Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Reproducible Workflow for Regionalisation of European Network Expansion Scenarios

Eric

eric.jahnke@kit.edu

Chair of Energy Economics, Institute for Industrial Production, Karlsruhe Institute of Technology
https://orcid.org/0009-0006-8901-8128

Research Article

Keywords: Data regionalisation, heuristics, network expansion scenarios

Posted Date: May 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9791175/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

A Reproducible Workflow for Regionalisation of European Network Expansion Scenarios

Eric Jahnke

Chair of Energy Economics, Institute for Industrial Production, Karlsruhe Institute of Technology, Karlsruhe, Germany; eric.jahnke@kit.edu

Table of Contents

Abstract 2
Abbreviations 2
Nomenclature 3
Introduction 5
General workflow 5
Regionalisation heuristics 7
RES 7
Hydro 8
Thermal units 8
Other RES and other non-RES 9
BESS 9
Load and DSR 10
Summary, Outlook and Limitations 10
References 10
Appendix 13
A1: Algorithms 13
Algorithm: Electrical grid reduction 13
Algorithm: RES regionalisation 13
Algorithm: Hydro regionalisation 14
Algorithm: Thermal units regionalisation 15
Algorithm: Other non-RES regionalisation 15
Algorithm: Other RES regionalisation 16
Algorithm: BESS regionalisation 16
Algorithm: Load regionalisation 17
Algorithm: DSR regionalisation 17
A2: Grid reduction 18

A2.1: Cluster budget allocation 18
A2.2: Spectral variant 18
A2: DC-effective-reactance variant 18
A2.3: Line equivalents and HVDC handling 19
A3: Raster-based RES and hydro profiles 19
A3.1: Availability masks 19
A3.2: Investment resource classes 20
A3: Load 20

Abstract

European power system planning and operation studies require spatially consistent input data that connect national scenario assumptions, weather-dependent renewable profiles, hydro inflows, and detailed grid representations. This paper presents a reproducible preprocessing workflow for deriving nodal input data for large-scale European unit commitment and optimal power flow analyses based on capacity and transmission expansion scenarios from the ten-year network development plan by ENTSO-E and ENTSO-G. The workflow combines TYNDP 2024 National Trends data with a *PyPSA*-based grid model, extended by selected grid expansion assumptions and maps all relevant demand, generation, storage, and flexibility data to the nodes of a reduced grid with user-specific size. The methodology covers two complementary spatial transformations. First, raster-based renewable energy potentials, capacity factors, and hydro inflow profiles are generated using an extended approach based on *Atlite* functionalities and aggregated to grid nodes using technology-specific availability masks. Second, national scenario quantities, including load, thermal generation, hydro capacities, other renewable and non-renewable technologies, battery storage, and demand-side response, are disaggregated to nodes using transparent allocation rules based on existing assets, population, GDP, distance, renewable potentials, load shares, and hydro shares. The resulting data pipeline produces consistent nodal timeseries, installed capacities, and technology attributes for subsequent optimisation. By separating grid reduction, geo-spatial profile generation, and technology-specific disaggregation, the workflow supports reproducible scenario construction, sensitivity analysis, and transparent validation of European power system optimisation studies.

Abbreviations

Abbreviation	Meaning
AC	Alternating current
BESS	Battery energy storage system
DC	Direct current
DSR	Demand-side response
EEZ	Exclusive economic zone

Abbreviation	Meaning
ENTSO-E	European Network of Transmission System Operators for Electricity
ENTSO-G	European Network of Transmission System Operators for Gas
GDP	Gross domestic product
HV	High voltage
NUTS3	Third-level Nomenclature of Territorial Units for Statistics region
OPF	Optimal power flow
PECD	Pan-European Climate Database
PEMMDB	Pan-European Market Modelling Database
PHS	Pumped hydro storage
PV	Photovoltaic
RES	Renewable energy sources
ROR	Run-of-river
TYNDP	Ten-Year Network Development Plan
UC	Unit commitment
WDPA	World Database on Protected Areas
WOFF	Wind offshore
WON	Wind onshore
WR	Water reservoir

Nomenclature

Symbol	Meaning
	Sets and indices
$\mathcal{M}, m$	Model countries or country aggregates and country index
$\mathcal{B}, b, b'$	AC buses or reduced AC buses and bus indices
$\mathcal{B}_m$	Reduced AC buses assigned to model country m
$\mathcal{E}, \ell$	AC lines/corridors and line index
$\mathcal{I}, i$	Raster cells and cell index
$\mathcal{I}_{b,r}, \mathcal{U}_{q,r}$	Eligible cells assigned to bus b and unsaturated cells in resource class q
$\mathcal{P}, p$	Population/GDP proxy regions and proxy-region index
$\mathcal{G}, g$	Grid-reduction groups and group index
$\mathcal{R}, r$	Renewable technologies and RES technology index
$\mathcal{H}, h$	Hydro plant types and hydro type index
t, y	Hour and weather-year indices
$j = (m, h, o)$	Hydro target tuple with model country m, hydro type h, and operating class o
$n = (m, f, e)$	Thermal target group with model country m, fuel type f, and

Symbol	Meaning
	technology type e
$d = (m, u, z)$	DSR price-band tuple with model country m, product class u, and price band z
s	Candidate-stage index in staged thermal allocation
	Grid
K, k_g, n_g	Requested total cluster budget, group cluster budget, and bus count of group g
$C_m^{\mathrm{gen}}, C_g^{\mathrm{gen}}$	Existing generation-capacity weights at country and group level
$\overline{L}_m$	Mean electrical load used for country-level cluster-budget weighting
α_G, α_L	Generation and load weights for distributing the cluster budget across countries
ω_m, ρ_g	Country and within-country group weights for cluster-budget allocation
$\boldsymbol{W}, \boldsymbol{D}, \boldsymbol{L}^{\mathrm{sym}}, \boldsymbol{B}$	Weighted adjacency, degree matrix, normalised graph Laplacian, and DC susceptance Laplacian
$w_{bb'}^{\mathrm{el}}, w_{bb'}^{\mathrm{eff}}, d_{bb'}^{\mathrm{eff}}$	Electrical similarity, DC-effective similarity, and DC-effective-reactance distance
	Regionalisation heuristics
P, E	Installed power capacity and storage energy capacity
$P^{\max}, P^{\mathrm{cur}}, P^{\mathrm{add}}, P^{\mathrm{scen}}$	Maximum, current, added, and scenario capacity
$h_{i,r}^{\mathrm{RES}}, \Delta P_{m,r}^{\mathrm{rem}}$	Renewable headroom in cell i and remaining national capacity addition
$A_{i,r}, M_{i,\cdot}, A_i^{\mathrm{cell}}, \kappa_r^{\mathrm{cap}}$	Cell availability, active availability masks, cell area, and capacity density
$cf_{i,t,r,y}, CF_{b,t,r,y}$	Cell-level and bus-level capacity factors
$G_{m,t,r,y}^{\mathrm{TYNDP}}, g_{b,t,r,y}$	National TYNDP RES generation and scaled bus-level generation
$L_{m,t,y}^{\mathrm{TYNDP}}, L_{b,t,y}$	National TYNDP load and disaggregated bus-level load
$\mu_{pb}, \lambda_{b,m}^{\mathrm{pop}}, \lambda_{b,m}^{\mathrm{gdp}}, \lambda_{b,m}$	Proxy-to-bus distance weight, proxy shares, and static bus load share
$\eta_{\mathrm{pop}}, \eta_{\mathrm{gdp}}, \gamma$	Population/GDP load-share weights and distance-decay exponent
$\sigma_{b,j}^{P}, \sigma_{b,j}^{E}$	Turbine-capacity and storage-energy shares for hydro target tuple j
$\chi_{b,n}^{(s)}, \chi_b^{\mathrm{RES}}, \chi_b^{\mathrm{BESS}}$	Thermal-stage, RES, and BESS allocation bases
$R_{b,f}^{\mathrm{nonRES}}, H_b$	Residual other-non-RES fuel basis and free renewable headroom at bus b
$P_{b,d}^{\mathrm{DSR}}, \overline{P}_{b,d,t}^{\mathrm{DSR}}$	Installed and hourly available DSR capacity at bus b and price band d

Symbol	Meaning
$q_{i,r}, Q_r, \tau$	Investment resource class, number of RES classes, and DC-effective similarity-kernel width

Introduction

Optimal power flow (OPF) with unit commitment (UC) studies are used to assess operating conditions under high renewable penetration, capacity expansion with changing hydro- and thermoelectric generation, transmission expansion and the impact of contingencies as well as the development of frequency response products [1, 2]. Such studies require a coherent nodal data set: installed capacities of thermal and hydro power plants, RES, BESS; transmission data; time series for electrical demand, RES capacity factors, hydro inflows; and more. Inconsistent spatial allocation of power system data makes it impossible to interpret optimization results or even find a feasible solution for the problem [3], [4], [5]. Existing nodal data sets in the literature are based on individual assumptions and modeling, reflect only the status quo or a specific historical capacity year, and cannot be compared directly with the TYNDP, as they are merely inputs for capacity and network expansion planning rather than the outcome of such planning [6, 7]. TYNDP scenario data is published at country, bidding-zone, control-area, or offshore-zone level. Therefore, a reproducible workflow for deriving nodal input data based on network expansion scenarios has been implemented. The target use case combines TYNDP 2024 National Trends scenario data for 2030 and 2040 [8]. The paper contributes a transparent data disaggregation concept with regionalisation heuristics, fallbacks, and diagnostics, that can also be adopted for other network expansion scenarios with national resolution.

General workflow

The workflow is designed in several layers and streams. The four streams consist of the development and preparation of the current European transmission grid; the calculation of capacity factors for PV, WON and WOFF, as well as country- and technology-specific inflow profiles for different historical weather years, including the derivation of investment resource classes and maximum installable capacities for PV, WON and WOFF in a grid with a user-defined spatial resolution. Additionally, site-specific information of current thermo- and hydroelectric power plant and battery-electric storage lists are updated and societal proxies about population and economic strength processed. The whole regionalisation workflow is shown in fig. 1.

First, a dataset of the physical transmission grid is taken from [9] and enriched with future grid expansion projects [10]. The grid can be reduced based on simple geo-based or graph-based clustering methods such as k-means or the maximum spanning tree [11] or more complex clustering methods such as spectral [12] or effective-resistance [13]. A target number of AC buses can be defined. DC terminals are not clustered themselves, but are assigned to their AC connection nodes; HVDC links are then re-routed between the reduced AC clusters or removed if both ends end up in the same cluster. The allocation of is a two-stage budget allocation: first, the regional budget is allocated

according to a weighted average generation capacity and load from 2025; then, the group budget within the region is allocated according to generation sites.

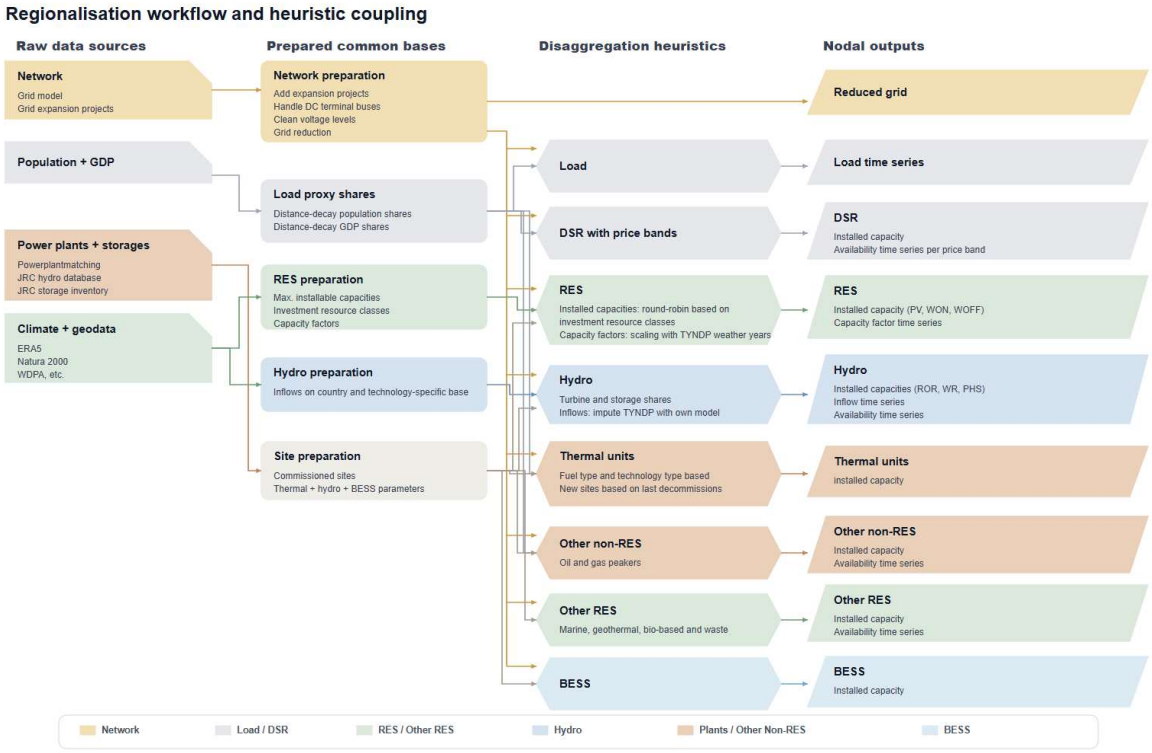


Fig. 1: Workflow of data disaggregation

Second, renewable capacity potentials, capacity factors, investment resource classes as well as country and technology-specific hydro inflow profiles are prepared from reanalysis from [14] for the weather years 1982 to 2016. The workflow is based on the *Atlite* package [15] and includes various datasets covering land use [16, 17], onshore and offshore topology [18], nature reserves [19] and other protected areas [20], marine traffic [21] and offshore zones [22].

Third, a power plant and storage list from the *powerplantmachting* package [23] is enriched with updated information from the *JRC* hydro database [24] and capacity data from [25] as well as the *JRC* energy storage inventory list [26] (last updated in May 2026) that covers various forms of storage.

Fourth, high-resolution population data from *JRC* [27] and estimated gridded GDP data [28] are pre-processed as proxies for the load disaggregation based on the parametrization from the *PYPSA* framework.

As there are significant gaps in the data for some countries (Moldova, Ukraine) or they cannot easily be considered in isolation for other reasons (Kosovo, Luxembourg), country aggregates can be defined so that capacities, load, etc. are cumulated for these artificial countries. Fig. 2 shows a reduced and regionalised network topology map for the target year 2030.

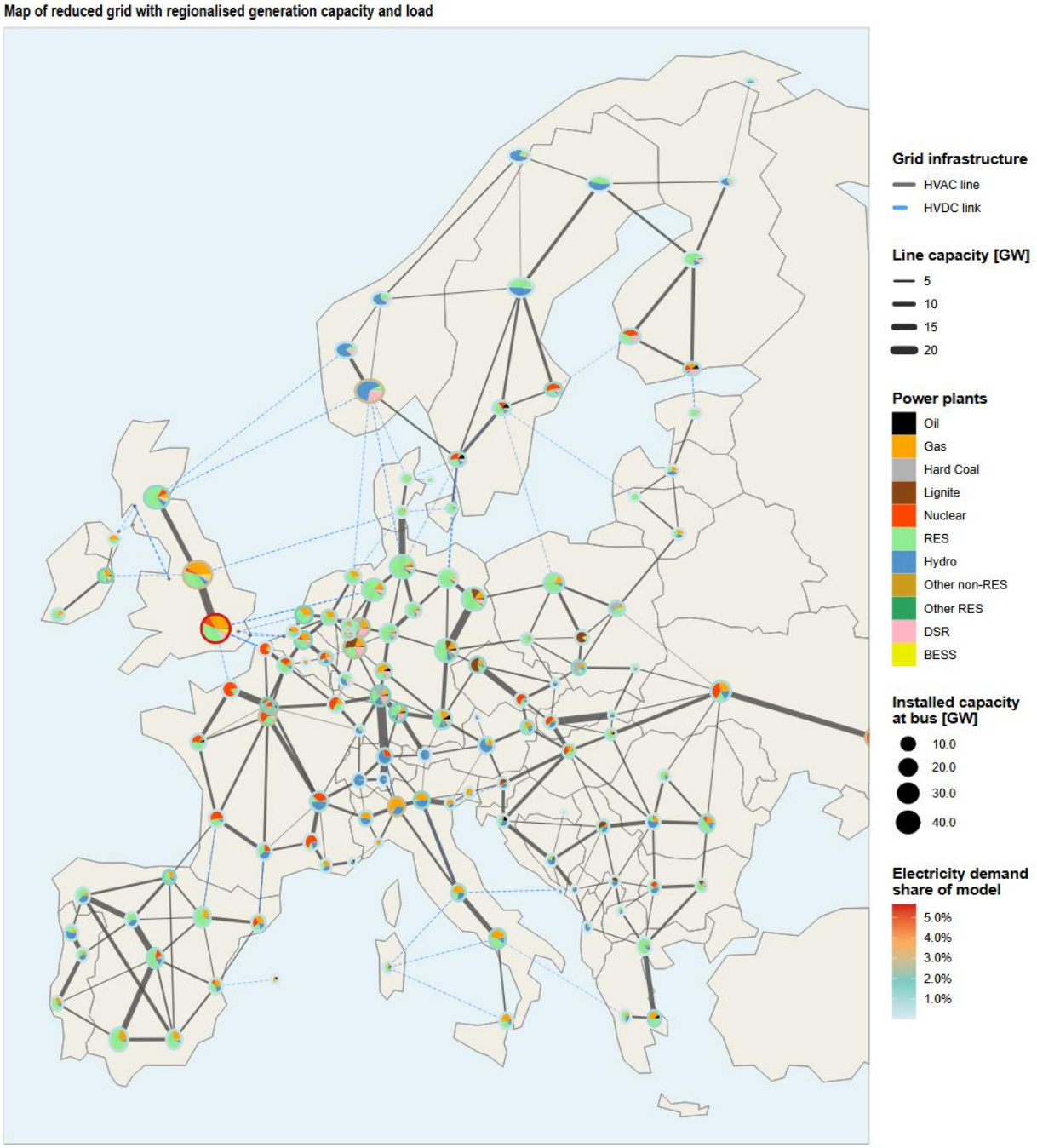


Fig. 2: Reduced network topology with 128 AC buses for the target year 2030

Regionalisation heuristics

RES

Let $\mathcal{I}_{b,r}$ be the eligible cells assigned to bus b. Existing capacity at the bus is first distributed to its cells proportional to the capacity potential:

$$P_{i,r}^{\text{cur}} = P_{b,r}^{\text{cur}} \frac{P_{i,r}^{\max}}{\sum_{\tilde{\imath}\in\mathcal{I}_{b,r}} P_{\tilde{\imath},r}^{\max}}, \qquad i \in \mathcal{I}_{b,r}. \tag{1}$$

If the potential denominator is zero, the current capacity is distributed equally over the available cells. Capacity is placed in cells by descending investment resource class. In class q, each unsaturated cell receives an equal increment capped by remaining headroom:

$$P_{i,r}^{\mathrm{add}} \leftarrow P_{i,r}^{\mathrm{add}} + \min\left(h_{i,r}^{\mathrm{RES}}, \frac{\Delta P_{m,r}^{\mathrm{rem}}}{|\mathcal{U}_{q,r}|}\right), \qquad h_{i,r}^{\mathrm{RES}} = P_{i,r}^{\max} - P_{i,r}^{\mathrm{cur}} - P_{i,r}^{\mathrm{add}}, \tag{2}$$

where $\mathcal{U}_{q,r}$ is the set of unsaturated cells in the current class and $\Delta P_{m,r}^{\mathrm{rem}}$ is the remaining country addition. If all classes saturate before the TYNDP target is reached, the remaining capacity is reported as unallocated rather than overbuilt. For every weather year y and hour t, a bus capacity factor is computed as a scenario-capacity-weighted average of the assigned cells:

$$CF_{b,t,r,y} = \frac{\sum_{i \in \mathcal{I}_{b,r}} cf_{i,t,r,y} P_{i,r}^{\mathrm{scen}}}{\sum_{i \in \mathcal{I}_{b,r}} P_{i,r}^{\mathrm{scen}}}. \tag{3}$$

The bus generation is calculated and scaled to TYNDP national generation:

$$g_{b,t,r,y} = G_{m,t,r,y}^{\mathrm{TYNDP}} \frac{(CF_{b,t,r,y} P_{b,r}^{\mathrm{scen}})}{\sum_{b' \in m} (CF_{b',t,r,y} P_{b',r}^{\mathrm{scen}})}. \tag{4}$$

If a country has a valid profile but the raw total is zero, no capacity fallback is used because the weather profile itself implies zero generation. If no valid profile exists, generation is distributed by scenario bus capacity.

Hydro

The pre-processing distinguishes between ROR, WR, PHS, and separates OL from CL technologies. A major problem with the regionalisation of PHS is that, even in comprehensive hydro datasets, there is no detailed information on whether a site receives natural inflows or is closed [24, 29]. PHS handling is therefore controlled by an optional parameter: If PHS is resolved, PHS-OL is mapped to WR, while PHS-CL is mapped to ROR if no WR but ROR exists, otherwise to WR. If PHS is not resolved and a country contains both OL and CL PHS, PHS-OL is collapsed into PHS-CL; former natural inflow is redirected to WR or ROR where available.

Hydro inflows from TYNDP are interpolated if possible, and otherwise imputed with self-generated inflows. As the TYNDP inflows are derived from the PECD and are based primarily on ERA5, the data sources and availability are very similar. The bus-level total inflow uses storage shares if any storage basis exists, otherwise turbine shares. If multiple hydro technologies at the same bus receive the bus total, the amount is split by technology-specific turbine shares.

Thermal units

The heuristic first searches the current plant list for a fuel–technology match and then for a fuel-only match. All candidates are restricted to buses with positive load share, so pure DC converter terminals cannot receive generic units. For candidates, bus weights are derived from current capacity and plant-count basis:

$$\chi_{b,n}^{(s)} = \frac{1}{2}\frac{P_{b,n}^{\text{basis}}}{\sum_{b'} P_{b',n}^{\text{basis}}} + \frac{1}{2}\frac{N_{b,n}^{\text{basis}}}{\sum_{b'} N_{b',n}^{\text{basis}}}, \quad (5)$$

when both denominators are positive. If only one denominator is positive, the available basis is used alone. If neither current-plant stage yields a positive basis, the workflow uses recently decommissioned sites before falling back to a country bus sequence. Usable decommissioned plants are limited to a configurable look-back window before the target year, and matched by a fuel-priority rule. For example, coal groups can substitute each other, while gas and oil-like groups first use matching sites and then oil/gas and selected coal sites. Remaining units are placed in a round-robin sequence over country buses sorted by load share, maximum current unit size, thermal capacity, and total plant capacity. Marginal costs are computed from fuel prices, emission factors, CO_2 prices, and technology parameters, with documented high-cost fallbacks where parameter mappings are missing.

Other RES and other non-RES

Other non-RES power plants consist of oil or gas-fueled “peaker” power plants. Therefore, raw oil and oil-shale rows are grouped as oil; all other rows are grouped as gas. The first allocation basis is the thermal residual that remains after the explicit thermal TYNDP groups have been placed. If the TYNDP entry contains a unit count, the target is split into equally sized units; otherwise, the country-fuel target is treated as one aggregate unit. All units are first assigned to matching residual thermal basis using the same fuel-priority logic as the decommissioning fallback. Remaining units are then assigned to decommissioned sites in the target-year look-back window. Any units still unplaced after this step are distributed round-robin over positive-load-share buses, starting with the lowest load shares. The bus-level output can aggregate all fuel-specific units into one generic other-non-RES unit per bus, while a technology-level file keeps the fuel group and allocation mode.

The TYNDP distinguishes marine, bio, waste, geothermal, and undefined residual technologies as other RES technologies. Marine capacity is allocated equally to positive-load-share buses that have either free offshore-wind potential or already installed offshore-wind scenario capacity. Bio-based, waste, and geothermal targets first use matching other-RES capacity from the plant list. If this basis is insufficient, the remaining capacity is placed on matching decommissioned sites in the target-year look-back window. The final fallback is a round-robin assignment over positive-load-share buses. Waste uses descending load shares, while bio, geothermal, and undefined technologies use ascending load shares; ties are broken by descending onshore-wind resource class and then by bus identifier. Undefined other-RES targets skip the decommissioning bases and therefore use only the round-robin fallback.

BESS

BESS are allocated by a combined siting basis rather than by a hard split into current and new capacity. Existing battery locations and renewable generation centers are both interpreted as evidence for future battery siting. For every eligible positive-load-share bus,

$$\chi_b^{\text{BESS}} = P_b^{\text{bat,cur}} + \chi_b^{\text{RES}}, \quad (6)$$

where $P_b^{\mathrm{bat,cur}}$ is current battery capacity and χ_b^{RES} is the configured RES scenario-capacity weight, by default PV and onshore wind. If the combined basis is positive, the full TYNDP discharging power target is allocated by

$$P_b^{\mathrm{BESS,dis}} = P_m^{\mathrm{BESS,dis,TYNDP}} \frac{\chi_b^{\mathrm{BESS}}}{\sum_{b' \in m} \chi_{b'}^{\mathrm{BESS}}}. \tag{7}$$

If no combined basis exists, the module falls back to static load shares and then to a uniform split over eligible country buses. Charging power and storage energy are scaled by the final country share $P_b^{\mathrm{BESS,dis}} / \sum_{b' \in m} P_{b'}^{\mathrm{BESS,dis}}$, while effective capacity is computed from discharging power and the reported efficiency.

Load and DSR

NUTS3 geometries are enriched with population from the *JRC* 2021 census grid [27] and estimated gridded GDP [28]. A distance-decay weight with equal splitting for coincident locations is computed with eq. [eq:distance_weight]. Population and GDP shares are aggregated from regions to buses with eq. [eq:load_proxy]. The final static load share is can be controlled with user-specified weighting parameters for the importance of GDP and population with eq. [eq:load_share]. The load disaggregation is based on the methodology applied in the *PyPSA-Eur* model [30].

The regionalisation of DSR capacity is conducted using static load shares from eq. [eq:load_share] as the siting basis. Unit counts, where available, are scaled by the same load share. Price-bands are distributed uniformly. Relative availability time series of DSR capacity are kept per price-band. This keeps both static capacity and hourly available capacity conservative, while preserving the TYNDP price-band structure.

Summary, Outlook and Limitations

This paper provides a reproducible pre-processing workflow for deriving nodal input data for European power system studies based on network expansion scenarios. Its main methodological feature is a consistent disaggregation pipeline: each module defines an allocation basis, a fallback hierarchy and conservation checks. Some modules are based on existing algorithms from the *PyPSA* framework, modifying or extending them. The regionalisation results can be controlled and adjusted by the user through parametrisation in certain modules or by selecting the level of detail during grid reduction. Future work must validate the allocated data with external data sources. The workflow can also be extended to additional TYNDP scenarios or updated grid data sets. Furthermore, a heuristic can be implemented to distribute hydrogen demand and electrolyzer capacities [31]. In addition, the hydrogen transport network can be added as a second-layer grid, to extend for energy system models.

References

[1] Constante-Flores, Gonzalo E., and Antonio J. Conejo. 2024. “Security-Constrained Unit Commitment: A Decomposition Approach Embodying Kron Reduction.” European Journal of Operational Research 319 (2): 427–41. https://doi.org/10.1016/j.ejor.2023.06.013.

[2] Zhang, Huaiyuan, Kai Liao, Jianwei Yang, Shunwei Zheng, and Zhengyou He. 2024. “Frequency-Constrained Expansion Planning for Wind and Photovoltaic Power in Wind-Photovoltaic-Hydro-Thermal Multi-Power System.” Applied Energy 356 (122401). https://doi.org/10.1016/j.apenergy.2023.122401.

[3] Andresen, Gorm B., Anders A. Søndergaard, and Martin Greiner. 2015. “Validation of Danish Wind Time Series from a New Global Renewable Energy Atlas for Energy System Analysis.” Energy 93: 1074–88. https://doi.org/10.1016/j.energy.2015.09.071.

[4] Slednev, Viktor, Manuel Ruppert, Valentin Bertsch, et al. 2017. “Regionalizing Input Data for Generation and Transmission Expansion Planning Models.” In Advances in Energy System Optimization, edited by Valentin Bertsch, Wolf Fichtner, Vincent Heuveline, and Thomas Leibfried. Springer International Publishing. https://doi.org/10.1007/978-3-319-51795-7_13.

[5] Peña-Sánchez, E. U., P. Dunkel, C. Winkler, et al. 2026. “Towards High Resolution, Validated and Open Global Wind Power Assessments.” Nature Communications 17 (1): 539. https://doi.org/10.1038/s41467-026-68337-z.

[6] Verschelde, Tars, and Troy Bober. 2026. Mopo: Pan-European Dataset for Energy System Planning. Zenodo. https://doi.org/10.5281/ZENODO.18593679.

[7] Hörsch, Jonas, Fabian Hofmann, David Schlachtberger, et al. 2026. Data Bundle for PyPSA-Eur: An Open Optimisation Model of the European Transmission System. Zenodo. https://doi.org/10.5281/ZENODO.18164492.

[8] ENTSO-E, and ENTSO-G. 2025. TYNDP 2024 Scenarios Methodology Report - Final Version. European Network of Transmission System Operators for Electricity; European Network of Transmission System Operators for Gas. https://2024.entsos-tyndp-scenarios.eu.

[9] Xiong, Bobby, Davide Fioriti, Fabian Neumann, Iegor Riepin, and Tom Brown. 2026. Prebuilt Electricity Network for PyPSA-Eur Based on OpenStreetMap Data. Zenodo. https://doi.org/10.5281/ZENODO.18619025.

[10] ENTSO-E und ENTSOG. 2020. TYNDP 2020 Scenario Report. European Network of Transmission System Operators for Electricity; European Network of Transmission System Operators for Gas.

[11] Strömer, Stefan. 2021. Graph Clustering Algorithms in European Power Grids. Master thesis. TU Vienna.

[12] Von Luxburg, Ulrike. 2007. “A Tutorial on Spectral Clustering.” Statistics and Computing 17 (4): 395–416. https://doi.org/10.1007/s11222-007-9033-z.

[13] Dorfler, Florian, and Francesco Bullo. 2010. “Synchronization of Power Networks: Network Reduction and Effective Resistance.” IFAC Proceedings Volumes 43 (19): 197–202. https://doi.org/10.3182/20100913-2-FR-4014.00048.

[14] Copernicus Climate Change Service (C3S). 2024. “ERA5 Hourly Data on Single Levels from 1940 to Present.” In ERA5 Hourly Data on Single Levels from 1940 to Present. https://doi.org/10.24381/cds.adbb2d47.

[15] Hofmann, Fabian, Johannes Hampp, Fabian Neumann, Tom Brown, and Jonas Hörsch. 2021. “Atlite: A Lightweight Python Package for Calculating Renewable Power Potentials and Time Series.” Journal of Open Source Software 6 (62): 3294. https://doi.org/10.21105/joss.03294.

[16] European Environment Agency. 2019. CORINE Land Cover 2000 (Raster 100 m), Europe, 6-Yearly - Version 2020_20u1, May 2020. European Environment Agency. https://doi.org/10.2909/DDACBD5E-068F-4E52-A596-D606E8DE7F40.

[17] Pigaiani, Cristian, and Filipe Batista E Silva. 2021. The LUISA Base Map 2018: A Geospatial Data Fusion Approach to Increase the Detail of European Land Use/Land Cover Data. Publications Office of the European Union. https://data.europa.eu/doi/10.2760/503006.

[18] GEBCO Bathymetric Compilation Group 2025. 2025. The GEBCO_2025 Grid - a Continuous Terrain Model for Oceans and Land at 15 Arc-Second Intervals. NERC EDS British Oceanographic Data Centre NOC. https://doi.org/10.5285/37C52E96-24EA-67CE-E063-7086ABC05F29.

[19] European Environment Agency, and European Commission. 2025. Natura 2000 (Vector) - Version End 2024. European Commission. https://doi.org/10.2909/91357F39-7866-41CE-B447-43905C364EC8.

[20] UNEP-WCMC and IUCN. 2025. Protected Planet: The World Database on Protected Areas (WDPA). Dataset. www.protectedplanet.net.

[21] Cerdeiro, D. A., A. Komaromi, Y. Liu, and M. Saeed. 2021. Global Shipping Traffic Density: IMF’s World Seaborne Trade Monitoring System. World Bank Group. https://datacatalog.worldbank.org/search/dataset/0037580/global-shipping-traffic-density.

[22] Flanders Marine Institute. 2023. Maritime Boundaries Geodatabase: Maritime Boundaries and Exclusive Economic Zones (200NM). https://doi.org/https://doi.org/10.14284/632.

[23] Gotzens, Fabian, Heidi Heinrichs, Jonas Hörsch, and Fabian Hofmann. 2025. Performing Energy Modelling Exercises in a Transparent Way - The Issue of Data Quality in Power Plant Databases. Zenodo. https://doi.org/10.5281/ZENODO.14785651.

[24] European Commission, Joint Research Centre. 2026. JRC Hydro-Power Database. European Commission, Joint Research Centre. https://doi.org/10.2905/JRC.JBF7S58.

[25] Quaranta, Emanuele, Robert M. Boes, Julian David Hunt, Sandor Szabò, Jacopo Tattini, and Alberto Pistocchi. 2024. “Considerations on the Existing Capacity and Future Potential for Energy Storage in the European Union’s Hydropower Reservoirs and Pumped-Storage Hydropower.” Journal of Energy Storage 104 (114431). https://doi.org/10.1016/j.est.2024.114431.

[26] European Commission. Joint Research Centre. 2025. Overview of the Energy Storage Deployment in Europe: An Analysis of Current Status and Policy Framework on Energy Storage. No. JRC141463. Publications Office. https://data.europa.eu/doi/10.2760/3846606.

[27] Batista E Silva, Filipe, Sergio Freire, and Cristian Pigaiani. 2024. JRC-ESTAT Census Population Grid 202. European Commission, Joint Research Centre. https://doi.org/10.2905/JRC.VP18KXG.

[28] Kummu, Matti, Maria Kosonen, and Sina Masoumzadeh Sayyar. 2026. Data for: Downscaled Gridded Global Dataset for Gross Domestic Product (GDP) Per Capita at Purchasing Power Parity (PPP) over 1990-2022 (Updated to 1990-2024). Zenodo. https://doi.org/10.5281/ZENODO.18429133.

[29] Shah, Jignesh, Jing Hu, Oreane Y. Edelenbosch, and Michelle T. H. Van Vliet. 2025. “Global Dataset Combining Open-Source Hydropower Plant and Reservoir Data.” Scientific Data 12 (1): 646. https://doi.org/10.1038/s41597-025-04975-0.

[30] Hörsch, Jonas, Fabian Hofmann, David Schlachtberger, and Tom Brown. 2018. “PyPSA-Eur: An Open Optimisation Model of the European Transmission System.” Energy Strategy Reviews 22: 207–15. https://doi.org/10.1016/j.esr.2018.08.012.

[31] Giannelos, Spyros, Ioannis Konstantelos, Danny Pudjianto, and Goran Strbac. 2026. “The Impact of Electrolyser Allocation on Great Britain’s Electricity Transmission System in 2050.” International Journal of Hydrogen Energy 202 (153097). https://doi.org/10.1016/j.ijhydene.2025.153097.

Appendix

A1: Algorithms

Algorithm: Electrical grid reduction

Input: AC buses and AC lines, HVDC links and converters, requested AC cluster budget K, weights α_G, α_L, clustering variant $v \in \{\text{kmeans}, \text{mst}, \text{electrical}, \text{dceffective}\}$

Output: Reduced AC buses and corridors, collapsed HVDC links, cluster map, diagnostics

Simplify AC grid to a 380 kV equivalent and contract tiny AC leaves
Build country–component groups $\mathcal{G}$ from AC buses
For each $g \in \mathcal{G}$:
    Compute country weight $\omega_{m(g)}$
    Compute within-country group share ρ_g
Assign integer budgets k_g by largest remainders with
$1 \le k_g \le n_g$ and $\sum_g k_g = K$
For each $g \in \mathcal{G}$:
    If $v = \text{electrical}$:
        Build weighted adjacency with $w^{\text{el}}_{bb'}$
    Else:
        Build $\boldsymbol{B}$, compute $\boldsymbol{B}^{+}$, derive $d^{\text{eff}}_{bb'}$, and convert
        distances to $w^{\text{eff}}_{bb'}$
    If the group graph is disconnected:
        Split k_g over connected components by largest remainders
    For each component with budget k:
        If the electrical similarity graph is usable:
            Run spectral clustering with k labels
        Else:
            Fall back to coordinate clustering and flag the component
Aggregate AC bus attributes and AC corridors
Collapse point-to-point HVDC topologies to AC–AC links when both terminals are
connected to reduced AC clusters
Prune one-sided DC remnants
Write reduced grid, cluster membership, and country membership

Algorithm: RES regionalisation

Input: Cell capacity factors $cf_{i,t,r,y}$, availability masks, current capacities, TYNDP capacities $P^{\text{TYNDP}}_{m,r}$, TYNDP hourly generation $G^{\text{TYNDP}}_{m,t,r,y}$

Output: Bus capacities, bus capacity factors, scaled hourly generation profiles

For each technology $r \in \{\text{PV}, \text{WON}, \text{WOFF}\}$:
    Build availability $A_{i,r}$
    Compute $P^{\max}_{i,r}$ and apply cell filters
    Compute investment resource class $q_{i,r}$
    Allocate current bus capacity to cells

For each country m:
Compute TYNDP capacity overhead and initialise $\Delta P_{m,r}^{\mathrm{rem}}$
For investment resource classes from best to worst:
While $\Delta P_{m,r}^{\mathrm{rem}} > 0$ and unsaturated cells remain in this class:
Fill equal cell increments capped by headroom
Update cell additions, headroom, and $\Delta P_{m,r}^{\mathrm{rem}}$
Report unallocated capacity if all headroom is saturated
Restrict buses to positive load-share AC buses
Aggregate scenario cell capacities to bus capacities
For each weather year $y = FROM, \ldots, TO$ and hour t:
Compute $CF_{b,t,r,y}$
If a valid profile exists and the raw country total is positive:
Compute raw generation and scale it to $G_{m,t,r,y}^{\mathrm{TYNDP}}$
Else:
If no profile exists but scenario bus capacity is positive:
Allocate generation by scenario bus capacity
Else:
Write zero generation and flag missing profile or zero weather output

Algorithm: Hydro regionalisation

Input: TYNDP hydro capacities, weekly constraints and inflows, plant list, load shares, self-generated hydro profiles, PHS-mode
Output: Nodal hydro capacities, constraints, weekly inflows

Restrict candidate buses to positive load-share AC buses
If PHS resolution is enabled:
Map PHS components to WR or ROR targets according to country
hydro presence
Else:
Keep PHS; if OL and CL PHS coexist, collapse OL PHS into
CL and prepare inflow redirects
Aggregate hydro capacity and constraints to target tuples $j = (m, h, o)$
Build current turbine and storage shares $\sigma_{b,j}^{P}$ and $\sigma_{b,j}^{E}$
For each target tuple $j = (m, h, o)$:
If exact current hydro basis exists:
Allocate turbine and storage targets by current hydro shares
Else:
If any hydro bus exists in m:
Allocate to the strongest hydro bus and flag strongest-bus fallback
Else:
Allocate by load shares and flag load fallback
Scale turbine-related constraints by turbine shares and storage-related
constraints by storage shares
Load TYNDP weekly inflows, interpolate gaps, and impute missing
weather years from weekly means
Aggregate self-generated hourly hydro country profiles to weekly totals
For each OL target (m, h, o, w, y):
Select national inflow source

Distribute the selected inflow to buses
Split bus inflow across hydro technologies by bus-level turbine shares
Set natural inflow to zero for CL targets
Write Nodal hydro capacities, constraints, weekly inflows

Algorithm: Thermal units regionalisation

Input: TYNDP thermal capacity and unit targets, plant list, bus country membership, positive load-share buses, fuel prices, revision-duration
Output: Generic thermal units, thermal groups, bus allocations, marginal costs

Filter plant list to active thermal units and restrict candidate buses to positive-load-share buses
Build decommissioned-site basis for the target-year look-back window and map sites
to reduced clusters
Build country bus priority from load share, maximum unit size, current thermal
capacity, and total plant capacity
Group TYNDP targets by country, fuel code, and technology
    For each target group $n = (m, f, e)$:
    Adjust target unit count if implied unit size falls below the minimum threshold
        For each candidate stage in fuel–technology, fuel-only:
        Select candidates in country m for the current stage
        Compute bus weights $\chi_{b,n}^{(s)}$
            If a positive candidate basis exists:
            Stop the stage search and keep this basis
        If no positive basis exists:
            Try to place all units on decommissioned sites using the thermal fuel-priority
        hierarchy
    Place remaining units by round-robin over the country bus-priority sequence
    Assign integer units by largest remainders
    Create one generic unit row per assigned unit
    Attach revision durations, group identifiers, CHP flags,
inertia parameters, and calculated marginal costs
Write thermal units, thermal groups, and bus allocations

Algorithm: Other non-RES regionalisation

Input: TYNDP other non-RES, plant list, thermal bus allocations, positive load-share buses
Output: Nodal other non-RES capacities, residual thermal basis, cost attributes

Restrict candidate buses to positive load-share AC buses
Classify raw other-non-RES targets into gas and oil fuel groups and retain raw
cost columns
Build thermal basis by bus and fuel group from the current plant list
Subtract assigned explicit thermal capacity and compute residual basis
$R_{b,f}^{\mathrm{nonRES}}$
Build decommissioned-site basis for oil, gas, lignite, hard-coal, and oil-shale
groups in the target-year look-back window
For each model country m and fuel group f:
    Split the fuel-group into equal unit sizes if a unit count is available
    Assign units to residual thermal basis using the fuel-priority hierarchy
    Assign units still remaining to decommissioned sites using the same priority

13. hierarchy
14. **If** units remain after residual and decommissioned bases:
15. Assign them round-robin over positive-load-share buses, starting with the
16. lowest load shares
17. Aggregate fuel-specific units to one generic other-non-RES unit per bus
18. Calculate and attach direct or calculated marginal-cost attributes
19. Write other non-RES units bus allocation

Algorithm: Other RES regionalisation

Input: TYNDP other RES, plant list, RES scenario capacities, RES potentials, WOFF eligibility, positive load-share buses, WON investment resource classes

Output: Nodal other RES capacities, and technology-level allocation

1. Restrict candidate buses to positive load-share AC buses
2. Compute free PV/WON renewable headroom H_b
3. Build marine eligibility from WOFF free capacity or installed
4. WOFF scenario capacity
5. Build other-RES basis by bus for bio, waste, and geothermal based on the plant list
6. Build matching decommissioned-site basis in the target-year look-back window
7. Load WON investment resource classes for round-robin tie breaking
8. **For each** model country m and other-RES technology:
9. Normalise the technology to marine, bio, waste, geothermal, or undefined
10. **If** technology is marine:
11. Allocate the target equally across marine-eligible buses
12. **Else**:
13. **If** technology is bio, waste, or geothermal:
14. Allocate by matching plant list, capped at the available basis sum
15. Allocate remaining capacity to matching decommissioned sites
16. Allocate any remaining capacity by round-robin over load-share buses
17. Use descending load share for waste and ascending load share otherwise; use
18. descending resource class as secondary priority
19. Record allocation mode, candidate-bus count, allocated capacity, and unallocated
20. capacity
21. Aggregate technology-level units to one generic other-RES unit per bus
22. Write other RES units bus allocation

Algorithm: BESS regionalisation

Input: TYNDP BESS discharging power, charging power, energy capacities, current battery sites from plant list, RES bus capacities, positive load-share buses

Output: Nodal BESS discharging power, charging power, energy capacity, effective capacity

1. Restrict candidate buses to positive load-share AC buses
2. Compute current battery shares from existing battery capacity
3. Compute RES shares from configured RES scenario-capacity weights
4. **For each** model country m:
5. Build combined BESS basis χ_b^{BESS}
6. **If** combined basis exists:
7. Allocate full discharging target and flag
8. battery-RES basis

9. **Else**:
10. **If** load shares exist:
11. Allocate full discharging target by load shares and flag load fallback
12. **Else**:
13. Allocate uniformly over eligible country buses and flag uniform fallback
14. Scale charging power, storage energy, and effective capacity by
15. $P_b^{\mathrm{BESS,dis}} / \sum_{b' \in m} P_{b'}^{\mathrm{BESS,dis}}$
16. Report unallocated discharge if no eligible bus exists
17. Write BESS bus capacities

Algorithm: Load regionalisation

Input: Reduced AC load buses, NUTS3 proxy geometries, population and GDP attributes, national hourly load series $L_{m,t,y}^{\mathrm{TYNDP}}$, weights $\eta_{\mathrm{pop}}, \eta_{\mathrm{gdp}}$, distance exponent γ

Output: Nodal load series $L_{b,t,y}$ and static load shares $\lambda_{b,m}$

1. Restrict buses to positive load-share AC buses
2. **For each** country m:
3. Select proxy regions and AC buses in m
4. **If** no proxy region exists:
5. Build one fallback proxy from the mean bus location and flag the country
6. Compute distance weights μ_{pb}
7. Aggregate population and GDP shares to buses
8. Combine both shares to $\lambda_{b,m}$
9. Normalise load shares of buses: $\sum_{b \in m} \lambda_{b,m} = 1$
10. **For each** national hourly load (m, t, y):
11. Allocate $L_{b,t,y} = \lambda_{b,m} L_{m,t,y}^{\mathrm{TYNDP}}$
12. Write nodal loads and static shares

Algorithm: DSR regionalisation

Input: TYNDP DSR capacity and availability time series per price band, static load shares $\lambda_{b,m}$

Output: Bus-level DSR price bands with available capacity

1. Restrict candidate buses to positive load-share AC buses
2. Normalise load shares of buses: $\sum_{b \in m} \lambda_{b,m} = 1$
3. **For each** price band $d = (m, u, z)$:
4. **For each** bus $b \in \mathcal{B}_m$ with $\lambda_{b,m} > 0$:
5. Allocate $P_{b,d}^{\mathrm{DSR}} = \lambda_{b,m} P_d^{\mathrm{DSR}}$
6. Scale unit count by $\lambda_{b,m}$ where unit counts are available
7. Allocate $\overline{P}_{b,d,t}^{\mathrm{DSR}} = \lambda_{b,m} \overline{P}_{d,t}^{\mathrm{DSR}}$
8. Write nodal available DSR capacity per price band

A2: Grid reduction

A2.1: Cluster budget allocation

Buses are first grouped by model country and AC component or sub-network (if no standardisation to a 380 kV grid is applied). Let $\mathcal{G}$ denote the resulting groups, n_g their bus counts, $m(g)$ the model country of group g, C_g^{gen} an existing generation-capacity weight, and $\overline{L}_m$ a mean load weight. Country relevance is measured by a convex combination of generation and load:

$$\omega_m = \frac{\alpha_G C_m^{\mathrm{gen}} / \sum_{\bar{m}} C_{\bar{m}}^{\mathrm{gen}} + \alpha_L \overline{L}_m / \sum_{\bar{m}} \overline{L}_{\bar{m}}}{\sum_{\bar{m}} \left(\alpha_G C_{\bar{m}}^{\mathrm{gen}} / \sum_{\bar{m}} C_{\bar{m}}^{\mathrm{gen}} + \alpha_L \overline{L}_{\bar{m}} / \sum_{\bar{m}} \overline{L}_{\bar{m}} \right)}. \tag{8}$$

Within each country, the group share is based on existing generation capacity, or bus count if no generation weight is available:

$$\rho_g = \begin{cases} C_g^{\mathrm{gen}} / \sum_{\bar{g}:m(\bar{g})=m(g)} C_{\bar{g}}^{\mathrm{gen}}, & \text{if } \sum_{\bar{g}:m(\bar{g})=m(g)} C_{\bar{g}}^{\mathrm{gen}} > 0, \\ n_g / \sum_{\bar{g}:m(\bar{g})=m(g)} n_{\bar{g}}, & \text{otherwise.} \end{cases} \tag{9}$$

The preliminary budget is $\tilde{k}_g = K \omega_{m(g)} \rho_g$. Integer budgets k_g are then obtained by largest remainders with $1 \le k_g \le n_g$ for non-empty groups and $\sum_g k_g = K$. This prevents countries or synchronous components with relevant assets from disappearing in the reduced grid.

A2.2: Spectral variant

The electrical spectral variant uses an undirected weighted graph over AC buses. For each AC edge (b, b'), the edge weight is the sum of circuit-count-adjusted inverse reactances:

$$w_{bb'}^{\mathrm{el}} = \sum_{\ell \in (b,b')} \frac{n_\ell}{|x_\ell|}. \tag{10}$$

Let be the weighted adjacency matrix, the degree matrix, and the normalised graph Laplacian. Spectral clustering embeds the graph using the low-dimensional eigenspace of and partitions the embedding into clusters [12, 13]. Equivalently, the heuristic approximates a normalised cut objective,

$$\min_{\mathcal{S}_1, \ldots, \mathcal{S}_{k_g}} \sum_{a=1}^{k_g} \frac{\mathrm{cut}(\mathcal{S}_a, \bar{\mathcal{S}}_a)}{\mathrm{vol}(\mathcal{S}_a)}. \tag{11}$$

The resulting clusters tend to preserve strong electrical neighbourhoods because lines with lower reactance and more circuits have higher similarity [12].

A2: DC-effective-reactance variant

The second variant converts electrical distance into a graph similarity through DC effective reactance, analogous to effective resistance concepts in network reduction [13]. For each connected AC component, the DC susceptance Laplacian is

$$\boldsymbol{B} = \mathrm{diag}(\boldsymbol{W}\mathbf{1}) - \boldsymbol{W}, \qquad \boldsymbol{W}_{bb'} = \sum_{\ell \in (b,b')} \frac{n_\ell}{|x_\ell|}. \tag{12}$$

With the Moore–Penrose inverse $\boldsymbol{B}^{+}$, the pairwise effective-reactance distance is

$$d_{bb'}^{\mathrm{eff}} = \boldsymbol{B}_{bb}^{+} + \boldsymbol{B}_{b'b'}^{+} - 2\boldsymbol{B}_{bb'}^{+}. \tag{13}$$

Distances are converted into a sparse similarity graph,

$$w_{bb'}^{\mathrm{eff}} = \exp(-d_{bb'}^{\mathrm{eff}}/\tau), \tag{14}$$

where controls the kernel width. To keep the graph sparse and robust, only nearest electrical neighbours are retained before spectral clustering. This variant places buses in the same cluster when an injection at one bus and withdrawal at the other would see a small DC equivalent impedance through the network [13].

A2.3: Line equivalents and HVDC handling

After bus clustering, parallel AC lines between reduced buses b and b' are combined by an electrical line equivalent. For a corridor $\mathcal{E}_{bb'}$,

$$Y_{bb'}^{\mathrm{eq}} = \sum_{\ell\in\mathcal{E}_{bb'}} \frac{1}{r_\ell + jx_\ell}, \qquad z_{bb'}^{\mathrm{eq}} = \frac{1}{Y_{bb'}^{\mathrm{eq}}}, \qquad s_{bb'}^{\mathrm{eq}} = \sum_{\ell\in\mathcal{E}_{bb'}} s_\ell. \tag{15}$$

The equivalent resistance and reactance are the real and imaginary parts of $z_{bb'}^{\mathrm{eq}}$, while the transfer capacity is summed. Length and metadata are aggregated by capacity-weighted averages or unique joins.

HVDC components are handled differently. Standard point-to-point HVDC topologies are collapsed to AC–AC links when both converter ends connect to meaningful AC clusters. Residual one-sided DC components, internal DC helper buses, and DC stubs without two AC terminals are pruned or reported. This keeps the nodal allocation basis aligned with AC dispatch and load-flow buses while preserving cross-border HVDC transfer corridors in the reduced network.

A3: Raster-based RES and hydro profiles

A3.1: Availability masks

The cell availability for a technology r is represented as a product of masks:

$$A_{i,r} = M_i^{\mathrm{shape}} M_{i,r}^{\mathrm{land}} M_i^{\mathrm{protected}} M_i^{\mathrm{ship}} M_i^{\mathrm{depth}} M_i^{\mathrm{shore}}. \tag{16}$$

Only the masks relevant for a technology are active. The maximum installable capacity of cell i is:

$$P_{i,r}^{\max} = A_{i,r}\, A_i^{\mathrm{cell}}\, \kappa_r^{\mathrm{cap}}, \tag{17}$$

where A_i^{cell} is cell area and κ_r^{cap} is the technology-specific capacity density. Cell-level filters remove cells with very small installable capacity or very low maximum capacity factor:

$$i \in \mathcal{I}_r \quad \Leftrightarrow \quad P_{i,r}^{\max} \geq \underline{P_r} \ \wedge \ \max_t cf_{i,t,r,y} \geq \underline{cf_r}. \tag{18}$$

This prevents negligible cells from receiving scenario capacity in later steps.

A3.2: Investment resource classes

Resource classes are used to place future renewable capacity preferentially in better cells without hard-coding a single best-cell ordering. For a selected resource-class year $y^{\star}$, mean capacity factors are split into Q_r equal-width classes between the minimum and maximum eligible cell values:

$$q_{i,r} = \sum_{a=1}^{Q_r-1} \mathbf{1}\left\{\overline{cf}_{i,r,y^{\star}} \geq \underline{cf_r} + a\,\frac{\overline{cf_r}^{\max} - \overline{cf_r}^{\min}}{Q_r}\right\}, \tag{19}$$

with $q_{i,r} = -1$ for unavailable cells. Optional, the same rule can be applied but computes the minimum and maximum within each availability region. The resource class therefore captures relative quality within the relevant planning region.

A3: Load

Distance-decay weight with equal splitting for coincident locations, with p as a proxy region and b as a reduced bus in the same country:

$$\mu_{pb} = \frac{d(p,b)^{-\gamma}}{\sum_{b' \in m} d(p,b')^{-\gamma}}, \tag{20}$$

Population and GDP shares aggregation from regions to buses:

$$\lambda_{b,m}^{\mathrm{pop}} = \sum_{p \in m} \frac{\mathrm{pop}_p}{\sum_{\bar{p} \in m} \mathrm{pop}_{\bar{p}}}\,\mu_{pb}, \qquad \lambda_{b,m}^{\mathrm{gdp}} = \sum_{p \in m} \frac{\mathrm{gdp}_p}{\sum_{\bar{p} \in m} \mathrm{gdp}_{\bar{p}}}\,\mu_{pb}. \tag{21}$$

The final static load share with controllable weighting parameters η:

$$\lambda_{b,m} = \eta_{\mathrm{pop}}\lambda_{b,m}^{\mathrm{pop}} + \eta_{\mathrm{gdp}}\lambda_{b,m}^{\mathrm{gdp}}, \qquad L_{b,t,y} = \lambda_{b,m} L_{m,t,y}^{\mathrm{TYNDP}}. \tag{22}$$