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

Bobby Xiong[1,*], Davide Fioriti[2], Fabian Neumann[1], Iegor Riepin[1], and Tom Brown[1]

[1]Technische Universität Berlin, Department of Digital Transformation in Energy Systems (Institute of Energy Technology), Berlin, Germany
[2]Università di Pisa, Department of Energy Systems, Territory and Construction Engineering, Pisa, Italy
[*]corresponding author: Bobby Xiong (xiong@tu-berlin.de)

ABSTRACT

This paper provides the background, methodology and validation for constructing a representation of the European high-voltage grid, including and above 200 kV, based on public data provided by OpenStreetMap. The model-independent grid dataset is published under the Open Data Commons Open Database (ODbL 1.0) licence and can be used for large-scale electricity as well as energy system modelling. 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 value of open and maintainable high-voltage grid data extends to the global context. By accessing the latest data through the the Overpass turbo API, the dataset can be easily reconstructed and updated within minutes. To assess the data quality, this paper further compares the dataset with official statistics and representative model runs using PyPSA-Eur based on different electricity grid representations.

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 models.[1] 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 map[2] of the European high-voltage grid. There are however, several limitations typical for these sources: i) there is no underlying, topologically connected dataset, ii) it is not released under an open licence, iii) nor updated frequently, and iv) its geographic detail is limited or highly stylised.

There are previous projects that have modelled the European high-voltage grid or its parts based on OpenStreetMap (OSM) data. Some institutions provide data for particular regions, however all of them come with their individual limitations: While the most trustworthy data comes from TSOs themselves, they are — with few regional exceptions[3] — not georeferenced[4] 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 Application Programming Interface (API)[8] that always allows to retrieve the latest OSM data. Second, an active OSM community as well as a large user base of PyPSA-Eur and integration into automated workflows mean frequent updates and validation of processed data. Debugging can be easily done with the help of the open source project OpenInfraMap[9] which renders the OSM energy infrastructure on an interactive map. Finally, the entire 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-Earth,[11] 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 continent.[12] The model is build on top of the open-source toolbox PyPSA[13] 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 powerplanmatching[14]) as well as renewable potentials and availability time series (processed with atlite[15]). 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, one notable being OSeMOSYS Global,[24] however it lacks the detailed geographical as well as electrical representation of the transmission grid that PyPSA-Eur provides. 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 rating[25] and adding projects under planning (e.g. European Ten-Year Network Development Plan[26] and the German Network Development Plan[27]).

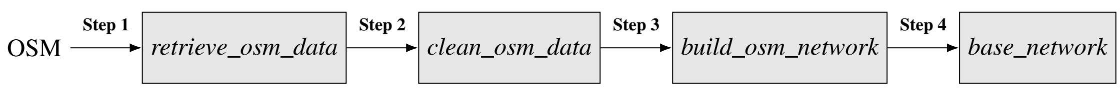

Figure 1. Process diagram for creating the European high-voltage grid from OSM data, representing the snakemake rules.

PyPSA-Eur is managed by a workflow management system called Snakemake.[28] 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. 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. To obtain a functioning, topologically connected representation of the European high-voltage grid based on OSM data, we take the following steps (see Figure 2 for an application to an example region).

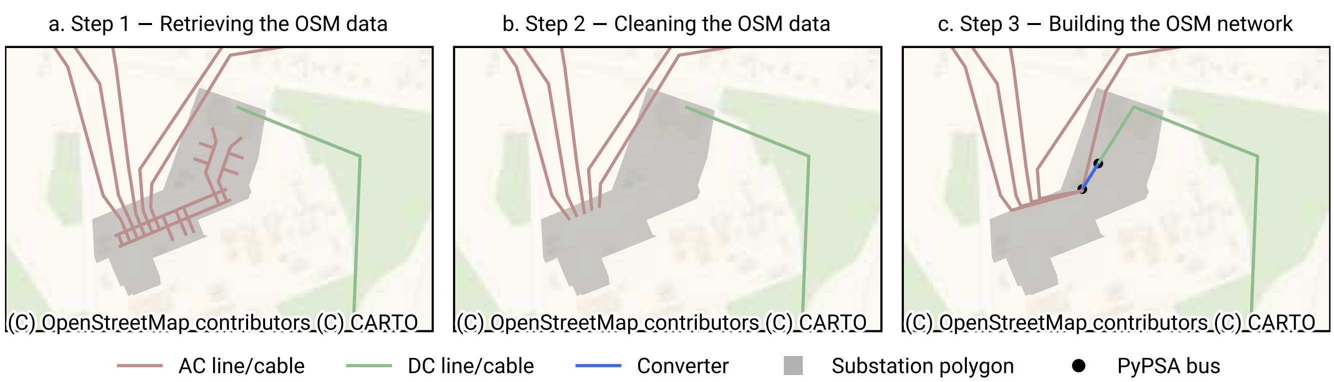

Figure 2. Illustration of the steps to create a PyPSA-ready network from OSM data. *Note that Step 4 does not make changes to the topology and is hence omitted from this illustration.*

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 relations.[7] For the purpose of this work, we extract ways for obtaining the outline of substations and AC power lines and cables. To obtain DC projects (usually cables, hereafter refered 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, whereas for DC links, given that there are much fewer of them in Europe, coverage is close to 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 API[8] (Figure 2.a). 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 avoid unnecessary load and re-use of data, we provide a prepared transmission grid for for download via Zenodo.[29]

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 a pandas dataframe,[30] 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 geopandas[31] 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 200 kV 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 are 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 of 2000 MW between buses of different voltages within the perimeter of the same substation. This is in line with previous approaches to obtain the ENTSO-E map based transmission grid.[12]

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 V) and 'power' (string: 'line' or 'cable').

Tag	Data type	Example
cables	numeric	9
circuits	numeric	3
frequency	numeric	50 Hz
power	string	line
voltage	numeric	380 000 V

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

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.

line id	**cables**	**circuits**	**frequency** (Hz)	**type**	**voltage** (V)
way/1		2	50	cable	380 000
way/2	3		50	cable	380 000
way/3	9	1;2	50	line	380 000; 220 000
way/4	9	3	50; 50	line	380 000; 220 000
way/5	8		50	line	110 000; 220 000
way/6			50	cable	300 000

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

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).[32] 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 nor circuits are available, we assume a single circuit, provided that a voltage level is given (e.g. way/6). Finally, we

line id	circuits	frequency (Hz)	type	voltage (V)
way/1	2	50	cable	380 000
way/2	1	50	cable	380 000
way/3-1	1	50	line	380 000
way/3-2	2	50	line	220 000
way/4-1	1	50	line	380 000
way/4-2	1	50	line	220 000
~~way/5-1~~	~~1~~	~~50~~	~~line~~	~~110 000~~
way/5-2	1	50	line	220 000
way/6	1	50	cable	300 000

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

remove all ways which represent bus bars and lines which are located fully inside of a substation outline (Figure 2.b), as they are considered internal elements of the substation and provide no additional information for the purpose of 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 are widely available and because there are fewer of them,[2,33] 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. In order for future DC projects to be traced by 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, respectively.

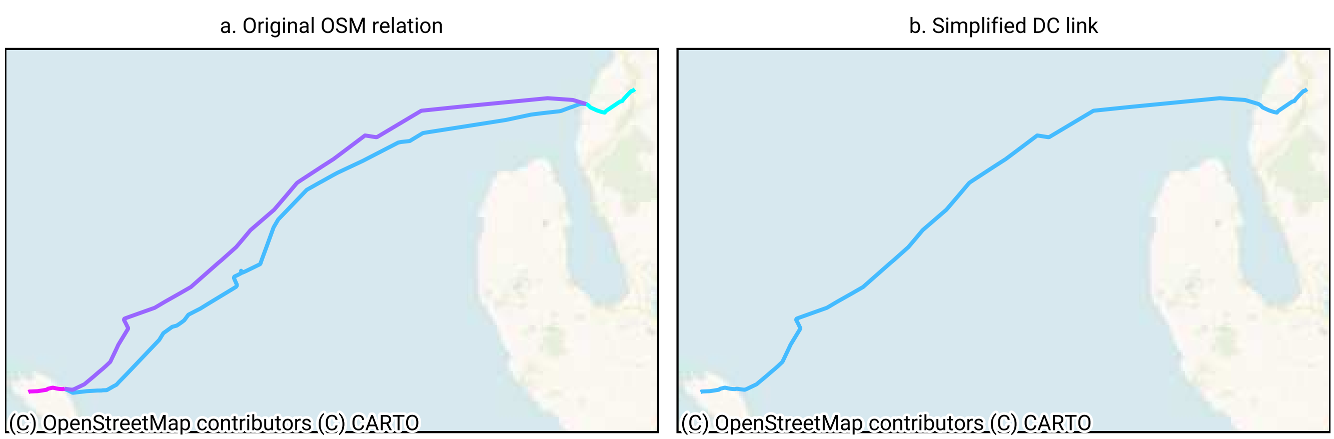

Figure 3. Example – Simplification of a DC link relation: Moyle interconnector (Scotland - Northern Ireland), OSM relation ID 6914309. *Different colors show the four segments, i.e. ways that compose the OSM relation.*

Note 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 (Figure 3.a) is simplified (Figure 3.b). 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. In analogy to how transformers are introduced to the network ex-post. i.e. connecting the terminals of the DC link, where the converter stations are also located on OSM and the closest AC bus in the transmission grid. As such, we guarantee that DC links are always topologically connected.

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 Figure 4.

First, we add the endpoints of both AC and DC lines to the set of buses (Figure 4.a). This step ensures that each line has

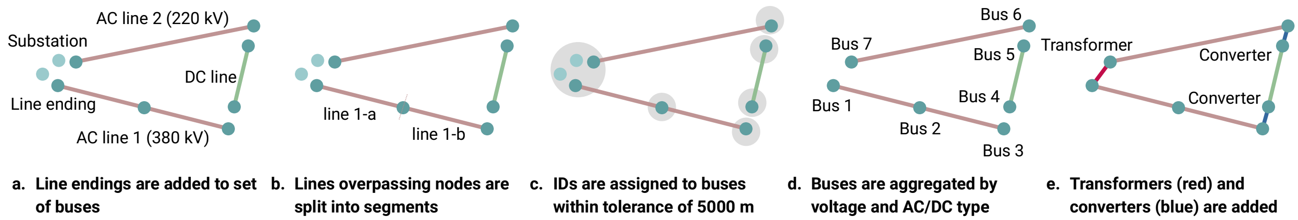

Figure 4. Illustration of building a topologically connected grid.

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 take of 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 (Figure 4.b), to i) preserve the uniqueness of each line and ii) still allow for tracing back the line on OSM or OpenInfraMap.[9] We update the endpoints of the split lines, accordingly. Third, we cluster all buses within a radius of 5000 m. We do this for several reasons, i) to avoid duplicates of buses that represent the same substation in reality, ii) to improve the computational efficiency of the model and iii) to improve the topological connectedness of the obtained network. In an interative procedure, buses within the given radius are assigned the same ID (Figure 4.c). Fourth, we cluster all buses assigned to the same station ID and AC/DC type. For each of the clusters, a bus is created. Connected lines are remapped, accordingly (Figure 4.d). 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 (Figure 4.e) of equal nominal rating, respectively.

The proposed methodology is an enhanced version of the data processing integrated in the PyPSA-Earth model[11] which has previously demonstrated the potential of using OSM data for global energy system modelling. Our proposed methodology improves the original implementation in efficiency, computational performance and enhancing the representation of the European high-voltage grid (see Technical Validation Section).

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 benefit of our our dataset is that it is provided in .csv format and easily readable. Hence, it may potentially be used outside our given use case, directly or with small adaptations.

Based on the voltage levels, we map each line to a standard line type library provided with PyPSA[12,13,34,35] with the closest voltage (e.g. 400 kV is mapped to a 380 kV line type). Using the standard grid model provided by 50Hertz,[3] we demonstrate in Figure 15 that this approach is effective for this particular region in the absence of more accurate data. Using the geometry line length and number of circuits, we calculate the 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 apply a factor of 0.7 to approximate the N-1 security margin (Eq. 2).[12,36] 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.

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

$$S_{n-1}^{AC} = 0.7 \cdot S_{nom}^{AC} \tag{2}$$

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 (Figure 5) is hosted online in .csv format and can be downloaded via the Zenodo repository.[29] The dataset includes the geographical scope of the ENTSO-E member states (without Cyprus, Iceland, Kosovo, and Turkey). We continuously update the dataset as the underlying OSM input data and the workflow is improved. As of the submission of this paper, the resulting network (Figure 5) contains 5848 buses, 7320 AC lines and cables, 36 simplified/aggregated DC links (and converters at their endpoints, respectively), and 1059 AC transformers, comprising a total of 261 757 km in total route length.

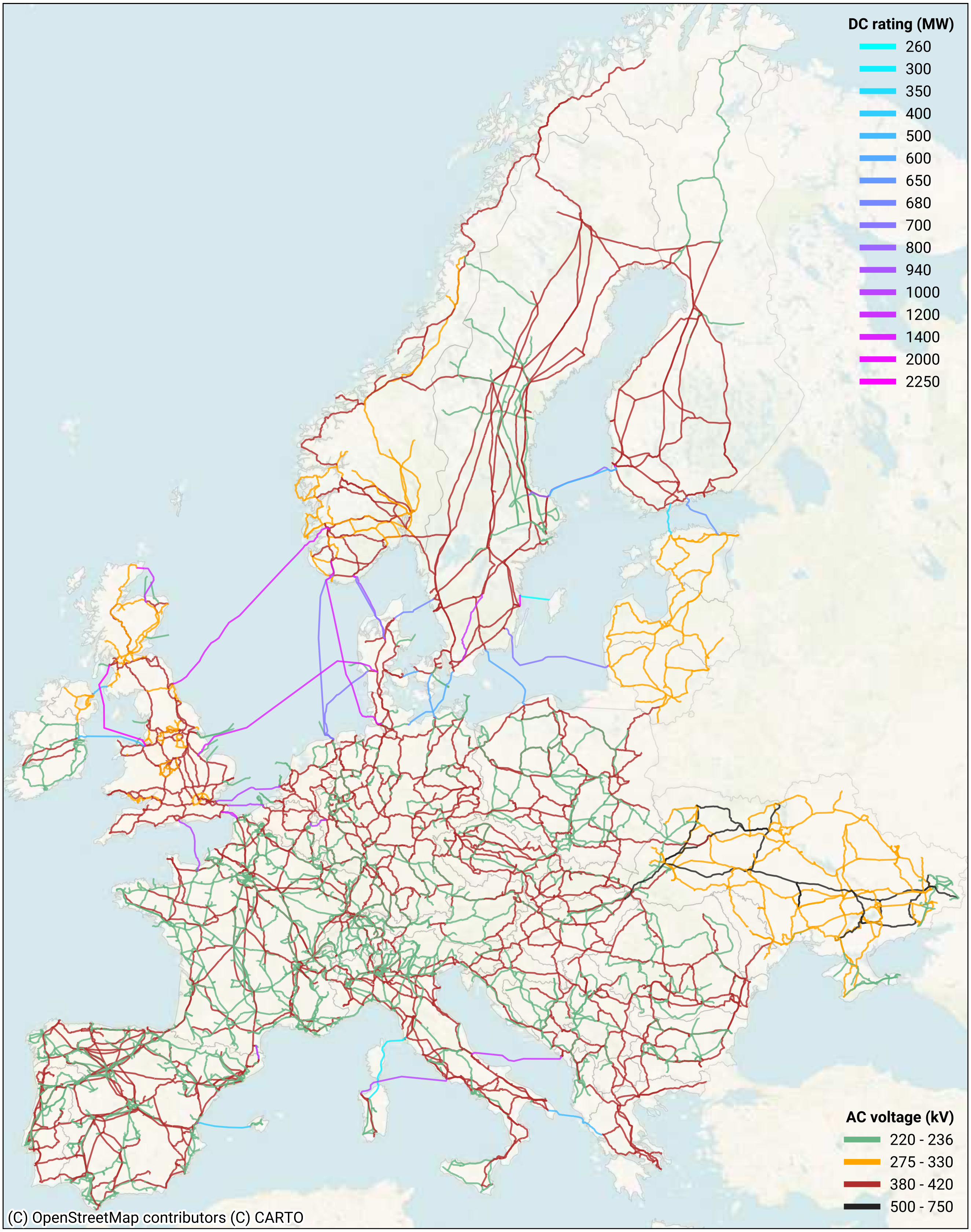

Figure 5. Map of the OSM-based European high-voltage grid. *This map was generated using the grid dataset provided with this publication.*[29]

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 5) and an extract from the online ENTSO-E map using GridKit tool (referred to as 'ENTSO-E map').[10] This network is currently being used by numerous PyPSA-Eur users, and is hence a good reference for comparison. In Figure 14, we further provide a comparison between a geo-referenced, official dataset by 50Hertz[3] and the OSM-based grid of the region, respectively.

Comparison with ENTSO-E statistics and map

Based on ENTSO-E's 2023 inventory of transmission,[37] we first compare the total route (a) and circuit lengths (b) of AC lines and cables on a per country level (Figure 6.). 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, we take reference data from third party sources.[38–40]

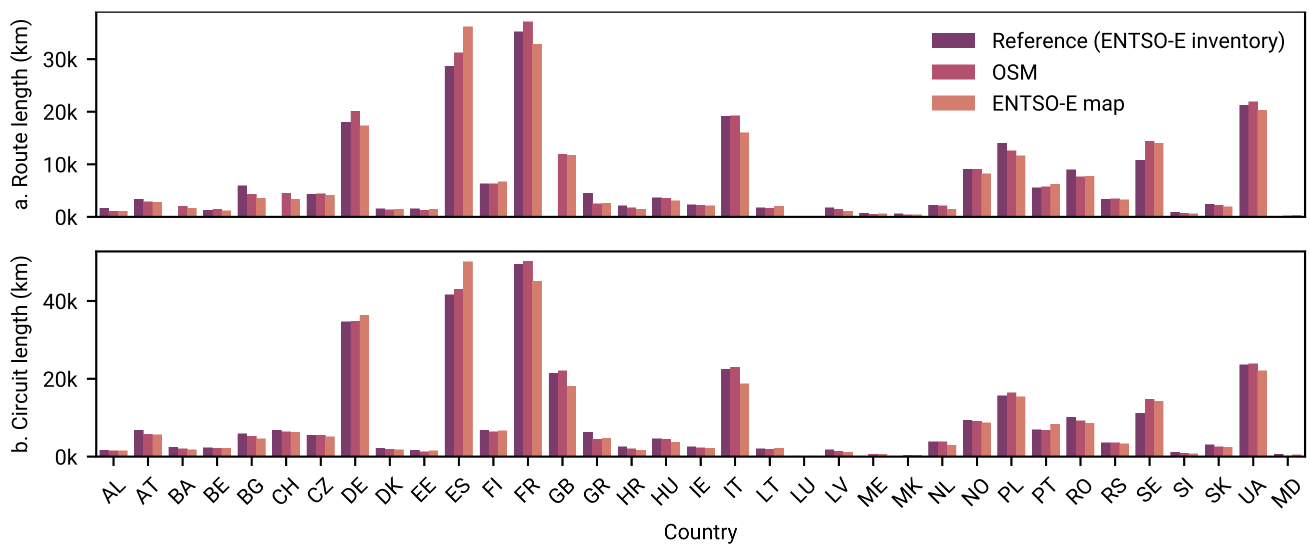

Figure 6. Comparison of total route and circuit lengths per country.

We find that our transmission grid based on OSM data is in agreement with the ENTSO-E inventory. Calculating the Pearson correlation coeffficient 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 ($\rho_{routes} = 0.9497$ and $\rho_{circuits} = 0.9862$) to OSM ($\rho_{routes} = 0.9636$ and $\rho_{circuits} = 0.9980$) in the reproduction of the high-voltage grid (220 kV 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 (Figure 7.a). Here we see strong similarities between the OSM-based and the extracted transmission grid of the ENTSO-E map, with few outliers ($\rho_{MVAkm} = 0.9484$). 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_{nom}^{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 (Figure 7.b) and number of line crossings (Figure 7.c) per NUTS1 border (Figure 7.b). While the two transmission grids strongly correlate, we observe notable differences at individual borders ($\rho_{MVA} = 0.8491$), the same is true for the absolute number of line crossings ($\rho_{crossings} = 0.8573$). 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 (Figure 7.b). 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

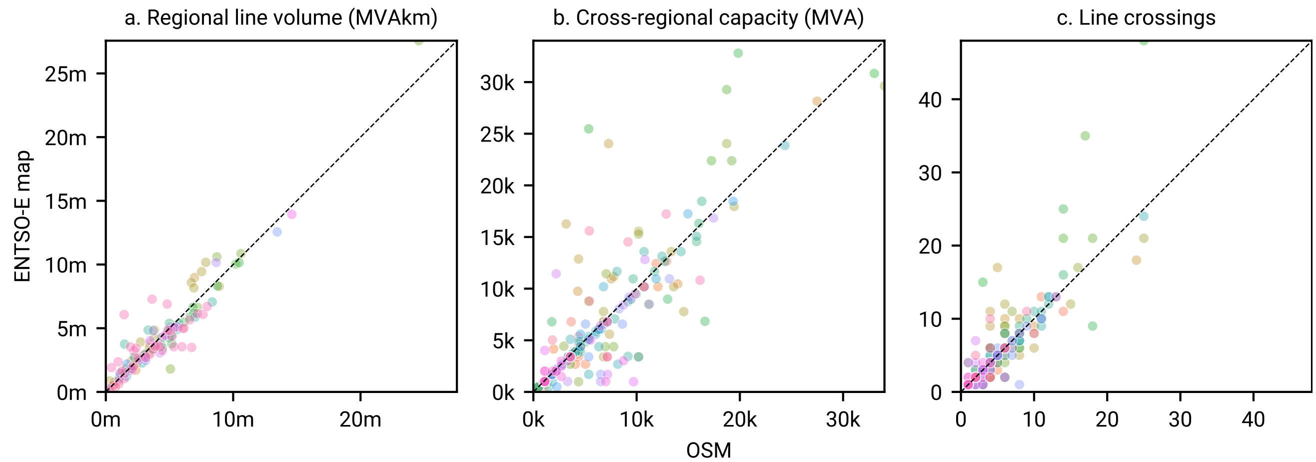

Figure 7. Comparison of line volume per NUTS1 region — Colors represent individual countries. *Line volume is the product of the nominal capacity and the length, summed over all lines within the region.*

data. In the case of Spain, 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 Spain[41] 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-Eur.[12] Note that the results shown in this publication are based on version 0.3 of the released prebuilt high-voltage grid representation on Zenodo.[29] 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. 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 CO_2 emitted.

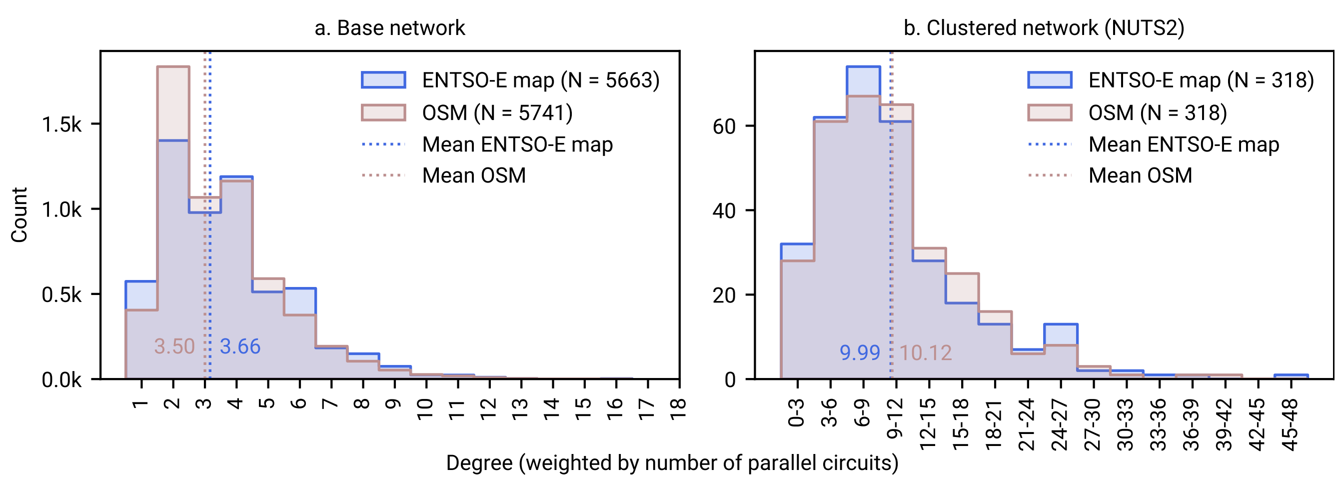

Figure 8. Comparison of the weighted degree distribution in both transmission grid representations before and after clustering (NUTS2). *Ukraine at geoBoundaries[42] administration level 1, Moldova in full bus resolution. A comparison at NUTS3 resolution is provided in Figure 11.*

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 used,[43] 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 (geoBoundaries[42]) 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 8 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:

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

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 $\rho_{degree,base} = 0.8518$). 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 $\rho_{clustered,base} = 0.8769$ (using unbinned data), $\rho_{clustered,base} = 0.9937$ (using binned data) — indicating that at NUTS2 resolution, the two networks are very similar in terms of connectivity and adequately represent the real grid.

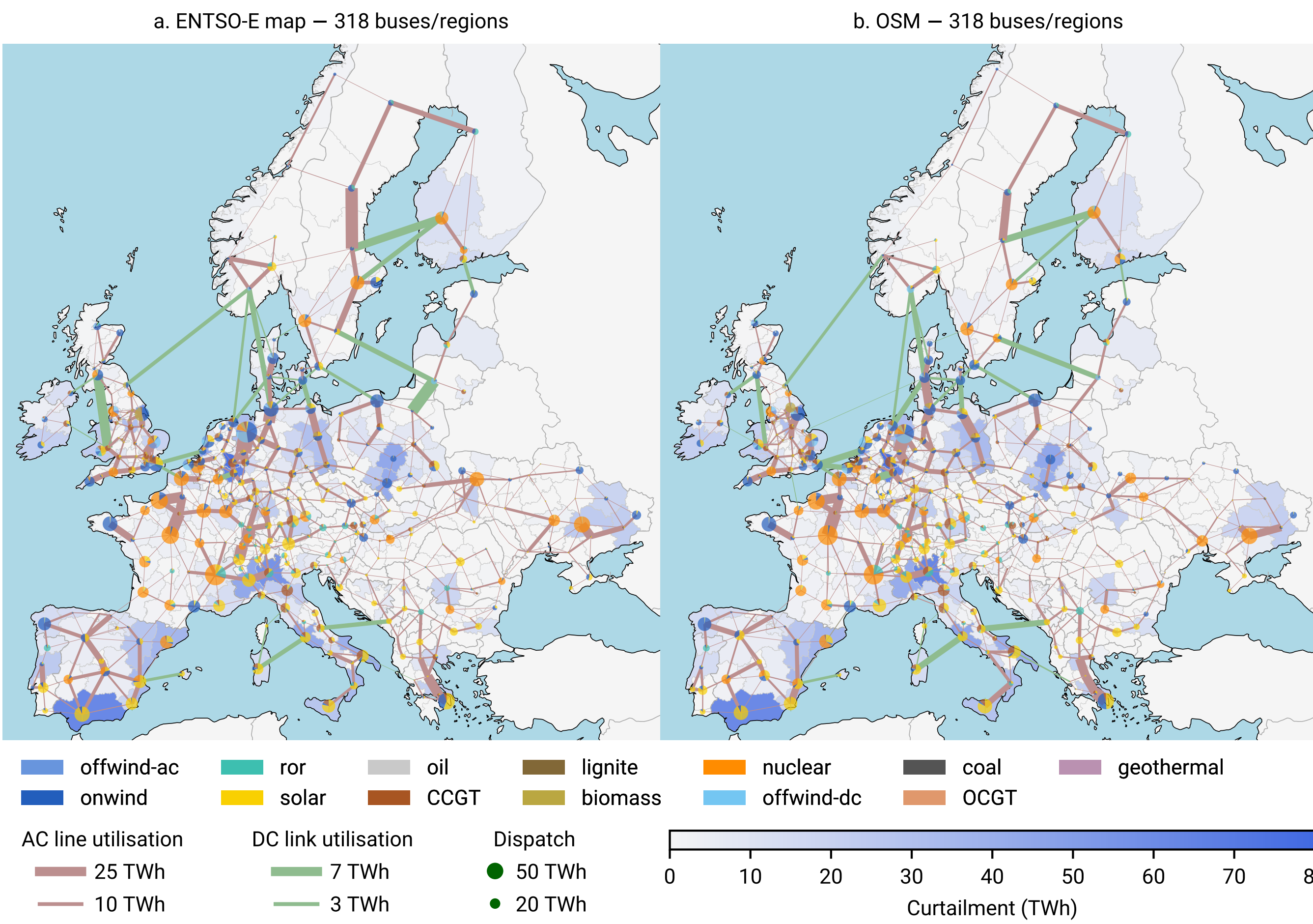

Figure 9. Regional dispatch, line utilisation and curtailment. *A map comparing nominal ratings of the two clustered grids is provided in the Figure 12.*

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 274 iterations to converge, translating into more than 18 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 (Figure 9). This is also true for the aggregated picture (Table 4). Total system costs drop from 312.94 bn. € in the ENTSO-E map based run to 311.7 bn. € in the OSM-based optimisation, corresponding to mere 0.40 % in difference. The higher weighted degree for both the base and clustered OSM-based high-voltage grid (Figure 8) 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 i) the line and link utilisation and ii) 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 Figure 10. 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 (especially decentral, semi-circles in the bottom half) and storage capacities, i.e. +4.9 GW in solar photovoltaics, +4.8 GW in onshore wind, +1.1 GW in offshore wind (AC), and +2.4 GW in battery storage. In the OSM-based model run, we see a slightly stronger build-out of DC connected offshore wind (+2.7 GW). We also provide an overview of average electricity prices, CAPEX and OPEX at nodal level in Figure 13.

	System costs (bn. €/a)	**CAPEX** (bn. €/a)	**OPEX** (bn. €)/a	**Curtailment** (TW h/a)	**Generation** (TW h/a)
ENTSO-E map	312.94	274.46	38.48	2177.91	3120.0
OSM	311.7	273.47	38.24	2175.57	3110.03
Delta (%)	-0.4	-0.36	-0.64	-0.11	-0.32

Table 4. Comparison of key result metrics between ENTSO-E map and OSM-based transmission grid.

Bottlenecks in both model runs are located in the same regions, contributing to an annual curtailment in the range of 2176 TW h to 2178 TW h. More prominent differences in line utilisation are visible in Norway and Poland from North to South, in the western region of Ukraine, southern and central Spain around the Madrid area, as well as southern Italy.

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 shown that the dataset is in good agreement with official statistics and the ENTSO-E map, and the results of a representative PyPSA-Eur model instance based on the two high-voltage grid datasets are very similar. Its core strengths lie in 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 provided openly. While we cannot guarantee the correctness of the data, as only TSOs have access to the real grid data, we believe that the dataset provides the best publicly available representation of the European high-voltage grid.

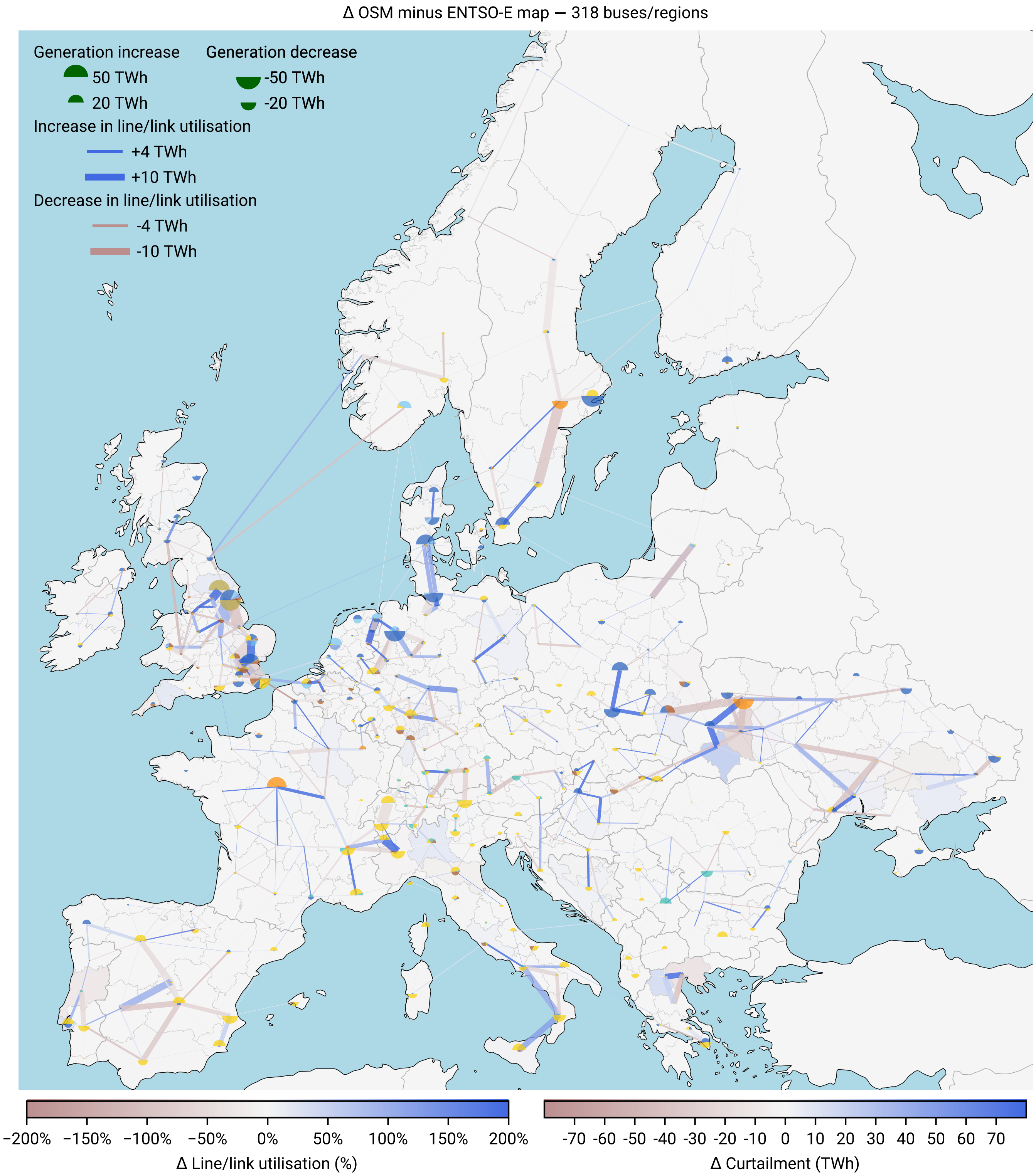

Figure 10. Regional dispatch, line utilisation and curtailment. *Blue indicates an increase in curtailment or line utilisation from the ENTSO-E map to the OSM-based transmission grid, while red indicates a decrease. For full transparency, note that this map shows an outer join of all transmission grid elements, including lines and links that are not present in the other network.*

Usage Notes

The published dataset is provided under the Open Data Commons Open Database License (ODbL) 1.0 licence. Geoinformation is encoded in the WGS84 (EPSG:4326) coordinate system. Although 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 an input for other energy 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 with the latest OSM data, the configuration file 'config.yaml': needs to be set to `base_network = 'osm-prebuilt'` 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 --call`.
- Per default, buses within the perimeter of a 5000 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.

To abide to the fair use policy of the OSM Overpass turbo API, we kindly encourage users to download the prebuilt network topology from the Zenodo repository and only rebuild the network, if necessary. We would further like to encourage readers to actively contribute to the OSM database.

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-Eur[12] on GitHub:
  https://github.com/pypsa/pypsa-eur
- Version 0.3 of the prebuilt network[29] based on OSM data can be retrieved via the Zenodo repository. This link will also point to future updates:
  https://zenodo.org/records/13358976

References

1. 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, 10.1109/EEM.2017.7982024 (2017).
2. ENTSO-E. ENTSO-E Transmission System Map. https://www.entsoe.eu/data/map/.
3. 50Hertz. Static grid model. https://www.50hertz.com/Transparency/GridData/Gridfigures/Staticgridmodel/ (2022).
4. JAO. Static Grid Model. https://www.jao.eu/static-grid-model (2023).
5. 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).
6. 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, 10.1109/PTC.2013.6652178 (2013).
7. Medjroubi, W., Müller, U. P., Scharf, M., Matke, C. & Kleinhans, D. Open Data in Power Grid Modelling: New Approaches Towards Transparent Grid Models. Energy Reports **3**, 14–21, 10.1016/j.egyr.2016.12.001 (2017).
8. Raifer, M. Overpass turbo. https://overpass-turbo.eu/ (2024).
9. Garrett, R. Open Infrastructure Map. https://openinframap.org (2024).
10. Wiegmans, B. GridKit extract of ENTSO-E interactive map, 10.5281/zenodo.55853 (2016).
11. Parzen, M. et al. PyPSA-Earth. A new global open energy system optimization model demonstrated in Africa. Appl. Energy **341**, 121096, 10.1016/j.apenergy.2023.121096 (2023).

12. Hörsch, J., Hofmann, F., Schlachtberger, D. & Brown, T. PyPSA-Eur: An open optimisation model of the European transmission system. Energy Strateg. Rev. **22**, 207–215, 10.1016/j.esr.2018.08.012 (2018).

13. Brown, T., Hörsch, J. & Schlachtberger, D. PyPSA: Python for Power System Analysis. J. Open Res. Softw. **6**, 10.5334/jors.188 (2018).

14. Gotzens, F., Heinrichs, H., Hörsch, J. & Hofmann, F. Performing energy modelling exercises in a transparent way - The issue of data quality in power plant databases. Energy Strateg. Rev. **23**, 1–12, 10.1016/j.esr.2018.11.004 (2019).

15. Hofmann, F., Hampp, J., Neumann, F., Brown, T. & Hörsch, J. Atlite: A Lightweight Python Package for Calculating Renewable Power Potentials and Time Series. J. Open Source Softw. **6**, 3294, 10.21105/joss.03294 (2021).

16. Neumann, F., Zeyen, E., Victoria, M. & Brown, T. The potential role of a hydrogen network in Europe. Joule **7**, 1793–1817, 10.1016/j.joule.2023.06.016 (2023).

17. Victoria, M., Zeyen, E. & Brown, T. Speed of technological transformations required in Europe to achieve different climate goals. Joule **6**, 1066–1086, 10.1016/j.joule.2022.04.016 (2022).

18. Brown, T. & Hampp, J. Ultra-long-duration energy storage anywhere: Methanol with carbon cycling. Joule **7**, 2414–2420, 10.1016/j.joule.2023.10.001 (2023).

19. Glaum, P., Neumann, F. & Brown, T. Offshore power and hydrogen networks for Europe's North Sea. Appl. Energy **369**, 123530, 10.1016/j.apenergy.2024.123530 (2024).

20. Riepin, I. & Brown, T. On the means, costs, and system-level impacts of 24/7 carbon-free energy procurement. Energy Strateg. Rev. **54**, 101488, 10.1016/j.esr.2024.101488 (2024).

21. Rahdan, P., Zeyen, E., Gallego-Castillo, C. & Victoria, M. Distributed photovoltaics provides key benefits for a highly renewable European energy system. Appl. Energy **360**, 122721, 10.1016/j.apenergy.2024.122721 (2024).

22. Grochowicz, A., van Greevenbroek, K. & Bloomfield, H. C. Using power system modelling outputs to identify weather-induced extreme events in highly renewable systems. Environ. Res. Lett. **19**, 054038, 10.1088/1748-9326/ad374a (2024).

23. TransnetBW. Stromnetz 2050 - Eine Studie der TransnetBW. Tech. Rep., TransnetBW GmbH (2022).

24. Barnes, T., Shivakumar, A., Brinkerink, M. & Niet, T. OSeMOSYS Global, an open-source, open data global electricity system model generator. Sci. Data **9**, 623, 10.1038/s41597-022-01737-0 (2022).

25. Glaum, P. & Hofmann, F. Leveraging the existing German transmission grid with dynamic line rating. Appl. Energy **343**, 121199, 10.1016/j.apenergy.2023.121199 (2023).

26. ENTSO-E. Ten-Year Network Development Plan (TYNDP) 2020 Main Report. Tech. Rep., ENTSO-E (2020).

27. BNetzA. Bestätigung des Netzentwicklungsplan Strom - NEP 2037/2045 (2023). Tech. Rep., Bundesnetzagentur (2024).

28. Mölder, F. et al. Sustainable data analysis with Snakemake, 10.12688/f1000research.29032.2 (2021). 10:33.

29. Xiong, B., Neumann, F. & Brown, T. Prebuilt Electricity Network for PyPSA-Eur based on OpenStreetMap Data, 10.5281/zenodo.12799201 (2024).

30. McKinney, W. Data Structures for Statistical Computing in Python. Proc. 9th Python Sci. Conf. 56–61, 10.25080/Majora-92bf1922-00a (2010).

31. Jordahl, K. et al. Geopandas/geopandas: V0.8.1. Zenodo, 10.5281/zenodo.3946761 (2020).

32. Kirschen, D. S. Power Systems: Fundamental Concepts and the Transition to Sustainability (John Wiley & Sons, 2024).

33. Pierri, E., Binder, O., Hemdan, N. G. A. & Kurrat, M. Challenges and opportunities for a European HVDC grid. Renew. Sustain. Energy Rev. **70**, 427–456, 10.1016/j.rser.2016.11.233 (2017).

34. Oeding, D. & Oswald, B. R. Elektrische Kraftwerke und Netze (Springer, Berlin, Heidelberg, 2016).

35. Thurner, L. et al. Pandapower—An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems. IEEE Transactions on Power Syst. **33**, 6510–6521, 10.1109/TPWRS.2018.2829021 (2018).

36. Shokri Gazafroudi, A., Neumann, F. & Brown, T. Topology-based approximations for N-1 contingency constraints in power transmission networks. Int. J. Electr. Power & Energy Syst. **137**, 107702, 10.1016/j.ijepes.2021.107702 (2022).

37. ENTSO-E. Inventory of Transmission - 2023 data. https://www.entsoe.eu/publications/data/power-stats (2024).

38. CIGRE. The Power System of Ukraine (2018).

39. GlobalData. Top five transmission line projects in the Ukraine (2023).

40. Moldelectrica. Technical and economic indicators. https://moldelectrica.md/ro/network/annual_report (2023).

41. Instituto Geografico Nacional. Energía - Mapa de red eléctrica española. 2016. http://atlasnacional.ign.es/wane/Energ%C3%ADa (2016).

42. Runfola, D. et al. geoBoundaries: A global database of political administrative boundaries. PLOS ONE **15**, e0231866, 10.1371/journal.pone.0231866 (2020).

43. 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, 10.1186/s42162-022-00187-7 (2022).

44. Egerer, J. Open Source Electricity Model for Germany (ELMOD-DE). Data Documentation (2016).

45. 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. 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 statement

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.

Competing interests

The authors declare no competing interests.

Figures & Tables

Grid element	Overpass turbo query
AC power lines and cables	way[‘power’=‘line’] and way[‘power’=‘cable’]
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	Geo-referenced	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[44]	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[45]	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	August 2024 (updated frequently)	Yes	Yes[29] (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 Figure 14.*

OSM relation 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	90
relation/3918230	Baltic Cable	DE	SE	450	600	260
relation/14126301	BritNed	GB-ENG	NL	450	1000	254
relation/8099179	Caithness-Moray Link	GB-SCT	GB-SCT	320	1200	136
relation/17631956	Conexión Mediterránea Transporte Alta tensión (COMETA)	ES	ES	250	400	238
relation/8185420	Copenhagen-Brussels-Amsterdam cable (COBRAcable)	DK	NL	320	700	328
relation/2127794	Cross-Channel (IFA 2000)	FR	GB-ENG	270	2000	70
relation/8184641	East-West Interconnector	IE	GB-WLS	400	500	257
relation/15772117	ElecLink	FR	GB-ENG	320	1000	53
relation/8184630	Estlink 1	EE	FI	150	350	105
relation/6886400	Estlink 2	EE	FI	450	650	175
relation/8184631	Fenno-Skan	FI	SE	400	500	229
relation/3391954	Fenno-Skan 2	FI	SE	500	800	305
relation/3392001	Gotlandslänken (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	231
relation/8185664	Italy-Greece	GR	IT	400	500	302
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	241
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	412
relation/6914309	Moyle interconnector	GB-NIR	GB-SCT	500	500	62
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	448
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	417
relation/3391931	Skagerrak (1-3)	DK	NO	350	940	241
relation/8184632	Skagerrak (4)	DK	NO	500	700	226
relation/3392010	SwePol	PL	SE	450	600	255
relation/8184633	SydVästlänken	NO	SE	300	1200	253
relation/15781671	Viking Link	DK	GB-ENG	525	1400	742
relation/15775538	Western HVDC Link	GB-SCT	GB-WLS	600	2250	401

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.*

U_{nom}^{OSM} (kV)	**Line type**	$I_{nom}^{pandapower}$ (A)	$n_{circuits}$	S_{nom}^{AC} (MV A)	**Total route length** (km)
220	Al/St 240/40 2-bundle 220.0	1.290	1	492	52897
220	Al/St 240/40 2-bundle 220.0	1.290	2	983	22719
220	Al/St 240/40 2-bundle 220.0	1.290	3	1475	524
220	Al/St 240/40 2-bundle 220.0	1.290	4	1966	37
225	Al/St 240/40 2-bundle 220.0	1.290	1	503	19355
225	Al/St 240/40 2-bundle 220.0	1.290	2	1005	4465
236	Al/St 240/40 2-bundle 220.0	1.290	1	527	19
275	Al/St 240/40 3-bundle 300.0	1.935	1	922	1097
275	Al/St 240/40 3-bundle 300.0	1.935	2	1843	2845
300	Al/St 240/40 3-bundle 300.0	1.935	1	1005	4127
300	Al/St 240/40 3-bundle 300.0	1.935	2	2011	20
330	Al/St 240/40 3-bundle 300.0	1.935	1	1106	17335
330	Al/St 240/40 3-bundle 300.0	1.935	2	2212	1115
380	Al/St 240/40 4-bundle 380.0	2.580	1	1698	13971
380	Al/St 240/40 4-bundle 380.0	2.580	2	3396	13989
380	Al/St 240/40 4-bundle 380.0	2.580	3	5094	362
380	Al/St 240/40 4-bundle 380.0	2.580	4	6792	259
400	Al/St 240/40 4-bundle 380.0	2.580	1	1787	56750
400	Al/St 240/40 4-bundle 380.0	2.580	2	3575	30977
400	Al/St 240/40 4-bundle 380.0	2.580	4	7150	84
412	Al/St 240/40 4-bundle 380.0	2.580	1	1841	32
420	Al/St 240/40 4-bundle 380.0	2.580	1	1877	4839
420	Al/St 240/40 4-bundle 380.0	2.580	3	5631	11
500	Al/St 240/40 4-bundle 380.0	2.580	1	2234	248
750	Al/St 560/50 4-bundle 750.0	4.160	1	5404	4148

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

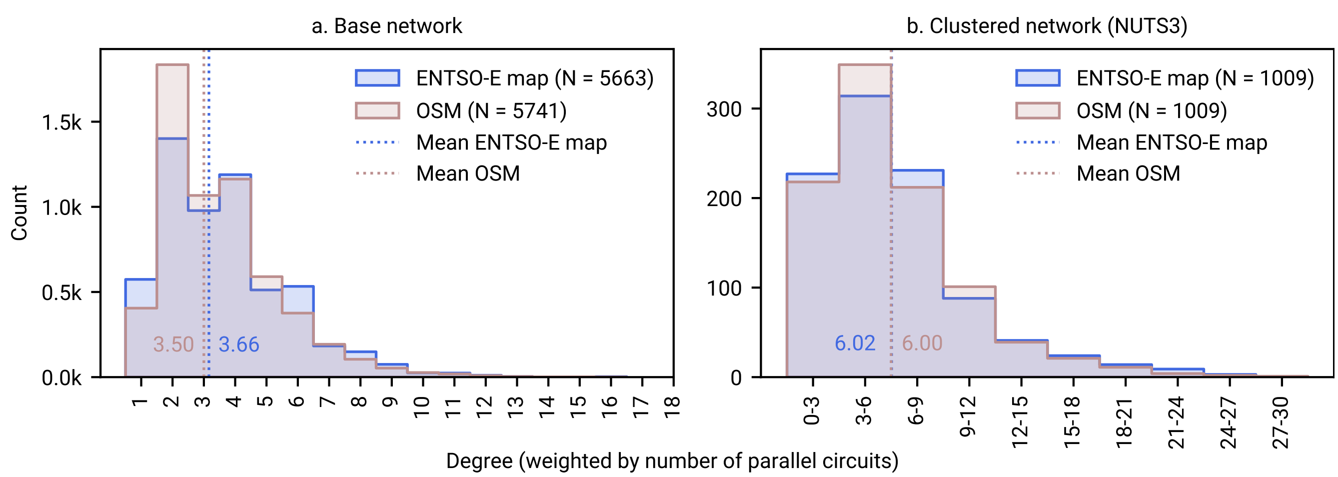

Figure 11. Comparison of the weighted degree distribution in both networks before and after clustering (NUTS3). *Ukraine at geoBoundaries*[42] *administration level 1, Moldova in full bus resolution.*

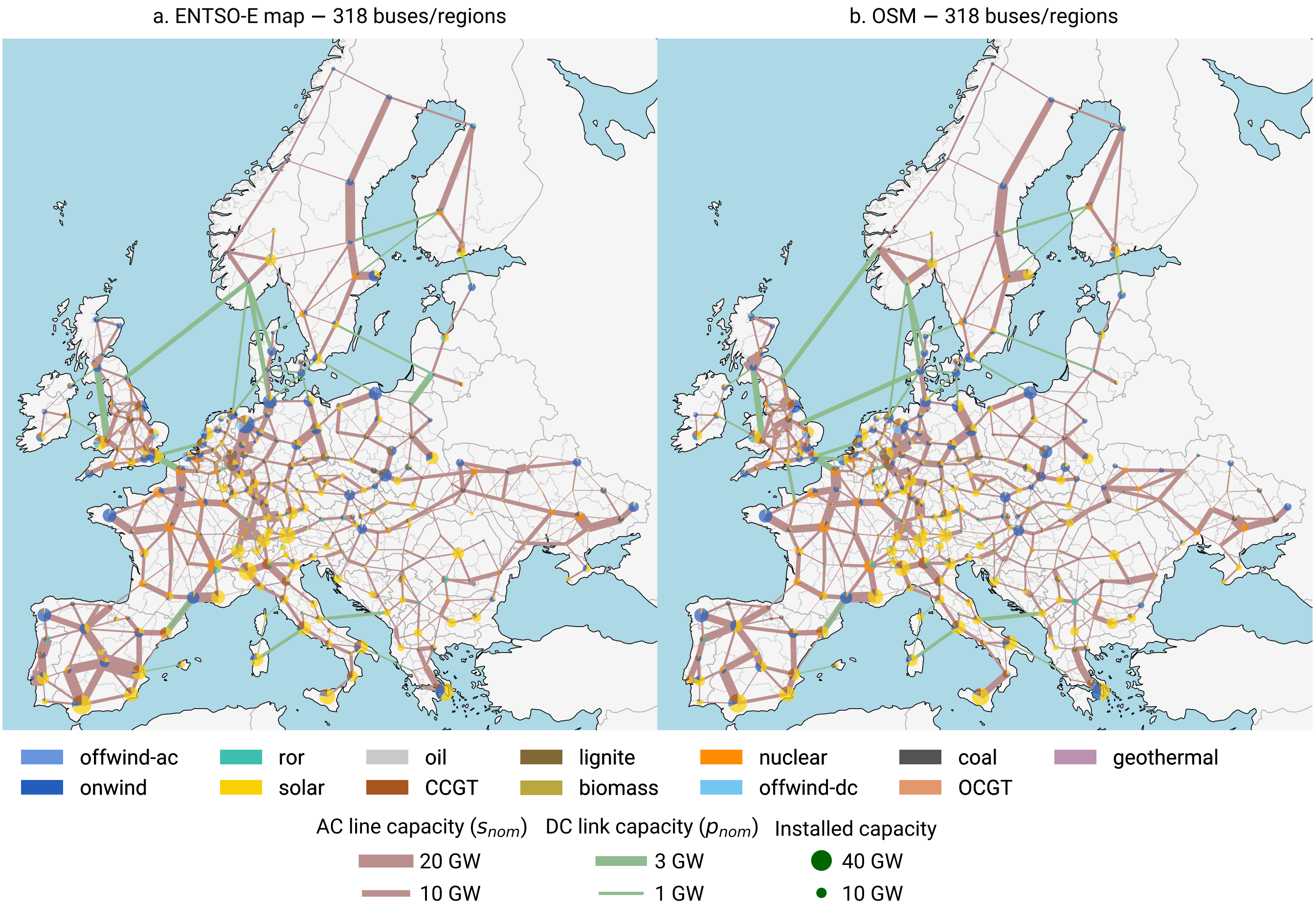

Figure 12. Clustered AC line capacities, DC link nominal ratings, and optimal generation capacities.

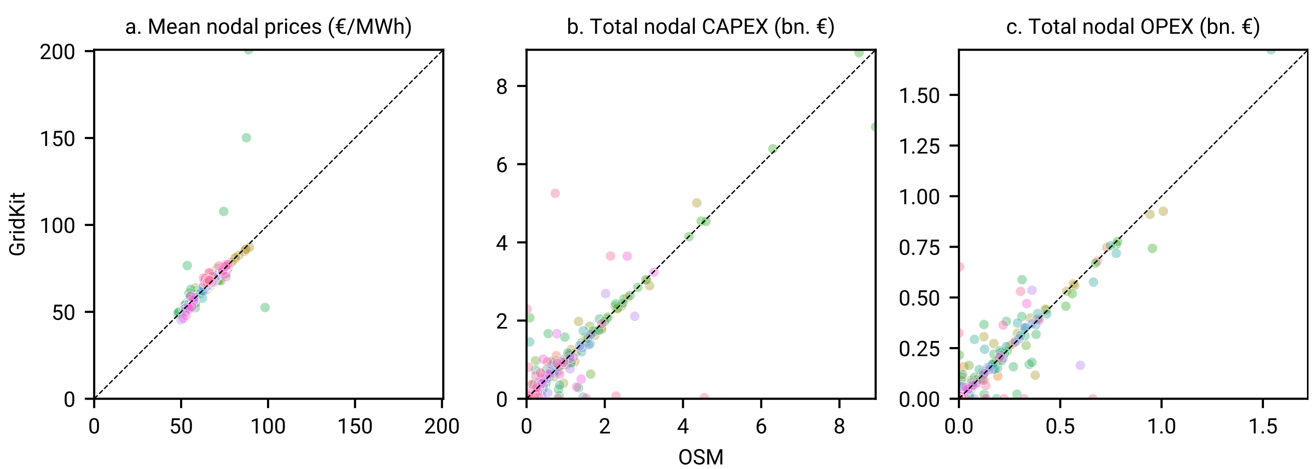

Figure 13. Comparison of nodal prices, capital expenditures (CAPEX) and operational expenditures (OPEX). *Outliers nodal prices (a): Great-Britain (green). Outliers nodal CAPEX (b): Ukraine (pink).*

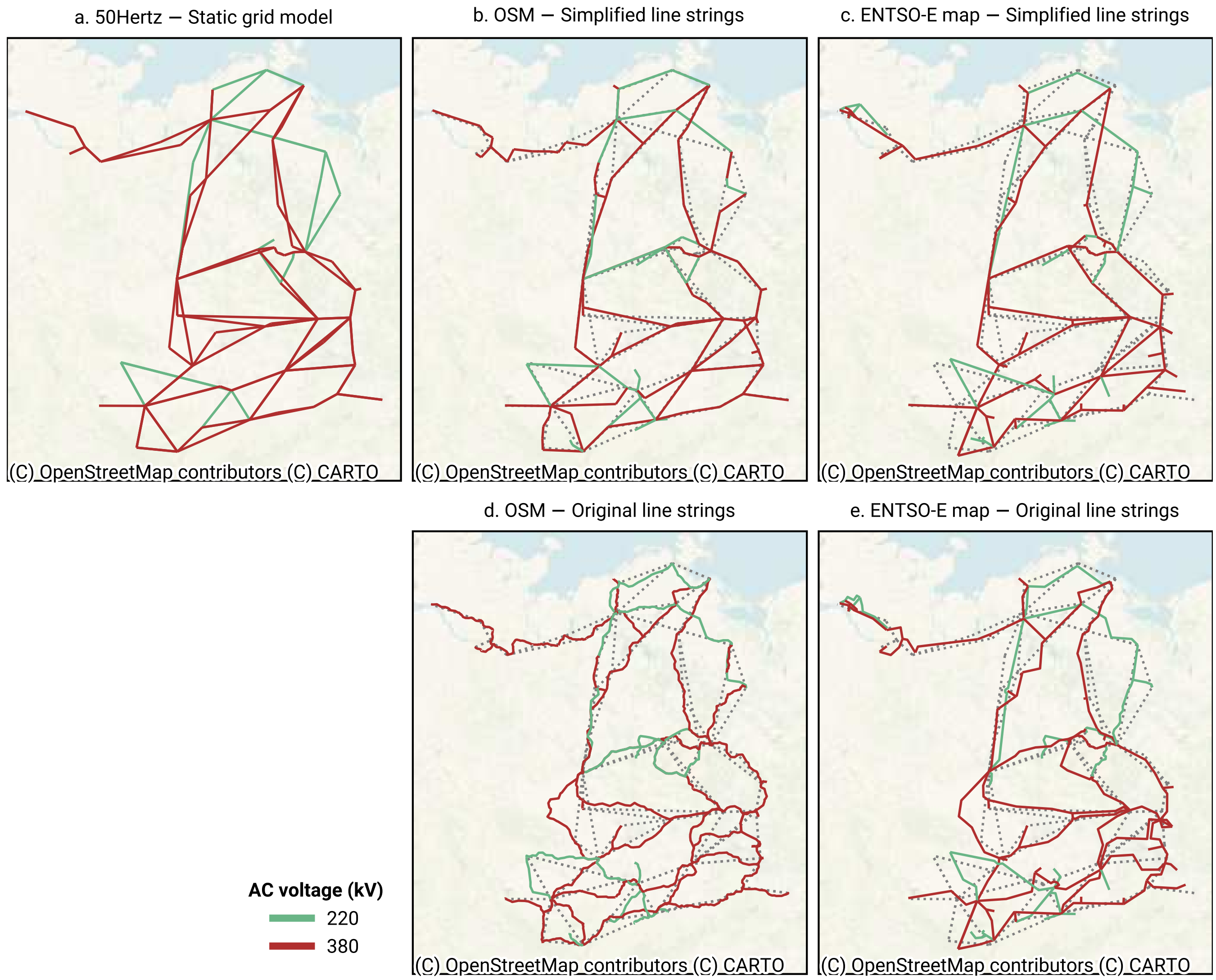

Figure 14. Comparison of OSM and ENTSO-E map-based transmission grid with reference 50Hertz static grid model.[3] *Dashed grey lines underneath show the 50Hertz static grid model for comparative purposes. Note that the geospatial data is provided in simplified, point-to-point form, only. Many of the dashed grey lines are also included in the OSM or ENTSO-E map-based grid representations, however, intermediary buses exists along the line. As such, in their simplification they are not simplified to the level of the 50Hertz static grid model.*

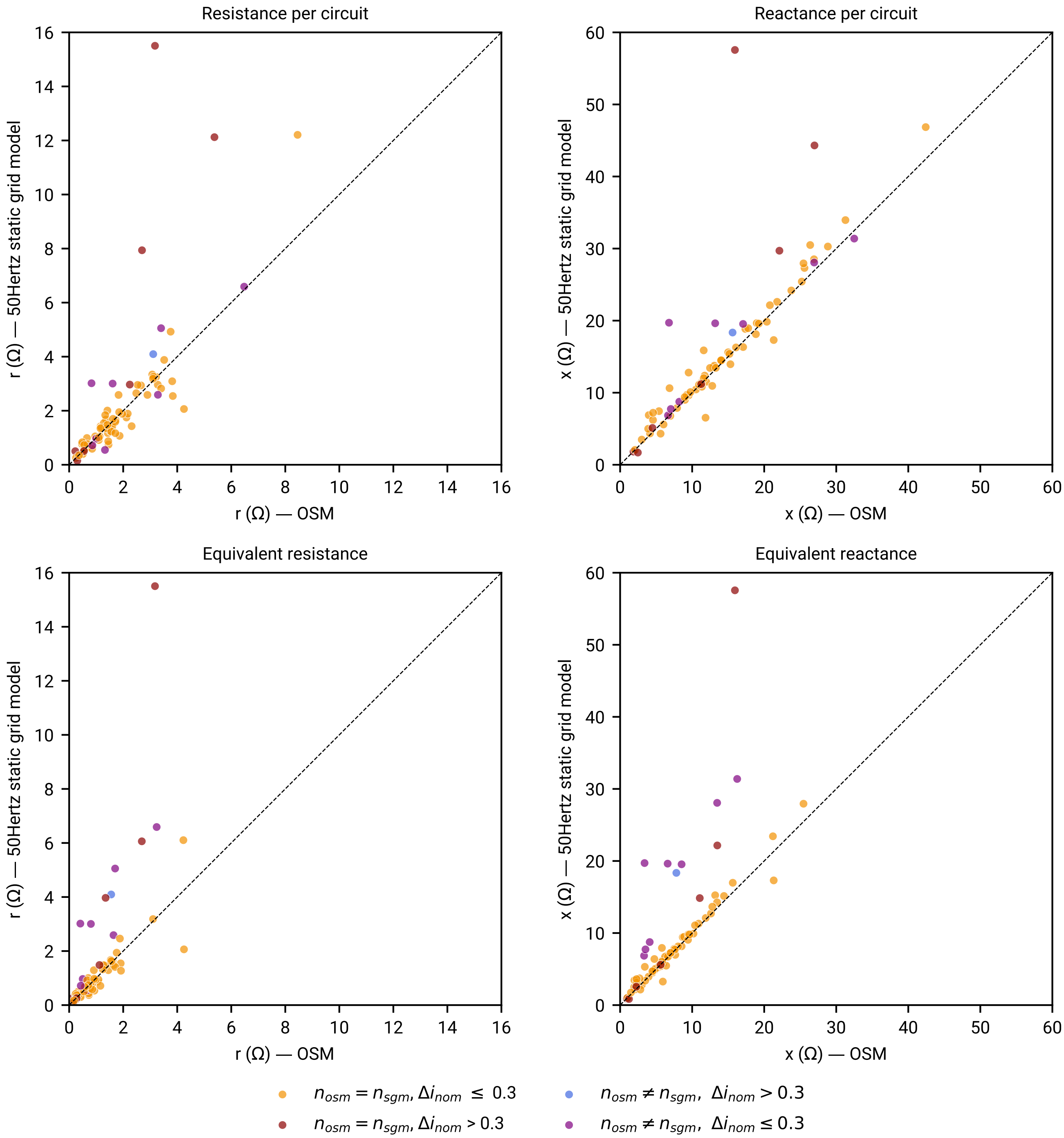

Figure 15. Comparison of AC line/cable resistance and reactance between the OSM-based transmission grid and reference 50Hertz static grid model.[3] *n_{osm} and n_{sgm} refer to the number of parallel circuits for a distinct line in each network, while $\Delta i_{nom} = \frac{|i_{nom,osm} - i_{nom,sgm}|}{i_{nom,sgm}}$ refers to the relative change in underlying nominal current.*

Figure 15 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 types[35] for calculating the resistance and reactance comes close to official data in the SGM (orange). Purple data points a discrepance 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 where SGM e.g. has a newer 380 kV (allowing higher higher currents) or weaker 220 kV line type.