that it is provided in .csv format, making it easily readable. Further, we include all original and mapped technical parameters of the grid components. Hence, it may potentially be used outside our given use case, directly or with small adaptations.

Based on the voltage levels, we assign each line to a standard line type from PyPSA^{12,13,37,38}, selecting the closest matching voltage (e.g. 400 kV is mapped to a 380 kV line type). Using the static grid model provided by 50Hertz³, we demonstrate in Fig. 18 that this approach is effective for this particular region in the absence of more accurate data. Using geometry line lengths and number of circuits, we calculate electric parameters, such as impedance, reactance and apparent power S_{nom}^{AC} (Eq. 1), as these are not contained in the original OSM input data. We point out that, while the mapped standard line type of a 380 kV and a 400 kV is the same, the calculated apparent power differs given the multiplication with their individual voltage levels. We apply a factor of 0.7 to approximate the N-1 security margin (Eq. 2)^{12,39}. Note that this factor can be individually set in the configuration file of PyPSA-Eur. We provide an overview of all resulting AC lines and cables in Table 8.

$$S_{nom}^{AC} = n_{circuits} \cdot \sqrt{3} \cdot U_{nom}^{OSM} \cdot I_{nom}^{pandapower} \quad (1)$$

$$S_{n-1}^{AC} = 0.7 \cdot S_{nom}^{AC} \quad (2)$$

For DC links, we use the provided length and nominal rating, directly. We assume that all DC links can be operated in both directions. Lastly, after transformers and converters have been added to the network, we remove all unconnected or islanded components.

Data Records

The compiled representation of the European high-voltage grid (Fig. 6) is hosted online in .csv format and can be downloaded via the Zenodo repository³¹. The dataset includes the geographical scope of the ENTSO-E member states (without Cyprus, Iceland, and Turkey), i.e., Albania (AL), Austria (AT), Belgium (BE), Bosnia and Herzegovina (BA), Bulgaria (BG), Croatia (HR), Czech Republic (CZ), Denmark (DK), Estonia (EE), Finland (FI), France (FR), Germany (DE), Greece (GR), Hungary (HU), Ireland (IE), Italy (IT), Kosovo (XK), Latvia (LV), Lithuania (LT), Luxembourg (LU), Moldova (MD), Montenegro (ME), Netherlands (NL), North Macedonia (MK), Norway (NO), Poland (PL), Portugal (PT), Romania (RO), Serbia (RS), Slovakia (SK), Slovenia (SI), Spain (ES), Sweden (SE), Switzerland (CH), Ukraine (UA), and the United Kingdom (GB).

We continuously update the dataset as the underlying OSM input data and the workflow improve. As of the submission of this paper, the resulting network (Fig. 6) contains 6737 buses—of which 4931 are original OSM

substations polygons and 1806 are virtual buses, 8994 AC lines and cables, 37 simplified/aggregated DC links (and converters at their endpoints, respectively), and 875 simplified AC transformers, comprising a total of 283974 km in total route length. In the following, we describe the structure, parameters, and units of the grid components which are stored as comma-separated value files of the same name. The description refers to the dataset version that forms the basis of this publication (i.e., version 0.6)³¹.

Buses (buses.csv). All AC and DC substations are represented as buses and stored in the ‘buses.csv’ file (0.9 MB). The file contains the following columns and properties:

- **bus_id** (string): Unique identifier of the bus. The name of a bus is derived in two ways: If a OSM relation or way was used to create the bus, the prefix starts with ‘way/’ or ‘relation/’, accordingly. As such, the original identifier is preserved. If the bus is a result of endpoints of connected lines or cables (‘virtual buses’), without an underlying OSM, the prefix is the two-letter ISO3166-1 alpha-2 country code, numbered from North to South and West to East. The suffix represents the voltage level of the bus.
- **voltage** (integer): Nominal voltage level of the bus in as extracted from the OSM data or connected lines and cables (in the case of virtual buses).
- **dc** (boolean): Boolean flag indicating whether the bus type is DC (True/t) or AC (False/f).
- **symbol** (string): String that describes the nature of the bus. By design, all buses are substations, hence the symbol is ‘Substation’.
- **under_construction** (boolean): Boolean flag indicating whether the bus is under construction (True/t) or not (False/f). This is needed to differentiate commissioned substations from planned transmission projects when activating extensions of the PyPSA-Eur model. By design, all buses included in this dataset are commissioned and hence not under construction.
- **tags** (string): Additional information on the bus, e.g. the prefix of the bus. This string can be used to access the underlying OSM way or relation, if available.
- **x** and **y** (float): Geographical coordinates of the bus in the WGS84 (EPSG:4326) coordinate system, calculated using the PoI approach³⁶.
- **geometry** (shapely Point): Geographical point of the bus in the WGS84 (EPSG:4326) coordinate system. This property provides the ‘x’ and ‘y’ formatted as Point object of the open-source Python library ‘shapely’⁴⁰. It can be imported using shapely.wkt.loads() function.

AC lines and cables (lines.csv). All AC lines and cables are stored in the ‘lines.csv’ file. Given the high resolution of the geographic linestrings, its file size is larger (19.7) and makes up the majority of the dataset’s content and size. The file contains the following columns and properties:

- **line_id** (string): Unique identifier of the AC line/cable. The name is directly derived from the underlying OSM way or relation, preserving the original identifier. The suffix includes the voltage level in. Merged lines include a ‘merged_’ prefix, followed by the OSM identifier of the longest contained line/cable as well as the number of additional lines/cables that were merged.
- **bus0** and **bus1** (string): Reference to the buses at the start (‘bus0’) and end (‘bus1’) of the line/cable, see ‘buses.csv’.
- **voltage** (integer): Nominal voltage level of the line/cable in as extracted from the OSM data.
- **i_nom** (float): Nominal current of the line/cable in kA as mapped using using pandapower’s standard line type library³⁸.
- **circuits** (integer): Number of parallel circuits of the line/cable, as extracted from the OSM data.
- **s_nom** (float): Nominal apparent power capacity of the line/cable in MV A as calculated using Eq. 1 based on pandapower’s standard line type library³⁸. Note that this value represents the maximum technical capacity (100%) taking into account the number of circuits.
- **r** (float): Line resistance in Ω as calculated using the length of the line/cable and pandapower’s standard line type library³⁸.
- **x** (float): Line reactance in Ω as calculated using the length of the line/cable and pandapower’s standard line type library³⁸.
- **b** (float): Line susceptance in as calculated using pandapower’s standard line type library³⁸.
- **length** (float): The length of the line/cable in as calculated using the linestring of the underlying OSM way or relation, transformed from the WGS84 (EPSG:4326) coordinate system to a metric distance projection, i.e., LAEA Europe (EPSG:3035).
- **underground** (boolean): Boolean flag indicating whether the element is underground (True/t) or overhead (False/f), hence an overhead line or a cable. This information is derived from the OSM data.
- **under_construction** (boolean): Boolean flag indicating whether the line/cable is under construction (True/t) or not (False/f). This is needed to differentiate commissioned transmission lines from planned transmission projects when activating extensions of the PyPSA-Eur model. By design, all lines included in this dataset are commissioned and hence not under construction.
- **type** (string): Type of the line as mapped using pandapower’s standard line type library³⁸. This property is the foundation for calculating ‘r’, ‘x’, ‘b’, and ‘s_nom’.
- **tags** (string): Additional information on the line/cable, e.g. the contained OSM way or relation. For merged lines/cables, contained elements are separated by semicolons.

- **geometry** (shapely LineString): Geographical line of the line/cable in the WGS84 (EPSG:4326) coordinate system. This property contains geographically highly spatially resolved linestring of the OSM way or relation. The endpoints have been cleaned to exactly end in ‘bus0’ and ‘bus1’. The property is provided as a shapely LineString object⁴⁰. It can be imported using shapely.wkt.loads() function.

DC lines and cables (links.csv). All DC lines and cables are stored in the ‘lines.csv’ file (0.4 MB). The file contains the following columns and properties:

- **link_id** (string): Unique identifier of the DC line/cable. The name is directly derived from the underlying OSM way or relation, preserving the original identifier. The suffix includes the voltage level in and a ‘DC’ identifier.
- **bus0** and **bus1** (string): Reference to the buses at the start (‘bus0’) and end (‘bus1’) of the line/cable, see ‘buses.csv’.
- **voltage** (integer): Nominal voltage level of the line/cable in as extracted from the OSM data.
- **p_nom** (float): Nominal real power capacity of the line/cable in as provided by the OSM relation.
- **length** (float): The length of the line/cable in as calculated using the linestring of the underlying OSM way or relation, transformed from the WGS84 (EPSG:4326) coordinate system to a metric distance projection, i.e., LAEA Europe (EPSG:3035).
- **underground** (boolean): Boolean flag indicating whether the element is underground (True/t) or overhead (False/f), hence an overhead line or a cable. This information is derived from the OSM data.
- **under_construction** (boolean): Boolean flag indicating whether the line/cable is under construction (True/t) or not (False/f). This is needed to differentiate commissioned transmission lines from planned transmission projects when activating extensions of the PyPSA-Eur model. By design, all lines included in this dataset are commissioned and hence not under construction.
- **tags** (string): Additional information on the line/cable, e.g. the contained OSM relation.
- **geometry** (shapely LineString): Geographical line of the line/cable in the WGS84 (EPSG:4326) coordinate system. This property contains geographically highly spatially resolved linestring of the OSM way or relation. The endpoints have been cleaned to exactly end in ‘bus0’ and ‘bus1’. The property is provided as a shapely LineString object⁴⁰. It can be imported using shapely.wkt.loads() function.

Transformers (transformers.csv). All transformers are stored in the ‘transformers.csv’ file (0.1 MB). The file contains the following columns and properties:

- **transformer_id** (string): Unique identifier of the transformer. The prefix is determined by the substation where the transformer is located. The suffixes contain the voltage ‘bus0’ and ‘bus1’ in kV.
- **bus0** and **bus1** (string): Reference to the buses at the start (‘bus0’) and end (‘bus1’) of the line/cable, see ‘buses.csv’.
- **voltage_bus0** and **voltage_bus1** (integer): Voltage level at the ‘bus0’ and ‘bus1’ side of the transformer in kV, respectively.
- **s_nom** (integer): Rounded (ceiling) apparent power of the transformer in. This is determined by the maximum throughput of the transformer, i.e., the maximum of the sum of connected lines/cables on either side.
- **geometry** (shapely LineString): Geometry of the transformer, represented by a linestring in (EPSG:4326) coordinate system, connecting two buses of different voltage levels at a substation. It can be imported using shapely.wkt.loads() function.

Converters (converters.csv). All converters are stored in the ‘converters.csv’ file (0.01 MB). The file contains the following columns and properties:

- **converter_id** (string): Unique identifier of the converter. The suffix is determined by the substation where the converter is located. The suffixes contain the voltage ‘bus0’ and ‘bus1’ in kV.
- **bus0** and **bus1** (string): Reference to the buses at the start (‘bus0’) and end (‘bus1’) of the line/cable, see ‘buses.csv’.
- **voltage**: Maximum voltage level of the converter in kV as extracted from OSM data of the connected DC lines/cables.
- **p_nom** (integer): Nominal real power capacity of the converter in MW, determined by the sum of real power capacities of the connected DC lines/cables.
- **geometry** (shapely LineString): Geometry of the converter, represented by a linestring in (EPSG:4326) coordinate system, connecting an AC and DC bus at a substation. It can be imported using shapely.wkt.loads() function.

Interactive map (map.html). The dataset is accompanied by an interactive map built using the open-source packages geopandas³³ and folium. The map contains all components of the dataset, i.e., buses, lines, transformers, and converters, their geometries and technical parameters. The map can be opened with any modern browser to

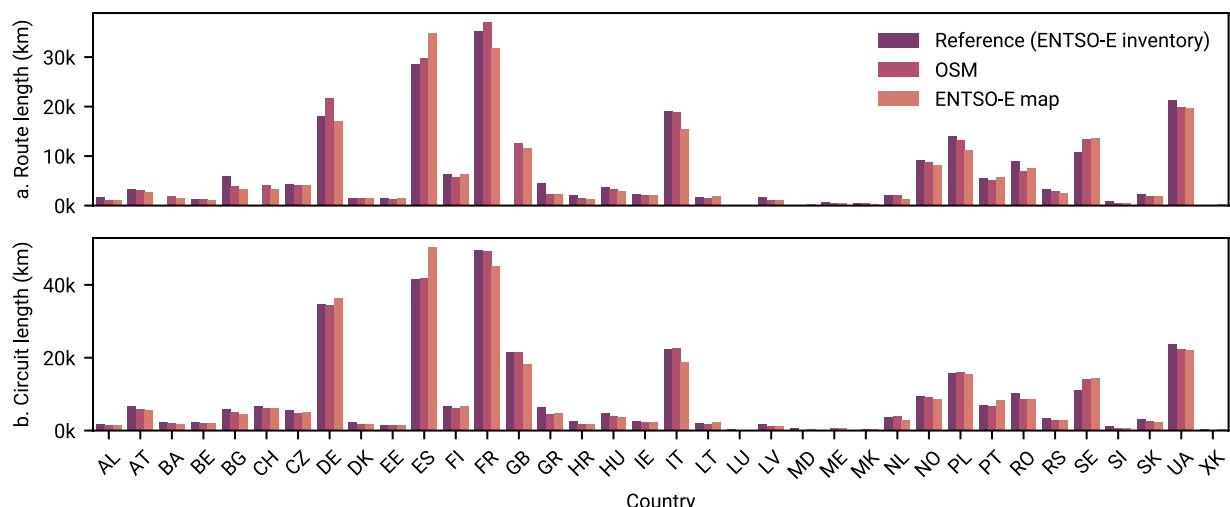


Fig. 7 Comparison of total route and circuit lengths per country.

visually explore the dataset. Different layers can be individually activated or deactivated — technical properties can be inspected by hovering over or clicking on the components. As the map is self-contained and stores all information of the dataset, it is considerably large at 53.1 MB and can be used offline. We provide a screenshot of the map in Fig. 6.

Technical Validation

We perform two validation steps for assessing the quality of the dataset. First, we compare the dataset with official inventory statistics provided by ENTSO-E. Second, we compare the results of an representative PyPSA-Eur model instance based on the two high-voltage grid datasets: OSM (presented in Fig. 6) and an extract from the online ENTSO-E map using GridKit tool (referred to as ‘ENTSO-E map’)¹⁰. This network is has been used by numerous PyPSA-Eur users, and is hence a good reference for comparison. In Fig. 17, we further provide a comparison between a geo-referenced, official dataset by 50Hertz³ and the OSM-based grid of the region, respectively.

Comparison with ENTSO-E statistics and map. Based on ENTSO-E’s 2023 inventory of transmission⁴¹, we first compare the total route (a) and circuit lengths (b) of AC lines and cables on a per country level (Fig. 7). Note that the inventory does not include all statistics for each country, i.e., route lengths are missing for Bosnia and Herzegovina, Switzerland, and Great Britain, while circuit lengths are missing for Montenegro and North Macedonia. While Ukraine and the Republic of Moldova have joined ENTSO-E on 1 January 2024 and 22 November 2023 as full and observing members, respectively, their inventory are not yet included in the dataset. For these two countries and Kosovo, we take reference data from third party sources^{42–45}.

We find that our transmission grid based on OSM data is in agreement with the ENTSO-E inventory. Calculating the Pearson correlation coefficient for both route and circuit lengths between the official statistics and the respective transmission grid representations, we see an overall improvement from the ENTSO-E map ($\rho_{\text{routes}} = 0.9489$ and $\rho_{\text{circuits}} = 0.9862$) to OSM ($\rho_{\text{routes}} = 0.9575$ and $\rho_{\text{circuits}} = 0.9980$) in the reproduction of the high-voltage grid (220 to 750 kV). One of the key reasons for these improvements is the much higher level of geographic detail of lines and cables in the OSM-based transmission grid compared to the stylised lines on ENTSO-E’s interactive map. We observe larger discrepancies for Sweden, where both transmission grid representations seem to overestimate the total lengths of the inventory.

As another validation step, we compare the total line volume on NUTS1 region level (Fig. 8a). Here we see strong similarities between the OSM-based and the extracted transmission grid of the ENTSO-E map, with few outliers ($\rho_{MVAkm} = 0.9674$). While a higher geospatial resolution of lines of the OSM-based transmission grid may contribute to the increase in line volume on average, we calculate $S_{\text{nom}}^{\text{AC}}$ using the more differentiated voltage levels given in the OSM data (see Table 8) as opposed to the clustered voltage levels given in the ENTSO-E map, i.e., 220 kV, 300 to 330 kV, 380 to 400 kV, 500 kV, and 750 kV. This may lead to a more accurate representation of the line volume in the OSM-based transmission grid.

To assess the transmission capacity across regions, we compare the capacity (Fig. 8b) and number of line crossings (Fig. 8c) per NUTS1 border (Fig. 8b). While the two transmission grids strongly correlate, we observe notable differences at individual borders ($\rho_{MVA} = 0.8441$), the same is true for the absolute number of line crossings ($\rho_{\text{crossings}} = 0.8575$). Due to different quality in geospatial information contained in both transmission grid representations, buses in one may be offset (or not even exist) in the other. Stronger outliers can primarily be traced back to buses close to NUTS1 borders (Fig. 8b). Notable outliers are located in Spain (light green), in Ukraine (pink), and southwestern parts of Germany (ocher). The reasons for discrepancies can be manifold, e.g. due to differences in exact locations of boundary nodes, missing, outdated or wrong data. In the case of Spain,

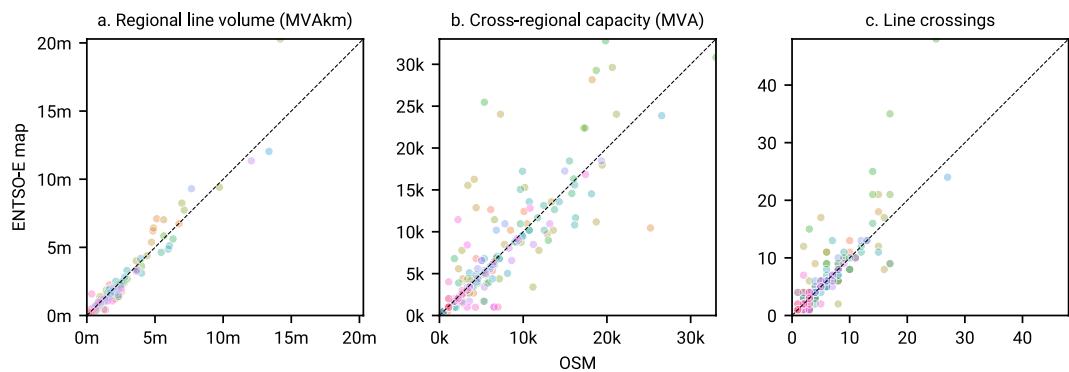


Fig. 8 Comparison of line volume per NUTS1 region — Colors represent individual countries. Line volume is the product of the nominal capacity and the length, summed over all lines within the region.

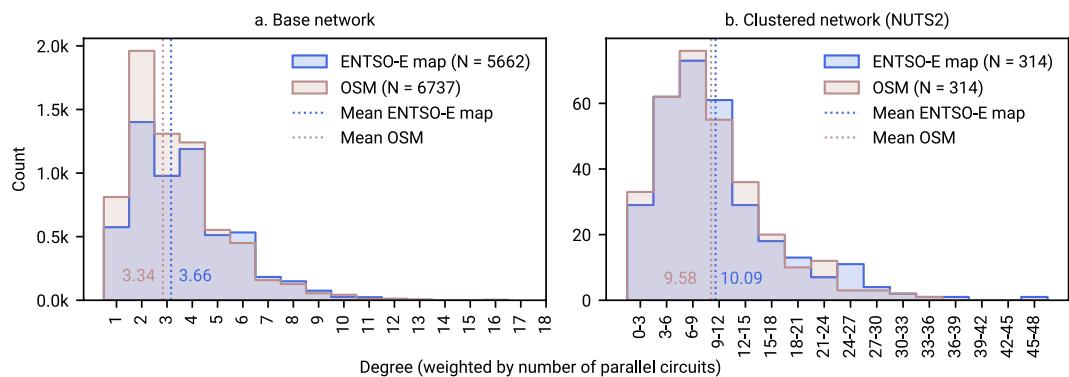


Fig. 9 Comparison of the weighted degree distribution in both transmission grid representations before and after clustering (NUTS2). Ukraine at geoBoundaries⁴⁸ administration level 1, Moldova at full bus resolution. A comparison at NUTS3 resolution is provided in Fig. 12.

the main reason for the large discrepancy lies in the representation of the high-voltage grid in and around the Madrid area, where the ENTSO-E map is stylised for clarity. A comparison with an official map provided by the National Geographic Institute of Spain⁴⁶ confirms the OSM topology to be more accurate.

Comparison of model results. In order to assess the impact of the new transmission grid representation, we compare the results of a representative PyPSA-Eur model run based on the OSM dataset to a run based on the ENTSO-E map which is currently being used in PyPSA-Eur¹². Note that the results shown in this publication are based on version 0.6 of the released prebuilt high-voltage grid representation on Zenodo (<https://doi.org/10.5281/zenodo.14144752>)³¹. We use the same model setup and input data in both model runs, except for the grid representation. We focus on the electricity sector, taking techno-economic assumptions projected for the year 2030 based on version 0.9.2 of the PyPSA technology dataset⁴⁷. We allow for capacity expansion in renewable energy as well as gas-fired generation capacities. To narrow down the effect of the transmission grid, we do not allow for grid expansion and disable dynamic line rating. We set the carbon price to 100 € per tonne of CO₂ emitted.

As the number and exact locations of buses differ, we cluster the networks to make them comparable (N = 318 regions/buses). In energy system modelling, clustering is often motivated by the spatial reduction of the optimisation problem. While oftentimes clustering algorithms based on grid topology or resource class are used⁴⁸, we are interested in the regional differences of the two grid representations. As such, we map the buses of both grids to clusters based on administrative boundaries, i.e., NUTS2. For non-NUTS countries such as Ukraine, we use the administration level 1 (geoBoundaries⁴⁹) and for Moldova we keep the full high-voltage substation resolution, i.e., 8 nodes. This yields 318 regions/buses for each of the clustered high-voltage grids, respectively. Note that the clustering process in PyPSA-Eur involves a transformation of all transmission lines and cables to the default voltage level of 380 kV. We then run the model at hourly resolution of the year 2030, yielding 8760 time steps.

Figure 9 compares the weighted degree distribution for the two network topologies. We weight the degree by the number of parallel circuits (Eq. 3) to account for potential different representations of lines and links connecting the same two buses (e.g., single lines with multiple number of circuits or multiple lines with single circuits). If $G = (V, E)$ is a weighted graph with vertex set V (buses) and edge set E (lines, links, converters, and transformers), and each edge $e \in E$ has a weight $w(e)$, then the weighted degree of a vertex $v \in V$ is given by:

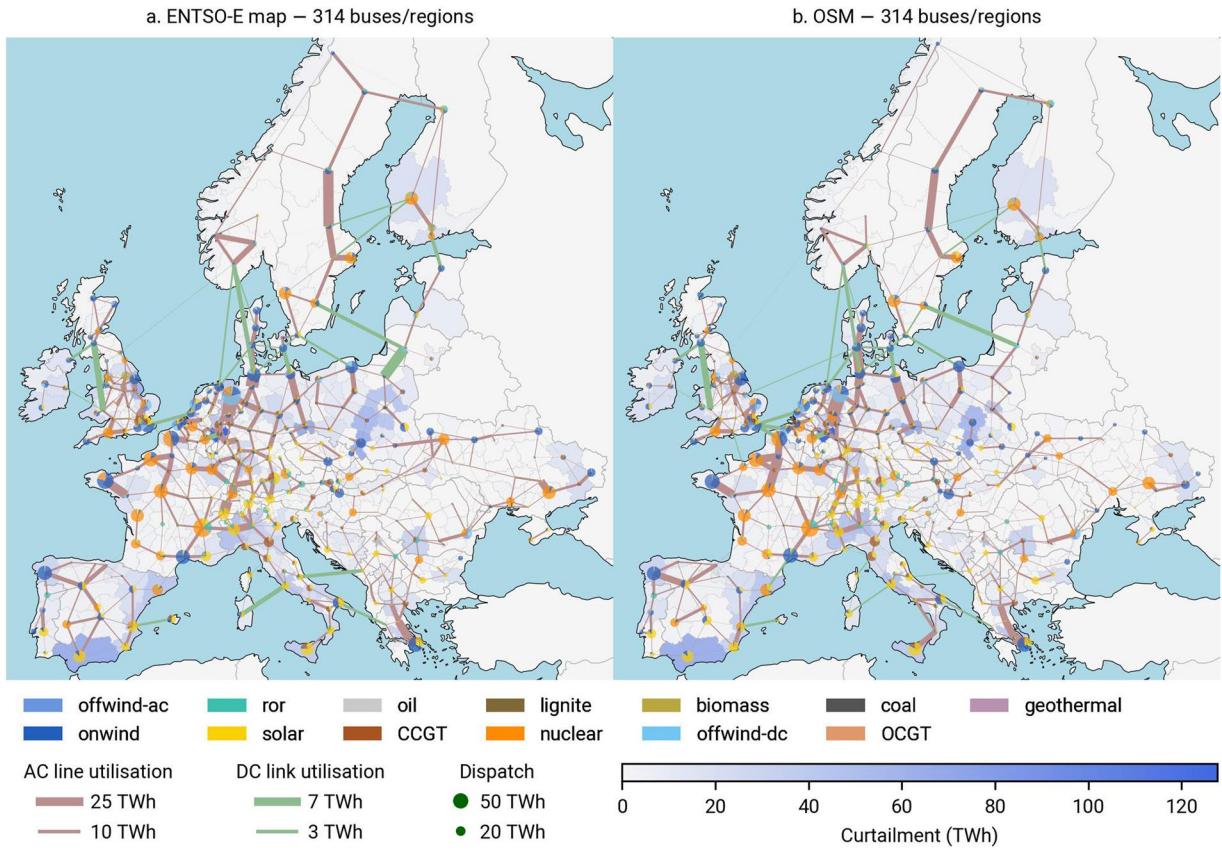


Fig. 10 Regional dispatch, line utilisation and curtailment. A map comparing nominal ratings of the two clustered grids is provided in the Fig. 13.

$$d_w(v) = \sum_{e \in \text{IncidentEdges}(v)} w(e) \quad (3)$$

where $\text{IncidentEdges}(v)$ is the set of edges incident to v . We find that the two base networks have a similar weighted degree distribution (Pearson correlation coefficient $\rho_{\text{degree}, \text{base}} = 0.9838$). Notably, the OSM-based transmission grid demonstrates a higher number of buses with degree 2. Clustering the two networks before running the optimisation problem will increase the Pearson correlation coefficient to $\rho_{\text{clustered}, \text{base}} = 0.9888$ (using binned data)—indicating that at NUTS2 resolution, the two networks are very similar in terms of connectivity.

We solve the two optimisation problems on a high-performance cluster (AMD EPYC 7543 32-Core processor) using up to 130 GB of memory. Each problem takes up to 280 iterations to converge, translating into 20 hours. Running the model for the two transmission grid representations, we find that regional results align closely, including the dispatch of generation assets, utilisation of lines and curtailment (Fig. 10). This is also true for the aggregated picture (Table 4). Total annual system costs drop from 319.28 bn. € in the ENTSO-E map based run to 318.19 bn. € in the OSM-based optimisation, corresponding to mere 0.34 % in difference. The higher weighted degree for both the base and clustered OSM-based high-voltage grid (Fig. 9) indicate a higher topological connectivity, potentially translating into higher degrees of freedom in the optimisation problem compared to the ENTSO-E map based run. This is confirmed when we look into the line and link utilisation and compare the investments. Since aggregate statistics over a continental area can be deceiving (smooth out errors), we show the differences in regional generation, line and link utilisation as well as curtailment in Fig. 11. Here, we can clearly see that the high-voltage grid in OSM is higher utilised than its ENTSO-E map based counterpart. We can also observe that the ENTSO-E map based model compensates by investing more into generation and storage capacities, i.e., +1.3 GW in offshore wind, +1.5 GW in combined cycle gas turbines, +0.6 GW in solar photovoltaics, and +2 GW in battery storage. We also provide an overview of key model results for sensitivity runs with varying CO₂ prices (200 and 300 €/t) in Tables 9–10. In Figs. 14–16, we compare average electricity prices, CAPEX and OPEX at nodal level.

Bottlenecks in both model runs are located in the same regions, contributing to an annual curtailment in the range of 2742 to 2149 TWh. More prominent differences in line utilisation are visible in Poland, Sweden, northern France, in the western region of Ukraine, southern and central Spain around the Madrid area, as well as southern Italy.

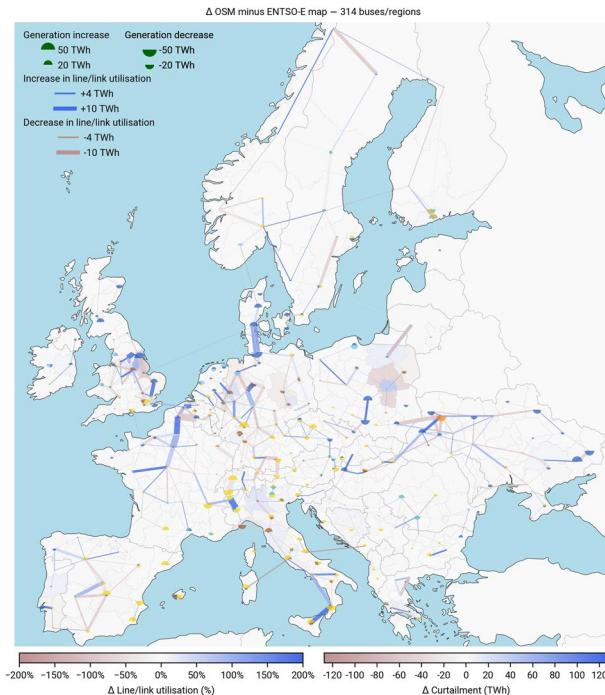


Fig. 11 Regional dispatch, line utilisation and curtailment (delta). Blue indicates an increase in curtailment or line utilisation from the ENTSO-E map to the OSM-based transmission grid, while red indicates a decrease. For full transparency, note that this map shows an outer join of all transmission grid elements, including lines and links that are not present in the other network.

Overall, the results of the two model runs are very similar, indicating that i) both grids seem to adequately represent reality and ii) the OSM-based transmission grid is a suitable replacement for the ENTSO-E map based grid. The higher utilisation of the OSM-based transmission grid is in line with the higher topological connectivity of the network. The differences in the investment decisions are marginal and can be attributed to the differences in grid topology.

We have demonstrated that the dataset aligns well with official statistics and the ENTSO-E map, and that the results of a representative PyPSA-Eur model instance based on both datasets are highly similar. Its core strengths stem from the high level of geographic detail and the continuous updates to the OSM database in combination with a strong PyPSA-Eur user base. The workflow is completely transparent and the data is openly accessible. Although we cannot guarantee the accuracy of the data, as only TSOs have access to the real grid data, we believe this dataset provides the best publicly available representation of the European high-voltage grid.

Usage Notes

The published dataset is provided under the ODbL 1.0 licence. Geoinformation is encoded in the WGS84 (EPSG:4326) coordinate system. While the dataset and workflow are provided as part of PyPSA-Eur, they are also suitable for a wide range of other applications, including network analyses, power flow calculations, as well as input for other energy system models and frameworks. Note that further validation and testing is needed for purposes outside the original scope of this work.

- In order to reproduce the network of this publication, the configuration file ‘config.yaml’: needs to be set to `base_network='OSM-prebuilt'`, `OSM-prebuilt-version: 0.6` and the command `snakemake base_network` -call needs to be executed.
- To rebuild the network from scratch, this setting can be changed to `base_network='OSM-raw'`, followed by the command `snakemake prepare_OSM_network_release -j 4`.
- Per default, buses within the perimeter of a 500 m radius are merged together. This value can be changed in the script, however this may change the topological connectedness of the obtained network.
- Networks can also be built for specific countries, regions or a subset of the countries within PyPSA-Eur by setting list of countries the configuration file.

Code availability

The code to replicate the entire workflow and dataset is provided as part of PyPSA-Eur and released as free software under the MIT licence. Different licences and terms of use may apply to the underlying input data.

- PyPSA-Eur¹² on GitHub: <https://github.com/pypsa/pypsa-eur>
- Version 0.6 of the prebuilt network³¹ based on OSM data can be retrieved via the Zenodo repository. This link will also point to future updates: <https://doi.org/10.5281/zenodo.14144752>
- An interactive map is bundled with the dataset on Zenodo³¹ (see Fig. 19). It can also be directly accessed via the PyPSA-Eur documentation⁵⁰: <https://pypsa-eur.readthedocs.io/en/latest/data-base-network.html>.

Received: 3 September 2024; Accepted: 28 January 2025;
Published online: 16 February 2025

References

- Hörsch, J. & Brown, T. The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios. In *2017 14th International Conference on the European Energy Market (EEM)*, 1–7, <https://doi.org/10.1109/EEM.2017.7982024> (2017).
- ENTSO-E. ENTSO-E Transmission System Map. <https://www.entsoe.eu/data/map/>.
- 50Hertz. Static grid model. <https://www.50hertz.com/Transparency/GridData/Gridfigures/Staticgridmodel/> (2022).
- JAO. Static Grid Model. <https://www.jao.eu/static-grid-model> (2023).
- Egerer, J. *et al.* Electricity sector data for policy-relevant modeling: Data documentation and applications to the German and European electricity markets. Research Report 72, DIW Data Documentation (2014).
- Hutcheon, N. & Bialek, J. W. Updated and validated power flow model of the main continental European transmission network. In *2013 IEEE Grenoble Conference*, 1–5, <https://doi.org/10.1109/PTC.2013.6652178> (2013).
- Medjroubi, W., Müller, U. P., Scharf, M., Matke, C. & Kleinhanß, D. Open Data in Power Grid Modelling: New Approaches Towards Transparent Grid Models. *Energy Reports* 3, 14–21, <https://doi.org/10.1016/j.egyr.2016.12.001> (2017).
- Raifer, M. Overpass turbo. <https://overpass-turbo.eu/> (2024).
- Garrett, R. Open Infrastructure Map. <https://openinframap.org> (2024).
- Wiegmans, B. GridKit extract of ENTSO-E interactive map, <https://doi.org/10.5281/zenodo.55853> (2016).
- Parzen, M. *et al.* PyPSA-Earth. A new global open energy system optimization model demonstrated in Africa. *Applied Energy* 341, 121096, <https://doi.org/10.1016/j.apenergy.2023.121096> (2023).
- Hörsch, J., Hofmann, F., Schlachtberger, D. & Brown, T. PyPSA-Eur: An open optimisation model of the European transmission system. *Energy Strategy Reviews* 22, 207–215, <https://doi.org/10.1016/j.esr.2018.08.012> (2018).
- Brown, T., Hörsch, J. & Schlachtberger, D. PyPSA: Python for Power System Analysis. *Journal of Open Research Software* 6, <https://doi.org/10.5334/jors.188> (2018).
- Gotzens, E., Heinrichs, H., Hörsch, J. & Hofmann, F. Performing energy modelling exercises in a transparent way - The issue of data quality in power plant databases. *Energy Strategy Reviews* 23, 1–12, <https://doi.org/10.1016/j.esr.2018.11.004> (2019).
- Hofmann, F., Hampp, J., Neumann, F., Brown, T. & Hörsch, J. Atlite: A Lightweight Python Package for Calculating Renewable Power Potentials and Time Series. *Journal of Open Source Software* 6, 3294, <https://doi.org/10.21105/joss.03294> (2021).
- Neumann, F., Zeyen, E., Victoria, M. & Brown, T. The potential role of a hydrogen network in Europe. *Joule* 7, 1793–1817, <https://doi.org/10.1016/j.joule.2023.06.016> (2023).
- Victoria, M., Zeyen, E. & Brown, T. Speed of technological transformations required in Europe to achieve different climate goals. *Joule* 6, 1066–1086, <https://doi.org/10.1016/j.joule.2022.04.016> (2022).
- Brown, T. & Hampp, J. Ultra-long-duration energy storage anywhere: Methanol with carbon cycling. *Joule* 7, 2414–2420, <https://doi.org/10.1016/j.joule.2023.10.001> (2023).
- Glaum, P., Neumann, F. & Brown, T. Offshore power and hydrogen networks for Europe's North Sea. *Applied Energy* 369, 123530, <https://doi.org/10.1016/j.apenergy.2024.123530> (2024).
- Riepin, I. & Brown, T. On the means, costs, and system-level impacts of 24/7 carbon-free energy procurement. *Energy Strategy Reviews* 54, 101488, <https://doi.org/10.1016/j.esr.2024.101488> (2024).
- Rahdan, P., Zeyen, E., Gallego-Castillo, C. & Victoria, M. Distributed photovoltaics provides key benefits for a highly renewable European energy system. *Applied Energy* 360, 122721, <https://doi.org/10.1016/j.apenergy.2024.122721> (2024).
- Grochowicz, A., van Greevenbroek, K. & Bloomfield, H. C. Using power system modelling outputs to identify weather-induced extreme events in highly renewable systems. *Environmental Research Letters* 19, 054038, <https://doi.org/10.1088/1748-9326/ad374a> (2024).
- Transnet, BW. Stromnetz 2050 - Eine Studie der TransnetBW. Tech. Rep., TransnetBW GmbH (2022).
- Hilpert, S. *et al.* The Open Energy Modelling Framework (oemof) - A new approach to facilitate open science in energy system modelling. *Energy Strategy Reviews* 22, 16–25, <https://doi.org/10.1016/j.esr.2018.07.001> (2018).
- Barnes, T., Shivakumar, A., Brinkerink, M. & Niet, T. OSeMOSYS Global, an open-source, open data global electricity system model generator. *Scientific Data* 9, 623, <https://doi.org/10.1038/s41597-022-01737-0> (2022).
- Tröndle, T., Lilliestam, J., Marelli, S. & Pfenninger, S. Trade-Offs between Geographic Scale, Cost, and Infrastructure Requirements for Fully Renewable Electricity in Europe. *Joule* 4, 1929–1948, <https://doi.org/10.1016/j.joule.2020.07.018> (2020).
- Glaum, P. & Hofmann, F. Leveraging the existing German transmission grid with dynamic line rating. *Applied Energy* 343, 121199, <https://doi.org/10.1016/j.apenergy.2023.121199> (2023).
- ENTSO-E. Ten-Year Network Development Plan (TYNDP) 2020 Main Report. Tech. Rep., ENTSO-E (2020).
- BNetzA. Bestätigung des Netzentwicklungsplan Strom - NEP 2037/2045 (2023). Tech. Rep., Bundesnetzagentur (2024).
- Mölder, F. *et al.* Sustainable data analysis with Snakemake, <https://doi.org/10.12688/f1000research.29032.2> (2021).
- Xiong, B., Fioriti, D., Neumann, F., Riepin, D. & Brown, T. Prebuilt Electricity Network for PyPSA-Eur based on OpenStreetMap Data. Version 0.6, <https://doi.org/10.5281/zenodo.14144752> (2024).
- McKinney, W. Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference* 56–61, <https://doi.org/10.25080/Majora-92bf1922-00a> (2010).
- Jordahl, K. *et al.* Geopandas/geopandas: V0.8.1. Zenodo <https://doi.org/10.5281/zenodo.3946761> (2020).
- Kirschen, D. *S. Power Systems: Fundamental Concepts and the Transition to Sustainability* (John Wiley & Sons, 2024).
- Pierri, E., Binder, O., Hemdan, N. G. A. & Kurrat, M. Challenges and opportunities for a European HVDC grid. *Renewable and Sustainable Energy Reviews* 70, 427–456, <https://doi.org/10.1016/j.rser.2016.11.233> (2017).
- Garcia-Castellanos, D. & Lombardo, U. Poles of inaccessibility: A calculation algorithm for the remotest places on earth. *Scottish Geographical Journal* 123, 227–233, <https://doi.org/10.1080/14702540801897809> (2007).
- Oeding, D. & Oswald, B. *R. Elektrische Kraftwerke und Netze* (Springer, Berlin, Heidelberg, 2016).
- Thurner, L. *et al.* Pandapower-An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems. *IEEE Transactions on Power Systems* 33, 6510–6521, <https://doi.org/10.1109/TPWRS.2018.2829021> (2018).
- Shokri Gazafroudi, A., Neumann, F. & Brown, T. Topology-based approximations for N-1 contingency constraints in power transmission networks. *International Journal of Electrical Power & Energy Systems* 137, 107702, <https://doi.org/10.1016/j.ijepes.2021.107702> (2022).

40. Gillies, S. *et al.* Shapely. Zenodo <https://doi.org/10.5281/zenodo.13345370> (2024).
41. ENTSO-E. Inventory of Transmission - 2023 data. <https://www.entsoe.eu/publications/data/power-stats> (2024).
42. CIGRE. The Power System of Ukraine (2018).
43. GlobalData. Top five transmission line projects in the Ukraine (2023).
44. Moldelectrica. Technical and economic indicators. https://moldelectrica.md/ro/network/annual_report (2023).
45. KOSTT. Transmission Development Plan. Tech. Rep., KOSTT Long-term Planning and Development (2022).
46. Instituto Geografico Nacional. Energía - Mapa de red eléctrica española. 2016. <http://atlasnacional.ign.es/wane/Energ%C3%ADA> (2016).
47. Zeyen, E. *et al.* PyPSA/technology-data: V0.9.2 <https://doi.org/10.5281/zenodo.13617294> (2024).
48. Frysztacki, M. M., Recht, G. & Brown, T. A comparison of clustering methods for the spatial reduction of renewable electricity optimisation models of Europe. *Energy Informatics* **5**, 4, <https://doi.org/10.1186/s42162-022-00187-7> (2022).
49. Runfola, D. *et al.* geoBoundaries: A global database of political administrative boundaries. *PLOS ONE* **15**, e0231866, <https://doi.org/10.1371/journal.pone.0231866> (2020).
50. Xiong, B., Fioriti, D., Neumann, F., Iegor, R. & Brown, T. Base network. <https://pypsa-eur.readthedocs.io/en/latest/data-base-network.html> (2024).
51. Egerer, J. Open Source Electricity Model for Germany (ELMOD-DE). *Data Documentation* (2016).
52. OSMTGmod Documentation 0.1.0. https://github.com/wupperinst/osmTGmod/blob/master/osmTGmod_documentation_0.1.0.pdf (2017).

Acknowledgements

The authors would like to thank the OSM community for their tremendous motivation and efforts in providing the data underlying this work. Further, the authors would like to thank the PyPSA-Earth community for setting the groundwork and initiating the move to grid data based on OSM. Last but not least, the authors would like thank Ekaterina Fedotova and Emmanuel Bolarinwa for early substantial developments and validations in integrating OSM into PyPSA-Earth and Philipp Glaum for technical sparring and support. Thanks to Lukas Trippe for embedding the interactive map to the PyPSA-Eur documentation. Map data copyrighted OpenStreetMap contributors and available from <https://www.openstreetmap.org>. This work was supported by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) under Grant No. 03EI4083A (RESILIENT) and Italian Ministry of University and Research (MUR) CUP I53C23002650007 (RESILIENT). This project has been funded by partners of the CETPartnership (<https://cetpartnership.eu/>) through the Joint Call 2022. As such, this project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement no. 101069750.

Author contributions

B.X.–Conceptualisation, Data curation, Methodology, Software/Programming, Model building and validation, Visualisation, Writing–original draft, review and editing. D.F.–Conceptualisation, Software/Programming, Writing–review and editing. F.N.–Conceptualisation, Discussion, Writing–review and editing. I.R.–Discussion, Writing–review and editing. T.B.–Supervision, Writing–review and editing.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to B.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

Figures & Tables

| Grid element | Overpass turbo query |
|---------------------------|---|
| AC power lines and cables | way['power' = 'line'], way['power' = 'cable'] and relation["route" = "power"]['frequency' != '0'] |
| DC links | relation['route' = 'power']['frequency' = '0'] |
| substations | way['power' = 'substation'] and relation['power' = 'substation'] |

Table 5. Overpass turbo queries used to extract the high-voltage grid elements from OSM.

| Project | Regional scope | Tools/data | Last update | Georeferenced | Data published |
|--|----------------------------------|--|------------------------------------|---------------|----------------------------------|
| 50Hertz static grid model ³ | Germany (50Hertz control area) | based on asset inventory | April 2022 (updated once a year) | Yes | Yes |
| ELMOD ⁵ | Europe | ENTSO-E interactive map ² plus manual changes | January 2014 (data not published) | Yes | No |
| ELMOD-DE ⁵¹ | Germany | VDE, TSO maps, and OSM | March 2016 (single release) | Yes | Yes (not reproducible) |
| Hutcheon & Bialek ⁶ | Europe (UCTE plus Balkan region) | PowerWorld model | June 2013 (updated once) | No | Yes (not reproducible) |
| JAO static grid model ⁴ | CORE capacity calculation region | based on CORE TSO asset inventory | April 2024 (updated frequently) | No | Yes |
| PyPSA-Eur ¹² | Europe | ENTSO-E interactive map ² using GridKit ¹⁰ plus manual changes | January 2022 (updated once) | Yes | Yes (reproduction complex) |
| osmTGmod ⁵² | Germany | OSM using Osmosis (SQL and Java) ¹⁰ | November 2017 (single release) | Yes | Yes (reproduction complex) |
| SciGrid (Power) ⁷ | Europe, Germany | OSM using GridKit ¹⁰ | November 2015 (updated once) | Yes | Yes (reproduction complex) |
| This publication | Europe | OSM using Overpass turbo API and Python | November 2024 (updated frequently) | Yes | Yes ³¹ (reproducible) |

Table 6. Notable projects and datasets modelling the high-voltage grid in Europe (alphabetical order). Note that the specific regional scope referring to 'Europe' may vary across the listed projects. A comparison of our dataset with the 50Hertz static grid model is shown in Fig. 17.

| OSM identifier | DC project name (sorted alphabetically) | From | To | Voltage (kV) | Rating (MW) | Calc. length (km) |
|--|--|--------|--------|--------------|-------------|-------------------|
| relation/8193755 | Aachen Lüttich Electricity Grid Overlay (ALEGrO) | BE | DE | 320 | 1000 | 89 |
| relation/3918230 | Baltic Cable | DE | SE | 450 | 600 | 260 |
| relation/14126301 | BritNed | GB-ENG | NL | 450 | 1000 | 253 |
| relation/8099179 | Caithness-Moray Link | GB-SCT | GB-SCT | 320 | 1200 | 161 |
| relation/17631956 | Conexión Mediterránea Transporte Alta tensión (COMETA) | ES | ES | 250 | 400 | 241 |
| relation/8185420 | Copenhagen-Brussels-Amsterdam cable (COBRAcable) | DK | NL | 320 | 700 | 328 |
| relation/2127794 | Cross-Channel (IFA 2000) | FR | GB-ENG | 270 | 2000 | 69 |
| relation/8184641 | East-West Interconnector | IE | GB-WLS | 400 | 500 | 255 |
| relation/15772117 | ElecLink | FR | GB-ENG | 320 | 1000 | 52 |
| relation/8184630 | Estlink 1 | EE | FI | 150 | 350 | 104 |
| relation/6886400 | Estlink 2 | EE | FI | 450 | 650 | 174 |
| relation/8184631 | Fenno-Skan | FI | SE | 400 | 500 | 228 |
| relation/3391954 | Fenno-Skan 2 | FI | SE | 500 | 800 | 304 |
| relation/3392001 | Gotlandslännen (Gotland 2 + 3) | SE | SE | 150 | 260 | 96 |
| relation/5487095 | Great Belt power link | DK | DK | 400 | 600 | 56 |
| relation/9934065; relation/9934066 | Interconnexion Electrique France-Espagne (INELFE) | ES | FR | 320 | 2000 | 64 |
| relation/10377412 | Interconnexion France-Angleterre 2 (IFA-2) | FR | GB-ENG | 320 | 1000 | 230 |
| relation/9982798 | Italy-Corsica-Sardinia (SACOI) — Corsica-Sardinia | FR | IT | 200 | 300 | 141 |
| relation/3391794 | Italy-Corsica-Sardinia (SACOI) — Italy-Corsica | FR | IT | 200 | 300 | 240 |
| relation/8185664 | Italy-Greece | GR | IT | 400 | 500 | 299 |
| relation/2505320 | KONTEK | DE | DK | 400 | 600 | 166 |
| relation/8184629 | Konti-Skan | DK | SE | 300 | 680 | 142 |
| relation/8185767 | Montenegro-Italy (MON.ITA) | IT | ME | 500 | 1200 | 411 |
| relation/6914309 | Moyle Interconnector | GB-NIR | GB-SCT | 500 | 500 | 61 |
| relation/8185487 | Nemo Link | BE | GB-ENG | 400 | 1000 | 139 |
| relation/9965201 | NorNed | NL | NO | 450 | 700 | 578 |
| relation/8185455 | NordBalt | LT | SE | 300 | 700 | 446 |
| relation/16213216 | NordLink | DE | NO | 525 | 1400 | 620 |
| relation/13295785 | North Sea Link | GB-ENG | NO | 515 | 1400 | 723 |
| relation/8185711 | Sardegna-Penisola Italiana (SAPEI) | IT | IT | 500 | 1000 | 418 |
| relation/15777152 | Shetland HVDC Link | GB-SCT | GB-SCT | 320 | 600 | 282 |
| relation/3391931 | Skagerrak (1-3) | DK | NO | 350 | 940 | 241 |
| relation/8184632 | Skagerrak (4) | DK | NO | 500 | 700 | 235 |
| relation/3392010 | SwePol | PL | SE | 450 | 600 | 254 |
| relation/8184633 | SydVästlänken | NO | SE | 300 | 1200 | 252 |
| relation/15781671 | Viking Link | DK | GB-ENG | 525 | 1400 | 741 |
| relation/15775538 | Western HVDC Link | GB-SCT | GB-WLS | 600 | 2250 | 398 |

Table 7. List of DC projects in the OSM-based transmission grid. Note that OSM relation identifiers are unique and persistent as long as the object is not deleted. Projects can be directly accessed via the OSM website by clicking on their respective relation identifier in the table.

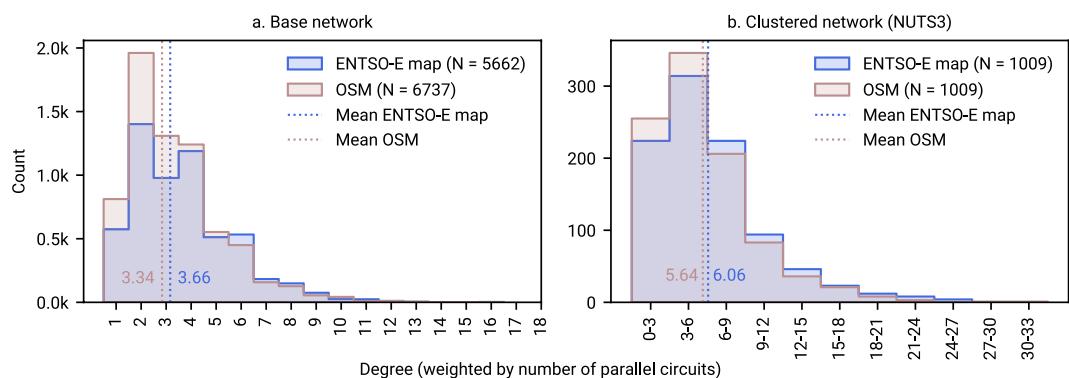


Fig. 12 Comparison of the weighted degree distribution in both networks before and after clustering (NUTS3). Ukraine at geoBoundaries⁴⁸ administration level 1, Moldova at full bus resolution.

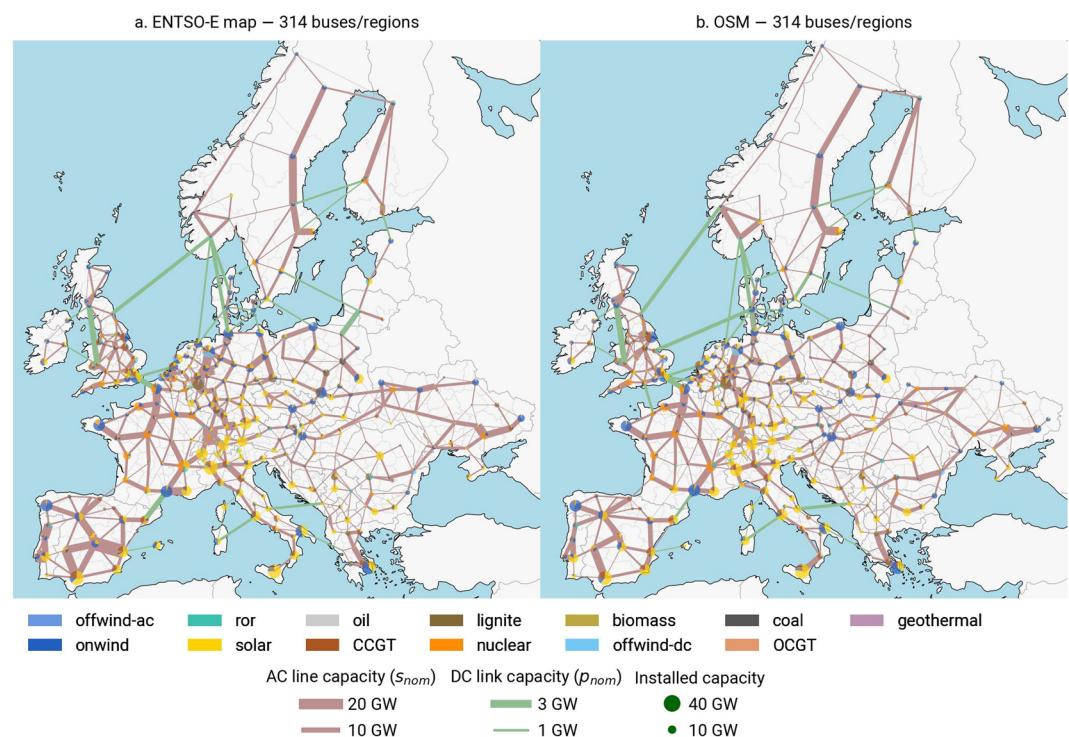


Fig. 13 Clustered AC line capacities, DC link nominal ratings, and optimal generation capacities.

| Voltage (kV) | Line type | Nominal current (kA) | Circuits | Rating (MVA) | Length (km) |
|--------------|-----------------------------|----------------------|----------|--------------|-------------|
| 220 | Al/St 240/40 2-bundle 220.0 | 1.29 | 1 | 492 | 59805 |
| 220 | Al/St 240/40 2-bundle 220.0 | 1.29 | 2 | 983 | 18612 |
| 220 | Al/St 240/40 2-bundle 220.0 | 1.29 | 3 | 1475 | 634 |
| 220 | Al/St 240/40 2-bundle 220.0 | 1.29 | 4 | 1966 | 79 |
| 225 | Al/St 240/40 2-bundle 220.0 | 1.29 | 1 | 503 | 24169 |
| 225 | Al/St 240/40 2-bundle 220.0 | 1.29 | 2 | 1005 | 1834 |
| 225 | Al/St 240/40 2-bundle 220.0 | 1.29 | 4 | 2011 | 9 |
| 236 | Al/St 240/40 2-bundle 220.0 | 1.29 | 1 | 527 | 18 |
| 275 | Al/St 240/40 3-bundle 300.0 | 1.935 | 1 | 922 | 3092 |
| 275 | Al/St 240/40 3-bundle 300.0 | 1.935 | 2 | 1843 | 1664 |
| 300 | Al/St 240/40 3-bundle 300.0 | 1.935 | 1 | 1005 | 4225 |
| 300 | Al/St 240/40 3-bundle 300.0 | 1.935 | 2 | 2011 | 30 |
| 330 | Al/St 240/40 3-bundle 300.0 | 1.935 | 1 | 1106 | 17107 |
| 330 | Al/St 240/40 3-bundle 300.0 | 1.935 | 2 | 2212 | 1016 |
| 380 | Al/St 240/40 4-bundle 380.0 | 2.58 | 1 | 1698 | 26384 |
| 380 | Al/St 240/40 4-bundle 380.0 | 2.58 | 2 | 3396 | 7802 |
| 380 | Al/St 240/40 4-bundle 380.0 | 2.58 | 3 | 5094 | 194 |
| 380 | Al/St 240/40 4-bundle 380.0 | 2.58 | 4 | 6792 | 230 |
| 400 | Al/St 240/40 4-bundle 380.0 | 2.58 | 1 | 1787 | 79590 |
| 400 | Al/St 240/40 4-bundle 380.0 | 2.58 | 2 | 3575 | 18395 |
| 400 | Al/St 240/40 4-bundle 380.0 | 2.58 | 3 | 5362 | 13 |
| 400 | Al/St 240/40 4-bundle 380.0 | 2.58 | 4 | 7150 | 68 |
| 420 | Al/St 240/40 4-bundle 380.0 | 2.58 | 1 | 1877 | 4936 |
| 420 | Al/St 240/40 4-bundle 380.0 | 2.58 | 2 | 3754 | 3 |
| 420 | Al/St 240/40 4-bundle 380.0 | 2.58 | 3 | 5631 | 12 |
| 500 | Al/St 240/40 4-bundle 380.0 | 2.58 | 1 | 2234 | 230 |
| 750 | Al/St 560/50 4-bundle 750.0 | 4.16 | 1 | 5404 | 4002 |

Table 8. Nominal capacities in the OSM base network AC lines and cables to pandapower standard type library.

| | System costs (bn. €) | CAPEX (bn. €) | OPEX (bn. €) | Curtailment (TWh) | Generation (TWh) |
|-----------------|----------------------|---------------|--------------|-------------------|------------------|
| GridKit | 326.83 | 295.71 | 31.12 | 2936.64 | 3131.02 |
| OSM | 325.58 | 294.66 | 30.92 | 2927.38 | 3128.45 |
| Delta (Percent) | -0.38 | -0.36 | -0.65 | -0.32 | -0.08 |

Table 9. Sensitivity run — CO₂ price: 200 €/t. Comparison of key result metrics between ENTSO-E map and OSM-based transmission grid.

| | System costs (bn. €) | CAPEX (bn. €) | OPEX (bn. €) | Curtailment (TWh) | Generation (TWh) |
|-----------------|----------------------|---------------|--------------|-------------------|------------------|
| GridKit | 331.81 | 301.79 | 30.03 | 3027.28 | 3135.37 |
| OSM | 330.36 | 300.83 | 29.54 | 3024.33 | 3132.4 |
| Delta (Percent) | -0.44 | -0.32 | -1.64 | -0.1 | -0.09 |

Table 10. Sensitivity run — CO₂ price: 300 €/t. Comparison of key result metrics between ENTSO-E map and OSM-based transmission grid.

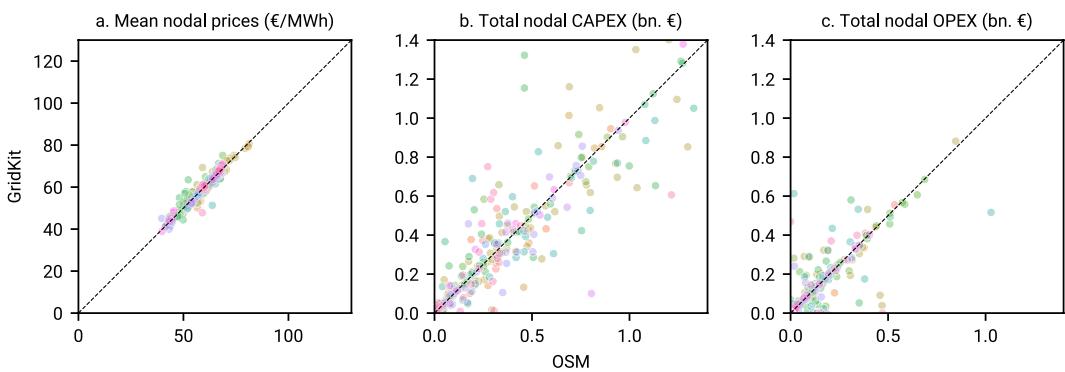


Fig. 14 Main results — CO₂ price: 100 €/t. Comparison of nodal prices, capital expenditures (CAPEX) and operational expenditures (OPEX).

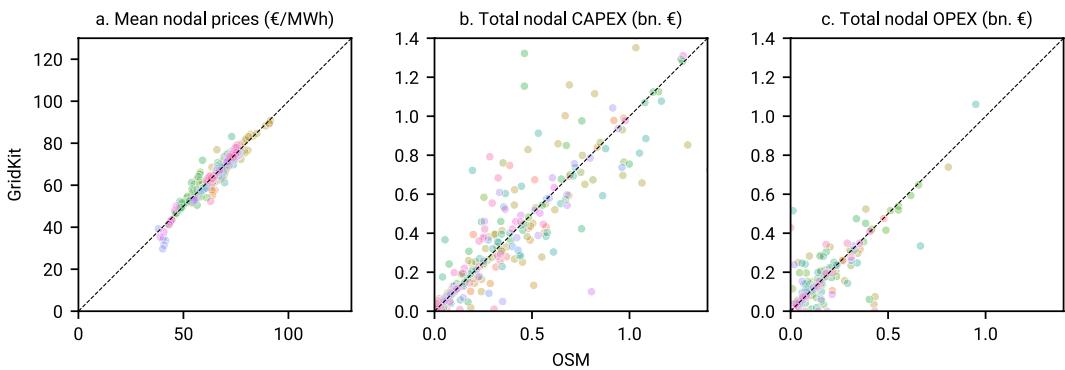


Fig. 15 Sensitivity run — CO₂ price: 200 €/t. Comparison of nodal prices, capital expenditures (CAPEX) and operational expenditures (OPEX).

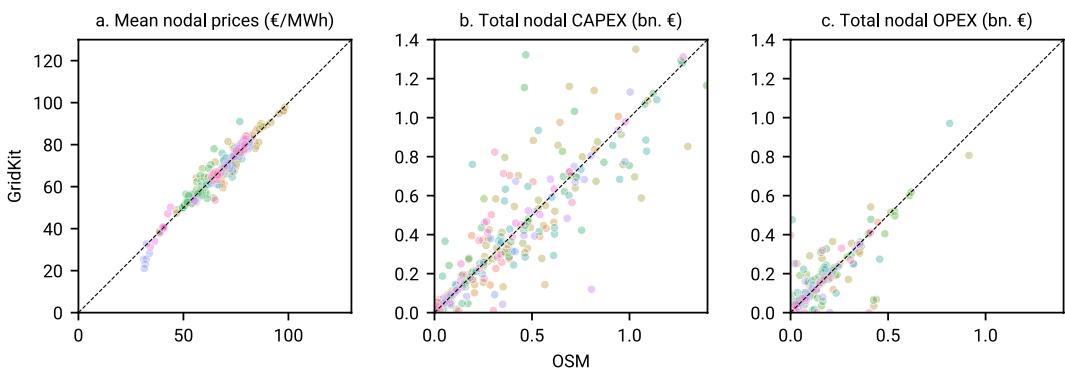


Fig. 16 Sensitivity run — CO₂ price: 300 €/t. Comparison of nodal prices, capital expenditures (CAPEX) and operational expenditures (OPEX).

a. 50Hertz – Static grid model b. OSM – Simplified line strings c. ENTSO-E map – Simplified line strings

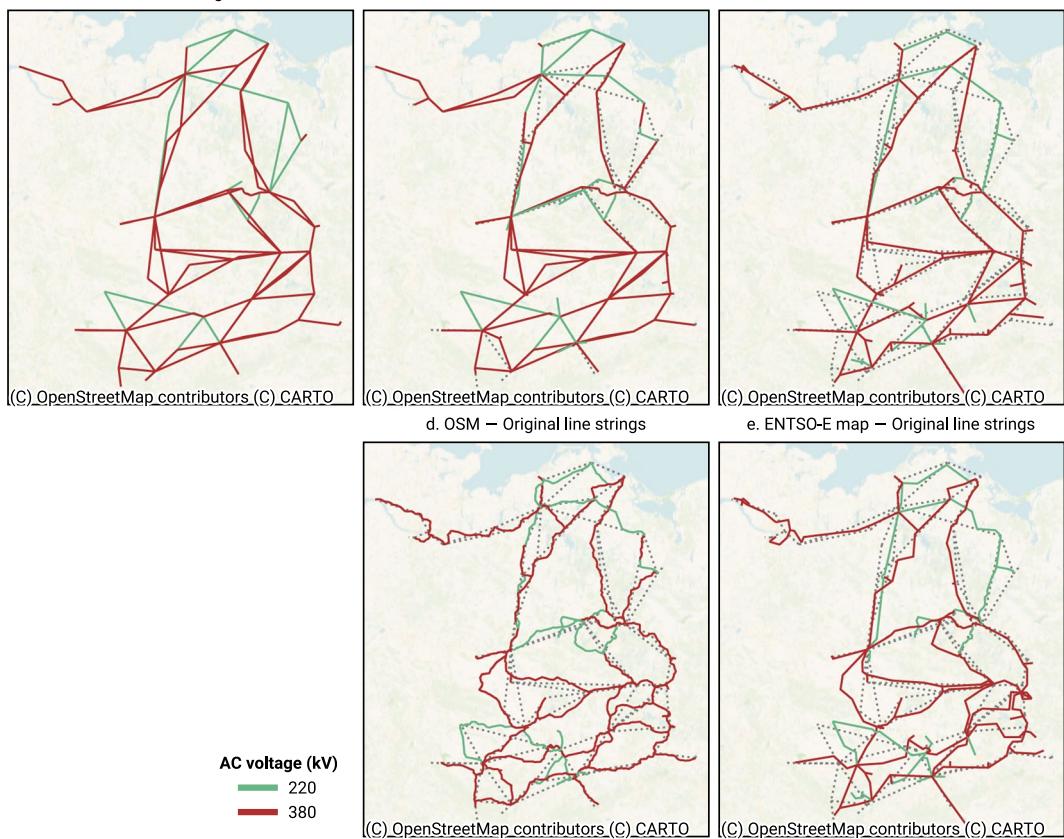


Fig. 17 Comparison of OSM and ENTSO-E map-based transmission grid with reference 50Hertz static grid model³. Dashed grey lines underneath show the 50Hertz static grid model for comparative purposes. Note that the geospatial data is provided in simplified, point-to-point form, only. Many of the dashed grey lines are also included in the OSM or ENTSO-E map-based grid representations, however, intermediary buses may exist along the line. As such, in their simplification they are not simplified to the level of the 50Hertz static grid model.

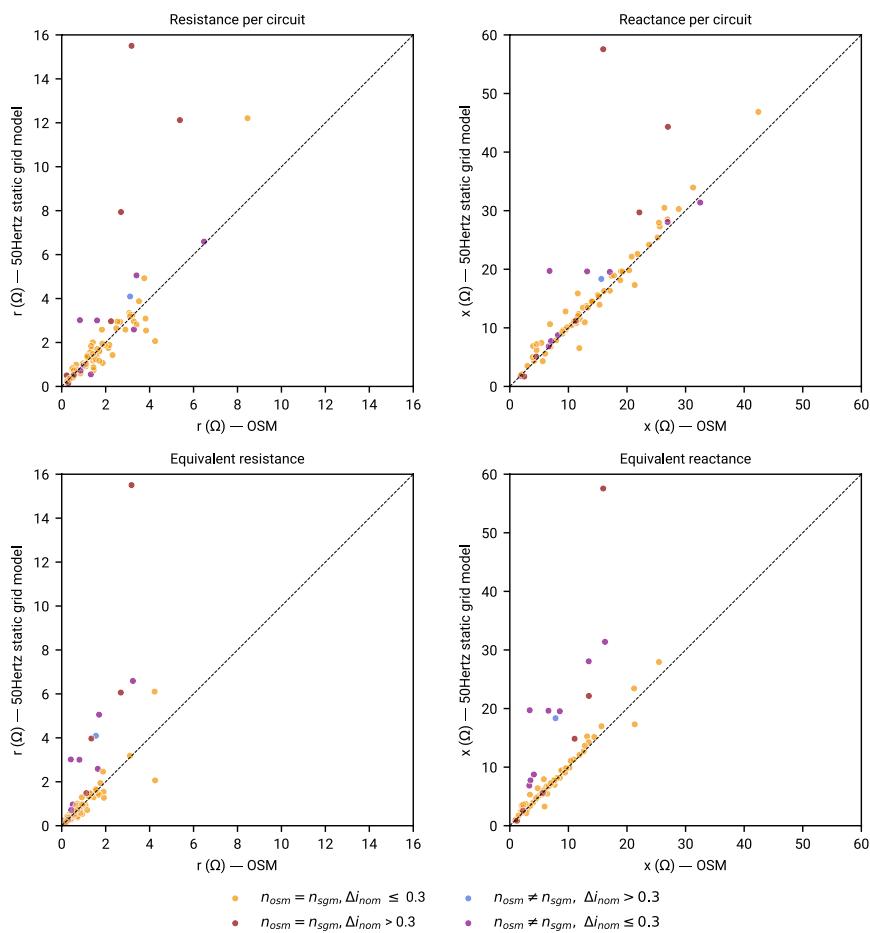


Fig. 18 Comparison of AC line/cable resistance and reactance between the OSM-based transmission grid and reference 50Hertz static grid model³. n_{osm} and n_{sgm} refer to the number of parallel circuits for a distinct line in each network, while $\Delta i_{nom} = \frac{|i_{nom,osm} - i_{nom,sgm}|}{i_{nom,sgm}}$ refers to the relative change in underlying nominal current.

Figure 18 was generated by mapping AC lines and cables of the OSM-based transmission grid to the 50Hertz static grid model (SGM) using OSM tags and SGM names (right join). Note that this data explains 4475 km of 5126 km in route length, as not all lines could mapped. For 79 % of the data, using pandapower's standard line types³⁸ for calculating the resistance and reactance comes close to official data in the SGM (orange). Purple data points show a discrepancy primarily due to unequal number of parallel circuits in both datasets (SGM data larger by factor 2). Red and blue data points indicate that underlying line types are entirely different. This is the case for some lines in SGM which are of a newer/stronger 380 kV (allowing higher currents) or weaker 220 kV line type.

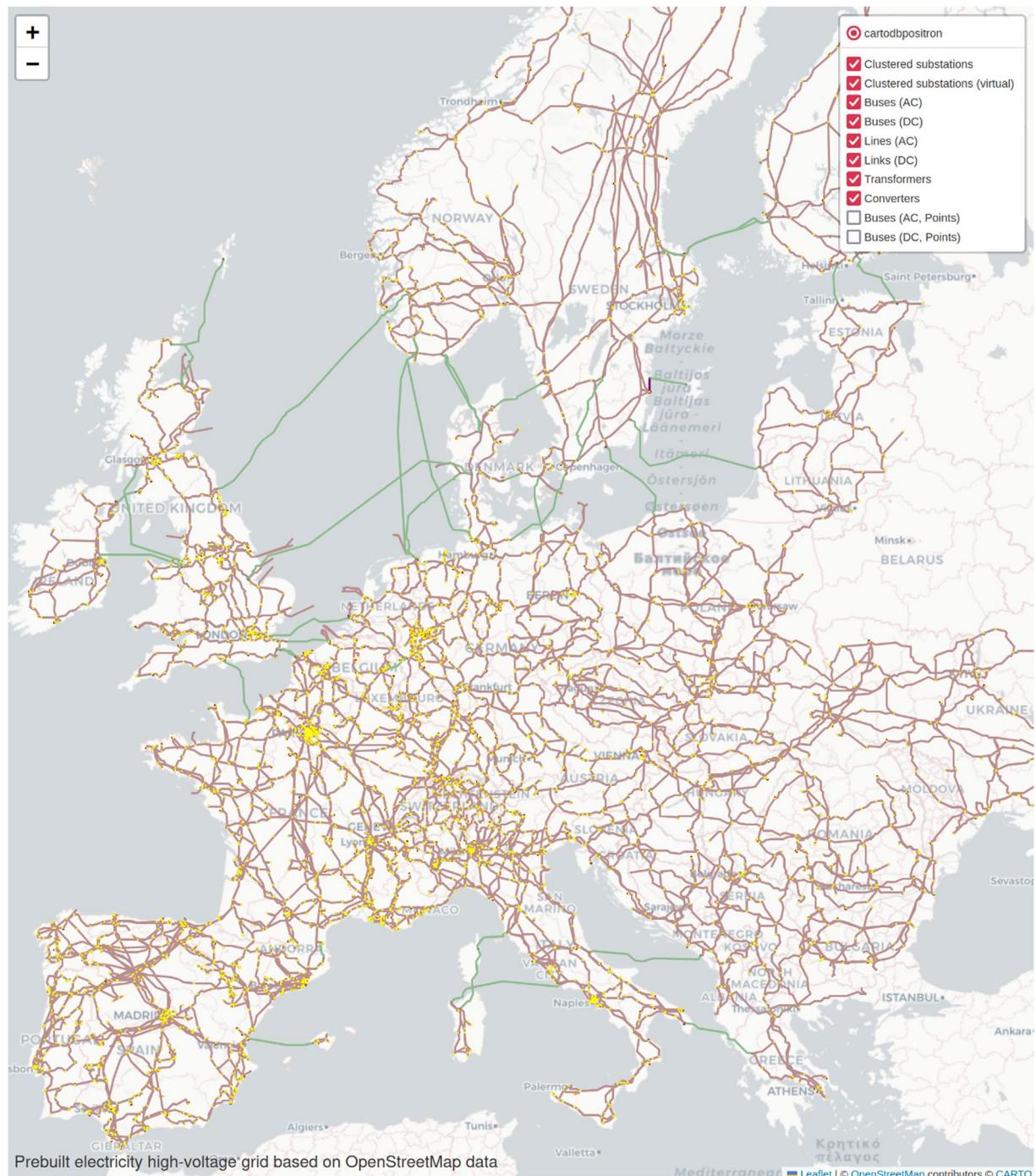


Fig. 19 Screenshot of the interactive map visualising the OSM-based transmission grid. The map is included in the dataset released on Zenodo (map.html)³¹.