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

Journal Club DTU WIND — Department of Wind and Energy Systems

Bobby Xiong*, Davide Fioriti, Fabian Neumann, Iegor Riepin & Tom Brown
*presenting author: xiong@tu-berlin.de
Technische Universität Berlin, Germany

July 4, 2025

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

RESILIENT — Work packages

WP1 – TUB
Project Leadership

WP2
Methods for Resilient Planning under Strategic Uncertainties

- Development of stochastic optimisation framework SMS++
- Development of multi-vector energy system model PyPSA-Eur-Sec

WP3
Datasets and Model Improvements on Industry, Biomass and E-Fuels

- Industry Transition Paths: Fuel and Process Switching
- Carbon Management and the Role of Biomass
- Global Green Fuel and Material Markets

WP4
Case Studies and Model Demonstrations for Need-Owners

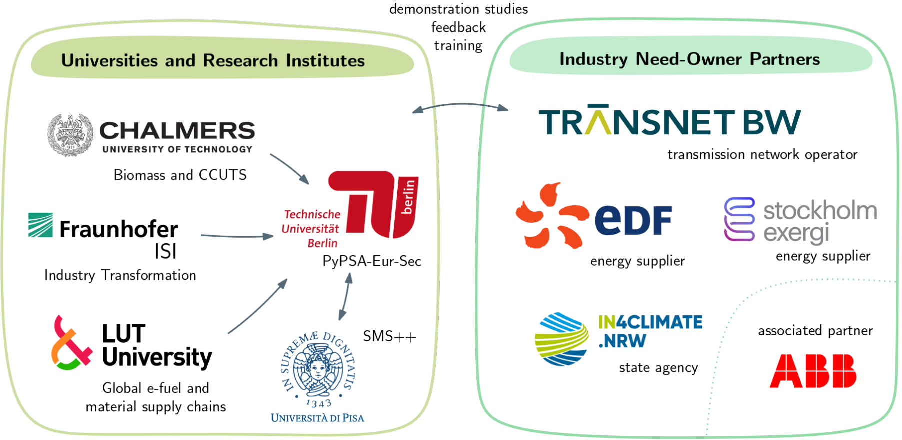
- France's future energy system in the European network
- Grid planning and industry transition in Western Germany
- Carbon and e-fuel strategies for Sweden and Finland

WP5
Outreach, Communication and Dissemination

- engagement with more need-owners
- training events and documentation

WP6
Reporting & Knowledge Community Standard WP

RESILIENT — Project partners



Funded via **CETPartnership 2022** Call — **BMWK** for all German partners.

Source: https://resilient-project.github.io/

Motivation

- Conclusion drawn from models only as good as underlying data and assumptions
- With few exceptions, existing datasets are either (i) not open or provide no licensing [4], (ii) not georeferenced [8], or (iii) are not up-to-date or too complex to update [3, 7, 9]
- Data provided on the ENTSO-E online map [4] is highly stylised in terms of line routes and topological connectivity. In some cases (e.g., Madrid, Spain), it contains clearly visible errors.


www.nature.com/scientificdata

scientific **data**

Check for updates

OPEN

DATA DESCRIPTOR

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

Bobby Xiong[1]✉, Davide Fioriti[2], Fabian Neumann[1], Iegor Riepin[1] & Tom Brown[1]

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 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) they are not released under an open licence, (iii) nor are they updated frequently, and (iv) their geographic detail is limited or highly stylised.

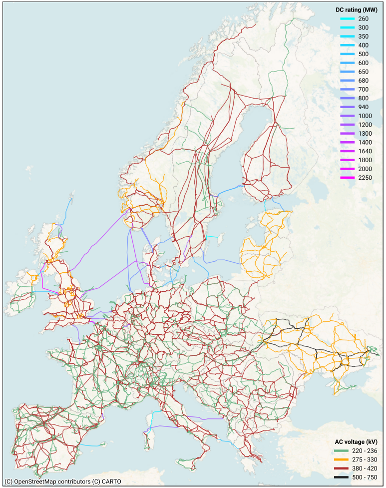
There are previous projects that have modelled the European high-voltage grid or its parts based on OSM data. Some institutions provide data for particular regions, however, all of them come with their own drawbacks: While the most trustworthy data comes from TSOs themselves, they are — with few regional 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 API[8] 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 OpenInfraMap[9] 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

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

Source: Screenshot taken from https://www.nature.com/articles/s41597-025-04550-7 [12]

Proposed solution

- Dataset [13] contains a topologically connected representation of the European high-voltage grid (220 kV to 750 kV) constructed using OpenStreetMap data, see OpenInfraMap for visualisation [5]
- Transparent, open-source and reproducible workflow completely implemented in Python/PyPSA-Eur
- By accessing the Overpass turbo API [10], the dataset can be updated anytime to include new AC and DC lines, substations, and transformers
- AC lines mapped using pandapower's standard line type library [11]. Nominal capacity is set to 70 % of the technical capacity to account for n-1 security approximation [2]
- Includes all 38 European HVDC connections with their nominal rating that are commissioned as of 2024



Source: Own illustration based on OSM data extracted using Overpass Turbo API https://openstreetmap.org

Technische Universität Berlin

Methodology

Process diagram

retrieve_osm_data.py

- overpass-turbo API requests to retrieve power features of OpenStreetMap per country

clean_osm_data.py

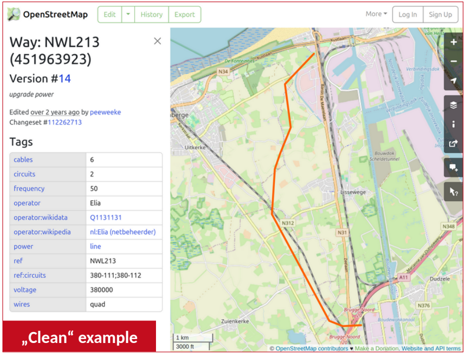
- Heuristic filtering + cleaning process to obtain single buses and lines (voltage, AC/DC, length, and number of circuits)

build_osm_network.py
(from PyPSA-Earth)

- Building the network on lines/links and aggregated buses;
- Constructing transformers + converters

base_network.py
(remaining PyPSA-EUR workflow)

- Creating PyPSA-ready base network base.nc
- TYNDP and NEP projects being added



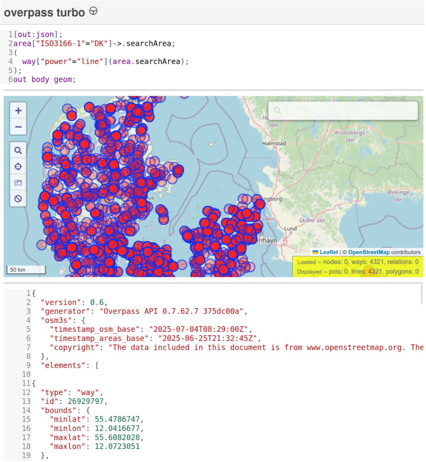
„Clean“ example

Source: Own illustration based on [12].

Steps — 1. Data retrieval

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: Overpass turbo API queries



Source: Screenshot taken from https://overpass-turbo.eu

Steps — 2. Data cleaning (before)

line id	**cables**	**circuits**	**frequency** (Hz)	**type**	**voltage** (V)
way/1		double	50	cable	380 kv
relation/1	3		50	cable	380 000 abc
way/3	9	1;2	50.0001	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: Illustrative example of AC lines and cables input data.

Source: Own illustration.

Technische Universität Berlin

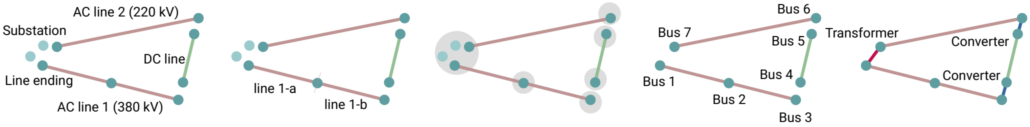
Steps — 2. Data cleaning (after)

line id	circuits	frequency (Hz)	type	voltage (V)
way/1	2	50	cable	380 000
relation/1	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: Illustrative example of AC lines and cables after cleaning. Changes highlighted in yellow.

Source: Own illustration.

Steps – 3. Network building: Creating a connected topology



a. Line endings are added to set of buses

b. Lines overpassing nodes are split into segments

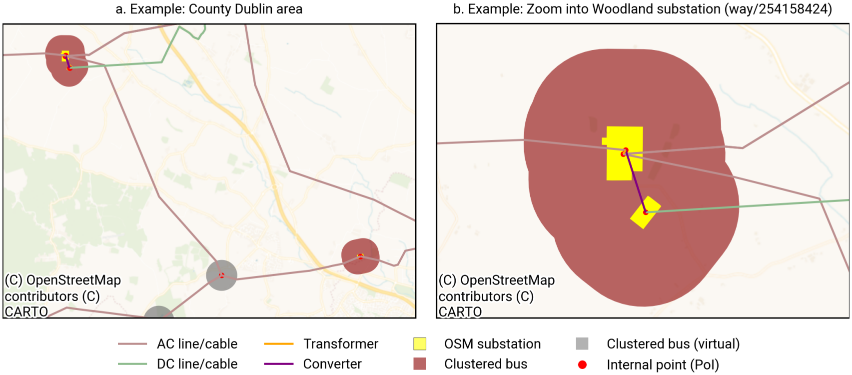
c. IDs are assigned to buses within tolerance of 500 m

d. Buses are aggregated by voltage and AC/DC type

e. Transformers (red) and converters (blue) are added

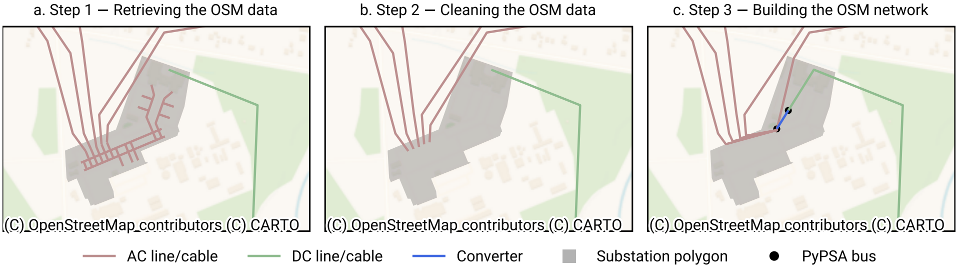
Source: Own illustration.

Steps — 3. Network building: Example zoom-in



Source: Own illustration.

Example of a completed section

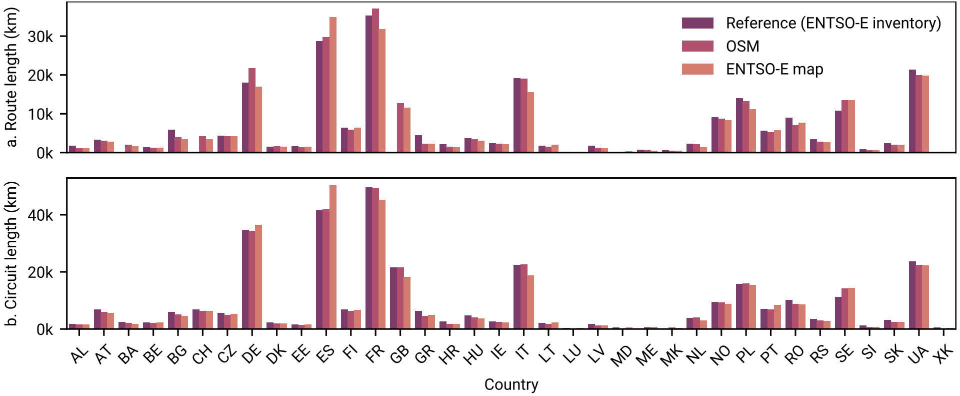


Source: Own illustration.

Results & Validation

Technische Universität Berlin

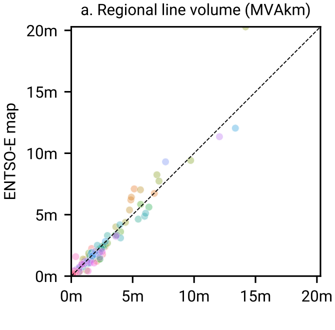
Route and circuit lengths



$\rightarrow$ Improvement from ($\rho_{routes} = 0.9489$, $\rho_{circuits} = 0.9862$) to ($\rho_{routes} = 0.9575$, $\rho_{circuits} = 0.9980$)

Source: Own illustration.

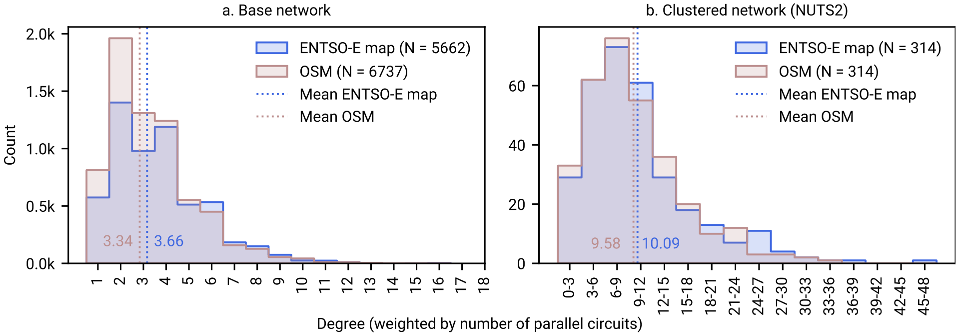
Line metrics based on original PyPSA paper [2]



→ At NUTS1, generally high correlation: $\rho_{MVAkm} = 0.9674$, $\rho_{MVA} = 0.8441$ $\rho_{crossings} = 0.8575$

Source: Own illustration.

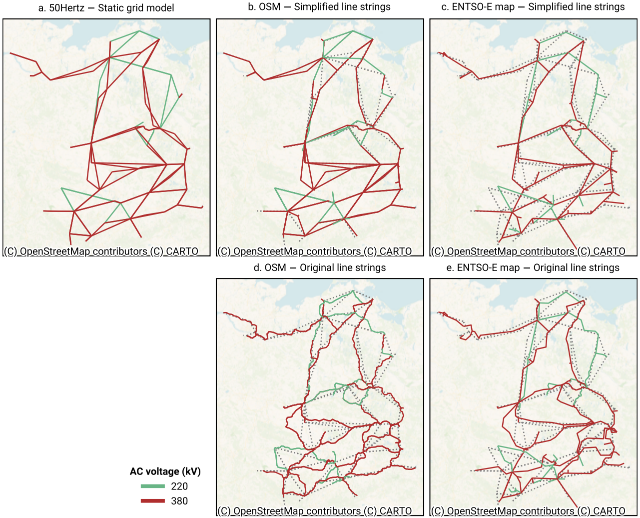
Graph-theoretic metrics using weighted degree distribution



In the base network, number of circuits can be tracked differently, i.e., (1×3) vs. (3×1) vs. $(2 \times 1 + 1)$. Hence, clustering accounts for different ways of counting and a ‘more accurate’ comparison.

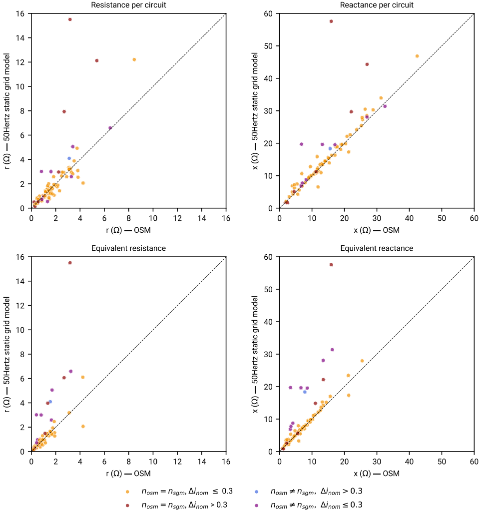
Source: Own illustration.

50Hertz SGM comparison — Topology



Source: Own illustration.

50Hertz SGM comparison — Electrical parameters

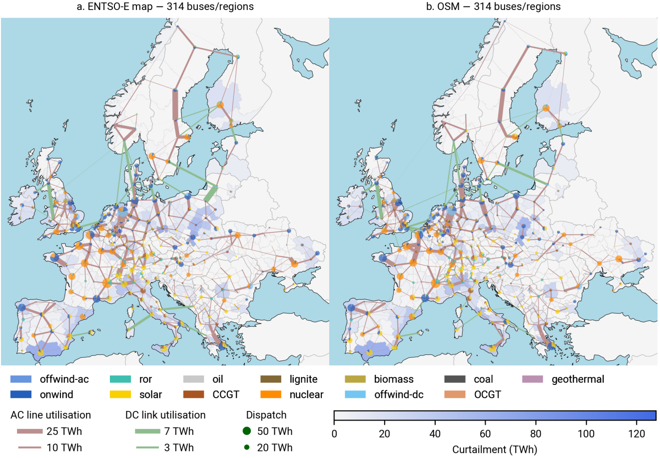


- Scatter plot generated by mapping AC lines and cables of the OSM-based transmission grid to the 50Hertz static grid model (SGM) [1] using OSM tags and SGM names (right join)
- Data explains 4475 km of 5126 km in route length, as not all lines could mapped
- For 79 % of the data, using pandapower's standard line types [11] for calculating the resistance and reactance comes close to official data in the SGM (orange)
- Purple data points show 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 in SGM which are of a newer/stronger 380 kV (allowing higher currents) or weaker 220 kV line type.

Source: Own illustration.

Technische Universität Berlin

Results from solving the electricity network model



Regional dispatch, line utilisation and curtailment.

Setup

- PyPSA-Eur electricity sector only for validation
- Year 2030, hourly resolution, clustered at NUTS2 (318 regions)
- No transmission line expansion allowed
- Generation and storage technologies expansion allowed
- CO_2 price set to 100 €/tCO_2

Source: Own illustration.

Results from solving the electricity network model

	System costs (bn. €)	CAPEX (bn. €)	OPEX (bn. €)	Curt. (TWh)	Gen. (TWh)
GridKit	319.28	283.47	35.81	2749.41	3119.82
OSM	318.19	283.09	35.10	2742.26	3118.50
Delta (%)	-0.34	-0.13	-1.99	-0.26	-0.04

Bottlenecks in both model runs are located in the same regions, contributing to an annual curtailment in the range of 2742 TW h to 2149 TW h. 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.

Setup

- PyPSA-Eur electricity sector only for validation
- Year 2030, hourly resolution, clustered at NUTS2 (318 regions)
- No transmission line expansion allowed
- Generation and storage technologies expansion allowed
- CO_2 price set to 100 €/tCO_2

Source: Own illustration.

Usage notes

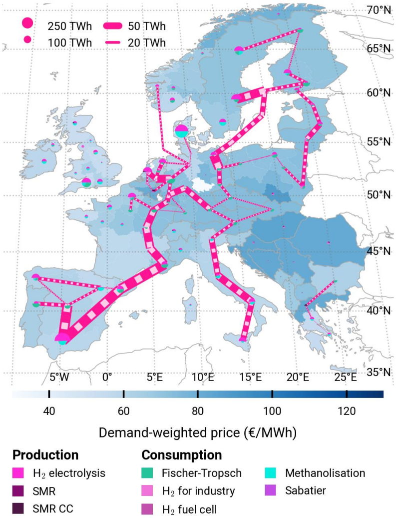
- In order to reproduce the network of this publication, the configuration file 'config.yaml': needs to be set to `base_network = 'osm-prebuilt'`, `osm-prebuilt-version: 0.6` and the command `snakemake base_network --call` needs to be executed.
- To rebuild the network from scratch, this setting can 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.

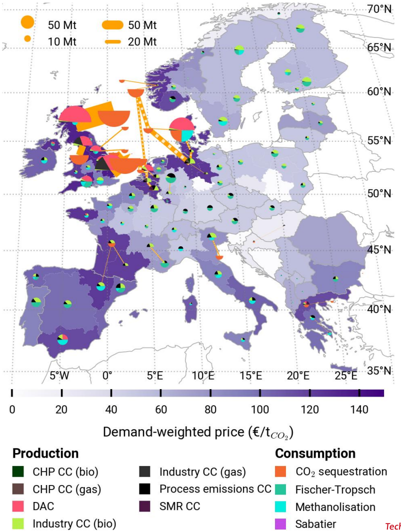
Code availability

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-Eur [6] on GitHub:
  https://github.com/pypsa/pypsa-eur
- Version 0.6 of the prebuilt network based on OSM data can be retrieved via the Zenodo repository [13]. This link will also point to future updates:
  https://doi.org/10.5281/zenodo.14144752
- An interactive map is bundled with the dataset on Zenodo[13]. It can also be directly accessed via the PyPSA-Eur documentation: https://pypsa-eur.readthedocs.io/en/latest/data-base-network.html

Ongoing work: Role of PCI-PMI CO_2 and H_2 projects





Source: Preprint uploaded to https://doi.org/10.48550/arXiv.2507.01860

References (excerpt)

[1] 50Hertz. *Static Grid Model*. Apr. 2022. (Visited on 07/31/2024).

[2] Thomas Brown, Jonas Hörsch, and David Schlachtberger. "PyPSA: Python for Power System Analysis". In: *Journal of Open Research Software* 6.1 (Jan. 2018). ISSN: 2049-9647. DOI: 10.5334/jors.188. (Visited on 05/30/2024).

[3] Jonas Egerer 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. (Visited on 07/31/2024).

[4] ENTSO-E. *ENTSO-E Transmission System Map*. https://www.entsoe.eu/data/map/. (Visited on 06/28/2024).

[5] Russ Garrett. *Open Infrastructure Map*. https://openinframap.org. 2024. (Visited on 07/08/2024).

[6] Jonas Hörsch et al. "PyPSA-Eur: An Open Optimisation Model of the European Transmission System". In: *Energy Strategy Reviews* 22 (Nov. 2018), pp. 207–215. ISSN: 2211-467X. DOI: 10.1016/j.esr.2018.08.012. (Visited on 03/20/2024).

[7] Neil Hutcheon and Janusz W. Bialek. "Updated and Validated Power Flow Model of the Main Continental European Transmission Network". In: *2013 IEEE Grenoble Conference*. June 2013, pp. 1–5. DOI: 10.1109/PTC.2013.6652178. (Visited on 07/30/2024).

[8] JAO. *Static Grid Model*. https://www.jao.eu/static-grid-model. Sept. 2023. (Visited on 07/30/2024).

[9] Wided Medjroubi et al. "Open Data in Power Grid Modelling: New Approaches Towards Transparent Grid Models". In: *Energy Reports* 3 (Nov. 2017), pp. 14–21. ISSN: 2352-4847. DOI: 10.1016/j.egyr.2016.12.001. (Visited on 06/06/2024).

[10] Martin Raifer. *Overpass Turbo*. https://overpass-turbo.eu/. 2024. (Visited on 07/01/2024).

[11] Leon Thurner et al. "Pandapower—An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems". In: *IEEE Transactions on Power Systems* 33.6 (Nov. 2018), pp. 6510–6521. ISSN: 1558-0679. DOI: 10.1109/TPWRS.2018.2829021. (Visited on 07/05/2024).

[12] Bobby Xiong et al. "Modelling the High-Voltage Grid Using Open Data for Europe and Beyond". In: *Scientific Data* 12.1 (Feb. 2025), p. 277. ISSN: 2052-4463. DOI: 10.1038/s41597-025-04550-7. (Visited on 02/19/2025).

[13] Bobby Xiong et al. *Prebuilt Electricity Network for PyPSA-Eur Based on OpenStreetMap Data. Version 0.6*. Nov. 2024. DOI: 10.5281/zenodo.14144752. (Visited on 11/13/2024).

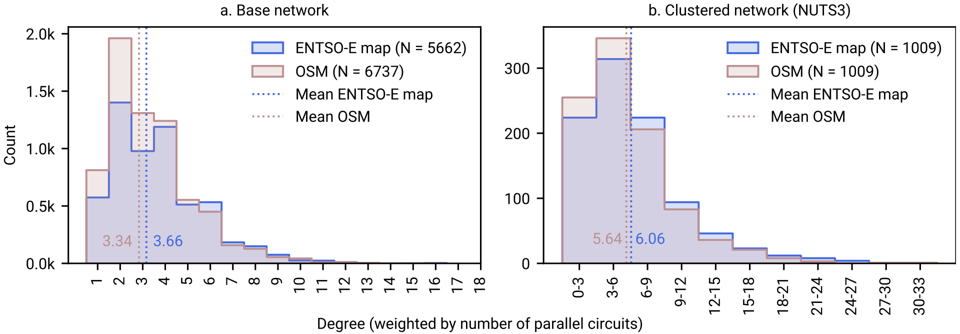
Thank you!

Potential collaborations? :)

Appendix

Technische Universität Berlin

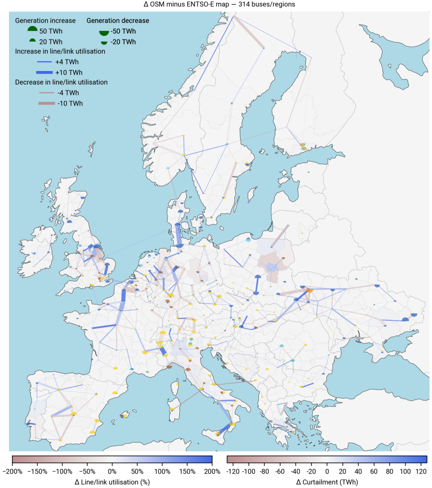
Graph-theoretic metrics using weighted degree distribution



In the base network, number of circuits can be tracked differently, i.e., (1×3) vs. (3×1) vs. $(2 \times 1 + 1)$. Hence, clustering accounts for different counting methods and a 'more accurate' comparison.

Source: Own illustration.

Delta regional dispatch, line utilisation and curtailment



Setup

- PyPSA-Eur electricity sector only for validation
- Year 2030, hourly resolution, clustered at NUTS2 (318 regions)
- No transmission line expansion allowed
- Generation and storage technologies expansion allowed
- CO_2 price set to 100 €/tCO_2

Source: Own illustration.