THE ROLE OF SPATIAL SCALE IN ELECTRICITY SYSTEM OPTIMISATION MODELS

Zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften

von der KIT-Fakultät für Informatik

des Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

M.Sc. Martha Maria Frysztacki

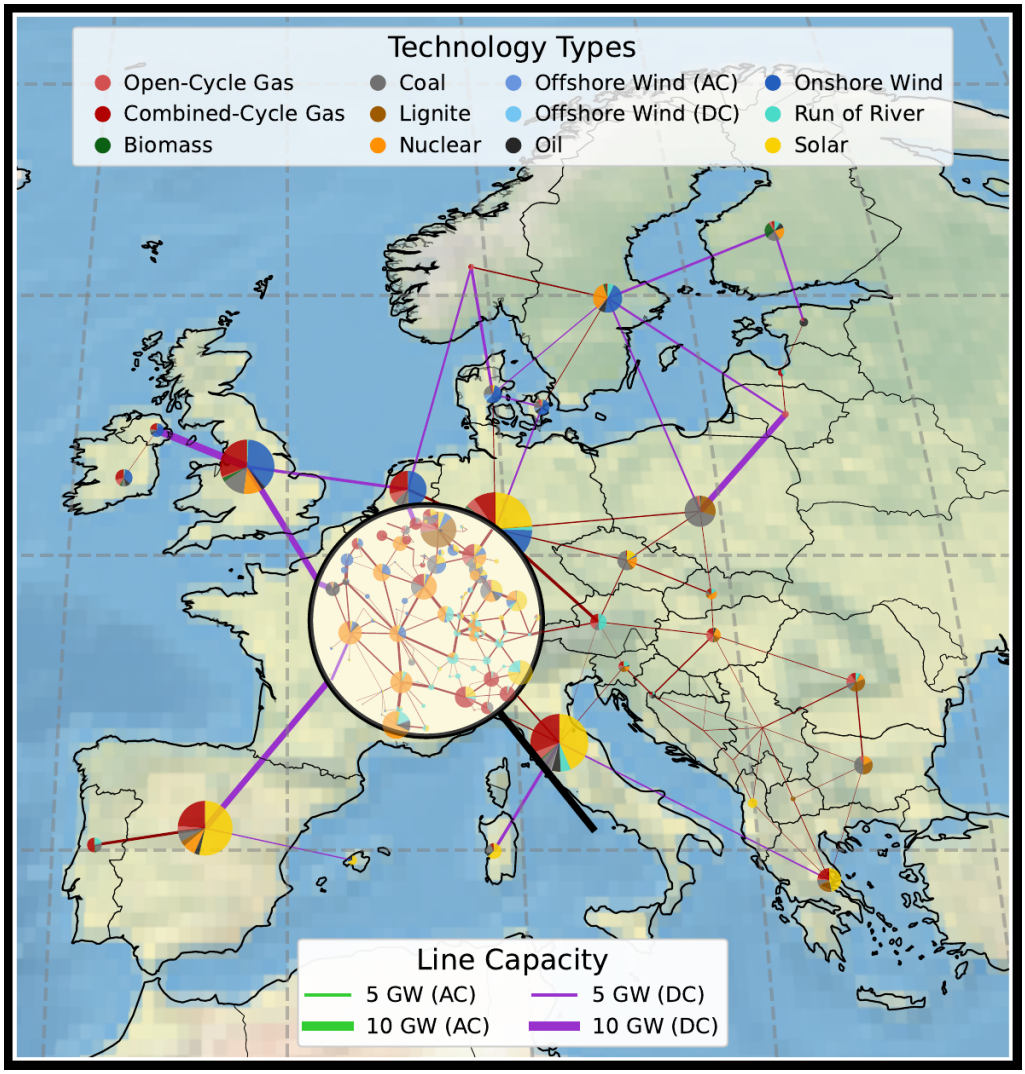
Tag der mündlichen Prüfung: 17.02.2023

1. Referent: Prof. Dr. Veit Hagenmeyer

2. Referent: Prof. Dr. Thomas Brown

THE ROLE OF SPATIAL SCALE IN ELECTRICITY SYSTEM OPTIMISATION MODELS

MARTHA MARIA FRYSZTACKI



December 2, 2022

As this dissertation is the result of a cumulative process of publications, different licences may apply for every chapter.

In real open source, you have the right
to control your own destiny.

— Linus Torvalds

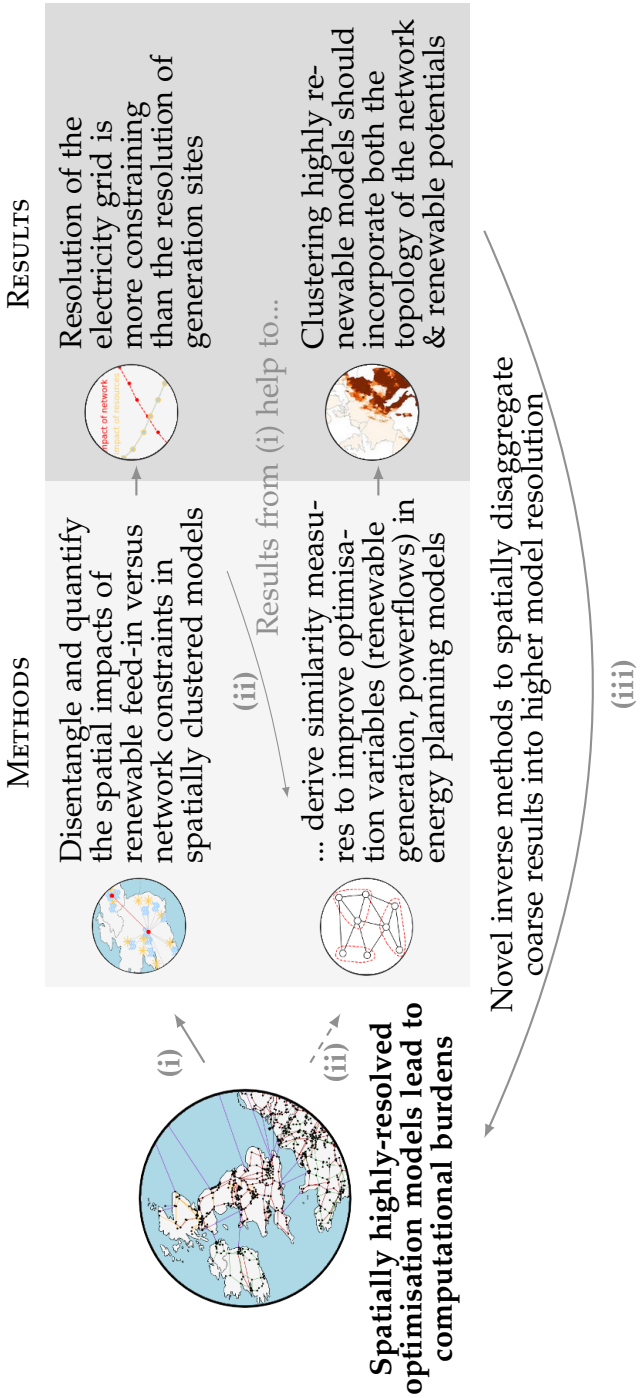
Dedicated to all enthusiastic
open source modellers

ABSTRACT

To investigate possible pathways to reduce greenhouse gas emissions in the electricity sector, researchers build optimisation models that typically minimise the total system costs such that all technical and physical constraints are met. For systems based on renewable energy, whose greatest expansion potentials are found for wind and solar generation, the chief challenge is dealing with their variability. To tackle this challenge, the optimisation models typically include large transmission networks to smooth renewable feed-in in space or storage technologies to smooth the variability in time. However, all aspects of the energy system at all levels of detail cannot currently be contained in a single model because of computational constraints. Instead, one must make simplifications and compromises that affect the optimality of the result from the point of view of the complete system. While reductions on the temporal scale and linearisation approaches of the model formulation have been previously analysed, in this thesis we focus on the quantification of the impact of the spatial scale. This is important because it is scientific practice to simplify models spatially while only little is known on the error made by the aggregation.

The contents of this dissertations spatial scale analysis are threefold and build upon one another: (i) A novel clustering methodology enables us to disentangle and quantify the error that is made by spatially aggregating generation sites where renewable electricity can be sourced versus the error made by aggregating transmission lines and, thus, electricity interactions between spatially distributed substations. By clustering the network on both features in tandem, we can verify the results and learn which of these two effects dominates the optimisation. (ii) Insights from (i) are used to improve existing spatial aggregation methods and to develop novel similarity measures to be applied for clustering electricity system models such that the spatially simplified model can better approximate the original, highly-resolved model with respect to renewable generation sites and the transmission grid. (iii) The prevailing best clustering method is applied on optimisation models with high shares of renewable generation to investigate if the spatially clustered low-resolved solutions are feasible with regard to the full, spatially highly-resolved model. To this end we propose novel inverse methods to spatially disaggregate the coarse optimisation solution in terms of the resulting, aggregated variables across the highly dimensioned model.

GRAPHICAL ABSTRACT



PUBLICATIONS

MAIN PUBLICATIONS

[1] Martha Frysztacki and Tom Brown. "Modeling Curtailment in Germany: How Spatial Resolution Impacts Line Congestion." In: *2020 17th International Conference on the European Energy Market (EEM)*. 2020, pp. 1–7. DOI: https://doi.org/10.1109/EEM49802.2020.9221886

[2] Martha Maria Frysztacki, Jonas Hörsch, Veit Hagenmeyer, and Tom Brown. "The strong effect of network resolution on electricity system models with high shares of wind and solar." In: *Applied Energy* 291 (2021), p. 116726. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2021.116726

[3] Martha Maria Frysztacki, Gereon Recht, and Tom Brown. "A comparison of clustering methods for the spatial reduction of renewable electricity optimisation models of Europe." In: *Energy Informatics* 5.4 (2022). ISSN: 2520-8942. DOI: https://doi.org/10.1186/s42162-022-00187-7

[4] Martha Maria Frysztacki, Veit Hagenmeyer, and Tom Brown. "Inverse methods: How feasible are spatially low-resolved capacity expansion modeling results when dis-aggregated at high resolution?" In: *submitted to Energy* (under review). DOI: https://doi.org/10.48550/arXiv.2209.02364

FURTHER PUBLICATIONS

[5] Kaleb Phipps, Maximilian Beichter, Martha Frysztacki, Ralf Mikut, Veit Hagenmeyer, and Nicole Ludwig. "Net load forecasting using different aggregation levels." In: *Energy Informatics*. Vol. 5(S1). 19. 2022. DOI: https://doi.org/10.1186/s42162-022-00213-8

[6] Amin Shokri Gazafroudi, Elisabeth Zeyen, Martha Frysztacki, Fabian Neumann, and Tom Brown. "Long-Term Benefits for Renewables Integration of Network Boosters for Corrective Grid Security." In: *submitted to International Journal of Electrical Power & Energy Systems* (under review). DOI: https://doi.org/10.48550/arXiv.2112.06667

[7] Maximilian Parzen, Hazem Abdel-Khalek, Ekaterina Fedorova, Matin Mahmood, Martha Maria Frysztacki, Johannes Hampp, Lukas Franken, Leon Schumm, Fabian Neumann, Davide Poli, Aristides Kiprakis, and Davide Fioriti. "PyPSA-Earth. A New Global Open Energy System Optimization Model Demonstrated in Africa." In: *submitted to Applied Energy* (under review). DOI: https://doi.org/10.48550/arXiv.2209.04663

[8] Franz Kaiser, Johannes Kruse, Philipp C. Böttcher, Martha Frysztacki, Tom Brown, and Dirk Witthaut. "Cascading Failures and Critical Infrastructures in Future Renewable Power Systems." In: (in preparation)

CONTENTS

ACRONYMS

ATC available transfer capacity.

Cbc coin-or branch and cut.
CC BY 4.0 Creative Commons Attribution 4.0 International.
CO_2 carbon dioxide.
CPU central processing unit.

DEA Danish Energy Agency.

EEA European Environment Agency.
ENTSO-E European Network of Transmission System Operators for Electricity.
ESM energy system model(ling).
ETIP European Technology and Innovation Platform.
EU European Union.

FLH full load hours.

GLPK GNU Linear Programming Kit.

HAC hierarchical agglomerative clustering.
HiGHS high performance software for linear optimisation.
HPC High Performance Cluster.
HV high voltage.
HVAC high voltage aternating current.
HVDC high voltage direct current.

IPM interior point method.

KCL Kirchhoff's current law.
KKT Karush-Kuhn-Tucker.
KVL Kirchhoff's voltage law.

LMP locational marginal price.
LV low voltage.

MSE mean squared error.

OPSD Open Power System Data.

PDF probability density function.
PTDF power transfer distribution factor.
PV photovoltaic.
PyPSA Python for Power System Analysis.
PyPSA-Eur Python for Power System Analysis in Europe.

RAM random access memory.
RE renewable energy.

TSO transmission system operator.
TYNDP ten year network development plan.

UCPTE Union for the Coordination of Production and Transmission of Electricity.
UCTE Union for the Co-ordination of Transmission of Electricity.

VRE variable renewable energy.

Table 1: Notation.

Abbrev.	Description
	general abbreviations
r	technology type (storage)
$\mathcal{R}$	set of all storage technologies
s	technology type (generators)
$\mathcal{S}$	set of all generating technologies
$\mathcal{S}^{re}$	subset of renewable technologies, $\mathcal{S}^{re} \subseteq \mathcal{S}$
$\mathcal{S}^{con}$	subset of conventional technologies, $\mathcal{S}^{con} \subseteq \mathcal{S}$
t	time discretisation
$\mathcal{T}$	set of all time-steps t
α	year, *optional*. If missing, we indicate a generic setting, or $\alpha = 2018$.
z	synchronous zone or country
$\mathcal{Z}$	set of all countries included in the model
v, w	(highly-resolved) nodes
d_t^z	temporally resolved electricity demand t for country z
$\mathcal{V}$	set of all original nodes in the network graph $\mathcal{G}$
(v, w)	(highly-resolved) line connecting nodes $v, w \in \mathcal{V}$
E	set of all original lines in the network graph $\mathcal{G}$
$\mathcal{G}$	original, fully-resolved network graph, $\mathcal{G} = (\mathcal{V}, E)$
K	number of clusters
K_z	number of clusters for the country-zone z
c, d	clusters, or (low-resolved) nodes
$\mathcal{V}_c$	set of nodes $v \in \mathcal{V}$, aggregated to form cluster c
$\mathcal{V}^K$	set of K clusters in the network graph $\mathcal{G}^K$
c_v	cluster $c \in \mathcal{V}^K$ that node $v \in \mathcal{V}$ is assigned to
(c, d)	(low-resolved) line connecting clusters $c, d \in \mathcal{V}^K$
$E_{(c,d)}$	set of lines $(v, w) \in E$, aggregated to form new line (c, d)
E^K	set of all aggregated lines of the network graph $\mathcal{G}^K$
$\mathcal{G}^K$	reduced network graph $\mathcal{G}^K = (\mathcal{V}^K, E^K)$ to approximate $\mathcal{G}$

Table 1: Notation (cont.).

Abbrev.	Description
	line attributes
$r^{\alpha}_{(v,w)}$	resistance of transmission line (v,w) in year α
$x^{\alpha}_{(v,w)}$	reactance of transmission line (v,w) in year α
$v^{\alpha}_{(v,w)}$	voltage applied at transmission line (v,w) in year α
$l_{(v,w)}$	length of transmission line (v,w)
$c_{(v,w)}$	capital costs of line (v,w)
$u_{(v,w)}$	underwater fraction ($\in [0,1]$) of line (v,w)
$F^{\alpha}_{(v,w)}$	capacity of transmission line (v,w) in year α
$f_{(v,w),t}$	electricity flow of transmission line (v,w) at time t
	nodal attributes
gdp_v	gross domestic product in node v
pop_v	population in node v
z_v	country of node v
x_v, y_v	coordinates of node v
lv_v	boolean, 1 indicates a grid connection to low voltage (LV)
$G^{\alpha}_{v,s}$	cost-optimal capacity of technology s in node v and year α
$G^{min}_{v,s}$	minimal capacity of technology s in node v
$G^{max}_{v,s}$	maximal install-able capacity of technology s in node v
$H^{\alpha}_{v,r}$	cost-optimal capacity of technology r in node v and year α
$H^{min}_{v,r}$	minimal capacity of technology r in node v
$H^{max}_{v,r}$	maximal install-able capacity of technology r in node v
$c_{v,s}$	capital costs of technology s in node v
$c_{v,r}$	capital costs of technology r in node v
$o_{v,s,t}$	variable costs of technology s in node v and time t
η	storage losses or efficiencies for technology
$d_{v,t}$	demand in node v and time t
$\bar{g}_{v,s,t}$	capacity factor for renewable technology s in time t
$g_{v,s,t}$	dispatch in node v of generator s in time t
	graph related attributes
$\mathcal{A}_{v,w}$	(weighted) adjacency matrix of the network graph $\mathcal{G}$
k_v	(weighted) degree of node $v \in \mathcal{V}$
$\mathcal{C}^{\alpha}$	cycle basis for graph $\mathcal{G}$ in year α
$\mathcal{K}^{\alpha}_{v,(v,w)}$	incidence matrix of graph $\mathcal{G}$ in year α
$\mathcal{L}^{\alpha}_{(v,w),c}$	cycle matrix in year α; here c is a cylce

1 INTRODUCTION

This dissertation thesis investigates the role of spatial granularity in European electricity system optimisation models at transmission level with a high share of renewable generation. Although we investigate only the European grid, results and concepts are generalisable to subsets of the European transmission network and other world regions. But let us first take a step back to understand how and when grid planning became relevant, and what the challenges to the grid are that are induced by the energy transition - particularly in regard to the integration renewable sources.

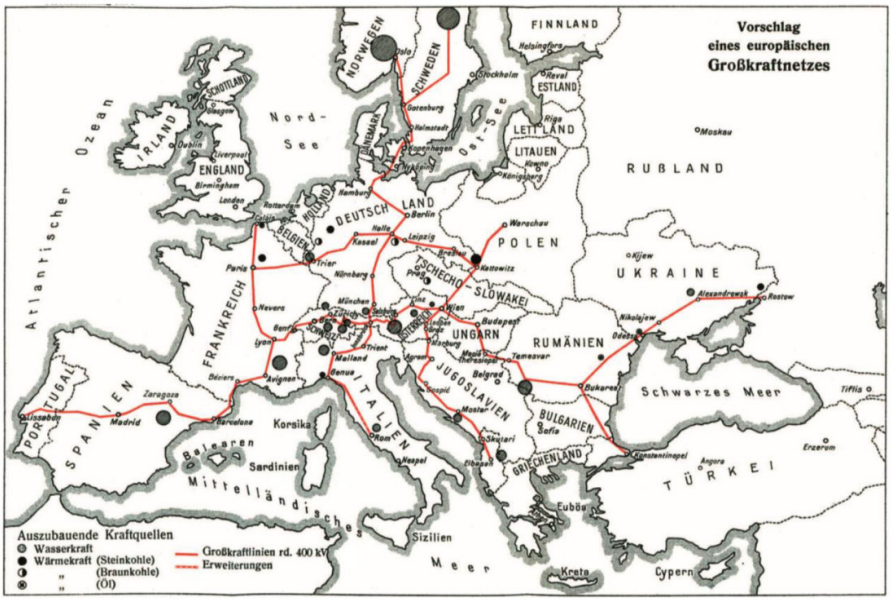


Figure 1: First European Electricity System Network Plan, as suggested by Oskar Oliven in 1930 [9].

In the early days of commercial electricity, transmission at the same LV as used in private households restricted the distance between generating plant and consumers because power lost from heat is proportional to the square of the current traveling in the cable. In contrast, according to Joule's Law, the same amount of power can be transmitted with relatively small losses at high voltage (HV). Therefore, in the early days of electric history, it seemed that the industry would develop into what is known today as a

distributed generation system with a large number of small generators with low capacity located near their loads [10]. But small local electricity generation was regarded as inefficient, thus LV to HV transformation was of interest to solve the problem of transmission over distance.

The invention of electrical transformers legitimised the construction of large power plants outside the direct vicinity of the area of electricity consumption and were connected by first long-distance transmission lines in the 1920s. Simultaneously, ideas to maximise the utilization of cheap energy sources across country borders emerged. A first 'general plan' for the creation of a *European electricity system* was suggested by Oskar Oliven at the World Power Conference in Berlin (1930). He envisioned a transmission grid extending from Norway to Italy and from Poland to Portugal [9]. The main objectives were to bring together the hydro power in Scandinavia and the Alps, together with coal sites in northern France, Belgium, Germany, Poland and the South of Russia to reduce costs for electricity [11]. The outline and routing of lines was mainly planned to connect major generators to major electrical loads (see Figure 1).

After several political and economical disruptions, forging an European electricity network referred back to the idea of 1930. In 1951, eight western European countries[1] forged an initiative[2] to foster economic development by efficient utilisation of primary energy sources through the construction of an interconnected network. Thus, the construction of the 380 kV transmission grid proceeded in the mid 1950s, with the main motivation to prevent the event of supply shortfalls. The initiative also provided an information system between member states, where they informed about water resources, energy storage reserves, the availability of thermal power stations or planned power plants. In this way, there progressively emerged a technical code for the expansion and operation of the bulk power system [12], allowing the electricity grid to expand in tandem with large coal, lignite, nuclear, oil and gas power stations to meet the growing demand for electricity that annually increased by significant rates [13] (see also Figure 2).

Plans to move away from these so-called conventional generators repeatedly came up through the course of history and for different motivations. A first concrete worked-out proposal to decarbonise the electricity grid came up only in 1975, suggested by Bent Sørensen. He issued the warning that extracting resources from lower-grade deposits will continue to sustain industrial ex-

1 Belgium, Federal Republic of Germany, France, Italy, Luxembourg, Netherlands, Austria and Switzerland

2 The Union for the Coordination of Production and Transmission of Electricity (UCPTE)

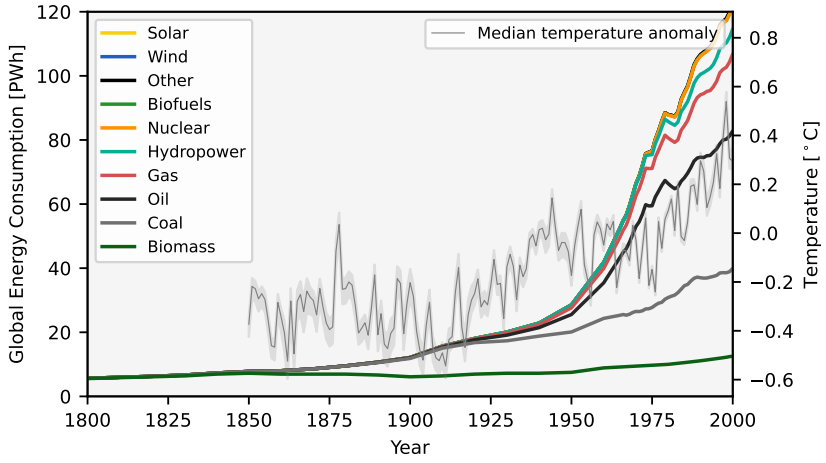


Figure 2: Global energy consumption by source since 1800 (left axis) [13] and global median temperature anomaly from 1961-1990 average since 1850 (right axis) [14].

pansion until either the environmental impacts become unacceptable or ultimate limits, such as climate disruptions, put an end to such growth. To prevent this outcome, he suggested an electricity grid that was fully powered by solar or wind energy on the sample region of Denmark that could be feasible by 2050 [15]. This first fully renewable electricity system plan contains many assumptions that are regarded important in energy system model(ling) (ESM) today, for example land-use restrictions[3], the variability of renewable sources or the orientation and cost assumptions for solar panels and windmills in order to incorporate economical considerations such as the cost of electricity. Nevertheless, climate change mitigation did not become an important goal of global politics until the 1990s.

In 1997, based on the scientific consensus that global warming is occurring and that human-made carbon dioxide (CO_2) emissions are driving it, the first legally binding commitments for developed countries to limit their greenhouse gas emissions were adopted in the Kyoto Protocol. In order to use the same benchmark, most modern studies refer to reductions of CO_2 compared to 1990, following the protocol[4]. The commitment periods of the Kyoto Protocol ended in 2020, and it was effectively replaced by the adoption of the Paris Agreement in 2015. The Agreement's long-term goal is to keep the rise in mean global temperature to well below 2°C above pre-industrial levels, and preferably limit the increase to 1.5°C, recognizing that this would substantially reduce the effects of climate change. Emissions should be reduced

3 He chose a ceiling such that only a small fraction of land area must be used for energy-collecting systems, in total less than 1% of the land.

4 We utilise the same convention in this dissertation.

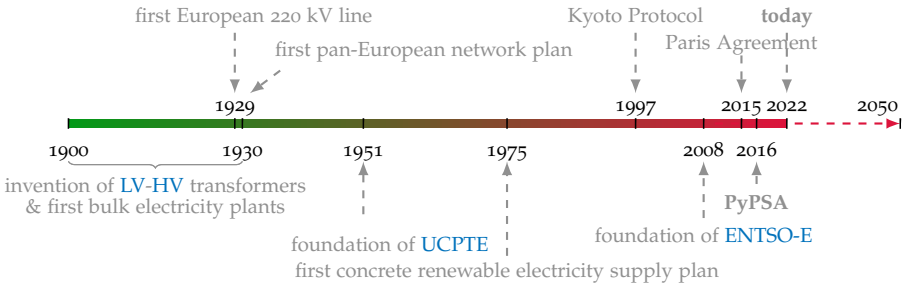


Figure 3: History of the expansion of the electricity grid.

as soon as possible and reach net-zero by the middle of the 21st century [16, 17].

Ever since the political anchoring of the energy transition, environmental topics gained increasing attention. For example, in the course of the third legislative package on the gas and electricity markets in the European Union (EU) in the year 2008 [18] all regional transmission grid associations were united as the European Network of Transmission System Operators for Electricity (ENTSO-E), effectively dissolving its predecessors[5]. ENTSO-E promotes closer cooperation across Europe to support the implementation of the EU energy policy and achieve Europe's energy and climate policy objectives [19], which are changing the intended nature of the power system to connect few large, predictable generators to major electricity demand.

To comply with various governmental regulations to achieve ambitious climate change mitigation policies, the carbon-intensive bulk electricity supply of the past decades must be replaced by either low or CO_2 neutral technologies. Currently the greatest expansion potentials are found for wind and solar generation because of their low emissions and cost-competitiveness with other electricity generation facilities. Moreover, their expansion has the potential to cover worldwide energy demands [20]. But for systems that are based on renewable energy (RE) whose generation yield is highly weather-dependent, the chief challenge is dealing with their variability. For example, the rising share of variable renewable energy (VRE) in recent years, whose layout is naturally designed as a *distributed generation system*, strains the transmission grid in a way that it was not constructed for, resulting in high levels of curtailment [21].

To tackle the challenge of replacing the plannable conventional generators by RE sources in the electricity grid, a growing number of scientific publications were issued that utilise *electricity system optimisation models*, for example [22–26]. Typically, the goal of the

5 such as the Union for the Co-ordination of Transmission of Electricity (UCTE) that has replaced the UCPTE in 1999

optimisation is to simulate how the integration of RE sources into the the transmission grid could be cost-efficiently implemented to achieve CO_2 neutrality while providing a secure and stable electricity supply. But modelling electricity systems with high shares of wind and solar photovoltaic (PV) generation is fundamentally different to modelling conventional power systems with plannable dispatchable generation [27]. While investments in conventional power plants can be dimensioned according to simple heuristics like screening curves [28], reliable investment recommendations for renewable generator, storage and transmission installations must accurately portray renewable potentials and capture the variability of wind and solar PV generation [29].

In order to manage the natural variability of the RE sources, the optimisation models are typically embedded with large transmission networks to smooth renewable feed-in in space or storage technologies to smooth the variability in time. Ideally, the optimisation model is allowed to jointly expand the transmission grid and the generation fleet such that the benefits of exploiting sites with the best renewable resources can be balanced against the network expansion costs [30].

But the assessment of wind and solar resources requires a high temporal and spatial resolution to capture their weather-driven variability. The need to assess investments in generation, transmission and flexibility options over thousands of representative weather and demand situations, as well as over thousands of potential locations, means that balancing model accuracy against computational resources has become a critical challenge. This balancing is the main topic of this dissertation.

1.1 STATE OF THE ART

To reduce the computational burden with regard to capacity planning optimisation models, a basic management option is to reduce the complexity of the model at hand. A general guide to handle model complexity in energy system optimisation models with a high share of RE is provided in [31], where the general approach is to reduce the number of variables and interdependencies. For capacity planning models, there are three main classes where we can reduce model complexity and gain computational advantages.

(i) The numerical description of the physical processes that are simulated by the optimisation model. To circumvent the need for large computational resources on this matter, there exists a broad range of methods to simplify the underlying mathematical description of, for example, non-linear and non-convex power flow calculations or unit commitment. Computational trade-offs with

respect to the model formulation of electricity system optimisation models, such as modelling losses, linearization approaches for electricity transmission [32] or efficiency curves for individual technologies [33], have already been carefully discussed in prior research.

(ii) The temporal resolution of the model that is needed to embed the variability of the weather and RE into the optimisation. For a high share of VRE in the model, the temporal resolution strongly impacts the resulting electricity yield. Effects of the temporal resolution have also been well examined in the electricity system planning literature for the integration of VRE carriers into the grid [20], including the need for at least hourly modelling resolution [27], the consequences of clustering representative conditions [34], and the need to include extreme weather events [35].

(iii) The spatial resolution of the model. At highest spatial resolution for models at transmission level, every substation is represented and interconnected to other substations via transmission lines. To circumvent computational limitations when modelling at such detailed spatial granularity, many studies pursue the approach to represent large political regions such as countries by a single region [36–39], use the full electricity substation level resolution for the transmission grid but only in selected regions [40], or reduce the full model to a smaller equivalent using spatial clustering methods [1, 22, 41–43]. They often make suggestions for the future energy system or the modelling process based on the results obtained by their reduced models. For models with a high share of VRE carriers, integrating renewable resources on a continental scale can smooth large-scale weather variations, particularly from wind [44], and avoid the need for temporal balancing. This smoothing effect has been found in studies of the benefits of grid expansion both in Europe, where the impact on balancing needs [45] and storage requirements [43, 46] has been analysed. Similar results have been found for modelling the United States [47]. When it comes to *clustering*, or, in other words, the representation of multiple regions of the model as a single one, there has been little research on the effects of spatial resolutions on planning results. This is partly due to the fact that collecting high-resolution spatial data is challenging, but mostly because of the fact that optimisation at a high spatial resolution over large areas is computationally demanding.

1.2 RESEARCH QUESTIONS AND CONTRIBUTIONS TO THE SCIENTIFIC COMMUNITY

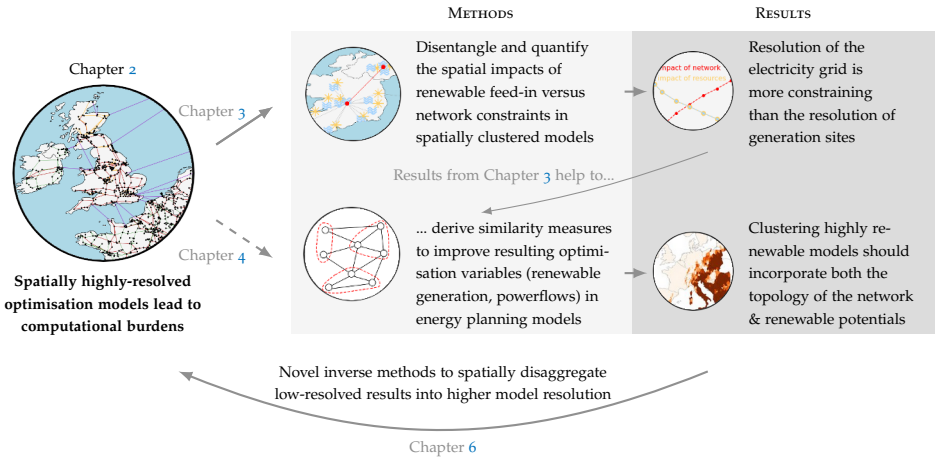
We provide a systematic approach to close the research gap presented in Section 1.1 and analyse the impacts of spatial resolution on the electricity system optimisation results. We focus on continental Europe and answer the following research questions:

- Is there a level of spatial resolution that is computationally tractable but also captures relevant detail? (Chapters 3 and 5)
- What are the main spatial quantities that drive the model results, and how strong is their respective impact? (Chapter 3)
- How do new transmission and new storage affect the spatial resolution discussions? (mainly Chapter 3, and Chapter 4 for storage)
- How can energy system models be aggregated in space and best preserve the properties that have the largest effect on investment optimisation results (e.g. grid bottlenecks, quantification of resource variation)? (Chapter 4)
- How well can a spatially reduced electricity system model at transmission level reproduce relevant quantities of the grid, such as historical transmission congestion, the generation-mix and curtailment? (Chapter 5)
- If a model is optimised at low resolution, how can the results be disaggregated into a highly-resolved model? (Chapter 6)
- Is the disaggregated, spatially highly-resolved model feasible with respect to the optimal low-resolved model solution? (Chapter 6)

1.3 STRUCTURE OF THE THESIS

Individual Chapters of this thesis are stand-alone but build upon one another as follows:

In Chapter 2 we first introduce the European dataset that is used throughout the whole dissertation. Then, we focus on foundations in graph theory and mathematical optimisation. These topics are fundamental for this thesis because they serve as key tools to turn the described dataset into a mathematical model of the (European) electricity grid. From this groundwork we derive the challenge of



computational burden, which in turn leads to the topic of cluster analysis.

After understanding why spatial clustering is required when modelling the European electricity grid, we analyse the main driving effects of the clustering on the planning results of the optimisation in Chapter 3. We examine the implications of the clustering on the European capacity expansion model or, in other words, analyse how the grouping of multiple renewable sites to a representative one affects the optimal solution of the model. We do this in three different scenarios: (i) We analyse the compound impact of all aggregation variables on the optimal results of the network. But as we are interested how certain isolated attributes impact the results of the aggregated model, we additionally examine (ii) how varying the resolution of renewable sites with a fixed number of transmission constraints affects the optimal solution. In contrast, in (iii), we examine how varying the resolution of transmission constraints with a fixed number of renewable sites affects the optimal solution.

Then we incorporate the lessons from the three scenarios of Chapter 3 to improve existing clustering algorithms and design novel aggregation methods or new functions to measure distance between nodes in Chapter 4 to improve the clustering. The improved and novel methods mainly focus on accurately portraying the connectivity of the network, or on accurately representing renewable potentials. We benchmark the results against established methods from previous studies.

In Chapter 5, we analyse if newly embedded data of renewable power stations into the European electricity grid model can reproduce historical numbers of curtailment. This purely operational case-study is carried out only for Germany. This choice of region allows us to overcome computational burdens that arise when modelling a larger geographical area. We can then examine

how modelling results behave when the spatial resolution of the model is varied[6]. We evaluate the spatially fully-resolved transmission grid results, and reduce the resolution of the graph that represents the grid in evenly-sized steps. Numbers of curtailment are statistically evaluated on a spatial and temporal scale. More specifically, we test if the model can capture curtailment in the right locations and during the right times, benchmarking against official numbers published by the German transmission system operators (TSOs). In this analysis we cover the historical years 2013-2018 (in total 6 years).

Finally, we test the feasibility of spatially low-resolved capacity planning results in Chapter 6. Here, feasibility means that spatially low-resolved modelling results are fed back into a model with a higher spatial resolution. The spatially higher dimensioned model is then carefully examined if electricity demand can be met in all places at all times with the additionally invoked constraints from modelling at higher resolution. From previous findings, we pick the clustering method that performed best, and examine how the low-resolved results perform when they are disaggregated at a high resolution. However, as there exist no evaluated disaggregated methods, we first need to research potential inverse functions. Here, we settle for three different options: (i) uniform distribution of the optimal result for the whole region across all nodes that are represented by the region, (ii) re-running a local optimisation problem on a subset of nodes with the full model formulation and additional constraints to incorporate results from the low-resolved optimisation variables and (iii) a novel approach that is based on a new objective function that is derived from prior research in Chapter 5. We compare the performance of these three methods in terms of computational resources and accuracy of modelling results. Model fidelity is measured in terms of feasibility of the spatially highly-resolved operational model. From this comparison and analysis we make implications on how well spatially low-resolved modelling results can approximate the original, highly-resolved system.

Conclusions are drawn in Chapter 7.

1.4 CONTRIBUTIONS TO THE OPEN SOURCE COMMUNITY

As a by-product of the methods of this dissertation, we have developed and contributed to open-source licensed models, datasets, toolboxes and packages. All novel packages and improvements to previously existing ones that are related to the methods of this dis-

6 A full European model would not be computationally feasible with our given computational resources, as we will learn in Chapter 2.

sertation are documented in the following alphabetically ordered Sections. Details on the contribution and further information can be found in Annex A.

1.4.1 *GridKit*

GridKit [48] is a power grid extraction toolkit, initially published in 2016. It processes the interactive map of the European transmission system, that is hosted online by the ENTSO-E to form complete topological connections. In its initial publication, the package included an error where lines that were provided in series were aggregated to a single representative line, such that certain substations were lost in the aggregation. This improvement mainly affects Chapter 5 of this thesis.

1.4.2 *Interannual Demand Calculator*

The *Interannual Electricity Demand Calculator* [49] was developed to account for the symbiosis of variations in weather conditions and electricity demand patterns. While electricity consumption time-series are scarce in the open-source community, records of weather information can be retrieved for a many historical years. To complete the dataset needed for ESM, we have developed the Interannual Demand Calculator to process weather data into an electricity consumption time-series. This novel tool can be utilised to show that all results of Chapters 3 - 6 are stable with respect to given weather conditions.

1.4.3 *Invers-E*

The *invers-e* is a novel package to disaggregate spatially low-resolved optimisation variables that represent regionalized results back at higher or full spatial resolution. At the current stage, it consists of three different disaggregation methods that we propose in Chapter 6, and is restricted to disaggregating only the aggregated low-resolved capacities of generators and storage units. Optimised transmission expansion results are currently not processed.

1.4.4 *NetworkX*

NetworkX is a package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. The main contribution here is the generalization of an existing function to process weighted graphs (i.e. graphs with edge weights), and a customization of the objective, such that all interim solutions

can be returned, not only the optimal one. We utilise these new features in Chapter 4.

1.4.5 *Powerplantmatching*

Powerplantmatching is a toolset for cleaning, standardizing and combining multiple power plant databases. As the transition in the electricity system is taking on, it becomes increasingly important for open-source packages to include renewable assets in their datasets, next to the existing database of conventional power plants such as coal, lignite or nuclear. We have contributed to this target by integrating a dataset of renewable generators provided Open Power System Data (OPSD) [50] to *Powerplantmatching*. These results are used in Chapter 5.

1.4.6 *Python for Power System Analysis (PyPSA)*

Python for Power System Analysis (PyPSA) is an open source modelling framework for simulating and optimising modern power systems. It is designed to scale well with large networks and long time series. The original functionality of the framework to spatially cluster a given network to a smaller approximation included only a weighted variation of the k-means partitioning algorithm. All novel spatial clustering methods developed during the course of this thesis have been contributed to PyPSA. We introduce and evaluate them in Chapter 4.

1.4.7 *Python for Power System Analysis in Europe (PyPSA-Eur)*

Python for Power System Analysis in Europe (PyPSA-Eur) is an open model dataset of the European power system at the transmission network level that covers the full ENTSO-E area. It is built upon PyPSA (section 1.4.6) to run today's or future electricity system simulations. Due to the coupling, the methods on spatially reducing the complexity of the model were initially limited. We have integrated a possibility to import and utilise the novel clustering methods and apply them in Chapter 4 and 6.

2

FOUNDATIONS IN MATHEMATICAL MODELLING FOR ELECTRICITY SYSTEM OPTIMISATIONS

There are several recurring theoretical subjects throughout this dissertation that we introduce in this Chapter before we move to the research content. They include topics from graph theory, mathematical co-optimisation tools and clustering algorithms. We discuss all topics from the point of view that is necessary for renewable investment planning applications at transmission level.

Graph theory is relevant in this thesis as the whole electricity grid can be interpreted as a graph. The main motivation to borrow definitions from graph theory is mainly for a consistent and coherent notation, but also to comprehensibly define physical and mathematical formulations.

The concept of mathematical optimisation is fundamental in this thesis because we analyse how the spatial representation of the underlying dataset of the European electricity grid impacts the optimisation results. We typically examine the optimal solution in terms of accurately representing the electricity system. This can be done by interpreting the optimisation result as a function of the spatial resolution of the model formulation. More specifically, we compare differences in the optimal solution of model runs at varying spatial resolutions against each other, and analyse model feasibility of a spatially low-resolved optimal solution with respect to a spatially highly-resolved modelling result. Understanding the mathematical optimisation also helps to improve the spatial representation of the dataset that is then fed into the model formulation. Thus, it is vital to provide a detailed description of the mathematical concepts prior to the research content.

Clustering algorithms are essential for this thesis, because they are the cornerstone of the spatial representation for the model. The data of the European electricity grid is given exogenous at a certain spatial resolution. However, as we explain later on, it is at too high spatial granularity to be processed by the mathematical optimisation formulation due to computational constraints. In sum, the mathematical problem at high spatial detail is too complex to be processed by modern computers with an appropriate amount of computational power in a reasonable time. To overcome this burden, we can spatially reduce, or *cluster*, the input data before feeding it into the mathematical model. But there are many possibilities how the given data can be clustered to a smaller, ideally

equivalent, data record. Therefore we discuss the general theory of clustering algorithms.

The remains of this chapter are structured as follows: First, we present the underlying datasets that we use for all modelling and research purposes of this thesis in section 2.1. Then we explain how the dataset can be mathematically interpreted to form a comprehensive model in section 2.2. Here, we introduce fundamentals of graph theory that serve as a basis for the notation that we use throughout the whole dissertation. From the formal definitions and mathematical model description, we can finally formulate an mathematical optimisation problem in section 2.3. Approaches to solve the resulting problem are discussed in section 2.4, where we reveal the resulting computational burden of modelling the whole European electricity grid. This big data induced difficulty leads us directly to the topic of clustering methods which we discuss in section 2.5.

2.1 A DATASET OF THE EUROPEAN ELECTRICITY GRID

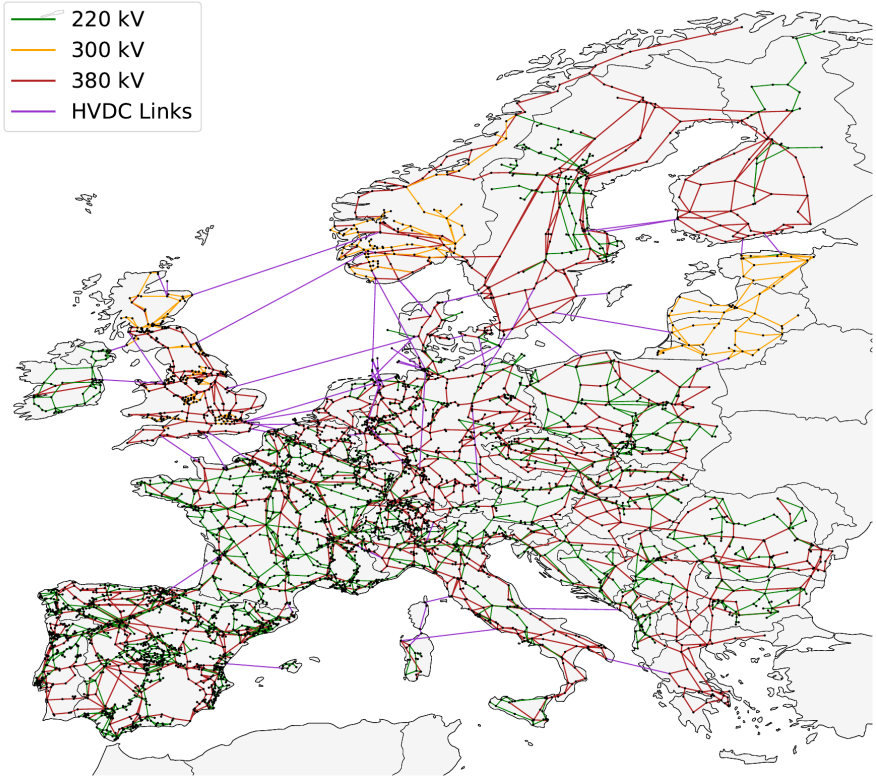


Figure 4: Visualisation of the topology of the European transmission grid at and above 220 kV.

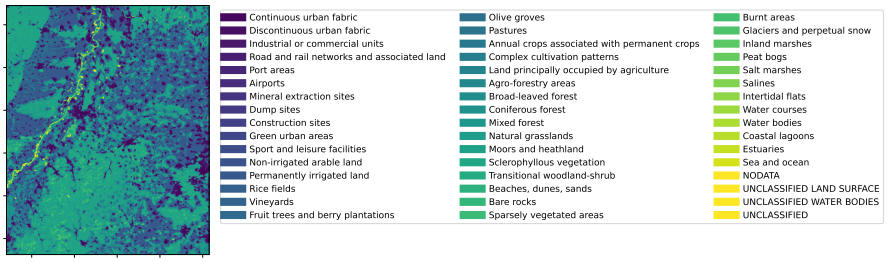


Figure 5: An exemplary cutout of the corine land-cover maps for the area around Karlruhe city and the disctrict of Karlsruhe.

All data used in this thesis for all modelling purposes is presented in this section. It is provided by the open-source model PyPSA-Eur [51], that bundles various sources of open data. The dataset contains more data than presented in the following, but we focus here on the part that is used in this dissertation. It contains:

- Information of the topology of the transmission grid covering the full ENTSO-E area at and above 220kV. This data is originally extracted from the ENTSO-E interactive map using *GridKit* [48]. See Figure 4 for a visualisation. In total, it contains 5399 substations, joints or locations of power plants, 6693 high voltage aternating current (HVAC) transmission lines, 70 high voltage direct current (HVDC) lines and 628 transformers.

- Data on conventional generators and storage units such as lignite or coal-fired power stations, nuclear power plants, open cycle gas turbines, biomass plants, geothermal power generators, pumped-storage hydroelectricity or run-of-river hydroelectricity. The dataset includes geospatial information on the locations of every plant, its capacity, commissioning dates, CO_2 intensity, efficiencies and whether the plants are in active use or not.

- Records of renewable generators, such as onshore or offshore wind assets or solar panels, extracted from the OPSD Project [50]. These records include geospatial information, capacity per asset, commissioning years and the expected lifetime.

- Time-series of the electricity consumption in hourly resolution for every country covered by the electricity grid, originally provided by OPSD [52] for the years $2010 - 2018$. See Figure 7a for an exemplary visualisation of the total electricity demand of Germany, Poland and Italy for two exemplary weeks in 2013.

- Hourly resolved time-series of renewable potentials, given as so-called "*capacity factors*". They indicate how much electricity can be produced in every hour if capacity was installed in a certain. They are derived using the open-source package *atlite* [53], are given in percent, and can be calculated for the years 1951-today.

- Classification of spatial data consisting of land cover maps of Europe that provide the land use status of 2012. This dataset is provided by the European Environment Agency (EEA) under the framework of the Copernicus programme. See Figure 5 for an exemplary cutout covering the area of Karlruhe city and the district of Karlsruhe.

- Geospatial information of protected areas, set up to ensure the survival of Europe's most valuable species and habitats. It includes bird sanctuaries and other landscape protection areas. It is updated on an annual basis. This dataset is originally provided by the EEA.

- Cost assumptions for a set of technologies for electricity generation. They are based on projections for the year 2030 and are derived according to suggestions from the Danish Energy Agency (DEA) [54] (wind), the German Institute for Economic Research [55] (conventional technologies, pumped hydro storage, hydro, run-of-river), Budischak et al. [56] (storage) and the European Technology and Innovation Platform (ETIP) for solar PV [57]. 2030 is chosen for the cost projections since this is the earliest possible time that such a system transformation might be feasible, and because the cost assumptions are conservative compared to projections for a later year. The projections for the cost assumptions are displayed in Table 2.

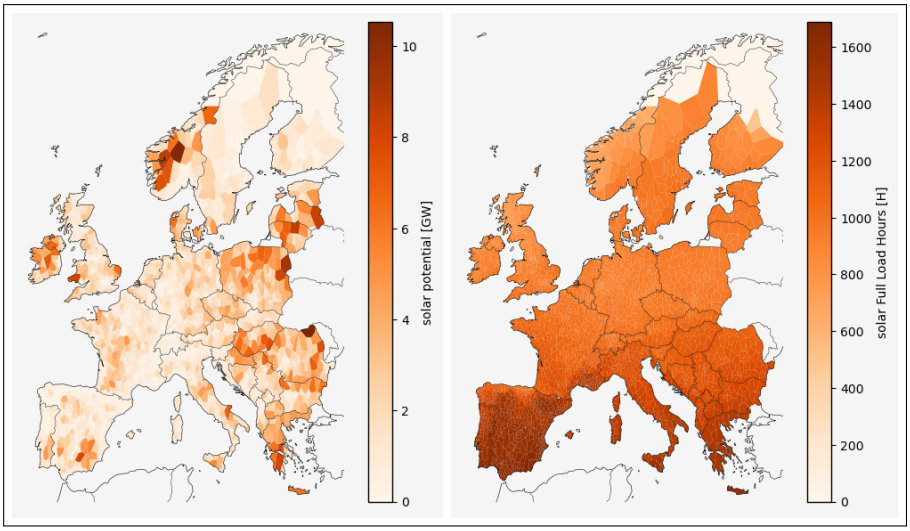
Some of the features, such as renewable potentials, of this dataset are visualised in Figure 6.

2.2 DATASET TO MODEL: SOME TOPICS FROM MATHEMATICS

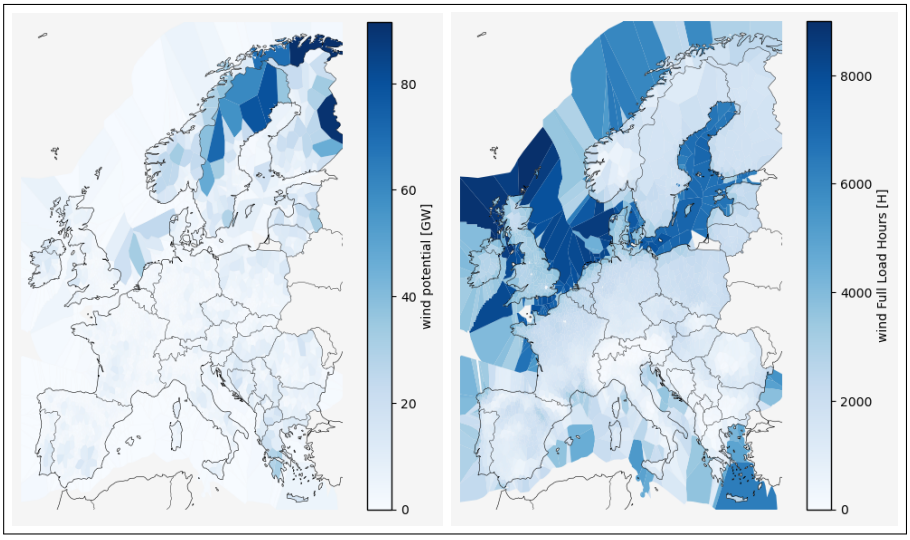
Here we formulate the mathematical model with conventions that are based on graph theory. First, let us provide some basic definitions:

Table 2: Technology investment cost assumptions.

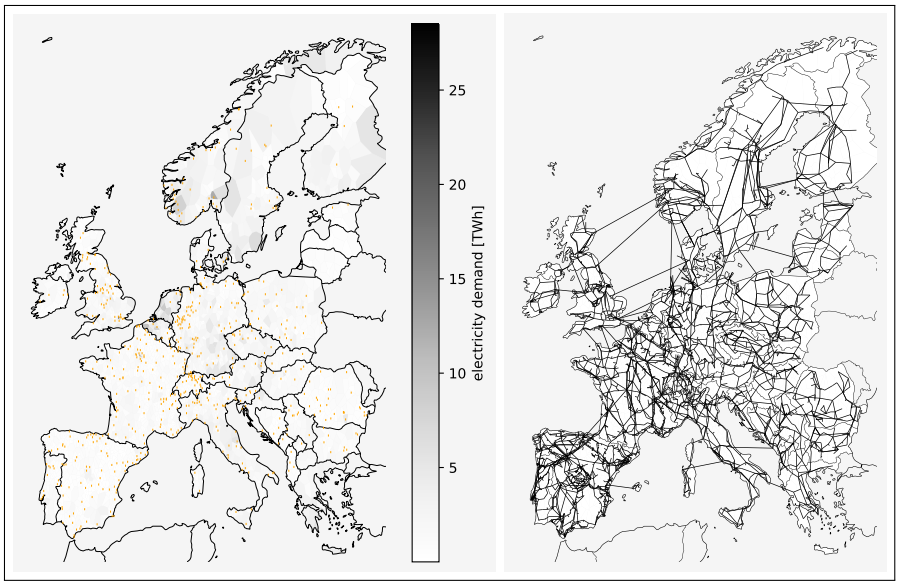
asset	cost	unit
onshore wind	1110	$\frac{€}{kW}$
offshore wind	1640	$\frac{€}{kW}$
(HVAC/HVDC grid connections separately)		
solar PV utility	425	$\frac{€}{kW}$
solar PV rooftop	725	$\frac{€}{kW}$
open cycle gas turbine	400	$\frac{€}{kW}$
run of river	3000	$\frac{€}{kW}$
pumped hydro storage	2000	$\frac{€}{kW}$
hydro storage	2000	$\frac{€}{kW}$
battery storage	192	$\frac{\$}{kW}$
battery power conversion	411	$\frac{\$}{kW_{el}}$
hydrogen storage	11.3	$\frac{\$}{kWh}$
hydrogen power conversion	689	$\frac{€}{kW_{el}}$
HVAC overhead transmission	400	$\frac{€}{MWkm}$
HVAC underground transmission	1342	$\frac{€}{MWkm}$
HVAC subsea transmission	2685	$\frac{€}{MWkm}$
HVDC underground transmission	1000	$\frac{€}{MWkm}$
HVDC subsea transmission	2000	$\frac{€}{MWkm}$



(a) Landuse potentials (left) and full load hours (FLH) (right) for solar PV.



(b) Landuse potentials (left) and FLH (right) for on- and offshore wind.



(c) Electricity demand (left) and network topology (right).

Figure 6: Visualisation of the PyPSA-Eur dataset.

Definition

A **graph** is a pair $\mathcal{G} = (\mathcal{V}, \mathsf{E})$, where $\mathcal{V}$ is a set whose elements are called *vertices* or *nodes* $v \in \mathcal{V}$, and E is a set of paired vertices $(v, w) \in \mathsf{E}$, whole elements are called *edges*, sometimes also *links* or *lines*.

Using this definition, we can formulate the grid data derived from the ENTSO-E interactive map, as visualised in Figure 4, as a graph where every substation, location of power plant or storage unit, converter station or joint is represented as a *node* v. The set of all nodes is denoted as $\mathcal{V}$.

Similarly, all transmission lines of the network graph $\mathcal{G}$ are given as the graph's *edges* (v, w) connecting the nodes $v, w \in \mathcal{V}$. We denote the set of all transmission lines as E. However, as the lines are *directed*, the simple definition of $\mathcal{G}$, as given above, is too vague. We need a specification for $\mathcal{G}$ to provide directions of the edges. This is critical to define the direction of electricity power flows.

Definition

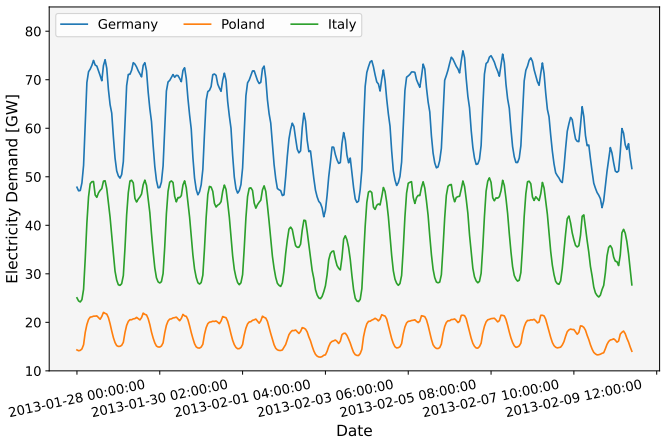
A **directed graph** or **digraph** is a graph in which the edges have orientations.

In this definition, the order of nodes in the notation of an edge is meaningful, i.e. $(v, w) \neq (w, v)$.

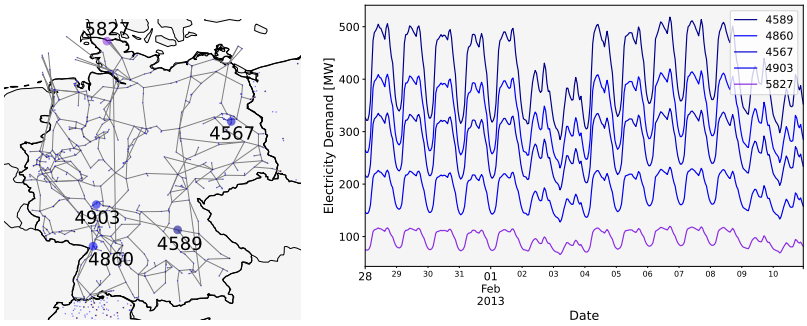
To supplement the formal mathematical description of the model, we assign characteristic attributes to every node $v \in \mathcal{V}$ in the graph, such as its geographical locations given as latitude $x_v \in [-180, 180]$ and longitude $y_v \in [-180, 180]$, the country z_v it is assigned to or a switch $lv_v \in \{0, 1\}$ to denote whether it is a substation, i.e. connected to the lower-voltage distribution grid. Every node with $lv_v = 1$ is assigned a temporally resolved electricity demand $d_{v,t} \in \mathbb{R}_0^+$, given in in MWh. However, as the electricity demand is exogenous only given per country z (see section 2.2), we first need to disaggregate d_t^z spatially to every substation. This is a challenge because as of today there exists no open-source spatially resolved data. We pursue a heuristic approach and disaggregate the electricity consumption proportional to local population (pop_v) and gross domestic product (gdp_v). Mathematically, the disaggregation can be written as:

$$d_{v,t} = d_t^{z_v} \cdot \left(0.6 \cdot \|gdp_v\|_{max} + 0.4 \cdot \|pop_v\|_{max}\right) \tag{1}$$

[58] has shown on a sample region in Italy, that this heuristic provides a good correlation with actual consumption. A visualisation



(a) Electricity demand retrieved from the OPSD [52] for three exemplary countries (Germany, Poland and Italy) and two exemplary weeks.



(b) Disaggregated electricity demand on five exemplary substations in Germany for the same two exemplary weeks.

Figure 7: Exemplary visualisation of the disaggregation of electricity demand that is given exogenous per country.

of the disaggregation of the electicity demand time-series is provided in Figure 7: Figure 7a shows exemplary electricity consumption profiles for Germany, Poland and Italy and Figure 7b shows the disaggregation of cumulative time-series for five sample nodes in Germany.

Further, every node is assigned a renewable installation potential $G_{v,s}^{max} \in [0, \infty)$ that is given in *MW* and is based on land cover maps, excluding for example nature reserves, cities or streets using the geospatial land availability toolkits "GLAES" [59] and "atlite" [53]. They calculate the share of land per region that can be used for the deployment of renewables either in % or in km^2. For onshore wind installations, we then assume a minimum distance of 1000m from streets or buildings, and for offshore installations a maximal water depth of 50m. The resulting renewable installation potentials are derived based on the resulting eligible area within the associated so-called *Voronoi region* of every node $v \in \mathcal{V}$ and the

different technology assumptions. We assume a capacity density of $1.7\frac{MW}{km^2}$ for solar PV installations, $3\frac{MW}{km^2}$ for onshore wind turbines and $2\frac{MW}{km^2}$ for offshore wind assets. Initial land use assumptions are taken from [60]. We then additionally assume that only 1% of the the already restricted area is available for solar PV panels, 20% for installations of offshore wind generators and 20% for onshore wind generators. We perform this additional fractioning due to competing land use and likely low public acceptance. An exemplary visualisation of the Voronoi regions to the associated nodes in Karlruhe city and the district of Karlsruhe is presented in Figure 8a. The resulting eligible area for onshore wind and solar PV deployment within these Voronoi regions is shown in Figure 8b.

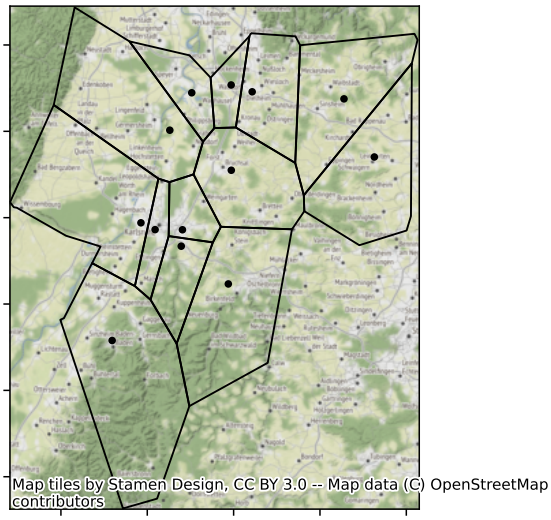
DEFINITION

The **Voronoi region** of a node $v \in \mathcal{V}$ is defined as the area that is closest to the node in the sense of a pre-defined metric. Here, we use the euclidean distance metric $d(v, w) := \|v - w\|_2$. Thus, the Voronoi region of $v \in \mathcal{V}$ is formally given as the set:

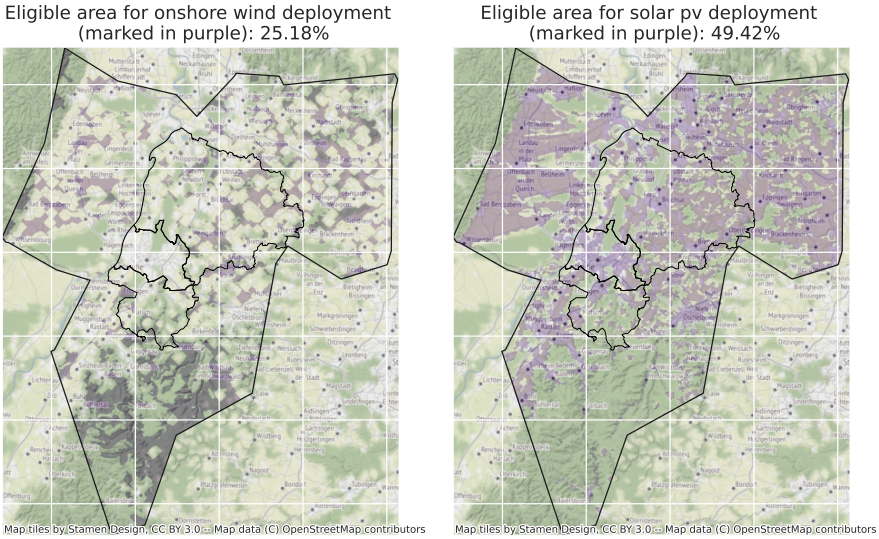
$$\{x \in \mathbb{R}^2 : d((x_v, y_v)^T, x) \leqslant d((x_w, y_w)^T, x) \quad \forall w \in \mathcal{V}\}.$$

Finally, every node v is assigned a per unit RE generation time-series $\bar{g}_{v,s,t} \in [0, 1]$ for its renewable carriers $s \in \{\text{solar PV, onshore wind, offshore wind}\}$. The factors $\bar{g}_{v,s,t}$ are derived from historical weather data, taking into account the solar irradiation and the wind speeds as well as technical properties of the assets, such as the orientation of solar panels (here: south orientation, tilt angle 35°) or the hub height of wind turbines (here: 80m). The capacity factors for wind are obtained from the ERA5 dataset with a spatial resolution of $0.281° \times 0.281°$ [61], and for solar from the SARAH-2 dataset [62], with a spatial resolution of $0.05° \times 0.05°$. Finally, $\bar{g}_{v,s,t}$ is derived from the Voronoi region of a node (excluding the one that is reserved for woodlands, rivers, streets etc.) by heuristically placing wind turbines and solar panels. The capacity factors for each location are taken from the characteristic power curves of the assets and are then averaged for the corresponding Voronoi region.

Similar as for the network nodes $v \in \mathcal{V}$, every transmission line $(v, w) \in E$ - or edge of the network graph $\mathcal{G}$ - has associated attributes that are relevant for the modelling. They include the lines individual resistance $r_{(v,w)} \in \mathbb{R}_0^+$ and reactance $x_{(v,w)} \in \mathbb{R}_0^+$, both given in Ω, as well as its transmission capacity $F_{(v,w)} \in \mathbb{R}_0^+$ given



(a) Exemplary Voronoi regions of nodes that lie in Karlsruhe city or the district of Karlsruhe. Nodes are depicted with black dots, and the corresponding Voronoi region of every node is depicted with a black boundary enclosing the associated node.



(b) Visualisation of land-use eligibility on the example of the Voronoi cells (outer black boundary) that contain Karlruhe city and the district of Karlruhe (inner black boundary). Area that is eligible for onshore wind (left) and solar PV (right) deployment is shaded in purple. Eligible land is derived from information provided by the EEA, which is displayed in Figure 5 for the same region.

Figure 8: Visualisation of how land-use restrictions are derived based on Voronoi regions and land cover maps.

in MW. Finally, the length $l_{(v,w)}$ of every line $(v,w) \in E$ is given in km.

In the context of transmission lines and power flows, we also need the definition of an incidence matrix.

Definition

The **incidence matrix** of a directed graph $\mathcal{G} = (\mathcal{V}, E)$ is a matrix $\mathcal{K} \in \{-1, 0, 1\}^{|\mathcal{V}| \times |E|}$. Its entries are defined as

$$\mathcal{K}_{u,(v,w)} = \begin{cases} -1 & \text{if } w = u \\ 1 & \text{if } v = u \\ 0 & \text{if } u \neq v,\ u \neq w \end{cases} \qquad \forall u \in \mathcal{V},\ (v,w) \in E \tag{2}$$

When deriving equations for electric circuits, cycle bases are a relevant term. They are useful for example to derive Kirchhoff's Laws. We need a couple of definitions to finally derive a cycle basis $\mathcal{C}$ and its associated cycle matrix $\mathcal{L}$. First, we need to understand what the term *cycle* means.

Definition

A **walk** of length $n \in \mathbb{N}\backslash\{0\}$ of a graph $G = (\mathcal{V}, E)$ is a finite sequence of n edges $\{(v_0, v_1), (v_1, v_2), \ldots, (v_{n-1}, v_n)\}$ with $(v_{i-1}, v_i) \in E$, $i = 1, \ldots, n$ and $v_{i=1,\ldots,n} \in \mathcal{V}$.
A **directed trail** is a directed walk where all edges are distinct.
A **directed circuit** is a non-empty directed trail with $v_0 = v_n$.
A **directed cycle** is a directed circuit with $v_i \neq v_j\ \forall i, j \in \{1, \ldots, n-1\}$, $i \neq j$.

Definition

A **cycle basis of a network** is a minimal collection of cycles c, such that any cycle in the network can be written as a sum of cycles in the basis. We denote the cycle basis with the letter $\mathcal{C}$.
Here, summation of cycles is defined as an *exclusive or* of the edges.

We note that the cycle basis of the Graph $\mathcal{G}$ that portrays the European transmission grid (see Figure 4) is a set $\mathcal{C}$ of length 1722. In other words, the basis consists of 1722 cycles, where every cycle in the basis is an own set of nodes. 50% of the cycles in the basis consist of 7 nodes or less (median), while the average length of a cycle in the basis consists of 50 nodes (mean). The shortest cycle contains 3 nodes, the longest one 561.

Definition

The cycle basis $\mathcal{C}$ of a graph $\mathcal{G} = (\mathcal{V}, E)$ can be used to define the graph's **cycle matrix** $\mathcal{L} \in \{0,1\}^{|E| \times |\mathcal{C}|}$. For every cycle $c \in \mathcal{C}$ its entries are defined as:

$$\mathcal{L}_{(v,w),c} = \begin{cases} 1 & \text{if } (v,w) \in c \\ 0 & \text{else} \end{cases}$$

Notation introduced in this section is summarised in Table 1.

2.3 MATHEMATICAL OPTIMISATION METHODS FOR ELECTRICITY SYSTEM INVESTMENT PLANNING

In this Section, we focus on representing the dataset described in Section 2.1 using the methods and notation introduced in Section 2.2 to formulate a mathematical co-optimisation model suited for renewable investment planning applications at transmission level [63].

The aim of the optimisation formulation is to minimise the yearly system costs. This can be achieved by formulating an adequate objective function that describes their composition. We propose to formulate it as a sum, that consists of various terms where each one accounts for a different part of the total system:

(i) The sum of capacity $G_{v,s}$ (given in MW) per node $v \in \mathcal{V}$ and technology $s \in \mathcal{S}$ multiplied by their annualised specific nodal and technological capital investment costs $c_{v,s}$ (given in $\frac{€}{MW}$), over all nodes $v \in \mathcal{V}$ and technologies $s \in \mathcal{S}$. This sum represents the total investment costs in new technologies.

(ii) To account for variable costs, such as costs for fuels and maintenance costs ($o_{v,s,t}$, given in $\frac{€}{MWh}$), the sum of (i) must also span over the time-dependent operation of every unit $g_{v,s,t}$ (given in MWh).

(iii) A similar sum must be taken into account for investments in new storage capacity $H_{v,r}$ (given in MW) for every node $v \in \mathcal{V}$ and storage technology $r \in \mathcal{R}$. To deduce the resulting system costs,

the capacity must again be multiplied with the technology and node specific costs $c_{v,r}$ (given in $\frac{€}{MW}$) for every unit of capacity.

(iv) Finally, we must consider investments into new transmission line projects. Their total annualised costs can be calculated as the sum of all transmission lines $(v,w) \in E$ over the capacity $F_{(v,w)}$ (given in MW) of every line, multiplied by the line-specific costs $c_{(v,w)}$ (given in $\frac{€}{MW}$).

Thus, the whole objective function can be formally written as

$$\min_{\substack{G_{v,s},\, H_{v,r} \\ g_{v,s,t},\, h^{\pm}_{v,r,t} \\ f_{(v,w),t}}} \Bigg[\sum_{v\in\mathcal{V},\, s\in\mathcal{S}} \Big(c_{v,s} G_{v,s} + \sum_{t\in\mathcal{T}} w_t o_{v,s} g_{v,s,t} \Big) + \sum_{v\in\mathcal{V},\, r\in\mathcal{R}} c_{v,r} H_{v,r} + \sum_{(v,w)\in E} c_{(v,w)} F_{(v,w)} \Bigg] \tag{3}$$

where w_t is without units and accounts for the time-weighting.

The objective function (3) is embedded with various constraints to account for socio-economic, electrical and physical aspects as well as weather-related uncertainty of the system we seek to model.

The total installable capacities $G_{v,s}$ for generators, $H_{v,r}$ for storage units and $F_{(v,w)}$ for transmission lines are constrained by upper and lower bounds, mathematically given as

$$G^{min}_{v,s} \leqslant G_{v,s} \leqslant G^{max}_{v,s} \quad \forall v \in \mathcal{V},\, s \in \mathcal{S}^{re} \tag{4}$$

$$H^{min}_{v,r} \leqslant H_{v,r} \leqslant H^{max}_{v,r} \quad \forall v \in \mathcal{V},\, r \in \mathcal{R} \tag{5}$$

$$F^{min}_{(v,w)} \leqslant F_{(v,w)} \quad \forall (v,w) \in E \tag{6}$$

The lower bounds $G^{min}_{v,s}$, $H^{min}_{v,r}$ and $F^{min}_{(v,w)}$ are set to the existing capacities of the year 2018, which was the most up-to-date open-source database at the time when this dissertation was written. The derivation of the upper bound $G^{max}_{v,s}$ is explained in section 2.2 and visualised in Figure 8. An upper bound for the transmission lines is not provided on a basis of individual lines. Instead it is given as a cumulative cap $\bar{F}$ given in percent of the volume of the existing grid, where the optimisation can freely distribute the extra capacity across the existing grid:

$$\sum_{(v,w)\in E} l_{(v,w)} F_{(v,w)} \leqslant \left(1 + \bar{F}^{max}\right) \sum_{(v,w)\in E} l_{(v,w)} \cdot \bar{F}^{min}_{(v,w)} \tag{7}$$

By this setting, no new transmission lines are considered and only the existing ones can be strengthened.

The nodal dispatch $g_{v,s,t}$ of generators has to be non-negative, and is constrained by the installed capacity $G_{v,s}$ for every conventional technology $s \in \mathcal{S}^{\text{con.}}$ and in each snapshot $t \in \mathcal{T}$:

$$0 \leqslant g_{v,s,t} \leqslant G_{v,s} \quad \forall v \in \mathcal{V},\, s \in \mathcal{S}^{\text{con.}},\, t \in \mathcal{T} \tag{8}$$

To account for weather-related uncertainty, we constrain the upper limit of renewable generators by an additional weather-related availability factor $\bar{g}_{v,s,t} \in [0,1]$:

$$0 \leqslant g_{v,s,t} \leqslant \bar{g}_{v,s,t} G_{v,s} \quad \forall v \in \mathcal{V},\, s \in \mathcal{S}^{\text{re}},\, t \in \mathcal{T} \tag{9}$$

The factors $\bar{g}_{v,s,t} \in [0,1]$ are derived from historical weather data (see section 2.2) and, thus, account for the availability of renewable resources. For example, for solar PV generation, $\bar{g}_{v,\text{solar},t} = 0\ \forall t \in \mathcal{N}_v \subset \mathcal{T}$, where $\mathcal{N}_v$ is a subset of $\mathcal{T}$ that contains only snapshots t where there is no solar radiation in the Voronoi region of v. Similarly, $\bar{g}_{v,\text{wind},t} = 0\ \forall t \in \mathcal{P}_v$, where $\mathcal{P}_v$ is a subset of $\mathcal{T}$ that contains only snapshots t where there is either too little or too strong wind for a wind turbine to generate electricity in the Voronoi region of $v \in \mathcal{V}$.

Similar constraints are defined for electricity storage. The charging $h^{+}_{v,r,t}$ and discharging $h^{-}_{v,r,t}$ process is constrained by the thermal rating of the storage unit $H_{v,r}$.

$$0 \leqslant h^{+}_{v,r,t},\, h^{-}_{v,r,t} \leqslant H_{v,r} \quad \forall v \in \mathcal{V},\, r \in \mathcal{R},\, t \in \mathcal{T} \tag{10}$$

To prevent simultaneous charging and discharging, a small artificial cost can be added as an additional constraint for the discharging $h^{+}_{v,s,t}$.

In contrast to generating units, we must embed additional constraints to account for consistency of the storage levels $e_{v,r,t}$. The storage level at snapshot $t \in \mathcal{T}\backslash\{0\}$ must equal to the amount that is charged $h^{+}_{v,r,t}$ plus the amount that naturally flows into the unit $h^{\text{inflow}}_{v,s,t}$ minus the amount that is discharged $h^{-}_{v,r,t}$ and the amount that is naturally spilled $h^{\text{spill}}_{v,s,t}$ plus the storage level at the previous snapshot $e_{v,r,t-1}$. All these individual processes are weighted with their respective efficiencies, such as the standing loss of the unit $\eta_{v,r,0}$ or the units charging $\eta_{v,r,+}$ and discharging $\eta_{v,r,-}$ efficiency.

$$\begin{aligned} e_{v,r,t} = {} & w_t \cdot \left(\eta_{v,r,+} \cdot h^{+}_{v,r,t} - \eta^{-1}_{v,r,-} \cdot h^{-}_{v,r,t} \right) \\ & + w_t \cdot \left(h^{\text{inflow}}_{v,r,t} - h^{\text{spill}}_{v,r,t} \right) + \eta^{w_t}_{v,r,0} \cdot e_{v,r,t-1} \\ & \forall v \in \mathcal{V},\, r \in \mathcal{R},\, t \in \mathcal{T}\backslash\{0\} \end{aligned} \tag{11}$$

The state of charge at the initial snapshot of the simulation $t = 0$ must be given exogenous. We additionally assume that all storage units must have the same final condition at the end of the simulation, i.e.

$$e_{v,r,0} = e_{v,r,|\mathcal{T}|} \quad \forall v \in \mathcal{V},\, r \in \mathcal{R} \tag{12}$$

The state of charge is bound by the capacity of the storage unit which is equal to the units technology-specific maximal duration to discharge power T_r multiplied with its power rating $H_{v,r}$.

$$0 \leqslant e_{v,r,t} \leqslant T_r H_{v,r,t} \quad \forall v \in \mathcal{V},\, r \in \mathcal{R},\, t \in \mathcal{T} \tag{13}$$

Now that all nodal constraints are defined, let us have a look at the power flow constraints of the network. Similar as the generation and storage charging and discharging, the power flows are constrained by the transmission line capacities. However, we allow the power flows to exploit only a certain margin of the full capacity in order to account for the so-called $(N-1)$-*security*. We apply a concrete 70% margin on all lines, following suggestions from previous literature, see [64, 65]. This additional margin is implemented as a measure to ensure that the grid has enough available capacity for reactive power flows in case a single circuit fails.

$$|f_{(v,w),t}| \leqslant 0.7 \cdot F_{(v,w)} \quad \forall (v,w) \in E,\, t \in \mathcal{T} \tag{14}$$

The sign of the power flow indicates its direction. If positive, electricity flows from node $v \in \mathcal{V}$ to node $w \in \mathcal{V}$. If negative, electricity flows in the opposite direction.

To account for a physical plausibility of the grid, we embed Kirchhoff's current law (KCL) and Kirchhoff's voltage law (KVL) laws into the optimisation problem. KCL essentially formulates a nodal power balance stating that the sum of all electricity flows into and out of a node must equal the amount that is locally consumed and charged minus the amounts that are locally discharged or generated at the node.

$$\sum_{w \in \mathcal{V}:\,(v,w) \in E} \mathcal{K}_{v,(v,w)} f_{(v,w),t} = \tag{15}$$
$$d_{v,t} + \sum_{r \in \mathcal{R}} \left(h^{+}_{v,r,t} - h^{-}_{v,r,t} \right) - \sum_{s \in \mathcal{S}} g_{v,s,t} \quad \forall v \in \mathcal{V},\, t \in \mathcal{T}$$

KVL on the other hand states that the directed sum of potential differences around any closed cycle is zero. The implementation of this rule is more challenging. A linearised formulation of it can

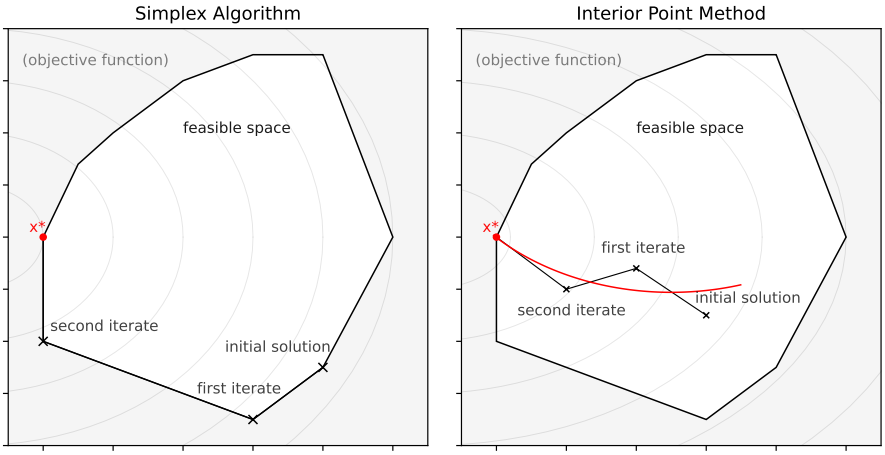


Figure 9: Simplified illustration of the simplex algorithm (left) and interior point method (IPM) (right).

be implemented using a cycle basis $\mathcal{C}$ of the network graph and its associated cycle matrix $\mathcal{L}$ as

$$\sum_{(v,w)\in c} \mathcal{L}_{(v,w),c} x_{(v,w)} f_{(v,w),t} = 0 \quad \forall t \in \mathcal{T},\, c \in \mathcal{C}$$

Finally, we include a constraint to cap the CO_2 emissions of the system.

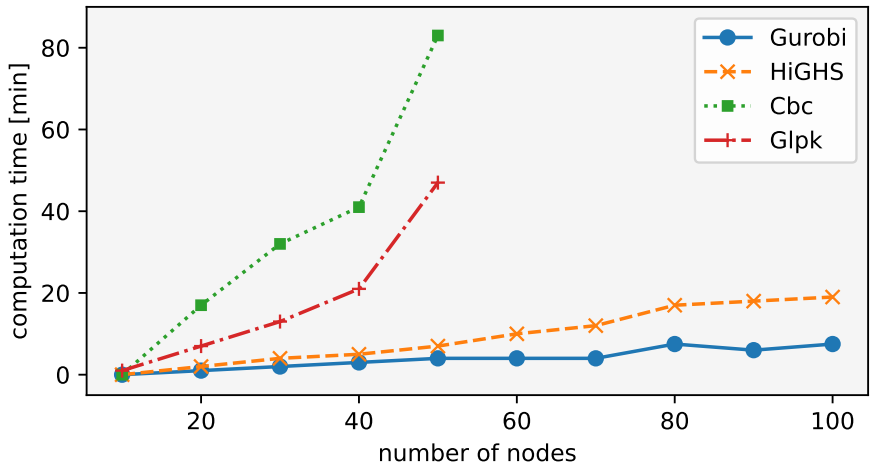
$$\sum_{v\in\mathcal{V},\, s\in\mathcal{S},\, t\in\mathcal{T}} \frac{1}{\eta_{v,s}} \rho_s w_t g_{v,s,t} \leqslant \Gamma_{CO_2} \cdot \sum_{z\in\mathcal{Z}} \gamma_z \tag{16}$$

Following the Kyoto protocol, this cap is given in percent of the emissions of the year 1990, i.e. $\Gamma_{CO_2} \in [0, 1]$.

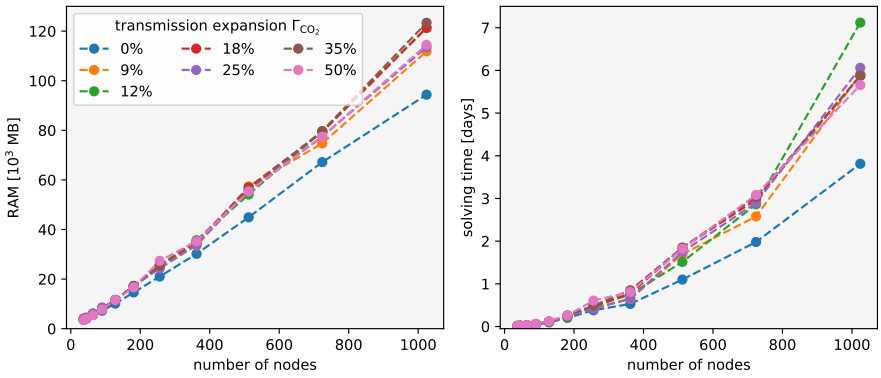
2.4 NUMERICAL SOLVING ALGORITHMS

The electricity system model for capacity planning derived in section 2.3 results in a linear convex optimisation problem. Currently, the standard state-of-the-art solving algorithms for this kind of problem are either the simplex algorithm or the class of interior point methods, which we have chosen for this dissertation.

The **simplex algorithm** generally performs efficiently in practice, but its worst-case efficiency has proven to be exponential [66], and it remains an open question if there is a variation of the algorithm that performs in polynomial time. The idea of the simplex algorithm is to explore the feasible space by descending along its edges, gradually approaching the optimal value. Guaranteed by the convexity of the feasible space, the so-found local minimum is also the global one and, thus, the solution of the problem at hand.



(a) Comparison of open-source solvers coin-or branch and cut (Cbc), GNU Linear Programming Kit (GLPK) and high performance software for linear optimisation (HiGHS) to the commercial solver Gurobi on small, modest sized PyPSA models [68].



(b) Experimental visualisation of time and space complexity of the Gurobi solving algorithm on model formulations as derived in the previous Section.

Figure 10: Complexity of solving algorithms.

IPMs are a group of algorithms that are efficient for solving linear and non-linear convex optimisation problems in polynomial complexity [67]. In contrast to the simplex method, here the main idea is to approach the optimal solution from the interior of the feasible set by an iterative search into the direction of the minimum[1]. An iteration that has diverged from the optimum or that has approached the boundary of the feasible space is penalised by the modification of the objective function with a barrier term.

For a simplified graphical illustration of the simplex algorithm and IPM see Figure 9.

software Although it is not in the spirit of open-source, we employ the highly efficient commercial solver Gurobi [69]. This

1 or maximum

choice of solver is mainly due to strong computational needs that incur by the aim of this dissertation to understand the dynamics of a spatially highly-resolved European grid. For academic purposes, the software provides a free license. For other, such as commercial, political, or non-profit applications, there exists a range of free and open-source solvers, such as Cbc [70] and the GLPK package [71]. However, for large modelling applications that are typical in the energy landscape, they scale badly and are wholly noncompetitive with commercial solutions. Fortunately, current efforts focus on developing the fully open-source solver HiGHS [72] that is competitive with the commercial solver Gurobi. While initial benchmarks show promising results (see Figure 10a), problems of practical interest are larger. For example, solving original electricity models currently results in 60-100 times slower solving times, compared to Gurobi. Considering that even Gurobi requires multiple days to solve huge practical problems (see Figure 10b) illustrates that the use of current open-source solvers is impractical [68].

complexity Let us now have a look on the time and space complexity of the Gurobi solver algorithm applied on the optimisation model that we have derived in Section 2.3, where we focus on the two most commonly analysed resources: time and memory. While we do not analyse the algorithms in detail in this dissertation, we can approximately determine the algorithms complexity based on experiments.

Our experimental data consists of 77 samples and motivates the hypothesis that the algorithm is of polynomial space and time complexity; mainly driven by the number of nodes in the model, see Figure 10b. From the Figure we can also see how much absolute resources we have needed on our High Performance Cluster (HPC) to solve the optimisation problem as a function of the number of nodes in the graph. We can see, that already a model with 1000 nodes requires approximately 120 GB RAM and solved within $3-7$ days, depending on the transmission expansion limit. Therefore, solving the spatially fully-resolved model that consists of more than 5000 nodes is computationally too challenging for most applications, even for our HPC. Therefore, we are in need of methods to reduce the European electricity system model to a lower complexity while retaining high accuracy of the results of the optimisation. This can be achieved by grouping nodes together and aggregating them to *clusters*.

2.5 CLUSTERING ALGORITHMS

Clustering is the task of grouping a set of nodes in a way that nodes that are assigned to the same *cluster* are more similar to each other than to those in other clusters, where the notion of *similarity* needs to be defined in a case-to-case basis.

Popular notions of clusters include grouping objects that are geographically close, dense areas of the data space, intervals or particular statistical distributions. But as the notion of *cluster* cannot be precisely defined, there exists a variety *clustering algorithms* or *clustering models* that vary significantly in their respective properties and the representation of a cluster [73]. In general, we can differentiate between three main categories of clustering. Within every category, there are multiple *clustering models* available.

hard clustering In hard clustering, every node is either assigned to one unique cluster, or it is not assigned at all. A finer distinctions can be the *strict partitioning*, where every node must have an assignment and outliers with no assignment are not allowed. In contrast, *strict partitioning with outliers* allows nodes that do not belong to any cluster. Probably the most famous examples of hard clustering are the k-means algorithm [74] that aims to partition a set of observations into k clusters in which each observation belongs to the cluster with the nearest mean, or spectral clustering methods [75, 76] that make use of the eigenvalues (or, in other words, the spectrum) of the similarity matrix of the data.

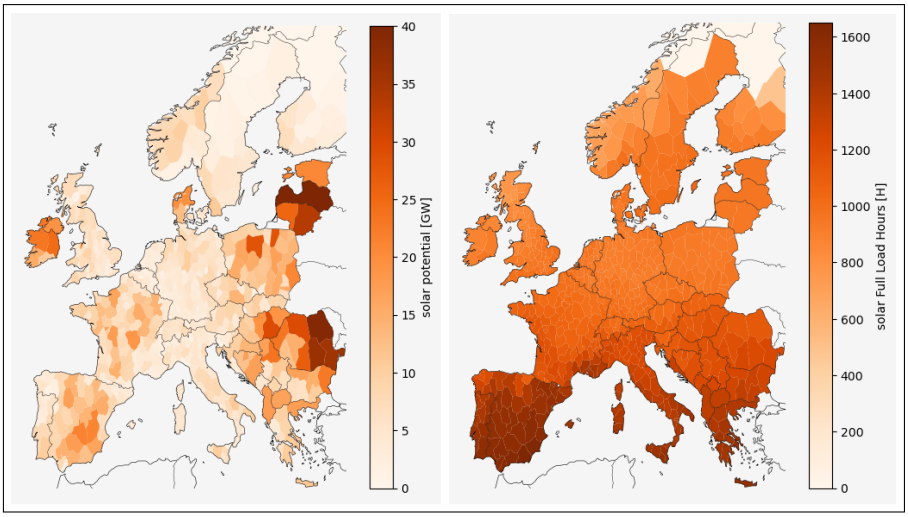
soft clustering In soft (also called fuzzy) clustering, every node belongs to a cluster to a certain degree. The degree can be for example described by a probability or likelihood. Such approaches are considered effective when the data space is very homogeneous and the clusters are interpreted as the centers of higher density of nodes. One example of soft clustering is the fuzzy-c-means algorithm. The degree to which a sample belongs to a cluster increases as the distance to the cluster center decreases [77]. Another popular soft clustering method is the expectation–maximisation algorithm which uses iterative statistical methods to derive the clusters [78].

overlapping clustering Overlapping clustering means that objects from a dataset can be assigned to more than one cluster. This category of node assignment usually involves hard clusters and is particularly relevant in medical datasets that inherently contain overlapping information [79]. One of the simplest and most efficient overlapping clustering methods is known as overlapping k-means, which is an extension of the traditional k-means algo-

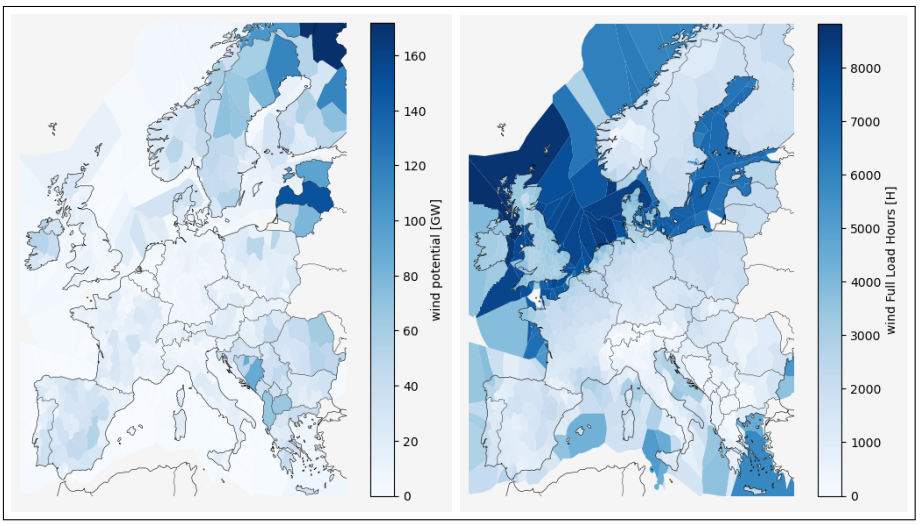
rithm.

In the context of ESM, the dataset that is spatially clustered is the set of substations. Note that they are interpreted as the nodes of a network graph and connected by transmission lines that are interpreted as the edges the graph (see Section 2.2). When it comes to clustering this set, it is impractical to pursue a probabilistic approach on the location of a substation that can draw and feed electricity into the grid, or to assume it on multiple locations simultaneously. Therefore it is common practice to choose one of the hard clustering methods. But which clustering model performs best in approximating the complexity of the spatially highly-resolved electricity grid is still subject to contemporary research. Another challenge is to define a useful term of *similarity*, because of the many features that are associated with a network node (such as electricity demand, renewable potential, grid congestion between nodes, connectivity of a node to other nodes, the topology of the graph, etc. See Section 2.2).

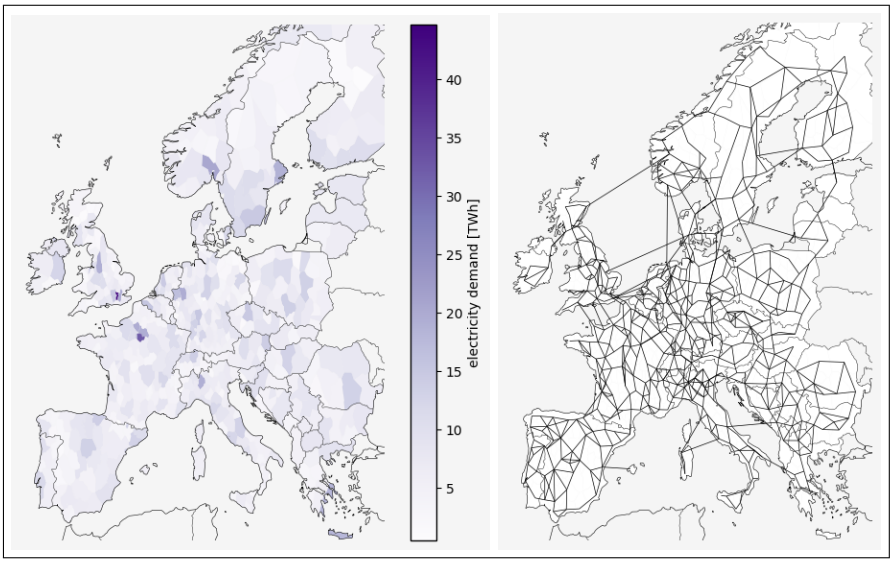
We visualise possible clustering results using the exemplary clustering method k-means (more details can be found in Chapter 3), where we cluster the original network to 512 nodes (Figure 11), 128 nodes (Figure 12) and to 64 nodes (Figure 13). As we can see from the Figures, an application of clustering on the original dataset leads to clusters that are similar in one attribute, but dissimilar in a different one, with respect to the chosen measure. Inevitably, the resulting renewable potentials and electricity demands lead to different solutions of the optimisation problem described in Section 2.3, particularly equation (3) and all associated constraints. We analyse these differences in the following Chapter.



(a) Landuse potentials (left) and FLH (right) for solar PV.

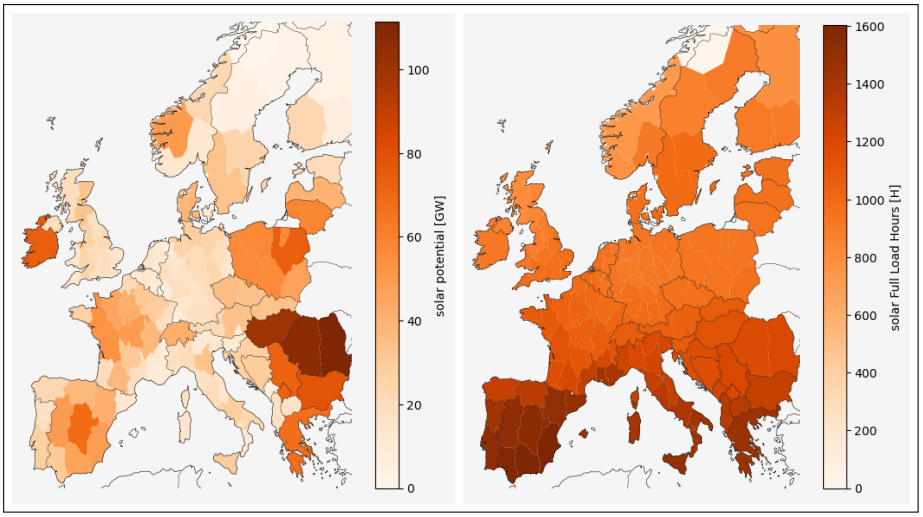


(b) Landuse potentials (left) and FLH (right) for on- and offshore wind.

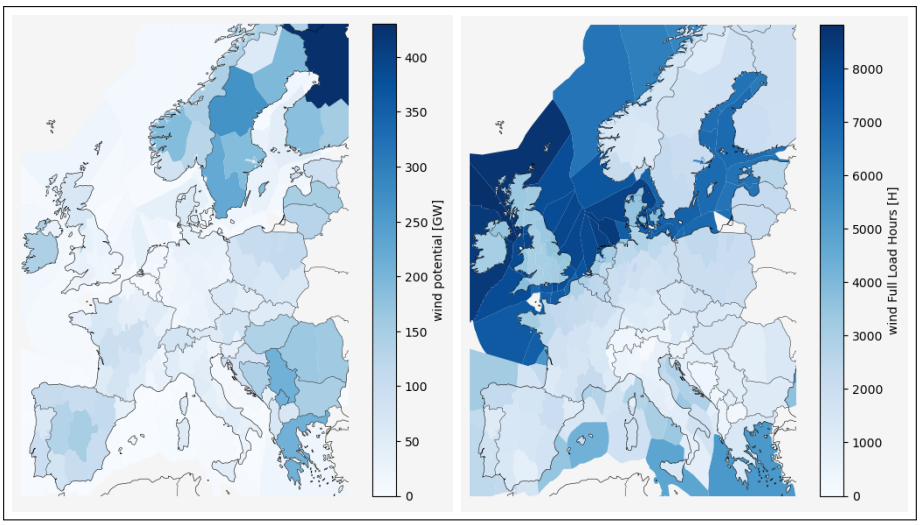


(c) Electricity demand (left) and network topology (right).

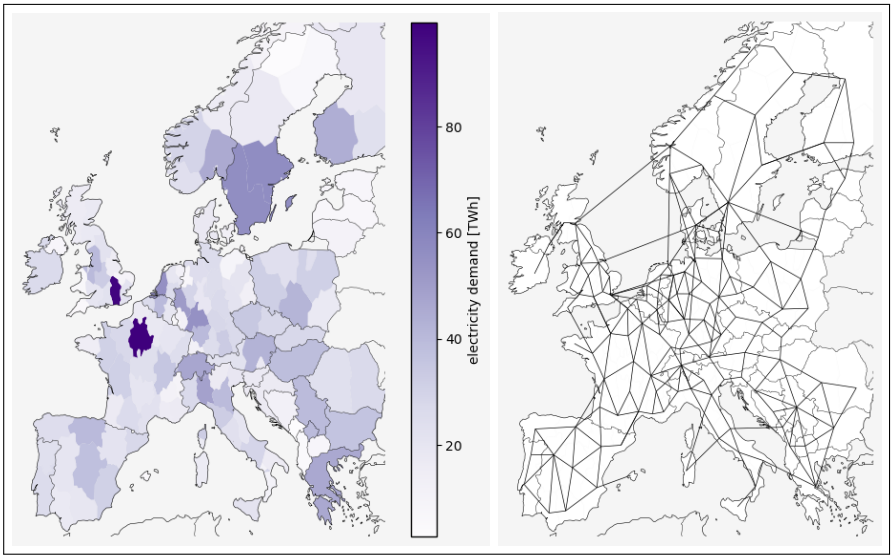
Figure 11: Visualisation of the clustered PyPSA-Eur dataset to an exemplary resolution of 512 nodes.



(a) Landuse potentials (left) and FLH (right) for solar PV.

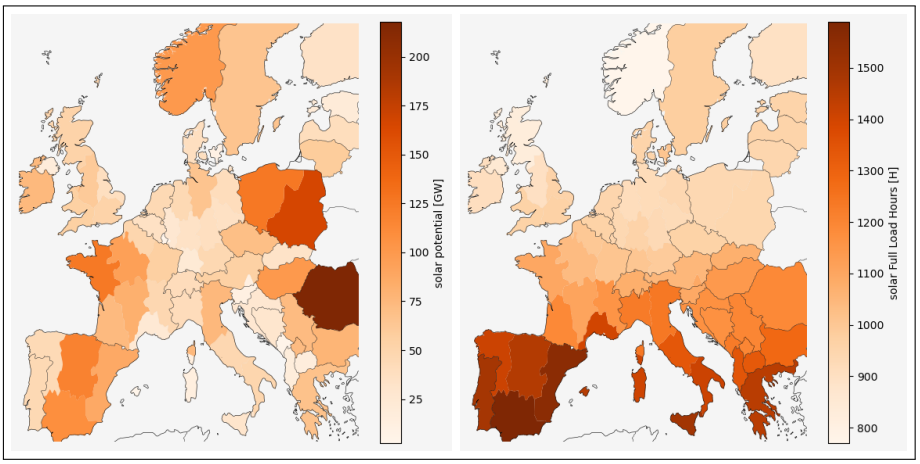


(b) Landuse potentials (left) and FLH (right) for on- and offshore wind.

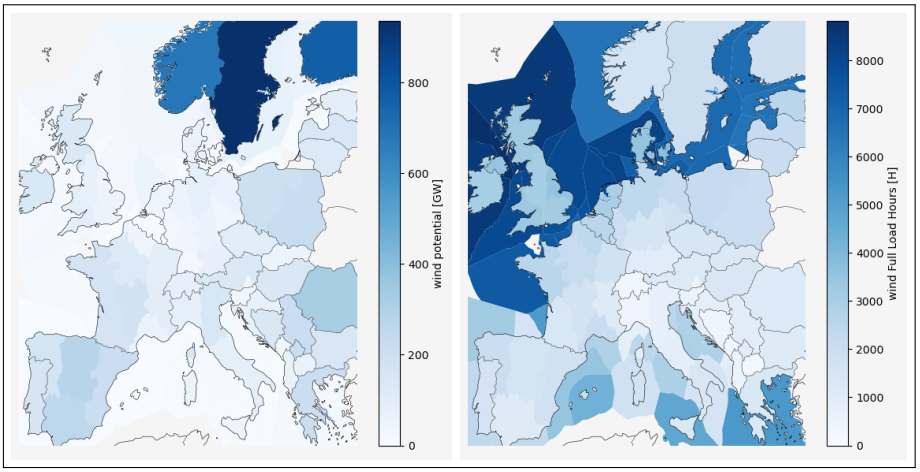


(c) Electricity demand (left) and network topology (right).

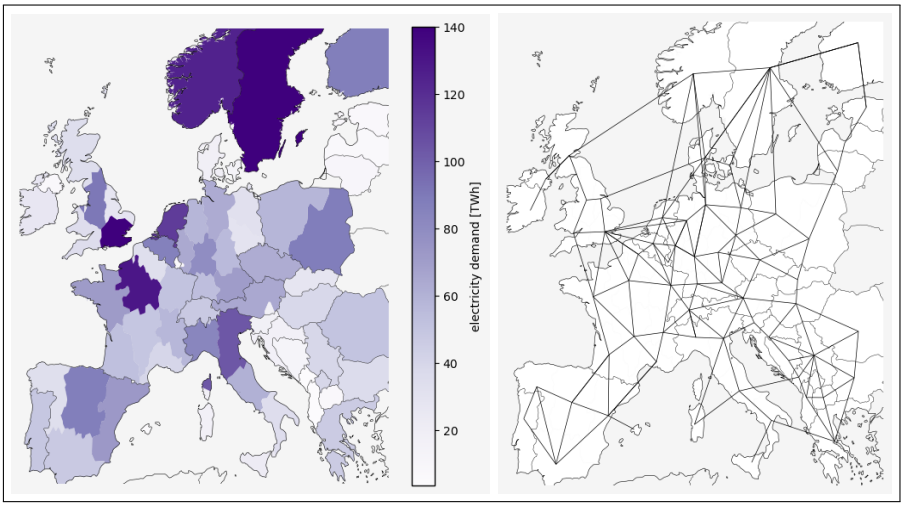
Figure 12: Visualisation of the clustered PyPSA-Eur dataset to an exemplary resolution of 128 nodes.



(a) Landuse potentials (left) and FLH (right) for solar PV.



(b) Landuse potentials (left) and FLH (right) for on- and offshore wind.



(c) Electricity demand (left) and network topology (right).

Figure 13: Visualisation of the clustered PyPSA-Eur dataset to an exemplary resolution of 64 nodes.

3

EFFECTS OF RESOURCE GRANULARITY AND TRANSMISSION GRID RESOLUTION

CONTENTS OF THIS CHAPTER ARE BASED ON

Martha Maria Frysztacki et al. "The strong effect of network resolution on electricity system models with high shares of wind and solar." In: *Applied Energy* 291 (2021), p. 116726. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2021.116726

3.1 INTRODUCTION

In this Chapter we analyse the distinct implications of clustering VRE resources and the transmission grid. We do that by introducing a novel methodology to disentangle these two competing spatial effects of resource and network resolution, so that for the first time their different impacts on system costs and technology choices can be quantified.

The common approach from the literature to cluster the model based on administrative boundaries such as country borders [36, 44, 45] fails to account for the variation of resources inside large countries like Germany or France because of two contrasting reasons: (i) Aggregating sites with low-yield together with sites that have high-yield takes away the opportunity to optimise generation placement, which distorts investment decisions and drives up costs. (ii) Aggregating diverse resources to single nodes tends to underestimate network-related costs, since the models are blind to network bottlenecks that might hinder the welfare-enhancing integration of renewable resources located far from demand centers. The effects of network restrictions are all the more important given the apparent low public acceptance for new overhead transmission lines, observed in Germany [80] and across Europe [81], and the long planning and construction times for new grid infrastructure [23].

Our novel methodology allows us to separate the effects of (i) and (ii) and to analyse their individual impacts on the compound optimisation. We demonstrate the methodology by running simulations in the proposed European electricity system of Chapter 2 with a higher spatial resolution than has previously been achieved

in the literature. Investments and operation of generation, storage and transmission is jointly optimised under a 95% reduction in CO_2 emissions compared to 1990, which is consistent with European targets for 2050 [82]. The European electricity system model is sequentially clustered from 1024 nodes down to 37 nodes in order to examine the effects on optimal investments in generation, transmission and storage.

state of the art Previous work in the engineering literature has focused on the effect of different network clustering algorithms [83], on the flows in single power flow simulations [84, 85], or used clustering algorithms that are dependent on specific dispatch situations [86–88] and therefore unsuitable when making large changes to generation and transmission capacities.

In the planning literature that considers a high share of renewables in the future energy system, the effects of clustering applied separately to wind, solar and demand were investigated in [89, 90], but neglected potential transmission line congestion within large regions. [91] extended the study by including a synthesised grid and renewable profiles, but it ignored the existing topology of the transmission grid. Effects of varying the resolution were not considered in either of the studies. Recent work has examined regional solutions for the European power system, but did not take into account existing transmission lines, potential low public acceptance for grid reinforcement or the grid flow physics [43]. Other studies have examined transmission grid expansion at substation resolution, but either the temporal resolution was too low to account for wind and solar variability [92, 93], only single countries were considered [93–95], or transmission expansion was not co-optimised with generation and storage [92, 96, 97]. The competing effect of clustering transmission lines versus variable resource sites on the share of renewables was also discussed in [98], but the report did not provide an analysis of how strongly the respective clustering impacts modelling and planning results. The effects of model resolution on system planning results were considered for the United States in [99], where a cost-benefit was seen for higher wind and solar resolution, but the resource resolution was not separated from the network resolution, and only a small number of time slices were considered to represent weather variations.

research contributions Advances in solver algorithms and code optimisation in the modelling framework PyPSA [63], as well as hardware improvements, allow us to achieve what was previously not possible in the literature: the co-optimisation of transmission, generation and storage at high temporal and spatial resolution across the whole of Europe, while taking into account

linearised grid physics, existing transmission lines and realistic restrictions on grid reinforcement. In previous work large effects of spatial resolution on investment results were seen [100, 101], but because the resource and network resolution were changed in tandem, it was not possible to analyse which effect dominates the results. Here, we present a novel study design that separates the effects of resource and network resolution, and demonstrate the substantial differences between the two effects using the high-resolution simulations enabled by recent software and hardware advances.

3.2 METHODS

In this Section we present an overview of the study design, particularly details on the clustering methodology.

3.2.1 *Clustering Study Design*

The nodes of the model introduced in Section 2.3 are successively clustered in space into a smaller number of representative nodes using the k-means algorithm [74]. This groups close-by nodes together, so that, for example, multiple nodes representing a single city are merged into one node. Nodes from different countries or different synchronous zones are not allowed to be merged; to achieve this, the overall number of desired nodes is partitioned between the countries and synchronous zones before the k-means algorithm is applied in each partition separately. In total there are 37 'country-zones' in the model, i.e. regions of countries belonging to separate synchronous zones.

Figure 14, Case 1 shows exemplary clustering results for Ireland and the United Kingdom (where Northern Ireland is in a separate synchronous zone to Great Britain). Once the nodes have been clustered, they are reconnected with transmission corridors representing the major transmission lines from the high-resolution model. Electricity demand, conventional generation and storage options are also aggregated to the nearest network node. More technical details on the clustering can be found in Subsection 3.2.4. An analysis of the effects of clustering on the network flows can be found in Annex B.1.

3.2.2 *Resource Versus Network Resolution Case Studies*

To separate the effects of the spatial resolution on the RE resources and the network, we consider three cases in which they are clustered differently. The three cases are summarised in Table 3 and

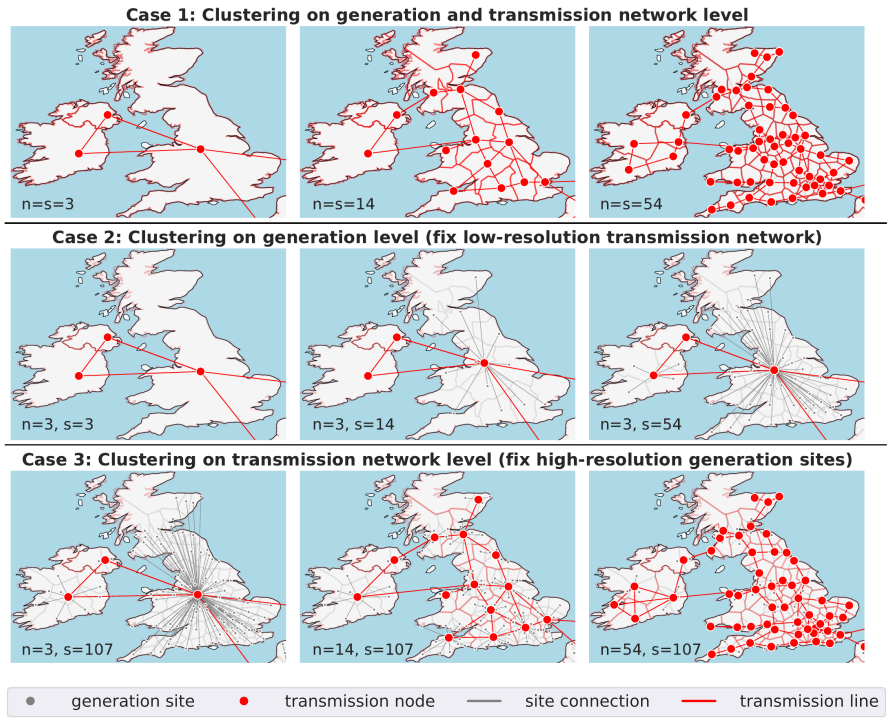


Figure 14: Clustering of network nodes (red, ν) and renewable sites (grey, s) in each of the cases (rows) for Ireland and the United Kingdom at different levels of clustering (columns).

shown graphically in Figure 14 for each case (rows) and for each level of clustering (columns).

In **Case 1** the wind and solar sites are clustered to the same resolution as the network. The number of clusters is varied between 37, the number of country-zones, and 1024, which represents the maximum resolution for which generation, transmission and storage investment can be co-optimised in reasonable time. The number of nodes is increased in half-powers of 2, so that eleven different resolutions are considered: $\mathcal{B} := \{37\} \cup \left\{ \lfloor \sqrt{2^i} \rfloor \right\}_{i=11,\dots,20}$.

In **Case 2** network bottlenecks inside each country-zone are removed so that there are only 37 transmission nodes, and only the resolution of the wind and solar generation is varied. Inside each country-zone, all wind and solar generators are connected to the central node. This allows the optimisation to exploit the best wind and solar sites available.

Finally in **Case 3** we fix a high resolution of renewable sites and vary the number of network nodes, in order to explore the effects of network bottlenecks. Each renewable site is connected to the nearest network node, where the transmission lines, electricity demand, conventional generators and storage are also connected.

For each case we optimise investments and operation for wind and solar power, as well as open cycle gas turbines, batteries, hy-

Case	Short name	Description
1	Simultaneous clustering	Successive increase in number of generation sites s and transmission nodes v. Every node v has exactly $\lvert\mathcal{S}^{re}\rvert$ RE sites.
2	Clustering on siting resolution	Fix the transmission network to one-node-per-country-zone, i.e. $\lvert\mathcal{V}\rvert = 37$ and increase the number of generation and storage sites $s \in \mathcal{B}$.
3	Clustering on network nodes	Maintain a high resolution of generation sites $s = 1024$ and successively increase the number of transmission nodes, $\lvert\mathcal{V}^K\rvert = K \in \mathcal{B}$

Table 3: Case descriptions. ($\mathcal{B} = \{37\} \cup \left\{\lfloor\sqrt{2^i}\rfloor\right\}_{i=11,\dots,20} = \{37, 45, 64, 90, 128, \dots 1024\}$)

drogen storage and transmission. Flexibility from existing hydroelectric power plants is also taken into account. The model is run with perfect foresight at a 3-hourly temporal resolution over a historical year of load and weather data from 2013, assuming a 95% reduction in CO_2 emissions compared to 1990. The temporal resolution is 3-hourly to capture changes in solar generation and electricity demand while allowing reasonable computation times. The technology selection is also limited for computational reasons. More details on the investment optimisation can be found in section 2.3 of the foundations chapter.

For each simulation we also vary the amount of new transmission that can be built, in order to understand the effect of possible grid reinforcements on the results. The model is allowed to optimise new transmission reinforcements to the grid as it was in 2018, up to a limit on the sum over new capacity multiplied by line length measured relative to the grid capacity in 2018. For example, a transmission expansion of 25% means that on top of 2018's grid, new lines corresponding to a quarter of 2018's grid can be added to the network. The exact constraint is given in equation (7).

3.2.3 *Network Preparation*

Before the clustering algorithm can be applied to the network, several simplifications are applied to the data.

In order to avoid the difficulty of keeping track of different voltage levels as the network is clustered, all lines are mapped to their

electrical equivalents at 380 kV, the most prevalent voltage in the European transmission system. If the original reactance of line (v, w) was $x_{(v,w)}$ at its original voltage $v_{(v,w)}$, the new equivalent reactance becomes

$$x'_{(v,w)} = x_{(v,w)} \left(\frac{380\ \text{kV}}{v_{(v,w)}} \right)^2 \tag{17}$$

This guarantees that the per unit reactance is preserved after the equivalencing.

The impedances and thermal ratings of all transformers are neglected, since they are small and cannot be consistently included with the mapping of all voltage levels to 380 kV.

Univalent nodes, also known as dead-ends, are removed sequentially until no univalent nodes exist. That is, if node v has no other neighbor than node w, then node v is merged to node w. We repeat the process until each node is multi-valent and update the merged node attributes and its attached assets (electricity demand, generators and storage units) according to the rules in Table 14.

HVDC lines in series or parallel are simplified to a single line (v, w) using the rules in Tables 15 and 16. Capital costs per MW of capacity for HVDC lines (v, w) with length $l_{(v,w)}$ and a fraction $u_{(v,w)} \in [0, 1]$ underwater are given by

$$c_{(v,w)} = 1.25 \cdot l_{(v,w)} \left(u_{(v,w)}\, c_{\text{marine}} + (1 - u_{(v,w)})\, c_{\text{ground}} \right)$$

where c_{marine} is the capital cost for a submarine connection and c_{ground} for an underground connection. The factor of 1.25 accounts for indirect routing and height fluctuations.

3.2.4 *Clustering Methodology*

Different methods have been used to cluster electricity system networks in the literature. Because k-means clustering [74] is one of the most commonly used algorithms to group objects, we have chosen a variant of it that is based on the geographical location of the original nodes in the Graph. We weight every node by its average load and conventional capacity, since this represents how the topology of the network was historically planned to connect major generators to major loads, see Chapter 1 (particularly Figure 1). It leaves the long transmission lines between regions, which are expensive to upgrade and are more likely to encounter low local acceptance, unaggregated, so that these lines can be optimised in the model. Regions with a high density of nodes, for example around cities, are aggregated together, since the short

lines between these nodes are inexpensive to upgrade and rarely present bottlenecks. Geographical k-means clustering has the advantage over other clustering methods of not making any assumptions about the future generation, storage and network capacity expansion.

Other clustering methods applied in the literature are not suitable for the co-optimisation of supply and grid technologies: these include clustering based on electrical distance using k-medoids [84, 102], a modified version of k-medoids to avoid assigning both end nodes of a critical branch to the same zone [22], hierarchical clustering [103], or k-decomposition and eigenvector partitioning [104] (which we do not use because we want to optimise new grid reinforcements that alter electrical distances), spectral partitioning of the graph Laplacian matrix [85] (avoided for same reason), an adaptation of k-means called k-means++ combined with a max-p regions algorithm applied to aggregate contiguous sites with similar wind, solar and electricity demand [89] (avoided since we want a coherent clustering of all network nodes and assets), hierarchical clustering based on a database of electricity demand, conventional generation and renewable profiles including a synthesised grid [91] (avoided for the same reason and because we do not want to alter the topology of the existing transmission grid), k-means clustering based on renewable resources as well as economic, sociodemographic and geographical features [105] (avoided because we need a clustering focused on network reduction), as well as clustering based on the zonal power transfer distribution factor (PTDF) to detect congestion zones [86], to yield the same flow patterns as the original network [106] or to analyse policy options and emissions [107] (avoided because they encode electrical parameters that change with reinforcement), available transfer capacity (ATC) [88] (avoided because they depend on pre-defined dispatch patterns) and locational marginal price (LMP) [87] (again avoided because they depend on pre-defined dispatch patterns).

We do not allow nodes in different countries or different synchronous zones to be clustered together, so that we can still obtain country-specific results and so that all HVDC between synchronous zones are preserved during the aggregation. This results in a minimum number of 37 clustered nodes for the country-zones. First we partition the desired total number K of clusters between the 37 country-zones, then we apply the k-means clustering algorithm within each country-zone.

In order to partition the K nodes between the 37 country-zones, the following minimisation problem is solved

$$\operatorname{argmin}_{\{K_z\}\in\mathbb{N}^{37}} \sum_{z=1}^{37} \left(K_z - \frac{d_z}{\sum_{y=1}^{37} d_y} K \right)^2 \quad \text{s.t.} \sum_{z=1}^{37} K_z = K \tag{18}$$

Then the k-means algorithm is applied to partition all nodes inside each country-zone z into K_z clusters. The algorithm finds the partition that minimises the sum of squared distances from the mean position of each cluster $(x_c, y_c)^T \in \mathbb{R}^2$ to the positions $(x_v, y_v)^T \in \mathbb{R}^2$ of its members $v \in \mathcal{V}_c$

$$\min_{\{(x_c,y_c)^T\in\mathbb{R}^2\}} \sum_{c=1}^{K_z} \sum_{v\in\mathcal{V}_c} w_v \left\| \begin{pmatrix} x_c \\ y_c \end{pmatrix} - \begin{pmatrix} x_v \\ y_v \end{pmatrix} \right\|_2 \tag{19}$$

The original formulation of k-means is designed without weighting w_v. However, we choose a weight proportional to nominal power $G_{v,s}$ for conventional generators s and averaged electricity demand $\langle d_{v,t} \rangle_t$:

$$w_v = \frac{\sum_{s\in\mathcal{S}^{\text{conv.}}} G_{v,s}}{\sum_{v\in\mathcal{V},\, s\in\mathcal{S}^{\text{conv.}}} G_{v,s}} + \frac{\langle d_{v,t}\rangle_t}{\sum_{v\in\mathcal{V}} d_{c,T}}, \quad \forall v \in \mathcal{V} \tag{20}$$

Then, w_v is normalised according to:

$$w_v \mapsto \left\lfloor \frac{100 \cdot w_v}{\|w\|_{\max}} \right\rfloor \tag{21}$$

The weight w_v is chosen such that it incorporates an approximation of the transmission system. It essentially represents how the topology of the network was historically planned to connect major generators to major loads, as we have outlined in the introduction of the thesis (see Chapter 1, particularly Figure 1).

The optimisation is run with 10^3 different centroid seeds, a maximum number of iterations for a single run of $3 \cdot 10^4$ and a relative tolerance with regards to inertia to declare convergence of 10^{-6} .

Attributes of the nodes in $\mathcal{V}_c$ and their attached assets are aggregated to the clustered node c according to the rules in Table 14.

Lines connecting nodes $v \in \mathcal{V}_c$ in cluster c with nodes $w \in \mathcal{V}_d$ in cluster d, given by the set $E_{(c,d)}$

$$E_{(c,d)} = \{(v,w) : v \in \mathcal{V}_c,\ w \in \mathcal{V}_d\}, \quad \forall c, d \in \mathcal{V}^K,\ c \neq d \tag{22}$$

are aggregated to a single representative line. The length of the representative line is determined using the haversine formula[1] multiplied by a factor of 1.25 to take indirect routing into account.

The representative line inherits the attributes of the lines $E_{(c,d)}$ as described in Table 16. If any of the replaced lines in $E_{(c,d)}$ had the attribute that their capacity was extendable, then the aggregated line inherits this extendability. An analysis of the effects of clustering on the network flows can be found in Annex B.1.

For Case 1, generators are clustered to the same resolution as the network. Time-series containing hourly resolved capacity factors $\bar{g}_{v,s,t} \in [0,1]$ for VRE generation are aggregated using a weighted average

$$\bar{g}_{c,s,t} = \frac{1}{\sum_{v \in \mathcal{V}_c} w_{v,s}} \sum_{v \in \mathcal{V}_c} w_{v,s}\, \bar{g}_{v,s,t} \tag{23}$$

$$\forall c \in \mathcal{V}^K,\ s \in \mathcal{S}^{re},\ t \in \mathcal{T}$$

The resulting capacity factor $\bar{g}_{c,s,t}$ is in $[0,1]$ by definition. For renewables, the weighting $w_{v,s}$ is proportional to the maximal yearly yield for technology s at node v, found by multiplying the maximal install-able capacity $G^{max}_{v,s}$ with the average capacity factor. In the case of conventional technologies the weightings are distributed equally, i.e $w_{v,s} = 1$. Note that there is per-se no relation between the weightings $w_{v,s}$ and the bus weightings w_v in equation (20).

For Case 2, the network is fixed at 37 nodes, and the wind and solar generators are merged in the aggregation step. Time series for VRE availability are aggregated according to equation (23) to their respective resolution.

For Case 3, the network is clustered, but wind and solar generators are not merged in the aggregation step. Their time series remain fixed at high resolution of 1024 nodes.

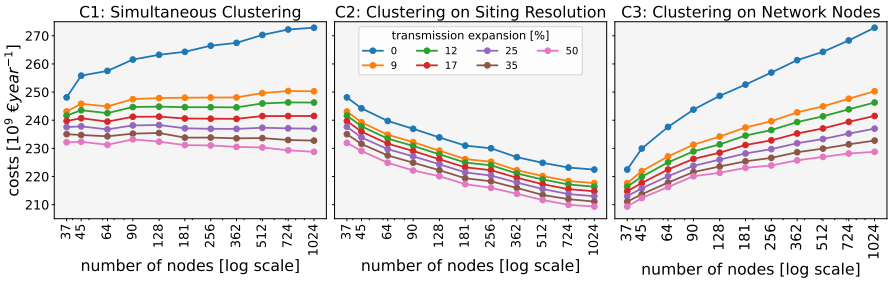


Figure 15: Total annual system costs as a function of the number of clusters for Cases 1, 2 and 3.

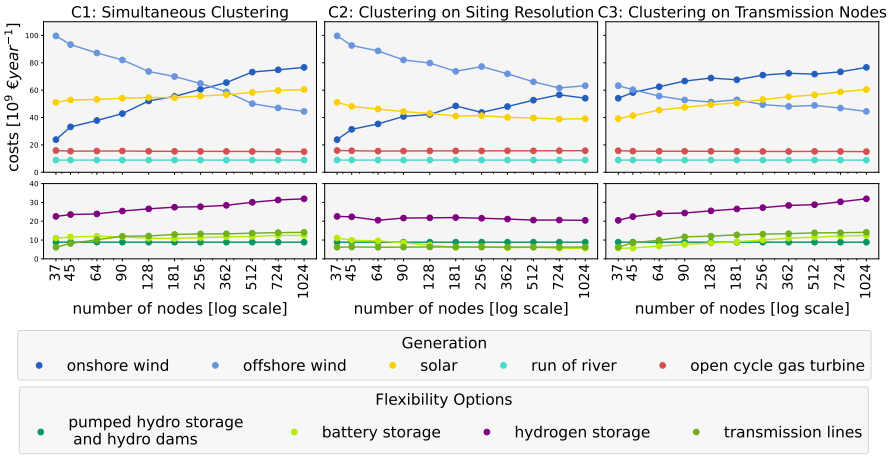


Figure 16: Breakdown of the annual system costs for generation (top) and flexibility options (bottom) as a function of the number of clusters for Cases 1, 2 and 3 when there is no grid expansion.

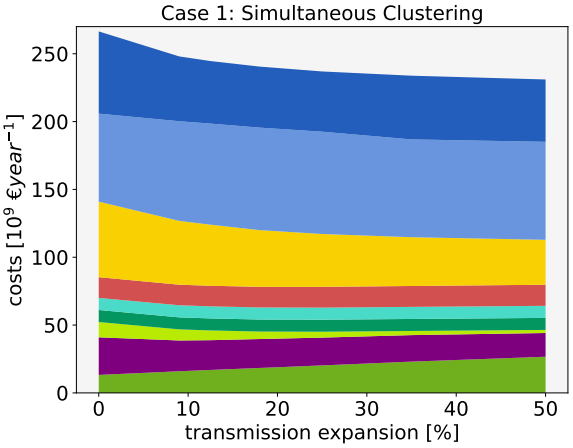


Figure 17: Costs as a function of the transmission expansion level for 256 nodes in Case 1.

3.3 RESULTS

3.3.1 *Increasing Number of both Generation Sites and Transmission Nodes*

If the resource and network resolutions increase in tandem according to Case 1 without grid expansion, the total annual system costs in Figure 15 rise gently with the increasing number of nodes, reaching a maximum of 273 billion euros per year at 1024 nodes, which is 10% more expensive than the solution with 37 nodes. This corresponds to an average system cost of 87 €/MWh. If some transmission expansion is allowed, costs are lower, and there is almost no change in total system costs as the number of nodes is varied.

However, the fact that costs are flat does not mean that the solutions are similar: a large shift from offshore wind at low resolution to onshore wind at high resolution can be observed in the left graph of Figure 16 (Case 1). This is an indication that spatial resolution can have a very strong effect on energy modelling results. To understand what causes this effect, we must examine Cases 2 and 3.

3.3.2 *Importance of Wind and Solar Resource Granularity*

In Case 2 we use the lowest network resolution of 37 nodes, corresponding to one-node-per-country-zone, and investigate the effect of changing the number of wind and solar sites on the results. As the resolution increases, total costs without grid expansion in Figure 15 drop by 10% from 248 to 222 billion euro per year. Although the slope of the cost curve appears constant, note that the x-axis is logarithmic, so that the rate of cost decrease slows as the number of sites increases.

The cost reduction is driven by strong changes in the investment between generation technologies, particularly the ratio between offshore and onshore wind (see Figure 16). At low spatial resolution, good and bad onshore sites are mixed together, diluting onshore capacity factors and making onshore a less attractive investment. Figure 6 or Figures 42a and 42b in Annex B.2 show how the capacity factors for wind and solar vary across the continent. While offshore is spatially concentrated and solar capacity factors are relatively evenly spread in each country-zone, onshore wind is stronger near coastlines. At high spatial resolution the model can choose to put onshore wind only at the best sites (within land

1 the haversine formula computes the great-circle distance between two points on a sphere

restrictions), increasing average capacity factors and thus lower the per-MWh-cost. (The increasing average capacity factors are plotted in Figure 44 in Annex B.4.) As a result, onshore wind investments more than double from 24 to 54 billion euros per year, while offshore investments drop 37% from 100 to 64 billion per year and solar by 23%. The biggest effect on the technology mix is when going from 37 to around 181 clusters; beyond that the changes are smaller.

3.3.3 *Impact of Transmission Bottlenecks*

In Case 3 we fix a high resolution of wind and solar generators (1024 sites) and vary the resolution of the transmission network to gauge the impact of transmission bottlenecks. With 37 network nodes many bottlenecks are not visible, so costs are lower, but as the resolution increases to 1024 nodes it drives up the costs by 23%. Note that because the x-axis is logarithmic, the highest rate of cost increase is when the number of nodes is small.

As can be seen from the breakdown in Figure 16, the rising transmission investments from the higher resolution only have a small contribution to the result. Instead, rising costs are driven by generation and storage. Unlike Cases 1 and 2, the ratio between the generation technologies does not change dramatically with the number of clusters, but the capacities for onshore wind, solar, batteries and hydrogen storage all rise.

The transmission bottlenecks limit the transfer of power from the best sites to the load, forcing the model to build onshore wind and solar more locally at sites with lower capacity factors. Average capacity factors of onshore wind and solar sink by 11% and 6% respectively with no grid expansion (see Figure 44 in Annex B.4), meaning that more capacity is needed for the same energy yield. Curtailment is generally low in the optimal solution (around 3% of available wind and solar energy) and has less of an effect on costs (see Figure 45 in Annex B.5).

Investment in battery and hydrogen storage rises with the number of network nodes since the storage is used to balance local wind and solar variations in order to avoid overloading the grid bottlenecks.

3.3.4 *Comparison of the Three Cases*

Separating the effects of resource resolution from network resolution reveals that the apparent stability of total system costs in Case 1 in Figure 15 as the number of clusters changes, as reported in [100], is deceptive. In fact, the sinking costs from the higher

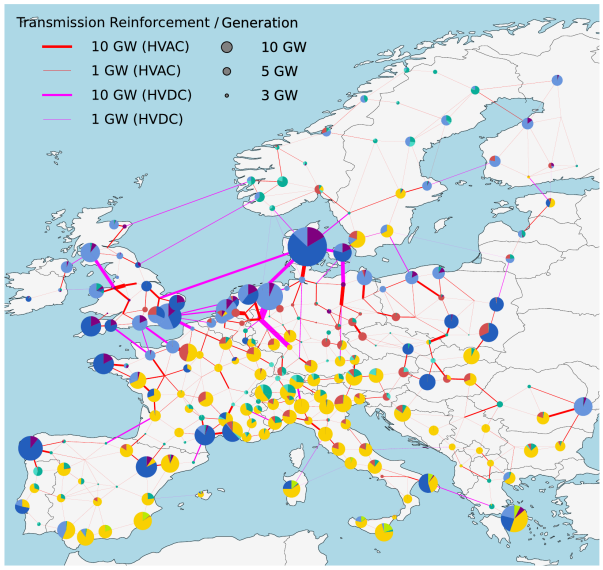


Figure 18: Example of investments with 25% grid expansion and 256 nodes in Case 1.

resource resolution are counter-acted by the rising costs from network bottlenecks. With no grid expansion, the system cost of network bottlenecks is double the benefit of the higher resource resolution.

While these two effects offset each other at the level of total system costs, they have very different effects on the technology mix. Resource resolution leads to much stronger investment in onshore wind, once good sites are revealed. Network bottlenecks have only a weak effect on the ratio of generation technologies, but lead to lower average capacity factors and drive up storage requirements.

3.3.5 *Benefits of Grid Expansion*

Grid expansion does not affect the main qualitative features of the different Cases, but it does have the overall effect of lowering total system costs. In Case 1, the total cost-benefit of grid expansion is highest at around 16% for a 50% increase in grid capacity, with the marginal benefit still increasing, but it is subject to diminishing returns (see Figure 47 in Annex B.7 for a comparison of the marginal benefit to the cost of transmission). The first 9% of additional grid capacity brings total cost savings of up to 8%, but for each extra increment of grid expansion, the benefit is weaker. There is more benefit from grid expansion at a higher number of nodes, since the higher network resolution reveals more critical bottlenecks in the transmission system.

The total savings from 25% and 50% grid expansion are around 36 and 44 billion euros per year respectively. In a 2018 study ENTSO-E examined scenarios with up to 75% renewable electricity in Europe in 2040 with and without planned ten year network development plan (TYNDP) grid expansions (corresponding to around 25% grid expansion), given fixed demand and a fixed generation fleet. They found that the grid reinforcements reduce generation costs by 43 billion euros per year. This is higher than our cost-benefit for 25% grid expansion, despite their study's lower level of renewable electricity, because in our simulations the generation and storage fleet can be re-optimised to accommodate the lower level of grid capacity, and because we subtract the costs of new grid reinforcement from the cost-benefit (a contribution of around 3.5 billion euros per year).

The breakdown of system cost as the grid is expanded for a fixed number of clusters (256), plotted in Figure 17, reveals how costs are reduced. Although the investment in transmission lines rises, generation and storage costs reduce faster as investment shifts from solar and onshore wind to offshore wind. Offshore wind reduces costs because of its high capacity factors and more regular generation pattern in time. It can be transported around the continent more easily with more transmission, and benefits from the smoothing effects over a large, continental area that grid expansion enables. The map of investments in Figure 18 shows how offshore wind is balanced by new transmission around the North Sea, which smooths out weather systems that roll across the continent from the Atlantic. Further transmission reinforcements bring energy inland from the coastlines to load centers. With more transmission, there is less investment in battery and hydrogen storage, as a result of the better balancing of weather-driven variability in space.

Turning to Case 3, we see that grid expansion mitigates the effect of network resolution by allowing bottlenecks to be alleviated. For a 50% increase in transmission capacity, total costs rise by only 4% from 90 nodes up to 1024 nodes. The distribution of investments between technologies also barely changes in this range (see Figure 43 in Annex B.3). This means that a grid resolution of around 90 nodes can give acceptable solutions for grid expansion scenarios if computational resources are limited, as long as the wind and solar resolution is high enough (as in Case 2, 181 generation sites would suffice). Without grid expansion, a higher grid resolution is needed to capture the effects of bottlenecks and achieve reliable results.

3.3.6 *Computation Times and Memory*

Besides the poor availability of data at high resolution, one of the main motivations for clustering the network is to reduce the number of variables and thus the computation time of the optimisation. In Annex B.8 Figure 48 the memory and solving time requirements for each Case are displayed as a function of the number of clusters. Both memory and solving time become limiting factors in Cases 1 and 3, with random access memory (RAM) usage peaking at around 115 GB and solving time at around 6 days for 1024 clusters. Beyond this number of clusters no consistent convergence in the solutions was seen.

Case 2, where the network resolution is left low and the resource resolution is increased, shows seven times lower memory consumption and up to thirteen times faster solving times compared to Cases 1 and 3 for the same number of clusters. It is therefore the network resolution rather than the resource resolution that drives up computational requirements, which it does by introducing many new variables and possible spatial trade-offs into the optimisation. Since Case 2 proved relatively reliable for estimating the ratio between technologies, if not their total capacity, it may prove attractive to increase the resource resolution rather than the network resolution if computational resources are limited.

3.3.7 *Further results*

Further results on curtailment, average capacity factors, the distribution of technologies between countries, maps, network flows and shadow prices can be found in Annex B.

3.4 CONCLUSIONS

From these investigations we can draw several conclusions. Modelers need to take account of spatial resolution, since it can have a strong effect on modelling results. In our co-optimisation of generation, storage and network capacities, higher network resolution can drive up total system costs by as much as 23%. Higher costs are driven by the network bottlenecks revealed at higher resolution that limit access to wind and solar sites with high capacity factors. On the other hand, resource resolution affects the balance of technologies by revealing more advantageous onshore wind sites. In both cases the system costs are driven more by the usable generation resources than investments in the grid or storage.

If grid expansion can be assumed, a grid resolution of 90 nodes for Europe is sufficient to capture costs and technology investments as long as the solar and onshore wind resolution is at least around 181 nodes. If grid expansion is not possible, a higher spatial resolution for the grid is required for reliable results on technology choices. Since grid expansion is likely to be limited in the future by low public acceptance, more attention will have to be paid to the computational challenge of optimising investments at high spatial granularity.

3.5 CRITICAL APPRAISAL

The need to solve the models at high spatial resolution and 3-hourly temporal resolution in reasonable time means that compromises have been made elsewhere: the conventional generation technologies are limited to hydroelectricity and gas turbines, the storage is limited to batteries and hydrogen storage, only a single weather year is modeled, and ancillary services, grid losses, discretisation of new grid capacities, distribution grids and forecast error are not modeled. This allows us to focus on the main interactions between wind, solar and the transmission grid; the effects of the other factors are expected to be small [20] since wind and solar investment dominates system costs. If it were cost-effective to build dispatchable low-carbon generators like nuclear or fossil generators with carbon capture and sequestration, then the effects of resource and network resolution would be dampened, since there would be less wind and solar investment.

Some of the quantitative conclusions may depend on the technology assumptions, such as the relative cost of solar PV, onshore wind and offshore wind. However, investigations of the sensitivities of similar models to generation costs [108] and of the near-optimal space of solutions [109] have shown that a large share of wind in low-cost scenarios for Europe is robust across many scenarios because of the seasonal matching of wind to demand in Europe. It is the interactions between wind and the transmission grid that drive the results in this paper.

The results may also change as additional energy sectors are coupled to the power sector, such as building heating, transport and non-electric industry demand. While extra flexibility from these sectors might offer an alternative to grid expansion, grid expansion is still expected to be cost-effective [110], while the effects of resource resolution on the optimal solution remain the same.

In the present paper different market structures to today's are assumed, namely nodal pricing to manage grid congestion, and a

high CO_2 price to obtain a 95% CO_2 reduction compared to 1990 levels.

We weighted the distribution of wind and solar inside each nodal region (Voronoi cell) proportional to the installable capacity and capacity factor at each weather grid cell [51]. This means good and bad sites are not mixed evenly, but skewed slightly towards good sites. This effect disappears at high resolution, where the capacity factor is more uniform inside each Voronoi cell.

Another approach would be to keep a low one-node-per-country network resolution and then have multiple resource classes defined not by region, like our Case 2, but by capacity factor [46, 111, 112] (e.g. a good class with sites with FLHs above 2000, a medium class between 1500 and 2000, and a bad class below 1500). This would also be beneficial but would not be compatible with the increasing grid resolution, since the generators in each class would be spread non-contiguously over the country.

4

IMPROVING SPATIALLY LOW-RESOLVED ELECTRICITY SYSTEM MODELS

CONTENTS OF THIS CHAPTER ARE BASED ON

Martha Maria Frysztacki, Gereon Recht, and Tom Brown. "A comparison of clustering methods for the spatial reduction of renewable electricity optimisation models of Europe." In: *Energy Informatics* 5.4 (2022). ISSN: 2520-8942. DOI: https://doi.org/10.1186/s42162-022-00187-7

4.1 INTRODUCTION

In this Chapter we dedicate towards improving the representation of spatially aggregated ESM. The overall aim is to develop clustering methods such that the result obtained from a spatially clustered model as introduced in Chapter 2 can approximate the solution of an highly-resolved model with an error in the optimal solution as small as possible. To this end, we focus on improving existing clustering methods from the literature and introduce novel ones by incorporating the lessons from Chapter 3 into suitable distance measures and choosing appropriate clustering models. The choice of clustering model is mainly driven by options to account for the network topology and electrical connectivity of substations. The distance (or similarity) measures are mainly chosen such that the spatio-temporal availability of VRE resources is incorporated into the optimisation considerations.

To compare the performance of the novel methods, we compare results obtained from the clustered models based on the different aggregation algorithms to each other and to established reduction methods from the literature. The obtained low-resolution optimal solutions are benchmarked against the optimal solution obtained from a spatially higher-resolved simulation. We explicitly devote significant attention towards an accurate power flow and the siting of renewable capacities and flexibility options, particularly solar PV, onshore and offshore wind, battery storage and hydrogen storage.

state of the art In the recent ESM literature, suggested solutions to spatially reduce high-resolution models to smaller equiv-

alents include different techniques that focus on individual features of the system. These solutions can essentially be categorised by whether they focus on (i) representing the network or (ii) the variability of renewable resources.

(i) Approaches that focus on the network representation and therefore on accurately approximating power flows include the following methods: the Ward equivalent [113], a hybrid method consisting of k-means and an evolutionary algorithm [102], clustering into zones based on the similarity of the PTDFs [107], k-medoids operating on a combination of electrical parameters of the grid as well as their geographical length [22], spectral partitioning taking into account the available transfer capability (ATC) [88] or density-based hierarchical clustering operating on the lines reactance [103]. All these approaches use distance or similarity measures on electrical parameters of the transmission grid, often referred to as electrical distance. These methods are designed for a good approximation of power flows and are mostly evaluated by comparing the power flows of a highly resolved model to power flows in a reduced model without changing the generation fleet. However, power flows are strongly impacted when moving away from conventional generation to other resources as shown in [114] (a study conducted with the Ward equivalent) for the example of switching from coal fired electricity generation to natural gas. Therefore it remains unclear if these methods are applicable when moving towards high shares of renewables as they are not designed to precisely approximate the spatio-temporal variability of wind and solar. This is especially true for models where the final installed capacity as well as its spatial distribution is subject to optimisation, such that no a-priori estimate of power flows can be made.

(ii) On the other hand, techniques that focus on an accurate representation of renewables include hierarchical clustering applied on a database of electricity demand, conventional generation and renewable profiles [91], max-p-regions applied on a database of wind and solar potential [115] or a combination of k-means++ with the max-p-regions algorithm applied separately on the FLH of wind, solar and electricity demand [89]. [116] proposes a novel screening routine that identifies relevant generation sites to be passed to a capacity expansion problem. All these methods include either a synthesised transmission grid or one at very low resolution. The downside of such approaches is that transmission bottlenecks within large regions cannot be identified, and therefore power flows within large regions are not considered at all. By ignoring the grid, transmission congestion within these large regions could hinder the exploitation of the best available resource sites and results of these models may not be feasible in reality.

In the most recent literature, hybrid approaches have been developed where both the representation of the grid as well as the representation of renewables was taken into account in the clustering. This was implemented by applying k-means on two stages: first on the renewable generation sites and second on the network nodes, such that every node can access multiple generation sites [2]. However, location-wise clustering has no inevitable correlation with either the transmission grid nor the renewable resources, hence requiring relatively large spatial resolutions to yield good results. Furthermore, using k-means on locations ignores the connectivity of the grid, and could end up aggregating two nodes that were previously disconnected, resulting in strong distortion of the network representation. In a second hybrid approach, the network was clustered using k-medoids equipped with an additional spatial contiguity constrained introduced by [117]. Here, on the first stage, the cumulative distance of all parameters of the network were considered in a distance measure (all 1 and 2-dimensional regional attributes and all 2-dimensional connection attributes between regions indicating how strongly they are connected). On the second stage the number of generation technologies within each region were clustered using a hierarchical clustering based on the similarity in their temporal profiles [118]. However, embedding all parameters of the network in the first stage does not highlight which ones are most important in terms of the network representation. Further, renewable profiles are double-weighted as they are included in the distance measure on both stages. Therefore, the question concerning the choice of an appropriate metric remains unresolved.

research contributions We focus on improvements in the clustering process to capture both renewable generation accurately as well as the transmission network. We then evaluate the proposed methods on both the representation of renewable generation and power flows. For this task we define metrics to determine if good renewable generation sites after the clustering are maintained while incorporating the electricity grid by aggregating only nodes connected by an existing transmission line using Ward's method, an hierarchical agglomerative clustering (HAC) model. This focus on the combination of renewable sites and the transmission network for the clustering is completely novel in the context of ESM. For this method, we distinguish between two features: The aggregated quantity of FLH (similar to [89] with the adaptation of incorporating the grid), and the full time series of renewables (a novel metric employed in the context of ESM). Using the full time series to define regions is motivated by the fact that regions with similar capacity factors may have very differ-

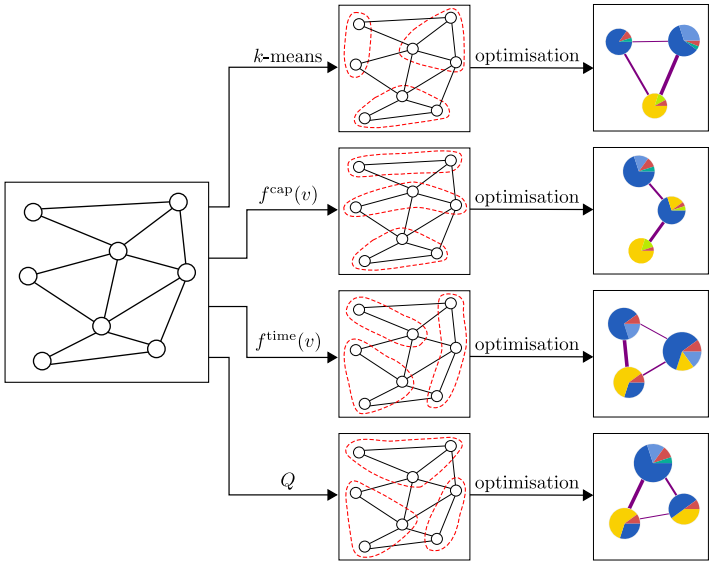


Figure 19: Graphical representation of the workflow of this Chapter (left to right).

ent time profiles, depending on how their generation is correlated over space; by using the full time series, we avoid aggregating sites with very different profiles.

We compare all results obtained by the same model using the same experimental setup and input data. This increases transparency and guarantees that differences in the results occur because of the clustering process, not because of differences in the data, the model formulation or other parameters of the model. To complete the comparison, we include two common reduction methods from the literature: k-means clustering on the coordinates of the network and clustering based on country borders.

remainder In Section 4.2 we introduce a novel pre-aggregation method of the data-set introduced in Chapter 2, that operates on a subset of network nodes and reduces the initial network size by a factor of approximately two. Then we present the application of the following clustering methods to energy system models for further model reduction: k-means, a benchmark clustering technique based on the coordinates of the network nodes that was used in several publications in the past (see also Chapter 3); Ward's method, for which we adapt the metric to a time-aggregated annual and on a time-resolved hourly feature of the system; and Modularity Maximisation, that involves considerations of electrical parameters of the model.

Results of the presented methods are presented in Chapter 4.3 and are divided into two main categories: In an a-priori analysis we show resulting regions obtained from the presented cluster-

ing algorithms before solving the optimisation problem. In an a-posteriori analysis we show the convergence of each method in a capacity expansion brownfield approach under a 60% and 100% CO_2 emission cap.

We draw conclusions in Section 4.4.

A visualisation of the outline is provided in Figure 19 using the abbreviations of Table 5. Table 5 additionally outlines the novelty of every proposed method.

4.2 METHODS

This study is performed using the ESM PyPSA-Eur [51], which is explained in in Chapter 2, particularly Sections 2.1, where we provide an overview of the underlying data, Section 2.2, where we introduce notation and finally Section 2.3 where we define the full optimisation problem.

4.2.1 *Clustering Methodology*

pre-aggregation Before applying a clustering method on the nodes of the model, several preparation steps are conducted to simplify the process. First, all lines are mapped to the voltage level of 380 kV, the prevalent level of the European transmission system. Second, all one-valent nodes are aggregated to their unique neighbors. This has only a weak effect on renewable generators because of the small cluster sizes, and power flows are not affected strongly because there is only one way for the power to flow from or to one-valent nodes. These two steps were described in higher detail in Chapter 3, Section 3.2.

In a final - and novel - pre-aggregation step, a shortest-path problem is solved using Dijkstra's algorithm $D((\mathcal{V}, E), l : E \to \mathbb{R}_0^+)$ on the nodes that are not substations (i.e. $lv_v = 0$, see Section 2.2 for Notation). Such nodes have no electricity demand, storage units or generators attached. Hence, the same amount of power that flows into such node must flow out as well, due to the fact that no power can be absorbed or generated. This follows directly from KCL (see equation (15)). Therefore, neither the power flows nor the generating assets are affected significantly when aggregating them to their electrically closest substations.

All these initial steps reduce the network by approximately a factor of 2 to 2435 nodes, 3673 HVAC and 42 HVDC lines. To further reduce the network size down to a desired number of clusters $37 \leqslant K \leqslant 2435$, a clustering method is applied. The lower bound represents the 37 countries and synchronous zones covered by the model. We therefore divide K into 37 integer summands K_z, each

representing the number of nodes within a unique associated synchronous zone or country. The clustering methods are respectively applied within each "country-zone" z. A formula to mathematically derive the summands K_z is provided in equation (18) in Section 3.2.

The lowest model resolution of 37 nodes represents the benchmark clustering method where every political region (here: every country-zone) is represented by one single node, regardless of the applied clustering method. Thus, each of the methods has the same properties at lowest resolution. When increasing the network resolution beyond 37 nodes, model results start to converge towards the solution at full resolution, where all the methods again yield the same solution, because no clustering is applied. At a resolution of 1250 nodes, the solutions of all discussed methods have sufficiently converged and are therefore taken as a benchmark to compare the low resolution solutions to. A detailed discussion on why 1250 nodes are a sufficient benchmark is conducted in Annex C.1.

k-means clustering Mathematical and technical details for the k-means clustering are already provided in Section 3.2. Note, that our k-means formulation of equation (19) is weighted with a nodal weighting w_v given in (20)-(21) to account for the original layout of the transmission grid to connect major generators and major electricity consumption (see Chapter 1, particularly Figure 1).

The average complexity of k-means is $\mathcal{O}(K|\mathcal{V}|i)$, with number of iterations i. In the worst case, $i = 2^{\Omega\left(\sqrt{|\mathcal{V}|}\right)}$, resulting in a super-polynomial complexity [119].

A drawback of k-means is that it is not possible to enforce a strict connectivity constraint based on the transmission grid. For example, two nodes that are close in space but not electrically connected can be aggregated to a single node, which can have a significant distorting effect on the power flows. Therefore, the other clustering methods are HAC models because it incorporates a connectivity constraint while clustering based a given feature of the data.

ward's method HAC is a bottom-up approach, initially treating each node as a singleton cluster. In each iteration two adjacent clusters are aggregated that have the most similar feature(s) with respect to a given similarity measure. Then, the aggregated cluster's feature is updated. HAC has greedy characteristics, as after the aggregation of the best suited clusters the decision is permanent, and has a running time of $\mathcal{O}(|\mathcal{V}|^2 \log^2 |\mathcal{V}|)$ [120].

As a distance measure we invoke a variance-minimising approach, similar to k-means. The distance $d : \mathcal{V} \times \mathcal{V} \to \mathbb{R}_0^+$ between two clusters $\mathcal{V}_c$ and $\mathcal{V}_d$ states how much the sum of squares will increase when merging:

$$d(\mathcal{V}_c, \mathcal{V}_d) = \frac{|\mathcal{V}_c||\mathcal{V}_d|}{|\mathcal{V}_c| + |\mathcal{V}_d|} \|\bar{\mathcal{V}}_c - \bar{\mathcal{V}}_d\|_2^2$$

where

$$\bar{\mathcal{V}}_c = \frac{1}{|\mathcal{V}_c|} \sum_{v \in \mathcal{V}_c} f(v)$$

with $f : \mathcal{V} \to \mathbb{R}^n$ being the feature of a node that can be of arbitrary dimension $n \in \mathbb{N}$. This choice of similarity measure is also known as *Ward's method* [121]. Recall that initially each node is treated as a single cluster, hence in the first iteration the distance between two nodes is

$$\frac{1}{2}\|f(v) - f(w)\|^2 \tag{24}$$

In this work, we consider two related, yet different features of the network: The renewable annual capacity factors $\bar{g}_{v,s}$ and the time-series $\bar{g}_{v,s,t}$ of each node, that we briefly present in the following and visualise in Figure 20.

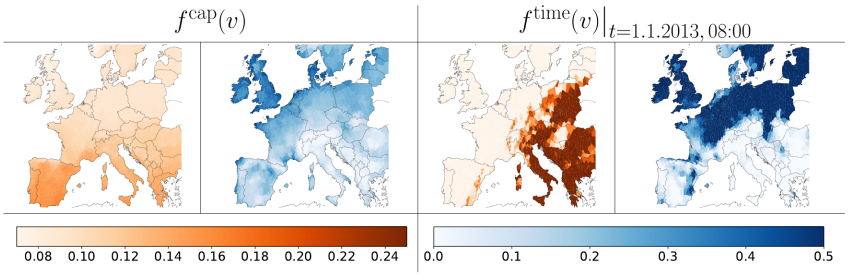


Figure 20: Examples of the different features for solar (orange) and wind (blue). All illustrations show a model resolution of 2435 nodes.

CAPACITY FACTOR AGGREGATION The annual capacity factor $\bar{g}_{v,s} \in [0, 1]$ is a unit-less ratio of the average actual energy output of an asset over its nominal capacity, i.e.

$$\bar{g}_{v,s} = \frac{\langle g_{v,s,t} \rangle_t}{G_{v,s}}$$

where $g_{v,s,t}$ is the energy dispatch of asset s in node v at time t. The average is taken over one year, as the name already suggests.

For Ward's method, we refer to the feature f in this case as f^{cap}, and define it as

$$f^{cap}(v) := \bar{g}_{v,s\in\{\text{solar, wind}\}} = \begin{pmatrix} \bar{g}_{v,\text{solar}} \\ \bar{g}_{v,\text{wind}} \end{pmatrix} \in [0,1]^2 \tag{25}$$

TIME SERIES AGGREGATION Resolved capacity factors in time form a series, in this case with a two-hourly resolution over an historical weather year. Without averaging the feed-in over the year, the variability of renewables is evident. For example, the energy production of a solar panel at night is typically zero, while during day time the power output is positive. While the annual capacity factor averages this fact and remains strictly positive for every region, a highly resolved time series captures fluctuations. Thus, additionally to a north-south gradient of the annual capacity factor for solar (higher irradiation in the south) an east-west gradient can be captured (day-night variation). In general, it holds $\bar{g}_{v,s} = \langle \bar{g}_{v,s,t} \rangle_t \ \forall v \in \mathcal{V}, \forall s \in \mathcal{S}^{re}$.

The feature f for Ward's method in this case is of high dimension, as every resolved time step has to be considered. We refer to it as f^{time} and define it via

$$\begin{aligned} f^{time}(v) :&= \bar{g}_{v,s\in\{\text{solar, wind}\},\, t\in\mathcal{T}} \\ &= \begin{pmatrix} \bar{g}_{v,\text{solar},1} \\ \cdots \\ \bar{g}_{v,\text{solar},|\mathcal{T}|} \\ \bar{g}_{v,\text{wind},1} \\ \cdots \\ \bar{g}_{v,\text{wind},|\mathcal{T}|} \end{pmatrix} \in [0,1]^{2|\mathcal{T}|} \end{aligned} \tag{26}$$

In this study, $|\mathcal{T}| = \frac{1}{2} \cdot 8760$, because we resolve our model with a temporal resolution of two hours and run the optimisation over one year (2013).

It is no curse of dimensionality to apply $f^{time}(v)$, because we solely measure the (high-dimensional) distance between two points; but we do not sample from this high-dimensional space to approximate it with insufficiently many data points.

clauset-newman-moore greedy modularity maximisation

The modularity maximisation approach aims to find community structures in large networks. It is a HAC method with approxi-

mately linear running time, $\mathcal{O}(|\mathcal{V}|\log^2|\mathcal{V}|)$ [122]. In each iteration, it greedily aggregates the two nodes v and w that increase modularity Q the most and continues to do so until the desired number of clusters is reached or until Q can not be further improved.

Q is formally defined as

$$Q = \frac{1}{2m} \sum_{v,w \in \mathcal{V}} \left(\mathcal{A}_{v,w} - \frac{k_v k_w}{2m} \right) \delta(c_v, c_w) \tag{27}$$

with the weighted adjacency matrix $\mathcal{A}_{v,w}$ of the network graph $\mathcal{G}$

$$\mathcal{A}_{v,w} := \begin{cases} w_{(v,w)} & \text{if } (v,w) \in E \\ 0 & \text{otherwise} \end{cases}$$

the sum of all edge weights in the graph m

$$m := \frac{1}{2} \sum_{v,w} \mathcal{A}_{v,w}$$

the weighted degree of node v k_v

$$k_v := \sum_{w} \mathcal{A}_{v,w}$$

and the Kronecker-Delta function, given as

$$\delta(c_v, c_w) := \begin{cases} 1 & \text{if} \quad c_v = c_w \\ 0 & \text{otherwise} \end{cases}$$

From the formal definition of modularity in equation (27), we can deduce that the sum in Q is only non-zero, if v and w belong to the same cluster. In its original publication [122], modularity was introduced without weights, i.e. $w_{(v,w)} = 1$, but we choose a different weighting to adapt the method to the network topology, similar to the suggestion in [103], but accounting for both the reactive and resistive components of the grid. We choose the absolute value of the admittance $|y_{(v,w)}|$ of each line (v,w), a measure of electrical distance that describes how easily a circuit allows power to flow. Admittance is defined as the inverse impedance, i.e.

$$y_{(v,w)} = \frac{1}{z_{(v,w)}} \tag{28}$$

Regarding the values of Q, $A_{v,w}$ is large and positive for a good division, i.e. when aggregating electrically close nodes v and w, and small or zero for a bad division, i.e. when the impedance is

$\mathbf{G_1 : A_{0,2} \approx 0.067 > \frac{k_0 k_2}{2m} \approx 0.022}$	$G_2 : A_{0,1} = 0.05 > \frac{k_0 k_1}{2m} \approx 0.018$
$G_3 : A_{1,5} = 0 < \frac{k_1 k_5}{2m} \approx 0.018$	$G_4 : A_{2,3} = 0.005 < \frac{k_2 k_3}{2m} \approx 0.026$

Table 4: Exemplary inital iterations of the Clauset-Newman-Moore Greedy Modularity Maximisation Algorithm associated with Figure 21.

high, or if the nodes are not connected at all. $\frac{k_v k_w}{2m}$ is a measure of (electrical) centrality: it tells us, how well the nodes v and w are interconnected in the graph, independent of each other. If the value is large, v and w are nodes with either many connections or they are connected by lines with low impedance. A small value indicates a sparse connection, i.e. either few edges or connections with high impedance. Thus, a (large) positive value of the difference $A_{v,w} - \frac{k_v k_w}{2m}$ marks v and w to be electrically closer than they are on average from other nodes in the network. Their aggregation therefore suggests a good grouping.

Consider an exemplary symmetric graph $\mathcal{G}_0$, such as given in Figure 21. All edges are weighted with a line reactance $x_{(v,w)}$ and without resistance. The four panels of Figure 21 show different first iteration choices of the weighted Clauset-Newman-Moore Algorithm into graphs $\mathcal{G}_i$, $i \in \{1,2,3,4\}$ marked in red. Due to the symmetry of G_0, other than the displayed choices for the fist iteration are equivalent. Without clustering, each node in G_0 can be interpreted as a singleton cluster, yielding the initial modularity of $Q_0 \approx -0.1677$. For the four displayed cases, we can calculate the increase of modularity if the marked nodes were merged, see Table 4. Hence, both G_1 and G_2 would improve the modularity, but G_1 is the better choice, as $A_{0,2} - \frac{k_0 k_2}{2m} > A_{0,1} - \frac{k_0 k_1}{2m}$. G_3 and G_4 are bad choices, reducing modularity and deteriorate the network community. However, if $x_{(2,3)}$ was much smaller, for example $x_{(2,3)} = 1$, then G_4 would be the best choice for the first iteration.

4.2.2 *Overview of Studied Clustering Algorithms*

We summarise the metrics and methods that are subject to our analysis in Table 5.

4.2.3 *Copperplate Aggregation*

After mapping every node v to a cluster $\mathcal{V}_c$, i.e. $v \mapsto \mathcal{V}_c$, all nodes within $\mathcal{V}_c$ are replaced by a single equivalent node, where the attributes of all nodes within $\mathcal{V}_c$ are aggregated to one equivalent. For example, demand and generation potentials are summed up,

Abbrev.	Clustering on ...	Novelty
'country-zones'	... political borders and synchronous zones.	benchmark (no novelty). The spatial resolution of 37 nodes is not variable and the lower bound for all other presented methods.
k-means	... geographic locations (coordinates) of graph nodes. Formulated in eq. (19)	Pre-Aggregation to substations (Dijkstra); otherwise benchmark (no novelty)
$f^{cap}(v)$	... annual capacity factors of nodes. Formulated in eq. (24) and (25). Hierarchical clustering.	Pre-Aggregation to substations (Dijkstra); thereafter similar to [89] with the following differences: considers network topology using HAC, simultaneous consideration of wind and solar capacity factors in each aggregation step, varying spatial resolution (27 nodes in [89])
$f^{time}(v)$	... hourly capacity factors (time-series) of nodes. Formulated in eq. (24) and (26). Hierarchical clustering.	Fully novel in the context of ESM.
Q	... electrical distance between two nodes. Formulated in eq. (24) and (27). Hierarchical clustering.	Pre-Aggregation to substations (Dijkstra); thereafter similar to [103], with the difference of accounting for both reactive and resistive parts of transmission lines and considering whole Europe not only Germany to make results comparable.

Table 5: Abbreviations and novelty declaration for the applied clustering methods. Each one is discussed in the methods Section 6.2.

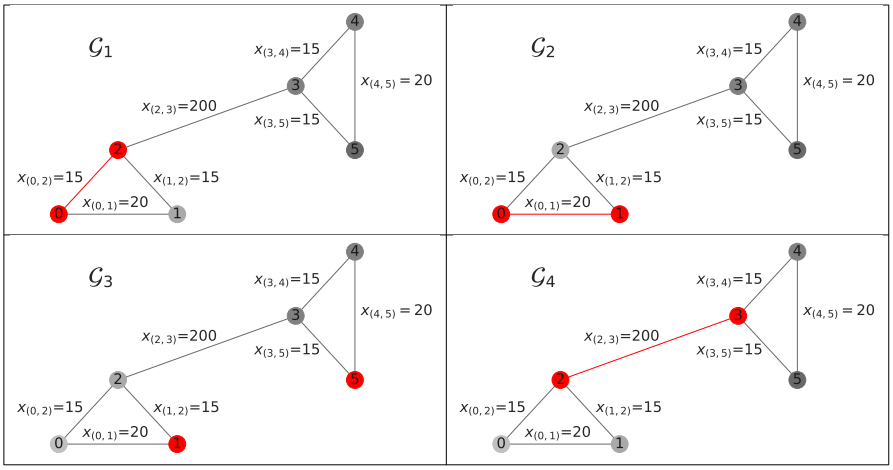


Figure 21: Example graph to illustrate the concept of the Clauset-Newman-Moore Greedy Modularity Maximisation Algorithm.

and capacity factors are averaged. This replacement is referred to as copperplate approach because it is equivalent to all nodes inside $\mathcal{V}_c$ being connected to a lossless copper plate. Finally, all lines (v, w) that connect nodes within the same cluster, i.e. $v, w \in \mathcal{V}_c$, are removed from the model, while lines connecting nodes in different clusters, i.e. $v \in \mathcal{V}_c \wedge w \in \mathcal{V}_d$ where $c \neq d$, are aggregated to an equivalent line.

4.2.4 *Capacity expansion problem*

We perform a brownfield capacity optimisation that builds on a system that exists as of 2018 for both the generating fleet according to the data-set introduced in Chapter 2, particularly the existing renewable generator fleet of [50], and the planned transmission grid in the TYNDP 2018 [23]. The optimisation is subject to two decarbonisation goals of 60% and 100% lower emissions compared to 1990. Missing capacities of renewables for the system to be feasible with respect to the decarbonisation goals are optimised in the sense that the total system costs are minimised.

4.3 RESULTS

4.3.1 *Evaluation of the Regions*

First of all we present resulting regions in Figure 22 for an exemplary spatial resolution of 67 nodes. Additionally, the ranges of cluster sizes are shown in Figure 23, displaying how many nodes

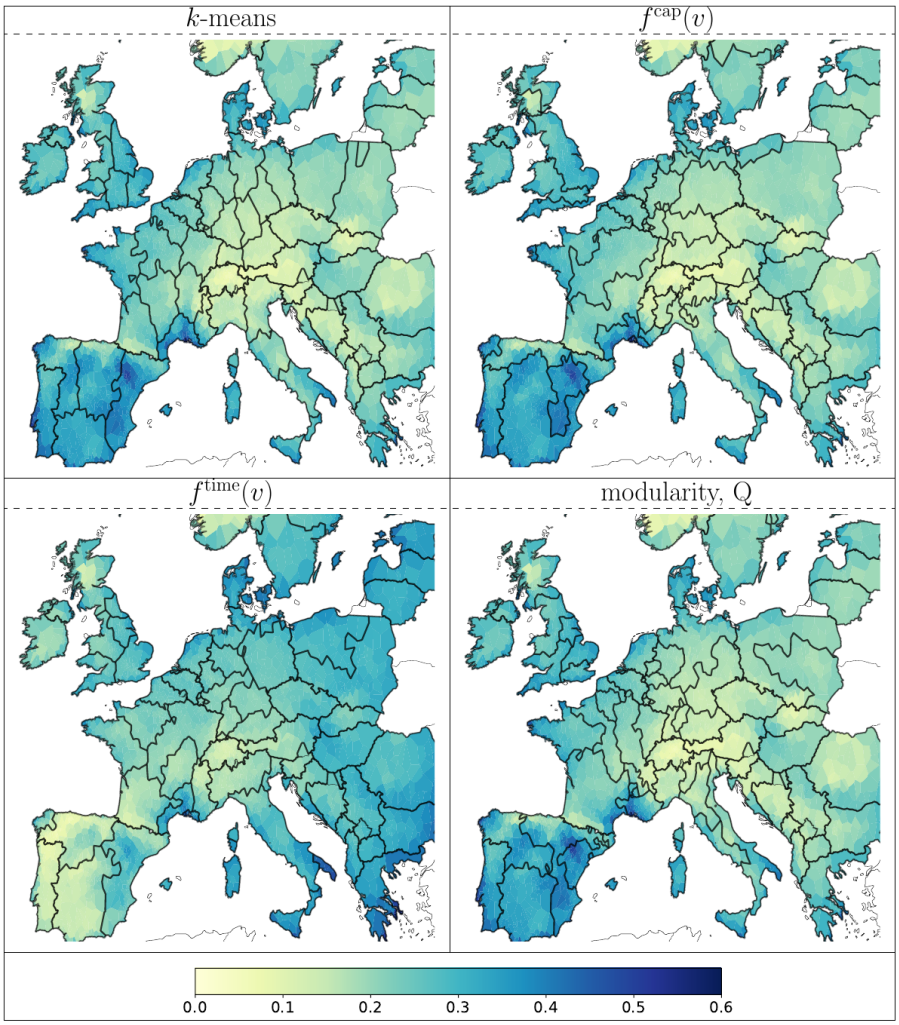


Figure 22: Resulting regions respective clustering method discussed in this Chapter at a resolution of 67 nodes.

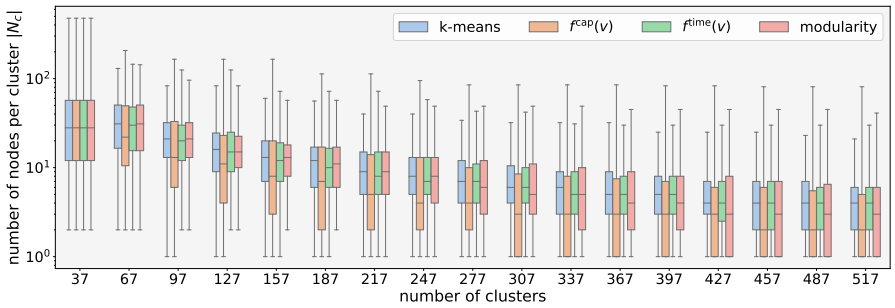


Figure 23: Resulting distribution of cluster-sizes respective clustering method as a function of the number of clusters.

were aggregated into one cluster for varying numbers of clusters in steps of 30. Results on the community structure, i.e. modularity Q given in equation (27), are shown in Figure 24 for all possible model resolutions starting at 37 and up to 2435 nodes.

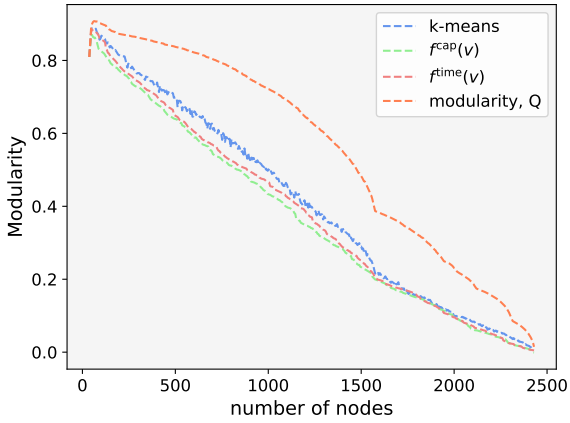


Figure 24: Resulting modularity respective clustering method as a function of the number of clusters.

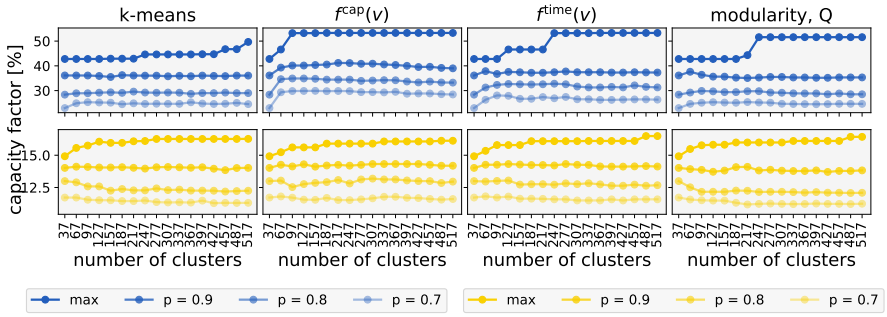


Figure 25: Resulting quantiles of capacity factors for wind (top) and solar (bottom) respective clustering method as a function of the number of clusters.

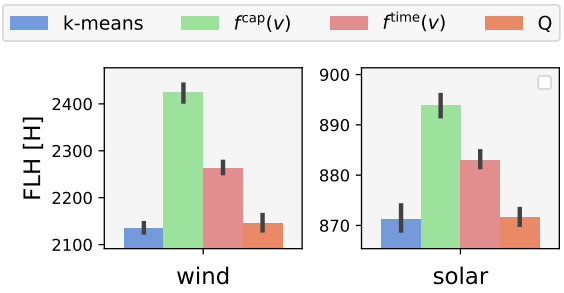


Figure 26: Average FLH of existing wind (left) and solar (right) assets according to equation (29) respective clustering method.

Capacity factors are evaluated in Figure 25 on a quantile base, because the optimisation places renewable capacity at its best available site whenever possible as more power can be generated there with the same cost penalty in the objective function, according to constraint (9). For a second perspective, we also present the average FLH of the renewable assets installed by 2018 in Figure 26:

$$\sum_{t\in\mathcal{T}} \left\langle \bar{g}_{v,s,t}\big|_{G^{2018}_{v,s}>0} \right\rangle_v \qquad s \in \mathcal{S}^{re} \tag{29}$$

4.3.2 *Evaluation of the Capacity Expansion model*

The main objective of applying different clustering techniques to the model is to reduce the model size for it to be computationally tractable. But at the same time, we want to obtain good estimates for all the optimization variables introduced in Chapter 2 (particularly Section 2.3), especially those of the model formulation in equation (3). It is desired that the low-resolved results (estimates) resemble the highly-resolved modelling results. For the power flow this means that the sum of flows f of highly-resolved lines (v, w) that are aggregated to the new line (c, d)

$$f_{(c,d),t} := \sum_{\substack{(v,w)\in E:\\ v\in\mathcal{V}_c \wedge w\in\mathcal{V}_d}} f_{(v,w),t} \quad \forall c, d \in \mathcal{V}^K,\ t \in \mathcal{T},\ K \in \mathbb{N}_{\geqslant 37}$$

in the low-resolved model (estimate $\hat{f}$) is the same:

$$\hat{f}_{(c,d),t} \overset{!}{=} f_{(c,d),t} \quad \forall c, d \in \mathcal{V}^K,\ t \in \mathcal{T},\ K \in \mathbb{N}_{\geqslant 37} \tag{30}$$

Note that we slightly deviate the notation in this Section from the remaining dissertation, where we do not interpret the low-resolved modelling results as estimates and do not mark them with a hat $\hat{f}$.

We proceed in a similar manner in case of the generation and storage capacities. The sum of optimised capacities G at nodes within a cluster $v \in \mathcal{V}_c$ should equal the one at the clustered node c at the low-resolved model (here, interpreted as estimated value $\hat{G}$):

$$\hat{G}_{c,s} \stackrel{!}{=} G_{c,s} := \sum_{v \in \mathcal{V}_c} G_{v,s} \quad \forall c \in \mathcal{V}^K,\, s \in \mathcal{S}^{re},\, K \in \mathbb{N}_{\geqslant 37} \tag{31}$$

The same is desired for the dispatch (or charging/discharging) of all generating (or storing) assets. But we restrict the discussion to those two samples of optimisation variables, as there exists a strong correlation between generation and capacity. This is because exploiting installed resources whenever possible is cheapest according to constraint (9) due to the low operational costs for renewables, $o_{v,s} \approx 0$.

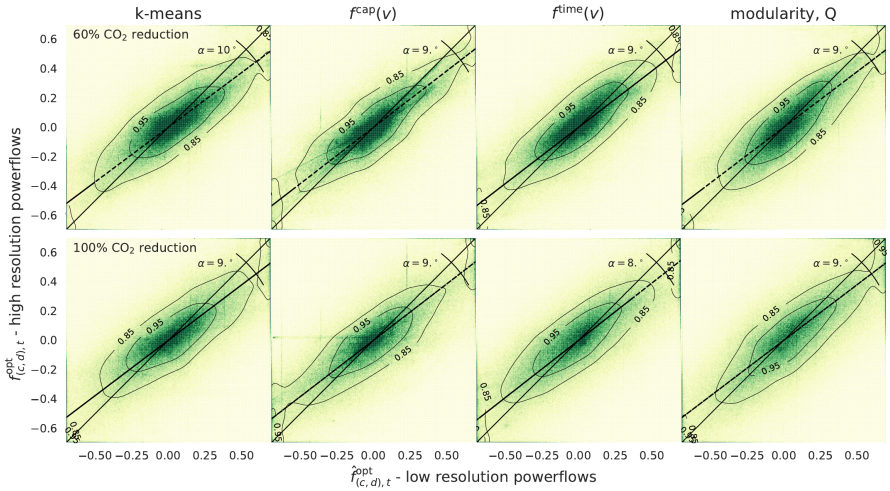


Figure 27: Normalised mapping of optimal power flows, additionally displaying the 95% and 85% percentiles of the corresponding probability density function (PDF) using contour plots.

As the model cannot be solved at full resolution for any of the clustering methods, the highly-resolved optimised capacities $G_{v,s}$ and power flows $f_{(v,w),t}$ are taken from a model resolution of 1250 nodes (see Annex C.1 for a justification why this benchmark resolution is sufficient), and the estimator quantities $\hat{G}_{c,s}$ and $\hat{f}_{(v,w),t}$ from a model with 97 nodes, the same resolution as in Figure 25 for the capacity factors. Analysing model results at the spatial resolution of 97 is because many studies choose a resolution of approximately 100 nodes for their research, such as the final report of the e-Highway 2050 project [22]. The mappings of optimal

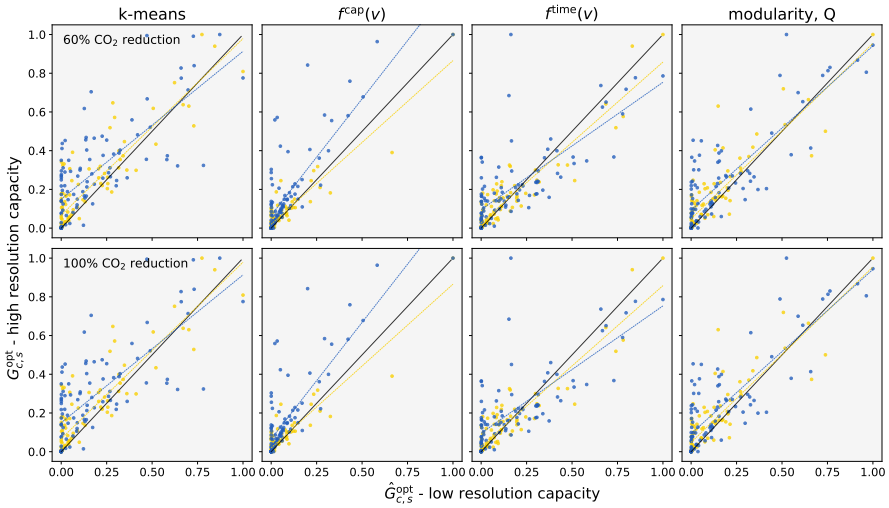


Figure 28: Normalised mapping of optimal capacities for wind (blue) and solar (yellow) including a linear fit to the respective data.

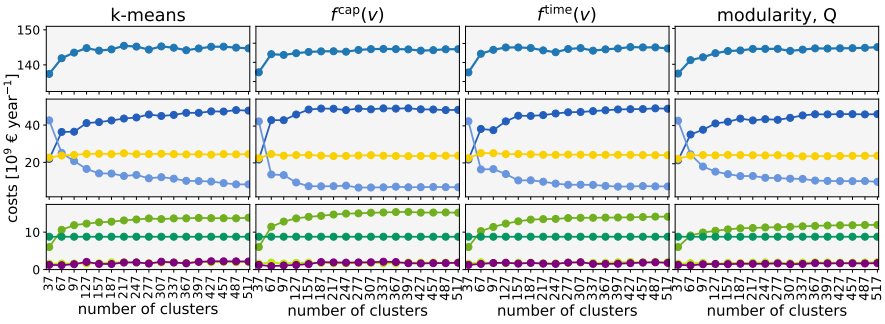
power flows in equation (30) and optimal capacities in equation (31) are shown in Figures 27 and 28 respectively.

Finally, Figures 29a and 29b display the resulting objective of the optimisation in equation (3) for the two considered carbon reduction targets of 60% and 100% for different model resolutions in steps of 30 nodes up to a model resolution of 397 nodes.

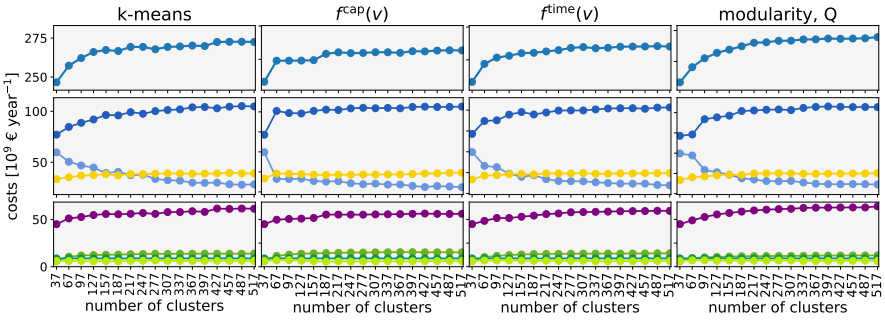
discussion of the results

4.3.3 *Discussion on the resulting clusters*

In Figure 22 it can be seen that k-means clustering and Ward's method with hourly capacity factors ($f^{time}(v)$) favor regions with similar size. For k-means this results from the objective to minimise geographical distances, see eq. (19), while for $f^{time}(v)$ the reason are spatially and temporally varying features that favor this outcome: There exists a north-south gradient for the solar capacity factors that constrains the clusters vertically, and the day-night variation of solar irradiation prevents clusters from being elongated from east to west by adding a high penalty in equation (24) when trying to merge "day"-nodes with "night"-nodes. This is visible in the east-west elongated clusters as well as large coastal regions that can be observed for $f^{cap}(v)$ because the annual capacity factors do not see these temporal variations. The cluster structure related to modularity Q is similar to $f^{cap}(v)$, resulting in some very small or elongated thin clusters. The structure of clusters continues so for higher resolutions, as can be seen in Figure 23. The outcome of long thin clusters is common for single-linkage



(a) Total system costs for the 60%-CO_2-reduction scenario



(b) [Total system costs for the 100%-CO_2-reduction scenario

Figure 29: Resulting total system costs for the two CO_2 scenarios respective clustering method.

HAC methods [123], but can be overcome with a profound choice of feature, as we demonstrate using $f^{time}(v)$.

In Figure 25 it can be seen that every of the three presented HAC techniques with different similarity/distance measures can capture annual capacity factors better than k-means clustering with respect to the same model resolution. Applying HAC with the similarity measure $f = f^{cap}(v)$ finds and maintains the best generation sites with an annual capacity factor of 53.3% for wind at a model resolution of 97 nodes. The competing clustering techniques are behind: The best generation site is reached at a model resolution of 247 when invoking hourly capacity factors $f^{time}(v)$ or modularity Q as a similarity or distance measure and 517 nodes for k-means. For Q as well as for k-means, the best generation site has a lower annual capacity factor of only 51.6% and 49.7% respectively. However, when siting solar assets, the behavior is different: The best site is available earliest for $f = f^{time}(v)$ and $f = Q$ for a model resolution of 487 nodes and capacity factors of 16.48% and 16.42%. Both k-means and $f = f^{cap}(v)$ perform worse, with lower capacity factors even at a model resolution of 512 nodes. This reflects also in the full load hours of existing assets $G^{2018}_{v,s}$ of the respective clustering methods (Figure 26).

Regarding the community structure of the resulting reduced graph, only HAC based on modularity performs significantly better than the competing methods. Although Ward's method takes into account the structure of the transmission grid by considering only adjacent neighbors, both algorithms perform slightly worse in terms of community structure than k-means, see Figure 24. Regardless of the method, when reducing the model resolution below a threshold of approximately 40 nodes, modularity suddenly drops to zero.

4.3.4 *Discussion of the Optimal Modelling Results*

First, we consider the power flow estimates introduced in equation (30). As we expect the low-resolved flow estimates $\hat{f}_{(c,d),t}$ to equal the aggregated optimal flow $f_{(c,d),t}$ derived from the highly-resolved system, the associated random variable should be distributed proportional to a two dimensional normal distribution:

$$\begin{pmatrix} \hat{f}_{(c,d),t} \\ f_{(c,d),t} \end{pmatrix} \sim \mathcal{N}_2(\mu, \Sigma)$$

where the covariance matrix Σ, and in particular the confidence ellipses provide insight of the correlation between the estimated and the optimal power flow. The length of the axes of the ellipses can be derived form the eigenvalues σ_i of the covariance matrix Σ, namely $r_i = \sqrt{\sigma_i}$. This approximately corresponds to the 40^{th} percentile, i.e. 40% of the data points lie within the ellipse and 60% outside of it. The narrower the minor axis r_2 is, the more data points are close to the origin. The major axis r_1 gives insight of the magnitude of the power flows. The larger the major axis, the more large power flows can be observed. Information on the relation between the actual power flow f and its estimate $\hat{f}$ can be gained from from the bivariate correlation coefficient ρ:

$$\rho = \frac{\Sigma_{01}}{\Sigma_{00}\Sigma_{11}} = \frac{\mathrm{COV}(\hat{f}_{(c,d),t}, f_{(c,d),t})}{\mathbb{V}(\hat{f}_{(c,d),t}) \cdot \mathbb{V}(f_{(c,d),t})} \tag{32}$$

It is a measure of correlation between the two variables. It is 1 for a perfect correlation, 0 for no correlation and -1 for a negative correlation. Resulting correlation factors ρ and minor axes r_2 can be taken from Table 6 (or Tables 17 and 18 in Annex C.3 for different low-resolved modelling results).

For the 60% carbon reduction target, $f^{\text{cap}}(\nu)$ as a similarity measure for Ward's method yields the best correlation factor ρ, but the features of annual capacity factors $f^{\text{time}}(\nu)$ and modularity Q de-

CO_2 reduction	60%		100%	
	ρ	r_2	ρ	r_2
k-means	0.746	0.165	0.755	0.175
$f^{cap}(v)$	0.769	0.160	0.768	0.173
$f^{time}(v)$	0.767	0.160	0.781	0.169
Q	0.757	0.164	0.757	0.179

Table 6: Bivariate correlation factor ρ and radius of the minor axis r_n of the PDF of power flows for each respective clustering method at a spatial resolution of 97 nodes.

viate from $f^{cap}(v)$ by only 0.26% and 1.56% respectively in terms of ρ. The minor axis of the confidence ellipses is also most narrow for the features $f^{time}(v)$ and $f^{cap}(v)$, but only 2.5% wider for Q. The distribution of k-means has an approximately 3% lower correlation factor of 0.746 and a 3.13% wider spread in terms of the minor axis of the confidence ellipse compared to $f^{cap}(v)$ and $f^{time}(v)$. These variations are clearly visible in Figure 27. With a higher carbon reduction target of 100%, the trend that clustering on siting capacity prevails over electrical distance when considering the power flow estimates. $f^{time}(v)$ yields a 3% better correlation ρ and $f^{cap}(v)$ a 1.45% better one than Q. In terms of the spread (r_2) the same can be observed: The distribution of power flows of $f^{time}(v)$ is 5.59% narrower than for Q and f^{cap}s distribution is 3.35% slimer than the one of Q. k-means performs similar as in the 60% reduction target. To make sure these results are not artificial for the resolution of 97 nodes, we provide the same table for a spatial resolution of 67 and 127 nodes in Annex C.3. These results are in-line with the ones we found for a resolution of 97 nodes, where it becomes even more evident that the error made by k-means is much larger than the one made by the competing methods.

Considering the mapping of optimal capacity according to equation (31), we could pursue the same approach as for the power flows, i.e. assuming a normal distribution, but due to the relatively low amount of data points, such an analysis would be inaccurate. Instead, we provide the mean squared error (MSE) in Table 7:

$$\text{MSE} = \frac{1}{K}\sum_{c=1}^{K}\left(G_{c,s} - \hat{G}_{c,s}\right)^2$$

CO_2 reduction	60%		100%	
technology	wind	solar	wind	solar
k-means	$0.37 + 3.82$	$0.01 + 2.80$	$0.51 + 3.33$	$0.12 + 1.23$
$f^{cap}(v)$	$0.21 + 0.60$	$0.03 + 1.00$	$0.01 + 2.22$	$0.11 + 0.15$
$f^{time}(v)$	$0.04 + 3.17$	$0.08 + 0.79$	$0.55 + 1.94$	$0.26 + 0.28$
Q	$0.36 + 1.31$	$0.47 + 1.17$	$0.25 + 1.98$	$0.17 + 0.78$

Table 7: MSE presented as a sum of over- and underestimated optimal estimates ($MSE^+ + MSE^-$) for a spatial resolution of 97 nodes.

We distinguish between the over- and underestimated optimal capacities to be able to make better judgement which clustering method is more conservative than another; i.e.

$$MSE = MSE^+ + MSE^- \tag{33}$$

where MSE^+ is defined to capture the error of all *overestimated* capacities, i.e. those of $\mathcal{K}^+ := \{c \in \{1, \dots, K\} \text{ s.t. } \hat{G}_{c,s} > G_{c,s}\}$

$$MSE^+ := \left(\frac{1}{|\mathcal{K}^+|} \sum_{c \in \mathcal{K}^+} \left(G_{c,s} - \hat{G}_{c,s} \right)^2 \right)$$

and analogously MSE^- is defined to capture the error of all *underestimated* capacities, i.e. those of $\mathcal{K}^- := \{c \in \{1, \dots, K\} \text{ s.t. } \hat{G}_{c,s} < G_{c,s}\}$

$$MSE^- := \left(\frac{1}{|\mathcal{K}^-|} \sum_{c \in \mathcal{K}^-} \left(G_{c,s} - \hat{G}_{c,s} \right)^2 \right)$$

While the clustered models tend to underestimate the need of renewable generation and storage capacity ($MSE^+ \ll MSE^-$) for any of the clustering methods, according to the resulting values presented in Table 7 and Figure 28, clustering based on f^{cap} performs best in the optimal placement of simultaneously placing wind and solar assets for every carbon reduction target. However, the methods of f^{time} and Q are not significantly worse and yield errors in the same order of magnitude. On the other hand, k-means performs significantly worse with an at least $0.21 - 2.39$ times higher MSE^- value compared to the competing methods. To make sure these results are not artificial for the resolution of 97 nodes, we provide the same Table for a spatial resolution of 67 and 127 nodes in Annex C.3. Results presented there are in-line with the ones we found for a resolution of 97 nodes.

In terms of storage technologies, no clear tendency can be derived. All methods equally under- and overestimate the need for storage technology ($\mathcal{O}(\text{MSE}^+) \approx \mathcal{O}(\text{MSE}^-)$) and all clustering methods perform equally well. Values for the MSE can be found in Table 21 in Annex C.2.

Regarding the total system costs presented in Figures 29a (60% reduction of carbon emissions) and 29b (100% reduction), we can observe substantially different convergence behaviors of the investment in different technologies and the total system costs. In all of the applied methods a big swing from offshore wind at low resolution of 37 nodes ('country-zones') to onshore wind can be observed, and all methods yield similar results in terms of generation capacity at a spatial resolution of approximately 320 nodes. Ward's method applied with $f = f^{cap}(v)$ converges fastest, where the total costs do not change substantially after reaching a model resolution of 157 nodes (60% reduction) and 67 nodes (100% reduction). At the side of flexibility options, the results need higher spatial resolution than provided to reach an equilibrium as they deviate from one another even at the highest spatial resolution. For the 60% reduction target, the investment in transmission lines is highest for $f^{cap}(v)$ and almost 25% cheaper for Q, because the assets are sourced more locally where demand is high, not exploiting the good sites as they are not available for this clustering. This can be taken from Figure 28. The same trend continues for the 100% reduction target, but here, for Q, it is clearly visible that 12% more hydrogen storage is needed compared to $f^{cap}(v)$, 8% more compared to k-means and 6% more compared to $f^{time}(v)$. This is because the transmission bottlenecks are better portrayed in Q than in the competing clustering techniques, while the good generation sites are not available to cover demand. This reflects well with Figure 27.

4.4 CONCLUSIONS

From this analysis several conclusions can be drawn. First of all, the choice of spatial resolution is crucial to obtain accurate model results, particularly to the ratio and distribution of renewable carriers. A model that is based on political borders such as countries is not advisable, as important transmission bottlenecks are neglected and good generation sites of onshore carriers are underestimated. When moving towards a higher spatial resolution where each country is represented by multiple nodes, modelers should consider carefully how the aggregation is conducted. For models that consist of conventional carriers (such as coal or lignite), an accurate estimate of power flows is more important than ac-

curately portraying renewable generation sites. Therefore, in this case we suggest a model reduction based on electrical distance such as the Clauset-Newman-Moore greedy modularity maximisation. However, modelling is mostly conducted to simulate future green scenarios that have high shares of renewable energy. In this case, Ward's method applied on the full time series prevails in terms of accurate siting of capacity and in terms of a good approximation of power flows. It is advisable not to choose annual capacity factors because it ignores correlations in time and leads to very elongated clusters. This tends to underestimate transmission bottlenecks within regions and, therefore, underestimates the need of renewable capacity. Inter-regional power flows in a model where Ward's method based on the full time series was applied for equivalencing are similarly well estimated as those obtained from a reduced model based on electrical distance. For higher shares of renewables the power flow approximation of the reduced model using Ward's method on the time-series is even more precise compared to results obtained from the reduced model based on electrical distance. Therefore, when modelling a highly renewable electricity system, we recommend using a hierarchical method with a similarity measure that entails spatio-temporal features of renewables, such as the renewable time-series. Model results obtained from clustering on the geographical locations of the nodes are less accurate than those from any of the three hierarchical methods both in terms of siting renewable capacities and an accurate estimate of power flows, so we advise against using this method in future.

4.5 CRITICAL APPRAISAL

Comparing modelling results retrieved from varying spatial resolutions is a computationally challenging task, meaning that additional simplifications had to be made to the model. This includes the bottleneck that we were able to compare low model results only to those obtained from a resolution of 1250 nodes, not the full network size. If more computational power or more efficient solvers become available, the low resolved solutions could be compared to the solved network at full model resolution.

Regarding our results on the network representation, we ignore the positive impacts that dynamic line rating could impose on the ampacity of the overhead transmission grid. In our simulations, we model severe high summer weather conditions such that the results are conservative. However, the ampacity of lines can be significantly increased, which might impair on our results conducted

on the electrical distance of the network, where the cooling effects of wind are not considered in the metric (27).

On the other hand, in terms of modelling renewables and particularly offshore wind, we did not model wake effects of wind turbines such that capacity factors for offshore wind are being overestimated. This might impact the strong preference towards offshore wind, particularly for models at low spatial resolution (see Figures 29a and 29b at low model resultion, particlularly 37 nodes). Another point on the representation of renewables is modelling only one technology per carrier, for example the model can build only wind turbines at a hub-height of 80m and place solar panels oriented only south with a specific tilt-angle of 35°. More technology options could be included in such study, particularly as including more technologies per node is not driving the computational complexity. It is the total number of nodes included in the model, as shown in [2].

Further simplifications include that the optimisation is run for a single weather year, only those technologies that are considered most substantial in the energy transition are included in the model, only a single set of cost assumptions is considered and the scope of the model is limited to the electricity system. The latter lacks the coupling of different sectors such as building heating, transport and non-electric industry demand, but including them might offer additional flexibility and interactions and change the results substantially. Nevertheless, a lot of research is conducted based on electricity-only models, such that our results are still valuable. A follow-up study could consider the interactions of spatial scale under different clustering methods in sector-coupled systems.

Finally, allowing grid-expansion relaxes many of the constraints imposed by the upper bound of line-capacities (14), which in turn will effect the results of this study. However, as found in [2], grid expansion does not affect the main qualitative features of the results, but it does have the overall effect of lowering the total system costs. Nevertheless, this study could be expanded upon which clustering method captures most of the congested lines and performs best in a planning transmission expansion study.

5

INTERMEZZO: CAN SPATIALLY LOW-RESOLVED MODELS CAPTURE CURTAILMENT? A GERMAN CASE-STUDY.

Contents of this chapter are based on

Martha Frysztacki and Tom Brown. "Modeling Curtailment in Germany: How Spatial Resolution Impacts Line Congestion." In: *2020 17th International Conference on the European Energy Market (EEM)*. 2020, pp. 1–7. DOI: https://doi.org/10.1109/EEM49802.2020.9221886

5.1 INTRODUCTION

In the previous Chapters we have seen how spatial clustering affects the optimality of planning results for systems with a high share of VRE supply. These models and mathematical optimisation formulations were always subject to optimising the generation fleet and flexibility options such as the transmission grid or dimensioning storage capacity to simulate possible future renewable grid scenarios. They were not aiming to model the operation of the electricity grid or to reproduce historical events, but it is important to ensure that the models utilised for planning applications are a good approximation of the real world, particularly after applying spatial reductions to the models.

The initial release of PyPSA-Eur [51] (see Chapter 2) provides a limited model validation, mainly focusing on the network total line lengths, the grid topology, the potentials for expansion of renewables and the linear optimal power flow. It was not analysed if the model is able to reproduce the feed-in or the curtailment of the existing fleet of renewable assets. But as the share of renewable generators is increasing considerably and expected to continue to do so, we want to ensure that the model can capture these dominant effects of the system. Using an evaluated model on curtailment for future planning studies has the advantage of more reliable investment decisions, because curtailing renewable electricity results in high costs for transmission grid operators due to political incentives for the expansion of renewable energy in Germany.

background Incentives for renewable energy in Germany are regulated by the Renewable Energy Sources Act [124–127] that is a key driving force for the expansion of renewable energy. It enforces that if the feed-in of electricity from an installation classified as a renewable energy source is reduced due to a grid system bottleneck, grid system operators must compensate operators affected by their measure for 95% of the lost revenues[1] [126]. Compensation payments for the curtailed energy in 2013 were 43.7 million euros and have increased ever since, reaching a maximum of 718.7 million euros in 2018 [21, 128]. The reason is that in the last few years significant amounts of wind have been curtailed because of congestion in the German transmission system.

At the same time, the German government has set a target that the share of renewables must be increased from 40% in 2019 to 65% by 2030. By 2050 the capacities of today must at least quintuple even in optimistic scenarios to meet the ambitious CO_2 reduction targets [129]. Thus, congestion is likely to continue to be present as shares of wind and solar rise, particularly given the delays in building new transmission projects. Various solutions have been proposed to mitigate congestion: flexibility from sector-coupling [110], the production of green hydrogen [130], innovative new technologies such as dynamic line rating [131] or fast-acting storage [132] could be introduced to the market to increase available green energy from existing assets.

research contributions In this Chapter we dedicate to validating historical curtailment in the German network in PyPSA-Eur [51] (see Chapter 2). We perform this exercise to demonstrate that the model at hand is capable to represent the existing electricity grid in terms of line congestion and curtailment. Particularly as the share of VRE carriers will increase over the decades to come, openly-available, validated models should increase transparency in testing new flexibility strategies and innovation in managing congestion [133]. Therefore, accurate modelling of the interactions between renewables and the grid is critical to assess future scenarios for the energy system. Within the validation, we particularly examine the spatial representation of the model by varying its resolution to understand at what granularity the most important bottlenecks are still captured. Results are validated against published feed-in management numbers by the German Federal Network Agency [134, 135] in time and space. Note that in this case, the model optimisation is performed as a pure operational dispatch, where no new transmission, generation or storage facilities can be built.

1 according to the version of the Act that was in force during the years we evaluate in this Chapter

remainder This Chapter is arranged as follows: In Section 5.2 we present adaptations made to the full optimisation problem with respect to all its constraints to model curtailment in Germany. In Section 5.3, we present our results of curtailment with respect to different spatial resolutions of the model. Based on these findings, results are discussed on an annual scale for the years 2013-2018, a spatial scale for the four distinct TSO regions in Germany and on a temporal scale discussing results per quarter of two of the considered years where curtailment rates are highest.

5.2 METHODS

5.2.1 *Optimisation problem*

While the original model formulation of PyPSA-Eur [51, 63] as presented in Chapter 2 (particularly Section 2.3) is capable of co-optimising investment in generation and transmission, this case-study aims to reproduce historical events of the past. Therefore, the objective function minimises solely the operational costs for a fixed generation and storage fleet and fixed transmission capacities. Historic capacities are given exogenously for each year, labeled by an index α. Thus, the initial objective function given in equation (3) simplifies to

$$\min_{\substack{g_{v,s,t},\, h^{\pm}_{v,r,t},\\ f_{(v,w),t}}} \left[\sum_{v \in \mathcal{V},\, s \in \mathcal{S},\, t \in \mathcal{T}} w_t\, o_{v,s}\, g_{v,s,t} \right] \tag{34}$$

in this Chapter.

As we have also already discussed in Section 2.3, electricity flows in the transmission grid are constrained by their respective line capacity multiplied by a factor of 0.7. By re-formulating the constrained optimisation formulation as a Lagrangian function, each constraint is associated with dual variable, also known as Karush-Kuhn-Tucker (KKT) multiplier.

$$\begin{aligned} |f_{(v,w),t}| &\leqslant 0.7 \cdot F^{\alpha}_{(v,w)} \\ &\leftrightarrow \begin{cases} \overline{\mu}^{\alpha}_{(v,w),t} \geqslant 0 \\ \underline{\mu}^{\alpha}_{(v,w),t} \geqslant 0 \end{cases} \forall (v,w) \in E,\, t \in \mathcal{T} \end{aligned} \tag{35}$$

The KKT multipliers $\overline{\mu}^{\alpha}_{(v,w),t}$ and $\underline{\mu}^{\alpha}_{(v,w),t}$ [€/MW] are zero, if the flow $f_{(v,w),t}$ is less or equal to 70% of its capacity, and strictly positive, if constraint (14) is binding, meaning that a better optimum of the overall annual costs according to the objective in equation (34) could be reached by increasing $f_{(v,w),t}$ beyond $0.7 \cdot F^{\alpha}_{(v,w)}$.

5.2.2 *Data Classification*

We model Germany with a maximum of 306 nodes, including all transmission lines from the ENTSO-E Interactive Transmission Map [136]. The network is adjusted according to annual reports from local TSOs [132, 137–140]. All manual adaptations to the lines are described in Annex D.1. Lines that were not build by the time of e.g. 2014, are removed from the optimisation for the simulations of later years. Lines that have been strengthened were reduced in capacity for the optimisation. Electricity demand data is derived from the OPSD [52] for the respective year and generation time series for hydroelectricity (run of river) are included and fixed. Capacities for conventional and renewable generators are taken from the new database provided by the German Federal Network Agency, the *Marktstammdatenregister* [141], see Figure 30 for their spatial distribution. Fuel costs, variable operation and maintenance costs per technology are based on historical market prices [55] and remain untouched compared to simulation results from different Chapters of this dissertation. Renewables have no marginal costs, but were given very small ones to set the curtailment order for wind and solar: 0 ct/MWh_{el} for run of river and geothermal, 1 ct/MWh_{el} for solar, 2 ct/MWh_{el} for onshore wind and 3 ct/MWh_{el} for offshore wind.

The dataset [141] contains geographic coordinates or equivalent information of all generators, where each one lies within a Voronoi cell. These cells are defined by a center point and cover the space that is closest in the sense of the euclidean metric, see Section 2.2. Therefore, we assign each generator to the Voronoi cell in which it lies which is represented by its central node v:

$$\operatorname{argmin}_{v \in \mathcal{V}} \sqrt{(x_v - x_g)^2 + (y_v - y_g)^2} \tag{36}$$

The data is filtered such, that the commissioning year of each generator matches the one of the demand time series. This is a simplification in the sense, that the register does not provide substations where the respective generator is attached to, only its geographical coordinates x_g and y_g, such that this assignment comes with errors.

To assess the assignment errors of both electricity demand and the generation fleet, we can quantify the amount of renewable energy available at the node which cannot be consumed locally or

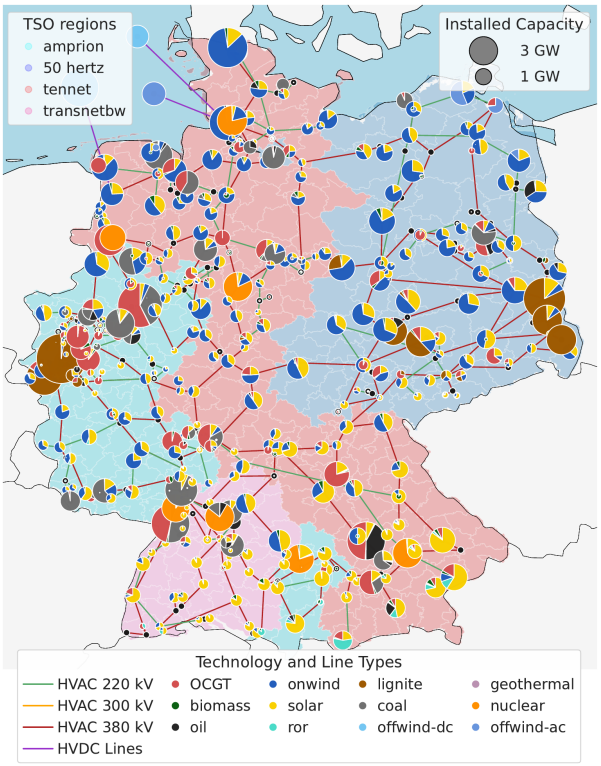


Figure 30: Original network model for Germany including all HVAC and HVDC transmission lines and powerplants (2018).

exported due to the constraint (35), i.e. the excess which necessarily must be curtailed:

$$\sum_{v \in \mathcal{V},\, s \in \mathcal{S}^{re},\, t \in \mathcal{T}} \left[A_{v,s,t} - d_{v,t} - \sum_{\substack{(u,w) \in E: \\ u=v \vee w=v}} 0.7 \cdot F^{\alpha}_{(u,w)} \right]^{+} \tag{37}$$

The bracket $[\cdot]^{+}$ denotes the positive part of a value; $\max(0, \cdot)$. Equation (37) captures an assignment imbalance in quantities of excess TWh: If the result is positive, it indicates that the installed potential at a local substation n is higher than local demand *and* higher than the transmission capacity. Building such a power plant is uneconomical, because it is known in advance that its power output cannot be used. Therefore, we assume either the assignment of g to v or the heuristic of spatially distributing electricity demand given in equation (1) to be inaccurate. Evaluation of (37) can be done a priori, i.e. without solving the optimisation problem (34) with its corresponding constraints (see Section 2.3).

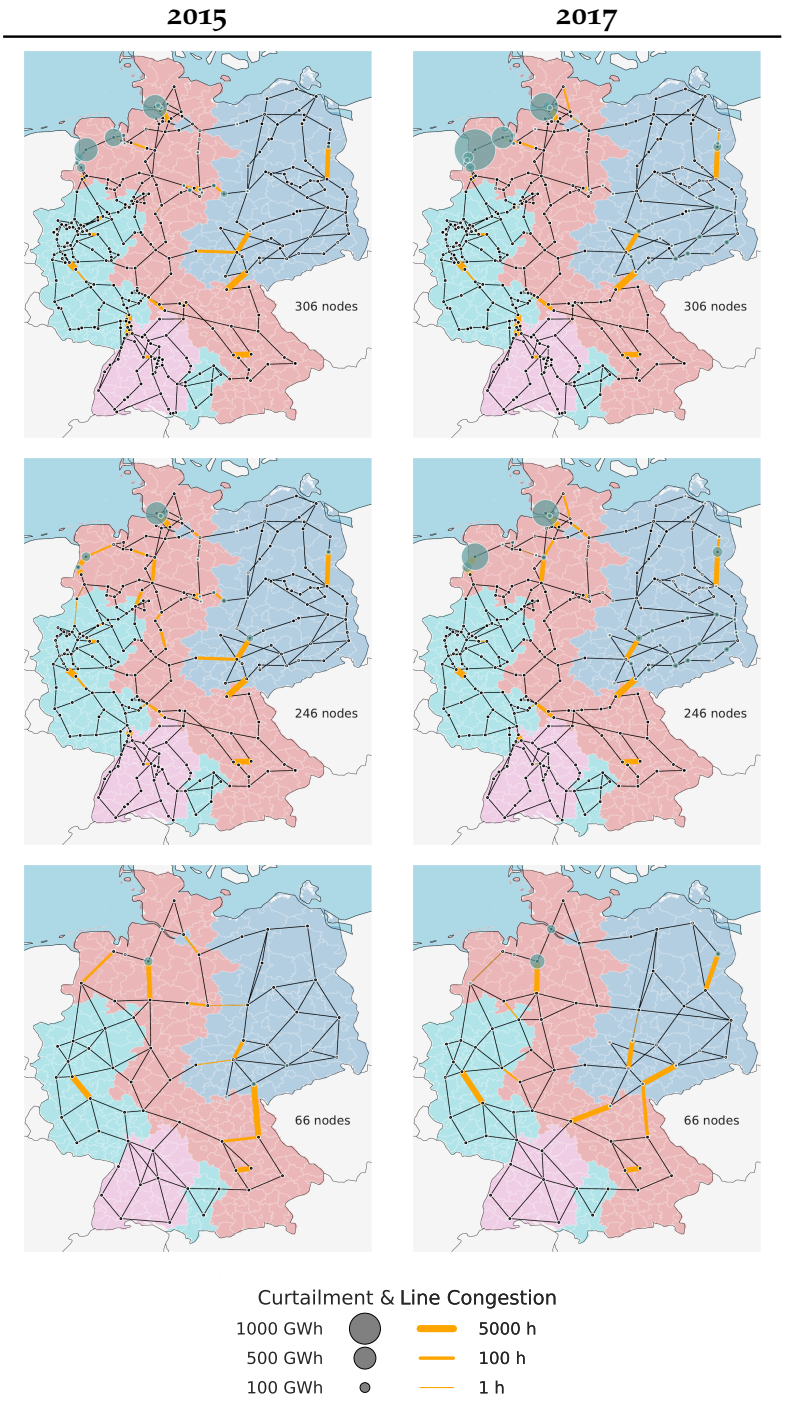


Figure 31: Clustered networks displaying the amount of curtailment for the years 2015 and 2017 for three exemplary network resolutions.

5.2.3 *Network Aggregation*

For this Chapter we chose the same version of k-means clustering as introduced in Chapter 3 of this dissertation that is based on the geographical location of the original substations in the network and weighted by the average electricity demand and conventional capacity at the substations, see Section 3.2. We also conduct the same network preparation where we map all voltage levels to 380 kV and aggregate univalent nodes. This method was as initially introduced in [100] and was used in many prior ESM studies conducted with PyPSA. We visualise three exemplary resolutions of the resulting clustering for two different years of the network in Figure 31.

For the remainder of this Chapter, we progressively cluster our highly-resolved German model with 306 nodes and 406-411 lines (2013/2018) in steps of 5 down to a 6 node network and compare the distinct results.

5.3 RESULTS

The original fully-resolved network model with assigned power plants is shown in Figure 30 and can be compared to three clustered down networks for the years 2015 and 2017 in Figure 31, that also shows additional spatial information of curtailment. Annual curtailment results for different network resolutions for the year 2017 are displayed in Figure 32.

Historical model results on curtailment in Germany are validated in three separate validation steps. First, on a cumulative scale where we present the total annual curtailment results in Figure 33. Second, on a spatial scale, where results per distinct control zone of the German TSOs are presented in Table 8. On the temporal scale we validate curtailment results per quarter of every year, see Table 9.

Finally, Figure 34 presents memory consumption and the duration to solve the simplified optimisation problem (34) with the same constraints as introduced in Section 2.3 as a function of the network nodes.

We have tested all results for stability by perturbing the assignment of generators g according to equation (36) with a probability of 5% to a node that is within the radius of 31km of the closest node v. 31km account for 5% of the longest east-west extension of Germany. The results are stable with deviations of below 5%.

5.3.1 *The Impact of Spatial Resolution on Curtailment*

Figure 32 displays the impact of clustering on curtailment results in Germany in the year 2017. The total amount of curtailment experiences four distinct stages, where in each stage the cumulative annual value is approximately steady, deviating from the mean by only 5%.

Those four stages can be distinguished by applying the excess energy measure discussed in (37): (i) First, at high model resolution, results are highly overestimated by more than 100% on average. Both curtailment and excess energy results are relatively stable with minor deviations of up to 5%. This is because both the assignment of electricity demand and power plants according to equations (1) and (36) in some cases is not precise, hence a mismatch emerges between low demand and high generation with lacking transmission capacities to transport the excess. (ii) Second, at intermediate model resolution between 250 and 150 nodes, curtailment results match the historical ones by on average 128%, i.e. deviating from historical numbers by 28%. At this stage, the effect of clustering overcomes the errors made by assigning electricity demand and generators to nodes, but at the same time, clustering preserves major transmission bottlenecks. Excess energy is still available due to the uncertainty of weather conditions, but is low at $0-1\%$ of available renewable energy. (iii) The third stage ranges from an intermediate to low resolution network (150 to 80 nodes) where both curtailment and excess results have large fluctuations. This high variance is because the probability to cluster important transmission lines becomes higher as fewer clusters are available for aggregation: Similar resolutions of plus or minus 10 nodes result in different minima of the k-means objective (19), and the choice of V_c to preserve major bottlenecks is crucial. (iv) In the final stage, the clustering technique overcomes the transmission constraints (15)-(35) and hence curtailment is highly underrated by $95\% \pm 5\%$: Without transmission constraints, all renewable energy is consumed because of its low marginal cost.

5.3.2 *Annual Curtailment Rates of a Spatially Aggregated Model*

Model results to simulate historical curtailment are shown in Figure 33, presenting a breakdown per carrier. It also displays the installed capacity of renewables.

Although the marginal costs of renewables were artificially assigned for the optimisation in (34) to fix the curtailment order, the energy-mix deviates from the historical mix only by 2.5% on aver-

Table 8: Curtailment per control zone. Results are extracted from models at a resolution of 246 nodes. The error compares model results to historical data of [134]

Year	TSO zone	Hist. share [%]	Model share [%]	Error [%]
2017	50Hertz	17.7	20.1	+1.6
	TenneT	81.6	80.0	−1.6
	Transnet	0.1	0	−0.1
	amprion	0.7	0	−0.7
2018	50Hertz	12	21.6	+9.6
	TenneT	87	78.4	−8.6
	Transnet	0.3	0	−0.3
	amprion	0.7	0	−0.7

age. We differentiate between solar, wind (onshore and offshore) and hydroelectricity.

The chosen resolution to model historical curtailment takes into account the analysis of the previous Section 5.3.1 to balance the mismatch of assigning input data to nodes versus clustering the transmission grid and overcoming important transmission bottlenecks. We choose a resolution of 246 nodes as the excess according to (37) is in the range of $0-10\%$ of the annual available renewable energy. Minor fluctuations are tolerated as they might not be compensated due to uncertain weather conditions. However, if excess energy accounts for more than 50% of the annual available energy, it would be a highly uneconomical location of the power plant, because it is known *in advance* that a large amount of the years energy can not be used.

A trend is seen that before 2017, the model tends to underestimate the total curtailment, with only 80% of the historical curtailment captured in 2016. However as wind generation grows, the model overestimates the congestion and therefore the curtailment, reaching 50% more than the historical numbers in 2018.

5.3.3 *Historical Curtailment Rates of a Spatially Aggregated Model per TSO Area*

To investigate the spatial distribution of curtailment across Germany, results per control area are presented in Table 8 for the years 2017 and 2018 in percent of annual curtailment. For consistency, the model resolution is chosen such as in Section 5.3.2. Results indicate, that curtailment numbers in our model deviate by up to 9% from historical values, and by 3.6% on average.

Table 9: Curtailment per quarter. Results are extracted from networks at a resolution of 246 nodes. The error compares simulation results to historical data from [135].

Year	Quarter	Historical share	Model share	Error
2015	I	24.0	33.4	+9.4
	II	15.6	10.6	-5.0
	III	17.3	14.2	-3.1
	IV	43.1	41.8	-1.3
2016	I	40.7	48.2	+7.5
	II	14.3	6.3	-8.0
	III	14.7	8.3	-6.4
	IV	30.3	37.2	+7.2
2017	I	25.6	23.3	-2.3
	II	24.7	19.7	-4.0
	III	7.9	8.7	+0.8
	IV	41.8	48.3	+6.5
2018	I	36.5	39.6	+3.1
	II	17.5	13.0	-4.5
	III	13.4	11.1	-2.3
	IV	32.6	36.3	+3.7

A validation of how these result change with the number of nodes representing the model show the same four stages as discussed in Section 5.3.1: In stage (i), where curtailment results are highly overestimated and electricity demand and generators were assigned to incorrect nodes, curtailment in 2017 is split 92 : 8 between TenneT and 50Hertz, 87 : 12 in 2018. These numbers deviate in both years by $\pm 1\%$ as the network resolution changes. The balancing of stage (ii) results in a stable 80 : 20 split between TenneT and 50Hertz in 2017, and 78 : 22 in 2018 with deviations of up to 2% in both years as the network resolution changes. Stage (iii) remains relatively random, which is true for stage (iv) as well, but in the latter, total annual curtailment is so low, such that the proportionality has no meaning.

5.3.4 *Historical Curtailment Rates of a Spatially Aggregated Model per Quarter*

Finally, we consider curtailment per quarter in the years 2015-2018. Results are displayed in Table 9 in percent of annual curtailment. The model resolution is chosen the same as in Section 5.3.2 for

reasons of consistency. Results indicate that the distribution of curtailment at an hourly resolution model reflect the historical distribution with an error of 4.7% on average.

Again, these results are validated on how they change with the cluster size. Here, a positive trend can be observed as the clustering happens mainly in space and not in time, such that results are stable for stages (i)-(iii). The average share reflects the number in Table 9, while it starts fluctuating towards stage (iv), where we know that the overall curtailment tends to 0, such that the proportionality has no meaning.

5.3.5 *Memory Consumption and Solving times*

Solving the full optimisation problem with a full resolution network of 306 nodes and more than 400 HVAC or HVDC lines requires approximately 20 GB RAM on our HPC and runs for almost an hour, while a clustered network of only 156 nodes is twice as fast and needs about 30% less RAM, while providing more accurate results.

Further results can be found in Annex D.

5.4 CONCLUSIONS

We have shown that historic curtailment in Germany can be reproduced in the open model PyPSA-Eur using the latest database for the locations of existing RE generators. Results agree well in time (curtailment per quarter) and space (curtailment per TSO region), provided a balancing resolution is used that is low enough to overcome assignment-errors and high enough to account for important transmission routes. We suggest to cluster the 306 node German network to well below 280 nodes, but not below 150 nodes. In this range, curtailment in the model provides the best match with historical data. A resolution below 100 nodes for Germany using a weighted k-means clustering scheme is not advisable.

5.5 CRITICAL APPRAISAL

This case-study covers Germany only, neglecting the fact that power can also be exchanged with bordering countries, such as France, Denmark, Poland, the Czech Republic, Austria or Switzerland, reducing the overall curtailment of renewables in Germany. Previous studies have pointed out, that international cooperation benefits renewable electricity markets [46].

Further, to avoid the difficulty of keeping track of different voltage levels as the network is reduced, all lines are mapped to their

electrical equivalents at 380 kV, the most prevalent voltage in the German transmission system. The electrical parameters and capacities of the lines use standard assumptions for 380 kV circuits whereas in reality they vary from line to line. In addition, we use constant summer thermal ratings for an outside temperature of 20 Celsius throughout the year and do not adapt them for lower temperatures or wind conditions. The use of winter ratings as well as dynamic line rating [131] on some congested lines today may account for the lower historical curtailment compared to our model.

Finally, the capacity factors for wind and solar from weather data overestimate historical production, so we linearly reduced them by a factor of 0.9 for wind and 0.8 for solar for each point in time and space.

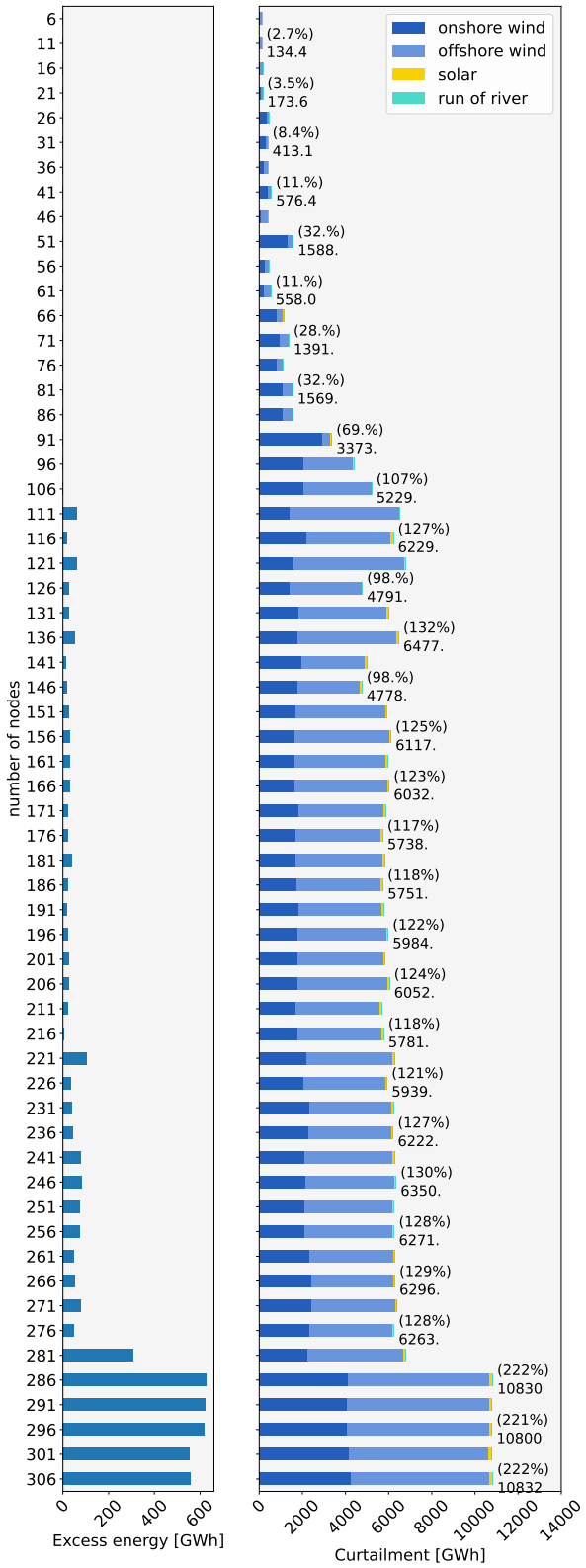


Figure 32: Excess energy according to equation (37) (left) and model curtailment in GWh (right) for the weather year 2017 respective network size.

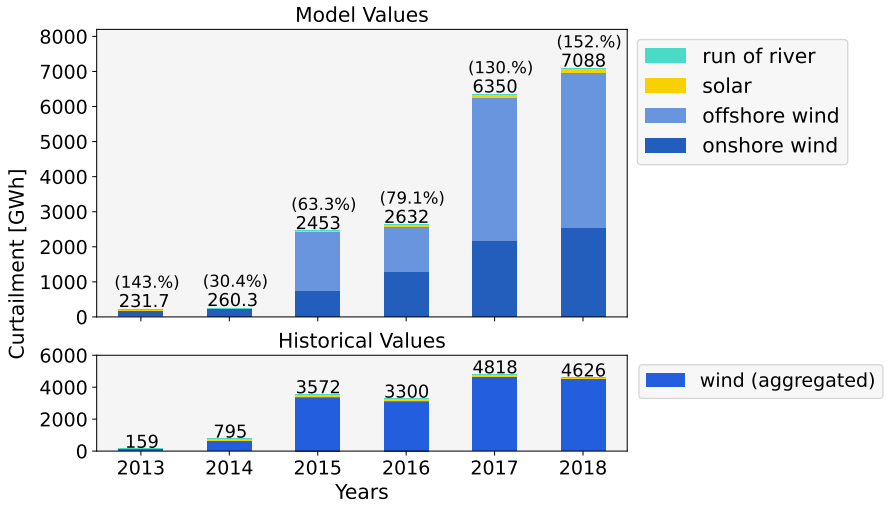


Figure 33: Model results on historical curtailment (top). The number in percent shows the agreement with historical data (bottom) from [128]. Results are extracted at a network resolution of 246 nodes.

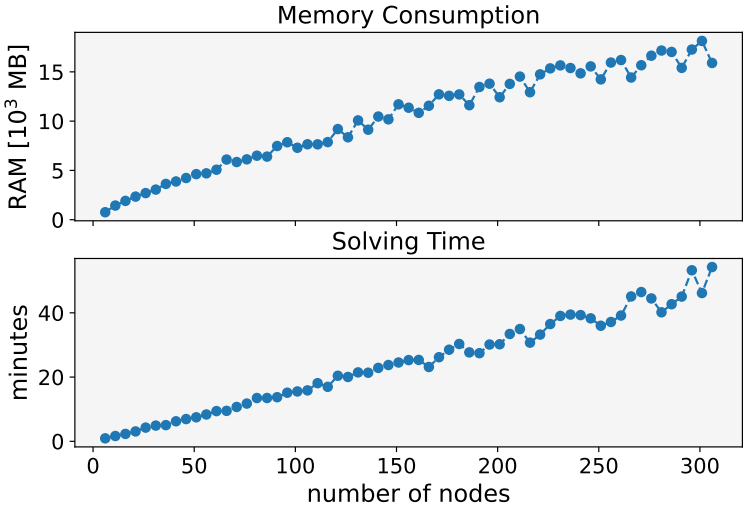


Figure 34: Memory consumption and solving time per cluster resolution, exemplary numbers for the year 2017.

6

HOW FEASIBLE ARE SPATIALLY LOW-RESOLVED MODELLING RESULTS?

CONTENTS OF THIS CHAPTER ARE BASED ON

Martha Maria Frysztacki, Veit Hagenmeyer, and Tom Brown. "Inverse methods: How feasible are spatially low-resolved capacity expansion modeling results when dis-aggregated at high resolution?" In: *submitted to Energy* (under review). DOI: https://doi.org/10.48550/arXiv.2209.02364

6.1 INTRODUCTION

In the previous Chapters, we have identified that a high spatial resolution is required for the modelling to produce an accurate representation of renewable generation. We know that modelling a large geographical area at the scale of Europe allows the model to exploit very good continental renewable potentials and strongly impacts the composition of the generation and storage fleet of individual regions [46] while, in the same time, it is relevant to include a high spatial granularity into models with heterogeneous regions [142]. This is particularly relevant for models with a high share of renewable generation. We have seen that the necessity for the models to include a high spatial resolution is essential to detect bottlenecks in the transmission grid [143], assess renewable potentials based on local weather conditions and to identify regional variations in electricity demand. The granularity of this data-driven information significantly improves the design and composition of the electricity mix and the routing of new grid infrastructure predicted by the model.

In Chapter 3, we have disentangled the effects of sourcing renewable generation versus routing the electricity to locations of high demand, revealing that routing dominates the system effects and forces the model to build renewable assets closer to demand centers at potentially worse capacity factors at a high spatial resolution. But the downside of higher resolution is that processing large data volumes inevitably results in a computational burden that arises from solving the associated mathematical formulation in the model, see Chapter 2, particularly Section 2.4. To obtain more reliable investment decisions, more effective cluster-

ing methods must be developed so that the spatially aggregated models can better represent the highly-resolved system. We have contributed to this target by using the lessons from Chapter 3 to design novel clustering methods in Chapter 4. To this end, we have focused on HAC methods that are based on different features of the model, such as VRE generation or electrical parameters of the grid, because these clustering methods can better approximate the topology of the transmission grid.

The overall consensus of the previous Chapters and most other studies is to model the European electricity system that scatters across an area of 10 million square kilometers and contains more than 5000 substations at and above 220 kV at a spatial model resolution of $100 - 200$ nodes. An exact recommendation depends on the model configuration, such as the technology-mix or the allowed amount for transmission reinforcements. It is therefore subject to an individual analysis of the set-up of the model.

state of the art All these previous efforts were dedicated to understanding and improving spatially low-resolved models. As of now, no study provided insights whether the spatial resolution of the models impacts only the optimal found solution, or whether these spatially simplified modelling results are feasible with respect to the original, spatially highly-resolved model. Moreover, approaches to disaggregate the simplified modelling results back into its original highly-resolved set-up are not well represented in previously published research. There exist only a few publications, such as from [144] who disaggregate simplified modelling results into higher spatial detail. But this approach was conducted only as means to distribute the resulting optimal dispatch from transmission level down to substation level in order to be processed at distribution level. Capacity expansion was not allowed, neither for generation or flexibility options. The study also did not analyse the overall feasibility at a higher spatial resolution. Thus, neither the assumptions on the disaggregation, or the resulting highly-resolved systems were analysed in detail with respect to their plausibility. Another disaggregation method is presented by [145], who propose an iterative spatial disaggregation process and allow transmission grid expansion in the final iteration, such that the highly-resolved model becomes feasible. Here, it remains unclear how feasible the disaggregated model is when omitting the final transmission reinforcements. Finally, [146] propose three disaggregation methods. However, in their case-study they pursue an iterative aggregation and disaggregation with the overall aim that their resulting model is designed to be feasible at any spatial resolution.

research contributions In this Chapter, we look back at the spatially highly-resolved electricity model and disaggregate the spatially coarse optimisation variables at higher resolution. To this end, we present three different disaggregation methods in our continent-scale model at hand. Two of them are adapted inverse methods from the literature, where both of them are very intuitive approaches. One of them provides a very vague estimation of how the low-resolved variables can be distributed, and the other one is potentially very demanding from a computational point of view. The third method is a completely novel algorithm that builds on a metric derived in Chapter 5 of this dissertation. It provides a compromise between the previously proposed methods where it makes use of a novel metric that describes the system by balancing the distribution of renewable yield with electricity consumption and existing grid infrastructure. In the same time, the method is considerably less computationally demanding. These inverse methods can be used to analyse if the spatially low-resolved optimisation models are suit for reliable investment decisions. For this analyis, we use the spatially disaggregated, highly-resolved model results to examine if the solution derived from the simplified models can solve the spatially more complex operational optimisation. Because now the optimisation is reduced by the investment decisions, we can solve it at higher granularity. This study design is completely novel in the ESM literature, as none of the disaggregation methods was previously analysed with respect to model fidelity.

remainder The remainder of this Chapter is organised as follows. We present a detailed description of three methods to disaggregate coarse model results in Section 6.2. Two of the disaggregations are improved methods that were proposed in earlier research, and one of them is completely novel. Results from the disaggregation are presented in section 6.3, where we test if the disaggregated highly-resolved models are feasible from a technical point of view. Conclusions are drawn in section 6.4. Limitations of this study are discussed in Section 6.5.

6.2 METHODS

We present the overall modelling process in Section 6.2.1, where we also highlight the novelty of this research. The three inverse methods on how spatially low-resolved capacity expansion model results can be disaggregated back into higher spatial detail are presented in sections 6.2.3-6.2.5 and are summarised Table 10. Treatment of inter-cluster power flows is discussed in Section 6.2.6.

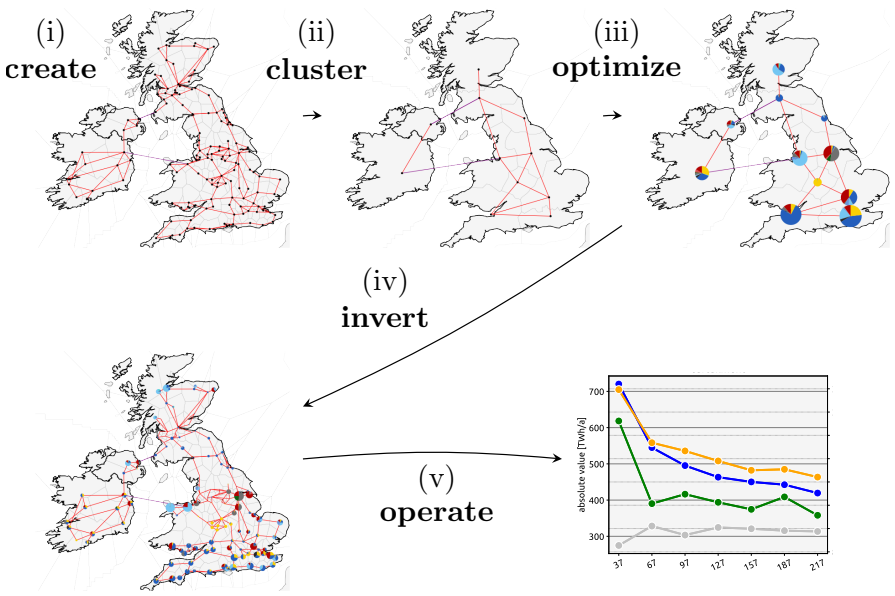


Figure 35: Illustration of the disaggregation study design of this Chapter.

The study design and evaluation of results is presented in Section 6.2.7.

6.2.1 *Modelling Overview*

Figure 35 displays the overall approach of electricity system modelling. It typically executes in the following order: (i) Creating the model. This includes collecting data of the system to be analysed (as we did in Section 2.1), for example the network topology of the transmission system, capacities of generators that are to be included in the model, land-use constraints, time-series of electricity demand, wind speeds, solar radiation, etc. and assigning the data to the correct locations. (ii) Formulating a set of mathematical equations associated with the problem at hand (Section 2.3), clustering the spatially highly-resolved network down to a smaller approximation to gain computational advantages and (iii) solving it.

Here, we introduce an additional fourth and fifth step to this queue: (iv) Disaggregating the low-resolved results (i.e. the resulting coarse renewable capacities $G_{c,s}$) back into high spatial resolution (i.e. the set $\{G_{v,s} : v \in \mathcal{V}\}$). (v) Running an optimisation with fixed capacity that is derived from step (iii). This is also referred to as operational optimisation. It allows to gain insight into the dynamics of the electricity system, and particularly analyse its feasibility.

But as the clustering

$$f : \mathbb{R}^m \to \mathbb{R}^n \quad \text{with} \quad n < m, \text{ often even } n \ll m$$

reduces the (spatial) dimension of the data, the mapping is surjective but not injective, hence not bijective. Therefore, finding an inverse in step (iv) that maps the results back into high dimension

$$f^{-1} : \mathbb{R}^n \to \mathbb{R}^m$$

is a challenging task and the inverse is not unique. We propose three different approaches to tackle the disaggregation of the generation and storage capacities for each technology from the low-resolved model to the highly-resolved model and suggest adequate inverse methods. The proposed methods are summarised in Table 10 and are explained in detail in the following three Sections.

6.2.2 *Model and Input Data*

As continuously done in this dissertation, we again employ the openly available European Electricity System Model at transmission substation level, PyPSA-Eur [51], as introduced in Chapter 2, particularly Sections 2.1 and 2.3. Variables appearing in this Chapter are listed in Table 1.

All simulations in this Chapter are carried out with a two-hourly resolution in time, i.e. the weight w_t is fixed to 2 in our application ($w_t \equiv 2$ in equation (3)). This approach has shown to yield good results compared to full temporal resolution while reducing the model size by a factor of 2 [108].

To enable computational feasibility, the whole model with more than 5000 nodes must be reduced to a computationally tractable size by spatially clustering the nodes. Then, the reduced optimisation problem given in equation (3), now with a smaller set of nodes $\mathcal{V}$, can be solved. As we have seen, there exist many approaches to spatially reduce the size of an ESM. We have demonstrated in Chapter 4 (see also [3, 89, 103]) that the best suited clustering method for capacity expansion models with highly renewable scenarios are of hierarchical nature. Therefore, we choose an HAC method for clustering the model spatially in this Chapter and apply the same similarity measure as previously analysed in Section 4.2.1. It is defined such that the aggregated nodes have the most similar renewable time-series $\bar{g}_{v,s,t}$ throughout the whole year, see equation (26).

Table 10: Summary of the proposed disaggregation methods.

Short name	Method description
	Optimal capacities retrieved from an optimised low-resolved model are distributed ...
uniform	... uniformly across all nodes within a cluster of the highly-resolved network while accounting for land-use restrictions by imposing an upper bound.
re-optimise	... anew by re-optimising capacities within each cluster with full formulation, while enforcing the same build-out capacity totals per technology as in the clustered model.
min excess	... according to a local optimisation which seeks to concentrate generation at nodes with higher demand and grid capacity and thus to minimise load-shedding.

6.2.3 *Uniform Disaggregation*

The first approach to disaggregate low-resolved modelling results is simple and computationally inexpensive. For each generation and storage technology, we distribute the capacity retrieved from the coarse model within every cluster uniformly across all highly-resolved nodes:

$$G_{c,s} \mapsto \frac{1}{|\mathcal{V}_c|} \begin{pmatrix} G_{c,s} \\ \dots \\ G_{c,s} \end{pmatrix} \in \mathbb{R}^{|\mathcal{V}_c|} \quad \forall c \in \mathcal{V}^K,\, K \in \mathbb{N}_{\geqslant 37} \tag{38}$$

An additional constraint is formulated to account for land-use constraints, see equation (4). To enforce the inequality, we select those generators where (4) is not satisfied and uniformly distribute the residual capacity over the remaining nodes within the cluster, i.e. across the following set of generators:

$$\{G_{(v,s)} : v \in \mathcal{V}_c \wedge G_{v,s} < G_{v,s}^{max}\} \tag{39}$$

The last step is repeated until all nodes satisfy equation (4).

6.2.4 *Regional Re-Optimisation*

The second considered method may be computationally challenging. Within each cluster, we re-optimise at high resolution the orig-

inal objective function (3) with all associated mathematical constraints. Additionally, we impose another set of constraints to incorporate the spatially low-resolved modelling results:

$$\sum_{v \in \mathcal{V}_c} G_{v,s} = G_{c,s} \quad \forall c \in \mathcal{V}^K,\, s \in \mathcal{S} \tag{40}$$

This set of constraints ensures that the amount of installed capacity is the same for every technology in every region.

Depending on the size of the cluster, the disaggregation may blow up the problem beyond the computational capacity of the machine and can result in an even larger problem than the clustered one. On the positive side, the re-optimisations for each cluster can be run in parallel.

6.2.5 *Minimal Excess Electricity*

Our third Ansatz for disaggregation is motivated by finding a compromise in terms of computational resources. It is not evident that solving the full optimisation problem is necessary to distribute renewable capacity. Instead, we define a simpler objective function that minimises (renewable) excess electricity. It is designed to spatially align renewable generation with demand and possible flexibility options inside the cluster and has proven in Chapter 5 (see also [1]), that it can determine a priori, meaning before solving the optimisation, if there is excess capacity in a node. The optimisation is formally defined as

$$\min_{G_{v,s}} \sum_{\substack{v \in \mathcal{V}_c,\\ s \in \mathcal{S},\, t \in \mathcal{T}}} \left[\bar{g}_{v,s,t} G_{v,s} - d_{v,t} - 0.7 \sum_{\substack{l_{(v,w)} \in E:\\ v=c \vee w=c}} F_{(v,w)} \right]^+ \tag{41}$$

in alignment with the excess defined in equation (37). The bracket $[x]^+ := \max\{0, x\}$ yields the positive part of the sum. Compared to the regional re-optimisation introduced in Section 6.2.4, the set of additional constraints to the optimisation problem is much smaller. We only impose the constraints (4) and (40).

Similarly to the previous methods, this disaggregation method can be run in parallel for each cluster.

6.2.6 *Modelling Power Flows between Clusters*

For two out of the three proposed disaggregation methods ('re-optimise' and 'min excess'), we add additional boundary constraints on the inter-cluster transmission lines to simulate electricity im- and exports. This can be done by extracting the optimal power

flows of the low-resoled network $f_{(c,d),t}$ and distributing them proportional to the capacities of the inter-cluster highly-resolved transmission lines $F_{(v,w)}$, following

$$f_{(v,w),t} = \frac{F_{(v,w)}}{F_{(c,d)}} f_{(c,d),t} \quad \forall (v,w) \in E_{(c,d)} \tag{42}$$

where the resulting power flow $f_{(v,w),t}$ is in the interval $\left[-F_{(v,w)}, F_{(v,w)}\right]$ by definition. A formal definition of the set $E_{(c,d)}$ is given in equation (22). The resulting power flow $f_{(v,w),t}$ at the spatially highly-resolved line (v,w) is modeled as additional demand imposed on all nodes $v \in V_c$ and $w \in V_d$ that are connected by a transmission line $(v,w) \in E$, with $c \neq d$, i.e.

$$\begin{aligned} d_{v,t} &\mapsto d_{v,t} + f_{(v,w),t} \quad \forall v \in V_c\,, \\ d_{w,t} &\mapsto d_{w,t} - f_{(v,w),t} \quad \forall w \in V_d \quad (c \neq d) \end{aligned}$$

where positive power flows represent electricity imports and negative ones electricity exports. These results do not deviate strongly from those where each region is treated as an island, meaning that no power flows retrieved from the coarse model are considered in the disaggregation. However it is plausible to include them when inverting modelling results, therefore in this article we focus on this approach. Islanded results can be found in Annex E.1.

6.2.7 *Study Design*

To investigate the quality of the proposed disaggregation methods, the model is solved as a pure operational problem, where no further capacity can be built. This is equivalent to solving equation (3) with its associated constraints, however the technology capacities $G_{v,s}$ are removed from the set of optimisation variables and replaced by a fixed number. We have solved the same operational optimisation in Chapter 5, see equation (34) that simplifies the originally introduced objective to a purely operational dispatch problem. Loadshedding generators with high but non-extendable capacity are added to the network to guarantee physical feasibility. The operational problem is computationally less extensive to solve because the capacity expansion has been removed from the problem. Therefore, solving a spatially highly-resolved model becomes feasible.

As main quality measures we consider the amount of loadshedding in the highly-resolved network with the disaggregated results and the amount of renewable curtailment. Curtailment describes how much abundant electricity the low-resolved model

Table 11: Investigated intra-cluster scenarios for each of the disaggregation methods. This means we additionally formulate constraints on the transmission lines within each cluster.

Short name	Scenario description
	The highly-resolved network with disaggregated capacities is solved as an operational problem where ...
regular	... no further adaptations are made.
copperplate	... the intra-cluster transmission capacity is infinitely high. Note, that the inter-cluster transmission capacity is still bound.

chooses to generate which, when highly resolved, cannot be transported to locations with high electricity demand. This is mostly due to an inaccurate choice of siting capacity due to missing information about possible transmission bottlenecks in the low-resolved model. Loadshedding is chosen because it indicates how much capacity is underestimated by the low-resolved model due to averaging capacity factors and removing grid bottlenecks from the network. Loadshedding could stem from different reasons: i) the disaggregation of solar and wind capacities to multiple sites with different capacity factors could result in a lower overall yield compared to the aggregated site, or ii) the grid bottlenecks inside the clusters could cause congestion, such that power generated at locations with surplus of electricity can not be transported to locations with high net load.

As loadshedding is a greater risk in terms of energy security, we design a test to better understand its origin. To rule out reason ii), we run a second operational scenario (which we refer to as 'copper-plate'), where the capacity of all transmission lines that have vanished in the low-resolved network due to aggregation are set to ∞, i.e.

$$F_{(v,w)} \to \infty \quad \forall (v,w) : v,w \in \mathcal{V}_c,\ c \in \mathcal{V}^K,\ K \in \mathbb{N}_{\geqslant 37} \tag{43}$$

This modification is only applied to solve the operational problem, not for the disaggregation of results.

Note that all inter-cluster transmission capacity is still finite, meaning

$$F_{(v,w)} \ll \infty \quad \forall (v,w) \in E_{(c,d)}. \tag{44}$$

A summary of the two considered scenarios ('regular' and 'copperplate') is provided in Table 11.

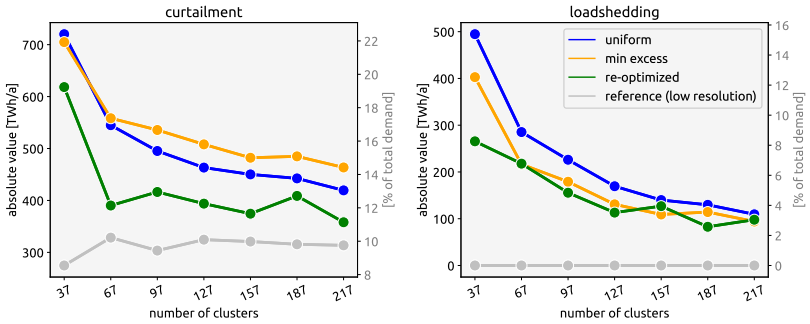


Figure 36: Amounts of annual loadshedding and curtailment of the highly-resolved operational problem. Transmission capacity within every cluster is not adjusted.

6.3 RESULTS

We first present the feasibility of low-resolved modelling results when disaggregated into high spatial resolution in section 6.3.1 using the three proposed disaggregation methods. We distinguish between the 'regular' set-up where the intra-cluster transmission capacity is not changed (section 6.3.1.1) where we additionally analyse where the curtailment and loadshedding measures are spatially located (section 6.3.1.2) and the 'copperplate' one, where the transmission capacity within clusters of the spatially highly-resolved model is set to infinity to approximate the clustered copperplate optimisation model (section 6.3.1.3). Finally, we discuss computational Trade-Offs of the presented disaggregation methods in section 6.3.2.

6.3.1 *Feasibility Considerations*

6.3.1.1 *Regular Intra-Cluster Transmission Capacity*

The results presented here are derived from the 'regular' set-up, meaning with no adjustment to the intra-cluster grid capacities in the high resolved optimisation model (see Table 11). The resulting amounts of loadshedding and curtailment are presented in Figure 36.

For any of the three proposed disaggregation methods we can see that the amount of both curtailment and loadshedding decreases as the resolution of the underlying low-resolved model increases. However, there are substantial differences in the performance of the disaggregation methods.

Curtailment rates of the different methods deviate by $1-3\%$ from one another on average, depending on the reference resolution. Distributing coarse investment variables across the spatially highly-resolved operational model results following the 'min ex-

cess' approach yields the highest curtailment rates of $11\% - 22\%$ of annual electricity demand, depending on the coarse reference resolution and the disaggregation method. The lowest reference resolution has the highest curtailment. Uniformly distributing results performs similar to the 'min excess' method at a very low reference resolution of 37 nodes (one node per country), resulting in 22% of curtailed electricity. The 're-optimise' method performs better in this regard, resulting only in 19% of curtailed electricity. But the curtailment rates decrease to approximately 14.5% ('min excess'), 13% ('uniform') and 11% ('re-optimised') of annual electricity demand, as the reference model resolution increases. Re-optimising the local problem yields the lowest curtailment for every reference model resolution, which is approximately $2 - 3\%$ lower compared to the results of the 'uniform' approach.

Regarding loadshedding, for a low-resolved network where every country is represented by a single node (37 clusters in Figure 36), 're-optimise' performs best as it results in the lowest loadshedding rates. Re-optimising yields approximately 265 TWh or 8.2% of the annual electricity demand that can not be covered by renewable generation. If this gap were filled with gas to satisfy electricity demand, annual carbon emissions would rise from 0% to 3.2% of 1990s levels. Compensating the unmet demand when disaggregating results with the 'min excess' method yields approximately 400 TWh of loadshedding, resulting in 4.8% of carbon emissions of 1990 (1.6% more compared to 're-optimised') if gas is dispatched for the loadshedding measure. When uniformly disaggregating renewable capacity within the clusters, the operational problem returns 500 TWh of loadshedding measures. Compensating with gas would result in 6% of carbon emissions of 1990, 1.2% more compared to 'min excess'.

When increasing the reference model resolution, the amount of loadshedding decreases for all the disaggregation methods. It can be seen that at a model resolution of 67 or more nodes, the amount of loadshedding is in the same range for the methods 're-optimised' and 'min excess' deviating by only 0.5% on average. When uniformly distributing the retrieved coarse capacities, loadshedding measures are higher than those of the competing disaggregation methods by initially 2.8% at a reference resolution of 37 nodes and linearly decreases as the resolution of the reference model increases. At around 187 nodes, the difference for all three methods is below 0.5% in terms of necessary loadshedding measures. At an underlying reference model resolution of 217 nodes, the amount of loadshedding is always within the range $94 - 110$ TWh, corresponding to $3 - 3.5\%$ of annual electricity demand in Europe.

6.3.1.2 *Localisation of Loadshedding and Curtailment*

In this section we analyse where curtailment and loadshedding is spatially localised. Recall that the highly-resolved model yields loadshedding measures because of i) disaggregating capacity factors results in a different overall yield of renewable electricity or ii) grid bottlenecks that did not occur in the low-resolved model, as described in detail in section 6.2.7.

Figure 37 displays the regions of curtailment and loadshedding spatially distributed after running the operational highly-resolved 1250 node model for all three disaggregation methods for a reference model resolution of 97 nodes. In all three cases it can be seen that the loadshedding is scattered in central Europe such as southern Poland, central and southern Germany, Switzerland and Austria and thus far from coastal areas and southern locations, such as e.g. northern Germany and France, Italy and Spain. In the same time, these locations have high amounts of curtailment. Transmission lines connecting regions with high amounts of curtailment and regions with high loadshedding show high congestion rates. Thus, the resulting network suggests that the low-resolved model favors investments in wind turbines at locations with good wind conditions at coastal areas, and in solar panels in the southern regions with good solar radiation, while it is blind to transmission bottlenecks that prohibit transporting the electricity to demand centers.

6.3.1.3 *Infinite Intra-Cluster Transmission Capacity*

To verify why loadshedding measures are necessary as well as to better understand the high curtailment rates, we now consider a setting where within each cluster the transmission capacity is set to infinity, following the description provided in the beginning of section 6.2.7, see equations (43)-(44). This means that in the highly-resolved model, only the capacity between clusters is limited. Results on loadshedding and curtailment are presented in Figure 38.

It can be seen that the amounts of renewable curtailment deviate by less than 5% from the curtailment rates of the reference model. They can mainly be explained by varying capacity factors. In the highly-resolved models, larger deviations of capacity factors within each clustered region become available compared to the reference model.

As the reference resolution increases, the amount of curtailment also tends to increase slightly. This can be explained by the fact that more total capacity is installed for a higher resolved reference resolution.

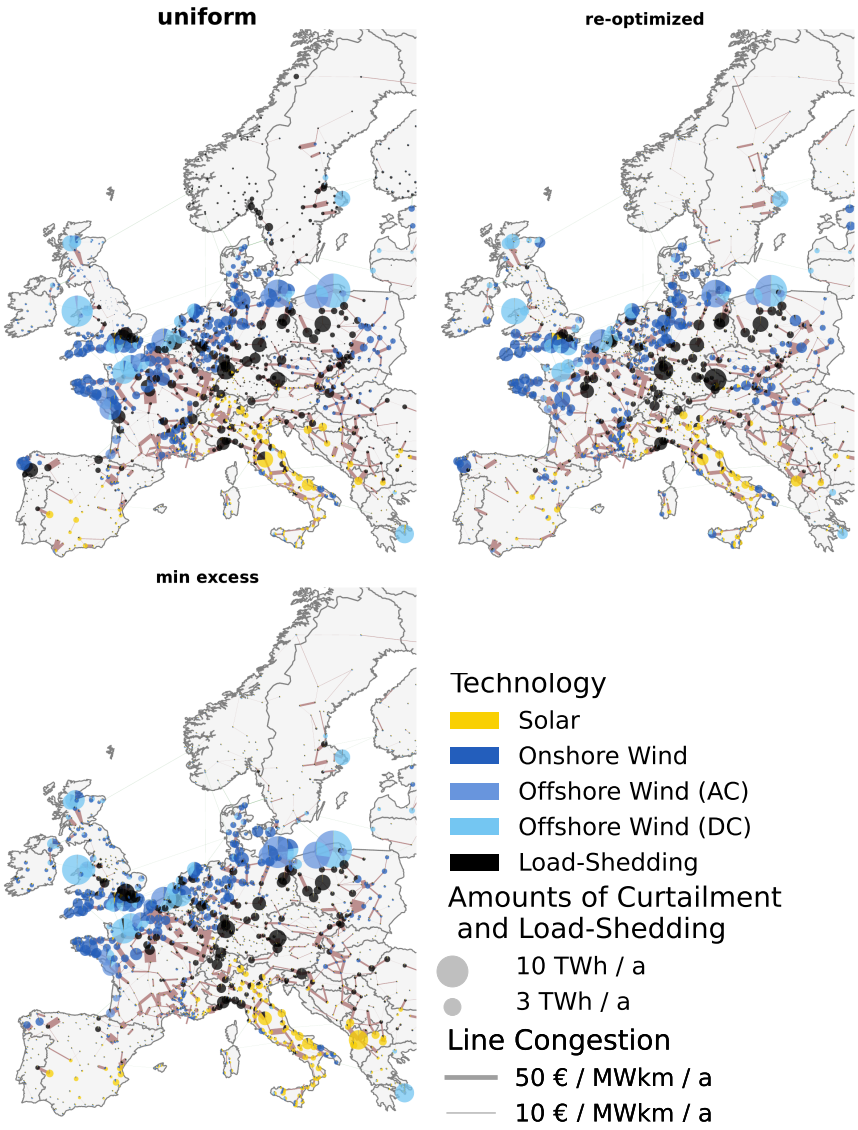


Figure 37: Spatial distribution of curtailment and loadshedding across Europe in the disaggregated, highly-resolved model. Capacity installations are taken from a reference model resolution of 97 nodes.

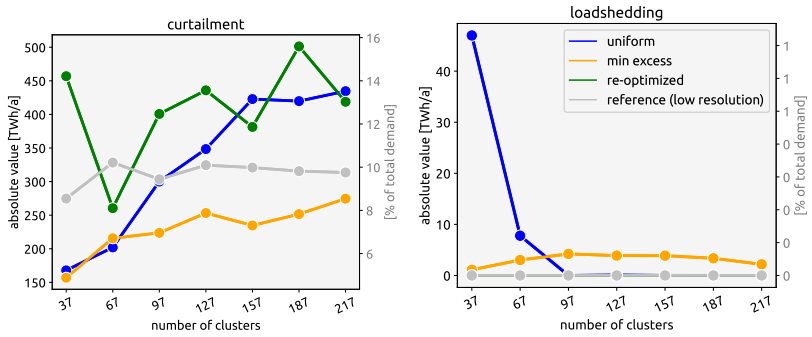


Figure 38: Amounts of annual loadshedding and curtailment of the disaggregated highly-resolved operational problem. Transmission capacity within every cluster is set to infinity.

Regarding loadshedding, the amount for uniformly disaggregating renewable capacity drops to 0 for a reference model resolution above 100 nodes. For a one-node-per-country reference model (37 nodes), there remains a relatively low amount of loadshedding of approximately 50 TWh, resembling about 1.5% of annual electricity demand. In case the 're-optimise' disaggregation Ansatz is invoked, loadshedding decreases to 0% of annual electricity demand for every low-resolved reference model. 'min excess' yields less than 5 TWh ($< 0.5\%$) of loadshedding measures for any reference low-resolved model. At peak (97 nodes) this amount of gas would emit around 800 kg of CO_2 (0.05% of 1990s emissions). We conclude that these results are consistent with the main cause of loadshedding being the transmission restrictions within the clusters.

6.3.2 *Trade-Offs of the Disaggregation Approaches*

There are four main qualities that we consider when evaluating trade-offs of the different disaggregation methods. First, the quality of results: How well do the proposed methods solve the problem at hand? Second and third, we consider the computational efforts. These mainly focus on the question: Are the proposed methods computationally legitimate for the considered problem? This consideration includes not only the memory requirements needed to solve the problem, but also the time it takes to solve. Fourth, depending on the results of the methods or the problem formulation, it might also be worth considering the efforts to implement a solver.

For the proposed methods in this Chapter, a summary of these four qualities is provided in Table 12.

The performance with respect to the quality of results of the proposed disaggregation methods was already discussed in section

Table 12: Trade-Offs of the three proposed approaches to disaggregate results. A ✓ indicates a reasonable trade-off, a ✗ indicates an inadequate compromise.

	Implementation	Solving Time	Memory (RAM)	Results Quality
uniform	✓	✓	✓	✗
min excess	✗	✓	✓	✓
re-optimise	✗	✗	✗	✓

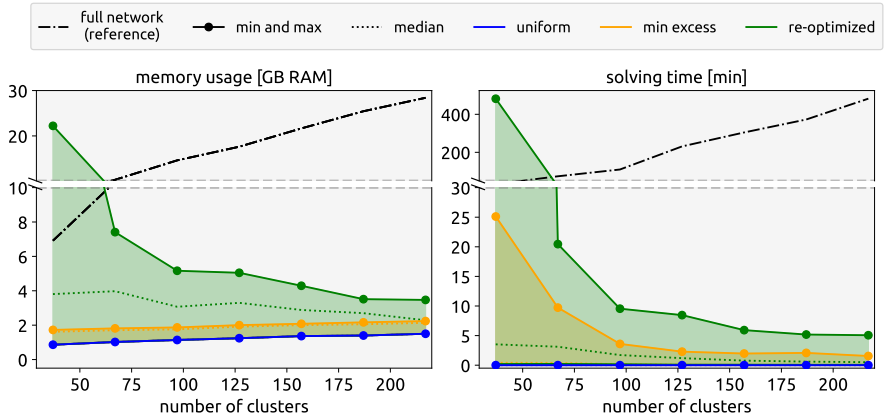


Figure 39: Memory requirements (left) and solving times (right) for executing the proposed disaggregation methods for individual regions.

6.3.1. Now, we analyse the performance of the proposed methods from a computational point of view. Rating the efforts of implementation is a subjective task, we therefore solely relate to the fact that uniformly distributing a number across a set of nodes does not involve mathematical optimisation. Therefore applying an uniform distribution is rated 'easier' than formulating a mathematical constraint to an existing optimisation problem as proposed in 're-optimise', or a whole optimisation problem including both objective function and associated constraints, as proposed in 'min-excess'.

Computational resources and solving times for disaggregating low-resolved model results back into high spatial resolution are presented in Figure 39 for every proposed method.

Resource-wise, re-optimising the local model consumes up to 13 times (1.7 times in average) the amount of resources compared to minimising a simpler objective in 'min excess', and up to 26

times (2.7 times in average) compared to uniformly distributing the capacity obtained from the low-resolved model ('uniform'). In absolute numbers, the method 're-optimise' consumes up to 22.2 GB RAM at peak, compared to 2.2 GB RAM for 'min excess' and only 1.5 GB RAM in case of 'uniform'. Today's average state-of-the-art personal computers are able to solve both the 'uniform' and 'min excess' problem formulations for any model resolutions, while solving the 're-optimise' approach needs more computational power and, therefore, requires a more advanced machine or even a high-computational cluster access. All local disaggregation runs were carried out in parallel.

Considering the computational times, these trade-offs are similar. The method 'uniform' is up to 4000 times faster at peak than 'min excess' and 70 times faster in average. In turn, 'min excess' is up to 20 times faster than 're-optimise' and 11 times faster in average. Note that computational times might change when allowing a lower accuracy of the results. Here, we choose a barrier convergence tolerance of 10^{-9} and a feasibility tolerance of 10^{-6}, which is not necessarily required. A tolerance of 10^{-3} might suffice in most applications. Lowering the tolerance of the solver will reduce solving times, but the memory consumption is likely to persist.

All experiments presented in this article were carried out on a HPC with 5 nodes, each having an allocatable capacity of 48 central processing units (CPUs) and 256 GB RAM memory.

6.4 CONCLUSIONS

From these results, we can draw conclusions on the methodology of the disaggregation methods as well as on the insights of disaggregating coarse modelling results into a higher spatial detail.

The presented methods to disaggregate optimal infrastructure investment of renewable generation technologies and flexibility options have significant differences in their quality of results, simplicity of implementation and computational resource consumption. We have shown that it is not necessary to locally solve the a full optimisation problem to disaggregate coarse results into higher spatial detail, as it was conducted in previous research. Instead, it can be sufficient to formulate a suitable alternative objective which reduces computational cost and is able to preserve the quality of the disaggregation. We have suggested a function 'min excess' that performs just as well, for lower computational burden. Further inverse functions could be considered in future research.

Regarding the insights of the disaggregated results, we formulate only conclusions retrieved from the method that performs

best as the presented disaggregation methods have deviations in their quantitative results. The 'best' method is considered the one which yields the lowest amount of loadshedding. From the findings presented in this Chapter, we can stress that modelling Europe at a resolution of one node per country is insufficient to retrieve reliable capacity expansion suggestions. Moreover, results retrieved from models clustered to a spatial resolution of around 100-200 nodes using state-of-the-art evaluated aggregation methods fail to cover approximately 100 TWh of Europe's electricity demand, approximating $3-5\%$ of its annual consumption. Instead of consuming the excess electricity, curtailment rates rise by approximately the shed amount, additional to what would have been expected from an economic optimum. Our analysis reveals that the electricity shortage is due to local transmission constraints. Spatially low-resolved models assume that power can be transferred without limit to all locations that are represented within a single node. Therefore, intra-nodal transmission constraints are ignored in the constraints of the aggregated model. Thus, disaggregated results into higher spatial detail are confronted with power flow restrictions, resulting in transmission congestion and imply necessary loadshedding measures, making the investment decisions retrieved from a coarse model technically infeasible. However, the insight from studies dealing with spatial clustering and disaggregation provide insights that might help improve modelling results by developing methods to calibrate the low-resolved models.

6.5 CRITICAL APPRAISAL

Removing the set of optimisation variables that accounts for the capacity expansion allows solving an operational dispatch model at a higher model resolution. Nevertheless, the presented operational model results are based on model runs retrieved from a model resolution of 1250 nodes, i.e. approximately 25% of the original network size, due to a persisting computational burden. The results presented in this Chapter are likely to intensify if the operational model was spatially highly-resolved and, thus, strengthen our main argument.

While we have analysed methods to disaggregate the suggested optimal generation fleet, this study did not investigate methods to disaggregate transmission capacity expansion modelling results, or how additional transfer capacity obtained from a transmission expansion problem could improve the overall results. Such an analysis could build on our presented methods and extend them on additional optimisation variables. Moreover, all results presented

in this Chapter carried out for a fully self-sufficient and fully renewable Europe. Lowering the carbon emission target could relax the findings and would not make as strong implications. Therefore, in a future study, different carbon emission targets could be analysed more carefully.

Results of this study are all based on the MIT-licenced models PyPSA v0.18.0 and PyPSA-Eur v0.3.0. Therefore, nearly all of the limitations that apply for this version of the model also apply for this study. These include for example retrieving optimal capacities that rely on weather data from a single weather year, applying only a linearised power flow model or neglecting dynamic line rating. Some of these simplifications might improve in the future.

7

CONCLUSIONS

Electricity system optimisation models at transmission level for planning applications are laborious to build due to many complex aspects of the system. This is partly because of the problem of (open-source) data-collection at high spatial and temporal granularity. But even after the formal model is built, computational restrictions hinder the modelling to be conducted at a scale that is desirable. To circumvent this difficulty with respect to the models spatial resolution, researchers often model large regions coarsely, meaning at a low spatial granularity to reduce the number of mathematical constraints on the computational side. However, spatial aggregation is a compromise whose impacts we have analysed in this dissertation.

RESEARCH CONTRIBUTIONS

In Chapter 3 we have studied the effects of spatial resolution on two dominating attributes of the electricity optimisation model: variable renewable resource sites and transmission lines. We have introduced a novel methodology together with a study-design that allowed us to disentangle and to quantify these individual effects on the compound system. This approach allows us to understand their simultaneous effect on the simulation results. We have learned that allowing the model to site renewable capacity at a highly-resolved model significantly reduces the resulting system costs by approximately 10%. This is because sites with good yield can be better saturated with renewable generators. On the other hand, when the model is spatially higher resolved at transmission level and is thus subject to more transmission constraints, system costs increase more strongly by approximately 20%. This is because in this setting the high-yield sites can not be saturated anymore due to congestion in the grid. When global expansion of the grid is allowed, some of the effects are less dominant and total system costs de-crease by approximately 15%.

Based on these findings, we have improved existing spatial clustering methods and developed novel ones. We have benchmarked them respectively against commonly used methods from the literature in Chapter 4. Our results indicate that applying the commonly used k-means algorithm performs significantly worse than methods that are tailored specifically to the electricity system model

at hand. Clustering methods that are better suited to represent the system at a lower spatial level are of hierarchical nature because they can better represent the topology of the network grid. With respect to a high carbon reduction target, we have seen that models that were clustered based on a feature that incorporates renewable potentials, such as wind speeds or solar radiation, perform better than models that were clustered based on a measure on electrical distance. When a model is built to simulate a system dominated by conventional generation, the clustering based on electrical parameters of the grid prevailed.

We have used the findings of the previous chapters to show that spatially low-resolved modelling results are infeasible when disaggregated at a high spatial resolution in Chapter 6. This is even true when we apply the most effective spatial clustering method. We measure infeasibilites as missing renewable electricity to cover demand. In other words, in a fully renewable electricity model, an infeasibility is defined as power generated from a gas turbine because no renewable asset (either generation or storage) is available to meet the consumption needs at any point in time of the simulation. This is although the same amount of installed renewable capacity was sufficient to cover demands in the spatially coarse model. In agreement with the results found in Chapters 3 and 4, the amount of the infeasibility strongly depends on the spatially low-resolved reference result and is mainly induced by the additional transmission constraints in the highly-resolved model formulation. As model resolution increases, the infeasibility rapidly decreases, but stabilises for European planning simulations at an approximate range of 90 to 150 nodes, depending on the simulated decarbonisation goal. When further increasing the spatial model resolution, the trade-off between additional accuracy of the modelling results and increased computational complexity becomes marginal. However, the spatially low-resolved results must still be treated with caution, as they indicate that $5-10\%$ of electricity consumption needs can not be covered in a spatially more complex system.

INFORMATICS CONTRIBUTIONS

We have learned that spatially low-resolved electricity system optimisation models for capacity planning applications imply a significant error. Particularly a model resolution of one node per country is incapable to deduce plausible, feasible and implementable solutions if no subsequent a posteriori adaptations, such as additional transmission expansion, can be assumed. To improve the spatially low-resolved modelling results, we have designed a novel clus-

tering design than can be applied on electricity system planning models to allow different resolutions on siting renewable capacity and the transmission grid. This novel design builds on existing clustering methods that are interchangeable on a case to case basis. We have contributed the code to the widely used open source modelling frameworks *PyPSA* and *PyPSA-eur*. First spin-off applications have been adopted in the affiliated modelling framework *PyPSA-Earth*. The novel clustering design allowed us to study dominant features of the electricity grid which we used to develop novel and improved clustering methods, specifically tailored for capacity planning applications. We contributed these novel methods again to the code basis of *PyPSA*, and have improved the functionalities of established methods of the free software library *NetworkX*. Finally, we have released a stand-alone spin-off python framework *InversE* to disaggregate spatially low-resolved modelling results derived from clustered networks using any clustering method. The framework provides methods that were previously introduced in the engineering literature, but that were never released in form of open source, reproduceable code. The framework also provides a completely novel disaggregation method that compromises computational resources and accuracy in the solution of the hihgly-resolved objective function, compared to previously presented inverse methods.

OUTLOOK AND FUTURE WORK

Moving away from spatially simplified and clustered models is implausible because of the persisting computational burden. This is particularly true if open-source solvers remain unfunded and scarce. But even if there will be advances in open-source solving algorithms, we have seen that even highly effective commercial solvers limit us from solving spatially highly-resolved models at a desirable spatial granularity due to high requirements of direct access storage and the time intensive solving process. But as shares of renewable generation will continue to rise over the years to come, more accurate investment decisions must be made to circumvent high and unnecessary expenses due to curtailment or redispatch measures, that in the worst case must use carbon intensive fuels to perform the management. Therefore, further investigation is necessary to derive a reliable future design of the grid. This could be achieved by improving modelling at a low spatial resolution using calibration methods to incorporate the lessons from, for example, this thesis into the spatially low-resolved model formulation.

To evaluate spatially low-resolved modelling results, we have proposed first suggestions of disaggregation methods for capacity planning models, but have omitted the disaggregation of coarse transmission expansion projects. Extended methods to invert the coarsely derived transmission reinforcements could be investigated in a follow-up study. Particularly if novel methods to calibrate the spatially low-resolved models emerge, because the disaggregation can be highly effective for model evaluation purposes.

In addition to model calibration and disaggregation methods for electricity system planning, similar approaches could be pursued for sector coupled models. It is likely that the conclusions derived in this dissertation also hold true for different sectors, particularly when decarbonisation is achieved through the electrification of other sectors. In the transportation sector, for example, decarbonisation could be achieved by switching from combustion engines to electric cars, or, in the heating sector, central heating that is mainly fueled by coal, oil, or natural gas could be replaced with electric heat pumps. Electrification of the different sectors have impacts on the patterns of electricity demand, and might consolidate peaks in the inelastic electricity demand, smooth them over time, for example if we can assume bidirectional charging of electric vehicles, or seasonally intensify particularly with respect to the heating sector. This changing behavior can have implications on the spatial resolution in the modelling process and call for detailed analyses to better quantify the impact.

The psychological profiling [of a programmer] is mostly the ability to shift levels of abstraction, from low level to high level. To see something in the small and to see something in the large. — **Donald E. Knuth**

ACKNOWLEDGMENTS

Along the course of this journey I have met a great number of people and would like to thank each and every one of them.
First and most, I thank my supervisor, Prof. **Thomas Brown**, for employing my younger and much more naïve self to work with his initially small *Energy System Modelling* group at KIT. Although they eventually moved away to grow a larger and welcoming community of researchers and open source modellers, I am proud to be part of the extended team. In addition to his professional and highly qualitative council, I would like to thank him for his support in all difficult situations over the last years.
I thank Prof. **Veit Hagenmeyer**, **Andreas Hofmann** and Prof. **Klemens Böhm** who have supported the transition when Prof. Brown left KIT. Particularly Prof. Hagenmeyer for taking over the supervision of my doctoral studies, all philosophical conversations, and his highly appreciated and helpful feedback on my research. I thank Prof. Böhm for associating me with the graduate school "Energy Status Data" which certainly has helped feeling more involved at KIT.
I thank **Fabian Neumann** for getting me started during my first weeks at the Institute and **Elisabeth Zeyen** for her lovely hospitality during many of my stays in Berlin.
I thank **Nicole Ludwig** and **Kaleb Phipps** for being dear friends during times when I felt lonely at the Institute, mostly during times when the 2nd floor in building 449 - the *Antarctic of the institute* - was desolated.
I thank my colleagues **Anthony Britto**, **Philipp Böttcher**, **Clara Büttner**, **Ilka Cussmann**, **Amin Shokri**, **Cagla Fadillioglu**, **Davide Fioriti**, **Phillip Glaum**, **Claudia Greceanu**, **Aleks Grochowicz**, **Koen van Greevenbroek**, **Sascha Gritzbach**, **Johannes Hampp**, **Fabian Hofmann**, **Jonas Hörsch**, **Richard Jumar**, **Franz Kaiser**, **Thorben Knust**, **Johannes Kruse**, **Lieke van der Most**, **Maximilian Parzen**, **Lisa Petani**, **Iegor Riepin**, **Maximilian Roithner**, **Dominique Sauer**, **Mirko Schäfer**, **Christoph Tries**, **Jan Unnewehr**, **Dirk Witthaut**, **Matthias Wolf** and all others I have met for conversations, comments, insights, small-talk, lunch and coffee.
I thank **Andreas Hahmann** for setting up the finances of my research visit. I thank Prof. **Jordan Kern** for welcoming me to his group at the North Carolina State University. Thank you for your support in organising seminars at NCSU, UNC and Duke during my visit. I thank **Ece Akdemir**, **Kerem Akdemir**, **Kyle Bradbury**, **Anna Cybulsky**, **James Glynn**, **Trey Gowdy**, **Cameron Lisy**, **Dalia Patino-Echeverri**, **Paul Sizaire** and **Henry Ssembatya** for valuable discussions and shaping my stay overseas.
On a more personal note, I thank **Ronja Brixel**, **Monja Wolf** and my dog **Snoopy** for making me feel at home in Karlsruhe.
I thank my polish family and my parents **Jolanta** and **Wojciech Frysztacki** for their endless hospitality, food, efforts in dog-sitting, and enduring any of my very volatile moods during this journey, ranging all across excitement, grief, concern, pride, anger, frustration, relief, self-doubt, surprise and *bliss*.

All of you have certainly shaped my work and personality positively.

Cheers!

[1] Martha Frysztacki and Tom Brown. "Modeling Curtailment in Germany: How Spatial Resolution Impacts Line Congestion." In: *2020 17th International Conference on the European Energy Market (EEM)*. 2020, pp. 1–7. DOI: https://doi.org/10.1109/EEM49802.2020.9221886.

[2] Martha Maria Frysztacki, Jonas Hörsch, Veit Hagenmeyer, and Tom Brown. "The strong effect of network resolution on electricity system models with high shares of wind and solar." In: *Applied Energy* 291 (2021), p. 116726. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2021.116726.

[3] Martha Maria Frysztacki, Gereon Recht, and Tom Brown. "A comparison of clustering methods for the spatial reduction of renewable electricity optimisation models of Europe." In: *Energy Informatics* 5.4 (2022). ISSN: 2520-8942. DOI: https://doi.org/10.1186/s42162-022-00187-7.

[4] Martha Maria Frysztacki, Veit Hagenmeyer, and Tom Brown. "Inverse methods: How feasible are spatially low-resolved capacity expansion modeling results when dis-aggregated at high resolution?" In: *submitted to Energy* (under review). DOI: https://doi.org/10.48550/arXiv.2209.02364.

[5] Kaleb Phipps, Maximilian Beichter, Martha Frysztacki, Ralf Mikut, Veit Hagenmeyer, and Nicole Ludwig. "Net load forecasting using different aggregation levels." In: *Energy Informatics*. Vol. 5(S1). 19. 2022. DOI: https://doi.org/10.1186/s42162-022-00213-8.

[6] Amin Shokri Gazafroudi, Elisabeth Zeyen, Martha Frysztacki, Fabian Neumann, and Tom Brown. "Long-Term Benefits for Renewables Integration of Network Boosters for Corrective Grid Security." In: *submitted to International Journal of Electrical Power & Energy Systems* (under review). DOI: https://doi.org/10.48550/arXiv.2112.06667.

[7] Maximilian Parzen et al. "PyPSA-Earth. A New Global Open Energy System Optimization Model Demonstrated in Africa." In: *submitted to Applied Energy* (under review). DOI: https://doi.org/10.48550/arXiv.2209.04663.

[8] Franz Kaiser, Johannes Kruse, Philipp C. Böttcher, Martha Frysztacki, Tom Brown, and Dirk Witthaut. "Cascading Failures and Critical Infrastructures in Future Renewable Power Systems." In: (in preparation).

[9] Oskar Oliven. "Europas Großkraftlinien: Vorschlag eines europäischen Höchstspannungsnetzes." In: *Zeitschrift des Vereines Deutscher Ingenieure* 74.25 (1930), 875–879.

[10] Matthew H. Brown and Richard P. Sedano. *Electricity Transmission. A Primer*. National Council on Electric Policy. ISBN: ISBN 1-58024-352-5.

[11] Vincent Lagendijk. "'To Consolidate Peace'? The International Electro-technical Community and the Grid for the United States of Europe." In: *Journal of Contemporary History* 47.2 (2012), pp. 402–426. DOI: https://doi.org/10.1177/0022009411431722.

[12] Gernot Nischler. "Zukunftsorientierte elektrizitätswirtschaftliche Netzentwicklung." deutsch. PhD thesis. Graz University of Technology, 2014. URL: https://graz.pure.elsevier.com/de/publications/zukunftsorientierte-elektrizit%C3%A4tswirtschaftliche-netzentwicklung.

[13] Vaclav Smil. *Energy Transitions: Global and National Perspectives*. 2022. URL: https://ourworldindata.org/global-energy-200-years.

[14] *Median temperature anomaly from 1961-1990 average*. 2020. URL: https://ourworldindata.org/grapher/temperature-anomaly.

[15] Bent Sørensen. "A plan is outlined according to which solar and wind energy would supply Denmark's needs by the year 2050." In: *Science* 189.4199 (1975), 255–260. DOI: https://doi.org/10.1126/science.189.4199.255.

[16] United Nations Treaty Collection. "Paris Agreement." In: Chapter XXVII 7. d (adopted 12/2015, inforced 11/2016). URL: https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-d&chapter=27&clang=_en (visited on 09/01/2022).

[17] European Comission. *What is the European Green Deal?* 2019. DOI: https://doi.org/10.2775/97540.

[18] Europäische Union. "Third Energy Package." In: *Amtsblatt der Europäischen Union* Nr. L 211 (adopted 07/2009, inforced 08/2009), pp. 1–136.

[19] European association for the cooperation of transmission system operators (TSOs) for electricity. *Transmission System Map*. 2022. URL: https://www.entsoe.eu/about/inside-entsoe/objectives/.

[20] T. Brown, T. Bischof-Niemz, K. Blok, C. Breyer, H. Lund, and B.V. Mathiesen. "Response to 'Burden of proof: A comprehensive review of the feasibility of 100% renewable-electricity systems'." In: *Renewable and Sustainable Energy Reviews* 92 (2018), pp. 834 –847. ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2018.04.113.

[21] Telekommunikation Post und Eisenbahnen Bundesnetzagentur für Elektrizität Gas. *Monitoringbericht 2019*. 2020.

[22] *eHighways 2050 Final Reports*. Tech. rep. ENTSO-E et al., 2015.

[23] European Network of Transmission System Operators for Electricity. *Ten-Year Network Development Plan (TYNDP) 2018*. Tech. rep. ENTSO-E, 2018.

[24] Jörg Mühlenhoff, Jonathan Bonadio, and Others. *Building a Paris Agreement Compatible (PAC) energy scenario. CAN Europe/EEB technical summary of key elements*. Tech. rep. Climate Achtion Network Europe and European Environmental Bureau, 2020. URL: https://www.pac-scenarios.eu/.

[25] Manish Ram et al. *Global Energy System based on 100Heat, Transport and Desalination Sectors*. Tech. rep. Lappeenranta-Lahti University of Technology (LUT University) and Energy Watch Group (EWG), 2019.

[26] IPCC. *Global warming of 1.5C*. Tech. rep. 2018.

[27] Adam Hawkes Stefan Pfenninger and James Keirstead. "Energy systems modeling for twenty-first century energy challenges." In: *Renewable and Sustainable Energy Reviews* 22 (2014), pp. 74–86. DOI: https://doi.org/10.1016/j.rser.2014.02.003.

[28] Darryl R. Biggar and Mohammad Reza Hesamzadeh, eds. *The Economics of Electricity Markets*. John Wiley & Sons Ltd, 2014. DOI: 10.1002/9781118775745. URL: https://doi.org/10.1002/9781118775745.

[29] Peter Lopion, Peter Markewitz, Martin Robinius, and Detlef Stolten. "A review of current challenges and trends in energy systems modeling." In: *Renewable and Sustainable Energy Reviews* 96 (2018), pp. 156–166. ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2018.07.045. URL: https://www.sciencedirect.com/science/article/pii/S1364032118305537.

[30] Jonas Hörsch, Henrik Ronellenfitsch, Dirk Witthaut, and Tom Brown. "Linear optimal power flow using cycle flows." In: *Electric Power Systems Research* 158 (2018), pp. 126–135. ISSN: 0378-7796. DOI: https://doi.org/10.1016/j.epsr.2017.12.034. URL: https://www.sciencedirect.com/science/article/pii/S0378779617305138.

[31] "A modeler's guide to handle complexity in energy systems optimization, journal = Advances in Applied Energy." In: 4 (2021), p. 100063. ISSN: 2666-7924. DOI: https://doi.org/10.1016/j.adapen.2021.100063.

[32] Fabian Neumann. "Computational and Near-Optimal Trade-Offs in Renewable Electricity System Modelling." PhD thesis. Karlsruhe Institute of Technology, 2021. DOI: https://doi.org/10.5445/IR/1000137267.

[33] Christian Milan, Michael Stadler, Gonçalo Cardoso, and Salman Mashayekh. "Modeling of non-linear CHP efficiency curves in distributed energy systems." In: *Applied Energy* 148 (2015), pp. 334–347. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2015.03.053.

[34] Leander Kotzur, Peter Markewitz, Martin Robinius, and Detlef Stolten. "Impact of different time series aggregation methods on optimal energy system design." In: *Renewable Energy* 117 (2018), pp. 474 –487. ISSN: 0960-1481. DOI: https://doi.org/10.1016/j.renene.2017.10.017.

[35] A. T. D. Perera, Vahid M. Nik, Deliang Chen, Jean-Louis Scartezzini, and Tianzhen Hong. "Quantifying the impacts of climate change and extreme climate events on energy systems." In: *Nature Energy* 5 (2 2020), pp. 150–159. DOI: https://doi.org/10.1038/s41560-020-0558-0.

[36] Hans Christian Gils, Yvonne Scholz, Thomas Pregger, Diego Luca de Tena, and Dominik Heide. "Integrated modelling of variable renewable energy-based power supply in Europe." In: *Energy* 123 (2017), pp. 173 –188. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2017.01.115.

[37] William Zappa, Martin Junginger, and Machteld van den Broek. "Is a 100% renewable European power system feasible by 2050?" In: *Applied Energy* 233-234 (2019), pp. 1027–1050. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2018.08.109.

[38] Michael Child, Claudia Kemfert, Dmitrii Bogdanov, and Christian Breyer. "Flexible electricity generation, grid exchange and storage for the transition to a 100% renewable energy system in Europe." In: *Renewable Energy* 139 (2019),

pp. 80–101. ISSN: 0960-1481. DOI: https://doi.org/10.1016/j.renene.2019.02.077.

[39] Elisabeth Zeyen, Veit Hagenmeyer, and Tom Brown. "Mitigating heat demand peaks in buildings in a highly renewable European energy system." In: *Energy* 231 (2021), p. 120784. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2021.120784.

[40] Jan-Phillipp Sasse and Evelina Trutnevyte. "Regional impacts of electricity system transition in Central Europe until 2035." In: *Nature Communications* 11 (2020). DOI: https://doi.org/10.1038/s41467-020-18812-y.

[41] Francesco Lombardi, Bryn Pickering, Emanuela Colombo, and Stefan Pfenninger. "Policy Decision Support for Renewables Deployment through Spatially Explicit Practically Optimal Alternatives." In: *Joule* 4.10 (2020), pp. 2185–2207. ISSN: 2542-4351. DOI: https://doi.org/10.1016/j.joule.2020.08.002.

[42] Fabian Neumann. "Costs of regional equity and autarky in a renewable European power system." In: *Energy Strategy Reviews* 35 (2021), p. 100652. ISSN: 2211-467X. DOI: https://doi.org/10.1016/j.esr.2021.100652.

[43] Tim Tröndle, Johan Lilliestam, Stefano Marelli, and Stefan Pfenninger. "Trade-Offs between Geographic Scale, Cost, and Infrastructure Requirements for Fully Renewable Electricity in Europe." In: *Joule* (4 2020), pp. 1–20. DOI: https://doi.org/10.1016/j.joule.2020.07.018.

[44] Gregor Czisch. "Szenarien zur zukünftigen Stromversorgung: Kostenoptimierte Variationen zur Versorgung Europas und seiner Nachbarn mit Strom aus erneuerbaren Energien." PhD thesis. University of Kassel, 2005.

[45] R.A. Rodriguez, S. Becker, G. Andresen, D. Heide, and M. Greiner. "Transmission needs across a fully renewable European power system." In: *Renewable Energy* 63 (2014), pp. 467–476. DOI: 10.1016/j.renene.2013.10.005. URL: https://doi.org/10.1016/j.renene.2013.10.005.

[46] David Schlachtberger, Tom Brown, Stefan Schramm, and Martin Greiner. "The benefits of cooperation in a highly renewable European electricity network." In: *Energy* 134 (2017), pp. 469–481. DOI: https://doi.org/10.1016/j.energy.2017.06.004.

[47] Alexander E. MacDonald, Christopher T. M. Clack, Anneliese Alexander, Adam Dunbar, James Wilczak, and Yuanfu Xie. "Future cost-competitive electricity systems and their impact on US CO2 emissions." In: *Nature Climate Change* 6 (2017), pp. 526–531. URL: https://doi.org/10.1038/nclimate2921.

[48] Bart Wiegmans. *GridKit extract of ENTSO-E interactive map*. 2016. DOI: https://doi.org/10.5281/zenodo.55853.

[49] Martha Frysztacki, Lieke van der Most, and Fabian Neumann. *Interannual Electricity Demand Calculator*. 2022. DOI: https://doi.org/10.5281/zenodo.7070438.

[50] Open Power System Data. *Data Package Renewable Power Plants*. DOI: 10.25832/renewable$\$power$\$plants/2020-08-25. URL: https://data.open-power-system-data.org/renewable_power_plants/.

[51] Jonas Hörsch, Fabian Hofmann, David Schlachtberger, and Tom Brown. "PyPSA-Eur: An Open Optimisation Model of the European Transmission System." In: *Energy Strategy Reviews* 22.v3 (2018), pp. 207–215. DOI: https://doi.org/10.1016/j.esr.2018.08.012.

[52] Open Power System Data. *Time Series of Load in Hourly Resolution*. 2019. DOI: 10.25832/time$_$series/2019-06-05. URL: https://doi.org/10.25832/time_series/2019-06-05.

[53] Fabian Hofmann, Johannes Hampp, Fabian Neumann, Tom Brown, and Jonas Hörsch. "atlite: A Lightweight Python Package for Calculating Renewable Power Potentials and Time Series." In: *Journal of Open Source Software* 6.62 (2021), p. 3294. DOI: https://doi.org/10.21105/joss.03294.

[54] *Technology Data for Generation of Electricity and District Heating, Energy Storage and Energy Carrier Generation and Conversion*. Tech. rep. Danish Energy Agency and Energinet.dk, 2019. URL: https://ens.dk/en/our-services/projections-and-models/technology-data.

[55] Andreas Schröder, Friedrich Kunz, Jan Meiss, Roman Mendelevitch, and Christian Hirschhausen. *Current and prospective costs of electricity generation until 2050*. Tech. rep. Chapter 2. Deutsches Institut für Wirtschaftsforschung (DIW), 2013. DOI: 10419/80348.

[56] Cory Budischak, DeAnna Sewell, Heather Thomson, Leon Mach, Dana E. Veron, and Willett Kempton. "Cost-minimized combinations of wind power, solar power and electrochemical storage, powering the grid up to 99.9% of the time." In: *Journal of Power Sources* 225 (2013), pp. 60 –74. ISSN: 0378-7753. DOI: 10.1016/j.jpowsour.2012.09.054. URL: https://doi.org/10.1016/j.jpowsour.2012.09.054.

[57] Eero Vartiainen, Gaëtan Masson, and Christian Breyer. *The True Competitiveness of Solar PV: A European Case Study*. Tech. rep. European Technology and Innovation Platform for Photovoltaics, 2017. URL: http://www.etip-pv.eu/fileadmin/Documents/ETIP_PV_Publications_2017-2018/LCOE_Report_March_2017.pdf.

[58] Qiong Zhou and J.W. Bialek. "Approximate model of European interconnected system as a benchmark system to study effects of cross-border trades." In: *IEEE Transactions on Power Systems* 20.2 (2005), pp. 782–788. DOI: https://doi.org/10.1109/TPWRS.2005.846178.

[59] David Severin Ryberg, Martin Robinius, and Detlef Stolten. "Evaluating Land Eligibility Constraints of Renewable Energy Sources in Europe." In: *Energies* 11.5 (2018). ISSN: 1996-1073. DOI: https://doi.org/10.3390/en11051246. URL: https://www.mdpi.com/1996-1073/11/5/1246.

[60] Yvonne Scholz. "Renewable energy based electricity supply at low costs - Development of the REMix model and application for Europe." PhD thesis. University of Stuttgart, 2012.

[61] European Centre for Medium-Range Weather Forecasts (ECMWF). *ERA5 Reanalysis*. URL: {https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation}.

[62] Uwe Pfeifroth, Steffen Kothe, Richard Müller, Jörg Trentmann, Rainer Hollmann, Petra Fuchs, and Martin Werscheck. "Surface Radiation Data Set - Heliosat (SARAH) - Edition 2." In: (2017). DOI: 10.5676/EUM$_$SAF$_$CM/SARAH/V002.

[63] Tom Brown, Jonas Hörsch, and David Schlachtberger. "PyPSA: Python for Power System Analysis." In: *Journal of Open Research Software* 6 (2018), p. 4. DOI: https://doi.org/10.5334/jors.188.

[64] Heinz Stigler. *Gutachten zur Ermittlung des erforderlichen Netzausbaus im deutschen Übertragungsnetz*. Tech. rep. Chapter 2. Technische Universität Graz, 2012. URL: https://www.

tugraz.at/fileadmin/user_upload/Institute/IEE/files/Endbericht_NEMO_II.pdf.

[65] Bernhard Fuchs, Andreas Guido Roehder, Moritz Mittelstaedt, Janek Massmann, Hendrik Natemeyer, and Armin Schnettler. *Studie zu Aspekten der elektrischen Systemstabilität im deutschen Übertragungsnetz bis 2023*. Tech. rep. Chapter 2. RWTH Aachen University, 2015. URL: http://publications.rwth-aachen.de/record/538613/files/.

[66] Victor Klee and George J. Minty. "How good is the simplex algorithm?" In: *Inequalities, III (Proc. Third Sympos., Univ. California, Los Angeles, Calif., 1969; dedicated to the memory of Theodore S. Motzkin)*. Academic Press, New York, 1972, pp. 159–175.

[67] N. Karmarkar. "A New Polynomial-Time Algorithm for Linear Programming." In: *Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing*. New York, NY, USA: Association for Computing Machinery, 1984, 302–311. ISBN: 0897911334. DOI: https://doi.org/10.1145/800057.808695.

[68] Maximilia Parzen, Julian Hall, Jesse Jenkins, and Tom Brown. *Optimization solvers: the missing link for a fully open-source energy system modelling ecosystem*. DOI: https://doi.org/10.5281/zenodo.6409432.

[69] Gurobi Optimzation. URL: https://www.gurobi.com/.

[70] John Forrest et al. *coin-or/Cbc*. URL: https://zenodo.org/record/6522795#.Y3dZJyYo9hE.

[71] Andrew Makhorin. *GLPK (GNU Linear Programming Kit)*. URL: https://www.gnu.org/software/glpk/glpk.html.

[72] Julian Hall, Ivet Galabova, and Michael Feldmeier. *HiGHS - high performance software for linear optimization*. URL: https://highs.dev/.

[73] Vladimir Estivill-Castro. "Why so Many Clustering Algorithms: A Position Paper." In: 4.1 (2002), 65–75. ISSN: 1931-0145. DOI: https://doi.org/10.1145/568574.568575.

[74] J. A. Hartigan and M. A. Wong. "Algorithm AS 136: A K-means clustering algorithm." In: *Applied Statistics* 28.1 (1979), pp. 100–108. DOI: https://doi.org/10.2307/2346830.

[75] A. J. Hoffman W. E. Donath. "Lower bounds for the partitioning of graphs." In: *IBM Journal of Research and Development* 17.5 (1973), pp. 420–425.

[76] M. Fiedler. "Algebraic connectivity of graphs." In: *Czechoslovak Mathematical Journal* 23.2 (1973), 298–305.

[77] James C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer New York, 1981.

[78] A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum Likelihood from Incomplete Data Via the EM Algorithm." In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (1977), pp. 1–22. DOI: https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.

[79] Sina Khanmohammadi, Naiier Adibeig, and Samaneh Shanehbandy. "An improved overlapping k-means clustering method for medical applications." In: *Expert Systems with Applications* 67 (2017), pp. 12–18. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2016.09.025.

[80] Ray Galvin. "Trouble at the end of the line: Local activism and social acceptance in low-carbon electricity transmission in Lower Franconia, Germany." In: *Energy Research & Social Science* 38 (2018), pp. 114 –126. ISSN: 2214-6296. DOI: https://doi.org/10.1016/j.erss.2018.01.022.

[81] Jed J. Cohen, Johannes Reichl, and Michael Schmidthaler. "Re-focussing research efforts on the public acceptance of energy infrastructure: A critical review." In: *Energy* 76 (2014), pp. 4 –9. ISSN: 0360-5442. URL: https://doi.org/10.1016/j.energy.2013.12.056.

[82] European Commission. "Energy Roadmap 2050." In: *COM. 2011* 885/2 (12/2011).

[83] A. K. Jain, M. N. Murty, and P. J. Flynn. "Data Clustering: A Review." In: *ACM Comput. Surv.* 31.3 (1999), pp. 264–323. ISSN: 0360-0300. DOI: https://doi.org/10.1145/331499.331504.

[84] S. Blumsack, P. Hines, M. Patel, C. Barrows, and E. Cotilla Sanchez. "Defining power network zones from measures of electrical distance." In: *2009 IEEE Power & Energy Society General Meeting*. 2009, pp. 1–8. DOI: https://doi.org/10.1109/PES.2009.5275353.

[85] Camille Hamon, Ebrahim Shayesteh, Mikael Amelin, and Lennart Söder. "Two partitioning methods for multi-area studies in large power systems." In: *International Transactions on Electrical Energy Systems* 25.4 (2015), pp. 648–660. ISSN: 2050-7038. DOI: http://doi.org/10.1002/etep.1864.

[86] Xu Cheng and T. J. Overbye. "PTDF-based power system equivalents." In: *IEEE Transactions on Power Systems* 20.4 (2005), pp. 1868–1876. ISSN: 0885-8950. DOI: https://doi.org/10.1109/TPWRS.2005.857013.

[87] H. K. Singh and S. C. Srivastava. "A reduced network representation suitable for fast nodal price calculations in electricity markets." In: *IEEE Power Engineering Society General Meeting, 2005*. 2005, 2070–2077 Vol. 2. DOI: https://doi.org/10.1109/PES.2005.1489092.

[88] E. Shayesteh, B. F. Hobbs, L. Söder, and M. Amelin. "ATC-Based System Reduction for Planning Power Systems With Correlated Wind and Loads." In: *IEEE Transactions on Power Systems* 30.1 (2015), pp. 429–438. ISSN: 0885-8950. DOI: https://doi.org/10.1109/TPWRS.2014.2326615.

[89] Kais Siala and Mohammad Youssef Mahfouz. "Impact of the choice of regions on energy system models." In: *Energy Strategy Reviews* 25 (2019), pp. 75–85. ISSN: 2211-467X. DOI: https://doi.org/10.1016/j.esr.2019.100362.

[90] Kais Siala. "Spatial Complexity in Energy System Modeling." PhD thesis. Technical University of Munich, 2021.

[91] Martin Küppers, Christian Perau, Marco Franken, Hans Heger, Matthias Huber, Michael Metzger, and Stefan Niessen. "Data-Driven Regionalization of Decarbonized Energy Systems for Reflecting Their Changing Topologies in Planning and Optimization." In: *Energies* 13 (Aug. 2020), p. 4076. DOI: https://doi.org/10.3390/en13164076.

[92] Egerer, J., Lorenz, C., and Gerbaulet, C. "European electricity grid infrastructure expansion in a 2050 context." In: *10th International Conference on the European Energy Market*. IEEE, 2013, pp. 1–7. DOI: https://doi.org/10.1109/EEM.2013.6607408.

[93] Denis Hess, Manuel Wetzel, and Karl-Kiên Cao. "Representing node-internal transmission and distribution grids in energy system models", journal = Renewable Energy." In: 119 (2018), pp. 874 –890. ISSN: 0960-1481. DOI: https://doi.org/10.1016/j.renene.2017.10.041.

[94] Christian Nabe. "Impacts of Restricted Transmission Grid Expansion in a 2030 Perspective in Germany." In: *Wind Integration Workshop, Berlin*. 2014.

[95] *Kombikraftwerk 2: Abschlussbericht*. Tech. rep. Fraunhofer IWES et al., 2014. URL: \url{http://www.kombikraftwerk.de/mediathek/abschlussbericht.html}.

[96] Brown, T., Schierhorn, P., Tröster, E., and Ackermann, T. "Optimising the European transmission system for 77% renewable electricity by 2030." In: *IET Renewable Power Generation* 10.1 (2016), pp. 3–9.

[97] M. Schäfer, S. Bugge Siggaard, Kun Zhu, C. Risager Poulsen, and M. Greiner. "Scaling of transmission capacities in coarse-grained renewable electricity networks." In: *EPL (Europhysics Letters)* 119.3 (2017), p. 38004. DOI: https://doi.org/10.1209/0295-5075/119/38004.

[98] Wesley Cole et al. *Variable Renewable Energy in Long-Term Planning Models: A Multi-Model Perspective*. Tech. rep. Nov. 2017. DOI: 10.2172/1416124. URL: https://www.osti.gov/biblio/1416124.

[99] V. Krishnan and W. Cole. "Evaluating the value of high spatial resolution in national capacity expansion models using ReEDS." In: *2016 IEEE Power and Energy Society General Meeting (PESGM)*. 2016, pp. 1–5. DOI: 10.1109/PESGM.2016.7741996.

[100] Jonas Hörsch and Tom Brown. "The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios." In: *14th International Conference on the European Energy Market (EEM)*. 2017, pp. 1–7. DOI: https://doi.org/10.1109/EEM.2017.7982024.

[101] Jonas Hörsch. "Spatial scaling in renewable energy networks." PhD thesis. Goethe University Frankfurt, 2018.

[102] Eduardo Cotilla-Sanchez, Paul D. H. Hines, Clayton Barrows, Seth Blumsack, and Mahendra Patel. "Multi-Attribute Partitioning of Power Networks Based on Electrical Distance." In: *IEEE Transactions on Power Systems* 28.4 (2013), pp. 4979–4987. DOI: https://doi.org/10.1109/TPWRS.2013.2263886.

[103] Wolfgang Biener and Klaus René Garcia Rosas. "Grid reduction for energy system analysis." In: *Electric Power Systems Research* 185 (2020), p. 106349. ISSN: 0378-7796. DOI: https://doi.org/10.1016/j.epsr.2020.106349.

[104] H.K. Temraz, M.M.A. Salama, and V.H. Quintana. "Application of partitioning techniques for decomposing large-scale electric power networks." In: *International Journal of Electrical Power & Energy Systems* 16.5 (1994), pp. 301–309. ISSN: 0142-0615. DOI: https://doi.org/10.1016/0142-0615(94)90034-5.

[105] Chiara Scaramuzzino, Giulia Garegnani, and Pietro Zambelli. "Integrated approach for the identification of spatial patterns related to renewable energy potential in European territories." In: *Renewable and Sustainable Energy Reviews* 101 (2019), pp. 1–13. ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2018.10.024.

[106] HyungSeon Oh. "A New Network Reduction Methodology for Power System Planning Studies." In: *IEEE Transactions on Power Systems* 25.2 (2010), pp. 677 –684. DOI: 10.1109/TPWRS.2009.2036183.

[107] Di Shi and Daniel J. Tylavsky. "A Novel Bus-Aggregation-Based Structure-Preserving Power System Equivalent." In: *IEEE Transactions on Power Systems* 30.4 (2015), pp. 1977–1986. DOI: https://doi.org/10.1109/TPWRS.2014.2359447.

[108] D.P. Schlachtberger, T. Brown, M. Schäfer, S. Schramm, and M. Greiner. "Cost optimal scenarios of a future highly renewable European electricity system: Exploring the influence of weather data, cost parameters and policy constraints." In: *Energy* 163 (2018), pp. 100 –114. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2018.08.070.

[109] Fabian Neumann and Tom Brown. "The near-optimal feasible space of a renewable power system model." In: *Electric Power Systems Research* 190 (2021), p. 106690. ISSN: 0378-7796. DOI: https://doi.org/10.1016/j.epsr.2020.106690.

[110] Tom Brown, David Schlachtberger, Alexander Kies, Stefan Schramm, and Martin Greiner. "Synergies of sector coupling and transmission reinforcement in a cost-optimised, highly renewable European energy system." In: *Energy* 160 (2018), pp. 720–739. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2018.06.222.

[111] Lina Reichenberg, Fredrik Hedenus, Mikael Odenberger, and Filip Johnsson. "The marginal system LCOE of variable renewables – Evaluating high penetration levels of wind and solar in Europe." In: *Energy* 152 (2018), pp. 914 –924. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2018.02.061.

[112] Niclas Mattsson, Vilhelm Verendel, Fredrik Hedenus, and Lina Reichenberg. *An autopilot for energy models – automatic generation of renewable supply curves, hourly capacity factors and hourly synthetic electricity demand for arbitrary world regions*. Tech. rep. 2020. DOI: https://doi.org/10.48550/arXiv.2003.01233.

[113] J. B. Ward. "Equivalent Circuits for Power-Flow Studies." In: *Transactions of the American Institute of Electrical Engineers* 68.1 (1949), pp. 373–382. DOI: https://doi.org/10.1109/T-AIEE.1949.5059947.

[114] Di Shi, Daniel L. Shawhan, Nan Li, Daniel J. Tylavsky, John T. Taber, Ray D. Zimmerman, and William D. Schulze. "Optimal generation investment planning: Pt. 1: network equivalents." In: *2012 North American Power Symposium (NAPS)*. 2012, pp. 1–6. DOI: https://doi.org/10.1109/NAPS.2012.6336375.

[115] Christian Etienne Fleischer. "Minimising the effects of spatial scale reduction on power system models." In: *Energy Strategy Reviews* 32 (2020), p. 100563. ISSN: 2211-467X. DOI: https://doi.org/10.1016/j.esr.2020.100563. URL: https://www.sciencedirect.com/science/article/pii/S2211467X20301164.

[116] David Radu, Antoine Dubois, Mathias Berger, and Damien Ernst. "Model Reduction in Capacity Expansion Planning Problems via Renewable Generation Site Selection." In: *2021 IEEE Madrid PowerTech*. 2021, pp. 1–6. DOI: 10.1109/PowerTech46648.2021.9495027.

[117] J. Oehrlein and J.H. Haunert. "A cutting-planemethod for contiguity-constrained spatial aggregation." In: *Journal of Spatial Information Science* 15 (2017), 89–120. DOI: https://doi.org/10.5311/JOSIS.2017.15.379.

[118] Shruthi Patil, Leander Kotzur, and Detlef Stolten. "Impact of Spatial and Technology Aggregation on Optimal Energy System Design." In: *arXiv* math.OC.1 (2021), pp. 1–18.

[119] David Arthur and Sergei Vassilvitskii. "How Slow is the k-Means Method?" In: *Proceedings of the Twenty-Second Annual Symposium on Computational Geometry*. 2006, 144–153. DOI: https://doi.org/10.1145/1137856.1137880.

[120] David Eppstein. "Fast Hierarchical Clustering and Other Applications of Dynamic Closest Pairs." In: *ACM Journal of Experimental Algorithmics* 5 (2001), 1–es. ISSN: 1084-6654. DOI: https://doi.org/10.1145/351827.351829.

[121] Joe H. Ward Jr. "Hierarchical Grouping to Optimize an Objective Function." In: *Journal of the American Statistical Association* 58.301 (1963), pp. 236–244. DOI: https://doi.org/10.1080/01621459.1963.10500845.

[122] Aaron Clauset, M. E. J. Newman, and Cristopher Moore. "Finding community structure in very large networks." In: *Physical Review E* 70 (6 2004), p. 066111. DOI: https://doi.org/10.1103/PhysRevE.70.066111.

[123] Brian S. Everitt, Sabine Landau, Morven Leese, and Daniel Stahl. "Hierarchical clustering, Agglomerative methods." In: *Cluster Analysis*. Chichester, West Sussex, U.K.: Wiley, 2011. Chap. 4.2, 73–84. ISBN: 978-0-470-74991-3.

[124] "Gesetz für den Ausbau erneuerbarer Energien." In: Bundesgesetzblatt Jahrgang 2000 Teil I.13 (inforced 04/2000), p. 305. URL: https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl100s0305.pdf#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl100s0305.pdf%27%5D__1669219767945 (visited on 0011–2022).

[125] "Gesetz für den Ausbau erneuerbarer Energien." In: Bundesgesetzblatt Jahrgang 2014 Teil I (inforced 08/2014), p. 1066. URL: https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl122s1726.pdf#page=12 (visited on 0011–2022).

[126] *Gesetz für den Ausbau erneuerbarer Energien*. 2014.

[127] "Gesetz für den Ausbau erneuerbarer Energien." In: Bundesgesetzblatt Jahrgang 2014 Teil I (inforced 01/2023), p. 1726. URL: https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl122s1726.pdf#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl122s1726.pdf%27%5D__1669219746922 (visited on 0011–2022).

[128] Telekommunikation Post und Eisenbahnen Bundesnetzagentur für Elektrizität Gas. *EEG in Zahlen 2018*. chapter 9.2.

[129] Julian Brandes, Markus Haun, Daniel Wrede, Patrick Jürgens, Christoph Kost, and Hans-Martin Henning. *Wege zu einem klimaneutralen Energiesystem. Die deutsche Energiewende im Kontext gesellschaftlicher Verhaltensweisen*. Tech. rep. Fraunhofer ISE, 2020.

[130] Bundesministerium für Wirtschaft und Energie. *Die Nationale Wasserstoffstrategie*. 2020. URL: https://www.bmbf.de.

[131] Carl Johan Wallnerström, Yalin Huang, and Lennart Söder. "Impact From Dynamic Line Rating on Wind Power Integration." In: *IEEE Transactions on Smart Grid* 6.1 (2015), pp. 343–350. DOI: https://doi.org/10.1109/TSG.2014.2341353.

[132] Kerstin Maria Rippel, Thomas Wiede, Mario Meinecke, and Regina König. *Netzentwicklungsplan Strom 2030*. Tech. rep. Version 2, Chapter 6. 50Hertz Transmission, Amprion, TenneT TSO & TransnetBW, V 2019. URL: https://netzentwicklungsplan.de/sites/default/files/paragraphs-files/NEP_2030_V2019_1_Entwurf_Teil1_1.pdf/.

[133] Stefan Pfenninger et al. "Opening the black box of energy modelling: Strategies and lessons learned." In: *Energy Strategy Reviews* 19 (2018), pp. 63–71. ISSN: 2211-467X. DOI: https://doi.org/10.1016/j.esr.2017.12.002. URL: https://www.sciencedirect.com/science/article/pii/S2211467X17300809.

[134] *Quartalsbericht zu Netz- und Systemsicherheitsmaßnahmen; Gesamtjahr und Viertes Quartal 2018*. Tech. rep. Chapter 3. Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, 2019.

[135] *Quartalsbericht zu Netz- und Systemsicherheitsmaßnahmen; Erstes Quartal 2019*. Tech. rep. Chapter 2. Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, 2019.

[136] European association for the cooperation of transmission system operators (TSOs) for electricity. *Transmission System Map*. 2018. URL: https://www.entsoe.eu/data/map/.

[137] Olivier Feix, Ruth Obermann, Marius Strecker, and Angela Brötel. *Netzentwicklungsplan Strom 2013*. Tech. rep. Version 2, Chapter 7. 50Hertz Transmission, Amprion, TenneT TSO & TransnetBW, 2013. URL: https://netzentwicklungsplan.de/sites/default/files/paragraphs-files/nep_2013_2_entwurf_teil_1_kap_1_bis_9.pdf.

[138] Olivier Feix, Ruth Obermann, Marius Strecker, and Regina König. *Netzentwicklungsplan Strom 2014*. Tech. rep. Version 2, Chapter 5. 50Hertz Transmission, Amprion, TenneT TSO & TransnetBW, 2014. URL: https://www.netzentwicklungsplan.de/sites/default/files/nep_2014_1_entwurf_teil1.pdf.

[139] Olivier Feix, Thomas Wiede, Marius Strecker, and Regina König. *Netzentwicklungsplan Strom 2025, Version 2015*. Tech. rep. Version 2, Chapter 5. 50Hertz Transmission, Amprion, TenneT TSO & TransnetBW, 2015. URL: https://netzentwicklungsplan.de/sites/default/files/paragraphs-files/NEP_2025_1_Entwurf_Teil1_0.pdf.

[140] Kerstin Maria Rippel, Thomas Wiede, Mario Meinecke, and Regina König. *Netzentwicklungsplan Strom 2030, Version 2017*. Tech. rep. Version 2, Chapter 5. 50Hertz Transmission, Amprion, TenneT TSO & TransnetBW, 2015. URL: https://netzentwicklungsplan.de/sites/default/files/paragraphs-files/NEP_2025_1_Entwurf_Teil1_0.pdf.

[141] Bundesnetzagentur. *Marktstammdatenregister: Stromerzeugungseinheiten*. 2019. URL: https://www.marktstammdatenregister.de/MaStR.

[142] Vahid Aryanpur, Brian O'Gallachoir, Hancheng Dai, Wenying Chen, and James Glynn. "A review of spatial resolution and regionalisation in national-scale energy systems optimisation models." In: *Energy Strategy Reviews* 37 (2021), p. 100702. ISSN: 2211-467X. DOI: https://doi.org/10.1016/j.esr.2021.100702.

[143] R. Martínez-Gordón, G. Morales-España, J. Sijm, and A.P.C. Faaij. "A review of the role of spatial resolution in energy systems modelling: Lessons learned and applicability to the North Sea region." In: *Renewable and Sustainable Energy Reviews* 141 (2021), p. 110857. ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2021.110857. URL: https://www.sciencedirect.com/science/article/pii/S1364032121001519.

[144] Ulf Philipp Müller, Birgit Schachler, Malte Scharf, Wolf-Dieter Bunke, Stephan Günther, Julian Bartels, and Guido Pleßmann. "Integrated Techno-Economic Power System Planning of Transmission and Distribution Grids." In: *Energies* 12.11 (2019). ISSN: 1996-1073. DOI: https://doi.org/10.3390/en12112091.

[145] Christiane Reinert, Theo Söhler, Nils Julius Baumgärtner, and André Bardow. "Optimization of Regionally Resolved Energy Systems by Spatial Aggregation and Disaggregation." In: *Proceedings of the 16th Symposium Energieinnovation (EnInnov)*. 2020. URL: https://www.tugraz.at/events/eninnov2020/nachlese/download-beitraege/stream-c/.

[146] Aleksander Grochowicz, Koen van Greevenbroek, Fred Espen Benth, and Marianne Zeyringer. "Intersecting near-optimal spaces: European power systems with more resilience to weather variability." In: *arXiv* (2022). DOI: https://doi.org/10.48550/arXiv.2206.12242.

[147] Lieke van der Most, Karin van der Wiel, Rene Benders, Winnie Gerbens-Leenes, Peter Kerkmans, and Richard Bintanja. "Extreme Events in the European Renewable Power System: Validation of a Modeling Framework to Estimate

Renewable Electricity Production and Demand from Meteorological Data." In: 2022. DOI: https://doi.org/10.2139/ssrn.4095758.

[148] *Quartalsbericht 2015 zu Netz- und Systemsicherheitsmaßnahmen; Viertes Quartal 2015 sowie Gesamtjahresbetrachtung 2015*. Tech. rep. Chapter 3. Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, 2016.

[149] *Quartalsbericht zu Netz- und Systemsicherheitsmaßnahmen; Viertes Quartal und Gesamtjahr 2016*. Tech. rep. Chapter 3. Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, 2017.

[150] *Quartalsbericht zu Netz- und Systemsicherheitsmaßnahmen; Gesamtjahr und Viertes Quartal 2017*. Tech. rep. Chapter 3. Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, 2018.

A

ANNEX

This Annex provides additional information on the contents of Chapter 1. We present details on contribution to open source packages in the following alphabetically ordered Sections.

GRIDKIT

We have fixed the *GridKit* error described in Section 1.4.1 by returning every single line. In total, the error affected approximately 30 transmission lines that are now provided separately. The new functionality is publicly available[1].

EXAMPLE

A representative set of lines that were affected by the described error includes the substation *Diele*, close to the North Sea in Germany. The correct set of transmission lines provided by ENTSO-E contains the two walks (Meeden, Diele), (Diele, Rhede), and (Meeden, Diele), (Diele, Conneforde). These lines are provided in series. Therefore, GridKit extracted the lines (Meeden, Rhede) and (Rhede, Conneforde), omitting the substation in Diele. As a result, the substation in Diele remains with no connection to the grid. Moreover, when taking only the subset of German substations from the GridKit extract, it completely misses a connection between Conneforde and Rhede, as Meeden is a dutch village in the municipality of Midden-Groningen. For electricity system modelling this missing transmission line is significant, because it connects the north sea region with many offshore wind sites to onshore regions in North Rhine-Westphalia, the largest urban area in Germany.

INTERANNUAL DEMAND CALCULATOR

Most electricity system optimisation models rely on limited historical electricity consumption data. While there exist a long record of weather data from 1950 to today [61] that provides e.g. wind speeds and temperature in at least hourly resolution per grid cell, hourly electricity demand time-series in [52] are limited to the years 2010-2018. The *Interannual Electricity Demand Calculator* is build to generate aligned electricity consumption time-series to account for the missing years.

The *Interannual Demand Calculator* is based on a previously developed package and analysis [147] that correlates daily historical electricity demand to temperature using a regression model. [147] was validated against data reported on the ENTSO-E transparancy platform. Subsequently, we dis-aggregate the cumulative daily electricity demands to an hourly profile that is sampled from a random historical day (that is the same weekday) from OPSD [52]. This is a novel package, currently at v0.1.

1 https://github.com/martacki/GridKit

NETWORKX

All contributions regarding *NetworkX* pertain functions that cluster a graph $\mathcal{G}$ such, that it maximises its modularity Q by using the Clauset-Newman-Moore greedy modularity maximisation that we use in Chapter 4. In each iteration of the algorithm, it greedily aggregates the two nods v and w that increase Q the most and continues to do so until Q cannot be further improved. Q is a benefit function that measures the quality of a particular division of $\mathcal{G}$ into clusters. Contributions are both on functionality and implementation style.

FUNCTIONALITY Originally, the integrated functionality of *modularity max* neglected the edge weights of $\mathcal{G}$. Instead of taking into account the edge weight, the algorithm always assumed the graph to be unweighted, where every edge had the same weight 1. The improved functionality was released in v2.6.

Later, we adapted the algorithm such that it terminates when a given number of clusters is reached, even when Q can still be further improved (v2.7) and, moreover, such that it continues iterating after the maximal modularity is reached. However, as dQ becomes negative, the process merges those nodes v and w that decrease Q the least (v2.8).

STYLE The original function switched from a regular python function to a generator interface. The generator interface provides a larger flexibility for additional functionalities of the method. The switch to a generator interface retained the original signature of the function. However, we introduced the novel iterator to construct its return value. That way, we get the iterator functionality through the generator interface, but we keep the old interface for ease of use and backward compatibility. These improvements were released in v2.8.

POWERPLANTMATCHING

Contributed changes were released in v0.5.

PYPSA

Major improvements were released in v0.19.0, where the input network can now be spatially clustered on a custom feature using hierarchical clustering, and in v0.19.3, allowing the greedy modularity maximisation to be applied on the network. The latter one builds on previously contributed code to NetworkX, see Section 1.4.4.

PYPSA-EUR

An initial attempt to allow more flexibility on spatial aggregation included a customised mapping of nodes, released in v0.3.0. After that, all contributions that I have made to PyPSA were also applied to PyPSA-Eur, such that the European grid can now be clustered used evaluated and modern methods. Both the hierarchical clustering as well as greedy modularity maximisation were released in v0.5.0. For the hierarchical clustering, I have provided two features, one on the so-called FLH of renewable power generators, and another on the full year-long renewable time-series of capacity factors. More details on the methods and similarity measures can be found in Chapter 4 of this dissertation.

B

ANNEX

This Annex provides additional information on the contents of Chapter 3.

B.1 PRESERVATION OF FLOW PATTERNS WITH CLUSTERING

To understand how well the k-means clustering preserves flow patterns, we took a fixed dispatch pattern for the assets in Europe at high resolution and examined how the network flows changed as the network was clustered.

The fixed dispatch was determined by solving the linearised optimal power flow problem for a 1024-node representation of today's European electricity system. The asset dispatch was then mapped into the clustered networks, and a regular linearised power flow was solved in the clustered network.

If lines $\ell \in N_{c,d}$ in the 1024-node network were mapped to a single representative line $\ell_{c,d}$ in the clustered network, the summed flows from the original network $\hat{f}_{c,d,t} = \sum_{\ell \in N_{c,d}} f_{\ell,t}$ ('microscopic flows') were then compared to the flow $f_{c,d,t}$ in line $\ell_{c,d}$ of the clustered network ('macroscopic flows').

Figure 40 shows the Pearson correlation coefficient between the flows $f_{c,d,t}$ of aggregated lines $\ell_{c,d}$ in the lower resolution network and the summed flows $\hat{f}_{c,d,t}$ of all lines in $N_{c,d}$ in the full resolution network. Red is a linear fit through the points. The distortion from linearity is due to a non-linear scale in the x-axis. Even at 37 nodes the correlation between the flows is good (Pearson correlation coefficient above 0.90) and shows an improving trend until at full 1024-node resolution the flows are once again perfectly equal.

Example density plots of the $\hat{f}_{c,d,t}$ against the $f_{c,d,t}$ for all lines and all times are plotted for different clustering levels in Figure 41. The match between the flows is better for higher resolution networks, with a near-diagonal line already for 362 nodes.

For a more probabilistic approach, we perform a kernel density estimation (KDE) by applying a fast Fourier transformation of aggregated flows of the higher resolved network versus the flows of the low resolution network. Aggregated flows $\hat{f}_{c,d,t}$ are considered an estimator for the flow $f_{c,d,t}$ in the representative lower resolution network. The resulting density functions from the KDE are displayed in Figure 41. For the low resolution network, the probability distribution has two different modes, while a higher resolution network approaches a Gaussian distribution. The variance of the probability density function for a low resolution network is higher than for a high resolution network, as each of the quantile isolines are broader.

B.2 MAPS OF CAPACITY FACTORS FOR WIND AND SOLAR

Figures 42a and 42b present average capacity factors over the weather year 2013 for solar, wind on- and off-shore respectively, mathematically

$$\bar{g}_{v,s} = \langle \bar{g}_{v,s,t} \rangle_t \quad \forall v \in \mathcal{V},\ s \in \mathcal{S}^{re}$$

where $s \in \{\text{solar, wind onshore, wind offshore}\}$. The capacity factors are shown in the Voronoi cells around each of the 1024 node of the original network, i.e. the set of points closest to each node.

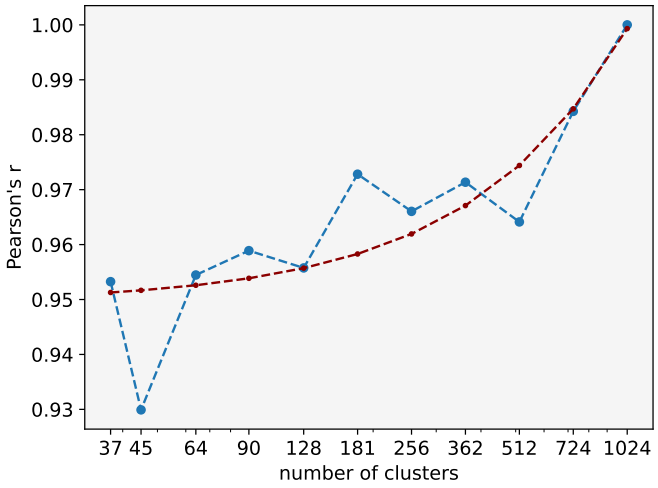


Figure 40: Pearson's correlation coefficient of mapped flows (blue). Note that the x-axis is non-linear, therefore we mark a linear fit to the data (red).

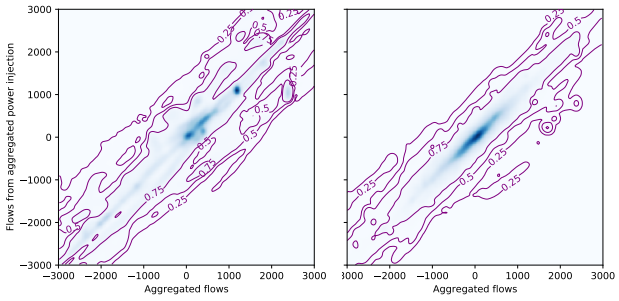


Figure 41: Kernel Density Estimation (KDE) of aggregated flows from a high resolution network grid with 1024 nodes on the x-axis and a low resolution grid with 45 nodes (left) and 362 nodes (right) on the y-axis. 0.25, 0.5 and 0.75 quantiles of the distribution are displayed as purple isolines around the KDE.

The graphics show that capacity factors for solar are decreasing from South to North while those for wind are increasing towards the North and Baltic Sea. The average capacity factors are spatially correlated, but as they are aggregated over larger and larger areas using the weighted average from the clustering approach in equation (23), they decline as bad sites are mixed with good sites. This is reflected in Figure 44, which shows how the average capacity factors per technology for the generation fleet optimised over the whole of Europe change with the clustering.

B.3 BREAKDOWNS FOR TRANSMISSION EXPANSION SCENARIOS

Figure 43 shows an extension of the cost breakdowns in Figure 16 from the scenario with no transmission to scenarios with 25% and 50% grid expansion. The general trends are the same as for the scenario without grid expansion, but grid expansion generally allows more wind capacity to be built, resulting in lower investment in solar, batteries and hydrogen storage, as was seen in Figure 17.

B.4 AVERAGE CAPACITY FACTORS PER TECHNOLOGY

To understand how the model exploits the best available resource sites per node, we examine a time-averaged technology-specific capacity factor $\bar{g}_{c,s}$. The clustered capacity factor is weighted by how much capacity $\sum_{v \in \mathcal{V}_c} G_{v,s}$ of technology s was built within each cluster c with the time-averaged capacity factor $\bar{g}_{c,s} = \langle \bar{g}_{c,s,t} \rangle_t$:

$$\bar{g}_{c,s} := \frac{\sum_{v \in \mathcal{V}_c} \bar{g}_{v,s} G_{v,s}}{\sum_{v \in \mathcal{V}_c} G_{v,s}} \quad \forall c \in \mathcal{V}^K, s \in \mathcal{S}^{re}$$

We present this technology-specific capacity factor in Figure 44 for all three cases with the no-expansion transmission scenario, i.e. where $F_{(v,w)} = F^{2018}_{(v,w)}$.

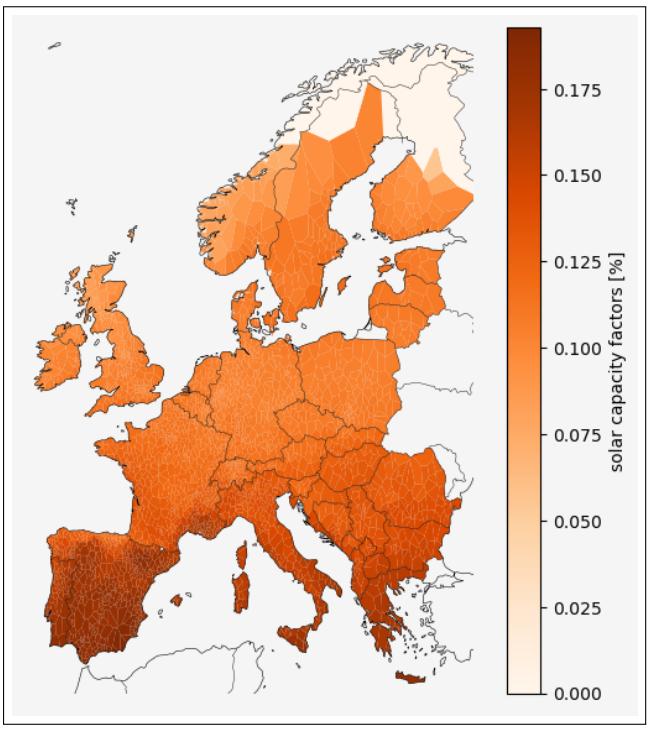
As the number of clusters increases, Case 2 has a larger variety of sites per node to choose where capacity should be installed optimally and is not restricted by transmission constraints beyond country-zones. Therefore, the more sites are available, the higher the weighted capacity factor is because it is not mixed with lower capacity factor sites in equation (23). The highest resolution of Case 2 is also the lowest resolution of Case 3: many resource sites and only one node per country-zone. As the number of nodes in Case 3 increases while the same sites are available, transmission bottlenecks force the model to build more capacity in locations of worse capacity factors. Therefore, the capacity factors drop again. For Case 1, where resource resolution and network resolution change in tandem, the resource resolution dominates and we see increasing capacity factors like in Case 2.

B.5 CURTAILMENT PER TECHNOLOGY

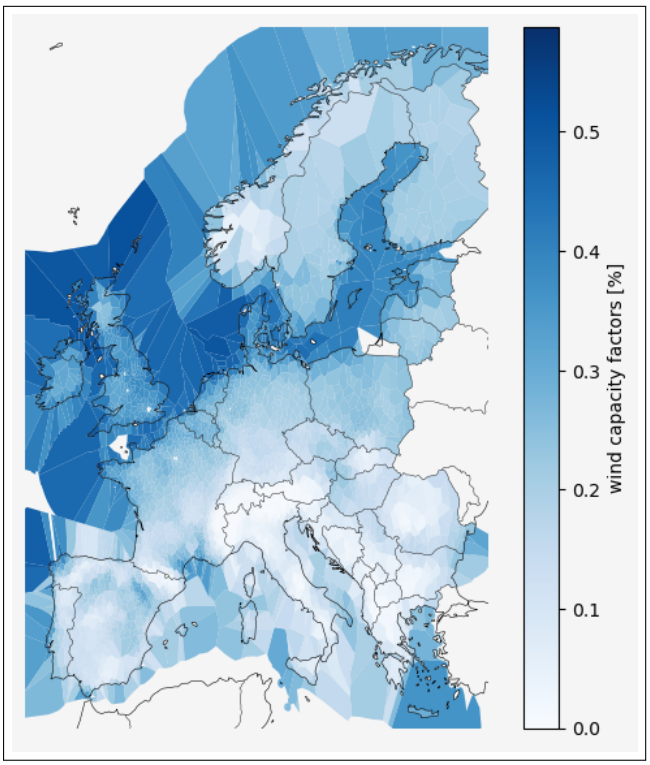
Curtailment is the amount of energy that is available in theory but cannot be injected into the grid because of transmission constraints or a lack of demand:

$$\bar{g}_{v,s,t} \cdot G_{v,s} - g_{v,s,t} \quad \forall v \in \mathcal{V}, s \in \mathcal{S}^{re}, t \in \mathcal{T}$$

Figure 45 shows total curtailment per technology in all Cases. Curtailment in all situations is low (less than 4% of total demand). Curtailment increases with higher network resolution in both the Cases 1 and 3 that incorporate transmis-



(a) Solar capacity factors in Europe for the weather year 2013 at full resolution.



(b) Wind on- and offshore capacity factors in Europe for the weather year 2013 at full resolution.

Figure 42: Solar (top) and wind (bottom, 42b) capacity factors in Europe for the weather year 2013 at full resolution.

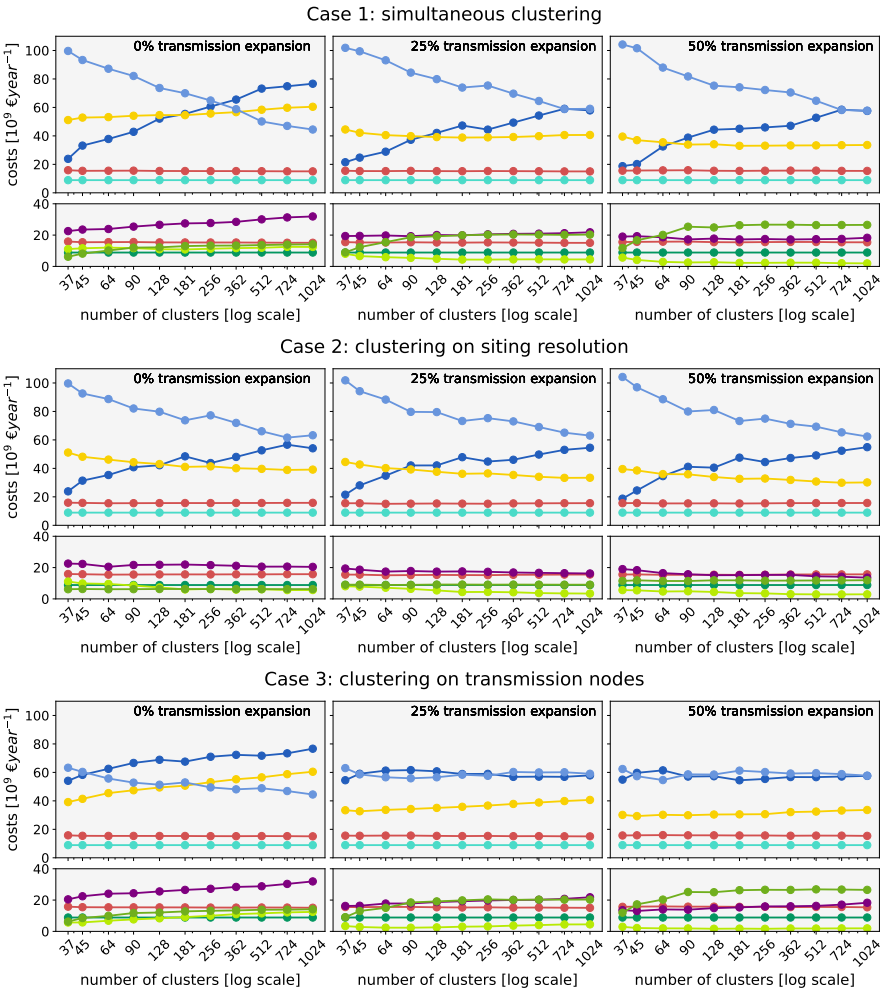


Figure 43: Technology breakdown of the annual system costs for generation (top) and flexibility options (bottom) as a function of the number of clusters.

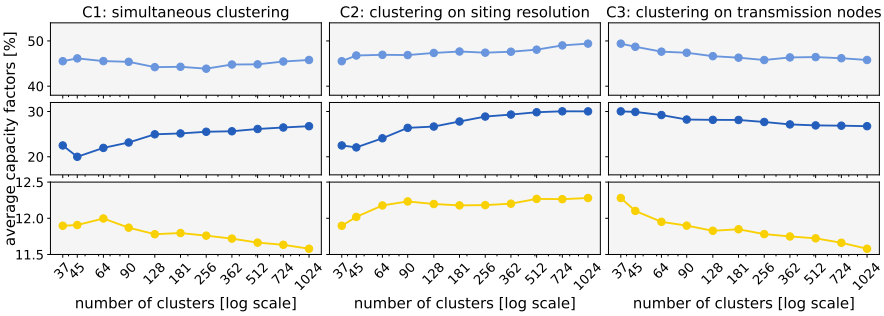


Figure 44: Average capacity factors for each technology for the no transmission expansion scenario in all three cases.

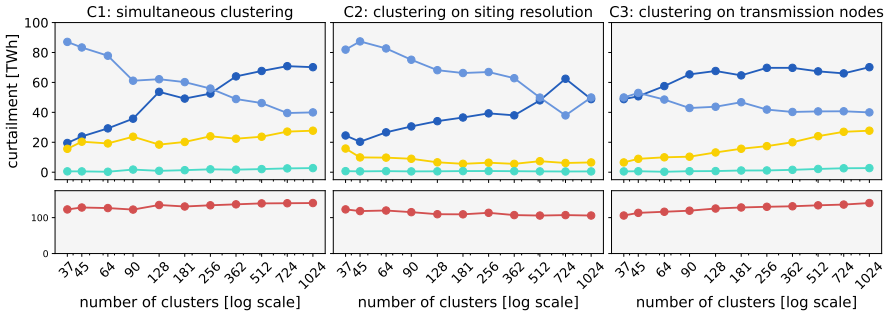


Figure 45: Curtailment for the no transmission expansion scenario in all three cases.

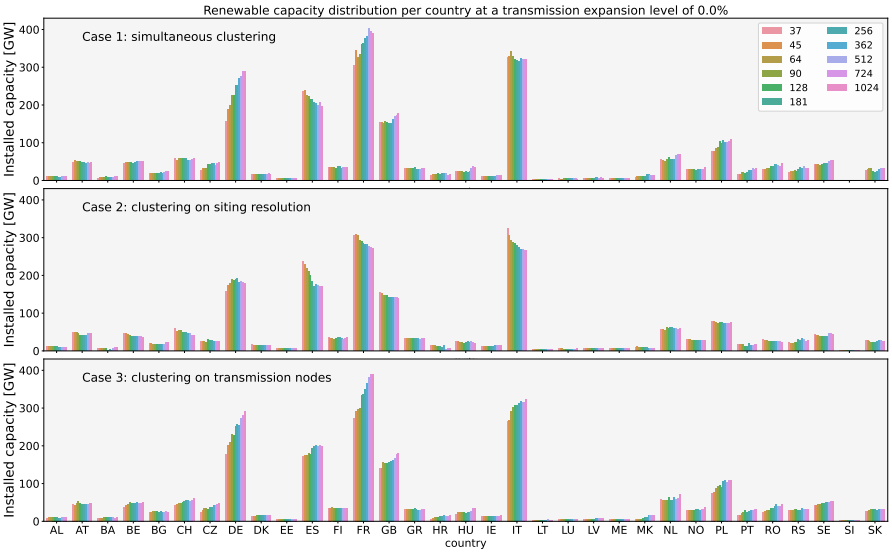


Figure 46: Capacities per country for the no transmission expansion scenario in all three cases.

sion constraints, while it is gently decreasing with resource resolution in Case 2 where there are only transmission constraints at the boundaries of country-zones.

B.6 BREAKDOWNS BY COUNTRY

Figures 16 and 43 show the breakdown of total costs by technology for the whole of Europe. However, it could be that for each technology, the spatial distribution is unstable, moving from country to country with the clustering changes.

For a better understanding of the spatial distribution of installed capacity, we examine the total installed renewable capacity per country in all Cases in Figure 46 with no transmission expansion. The general trend is that the total installed capacity per country is relatively stable with cluster resolution. In Case 2 capacity decreases with resolution, since the exploitation of better resource sites means that less capacity is needed for a given energy yield. The opposite effect is seen in Case 3, while Case 1 reveals a mix of the effects of Case 2 and 3.

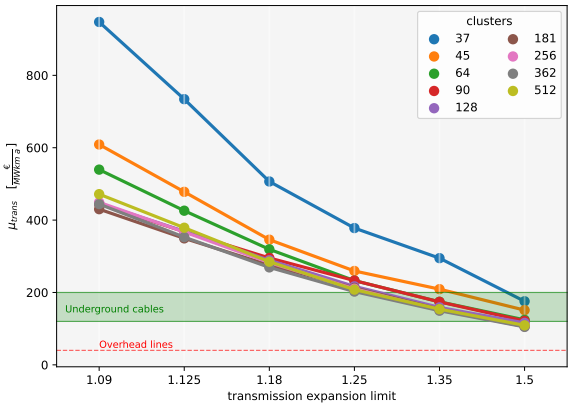


Figure 47: Shadow (dual) price of the line volume constraint.

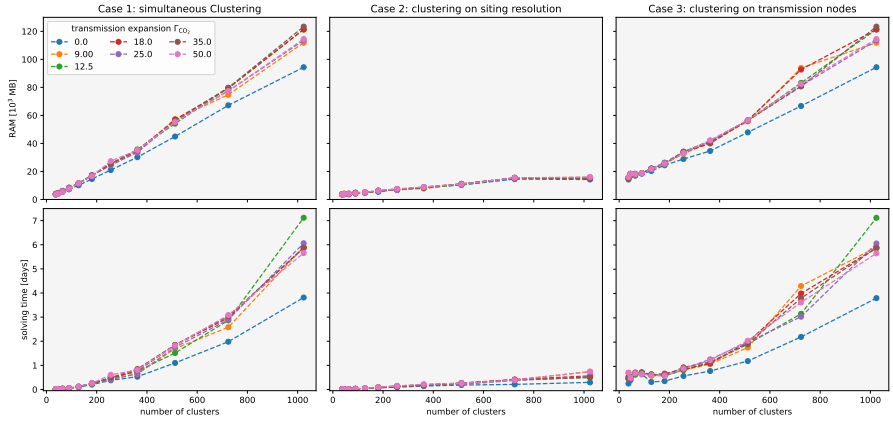


Figure 48: Memory consumption and solving time.

B.7 SHADOW PRICE OF LINE VOLUME CONSTRAINT

The shadow price μ_{trans} of the transmission expansion constraint in equation (14) corresponds to the system cost benefit of an incremental MWkm of line volume. Read another way, it is the line cost required to obtain the same solution with the constraint removed (i.e. lifting the constraint into the objective function as a Lagrangian relaxation).

We present the resulting shadow prices in Figure 47, where they are compared with the annuity for underground and overhead lines. Using the cost of underground cables, the cost-optimal solution would give a grid expansion of 25-50% at high resolution. For overhead transmission, the cost optimum would be over 50%.

B.8 RESOURCE REQUIREMENTS

Memory consumptions and solving times for the three cases discussed in the main body of this dissertation are displayed in Figure 48

B.9 CAPACITY FACTORS WITHIN EACH CLUSTER

In this subsection we analyse the homogeneity of time-average capacity factors for wind and solar within each cluster region as the number of clusters changes.

n clusters	solar	wind onshore	wind offshore
1024	$1.9 \cdot 10^{-3}$	$2.2 \cdot 10^{-2}$	$4.3 \cdot 10^{-2}$
724	$2.3 \cdot 10^{-3}$	$2.5 \cdot 10^{-2}$	$4.5 \cdot 10^{-2}$
512	$2.7 \cdot 10^{-3}$	$2.8 \cdot 10^{-2}$	$4.9 \cdot 10^{-2}$
362	$3.2 \cdot 10^{-3}$	$3.3 \cdot 10^{-2}$	$5.1 \cdot 10^{-2}$
256	$3.7 \cdot 10^{-3}$	$3.6 \cdot 10^{-2}$	$5.3 \cdot 10^{-2}$
181	$4.2 \cdot 10^{-3}$	$3.9 \cdot 10^{-2}$	$5.7 \cdot 10^{-2}$
128	$4.5 \cdot 10^{-3}$	$4.3 \cdot 10^{-2}$	$5.8 \cdot 10^{-2}$
90	$5.0 \cdot 10^{-3}$	$4.6 \cdot 10^{-2}$	$5.9 \cdot 10^{-2}$
64	$6.1 \cdot 10^{-3}$	$4.9 \cdot 10^{-2}$	$6.2 \cdot 10^{-2}$
45	$6.1 \cdot 10^{-3}$	$4.9 \cdot 10^{-2}$	$6.2 \cdot 10^{-2}$
37	$6.2 \cdot 10^{-3}$	$4.9 \cdot 10^{-2}$	$6.2 \cdot 10^{-2}$

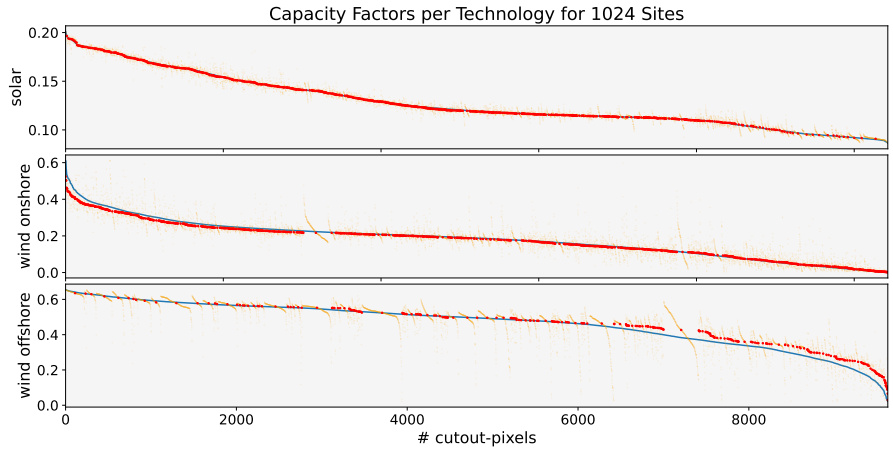
Table 13: average standard deviation of the capacity factor (per unit) per region for a network resolution of 1024, 256 and 37 sites.

Duration curves of the capacity factors in each of the $0.3^\circ \times 0.3^\circ$ weather pixels of the original ERA5 reanalysis dataset [61] for the European area ('cutout') are plotted in blue in Figure 49. In addition, the duration curves for the pixels in each cluster are plotted in orange, with the median for each cluster in red. This reveals how much the capacity factors of wind and solar vary within each cluster region, compared to the whole of Europe. Table 13 presents the average standard deviation with each cluster region for each technology and resolution.

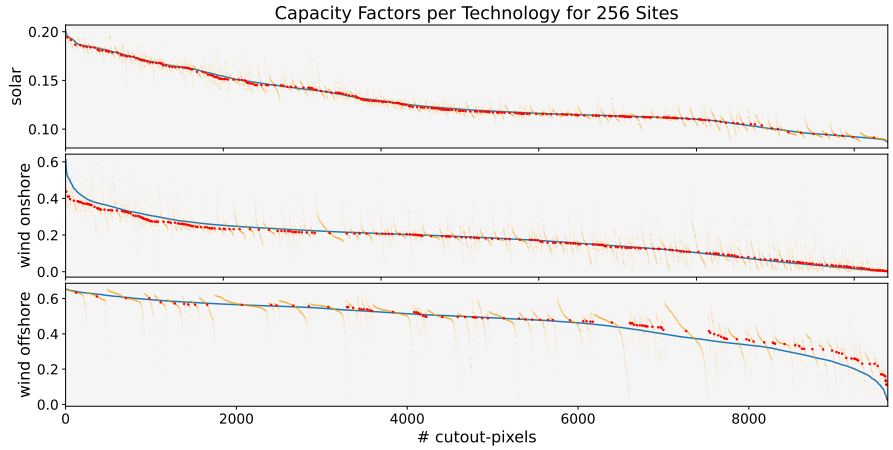
For a high resolution of 1024 clusters, we observe that the median values (red dots) for solar lie very close to the representative values of Europe (black line) with a relatively small average standard deviation of $1.9 \cdot 10^{-3}$ inside each cluster region (scattering of the orange dots). In the case of onshore wind, the high capacity factors are underestimated by the median value, while intermediate and low capacity factors are represented with a minor difference between median and representative European value. For onshore wind, the average standard deviation of the capacity factors within each region is larger than for solar by one magnitude ($\mathcal{O}(10^{-2})$, represented by the scattering of orange dots). The largest variance can be observed in offshore regions, where the average standard deviation is $4.3 \cdot 10^{-2}$, twice as large as for onshore regions, and the low capacity factors are overestimated by their representative median values.

In the case of 256 clusters, the standard deviation per region (scattered orange dots) doubles compared to a resolution of 1024 sites for solar and increases by $\sim 50\%$ for onshore and offshore wind. However, the median values (red dots) per site do not change much compared to the higher resolution case. Only at very low resolutions or, in the extreme, one site representing one country-zone, the median values (red dots) do not agree with the European curve (black line), and the capacity values per site (orange scattered dots) cover a wide range of values (for example $0 - 0.5$ for wind onshore, or $0.11 - 0.0.18$ for solar). At 37 nodes, the average standard deviation is three times larger for solar compared to a resolution of 1024 sites and twice as large for onshore wind.

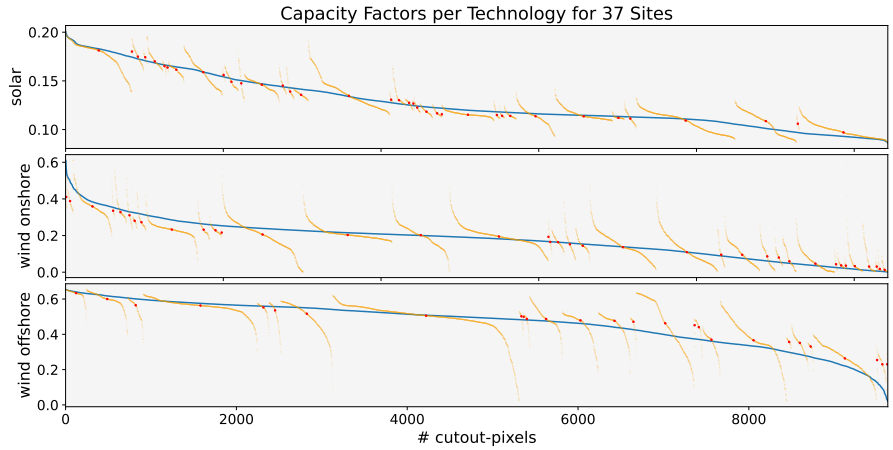
From this analysis we can conclude that a resource resolution of at least several hundred nodes is required to adequately capture the resource variation within Europe, with a higher resolution required for wind than for solar.



(a) Breakdown of capacity factors per technology for the weather cutout pixels inside each of the 1024 clusters presented as a duration curve (orange), with the median marked in red. The overall duration curve for the whole of Europe is plotted in blue.



(b) The same breakdown as in 49a but for 256 clusters.



(c) The same breakdown as in 49a but for 37 clusters.

Figure 49: Duration curves for three exemplary network resolutions of 1024 nodes (top), 256 nodes (middle) and 37 nodes (bottom).

Table 14: Aggregation rules for attributes of nodes and attached assets

attribute	aggregated attribute	mapping	values or units
latitude & longitude	$(x_c, y_c)^T$	$\frac{1}{\lvert\mathcal{V}_c\rvert} \sum_{v \in \mathcal{V}_c} (x_v, y_v)^T$	$\mathbb{R}^2$
power capacity	$G_{c,s}$	$\sum_{v \in \mathcal{V}_c} G_{v,s}$	MW
installable potential	$G^{max}_{c,s}$	$\sum_{v \in \mathcal{V}_c} G^{max}_{v,s}$	MW

Table 15: Aggregation rules for attributes of lines in series

attribute	agg. attribute	mapping	values or units
length (HVDC lines)	$l_{(c,d)}$	$\min_{(v,w) \in E_{(c,d)}} l_{(v,w)}$	km
power capacity	$F_{(c,d)}$	$\sum_{(v,w) \in E_{(c,d)}} F_{(v,w)}$	MVA
length underwater	$u_{(c,d)}$	$\frac{1}{l_{(c,d)}} \sum_{(v,w) \in E_{(c,d)}} (l \cdot u)_{(v,w)}$	p.u.

B.10 AGGREGATION RULES

Aggregation rules for nodal attributes are shown in Table 14, for lines in series in Table 15 and for lines in parallel in Table 16.

Table 16: Aggregation rules for attributes of lines in parallel

attribute	agg. attribute	mapping	values or units
power capacity	$s^{nom}_{(c,d)}$	$\sum_{(v,w)\in E_{(c,d)}} s^{nom}_{(v,w)}$	MVA
power capacity maximum	$s^{min}_{(c,d)}$	$\sum_{(v,w)\in E_{(c,d)}} s^{min}_{(v,w)}$	MVA
power capacity minimum	$s^{max}_{(c,d)}$	$\sum_{(v,w)\in E_{(c,d)}} s^{max}_{(v,w)}$	MVA
number of parallel lines	$n^{parallel}_{(c,d)}$	$\sum_{(v,w)\in E_{(c,d)}} n^{parallel}_{(v,w)}$	$\mathbb{R}$
terrain factor for capital costs	$t_{(c,d)}$	$\lvert E_{(c,d)}\rvert^{-1} \sum_{(v,w)\in E_{(c,d)}} t_{(v,w)}$	p.u.

C

ANNEX

This Annex provides additional information on the contents of Chapter 4.

C.1 SUFFICIENT BENCHMARK RESOLUTION AT 1250 NODES

Computational model feasibility remains a problem even after applying linearisation to the model formulation and spatial/temporal aggregation. Therefore all low-resoltuion model results were compared against a higher resolved model with a spatial resolution of 1250 nodes. Requirements for solving this model size were a runtime of up to 24 days and 240 GB of RAM capacity.

1250 nodes portray approximately 51% of the pre-aggregated model size and 24% of the original full-resolution problem. Results in [1] indicate that model results are stable when the spatial resolution is at least 49% of the pre-aggregated model and 26% of the original model. At this or higher resolutions only minor fluctuations of $4-6\%$ occur in terms of optimal dispatch and power flows. As this could potentially be wrong for a capacity expansion problem, we make further justifications for this benchmark by providing the average deviation from the mean as well as the correlation factors for every considered carrier (solar, wind, battery and hydrogen) evaluated on the lowest common region size and power flows between these regions in a 4×4 correlation-matrix. The lowest common regions turn out to be the countries and synchronous zones, which is in line with the benchmark-setting of equation (18).

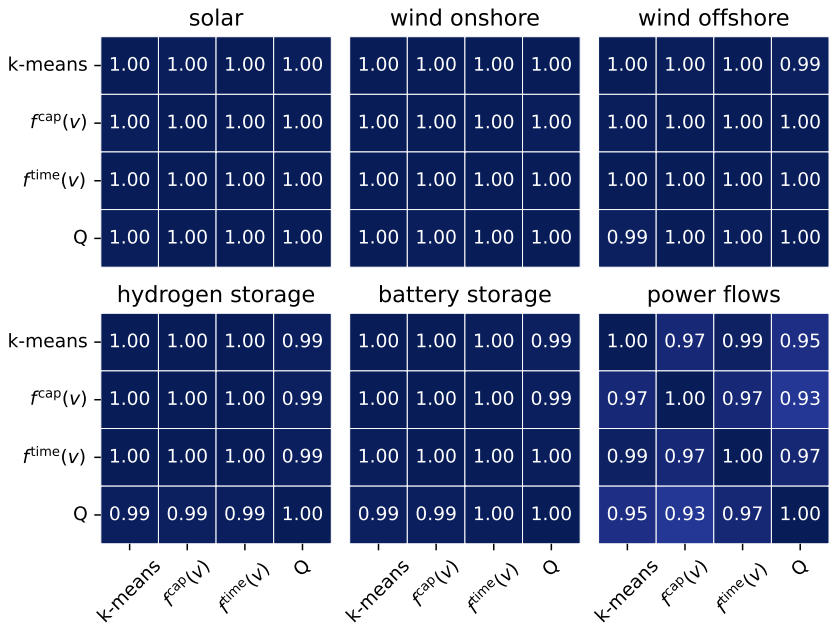
For the 60% carbon reduction target, optimal investments have small deviations from the mean of up to 5% for offshore wind. Onshore wind and solar installations are more stable with lower cross-deviations. However, optimal installation for battery storage deviates by more than 10% when comparing the 1250 node results of the clustered model with Q to the other clustered model results. But as battery storage for this carbon level is low in general (only 2% of total installed capacity), the relative deviation gives the wrong impression of having strong impact on the optimal result. These results are graphically illustrated in Figure 50 (50a and 50b). It shall also be noted that shifting capacity from one carrier to another might have only small impacts on the objective function, see [109].

For the 100% carbon reduction target the worst deviation from the mean can be observed for the optimal investment in offshore wind assets with deviations of up to 8%; other technologies as well as power flows have smaller deviations. These results are graphically illustrated in Figure 51 (51a and 51b).

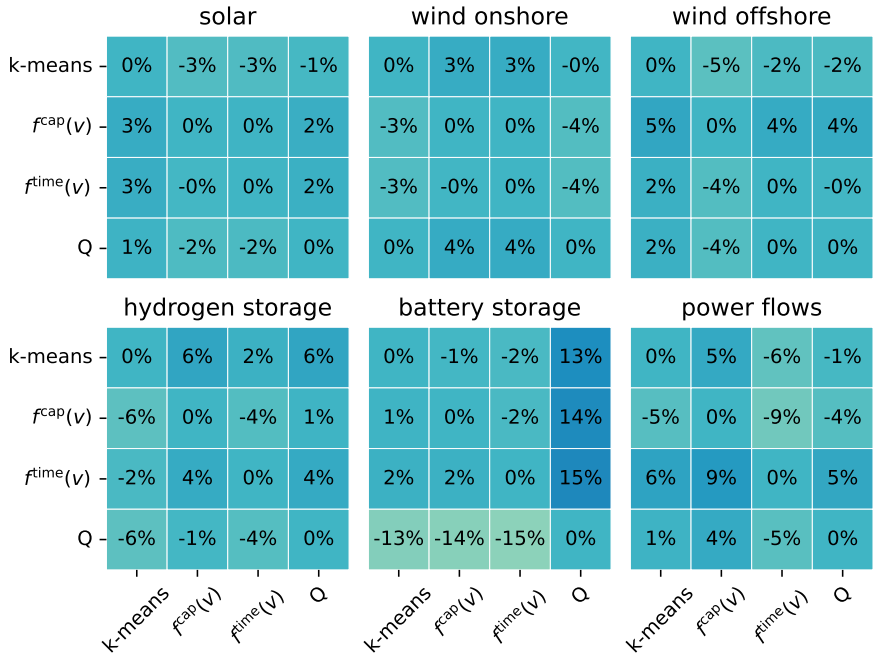
In both scenarios, the Pearson's correlation coefficients are ≈ 1 except for power flows where the coefficients are lower but still > 0.9, indicating a linear correlation between the results.

C.2 MEAN SQUARED ERROR VALUES FOR STORAGE

We provide the mean squared error values $\mathrm{MSE} = \mathrm{MSE}^{+} + \mathrm{MSE}^{-}$ for storage technologies for a spatial resolution of 97 nodes in Table 21, and additionally for 67 nodes in Table 22 and 127 nodes in Table 23.

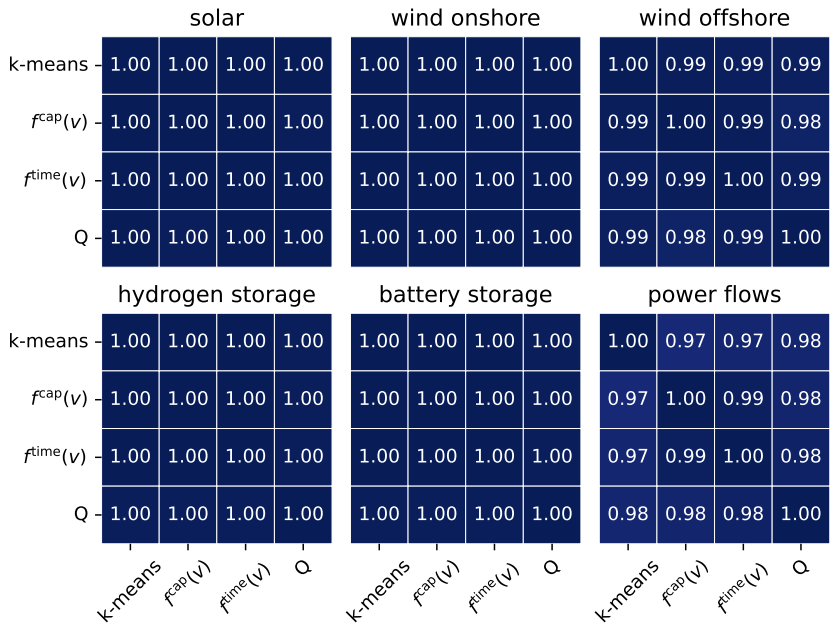


(a) Correlation factors for every considered carrier (solar, wind, battery and hydrogen) and power flows between the different model results respective clustering.

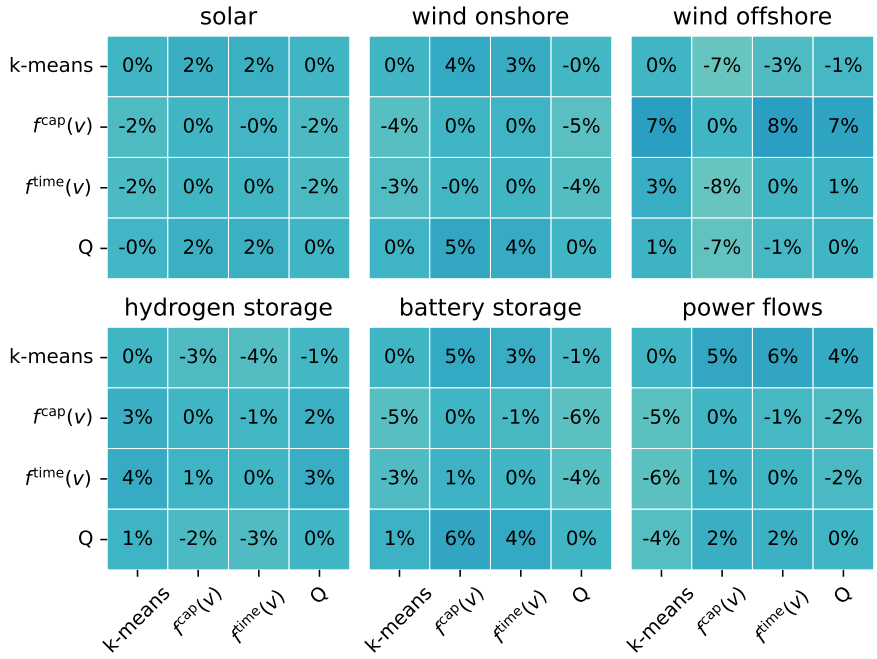


(b) Average deviations from the mean for every considered carrier (solar, wind, battery and hydrogen) and power flows between the different model results respective clustering.

Figure 50: Correlation factors and average deviations from the mean for modelling results at 1250 nodes and 60% CO_2 reduction compared to 1990.



(a) Correlation factors for every considered carrier (solar, wind, battery and hydrogen) and power flows between the different model results respective clustering.



(b) Average deviations from the mean for every considered carrier (solar, wind, battery and hydrogen) and power flows between the different model results respective clustering.

Figure 51: Correlation factors and average deviations from the mean for modelling results at 1250 nodes and 100% CO_2 reduction compared to 1990.

To show that the results of Tables 6 and 7 are no artifacts of a resolution of 97 nodes and change substantially when varying the spatial resolution, we additionally provide the equivalent tables for modelling results at 67 nodes (Tables 17 and 19) and 127 nodes (Tables 18 and 19).

Table 17: Analogous to Table 6 but with a spatial resolution of 67 nodes.

CO_2 reduction	60%		100%	
	ρ	r_2	ρ	r_2
k-means	0.704	0.188	0.725	0.195
$f^{cap}(v)$	0.754	0.174	0.759	0.187
$f^{time}(v)$	0.749	0.173	0.765	0.181
Q	0.739	0.173	0.740	0.187

Table 18: Analogous to Table 6 but with a spatial resolution of 127 nodes.

CO_2 reduction	60%		100%	
	ρ	r_2	ρ	r_2
k-means	0.735	0.164	0.772	0.166
$f^{cap}(v)$	0.802	0.144	0.786	0.163
$f^{time}(v)$	0.782	0.147	0.808	0.152
Q	0.789	0.152	0.792	0.165

Table 19: Analogous to Table 7 but with a spatial resolution of 67 nodes.

CO_2 reduction	60%		100%	
technology	wind	solar	wind	solar
k-means	$0.33 + 2.65$	$0.01 + 2.34$	$0.22 + 2.43$	$0.25 + 0.71$
$f^{cap}(v)$	$0.23 + 0.79$	$0.05 + 0.31$	$0.14 + 1.12$	$0.05 + 0.12$
$f^{time}(v)$	$0.02 + 2.26$	$0.07 + 0.99$	$0.51 + 1.63$	$0.06 + 0.24$
Q	$0.42 + 1.45$	$0.16 + 0.71$	$0.61 + 1.76$	$0.07 + 0.48$

Table 20: Analogous to Table 7 but with a spatial resolution of 127 nodes.

CO_2 reduction	60%		100%	
technology	wind	solar	wind	solar
k-means	$0.42 + 5.34$	$0.06 + 2.17$	$0.51 + 2.22$	$0.21 + 1.03$
$f^{cap}(v)$	$0.79 + 0.86$	$0.02 + 0.82$	$0.2 + 1.14$	$0.11 + 0.15$
$f^{time}(v)$	$0.81 + 2.74$	$0.02 + 1.45$	$0.14 + 2.38$	$0.24 + 0.75$
Q	$0.36 + 1.31$	$0.47 + 1.17$	$0.24 + 2.2$	$0.36 + 1.07$

Table 21: MSE presented as a sum of over- and underestimated optimal estimates according to equation (33) respective clustering method, storage technology and CO_2 reduction target for a spatial resolution of 97 nodes. Graphically presented in Figure 28.

CO_2 reduction	60%		100%	
technology	hydrogen	battery	hydrogen	battery
k-means	$0.62 + 0.28$	$1.0 + 0.76$	$0.74 + 1.54$	$0.28 + 0.32$
$f^{cap}(v)$	$1.37 + 0.04$	$0.41 + 1.39$	$0.24 + 0.68$	$2.29 + 0.67$
$f^{time}(v)$	$0.64 + 0.51$	$0.57 + 0.28$	$0.51 + 2.76$	$0.99 + 0.37$
Q	$0.58 + 0.07$	$0.82 + 1.45$	$0.08 + 2.8$	$0.09 + 0.26$

Table 22: Analogous to Table 21 but with a spatial resolution of 67 nodes.

CO_2 reduction	60%		100%	
technology	hydrogen	battery	hydrogen	battery
k-means	$0.81 + 0.47$	$0.34 + 0.26$	$0.44 + 0.64$	$0.58 + 0.52$
$f^{cap}(v)$	$0.89 + 0.33$	$1.18 + 1.03$	$0.13 + 0.86$	$0.43 + 0.01$
$f^{time}(v)$	$0.16 + 0.19$	$0.46 + 0.26$	$0.52 + 0.65$	$0.3 + 0.08$
Q	$0.0 + 0.1$	$0.75 + 0.44$	$0.12 + 0.78$	$0.02 + 0.19$

Table 23: Analogous to Table 21 but with a spatial resolution of 127 nodes.

CO_2 reduction	60%		100%	
technology	hydrogen	battery	hydrogen	battery
k-means	$0.23 + 1.55$	$0.62 + 1.09$	$0.34 + 2.75$	$0.17 + 0.69$
$f^{cap}(v)$	$0.24 + 0.96$	$0.24 + 1.07$	$2.1 + 1.29$	$1.92 + 0.96$
$f^{time}(v)$	$0.23 + 0.38$	$0.27 + 0.61$	$0.94 + 1.93$	$0.75 + 0.69$
Q	$0.94 + 0.06$	$0.34 + 1.59$	$0.35 + 1.35$	$0.52 + 0.67$

D

ANNEX

This Annex provides additional information on the contents of Chapter 5.

D.1 ADAPTIONS OF THE TRANSMISSION NETWORK

The transmission network for the model was extracted from the 2018 interactive ENTSO-E map [136], such that the final simulation for 2018 uses the unmodified transmission grid. For all other years, the following adaptions were made:

$\leqslant$ **2013**
strengthened lines: Ackerstraße - Matterbusch, Mengede - Wanne, Goldshöfe - Niederstotzingen, Dellmensingen - Niederstotzingen

$\leqslant$ **2014**
strengthened lines: Gütersloh-Friedrichsdorf, Sechtem-Neuenahr, Großgartach-Hüffenhardt;
new lines: Kriftel-Eschborn, Bruchsal-Forst

$\leqslant$ **2015**
strenghtened lines: Dellmensingen - Niederstrotzingen, Bärwalde - Schmölln, Mittelbexbach - St.Barbara Westerkappeln - Hambüren, Friedrichsdorf - Bielefeld;
new lines: Vieselbach - Altenfeld (Thüringer Strombrücke Part I), Abzweig Förderstedt, Görries - Parchim/Lübz

$\leqslant$ **2017**
strenghtened lines: Audorf/Süd - Hamburg/Nord, Hoheneck - Rommelsbach, Niederrhein - Lackhausen, Herne - Wanne;
new lines: Altenfeld - Redwitz, (Thüringer Strombrücke Part II), Kriftel - Obererlenbach

Strenghtened lines were reduced in capacity by a factor 2, new lines were removed from the grid.

D.2 LINE CONGESTION

We shortly discuss which lines were historically subject to redispatch measures with a duration of more than 500 hours according to [134, 148–150] and how these lines are affected by our model simulations. Lines affected by our simulation are filered wherever the shadow prices in equation 35 are positive.

Affected lines in 2015 were: (RR) Remptendorf - Redwitz, (NV) Neuenhagen - Vierraden, (NP) Neuenhagen - Pasewalk, (MLGGG) Mehrum - Lehrte - Godenau - Göttingen, (CU) Conneforde - Unterweser, (CH) Conneforde-Huntdorf and (BH) Brunsbüttel-Hamburg Nord. Our model captures congestion with a duration of (RR) 5446 hours, (NP) 730 hours, (MLGGG) 46 hours and (BH) 1344 hours. (CU) and (CH) are not overloaded at any time, however a neighboring lines are congested for 14 and 24 hours respectively and (NV) is lost due to aggregation.

In 2016, redispatch measures affected (RR), (NV), (NP) and (BH). Model congestion affects (RR) with a duration of 7225 hours, (NP) with 526 hours and (BH) with 1235 hours. (NV) is lost due to aggregation

Routes affected in 2017 include (RR), (BH), (SAP) Sittling - Altheim - St. Peter, (PP) Pleiting - St. Peter and (DNMH) Dörpen West - Niederlangen - Meppen

- Hanekenfähr. These lines were congested in our simulations for a duration of (RR) 7487 hours, (BH) 1017 hours and (DNMH) 2591 hours. (SAP) is not affected by congestion, however a neighboring line is congested for 1766 hours and (PP) is lost due to aggregation.

Reports for 2018 list (DG) Dipperz - Großkrotzenburg, (SW) Scottrum - Wechold, (SL) Scottum - Landesbergen, (DNMH), and (SAP). There lines are congested in our model simulations for a duration of (SW)17 hours, (SL) 440 hours and (DNMH) 744 hours. (DG) and (SAP) are never subject to congestion, however neighbouring lines are congested with 139 and 1359 hours respectively.

Note that all these lines underlie clustering inaccuracies that include both routing and available capacity. Hence a line mapping from the original grid to the model grid is inaccurate.

D.3 LOADSHEDDING

As the generation and transmission fleet of the model problem is fixed in for the optimisation (34), the conditions (35) and (15) might be too constraining to satisfy demand everywhere at any time. For this reason, an auxiliary generator with infinite capacity is added to each node $v \in \mathcal{V}$ to fill up for lacking generation and flow capacities. The goal of course is, to keep its dispatch to a low amount, 0 kWh in the best case, therefore its operation is penalised by extremely high costs of 10^3 €/kWh.

Results indicate, that the amount this auxiliary generator is required, is less than 0.2% in any simulation. The year with its highest total dispatch is 2013 with a total of 1214 GWh at a model resolution of 306 nodes, portraying 0.13% of the annual demand. For all other years and network resolutions, the dispatch of this generator is even lower, see Figure 52. As the installed capacity increases over the years, loadshedding decreases from its peak at a resolution of 266 nodes with approximately 1200 GWh in 2013 to below 2 GWh in 2018 at the same resolution. In both cases, the mismatch between electricity demand and generation is well below 1%.

D.4 ADDITIONAL EVALUATION RESULTS

D.4.1 *Cumulative Results*

Analogous results on clustering as for 2017 in section 5.2.3 in Figure 32 are displayed in Figures 53, 54 and 55 for the years 2013-2018.

D.4.2 *Curtailment per TSO area*

Stability of the results of section 5.3.3 as the network resolution changes are displayed in Figure 56.

D.4.3 *Curtailment per Quarter*

Stability of the results of section 5.3.4 as the network resolution changes are displayed in Figure 57.

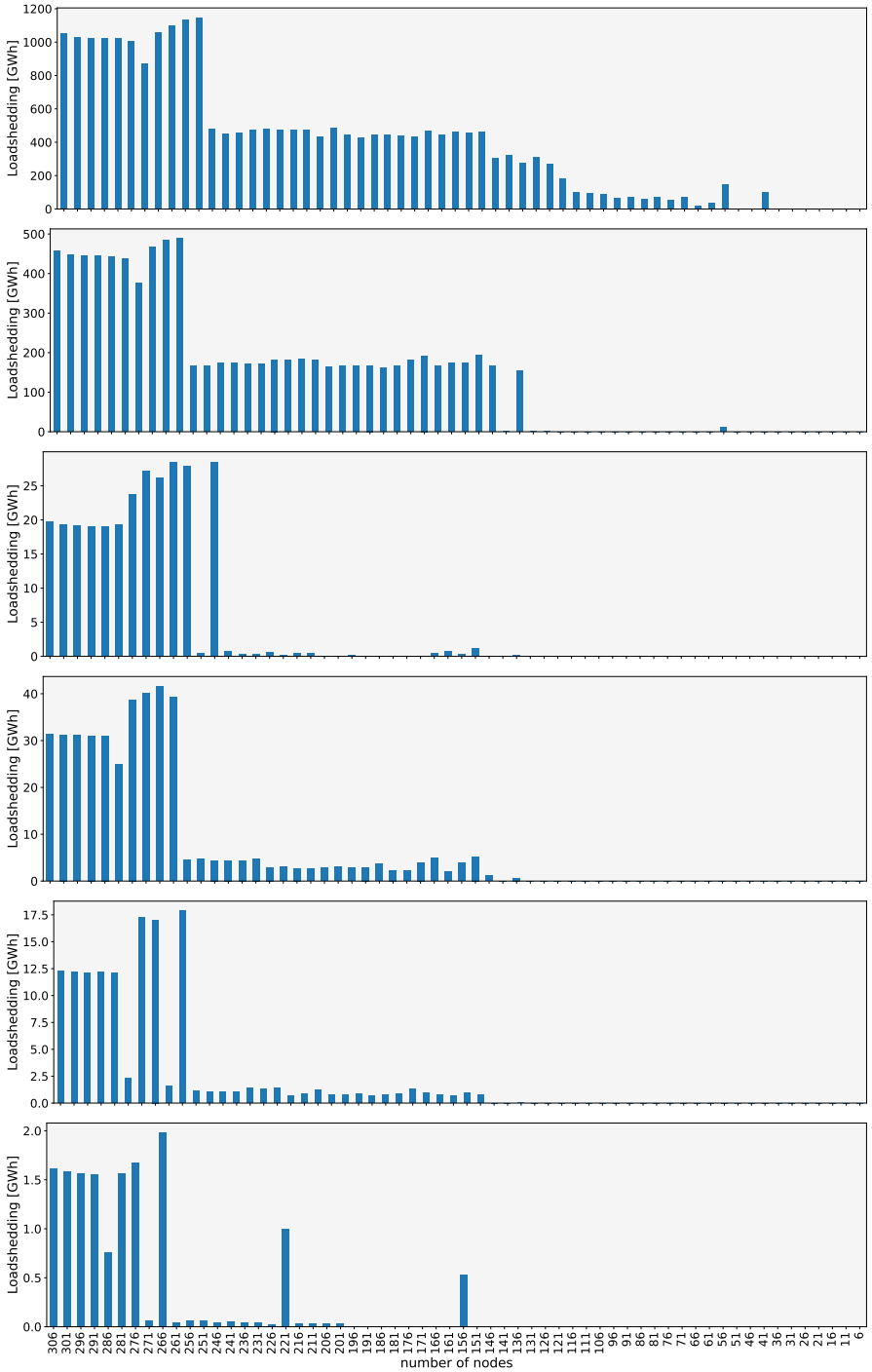


Figure 52: Model results on loadshedding for the years 2013-2018 and different cluster sizes.

D.4.4 *Sensitivity of the results*

Stability of the results of sections 5.3.1-5.3.4 as the generators are assigned with a probability of 5% to node that lies within a radius of 43km of its closest one, are displayed in Figures 58 and 59.

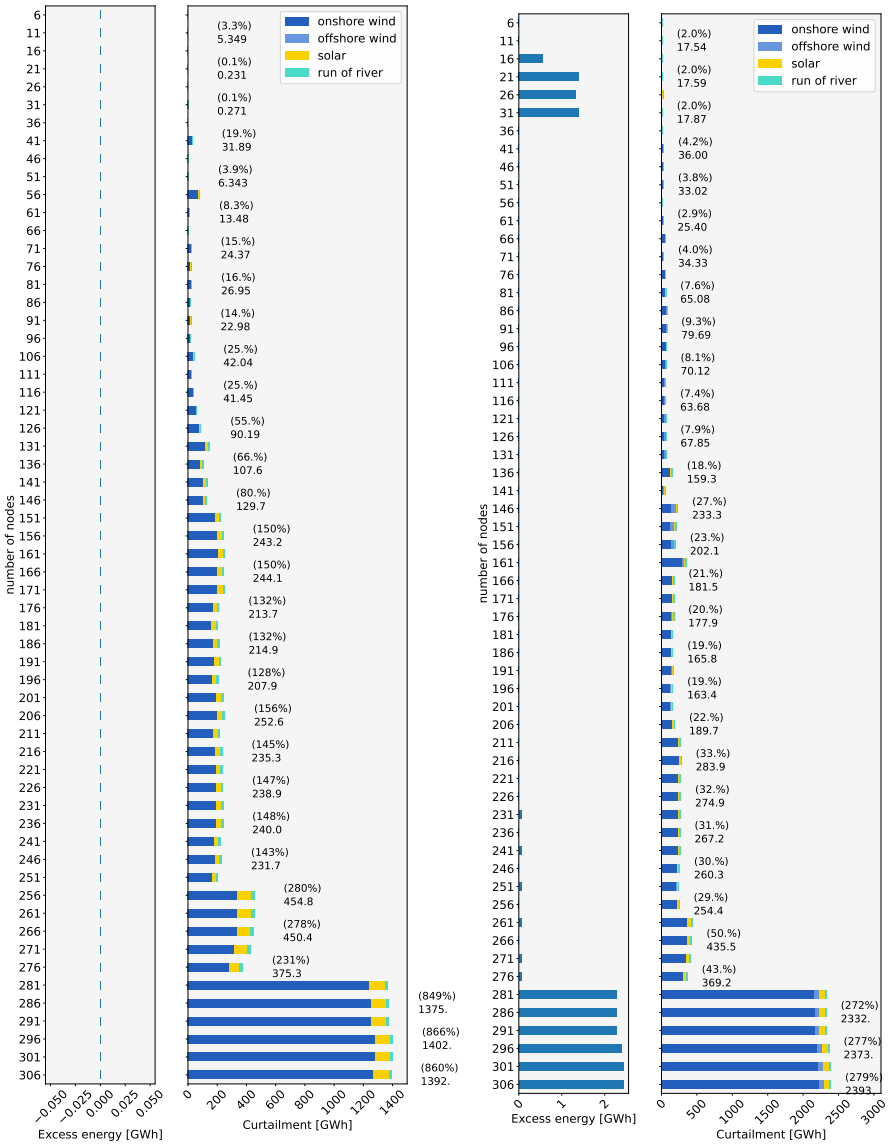


Figure 53: Cumulative historical results on curtailment for the years 2013 (left) and 2014 (right) as the network resolution changes.

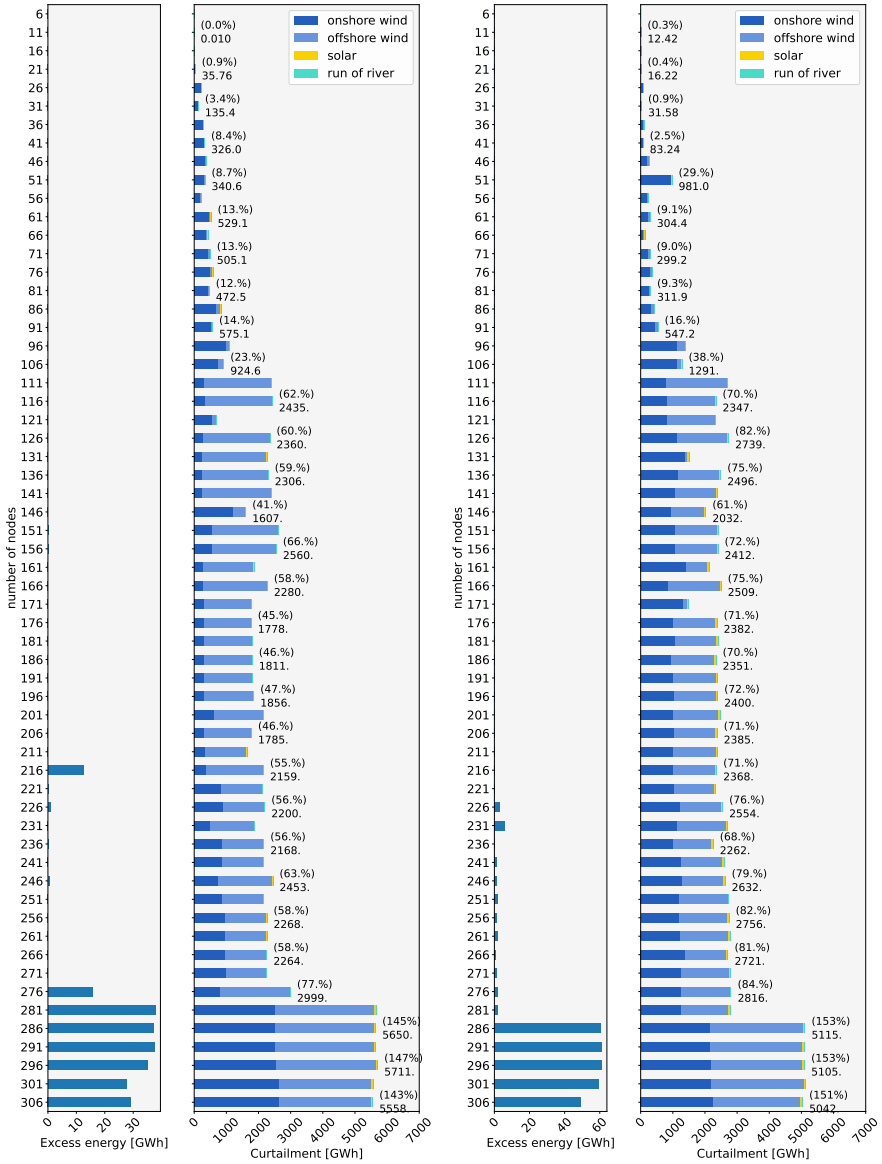


Figure 54: Cumulative historical results on curtailment for the years 2015 (left) and 2016 (right) as the network resolution changes.

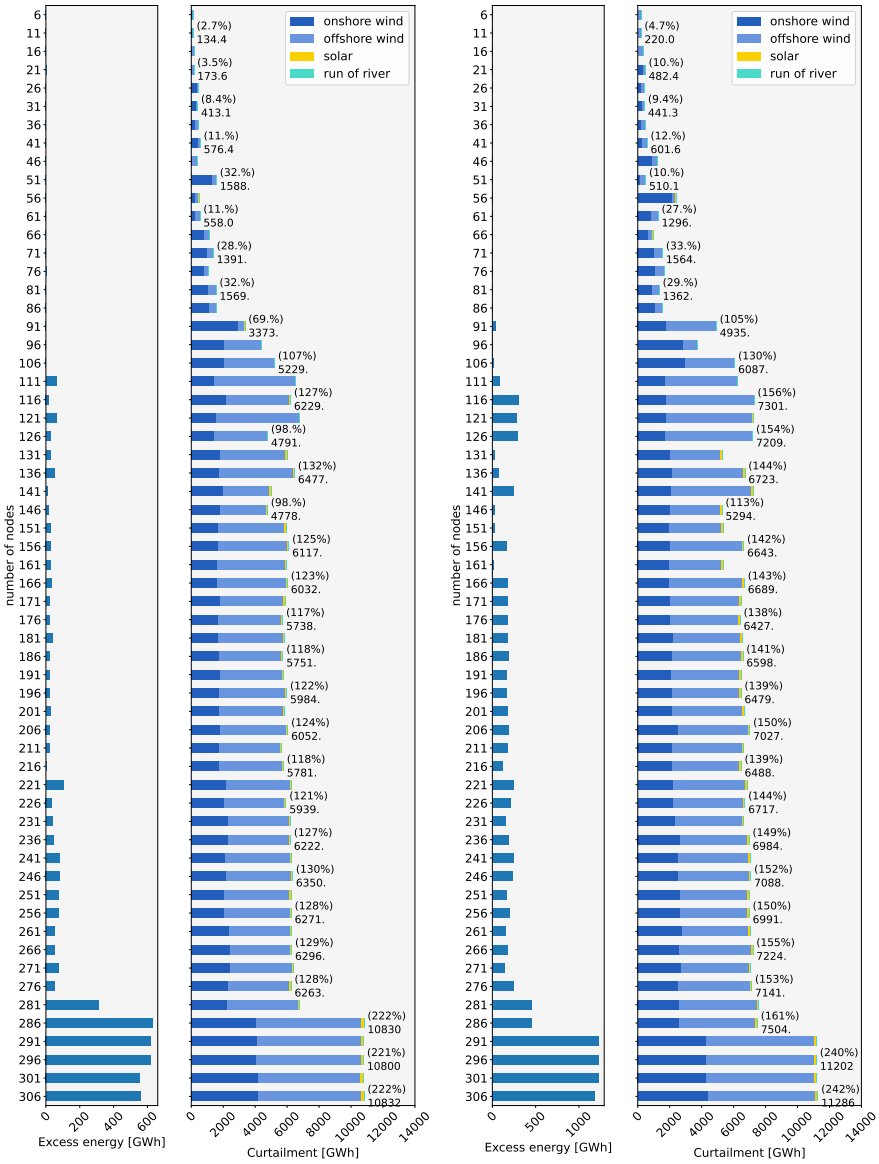


Figure 55: Cumulative historical results on curtailment for the years 2017 (left) and 2018 (right) as the network resolution changes.

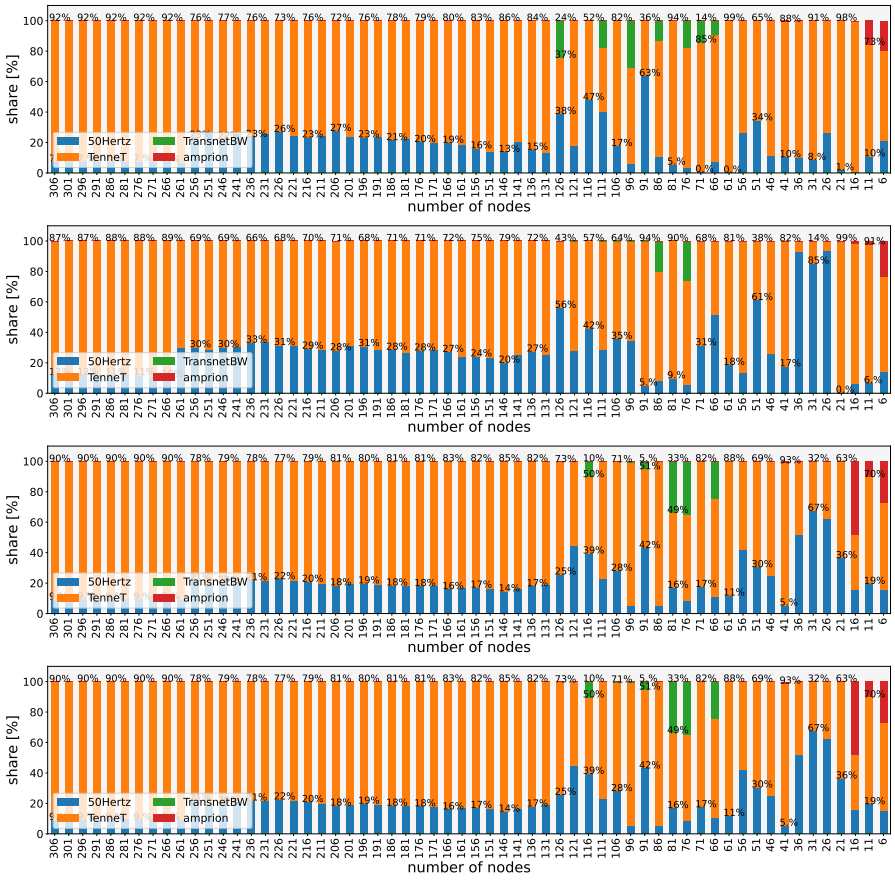


Figure 56: Model results of curtailment in in percent per control zone of German TSOs for the years 2015-2018.

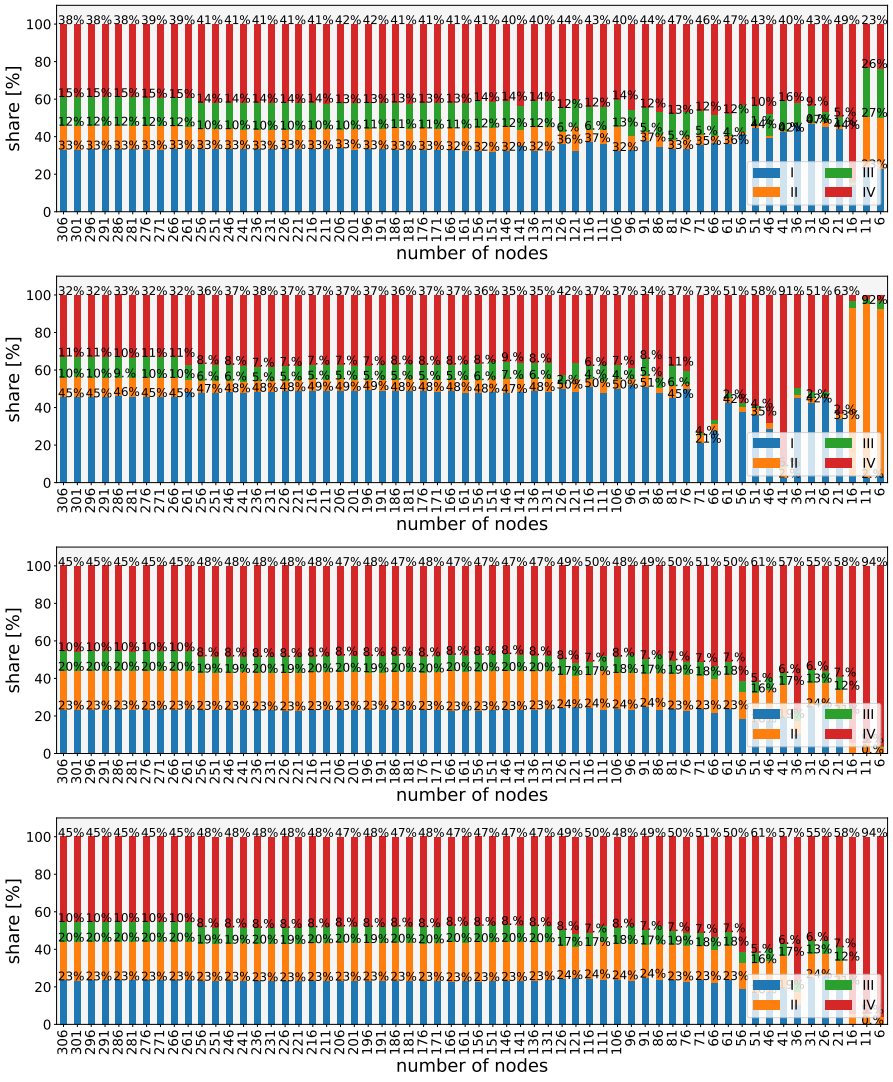


Figure 57: Model results of curtailment distributed per quarter for the years 2015-2018 as the network resolution changes.

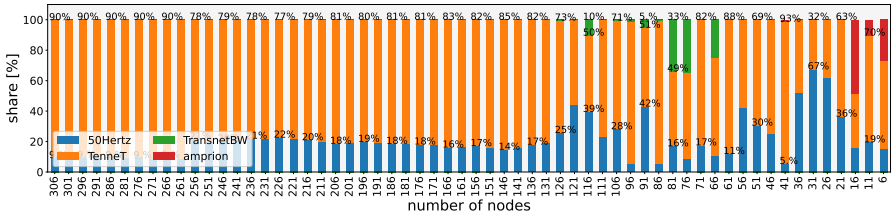


Figure 58: Model results of curtailment distributed per control zone of German TSOs for the years 2017 as the network resolution changes with a 5%-randomness in the assignment of generators.

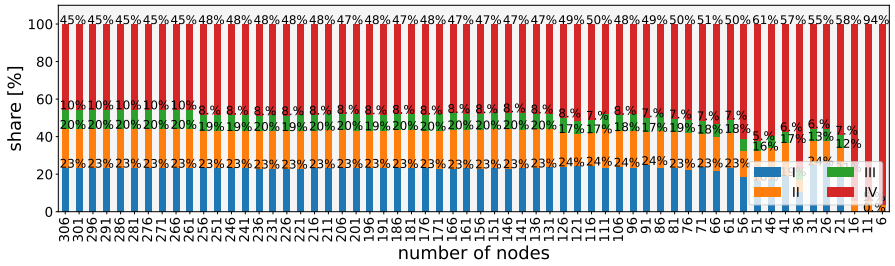


Figure 59: Model results of curtailment distributed per quarter for the year 2017 as the network resolution changes with a 5%-randomness in the assignment of generators.

E

ANNEX

This Annex provides additional information on the contents of Chapter 6.

E.1 RESULTS OF THE ISLANDED DISAGGREGATION METHOD

In this setting each cluster is treated as an island, meaning that no electricity trade between other clusters is considered for the dis-aggregation. Results on loadshedding and curtailment for this scenario are displayed in Figure 60.

The overall trend of the results is similar to the simulations where inter-cluster power flows were considered in the simulations. However, there are minor differences mainly affecting the "re-optimise" results. These result in an overall higher curtailment of $2-3\%$, and lower loadshedding of $1-2\%$.

E.2 ANALYSING THE SOURCE FOR LOADSHEDDING

Figure 62 additionally displays if loadshedding measures occur at times where the curtailment of the high-resolution model is higher compared to the lower resolved reference results. If true, this indicates that the loadshedding measures are due to underestimated within-cluster transmission bottlenecks. To precisely evaluate this statement, the Figure displays the following hypothesis:

$$\delta_{\{t:(\Upsilon_V-\Upsilon_C)(t)>0\}}\cdot(\Upsilon_V-\Upsilon_C)(t)\geqslant 0.5\cdot\Delta_V(t) \tag{45}$$

$$\delta_{\{t:(\Upsilon_V-\Upsilon_C)(t)<0\}}\cdot(\Upsilon_C-\Upsilon_V)(t)\leqslant 0.5\cdot\Delta_V(t) \tag{46}$$

where $\Delta_V(t)$ represents the amount of loadshedding measures in the high-resolution dis-aggregated model at snapshot t, and

$$\Upsilon_V(t) := \sum_{\substack{s\in S\\ v\in\mathcal{V}}} (\bar{g}_{v,s,t}G_{v,s} - g_{v,s,t})$$

the amount of curtailment in the high-resolution dis-aggregated model at snapshot t. Accordingly, $\Upsilon_C(t)$ represents the amount of curtailment in the lower resolved reference model.

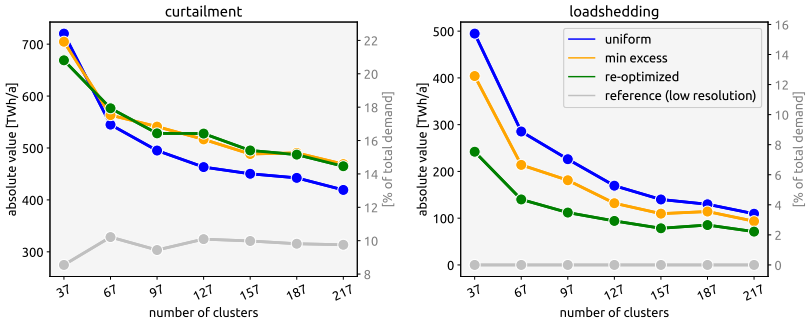


Figure 60: Results as displayed in Figure 36 of an island model, meaning that no inter-cluster electricity imports or exports are considered for the dis-aggregation.

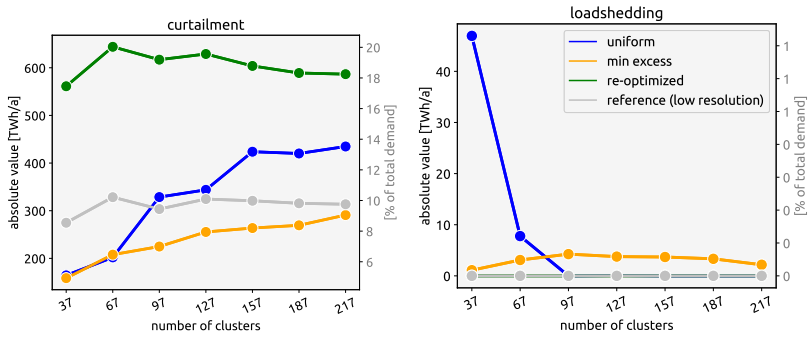


Figure 61: Results as displayed in Figure 61 of an island model, meaning that no inter-cluster electricity imports or exports are considered for the dis-aggregation.

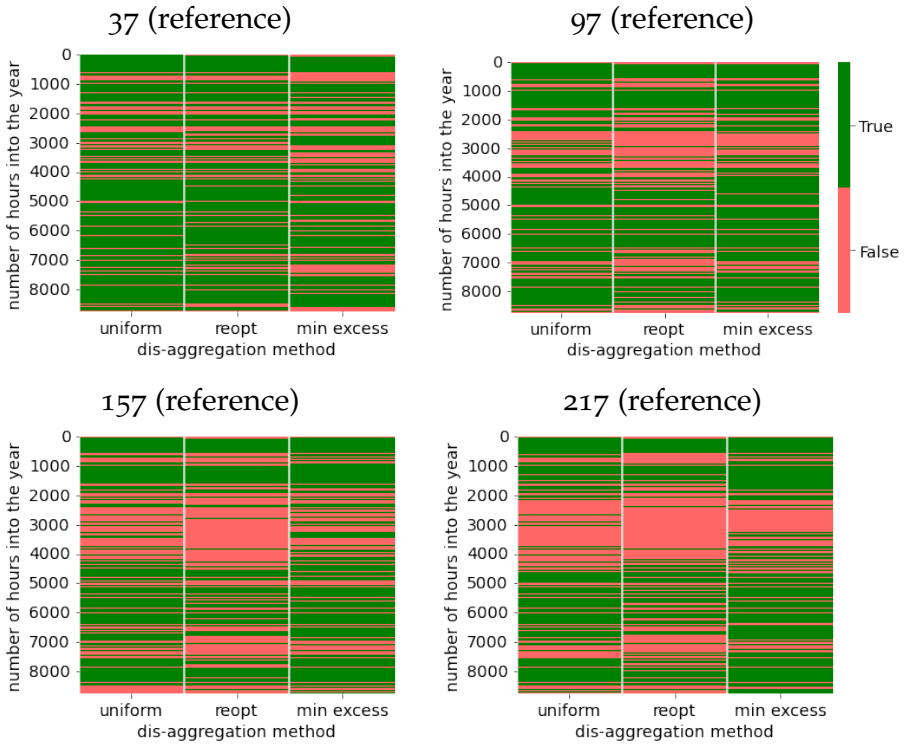


Figure 62: Evaluation if loadshedding measures of the dis-aggregated high-resolution models occur at times with higher curtailment compared to the lower resolved reference model, see equations (45)-(46).

It can be seen that, as the reference model resolution increases, there are more and more times t where the hypothesis is wrong. The amount of curtailed electricity is higher than loadshedding in 85% of the times on average for all of the three dis-aggregation methods for a very low-resolved reference model of 37 nodes. As the reference model resolution increases to 217 nodes, the statement is only true in average for 65% of the times for all three dis-aggregation methods. This indicates that transmission resolution is starting to saturate, however is still the major bottleneck preventing to feed-in the extra green electricity that is being curtailed.

FUNDING

This work was funded by the Helmholtz Association under grant no. VH-NG-1352 and the program "Energy System Design".

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/