Applied Energy 341 (2023) 121096

Contents lists available at ScienceDirect

Applied Energy

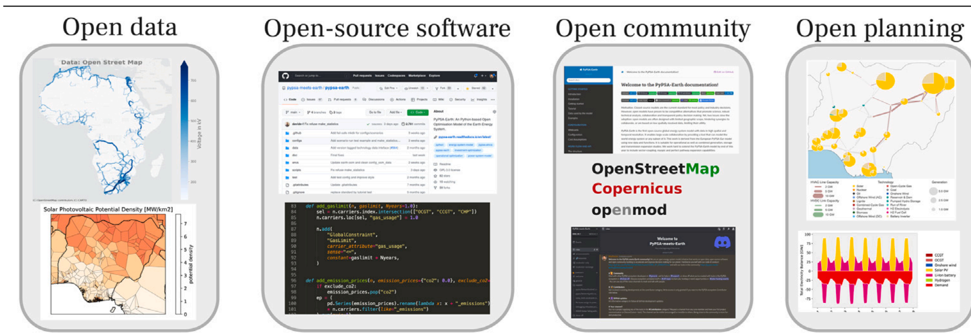
journal homepage: www.elsevier.com/locate/apenergy

PyPSA-Earth. A new global open energy system optimization model demonstrated in Africa

Maximilian Parzen [a,*], Hazem Abdel-Khalek [b], Ekaterina Fedotova [c], Matin Mahmood [a], Martha Maria Frysztacki [e], Johannes Hampp [d], Lukas Franken [a], Leon Schumm [h,g], Fabian Neumann [g], Davide Poli [f], Aristides Kiprakis [a], Davide Fioriti [f,*]

[a] *University of Edinburgh, Institute for Energy Systems, EH9 3DW Edinburgh, United Kingdom*
[b] *Fraunhofer Research Institution for Energy Infrastructures and Geothermal Systems IEG, Cottbus, Germany*
[c] *Kutuzovskaya 25 263, 143005, Odintsovo, Moscow region, Russia*
[d] *Justus-Liebig-University Gießen, Center for international Development and Environmental Research, Gießen, Germany*
[e] *Karlsruhe Institute of Technology, Institute for Automation and Applied Informatics, 76344 Eggenstein-Leopoldshafen, Germany*
[f] *University of Pisa, Department of Energy, Systems, Territory and Construction Engineering, Largo Lucio Lazzarino, 56122 Pisa, Italy*
[g] *Department of Digital Transformation in Energy Systems, Institute of Energy Technology, Technische Universität Berlin, Fakultät III, Einsteinufer 25 (TA 8), 10587 Berlin, Germany*
[h] *Research Center on Energy Transmission and Storage (FENES), Faculty of Electrical and Information Technology, University of Applied Sciences (OTH) Regensburg, Seybothstrasse 2, 93053 Regensburg, Germany*

GRAPHICAL ABSTRACT



ARTICLE INFO

Dataset link: https://github.com/pz-max/pypsa-earth-paper

Keywords:
Macro-energy systems
Optimization
OpenStreetMap
PyPSA-Earth
PyPSA-Africa
PyPSA meets Earth

ABSTRACT

Macro-energy system modelling is used by decision-makers to steer the global energy transition towards an affordable, sustainable and reliable future. Closed-source models are the current standard for most policy and industry decisions. However, open models have proven to be competitive alternatives that promote science, robust technical analysis, collaboration and transparent policy decision-making. Yet, two issues slow the adoption: open models are often designed with particular geographic scope in mind, thus hindering synergies from collaborating, or are based on low spatially resolved data, limiting their use. Here we introduce PyPSA-Earth, an open-source global energy system model with data in high spatial and temporal resolution. It enables large-scale collaboration by providing a tool that can model the world's energy system or any subset of it. The model is suitable for operational as well as combined generation, storage and transmission expansion studies. In this study, the novel power system capabilities of PyPSA-Earth are highlighted and demonstrated. The model provides two main features: (1) customizable data extraction and preparation with global coverage and (2) a PyPSA energy modelling framework integration. The data includes electricity demand, generation

* Corresponding authors.
E-mail addresses: max.parzen@ed.ac.uk (M. Parzen), davide.fioriti@unipi.it (D. Fioriti).

https://doi.org/10.1016/j.apenergy.2023.121096
Received 10 September 2022; Received in revised form 5 March 2023; Accepted 6 April 2023
Available online 19 April 2023
0306-2619/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

and medium to high-voltage networks from open sources, yet additional data can be further integrated. A broad range of clustering and grid meshing strategies help adapt the model to computational and practical needs. Data validation for the entire African continent is performed and the optimization features are tested with a 2060 net-zero planning study for Nigeria. The demonstration shows that the presented developments can build a highly detailed power system model for energy planning studies to support policy and technical decision-making. We anticipate that PyPSA-Earth can represent an open reference model for system planning, and we welcome joining forces to address the challenges of the energy transition together.

1. Introduction

1.1. Motivation

Energy system planning models are broadly adopted around the world. They are used as instruments to inform policy and investment decision-making, such as operational, supply diversification, and long-term infrastructure planning studies. Inscrutable 'black-box' models, despite being criticized in academia [1], are still the standard for high-impact modelling such as the African Continental Power System Plan [2]. This prevents transparent decision-making while having other major drawbacks, as described in [1]. Open-source models evolved to overcome these typical black-box model problems and can perform equivalent or even more tasks, but at no charge, while additionally supporting transparent and robust analyses [3]. In many examples, the European Commission applies open tools and requests their use in funded projects, proving its belief in the benefits of openness and transparency [4]. Now with the encouraging rise of more than 31 models in 2019 [5], simultaneously, concerns of failed collaboration and duplication are arising that cost taxpayer money [6]. As a result, it becomes increasingly important to avoid duplication and provide modelling solutions that allow global united efforts. For these reasons, in this study, we propose an open-source community-backed flexible energy system model able to represent any arbitrarily large region of the world power system in high spatial and temporal resolution that leverages other existing open-source projects to serve industry, policymakers, and researchers.

1.2. Literature analysis

In general, models are idealized representations of real physical systems. To ease building idealized systems, 'frameworks' have been developed to provide pre-compiled equations, algorithms, solver interfaces and/or input/output features. A framework becomes a 'model' only when data is added that describes real physical systems [7]. In this view, PyPSA is a framework and PyPSA-Eur and PyPSA-Earth are models for the European and Earth energy system, respectively. Nowadays, the open-source community is rich in energy system modelling frameworks that can provide similar functionalities. Table 1 compares some available functionalities across selected widely-adopted modelling frameworks [3,8,9]. Undoubtedly, each developer team might be capable of filling in missing features, but the functionality of the frameworks is only one important part of models, the other one, often even more relevant, is the integration of data.

Existing models are often designed to implement data with limited geographical coverage, such as a specific province, country or continent [22]. Continental models with implemented high-resolution data have proven to be the most maintained and active, possibly by covering many regions of interest and giving the user options for the aggregation level [23,24]. In contrast, there are several examples where single-country models have soon become outdated, poorly documented or inactive [25–27]. While global energy system models exist, they currently need to improve. The pioneering global energy system models that impacted policy and research discussions are closed source; for instance, LOADMATCH from Jacobson et al. [28] and LUT model from Breyer et al. [14]. GlobalEnergyGIS [29], an open-source tool which can create energy system model data for any arbitrary region, is used in the Supergrid model [30], but misses network data or a workflow management system that are important for flexible and reproducible data processing [31]. Similarly, the GENeSYS-MOD model is a global open-source model; however, it is written in GAMS, preventing free use and offers no data processing workflows [32]. Another promising candidate is the recently released OSeMOSYS Global model, which includes a workflow management system but misses network topology data as well as unit-commitment and power flow constraints that have been shown to strongly affect model results [9,33].

Similarly, existing PyPSA models are geographically limited. While PyPSA as a framework is adopted worldwide by many companies, non-profit organizations and universities (see example studies in [34]), no global model solution is available. Providing a global energy system model ecosystem solution has the potential to unlock collaboration potentials that accelerate and improve the energy transition planning.

1.3. Contributions

This paper presents PyPSA-Earth, an open-source global energy system model with data in high spatial and temporal resolution. We classify PyPSA-Earth as an energy system model even though the manuscript focuses on the electricity sector. We do this for three reasons: first, the underlying model framework allows sector-coupling, meaning modelling not only electricity, but also other sector such as heating and transport; second, the model can already build hydrogen networks and supply hydrogen demands which is not demonstrated in this study but visible in the source code; and finally, existing work is ongoing to implement the data for the other sectors. Across the manuscript high spatial resolution implies the ability to represent a regional (e.g. country) energy system with a flexible number of sub-regions, each of them describing an arbitrary large (e.g. counties) or small (e.g. provinces) proportion of the region under consideration. Similarly, high temporal resolution means that the time series used in dispatch analyses can be hourly, sub-hourly or larger than an hour, in agreement with the requirements of the energy model. Details on the data and methods are explained in Section 3.

Users can flexibly model the world or any subset of it. Using an automated workflow procedure, they can (a) generate energy system model relevant data and (b) perform planning studies. The novel contributions of PyPSA-Earth are detailed as follows:

1. New model creates arbitrary high spatial and temporal resolution representation of power systems around the world
2. Automated workflow generates national, regional, continental or global model-ready data for planning studies based on open or optionally closed data
3. Integration and linking of multiple data sources and open-source tools to process raw data, e.g. OpenStreetMap
4. Provision of new spatial clustering strategies to simplify the high-resolution model
5. Data and model validation for the African continent and Nigeria
6. Development of 2060 net-zero energy planning study for Nigeria

PyPSA-Earth includes several novel modelling features that make this manuscript unique. While the model leverages previous open-source tools, such as PyPSA-Eur [23], the aim of modelling the globe with flexible spatial and time resolution, the filtering methodology,

Table 1
Comparison of selected features for energy system modelling frameworks that are applied in Africa.

Software	Version	Citation	Language	Free and open	Power flow	Transport model	LOPF[d]	SCLOPF[e]	Unit commitment	Sector-coupling	Pathway optimization[f]
Calliope	v0.6.8	[10]	Python	✓		✓			✓	✓	
Dispa-SET	v2.4	[11]	GAMS	✓		✓			✓		
GridPath	v0.14.1	[12]	Python	✓		✓	✓		✓		✓
LEAP	2020.1.63	[13]	NA[a]							✓	
LUT	2021	[14]	GNU[b]			✓	✓			✓	✓
NEMO	v1.7	[15]	Julia	✓	✓	✓	✓		✓		
OSeMOSYS	2022	[16]	GNU[c]	✓		✓				✓	✓
PLEXOS	9	[17]	NA[a]			✓	✓	✓	✓	✓	✓
PYPOWER	5.15.5	[18]	Python	✓	✓	✓	✓		✓		
PyPSA	v0.20.0	[3]	Python	✓	✓	✓	✓	✓	✓	✓	✓
SPLAT-MESSAGE	2022	[19]	GAMS			✓			✓	✓	
TIMES[g]	2022	[20]	GAMS			✓	✓		✓	✓	✓

[a]NA = no information available.
[b]Mix of GNU-Mathprog and Matlab.
[c]Available in GNU Mathprog, Python and GAMS.
[d]Linearized optimal power flow [3].
[e]Security constrained linearized optimal power flow [3].
[f]Includes myopic and perfect foresight optimizations over multiple years [21].
[g]Times is open source but not free as licensing for GAMS is required to operate the model.

the automatic data fetching by OpenStreetMap, and the data workflow procedures here detailed are a novelty. Accordingly, PyPSA-Earth is not an extension of any previous models but a novel approach and tool that is first presented in this manuscript with a focus on the African continent. In the following, PyPSA-Earth is presented and quantitatively validated for the African continent and Nigeria; its optimization features are finally tested for a 2060 net-zero energy planning study for Nigeria's electricity sector.

All code and validation scripts are shared open-source under GPL 3.0 license. The data, often extracted by python script activation, is available under multiple open licenses. For a detailed license listing, see [35].

1.4. Organization of the paper

The rest of the paper is organized as follows. Section 2 introduces the novel PyPSA-Earth model that can perform large-scale modelling studies. The data processing novelties are detailed in Section 3. Data validation for the African continent is performed in Section 4, and a quantitative case study on Nigeria is discussed in Section 5. Finally, the limitations of the model are discussed, and the conclusions are drawn.

2. PyPSA-Earth model

This section describes the scope of the PyPSA-Earth model, its features as well as the role of the initiative that is facilitating the model developments.

2.1. Scope

The PyPSA-Earth model is a novel open-source data management and optimization tool that aims to provide policymakers, companies and researchers with a shared platform for a wide range of macro-energy system analyses needed to achieve the energy transition together. The option to create a tailored country, continental or global model under a unique code repository maximizes synergies and wider user benefits. For instance, one user in Africa can implement new features and data, improve the documentation or implement bug fixes that immediately benefit all other users worldwide.

By leveraging on existing methodologies and PyPSA-based tools [34], PyPSA-Earth has been developed to enable, among others, the following studies in any country or region on the planet:

1. energy system transition studies
2. power system studies
3. technology evaluation studies (e.g. energy storage, synthetic fuels and hydrogen pipelines)
4. technology phase-out plans (e.g. coal and nuclear)
5. supply diversification studies
6. electricity market simulations.

In this manuscript, we focus on the power system modelling features of PyPSA-Earth.

2.2. Features

The following features are implemented in PyPSA-Earth:

1. flexible model scope: from Earth to any subregion
2. high temporal and spatial resolution
3. model-ready data creation
4. co-optimization of investment and operation
5. single or multi-year optimization
6. flexible addition of arbitrary optimization constraints, e.g. socio-economic, technical, or economic.

Moreover, the PyPSA-Earth model has been developed with the following non-functional requirements:

1. easy to use and learn
2. highly customizable and flexible
3. modular to include new features and data
4. fully reproducible.

The proposed features of PyPSA-Earth are a novelty compared to the literature in Section 1. Furthermore, new features can be created in or adopted from other PyPSA-based models that share a similar backbone. Examples are the work on endogenous learning with pathway optimization and multiple investment periods [36], dynamic line rating constraints based on spatially differing environmental conditions [37], the implementation of generic constraint settings that enable equity constraints such as applied in [38] and uncertainty analyses by input parameter sweeps or by exploring the near-optimal solution space [39].

The data and methods Section 3 presents more details on the presented features.

2.3. PyPSA meets Earth initiative

The PyPSA meets Earth initiative is an independent research initiative that aims to improve energy system planning with open solutions. It supports, builds and maintains the PyPSA-Earth model and is therefore briefly introduced. The initiative's vision is to support transparent

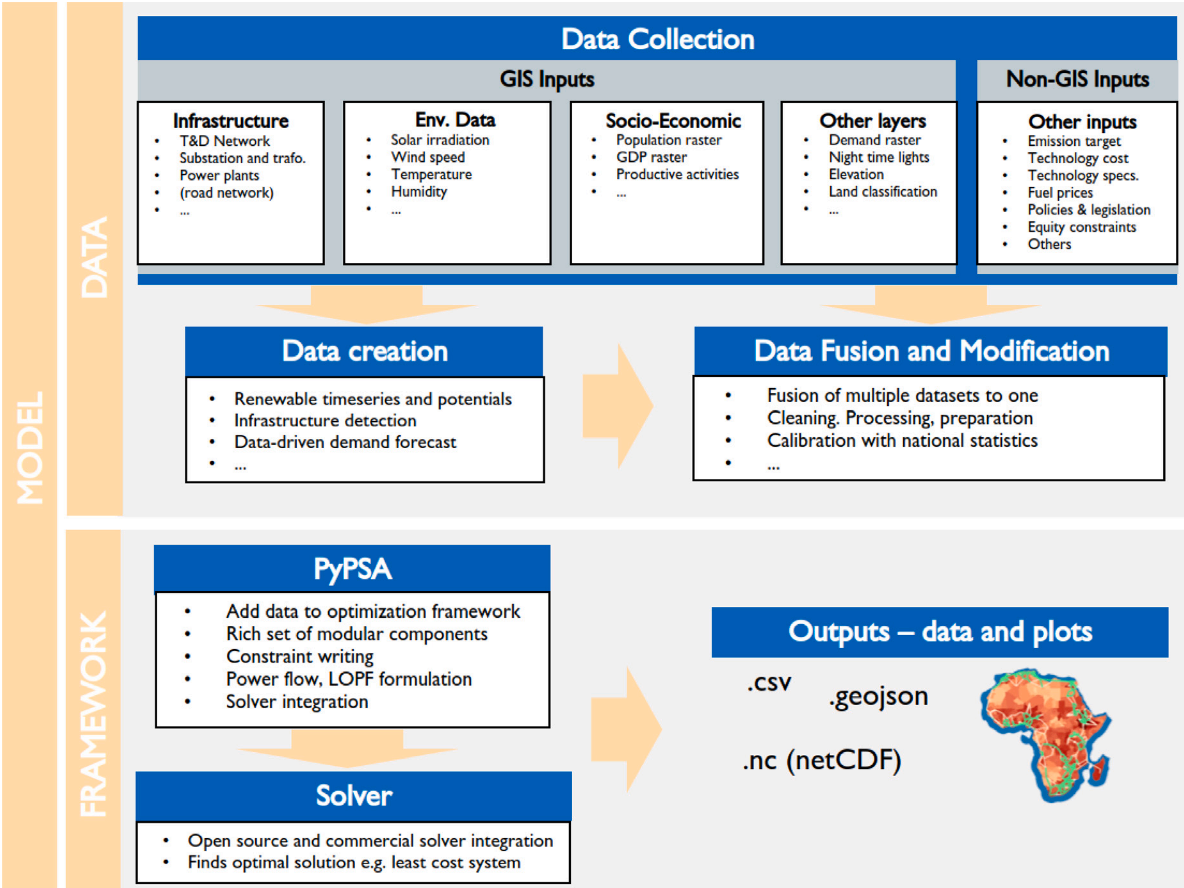


Fig. 1. PyPSA-Earth model design. After providing the configuration parameters and countries of interest, data is collected and processed to be then fed into the PyPSA model framework which enables to perform the desired optimization studies such as least-cost system transition scenarios.

and debatable decision-making on the energy matter that cannot be achieved with the status quo ruled by commercial inscrutable closed-source "black-box" tools. Current research activities in the initiative can be categorized into three distinct groups:

- open data
- open energy system model
- open-source solver.

First, the open data activities focus on open data creation, collection, fusion, modification, prediction and validation for energy system models. These data activities are not limited to aggregated country information but prioritize work on high spatial and temporal resolution data, which is fundamental for scalable and accurate mini-grid and macro-energy system model solutions. Second, the open energy system modelling activities focus on implementing new functions and data streams into the model, such as building a sector-coupled model with multi-horizon optimization that is useful across the globe. Third, open-source solver-related activities deal with benchmarks and efficient interfaces that help to adopt and develop open-source solvers. For instance, we created a benchmark that became a successful public funding proposal and attracted sufficient funding for the open-source solver HiGHS [40]. This activity pushes breakthroughs in large-scale optimization performance required for energy system models, which were until now reserved only for people that can afford commercial proprietary solvers.

In order to assure a continuous inflow of people that maintain, improve and use the software, as needed by open-source software [41], the initiative supports a free and open community where anyone can contribute. The initiative adopts:

- *GitHub* to publicly record issues, requests, solutions or source code-based discussions;
- *Discord* as a voice channel and messaging social platform for regular public meetings and exchanges;
- *Google Drive* to publicly store files and meeting notes.

Together, these tools provide the backbone of the open community supporting the initiative goals and activities (data, model, solver).

3. Data and methods

In this section, the novelties of PyPSA-Earth methodology are highlighted and described in detail. As depicted in Fig. 1, first, we introduce the workflow management tool that supports the model user experience. Then, data creation and processing approaches are discussed considering the main data blocks used by the PyPSA-Earth model: power grid topology and spatial shapes, electricity demand, renewable potential and power plant locations. Further, we describe some advanced pre-processing techniques, such as clustering and line augmentation, used to introduce data into the model in a robust and efficient way. Finally, we describe the energy system modelling and optimization framework with its solver interfaces.

3.1. Workflow management tool

First of all, similarly to PyPSA-Eur [23], PyPSA-Earth relies on the 'Snakemake' workflow management [42] that decomposes a large software process into a set of subtasks, or 'rules', that are automatically chained to obtain the desired output. Accordingly, 'Snakemake' helps sustainable software design that enables reproducible, adaptable and transparent science, as described in [43]. The whole PyPSA-Earth workflow was implemented as a new set of 'Snakemake' rules. For more details, Figure B.13 in Appendix B represents a workflow of PyPSA-Earth automatically created by 'Snakemake' for which the user can execute any part of the workflow with a single line of code. That is expected to improve the user and developer experience, as complex tasks are decomposed into multiple modular smaller problems that are easier to handle and maintain.

3.2. Network topology and model

The electricity network topology is one of the main inputs needed to build an energy system model which accounts for realistic power flow approximations across regions. The most comprehensive and accurate data on power grids are curated by the transmission system operator. In practice, the availability of open power grid data is still relatively low for many parts of the world, with the situation in Africa being extremely sparse.

A natural way to address the lack of power grid data is to utilize open geospatial datasets. Currently, a few open-source packages have been published to extract and build networks from such datasets (e.g. Gridkit [44], Transnet [45], SciGrid [44]). However, each package focuses on applications for a particular world region rather than on the global coverage and there still needs to be a ready-to-use solution which could be implemented into a global model. To fill this gap, we have developed an original approach which reconstructs the network topology by relying solely on open globally-available data. The developed approach is based on the OpenStreetMap (OSM) datasets that are a crowd-sourced collection of geographic information, which is daily updated and includes geolocation Refs. [46].

The electricity network topology is created in three novel steps: (i) downloading, (ii) filtering and cleaning the data, and (iii) building a meshed network dataset with transformer, substation, converter and high voltage alternating current (HVAC) as well as high voltage direct current (HVDC) components. Figure B.14 shows sample raw and cleaned networks along with the options for clustering and line augmentation that are introduced in Sections 3.7 and 3.8, respectively.

For the download step, the *esy-osm* tool is used to allow fast retrieval of OSM data through multi-threaded processing [47]. Appropriate OSM features are used to extract all necessary network components, including substations, transformers and power lines. Their geospatial description was cleaned in this process, and the data structure aligned with the PyPSA framework requirements.

Beyond that, to build the network, an approach has been developed to improve the quality of the OSM-extracted grid topology by accounting for a reasonable tolerance of OSM-derived coordinates.

3.3. Fundamental shapes

Fundamental shapes represent the smallest defined regions that gather various data types to characterize the energy system, such as in Fig. 2 or Figure B.15 in Appendix B. Before being ready for the model-framework execution, data is often provided in many different ways, e.g. geo-referenced point locations and raster data. To properly execute the modelling, such information is gathered and aggregated at the level of the fundamental shapes. These shapes can represent either administrative zones or spatial zones generated from the grid structure as shown in Figure B.15 in Appendix B, which is in the following discussed in more detail for onshore and offshore shapes.

For onshore regions, the model provides two ways to build fundamental data shapes. The first retrieves the so-called Global Administrative Areas (GADM) that represent administrative zones at various levels of detail (e.g. national, regional, province, municipality) [48]. The second one uses the substation GIS location to create Voronoi partitioned areas for each substation, which boundary is defined as equidistant to the centroid of the nearest sites [49]. The latter approach is beneficial to replicating the network accurately, while the former helps communicate results.

For offshore regions, the model uses only Voronoi partitioned areas to create fundamental shapes. These Voronoi areas are built from high voltage onshore nodes and are limited to the offshore extent by the Maritime Boundaries and Exclusive Economic Zones (EEZ) data for each country [50].

3.4. Electricity consumption and prediction

The model currently provides globally hourly demand predictions considering 'Shared Socioeconomic Pathways' [51] scenarios for 2030, 2040, 2050 and 2100 and weather years of 2011, 2013 and 2018. The demand time series is created using the new contributed *synde* package [52], which implements a workflow management system to extract the demand data created with the open-source Global-Energy GIS (GEGIS) package [29].

In principle, GEGIS produces hourly demand time series by applying machine learning methods [29] using as predictors temperature, population, GDP, industrial structure, heating, cooling technologies, among others, similarly to [53]. The observed absolute error of GEGIS in the validation test is considered acceptable for energy studies as it is 8% across 44 countries, yet with generally worse performance in low-income countries [29].

The coverage of the *synde* package is currently limited. Figure B.16 in Appendix B shows no data outputs for especially low-demand countries. A heuristic creates data for the countries with missing data by scaling the Nigerian demand time series proportionally to population and GDP. We validate this approach in Section 4.2.

3.5. Renewable energy sources

Renewable energy sources such as solar, wind and hydro time series are modelled with the open-source package *Atlite* [54]. *Atlite* (i) creates cutouts that define spatio-temporal boundaries, (ii) prepares cutouts, which means that environmental and weather data is added to geospatial boundaries by matching various datasets (ERA5 reanalysis data [55], SARAH-2 satellite data [56], and GEBCO bathymetry [57]), and finally, (iii) applies conversion functions to produce technology-specific spatially resolved time series and potentials [54]. Currently, the *PyPSA-Earth* model framework implements solar photovoltaic, on- and offshore wind turbines, hydro-runoff, reservoir and dam power resources. In the case of hydro, the runoff time series are obtained by *Atlite* for each powerplant location, as described in Section 3.6. As our new contribution, the hydro power output is thereby proportionally rescaled to match the total energy production of existing power plants by country [58]. At the time of writing, available in *Atlite* but not yet implemented in PyPSA-Earth are potentials and time series for concentrated solar power, solar thermal collectors, heat demand and dynamic line rating with a wide range of technology options. For details on the model implementation for each technology, we refer the reader to the *PyPSA-Eur* publication which the presented model mostly builds-upon [23]. A brief concept demonstration of *Atlite* is provided in Fig. 3.

3.6. Generators

Given the limitation of reliable datasets for power plants for the African region, the existing powerplantmatching tool [59] has been extended to include additional datasets, such as OpenStreetMap, to fine-tune the African model and validate the results with the final goal of maximizing accuracy and quality of the result.

Powerplantmatching has been successfully proposed to estimate the location and capacity of power plants in Europe. The validation performed with respect to the commercial World Electric Power Plants Database (WEPP) by Platts and the dataset by the Association of European Transmission System Operators (ENTSO-E) reaches an accuracy

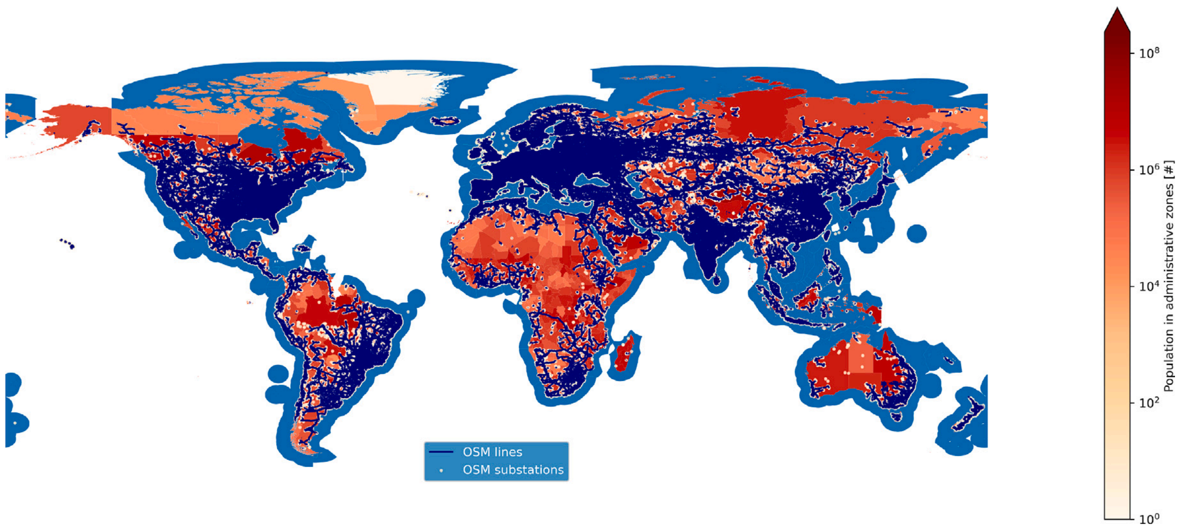


Fig. 2. Representation of transmission networks by Open Street Map (OSM) and shapes produced by PyPSA-Earth.

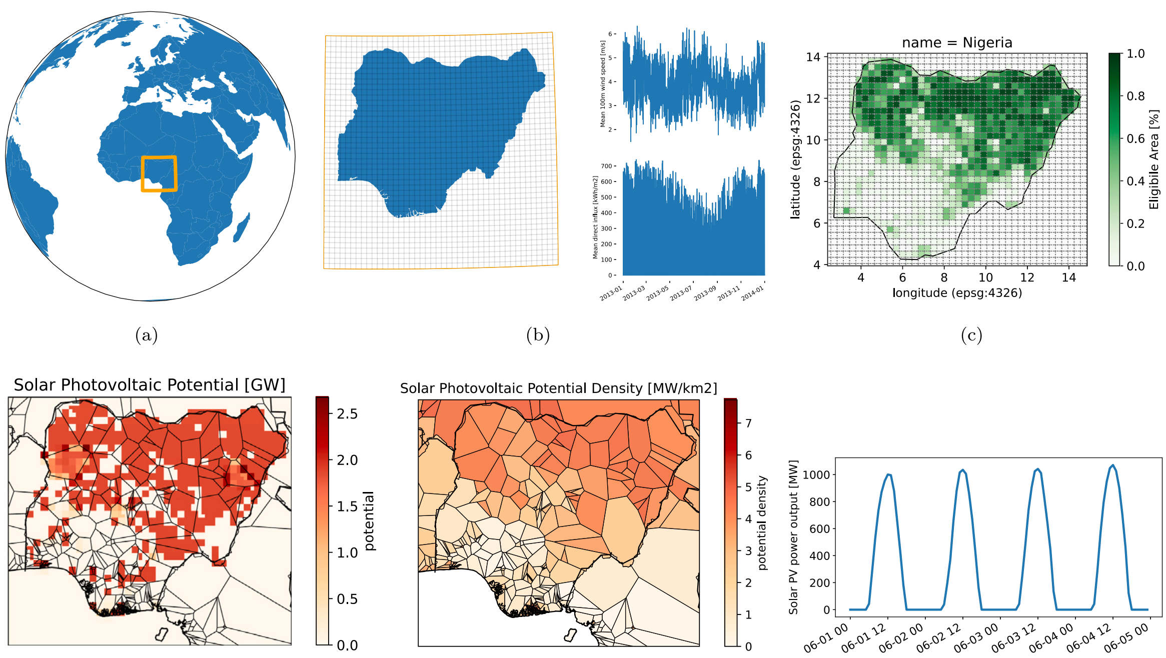


Fig. 3. A concept demonstration of Atlite for Nigeria. (a) Shows that environmental and weather data is extracted in a *cutout* for the region of interest. (b) The cutout is split in a raster of $(0.25°)^2$ or roughly $(27.5\ \mathrm{km})^2$ (length varies along latitude), whereby each cell contains static or hourly time series data. The example wind speed and direct irradiation influx time series are shown for one cutout cell that contains an *ERA5* extract of the Copernicus Data Store [55]. (c) Shows the eligible area raster, which is built by excluding protected and reserved areas recorded in *protectedplanet.net* and excluding specific land-cover types from *Copernicus Global Land Service* whose eligibility can vary depending on the technology. (d) Illustrates the maximal installable power raster, which is calculated by the eligible area and the socio-technical power density of a technology e.g. $4.6\ \mathrm{MW/km^2}$ for solar photovoltaic. (e) The raster is then downsampled to the region of interest or fundamental shape by averaging the proportion of the overlapping areas. (f) Finally, by applying a PV technology model to (b) and combining it with (e) we can define per region the upper expansion limit and the maximal hourly availability constraint for a given technology.

of around 90% using only open data [59]. By default, various open data sources are included, such as [60], ENTSO-E [61], GEO [62], and renewable statistics by IRENA [63] among others. The approach applied for *powerplantmatching* is based on the procedure depicted in Fig. 4 available in Appendix B, where the raw datasets are first downloaded, then filtered to remove missing or damaged data, and aggregated. Once the refined data are obtained, the datasets are pairwise compared to identify duplicated entries. Finally, non-duplicated data are merged into a unique dataset and used as a source for PyPSA-Earth. Only a few of these datasets have global scope (GEO, GPD and IRENA) and have been validated for Africa. In particular for Africa, where data is lacking, including all available open data can be critical to maximizing the accuracy of the results. Therefore, inspired by future work suggested in [64], we have extended the *powerplantmatching* tool to optionally

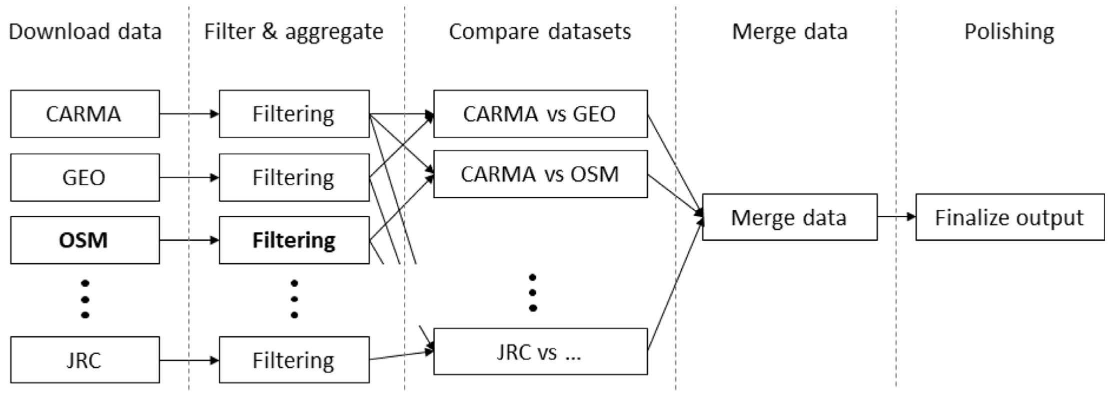


Fig. 4. Flowchart of the *powerplantmatching* procedure, including the novel OSM input (in bold) which was developed for PyPSA-Earth.

include and process OpenStreetMap data to improve the quality of outputs.

3.7. Spatial clustering approach

In order to tackle the computational complexity of solving a co-optimization problem of transmission and generation capacity expansion, the model offers state-of-the-art spatial clustering methods adapted from the PyPSA package and PyPSA-Eur model [3,65]. Spatial clustering allows aggregating nodes of the system to reduce the complexity of the model, which is essential for reducing the computational needs.

The available clustering methods provide a focus on (i) conserving the representation of renewable potentials as well as the topology of the transmission grid, (ii) accurately representing the electrical parameters to improve estimates of electrical power flows in an aggregated model, (iii) aggregating spatially close nodes disregarding other a-priori information of the network, or (iv) according to their location with regards to the country's subdivisions facilitating results interpretation for policy recommendations. An analysis of suitable clustering methods that depend on the modelling application is provided in [66].

In summary, (i) the clustering approach that focuses on a better representation of variable sources or sinks of the model is inspired by [67]. It includes variable potentials, i.e. capacity factors or full load hours for solar and wind, or the variable electricity demand as a distance metric between nodes. This is combined with a hierarchical clustering approach, similar to the suggestions provided in [68]. However, we only allow nodes to be aggregated when a physical transmission line connects them instead of assuming a synthesized grid in contrast with [68]. (ii) The clustering method that focuses on a better representation of the transmission grid was initially suggested by [69] to be applied to the case of electricity system modelling. It is a density-based hierarchical clustering operating on the line impedance. (iii) The network can also be reduced using a weighted k-means algorithm on the locations of the network nodes as explained in detail in [49]. (iv) Finally, using the GADM shapes allows aggregating all nodes in the same shape.

Any of these methods can be applied in single or two distinct iterations, as displayed in Fig. 5 for Nigeria. In each of these two iterations, a different method can be applied, choosing from (i)-(iv). In the first iteration, all nodes are clustered to a desired number of representative nodes, aggregating generators, flexibility options (electricity storage and transmission lines) and electrical demand. The second iteration is optional and allows the remaining nodes to be clustered again. However, now only the transmission network is effectively reduced such that the representation of renewable resources is fixed to the resolution of the previous iteration (compare the first row and second row of Fig. 5). The spatial resolution of the transmission network must always be larger or equal to the resource resolution, i.e. the clustering of the first iteration sets an upper bound.

3.8. Augmented line connection

The African network is often not well interconnected. This is due to isolated national planning data or the presence of isolated mini-grids that are popular electrification measures [70]. Therefore, we propose an algorithm to mesh a given network and assess different grades of connectivity. To investigate the benefits of meshed networks, PyPSA-Earth can perform a k-edge augmentation algorithm that guarantees every node has a modifiable number of connections to other nodes. Only if nodes do not already fulfil the connectivity condition, the algorithm will create new lines to the nearest neighbour by a minimum spanning tree. The new 'augmented' lines can be set to an insignificant size (e.g. 1 MW) to create new options for line expansion in the investment optimization. For example, Figure B.14 in Appendix B compares between the standard clustered network with 420 nodes and its augmented version. Only the model that includes augmented line connections can explore an interconnected continent.

3.9. Model framework and solver interface

The PyPSA-Earth model integrates the PyPSA model framework with its solver interfaces to perform planning studies; details on the mathematical modelling are provided in Appendix A for the sake of brevity. Using PyPSA has several benefits compared to other tools that are briefly introduced in the following. First, PyPSA enables large-scale optimization in Python. Python is well known for being user-friendly, but when analysing the memory consumption and speed for building optimization problems, it was considered non-competitive compared to tools based on the programming language Julia or C++ [71] – a bottleneck which also hinders large-scale optimization required for PyPSA-Earth. As a reaction, developers in the PyPSA ecosystem built *nomopyomo* overcoming the bottlenecks [5]. More recently, the same group has been working on a general package called *Linopy* that promises a 4–6 runtime speed up and a 50% improvement in memory consumption compared to the optimization problem formulator Pyomo, possibly making it also more memory efficient than the Julia alternative *JuMP* as indicated in [71]. Another point making the PyPSA dependency attractive is that it is one of the most popular tools, as suggested by GitHub stars in the GPST benchmark [72], possibly due to its standard component objects and the continuously maintained documentation [34]. Finally, the framework offers several solver interfaces (HiGHS, Cbc, GLPK, Gurobi, among others), providing flexibility in solving various optimization problems with open-source and proprietary solutions.

4. Validation

The data validation section aims to assess the data quality with publicly available data: at a continental level in Africa and a country level in Nigeria.

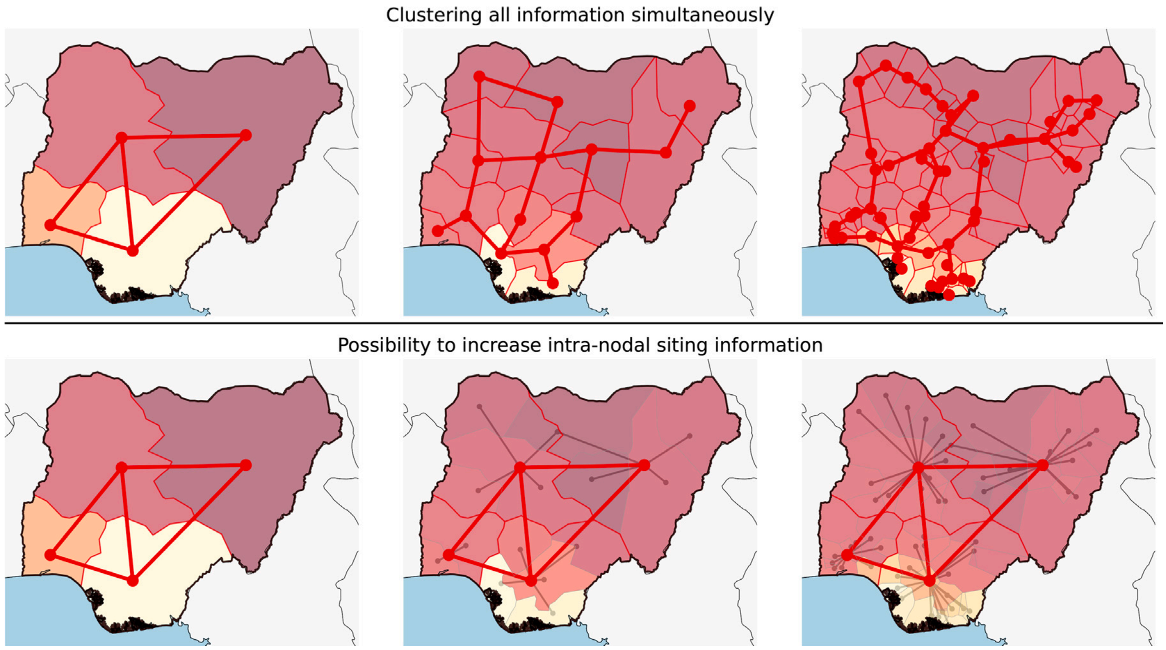


Fig. 5. Illustration of the clustering methodology applied for the transmission network (red nodes and edges) and resource resolution (grey nodes and edges). In the first row, we show how nodal data (i.e. generators, storage units, electrical loads etc.) is aggregated in tandem with the resolution of the transmission network. Three exemplary resolutions of the network for Nigeria are displayed here: 4, 14 and 54 nodes from left to right. The second row shows how the clustering also allows modelling the transmission grid at a different resolution than the resources. In this example, the transmission network contains 4 nodes connected by 4 lines (all in red) at every resolution, while 4, 14 and 54 generation sites become available (left to right). The background colour represents exemplary capacity factors in shades of red, for an arbitrary technology. The darker the colour, the higher the capacity factor.

4.1. Network topology and length

Validating the African power grid is challenging. Different from Europe, where ENTSO-E [73] provides reliable open data with continental scope, such a transparent data source is lacking in Africa, and only a few utilities release open data. The self-proclaimed most complete and up-to-date open map of Africa's electricity network is offered by the World Bank Group, which implements Open Street Map data, as well as indicative maps data from multiple sources [74]. However, the World Bank data should not be used as a single validation set because it may report outdated data, given that it has yet to be updated after 2020, and is partially based on indicative maps rather than on geo-referenced data, making the post-validation time-consuming. Conversely, PyPSA-Earth builds its grid topology directly from daily updated Open Street Map data. Finally, the World Bank data also provides less detailed information than Open Street Map; for instance, it does not give any information on the frequency, circuit or cable number, limiting the information that can be used for validation. The following compares grid statistics and topology on a Nigerian and African scale, including nationally reported data from the Nigerian energy commission.

First, the transmission lines are validated by comparing the total circuit lengths at different alternate current (AC) voltage levels. Transmission lines can carry one or more 3-phase circuits, each with at least three cables. Instead of looking only at the line length, which is the distance between high voltage towers, it is common to report the total circuit length, which multiplies each line length, e.g. distance from tower to tower, with the number of circuits [23]. Table 2 indicates that the Nigerian network length reported at the World Bank aligns approximately with the official transmission company statistics [75], suggesting that the World Bank data is either accurate in Nigeria or used as a reference by the transmission system operator. This officially reported total circuit length is approximately 35% longer than the original Open Street Map data or the modified and cleaned PyPSA-Earth derivative. Conversely, on a continental scale, Open Street Map provides approximately a 117% longer total circuit length than the reported World Bank data. To summarize, while Open Street Map data is qualitatively less available in Nigeria by looking at the statistics, it offers significantly more data on a continental scale.

To further compare and validate the data, Fig. 6 highlights good agreement between the network topology in Nigeria and Africa of Open Street Map and World Bank sources. However, in central and south Nigeria, the World Bank covers more power lines. On the African scale, the opposite is observed. Open Street Map covers more network structures in East and North Africa.

4.2. Electricity consumption

This subsection validates the demand prediction on the example year 2030 for every African country by comparing the individual country consumption for 2020 and 2030 with official continental annual electricity consumption used in PyPSA-Earth. Fig. 7 shows 2020 reported electricity consumption data per country, published from *Our World in Data* that is additionally refined by data from the global energy think-tank Ember and BP's statistical review of world energy [76]. The used electricity demand data in PyPSA-Earth roughly doubled from 2020 to 2030, indicating demand growth. While national demand predictions are often unavailable, the demand prediction is further validated by comparing it to other more common continental demand predictions. In Africa, *Our World in Data* reported an electricity consumption in 2020 of 782 TWh/a. For 2030, [77] predicted 1924TWh/a, [78] 1877 TWh/a and the PyPSA-Earth model data 1866 TWh/a, predicting more than a doubling of Africa's electricity consumption by 2030. In summary, looking at the total African electricity consumption suggests that the data used in the global PyPSA model is in the range of others.

4.3. Solar and wind power potentials

The validation of solar and wind potential is performed by comparing statistics by international organizations, such as IRENA, with

Table 2
HVAC and HVDC circuit line lengths of Nigeria and Africa from different sources.

Circuit lengths in 1000 km	Nigeria			Africa			Ref
	110–220 kV	220–380 kV	>380 kV	110–220 kV	220–380 kV	>380 kV	
World Bank Group[a]	9.3	12.1	0.0	59.4	63.5	41.0	[74]
Open Street Map (OSM)	6.3	9.1	0.0	87.9	180.7	76.7	[46]
Transmission Company of Nigeria	More than 20			–	–	–	[75]
PyPSA-Earth (cleaned OSM)	6.7	9.1	0.0	88.3	183.7	82.9	

[a]Information about circuits is missing.

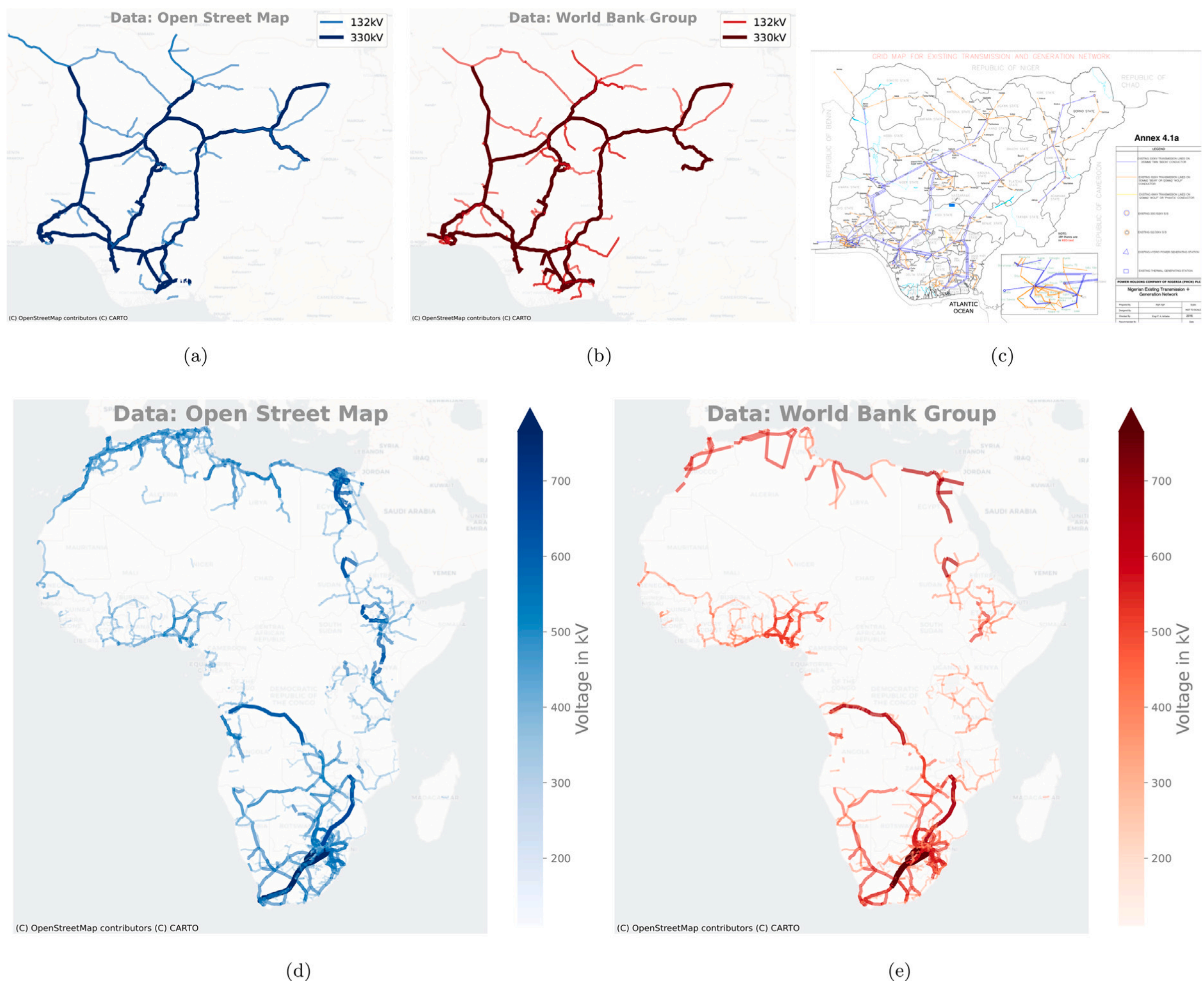


Fig. 6. Network topology of open available transmission network data (above 110 kV) from (a) & (d) Open Street Map, (b) & (e) World Bank Group and (c) the Nigerian Transmission Company. On the African scale, the voltage ranges from 110–765 kV in both data sets. The line format varies with the voltage level and includes transparency, thickness and colour.

the outputs of the PyPSA-Earth model, both including total generation capabilities and the specific power densities per unit of available land.

Solar and wind potentials are well-reported across the African continent. In 2021, the Global Wind Energy Council estimated for Africa a technical potential for wind generation of 180.000 TWh (PyPSA-Earth: 108.700 TWh), sufficient to electrify the continent 250 times relative to the 2019 demand [79]. Similarly, the International Renewable Energy Agency estimated in 2014 that the technical potential in Africa is 660.000 TWh (PyPSA-Earth: 122.200 TWh), which is sufficient to electrify the continent 916 times [80]. The discrepancy between the technical potentials observed in the PyPSA-Earth model and the institutional reports is due to the underlying assumptions. How many renewables can be installed in a region depends on two main assumptions: the excluded areas [km^2] and the power density per technology [MW/km^2], both discussed in the following.

While we define the available areas in a data-driven way similar to [23,80] (see details in Section 3.5), the remaining eligible area quantifies the technical potential per technology through the technical power density factor. However, this density applies only to land specifically and uniquely allocated to renewable production, yet this cannot easily be generalized to all non-protected land areas at the country level. Land areas are also necessary for non-technical activities such as economic activities, industries, farming, well-being, and housing. Accordingly, in PyPSA-Earth we considered a more conservative power density coefficient to account for such socio-economic considerations.

Focusing on solar photovoltaic power plants, we assessed the power density of the 41 largest installations in the world [81,82]: the average power density is 46.4 MW/km^2, the minimum 10.41 MW/km^2, and the maximum 150.0 MW/km^2. The type of solar module and the design of the solar photovoltaic plant drive this extensive range of values. For instance, the Cestas Solar Park in France uses high-performing solar

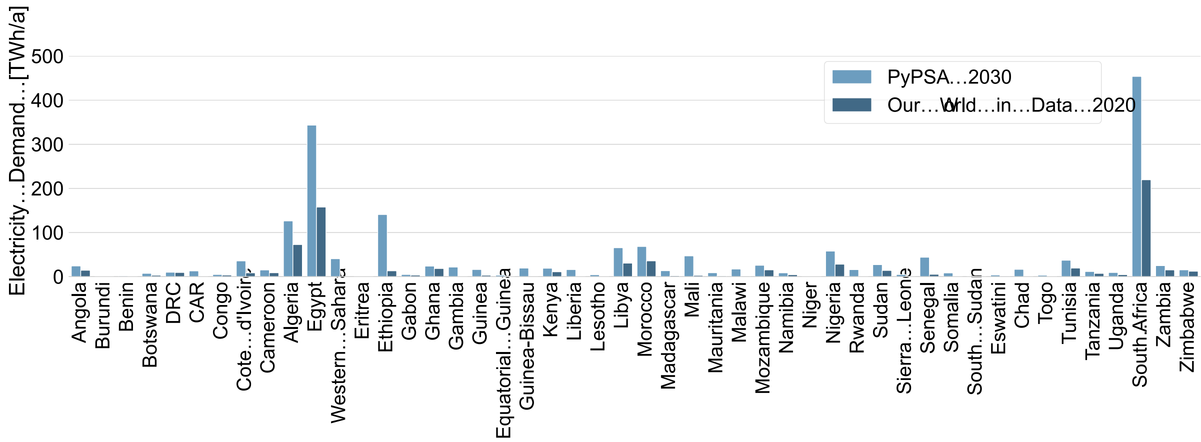


Fig. 7. Comparison of reported [76] and predicted annual electricity consumption data across African countries indicate in every country demand growth. For 2030, the African total electricity consumption of PyPSA aligns with other predictions from [77,78].

modules and additionally contains a compact east–west orientation solar field design leading to the 150.0 MW/km^2 extreme. Similarly to [23], we reduced the technical power density to 10% of the average power plant density to represent the socio-technical limit: 4.6 MW/km^2 for solar photovoltaic.

For onshore and offshore wind farm technologies, we verify power density assumptions by analysing seven existing utility-scale wind farms. The observed average technically feasible power density for onshore wind farms is 6.2 MW/km^2, and for offshore wind farms, 4 MW/km^2 [83]. Using the same approach as [23], we reduced these values from 6.2 to 3 MW/km^2 and from 4 to 2 MW/km^2 for onshore and offshore wind farms, respectively, to represent socio-technical power densities and give wind farms space to lower generation reducing wake-effects.

Currently, we apply the same socio-technical power density irrespective of the land cover type. However, roughly about 43% of the continent is characterized as extreme deserts [84], allowing these regions to be less conservative about the social-technical power density.

4.4. Power plant database

This section compares the site-specific power plant database used in PyPSA-Earth to national statistics.

Data on existing power plants is critical for accurate energy simulations as they affect long-term investments, dispatch, and stability of the energy systems. For validation purposes, relevant country statistics are provided by IRENA [85] and USAID [86]. While the PyPSA-Earth data is geo-referenced, including each power plant's location, type and nominal capacities of each power plant, the other sources only provide country statistics. Therefore, data used in PyPSA-Earth is of higher quality, especially for energy system modelling in high spatial resolution that would be impossible to perform only with the IRENA and USAID sources.

Fig. 8(b) shows that the PyPSA-Earth model matches the largest fraction of the installed capacity of existing databases, with 165 GW out of the 229 GW reported by IRENA. Most technologies are matched with adequate accuracy (2%–15% error), yet larger differences occur, especially for coal and gas power plants, partially due to the recent installation of power plants over the last 3–4 years, whose data has not been updated by the sources described in Section 3.6. In future work, adding more recent data sources may improve the data situation [87]. Furthermore, we note that, although the current PyPSA-Earth procedure does not include geothermal and CSP technologies, their capacity can be relevant for certain countries (e.g. Kenya, Morrocco and South Africa), but at an African scope, these technologies still represent a small fraction of the installed capacity. Therefore, the proposed validation is considered of good accuracy, supporting the appropriateness of the PyPSA-Earth model.

5. Demonstration of optimization capabilities in Nigeria

At COP26, Nigeria's president Buhari committed to net zero emissions by 2060 [88]. To demonstrate that the presented model can be useful for Nigeria's energy planning activities, we showcase the optimization capabilities of PyPSA-Earth. In particular, this section covers two least-cost power system optimizations, one for 2020 to reproduce the historical system behaviour and one representing a decarbonized 2060 scenario (see Figs. 10 and 11).

5.1. Nigeria 2020 - Dispatch validation

The 2020 scenario applies a dispatch optimization with linear optimal power flow constraints to simulate and validate the optimization results for Nigeria. Accordingly, only the operation of existing infrastructure is optimized for the lowest system cost, excluding any infrastructure expansion e.g. generation or transmission line expansion (see Fig. 10).

Starting with the scenario design. The power grid retrieved from OpenStreetMap is clustered into 54 nodes, representing the aggregation zones for the demand and supply. Since the existing network is more meshed than the OpenStreetMap based PyPSA-Earth network (see Fig. 6), a few augmented line connections with a negligible minimal capacity of 1 MW are added such that every node has at least two line connections, see (b) in Fig. 10, to overcome missing network data. A total demand of 29.5 TWh is considered for 2020 using the national demand profiles provided in PyPSA-Earth. The magnitude aligns with reports from *Our World in Data* (28.2 TWh) [76]. The demand profiles are distributed across all nodes proportional to GDP and population. With the available hourly electricity demand time series and the existing 2020 power plant fleet (validated in Section 4.4), the model calculates the optimal generator dispatch considering power flow constraints.

The dispatch validation shown in Table 3 compares the generation shares of the PyPSA-Earth results to those reported at *Our World in Data* [76]. The comparison highlights that PyPSA-Earth adequately represents the total electricity production shares by source in Nigeria with acceptable accuracy. Model results for the solar generation have a 100% accuracy compared to data provided by *Our World in Data*, gas generation is 2 TWh (10%) higher than the benchmark, while hydro generation is 0.3 TWh (5%) lower. These deviations could be explained by the 1.3 TWh (4%) higher assumption of total electricity demand and differences in the specific marginal costs of resources. Using the cost assumptions from [24], we derive an average marginal price for electricity of 59 €/MWh, which aligns with reported production costs in the range of 45 – 70 €/MWh [89].

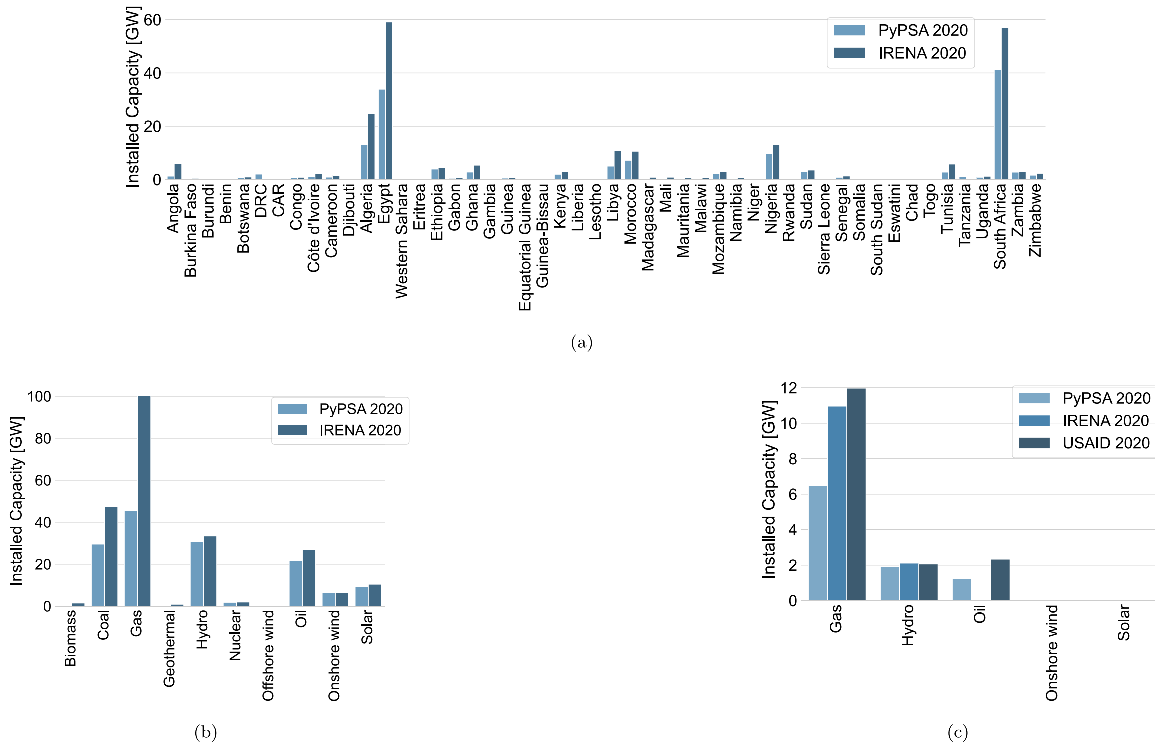


Fig. 8. Total installed generation capacity in Africa by (a) country and (b) technology, including a focus on (c) the installed capacities in Nigeria.

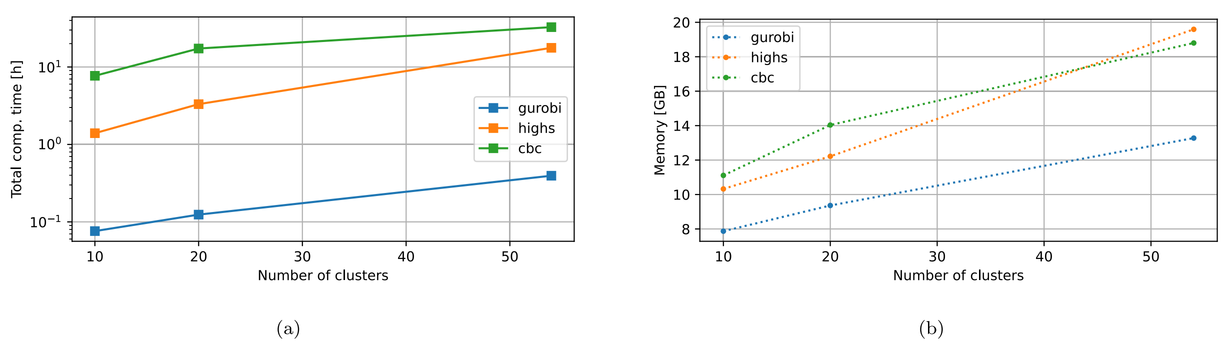


Fig. 9. Solution time (a) and memory requirements (b) for the 2020 Nigeria dispatch optimization for Gurobi 9.5.1, HiGHS 1.2.1 and CBC 2.10.8 solvers at different spatial resolution; solution time is weighted by threads.

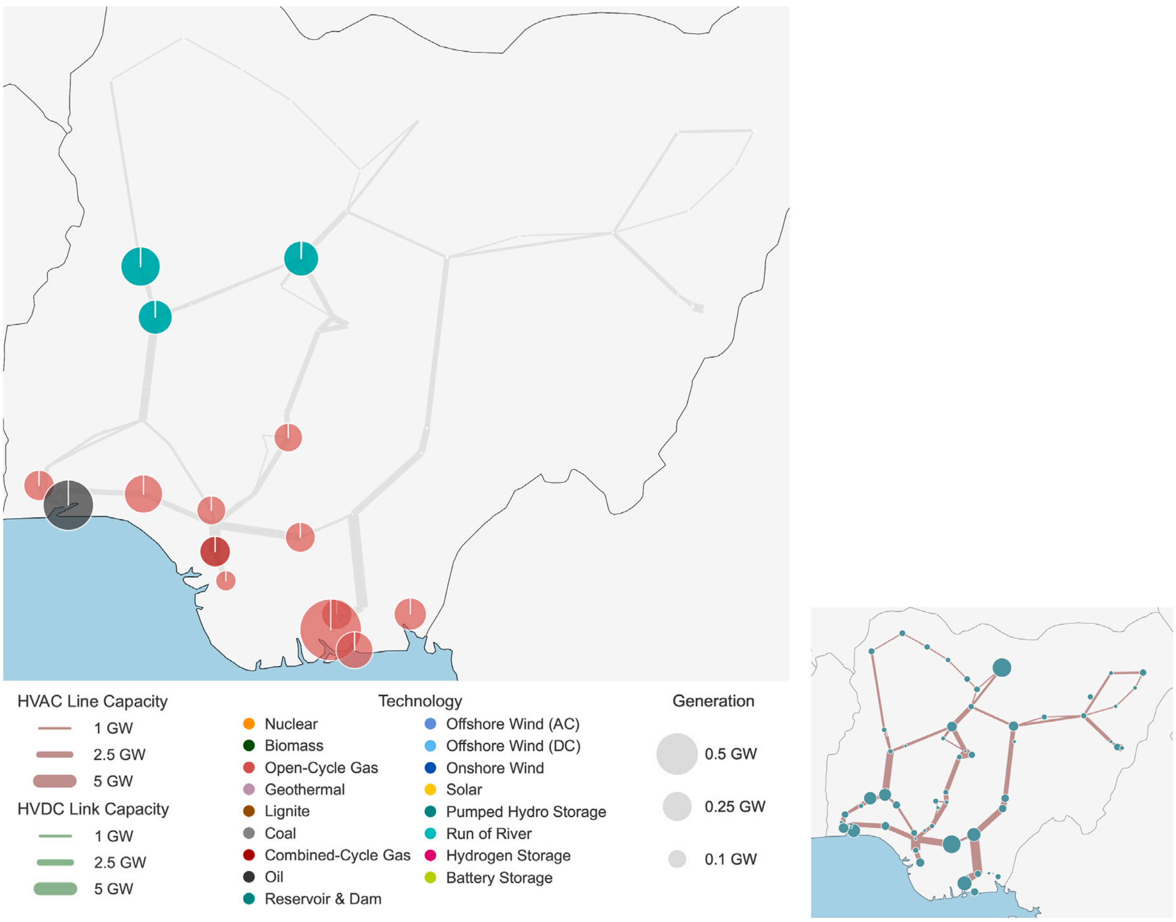
Table 3
Nigeria 2020 dispatch comparison.

	Total	Hydro	Coal	Gas	Wind	Solar
PyPSA-Earth [TWh]	29.5	5.8	–	23.6	0	0.04
Our World in Data [TWh]	28.2	6.1	0.6	21.4	0	0.04

The computational needs for this scenario in terms of total solving time (computational time times the average load of the processors) and memory are shown in Fig. 9. We used four threads with Gurobi 9.5.1 solver while only a single-core with HiGHS 1.2.1 and CBC 2.10.8, since their parallel solving capabilities are currently limited. While the commercial Gurobi solver is very efficient, the results in Fig. 9 confirm that the open-source HiGHS solver can also optimize the network below one day with memory requirements that are available for laptops. Given the expected improvements for open-source solvers, the computational requirements are likely to decrease significantly [40].

5.2. Nigeria 2060 - Net-zero study

In the 2060 net-zero scenario, we perform a brownfield capacity expansion optimization. This means new renewable energy and transmission capacity can be built on top of existing infrastructure. Simultaneously, a dispatch optimization is performed subject to linear optimal power flow constraints. To explore new transmission grid structures, the meshing strategy is modified such that each node connects HVAC lines to at least three nearest neighbouring nodes and that a random selection of far distance nodes above 600 km connects HVDC lines, see (b) in Fig. 11. Using a random selection for long-distance HVDC can help identify valuable line connections before applying any heuristic that might not find these. In addition to the net-zero emission constraint, the 2060 total demand has been calibrated in agreement to [90] to about 250 TWh by linear interpolating the stated energy policies of IEA for Nigeria [63]. Observing the optimized infrastructure in Fig. 10, the overall optimal least-cost power system can be mostly supplied with solar energy and a mix of battery energy storage. Hydrogen energy storage with steel tanks is included as an expansion



(a) (b)

Fig. 10. Optimization results of Nigeria's (a) 2020 power system. The coloured points represent installed capacities. (b) Shows all network options on a different scale as (a) with the total electricity consumption per node.

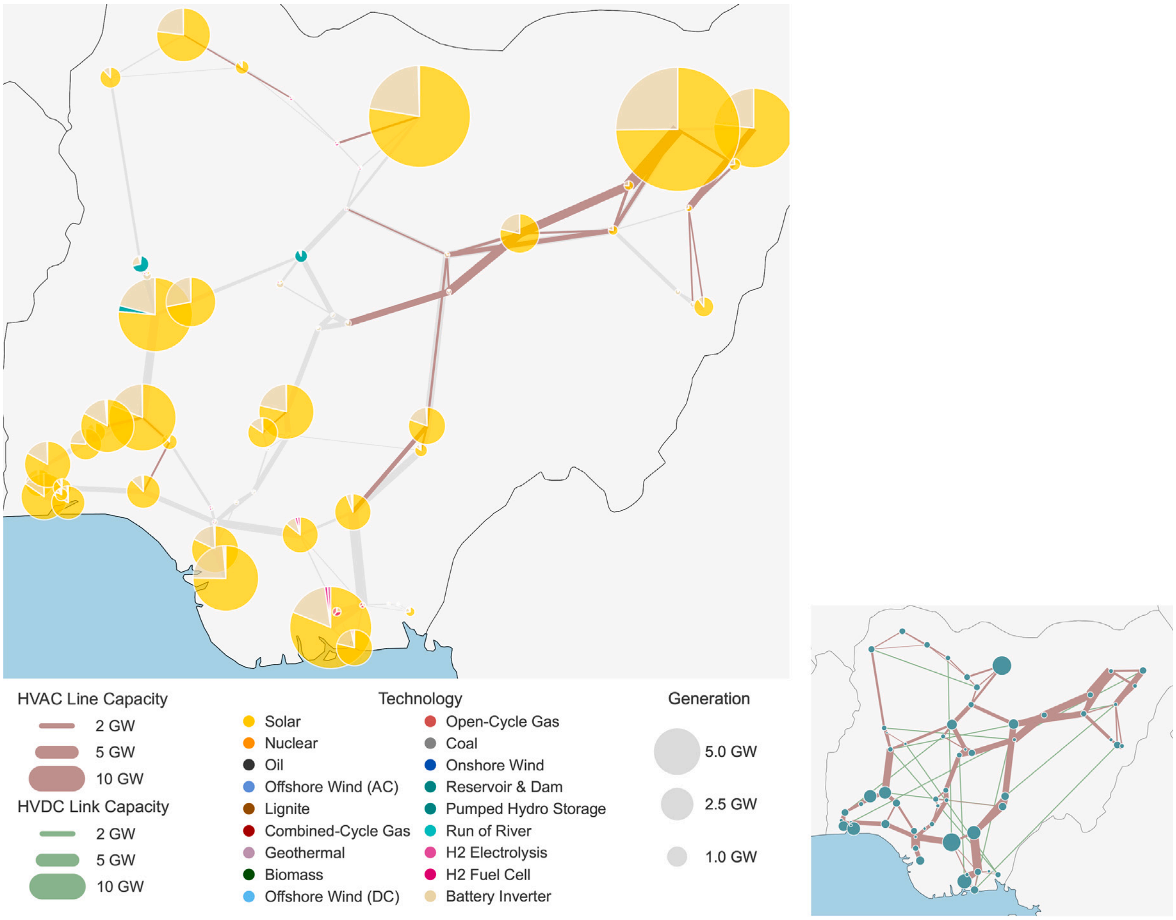
option. However, it is not significantly optimized, probably because we ignore fuel trade with other countries and the unique geo-location of the country. Nigeria lies close to the equator, where solar irradiation is homogeneous across the year, requiring less seasonal energy storage. The battery storage consists of an inverter component [€/MW] and a Li-Ion battery stack [€/MWh] that can be independently scaled by the model such as applied in [91]. The energy-to-power ratio (EP) indicating the sizing between these storage components is optimized in the range 4.5 h – 15.0 h with an average of 6.75 h. The total optimized Li-Ion battery storage discharging capacity and energy capacity is 67.9 GW and 459.7 GWh, respectively. The optimal solar capacity distribution is spatially uneven. Most solar is expanded in the country's north, where the solar potential is significantly higher [92]. It is also cost-optimal to build new transmission routes in the north and east of Nigeria, enabling the spatial distribution of electricity. The total optimized solar PV capacity is 256.9 GW, whereas about 20% of the solar energy is curtailed on an annual average. The HVDC options are not used significantly, indicating it is not cost-optimal in the scenario. Fig. 12 shows the dispatch profiles outputs of the optimization, which illustrates that most batteries charge during the day and discharge at night. Notably, to be conservative, with cost assumptions for 2050 [93], the average marginal prices reach only 51 €/MWh, compared to 59 €/MWh in the 2020 scenario. While we do not address in this demonstration uncertainty that could strengthen the results, future work can add modelling to generate alternatives and structured parameter sweeps [39] as well as robust optimization techniques [94]. Nevertheless, the results of the demonstration indicate that the optimized renewable electricity future for Nigeria could be cheaper than today.

6. Limitations and future opportunities

6.1. Missing network topology data

Modelling can only be as good as underlying data – the same applies to PyPSA-Earth. By relying on open sources to model energy systems, their data quality is a concern that we also acknowledge for the present paper. Yet, we also describe possible image recognition procedures to improve the data situation in PyPSA-Earth and potentially all energy models. This effort may complement the traditional effort by public institutions that disclose data of public relevance, such as installed network infrastructure, as performed by ENTSO-E [61].

Compared to Europe or North America, institutions that provide infrastructure data with geolocation for modelling have no analogues in Africa. The missing network data situation is limiting the use of energy system models. However, other types of data from which energy system components can be inferred exist on a much larger scale. Satellite imagery is one such data type. As part of the PyPSA meets Earth initiative, we are exploring opportunities to use neural network-based object detection applied to satellite imagery to enrich the existing datasets on energy infrastructure. In the past, such efforts have either been hard to apply on larger scales due to high requirements on manual input [95] or are coarse approximations to the true grid structure [96]. Under the umbrella of this initiative, we aim to develop precise and scalable methods and base our efforts on recent advances in the field.



(a) (b)

Fig. 11. Optimization result represent Nigeria's (a) 2060 power system. The coloured points represent installed capacities. Light grey and dark grey lines are existing and newly optimized transmission lines, respectively. (b) Shows all network options on a different scale as (a) with the total electricity consumption per node.

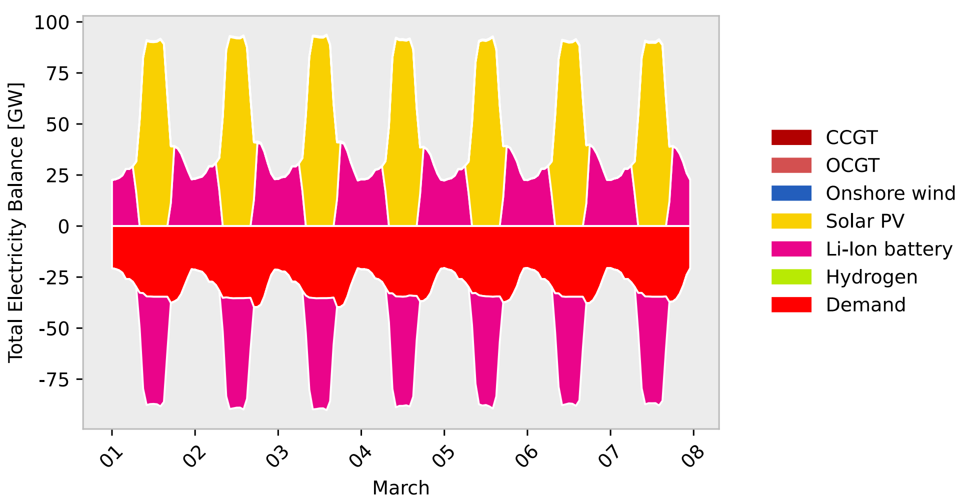


Fig. 12. Total electricity balance for the 2060 scenario. The time-series is hourly sampled for selected days in March including electricity supply (above zero) and consumption (below zero). CCGT and OCGT stands for closed and open cycle gas turbines, respectively.

6.2. Missing demand time series and prediction biases

Demand is a significant uncertainty factor in Africa due to the growth in magnitudes over the following decades that has implications on results created by PyPSA-Earth. Therefore, improving demand predictions is essential. We acknowledge this limitation in the data fed into our model and highlight research opportunities to address this challenge. PyPSA-Earth demand data is limited, as indicated in Section 3.4, by poor prediction performance for low-income countries due to input data biases and by missing machine learning output data for some low-demand countries due to software bugs. Additionally, while the open-source *synde* package [52] used in PyPSA-Earth extended the original GEGIS package for demand prediction by a workflow, there are opportunities to create a package focusing only on demand prediction substituting the GEGIS design that provides all energy model data in one package [29].

Developing a package focusing on sector-coupled energy demand forecast for macro-energy system modelling worldwide is an opportunity to improve the status quo of existing tools, not limited to the power sector. Further, instead of validating the demand with annual means, one should validate with hourly officially reported time series since the data quality is better indicated by the latter.

6.3. Imprecise global data

PyPSA-Earth relies on open data with global scope. This means that sometimes data is used that approximate country-specific details needed for national energy planning studies. While improving the global open data situation is one opportunity [97], another is to enable the integration of national and regional more precise data that can also be used as a source for validation. Therefore, PyPSA-Earth does not only use global data as default but will allow the integration of national or regional more precise data with specialized functions, here named "linkers".

6.4. Additional technologies

PyPSA-Earth includes the major transmission, generation and storage technologies. However, some system components are not yet included. Examples of not implemented generation technologies are Concentrating Solar Power (CSP), location-based geothermal, and other secondary technologies such as wave/tidal energy harvesting. While at a global scale these technologies represent a minor fraction, for country-specific analyses, they may have substantial implications, such as in the case of Kenya for geothermal or Morocco for CSP. Moreover, while currently only lithium-ion batteries and hydrogen energy storage are considered, other technologies may be considered and tested, such as the well-known Redox Flow batteries, Compressed-Air Energy Storage (CAES), Liquified-Air Energy Storage (LAES), that can have a large market in the future. Moreover, the dynamic calculation of the transmission capacity as a function of weather conditions [23], also known as Dynamic Line Rating (DLR) [98], still needs to be included.

These limitations, at the time of writing, represent future opportunities to improve the model and capture relevant technologies to perform detailed energy studies for all countries.

7. Conclusions

This paper presents the PyPSA-Earth model, which is an open-source global energy system model in high spatial and temporal resolution. It is making high-resolution modelling accessible to countries which so far had not detailed energy planning scenarios developed. Using a novel comprehensive workflow procedure PyPSA-Earth automatically downloads open data, provides model-ready data and integrates optimization features to address large-scale energy system planning. In agreement with the open-source spirit, the model is not built from scratch but derived from the European-focused PyPSA-Eur model adding global data as well as several new features.

The methodology is confirmed to be flexible and accommodate a high temporal and spatial resolution power model for national and regional energy planning with global scope. The validation performed for the African continent highlighted that PyPSA-Earth successfully provides power network and installed generation data that match trustworthy third-party national data with adequate accuracy, hence suggesting PyPSA-Earth to be a reliable model for energy planning. The 2020 and 2060 planning studies for Nigeria have further confirmed that net-zero emission scenarios for the electricity sector can be performed using PyPSA-Earth, leading to realistic results comparable with similar studies but in higher spatial detail. That further stresses the robustness of the approach and the flexibility of the methodology to be used in practical projects.

Given the need for reliable tools to foster the energy transition and the efficient use of resources, PyPSA-Earth can successfully support policymakers, utilities, and scholars in providing reliable, transparent, and efficient decision-making on energy studies. While several open-source projects are developed but discontinued, the authors of this paper and other PyPSA-Earth developers aim to foster collaborative energy system modelling on the same code-base to provide a well-maintained and robust tool rather than disperse resources across multiple models that get quickly outdated. Given the flexibility of the approach, additional improvements can be integrated, and scholars interested in contributing are invited to contact the model initiative team to join forces. Accordingly, this paper and the proposed tool can serve as a backbone, enabling further research and business activities to built on top to meet various energy transition planning needs that must be cheap and fast to develop for every nation and community on Earth.

Further studies may address the sector-coupled version of PyPSA-Earth. In addition to power and hydrogen data, sector-coupling requires various electric as well as non-electric demand and supply side data from other sectors, including heat, industry and/or transport. Other future work may interface energy modelling with economics modelling, add robust-optimization capabilities, improve demand forecasts in alignment with climate change scenarios, adapt imprecise global network data using object detection on satellite images or validate the model in other regions.

CRediT authorship contribution statement

Maximilian Parzen: Conceptualization, Project administration, Software development, Validation, Figure production, Writing and revising the manuscript. **Hazem Abdel-Khalek:** Software development, Validation, Figure production, Writing and revising the manuscript. **Ekaterina Fedotova:** Software development, Validation, Figure production, Writing and revising the manuscript. **Matin Mahmood:** Software development, Validation, Figure production, Writing and revising the manuscript. **Martha Maria Frysztacki:** Software development, Validation, Figure production, Writing and revising the manuscript. **Johannes Hampp:** Software development, Validation, Figure production, Writing and revising the manuscript. **Lukas Franken:** Writing and revising the manuscript. **Leon Schumm:** Software development, Validation, Figure production, Writing and revising the manuscript. **Fabian Neumann:** Conceptualization, Writing and revising the manuscript. **Davide Poli:** Writing and revising the manuscript. **Aristides Kiprakis:** Funding acquisition, Writing and revising the manuscript. **Davide Fioriti:** Conceptualization, Project administration, Software development, Validation, Figure production, Writing and revising the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Code and data to reproduce results and illustrations are available by using PyPSA-Earth v0.1 [34]. Further, instruction and configurations to reproduce the scenarios and plots are provided here: https://github.com/pz-max/pypsa-earth-paper.

Acknowledgements

This research was supported by UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/P007805/1 for the Centre for Advanced Materials for Renewable Energy Generation (CAMREG) and EPSRC, United Kingdom grant EP/V042955/1 DISPATCH. Maximilian Parzen would like to thank Tom Brown, Aminu Haruna Isa, Dahunsi Okekunle, Matija Pavičević, Michael Dioha and every one of our continuous supporters for their helpful comments and inspiring discussion.

Appendix A. The mathematical modelling

The following paragraphs are an extract of [91], which formulates PyPSA-Earth based on [3,23,99,100].

The objective of PyPSA-Earth is to minimize the total Annualized Costs (AC) of the system shown in (A.1), comprised of the annualized capital and operational expenditures. Capital expenditures include capacity-related, long-term investment costs c at location i for generator $G_{i,r}$ of technology r, storage energy capacity $H^{store}_{i,s}$, charging capacity $H^{+}_{i,s}$ and discharging capacity $H^{-}_{i,s}$ of technology s and transmission line F_l. Operational expenditures include energy-related variable cost o for generation $g_{i,r,t}$ and storage charging $h^{+}_{i,r,t}$ and discharging $h^{-}_{i,r,t}$, as well as energy-level related storage cost $e_{i,s,t}$. Thereby, the operation depends on the time steps t that are weighted by duration w_t that sums up to one year $\sum_{t=1}^{T} w_t = 365 day * 24 \frac{h}{day} = 8760$ h.

$$
\begin{aligned}
\underset{G,H,F,g,h,e}{\text{minimize}}\ AC = \Bigg[& \sum_{i,r} (c_{i,r} \cdot G_{i,r}) + \sum_{l} (c_l \cdot F_l) \\
& + \sum_{i,s} (c^{store}_{i,s} \cdot H^{store}_{i,s} + c^{-}_{i,s} \cdot H^{-}_{i,s} + c^{+}_{i,s} \cdot H^{+}_{i,s}) \\
& + \sum_{i,r,t} (o_{i,r} \cdot g_{i,r,t} \cdot w_t) + \sum_{i,s,t} \left((o^{+}_{i,s} \cdot h^{+}_{i,s,t} + o^{-}_{i,s} \cdot h^{-}_{i,s,t}) \cdot w_t \right) \\
& + \sum_{i,s,t} (o^{store}_{i,s} \cdot e_{i,s,t} \cdot w_t) \Bigg]
\end{aligned}
\tag{A.1}
$$

The objective function is subject to multiple linear constraints to make scenarios more realistic, leading to a convex linear program with continuous variables. The constraints explained in the following in more detail consist of (i) demand equals supply constraint, (ii) geophysical and operational constraint for generators, storage units as well as power lines, (iii) Kirchhoff's current and voltage law constraints that represent the physics of electric energy flows in the power network, (iv) a recovering cyclic energy storage constraint and finally, (v) greenhouse gas emissions reduction constraint. Such linear problems have in general one unique objective value with sometimes multiple non-unique operational solutions [100], making complex problems solvable in reasonable amount of time (sometimes multiple days).

Firstly, the PyPSA-Earth model guarantees the energy balance by the linearized power flow in Eqs. (A.2) and (A.3) to guarantee the energy balance of the power network, which is very distinctive feature compared to most other planning models [3]. This is done by including Kirchhoff's Current Law and Kirchhoff's Voltage Law constraints. The equivalent formulation of Kirchhoff's Current Law in (A.2) guarantees that the net energy balance for each substation i match, accounting for local generation, storage, transmission system and inelastic electricity demand $d_{i,t}$. The variable $f_{l,t}$ identifies the power flow in the line l having networks' incident matrix $K_{i,l}$.

$$
\sum_{r} g_{i,r,t} - \sum_{s} h^{+}_{i,s,t} + \sum_{s} h^{-}_{i,s,t} + \sum_{l} K_{i,l} \cdot f_{l,t} = d_{i,t} \quad \forall i, t \tag{A.2}
$$

While Kirchhoff's Current Law accounts for both, AC and controllable DC lines, the Kirchhoff's Voltage Law only additionally constraints AC power lines. In PyPSA-Earth the voltage angle difference around every closed cycle in the network must add up to zero. PyPSA-Earth formulates this constraint using linearized load flow assumptions, in particular, cycle basis $C_{l,c}$ of the network graph where the independent cycles c are expressed as directed linear combinations of lines [101]. This leads to the constraints (A.3), where x_l is the series inductive reactance of line l [99]. As might be noted, the linearized power flow assumptions completely disregard the resistance. These assumptions introduce negligible errors when (i) the reactance is much larger than the resistance, such as for high voltage lines, and (ii) the voltage angle differences are small i.e. $sin(\delta) = \delta$ [101].

$$
\sum_{l} C_{l,c} \cdot x_l \cdot f_{l,t} = 0 \quad \forall c, t \tag{A.3}
$$

Secondly, since generator and storage units as well as transmission lines can experience geographical restriction, PyPSA-Earth can constrain the installed capacities and gives the options for lower as well as upper limits. Equations from (A.4) to (A.6) specify the power capacities for generators, storages and lines, respectively.

$$
\underline{G}_{i,r} \le G_{i,r} \le \overline{G}_{i,r} \quad \forall i, r \tag{A.4}
$$

$$
\underline{H}_{i,s} \le H_{i,s} \le \overline{H}_{i,s} \quad \forall i, s \tag{A.5}
$$

$$
\underline{F}_{l} \le F_{l} \le \overline{F}_{l} \quad \forall l \tag{A.6}
$$

Thirdly, while the previous constraint only limits the installations, some components require time-varying operational limits. Examples for such technologies are renewable generators, described by the subset RG, and power lines with dynamic line-rating (DLR) [102], whose operation highly depend on the weather signals. With roughly 20x20 km globally rasterized era5 weather data that are available for more than 30 years, calculated using Atlite, PyPSA-Earth can limit the rated power of generators $G_{i,r}$ and lines F_l by a location and time dependent variable, i.e. temperature, wind speed, humidity and solar irradiation, as mathematically described in (A.7) and (A.8).

$$
0 \le g_{i,r,t} \le \overline{g}_{i,r,t} \cdot G_{i,r} \quad \forall i, r \in RG, t \tag{A.7}
$$

$$
0 \le f_{l,t} \le \overline{f}_{l,t} \cdot F_l \quad \forall l, t \tag{A.8}
$$

Fourth, describing storage constraints. Storage charging $h^{+}_{i,s,t}$ and discharging $h^{-}_{i,s,t}$ are both positive variables and limited by the installed capacity $H^{+}_{i,s,t}$ and $H^{-}_{i,s,t}$ in (A.9) and (A.10).

$$
0 \le h^{+}_{i,s,t} \le H^{+}_{i,s} \quad \forall i, s, t \tag{A.9}
$$

$$
0 \le h^{-}_{i,s,t} \le H^{-}_{i,s} \quad \forall i, s, t \tag{A.10}
$$

This formulation keeps the feasible solution space convex, though does not prevent simultaneous charging and discharging, which is often an unrealistic effect that can heavily distort modelling results in net-zero scenarios. Setting adequate variable cost parameter solves this modelling artefact while keeping the problem formulation linear [100].

The storage energy level $e_{i,s,t}$ is the result of a balance between energy inflow, outflow and self-consumption described in (A.11). Additional to directed charging and discharging with its respective efficiencies $\eta_{i,s,+}$ and $\eta_{i,s,-}$, natural inflow $h^{inflow}_{i,s,t}$, spillage $h^{spillage}_{i,s,t}$ as well as standing storage losses that reduces the storage energy content of the previous time step by a factor of $\eta_{i,s,+}$ are considered.

$$
\begin{aligned}
e_{i,s,t} \quad & = \eta_{i,s,+} \cdot e_{i,s,t-1} + \eta_{i,s,+} \cdot w_t \cdot h^{+}_{i,s,t} - \eta^{-1}_{i,s,-} \cdot w_t \cdot h^{-}_{i,s,t} \\
& + w_t \cdot h^{inflow}_{i,s,t} - w_t \cdot h^{spillage}_{i,s,t} \quad \forall i, s, t
\end{aligned}
\tag{A.11}
$$

The amount of energy that can be stored is limited by the energy capacity of the installed store unit $H^{store}_{i,s}$ [MWh], which allows independent storage component scaling shown in (A.12).

$$
0 \le e_{i,s,t} \le H^{store}_{i,s} \quad \forall i, s, t \tag{A.12}
$$

To fix the storage technology design, in (A.13) a technology-specific energy to discharging power ratio $\overline{T}_s$ is multiplied with the capacity of the discharging unit $H^{-}_{i,s}$ to define the upper energy limit per installed storage.

$$0 \leq e_{i,s,t} \leq \overline{T}_s \cdot H^{-}_{i,s} \quad \forall i, s, t \tag{A.13}$$

Further, Eq. (A.14) guarantees that the energy storage units are operated cyclically: the state of charge at the first and last period of the optimization period T (i.e. 1 year) must be equal.

$$e_{i,s,0} = e_{i,s,T} \quad \forall i, s \tag{A.14}$$

This cyclic definition is not mandatory but helps with the comparability of model results. It further avoids the free use of storage energy endowment, meaning that the model could prefer to start with a higher and end with a lower storage level to save costs.

Finally, PyPSA-Earth can constrain the total emissions. These emissions are tracked by a variable at each generator unit r and marginal emission intensity γ_r to constrain the total emission by a limiting parameter $\overline{GHG}$ as in (A.15).

$$g_{i,r,t} \cdot \gamma_r \leq \overline{GHG} \quad \forall i, r, t \tag{A.15}$$

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.apenergy.2023.121096.

References

[1] Pfenninger S. Energy scientists must show their workings. Nature 2017;542:393.
[2] International Renewabe Energy Agency. IRENA and IAEA to help African union develop continental power master plan with EU support. 2021, URL https://irena.org/newsroom/articles/2021/Sep/IRENA-and-IAEA-Selected-to-Help-African-Union-Develop-Continental-Power-Master-Plan-with-EU-support.
[3] Brown T, Hörsch J, Schlachtberger D. PyPSA: Python for power system analysis. J Open Res Softw 2018;6(4).
[4] European Commission. OPEN source software strategy 2020 – 2023 - Think open. Tech. rep., European Commission; 2020.
[5] Groissböck M. Are open source energy system optimization tools mature enough for serious use? Renew Sustain Energy Rev 2019;102:234–48.
[6] Kiviluoma J. Trends of open-sourcing models in the energy space. 2022, URL https://www.esig.energy/trends-of-open-sourcing-models-in-the-energy-space/.
[7] Pfenninger S, Hirth L, Schlecht I, Schmid E, Wiese F, Brown T, Davis C, Gidden M, Heinrichs H, Heuberger C, Hilpert S, Krien U, Matke C, Nebel A, Morrison R, Müller B, Pleßmann G, Reeg M, Richstein JC, Shivakumar A, Staffell I, Tröndle T, Wingenbach C. Opening the black box of energy modelling: Strategies and lessons learned. Energy Strategy Rev 2018;19:63–71.
[8] Victoria M, Zhu K, Brown T, Andresen GB, Greiner M. Early decarbonisation of the European energy system pays off. Nature Commun 2020;11(1):6223.
[9] Neumann F, Hagenmeyer V, Brown T. Assessments of linear power flow and transmission loss approximations in coordinated capacity expansion problems. Appl Energy 2022;314.
[10] Pfenninger S, Pickering B. Calliope: a multi-scale energy systems modelling framework. J Open Source Softw 2018;3(29):825.
[11] Dispa-SET contributors. Dispa-SET. 2022, URL https://github.com/energy-modelling-toolkit/Dispa-SET. [Accessed 17 May 2022].
[12] Blue Marble Analytics. Gridpath: Github repository. 2022, URL https://github.com/blue-marble/gridpath. [Accessed 30 July 2022].
[13] Heaps C. LEAP: The low emissions analysis platform. Stockholm Environment Institute; 2022, URL https://leap.sei.org. [Accessed 17 May 2022].
[14] Bogdanov D, Ram M, Aghahosseini A, Gulagi A, Oyewo AS, Child M, Caldera U, Sadovskaia K, Farfan J, De Souza Noel Simas Barbosa L, Fasihi M, Khalili S, Traber T, Breyer C. Low-cost renewable electricity as the key driver of the global energy transition towards sustainability. Energy 2021;227:120467.
[15] Stockholm Environment Institute. NEMO: Next energy modeling system for optimization. 2022, URL https://sei-international.github.io/NemoMod.jl/stable/. [Accessed 17 May 2022].
[16] Howells M, Rogner H, Strachan N, Heaps C, Huntington H, Kypreos S, Hughes A, Silveira S, DeCarolis J, Bazillian M, Roehrl A. OSeMOSYS: The open source energy modeling system: An introduction to its ethos, structure and development. Energy Policy 2011;39(10):5850–70, Sustainability of biofuels.
[17] Energy Exemplar. PLEXOS. 2022, URL https://www.energyexemplar.com/plexos. [Accessed 18 May 2022].
[18] Lincoln R. PYPOWER. 2022, URL https://github.com/rwl/PYPOWER,2017. [Accessed 02 November 2022].
[19] International Renewable Energy Agency. SPLAT-MESSAGE. 2022, URL https://irena.org/energytransition/Energy-System-Models-and-Data/System-Planning-Test-Model. [Accessed 17 May 2022].
[20] Loulou R, Goldstein G, Kanudia A, Lettila A, Remme U. Documentation for the TIMES model - PART I 1–78. Tech. rep., ETSAP; 2021.
[21] Victoria M, Zhu K, Brown T, Andresen GB, Greiner M. Early decarbonisation of the European energy system pays off. Nature Commun 2020;11(1):6223, Number: 1 Publisher: Nature Publishing Group.
[22] Ringkjøb HK, Haugan PM, Solbrekke IM. A review of modelling tools for energy and electricity systems with large shares of variable renewables. Renew Sustain Energy Rev 2018;96:440–59.
[23] Hörsch J, Hofmann F, Schlachtberger D, Brown T. PyPSA-Eur: An open optimisation model of the European transmission system. Energy Strategy Rev 2018;22:207–15.
[24] PyPSA-Eur. Github repository of PyPSA-Eur. 2021 [Accessed 15 August 2022].
[25] Hörsch J, Calitz J. PyPSA-ZA: Investment and operation co-optimization of integrating wind and solar in South Africa at high spatial and temporal detail. 2017, URL https://arxiv.org/abs/1710.11199.
[26] Schlott M, Schyska B, Viet DT, Phuong VV, Quan DM, Khanh MP, Hofmann F, Bremen Lv, Heinemann D, Kies A. PyPSA-VN: An open model of the Vietnamese electricity system. In: 2020 5th international conference on green technology and sustainable development. 2020, p. 253–8.
[27] Calliope-Kenya. Github repository of Calliope-Kenya. 2019 [Accessed 15 August 2022].
[28] Jacobson MZ, von Krauland A-K, Coughlin SJ, Palmer FC, Smith MM. Zero air pollution and zero carbon from all energy at low cost and without blackouts in variable weather throughout the U.S. with 100% wind-water-solar and storage. Renew Energy 2022;184:430–42.
[29] Mattsson N, Verendel V, Hedenus F, Reichenberg L. An autopilot for energy models – Automatic generation of renewable supply curves, hourly capacity factors and hourly synthetic electricity demand for arbitrary world regions. Energy Strategy Rev 2021;33.
[30] Reichenberg L, Hedenus F, Mattsson N, Verendel V. Deep decarbonization and the supergrid – Prospects for electricity transmission between Europe and China. Energy 2022;239.
[31] Kiviluoma J, Pallonetto F, Marin M, Savolainen PT, Soininen A, Vennström P, Rinne E, Huang J, Kouveliotis-Lysikatos I, Ihlemann M, Delarue E, O'Dwyer C, O'Donnel T, Amelin M, Söder L, Dillon J. Spine Toolbox: A flexible open-source workflow management system with scenario and data management. SoftwareX 2022;17.
[32] GENeSYS-MOD. Github repository of GENeSYS-MOD. 2022, URL https://git.tu-berlin.de/genesysmod/genesys-mod-public. [Accessed 06 May 2022].
[33] Barnes T, Shivakumar A, Brinkerink M, Niet T. OSeMOSYS Global, an open-source, open data global electricity system model generator. Sci Data 2022;9(1):623.
[34] PyPSA. Documentation: Users of PyPSA with case studies. 2022, URL https://pypsa.readthedocs.io/en/latest/users.html. [Accessed 01 July 2022].
[35] PyPSA-Earth. PyPSA-Earth GitHub repository. 2022, URL https://github.com/pypsa-meets-earth/pypsa-earth. [Accessed 30 August 2022].
[36] Zeyen E, Victoria M, Brown T. Endogenous learning for green hydrogen in a sector-coupled energy model for Europe. 2022, URL https://arxiv.org/abs/2205.11901.
[37] Glaum P, Hofmann F. Enhancing the german transmission grid through dynamic line rating. 2022, URL https://arxiv.org/abs/2208.04716.
[38] Neumann F. Costs of regional equity and autarky in a renewable European power system. Energy Strategy Rev 2021;35:100652.
[39] Neumann F, Brown T. Broad ranges of investment configurations for renewable power systems, robust to cost uncertainty and near-optimality. 2021, URL https://arxiv.org/abs/2111.14443.
[40] Parzen M, Hall J, Jenkins J, Brown T. Optimization solvers: The missing link for a fully open-source energy system modelling ecosystem. Zenodo; 2022.
[41] Steinmacher I, Treude C, Gerosa MA. Let me in: Guidelines for the successful onboarding of newcomers to open source projects. IEEE Softw 2019;36:41–9.
[42] Koster J, Rahmann S. Snakemake–a scalable bioinformatics workflow engine. Bioinformatics 2012;28:2520–2.
[43] Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S, Twardziok SO, Kanitz A, Wilm A, Holtgrewe M, Rahmann S, Nahnsen S, Köster J. Sustainable data analysis with snakemake. F1000Research 2021;10:33.
[44] Medjroubi W, Müller UP, Scharf M, Matke C, Kleinhans D. Open data in power grid modelling: New approaches towards transparent grid models. Energy Rep 2017;3:14–21.
[45] Rivera J, Nasirifard P, Leimhofer J, Jacobsen H-A. Automatic generation of real power transmission grid models from crowdsourced data. IEEE Trans Smart Grid 2018;10(5):5436–48.
[46] OpenStreetMap contributors. Planet dump retrieved. 2022, URL https://www.openstreetmap.org. [Accessed 06 May 2022].
[47] Pluta A, Lünsdorf O. Esy-osmfilter – A python library to efficiently extract OpenStreetMap data. J Open Res Softw 2020;8.

[48] GADM. Database of Global Administrative Areas. 2022, URL https://gadm.org/. [Accessed 15 August 2022].
[49] Frysztacki MM, Hörsch J, Hagenmeyer V, Brown T. The strong effect of network resolution on electricity system models with high shares of wind and solar. Appl Energy 2021;291.
[50] GADM. Maritime boundaries v11. 2019, URL https://www.marineregions.org/downloads.php. [Accessed 15 August 2022].
[51] Keywan R, et al. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Global Environ Change 2017;42:153–68.
[52] Hampp J. Synde. An open source package to create synthetic demand time-series. 2022, URL https://github.com/euronion/synde. [Accessed 07 June 2022].
[53] Toktarova A, Gruber L, Hlusiak M, Bogdanov D, Breyer C. Long term load projection in high resolution for all countries globally. Int J Electr Power Energy Syst 2019;111:160–81.
[54] 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 2021;6(62):3294.
[55] Hans H, et al. The ERA5 global reanalysis. Q J R Meteorol Soc 2020;146:1999–2049.
[56] Pfeifroth U, Kothe S, Müller R, Trentmann J, Hollmann R, Fuchs P, Werscheck M. Surface radiation data set - heliosat (SARAH) - edition 2. 2017, URL https://wui.cmsaf.eu/safira/action/viewDoiDetails?acronym=SARAH_V002. [Accessed 27 July 2022].
[57] GEBCO Compilation Group. GEBCO 2022 grid. 2022, URL https://www.gebco.net/data_and_products/gridded_bathymetry_data/. [Accessed 27 July 2022].
[58] Energy Information Administration. International - U.S. Energy Information Administration (EIA). 2022, URL https://www.eia.gov/international/overview/world. [Accessed 15 August 2022].
[59] 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 Strategy Rev 2019;23:1–12.
[60] Center For Global Development. Carbon monitoring for action | center for global development | ideas to action. 2012, URL https://www.cgdev.org/topics/carbon-monitoring-action. [Accessed: 15 August 2022].
[61] ENTSO-E. ENTSO-E - home. 2022, URL https://www.entsoe.eu/.
[62] Global Energy Observatory. Power plants data. 2018, URL http://globalenergyobservatory.org/docs/HelpGeoPower.php.
[63] IRENA. Data and statistics. 2022, URL https://www.irena.org/statistics. [Accessed 15 August 2022].
[64] Powerplantmatching. Adding Open Street Map (OSM) as a source · issue #12 · FRESNA/powerplantmatching. 2019, URL https://github.com/FRESNA/powerplantmatching/issues/12. [Accessed 15 August 2022].
[65] Hörsch J, Neumann F, Hofmann F, Schlachtberger D, Brown T. PyPSA-Eur: An open optimisation model of the European transmission system (code). 2020, URL https://doi.org/10.5281/zenodo.3885701#.XzKzdXCPvKE.mendeley.
[66] Frysztacki MM, Recht G, Brown T. A comparison of clustering methods for the spatial reduction of renewable electricity optimisation models of Europe. Energy Inf 2022;5(1):4.
[67] Siala K, Mahfouz MY. Impact of the choice of regions on energy system models. Energy Strategy Rev 2019;25:75–85.
[68] Kueppers M, Perau C, Franken M, Heger HJ, Huber M, Metzger M, Niessen S. Data-driven regionalization of decarbonized energy systems for reflecting their changing topologies in planning and optimization. Energies 2020;13(16):4076.
[69] Biener W, Garcia Rosas KR. Grid reduction for energy system analysis. Electr Power Syst Res 2020;185:106349.
[70] Fioriti D, Lutzemberger G, Poli D, Duenas-Martinez P, Micangeli A. Coupling economic multi-objective optimization and multiple design options: A business-oriented approach to size an off-grid hybrid microgrid. Int J Electr Power Energy Syst 2021;127.
[71] Linopy. Benchmark of Linopy: A package doing linear optimization with N-D labeled arrays in Python. 2022, URL https://linopy.readthedocs.io/en/latest/benchmark.html. [Accessed 01 July 2022].
[72] Global Power System Transformation Consortium. List of open-source tools for energy system modelling. 2022, URL https://g-pst.github.io/tools/. [Accessed 19 August 2022].
[73] ENTSO-E. ENTSO-E transparency platform. 2020, URL https://transparency.entsoe.eu/dashboard/show.
[74] World Bank Group. Africa - Electricity transmission and distribution grid map. 2022, URL https://energydata.info/dataset/africa-electricity-transmission-and-distribution-grid-map-2017. [Accessed 06May 2022].
[75] Nigerian Electricity Regulatory Commission. Nigerian transmission system information. 2022, URL https://nerc.gov.ng/index.php/home/nesi/404-transmission. [Accessed 06 May 2022].
[76] Ritchie H, Rosado P, Mathieu E, Roser M. Data on energy by our world in data. 2022, URL https://github.com/owid/energy-data/tree/master. [Accessed 20 May 2022].
[77] IRENA. Africa 2030: Roadmap for a renewable energy future. Tech. rep., IRENA; 2015.
[78] Alova G, Trotter PA, Money A. A machine-learning approach to predicting Africa's electricity mix based on planned power plants and their chances of success. Nat Energy 2021;6:158–66.
[79] Global Wind Energy Council. Exploring Africa's untapped wind potential. 2022, URL https://gwec.net/wp-content/uploads/2021/04/IFC-Africa-Wind-Technical-Potential-Oct-2020-1.pdf. [Accessed 27 August 2022].
[80] IRENA. Estimating the renewable energy potential in Africa: A GIS-based approach. Tech. rep., International Renewable Energy Agency; 2014 [Accessed 20 May 2022].
[81] Bolinger M, Bolinger G. Land requirements for utility-scale PV: An empirical update on power and energy density. IEEE J Photovolt 2022;12(2):589–94.
[82] Wikipedia. List of photovoltaic power stations. 2022, URL https://en.wikipedia.org/wiki/List_of_photovoltaic_power_stations. [Accessed 04 June 2022].
[83] PyPSA-Earth authors. Validation of the solar and wind potential. 2022, URL https://github.com/pypsa-meets-africa/pypsa-africa/blob/3f01c936850ceaa48774ea21a37a9e0bb6462d92/notebooks/validation/renewable_potential_validation.ipynb. [Accessed 27 August 2022].
[84] Reich P, Numbem S, Almaraz R, Eswaran H. Land resource stresses and desertification in Africa. Agro-Sci. 2004;2.
[85] IRENA. Installed electricity capacity (MW) by Country/area, Grid connection, Technology and Year. PxWeb. 2022, URL https://pxweb.irena.org/pxweb/en/IRENASTAT/IRENASTAT__Power%20Capacity%20and%20Generation/ELECCAP_2022_cycle1.px/. [Accessed 15 August 2022].
[86] USAID. Nigeria – Power Africa fact sheet. 2022, URL https://www.usaid.gov/powerafrica/nigeria.
[87] Global Energy Monitor. Global energy monitor Wiki. 2022, URL https://www.gem.wiki/Main_Page. [Accessed 28 August 2022].
[88] Climate Analytics and New Climate Institute. Climate action tracker. Report on 11.02.2022 for Nigeria. 2022, URL https://climateactiontracker.org/countries/nigeria/. [Accessed 15 Julyb 2022].
[89] Nigerian Economic Summit Group. Comparison of costs of electricity generation in Nigeria. Tech. rep., Nigerian Economic Summit Group; 2017.
[90] Kim J, Abdel-Hameed A, Joseph SR, Ramadhan HH, Nandutu M, Hyun JH. Modeling long-term electricity generation planning to reduce carbon dioxide emissions in Nigeria. Energies 2021;14.
[91] Parzen M, Neumann F, Van Der Weijde AH, Friedrich D, Kiprakis A. Beyond cost reduction: improving the value of energy storage in electricity systems. Carbon Neutrality 2022;1(1):26.
[92] Hamisu Umar N, Bora B, Banerjee C, Gupta P, Anjum N. Performance and economic viability of the PV system in different climatic zones of Nigeria. Sustain Energy Technol Assess 2021;43:100987.
[93] lisazeyen, euronion, Neumann F, Brown T, martavp, lukasnacken. PyPSA/technology-data: Technology Data v0.4.0. Zenodo; 2022.
[94] Riepin I, Schmidt M, Baringo L, Müsgens F. Adaptive robust optimization for European strategic gas infrastructure planning. Appl Energy 2022;324:119686.
[95] Wronkiewicz M. Mapping the electricity grid from space. Development Seed and World Bank; 2018, URL http://devseed.com/ml-grid-docs/. [Accessed 15 August 2022].
[96] Arderne C, Zorn C, Nicolas C, Koks EE. Predictive mapping of the global power system using open data. Sci Data 2020;7:19.
[97] Ritchie H. COVID's lessons for climate, sustainability and more from our world in data. Nature 2021;598:9.
[98] Poli D, Pelacchi P, Lutzemberger G, Baffa Scirocco T, Bassi F, Bruno G. The possible impact of weather uncertainty on the dynamic thermal rating of transmission power lines: A Monte Carlo error-based approach. Electr Power Syst Res 2019;170:338–47.
[99] Neumann F, Brown T. The near-optimal feasible space of a renewable power system model. Electr Power Syst Res 2021;190.
[100] Parzen M, Kittel M, Friedrich D, Kiprakis A. Reducing energy system model distortions from unintended storage cycling through variable costs. IScience 2023;26(1):105729.
[101] Hörsch J, Ronellenfitsch H, Witthaut D, Brown T. Linear optimal power flow using cycle flows. Electr Power Syst Res 2018;158:126–35.
[102] Glaum P, Hofmann F. Enhancing the german transmission grid through dynamic line rating. In: 2022 18th international conference on the European energy market. IEEE; 2022, p. 1–7.