

Author Response to Reviews of

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

Bobby Xiong, Davide Fioriti, Fabian Neumann, Iegor Riepin, Tom Brown

Scientific Data — Nature

Revision #1: November 14, 2024

EC: Editor Comment, RC: Reviewer Comment, AR: Author Response, ■ *Manuscript text*

We thank the reviewers for their constructive comments and acknowledge the time and effort they have spent in assessing our work. We have revised the paper based on the feedback from the reviewers and hope that we can adequately address their primary concerns.

To follow the revisions made, we have highlighted differences compared with the previous submitted version of the paper in blue and red text in an attached file. Figure numbers in this response letter refer to the numbering in the revised version of the manuscript.

Editor comments

EC: Please provide an additional copy of your manuscript file under 20MB as a *Related Manuscript File* in your next submission for internal use.

AR: We will provide an additional copy of the manuscript file under 20 MB, as requested.

EC: A data descriptor provides a description of a static dataset, rather than any novel methods or computational validation schemes. Please edit the text of your abstract to focus more on the dataset that you have generated, rather than a method for constructing the dataset.

AR: Thank you for your feedback and for helping us to improve the clarity and focus of our manuscript. We have revised the abstract to highlight the dataset itself that we have generated, rather than the methodology for its construction.

EC: Because at least some of your data looks to originate from third parties, please check the following and confirm/state what has been done in your *response to reviewers* document:

1. All sources are clearly described in either the main text. For general mentions of services, databases, or other platforms, please name the resource in the main text and provide a URL in (). Instructions should be sufficient for the exact input data to be retrieved, so please

provide specific links, accession numbers, or the exact search term/query when discussing any data from general databases. For specific datasets downloaded from repositories, or other items with formal metadata, please use a full citation in the reference list containing the relevant DOI. Note this should be the dataset, rather than any relevant publication, however these should be cited as well if required, or if the data was newly extracted from a document.

2. Confirm all the sources are openly available — i.e. your re-use or re-distribution is compliant with the terms and conditions/licenses for data sharing. If you cannot see a fully open licence at the source please check with the data provider. Please note it is your legal responsibility to ensure all data have been used or re-distributed appropriately and we cannot support publication of descriptions of data obtained or re-shared without this check.

AR: response

EC: There are multiple versions of your dataset at Zenodo. Please cite the specific accession (citation 29) that is relevant for this work.

AR: response

EC: Please provide further description in the Data Record on the contents of this dataset (e.g. variables used, column headings, file names, etc) and any folder structure if relevant. The purpose of the Data Record section is to describe your data contents at a level of detail so that others may re-use them.

AR: response

EC: The description of the data in this article is minimal, making the interpretation of the provided data files hard and even leading to potential discrepancies between the description and the dataset.

AR: response

Reviewer #1

RC: Dear Authors, the workflow and datasets described in this manuscript are very valuable to the energy modelling community. It builds and extends already existing models and datasets. I have some notes:

AR: response

RC: Oemof is missing as an open source model in the introduction.

AR: response

RC: The code will be of more value if it is shared as a stand alone code and not as part of the PyPSA-Eur code.

AR: response

RC: Is there a reason behind clustering nodes in a 5000 meters radius? did you check for sensitivity to this value?

AR: response

RC: What happens to this model if overpass turbo is not maintained anymore?

AR: response

RC: Do you also search for the tag *station*, *sub_station* and the like? some substation in OSM are not well tagged (i.e. do not all have *substation* as tag).

AR: response

RC: Did you compare the relations data and the cable/line data representation for DC lines? The same question for AC lines.

AR: response

RC: There are small typos in the text, which should be corrected. Best regards

AR: response

Reviewer #2

- RC:** The paper presents a description of several csv datafiles that enable construction of the European high-voltage grid for voltages above 200 kV and a method for obtaining these.
- RC:** The data is derived from open-street maps and presents new annotations compared to the primary sources via a cleaning procedure and attachment of electrical parameters. As such, the data becomes usable for studies such as energy system modelling, or mere electric power system studies.
- RC:** The aim of the dataset is to have an openly available and up-to-date set of lines of the electric power system. Therefore, an automated extraction and cleaning method for data obtained from Open Street Maps is proposed by the authors. The quality of the dataset therefore depends on the Open Street Map contributions. Several heuristics are used to clean the data obtained from Open Street Maps, which are methodologically sound. One of the main assumptions is the use of “standard” transmission line types with associated impedances and nominal powers. With the advent of new transmission line types (e.g., high-temperature low-sag conductors), such an approach may not be suitable for future updates and may require updates to the method.
- RC:** The technical quality of the dataset is supported in two ways: (i) through comparison against official statistics from ENTSO-E on European transmission grid line lengths and against data obtained from a German TSO, (ii) through a comparison of an optimization outcome against a benchmark system.
- RC:** The first method supports the technical quality of the dataset and proves that the method of retrieving data (using Open Street Maps) can be more reliable than other data sources.
- AR:** response
- RC:** The first conclusion of the second method, that *both grids adequately represent reality* cannot be stated as such as (i) a future scenario is being evaluated, and (ii) none of both grids represents reality. The reviewer therefore suggests removing this statement. Furthermore, the reviewer suggests to provide more information of the benchmarking case (e.g., the techno-economic assumptions projected for the year 2030), as these are not described. It is not possible to determine in how much the outcomes of this validation case study depend on these assumptions, e.g., whether other assumptions would lead to larger deviations. Could the authors clarify this?
- AR:** response
- RC:** I am a bit surprised by the statement on the data repository itself:

 | *The authors of this dataset do not claim correctness, completeness, or any other form of guarantee regarding the information presented.*

Although I understand the reasoning behind this statement, this actually undermines the validation

of the dataset provided in this article, and conflicts with the statement that *the grids adequately represents reality*.

AR: response

RC: The descriptions of the data files in the proposed article is minimal, mainly in that the authors claim that csv files are provided. The authors could provide a more-in depth description of (i) the meaning of the separate data files on the repository, (ii) the structure of each data file, (iii) the size of the data file.

AR: response

RC: As the description of the data file is minimal, it is not possible to verify whether the data files submitted to the repository are complete. First, a discrepancy between the description of the datafiles (e.g., 400 kV is mapped onto 380 kV) was noted, as 400 kV is still included in the *voltage* column of buses.csv. Could the authors please verify this? Second, when verifying e.g., *lines.csv*, it is not possible to find, e.g., electrical parameters such as nominal current or impedance. Could the authors include an improved description of this data in the article?

AR: response [1]

References

- [1] me, who, what 99 (1) (2000).