

Computational and Near-Optimal Trade-Offs in Renewable Electricity System Modelling

Zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

von der KIT-Fakultät für Informatik
des Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

M. Sc. Fabian Neumann

Tag der mündlichen Prüfung:

29. Juli 2021

Referenten:

Prof. Dr. Veit Hagenmeyer
Karlsruher Institut für Technologie (KIT)

Prof. Dr. Thomas Brown
Technische Universität Berlin (TUB)

Prof. Dr. Joseph DeCarolis
North Carolina State University (NCSU)

Different licences apply for different parts of this thesis:



The abstract (p. v), [Chapters 1 to 3](#) and [6 to 9](#), and the transitional *overview* and *summary* sections (pp. [25](#), [95](#), [99](#) and [145](#)) are licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0): <https://creativecommons.org/licenses/by/4.0/deed.en>

[Chapter 4](#) is a reprint of Neumann, F. & Brown, T. *Transmission Expansion Planning Using Cycle Flows* in *Proceedings of the Eleventh ACM International Conference on Future Energy Systems (ACM e-Energy 2020)* (2020), 253–263. doi:[10/d3qk](#). arXiv: [2004.08702](#). ©2020 Fabian Neumann and Tom Brown. Publication rights licensed to ACM. Reprinted with permission.

[Chapter 5](#) is a reprint of Neumann, F. & Brown, T. *Heuristics for Transmission Expansion Planning in Low-Carbon Energy System Models* in *16th International Conference on the European Energy Market* (2019), 1–8. doi:[10/d295](#). arXiv: [1907.10548](#). ©2019 IEEE. Reprinted with permission.

In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of Karlsruhe Institute of Technology's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

Fabian Neumann

Computational and Near-Optimal Trade-Offs in Renewable Electricity System Modelling

Doctoral Thesis

Revision: September 13, 2021

Date of examination: July 29, 2021

Reviewers: Prof. Dr. Veit Hagenmeyer, Prof. Dr. Thomas Brown, Prof. Dr. Joseph DeCarolis

Karlsruhe Institute of Technology (KIT)

KIT Department of Informatics

Institute of Automation and Applied Informatics (IAI)

Hermann-von-Helmholtz-Platz 1

76344 Eggenstein-Leopoldshafen

Abstract

In the decades to come, the European electricity system must undergo an unprecedented transformation to avert the devastating impacts of climate change. To devise various possibilities for achieving a sustainable yet cost-efficient system, in the thesis at hand, we solve large optimisation problems that coordinate the siting of generation, storage and transmission capacities. Thereby, it is critical to capture the weather-dependent variability of wind and solar power as well as transmission bottlenecks. In addition to modelling at high spatial and temporal resolution, this requires a detailed representation of the electricity grid. However, since the resulting computational challenges limit what can be investigated, compromises on model accuracy must be made, and methods from informatics become increasingly relevant to formulate models efficiently and to compute many scenarios.

The first part of the thesis is concerned with justifying such trade-offs between model detail and solving times. The main research question is how to circumvent some of the challenging non-convexities introduced by transmission network representations in joint capacity expansion models (CEMs) while still capturing the core grid physics. We first examine tractable linear approximations of power flow and transmission losses. Subsequently, we develop an efficient reformulation of the discrete transmission expansion planning (TEP) problem based on a cycle decomposition of the network graph, which conveniently also accommodates grid synchronisation options. Because discrete investment decisions aggravate the problem's complexity, we also cover simplifying heuristics that make use of sequential linear programming (SLP) and retrospective discretisation techniques.

In the second half, we investigate other trade-offs, namely between least-cost and near-optimal solutions. We systematically explore broad ranges of technologically diverse system configurations that are viable without compromising the system's overall cost-effectiveness. For example, we present solutions that avoid installing onshore wind turbines, bypass new overhead transmission lines, or feature a more regionally balanced distribution of generation capacities. Such alternative designs may be more widely socially accepted, and, thus, knowing about these degrees of freedom is highly policy-relevant. The method we employ to span the space of near-optimal solutions is related to modelling-to-generate-alternatives (MGA), a variant of multi-objective optimisation (MOO). The robustness of our results is further strengthened by considering technology cost uncertainties. To efficiently sweep the cost parameter space, we leverage multi-fidelity surrogate modelling techniques using sparse polynomial chaos expansion (PCE) in combination with low-discrepancy sampling and extensive parallelisation on high-performance computing infrastructure.

Publications

Main Publications:

- Neumann, F. & Brown, T. *Transmission Expansion Planning Using Cycle Flows in Proceedings of the Eleventh ACM International Conference on Future Energy Systems (ACM e-Energy 2020)* (2020), 253–263. doi:[10/d3qk](https://doi.org/10/d3qk). arXiv: [2004.08702](https://arxiv.org/abs/2004.08702). **peer-reviewed**, cited as [1].
- Neumann, F. & Brown, T. The Near-Optimal Feasible Space of a Renewable Power System Model. *Electric Power Systems Research* **190**, 106690. doi:[10/ghcpr2](https://doi.org/10/ghcpr2). arXiv: [1910.01891](https://arxiv.org/abs/1910.01891) (2021). Presented at *21st Power Systems Computation Conference 2020 (PSCC)*, **peer-reviewed**, cited as [3].
- Neumann, F. Costs of Regional Equity and Autarky in a Renewable European Power System. *Energy Strategy Reviews* **35**, 100652. doi:[10/gjr2fb](https://doi.org/10/gjr2fb). arXiv: [2007.08379](https://arxiv.org/abs/2007.08379) (2021). **peer-reviewed**, cited as [4].
- Neumann, F. & Brown, T. *Heuristics for Transmission Expansion Planning in Low-Carbon Energy System Models in 16th International Conference on the European Energy Market* (2019), 1–8. doi:[10/d295](https://doi.org/10/d295). arXiv: [1907.10548](https://arxiv.org/abs/1907.10548). **peer-reviewed**, cited as [2].
- Neumann, F., Hagenmeyer, V. & Brown, T. Approximating Power Flow and Transmission Losses in Coordinated Capacity Expansion Problems. arXiv: [2008.11510](https://arxiv.org/abs/2008.11510) (2020). **under review**, cited as [5].
- Neumann, F. & Brown, T. Broad Ranges of Investment Configurations for Renewable Power Systems, Robust to Cost Uncertainty and Near-Optimality. eprint: <https://bit.ly/3fxGEAb> (2021). **in preparation**, cited as [6].

Further Publications:

- Rose, P. K. & Neumann, F. Hydrogen refueling station networks for heavy-duty vehicles in future power systems. *Transportation Research Part D: Transport and Environment* **83**, 102358. doi:[10/d7tn](https://doi.org/10/d7tn). arXiv: [1908.10119](https://arxiv.org/abs/1908.10119) (2020). **peer-reviewed**, cited as [7].

- Sun, W., Neumann, F. & Harrison, G. P. Robust scheduling of Electric Vehicle Charging in LV Distribution Networks under Uncertainty. *IEEE Transactions on Industry Applications* **56**, 5785–5795. doi:[10/fccs](https://doi.org/10/fccs) (2020). **peer-reviewed**, cited as [8].
- Hofmann, F., Hampp, J., Neumann, F., Brown, T. & Hörsch, J. atlite: A Lightweight Python Package for Calculating Renewable Power Potentials and Time Series. *Journal of Open Source Software* **6**, 3294. doi:[10/gn4b](https://doi.org/10/gn4b) (2021). **peer-reviewed**, cited as [9].
- Parzen, M., Neumann, F., van der Weijde, A. H., Friedrich, D. & Kiprakis, A. Beyond cost reduction: Improving the value of energy storage in electricity systems. arXiv: [2101.10092](https://arxiv.org/abs/2101.10092) (2021). **under review**, cited as [10].
- Gazafroudi, A. S., Neumann, F. & Brown, T. Topology-based Approximations for $N - 1$ Contingency Constraints (2021). **under review**, cited as [11].

Co-Published Versions of Open-Source Software and Datasets:

- PyPSA: Python for Power System Analysis, *versions 0.14.0–0.18.0*, github.com/pypsa/pypsa, doi:[10.5281/zenodo.786605](https://doi.org/10.5281/zenodo.786605), cited as [12].
- PyPSA-Eur: An Open Optimisation Model of the European Transmission System, *versions 0.1.0–0.3.0*, github.com/pypsa/pypsa-eur, doi:[10.5281/zenodo.3601881](https://doi.org/10.5281/zenodo.3601881), cited as [13].

Contents

1	Introduction	1
2	Foundations	9
2.1	Power System Planning Problem	9
2.2	Optimisation Theory and Algorithms	12
2.3	Model of the European Power System	15
I	Model Detail versus Computational Speed	23
	Overview Part I	25
3	Approximating Power Flow and Transmission Losses	27
3.1	Introduction	27
3.2	Nonlinear Power Flow and Losses	28
3.3	Linearised Power Flow Models	31
3.4	Simulation Setup	40
3.5	Results and Discussion	43
3.6	Critical Appraisal	54
3.7	Conclusion	54
4	Transmission Expansion Planning Using Cycle Flows	57
4.1	Introduction	57
4.2	Linear Optimal Power Flow	59
4.3	Transmission Expansion Planning	63
4.4	Simulation Setup	73
4.5	Results and Discussion	74
4.6	Conclusion	78
5	Transmission Expansion Planning Using Heuristics	79
5.1	Introduction	79
5.2	Problem Description	80
5.3	Heuristic Approaches	84
5.4	Case Study of German Transmission System	86
5.5	Results and Discussion	87
5.6	Critical Appraisal	92
5.7	Conclusion	93
	Summary Part I	95

II Cost-Optimality versus Near-Optimal Alternatives	97
Overview Part II	99
6 Near-Optimal Solutions	101
6.1 Introduction	101
6.2 Methods	102
6.3 Results and Discussion	107
6.4 Critical Appraisal	114
6.5 Conclusion	114
7 Robust Near-Optimal Solutions	115
7.1 Introduction	115
7.2 Methods	116
7.3 Results and Discussion	126
7.4 Critical Appraisal	135
7.5 Conclusion	135
8 Regional Equity and Autarky	137
8.1 Introduction	137
8.2 Model Inputs and Simulation Setup	139
8.3 Results and Discussion	139
8.4 Critical Appraisal	143
8.5 Conclusion	144
Summary Part II	145
9 Conclusions and Outlook	147
Bibliography	153
List of Abbreviations	175
List of Figures	177
List of Tables	179

Introduction

As governments across the world are planning to increase their shares of renewable energy supply, energy system models have become a pivotal instrument to find cost-efficient system layouts that satisfy ambitious climate change mitigation targets [14–17]. The commitment to reaching net-zero emissions by mid-century to prevent the calamitous consequences of global warming means we need to find answers today to what infrastructure is required, where it should be built, and how much it costs [18]. Thereby, the speed of change needed to achieve these goals is unprecedented. Because a sustainable energy system will radically differ from today's system and the transition will entail exceptional financial outlay throughout the upcoming decades, detailed modelling becomes crucial to address these questions systematically.

Most analysis indicates that future energy systems will primarily be based on wind and solar generation because they represent two low-emission technologies that unite rapidly falling costs with more than enough expansion potential to cover worldwide energy demands [19]. But the chief challenge that arises for systems with high shares of weather-dependent renewable energy sources is dealing with their variability. Throughout a year, a system based on wind and solar power is exposed to many different supply situations.

In recent years, the rising shares of variable renewable energy have increasingly put power transmission grids under strain. The connection of wind turbines to the grid far from demand has led to frequent occasions of high network loading in countries such as Denmark, Germany and the United Kingdom, resulting in high levels of wind curtailment. In 2019, grid congestion caused curtailment in Germany of around 3% of renewable energy produced [20]. This raises questions about the system integration of more and more renewables.

Fortunately, there is a range of different choices that help us harmonise the rising spatio-temporal mismatch patterns of supply and demand. Preferable to maintaining fossil backup power plants and wastefully curtailing renewable generation, in the electricity sector, we have the options to balance renewable generation in space by transmitting energy across long distances or in time by storing energy. Continent-spanning transmission networks can smooth over multiple weather systems and connect remote high yield sites to demand centres. Conversely, storage technologies shift energy from times of surplus to times of deficit, whereas demand-side management works the other way around.

Thereby, different balancing technologies satisfy different balancing needs. For instance, batteries and pumped-hydro storage can deal well with the dominant daily variations of solar generation. In contrast, the characteristic weekly variability of wind power tends to pair better with hydrogen storage or hydro reservoirs. Both wind and solar variability also feature antithetical seasonal components that partially balance out if mixed [21]. Another prominent example concerns the spatial allocation of renewable generation capacities, namely whether it is better to build offshore wind turbines and the transmission network required to deliver the electricity or to locate renewable generation closer to where demand is while accepting a setback in productivity and possibly using energy storage. Both electricity networks and storage also induce an indirect cost from transmission and conversion losses, which affect the cost-optimal amount and location of renewable capacities.

Because the interactions between demand, generation, and balancing options are complex and cost-efficient solutions are not obvious, we build optimisation models of the energy system to be able to assess the different options thoroughly. These are known as investment planning models, capacity expansion models (CEMs) or, more recently, macro-energy systems models [22]. Their purpose is to weigh the many possible investment decisions in planning sustainable energy systems against each other.

Broadly speaking, in these models, we adopt the perspective of a social planner. We ask what is the cost-minimal system design and operation that satisfies emission reduction targets given that there are some technical constraints we must obey. These include meeting the demand at each location and hour of the year whilst considering the spatio-temporal availability and geographical potentials of variable renewables like wind and solar. Moreover, we must account for the limits of transmission capacities and the physics of power flow. In the context of these formalised mathematical models, we often initially abstract from market integration aspects, regulation and other policy instruments (e.g. subsidies and taxes) and, following equilibrium theory, idealise by assuming perfect competition among the technologies as well as long-term dynamic efficiency. Put into an economist's words, we search for the system composition that maximises the long-run social welfare.

The modelling requirements for systems with high shares of renewables are radically different than for conventional systems with only dispatchable power generation. In the absence of weather-dependent renewables, simple heuristics like screening curves used to suffice to dimension fossil-fuelled power plants [23]. However, the focus on variable renewables needs much higher model resolutions and an integrated system perspective: it has been demonstrated that such models require coordinated planning of generation, storage, and transmission infrastructure to capture their complex interactions and strong dependencies [24, 25]; high temporal resolution and scope to account for extreme weather events, storage

operation, and investments shaped by the characteristic daily, synoptic and seasonal patterns of renewables and load [26, 27]; high spatial resolution and scope to also capture the spatio-temporal patterns, such as correlations of wind speeds across the continent, and to represent grid bottlenecks [28, 29].

Many of the existing models have not been able to meet these modelling needs. Frequently, there is no joint planning which ignores the trade-offs between the siting of generation capacity, grids and storage [30, 31]. Most often, analyses are based on fragmented, insufficiently many typical days [32–34], which cannot capture the full variability of renewables and demand, or extreme weather events that test the reliability of the systems. Many models also do not surpass a country-level model resolution [35–38]. Such a low spatial resolution also means that only a simple representation of power flows based on net transfer capacities (NTCs) is applicable [14]. All too often, these are moreover closed-source models, which means that results are neither transparent, reproducible, nor verifiable for outsiders (e.g. LEAP [39] and, until 2020, TIMES [40]).

On the contrary, the open-source power system model, `PyPSA-Eur` [13], which will be used in this thesis, fulfils many of the prerequisites mentioned above. By combining detailed grid modelling with joint spatially explicit capacity expansion planning across thousands of operating conditions, it builds a bridge between broad energy system models (e.g. `calliope` [41], `oemof` [42], `OSeMOSYS` [43], `Temoa` [44]), for which power flow modelling does not exceed NTCs, and detailed power system analysis tools (e.g. `pandapower` [45], `PYPOWER` [46], `PowerModels.jl` [47]), that conversely do not perform coordinated investment optimisation.

Indeed, there has never been a more opportune time for raising model detail since Sørensen’s pioneering efforts in renewable energy system modelling in the 1970s [48, 49]. Driven by the growing urgency of climate change mitigation and encouraged by the plummeting costs of renewables in the past decade [50–52], we can now avail of decades-spanning historical weather datasets [53], substantial advances in algorithms to solve large-scale problems [54], and extensive performance optimisation in open-source modelling software like `PyPSA` [12]. Moreover, parallel computation of optimisation runs on high-performance clusters with modern hardware has become widely accessible, allowing for various model configurations and scenarios to be studied even at high resolution.

Nonetheless, solving sufficiently detailed CEMs still presents a computational challenge, even with state-of-the-art software. Besides the need for high spatial and temporal detail and the joint planning, which drives up the overall problem size, representing electricity transmission networks adds considerably to the problem complexity [55–57]. This is because correctly accounting for the physics of prevalent AC power flow and transmission losses introduces nonlinear constraints, and aspects of

transmission expansion planning (TEP) typically involve discrete decisions, each of which makes the optimisation model non-convex and, hence, NP-hard to solve [58].

Consequently, all aspects of the energy system at meticulous levels of detail cannot currently be contained in a single model. Instead, we must make simplifications. For instance, in the present thesis, we limit the scope of analysis to the electricity sector of Europe rather than modelling the entire global energy system, adjust the model resolution by spatial and temporal aggregation according to limits and needs, and compromise on how accurately we represent power flows. One part of the thesis is about justifying such compromises between model accuracy and computational speed. Focussing on the representation of electricity transmission networks in coordinating CEMs, we investigate what model details can be simplified, or formulated more efficiently, to remain sufficiently accurate within acceptable computation times.

Quantifying these modelling trade-offs is highly relevant because, with a justifiable complexity reduction in one dimension, we would be able to increase model detail in another, where coarse approximations may have a bigger effect [59, 60]. Also, even though capacity expansion planning is a strategic problem, computational speed matters for reasons beyond mere solvability. As many influential model parameters, like technology cost assumptions, reference weather years, technology choices, and resource potentials contain a high level of uncertainty, performing sensitivity analyses is crucial but often limited by computational resources and time at hand. The resulting common focus of design studies on a single or few scenarios misses not only parametric but also structural modelling uncertainties that result from the choices made to simplify the model's representation of reality [61].

Hence, it is highly advisable to also look beyond single least-cost solutions, which can easily give a false sense of exactness [62, 63]. In retrospective analyses, it has been shown that developments in the electricity sector have rarely followed cost-optimal paths [64, 65]. Additionally, modelling results are known to be sensitive to small perturbations of input parameters because the space of feasible solutions near the optimum is very flat. It is, therefore, widely uncontroversial that while we can reveal insights about the key technology interactions that shape investments, given the modelling uncertainties involved, there is little we can tell with certainty about the exact capacities to be built [66, 67]. Furthermore, pure least-cost optimisation rarely reflects societal preferences and attitudes towards renewable energy infrastructures. Local communities may oppose new overhead transmission lines and onshore wind turbine developments for various reasons. Because social acceptance is critical for a successful energy transition, it would be negligent not to contemplate alternative system designs which would be more widely accepted, even if they entail a limited additional cost. Such trade-offs are investigated in the second half of this thesis.

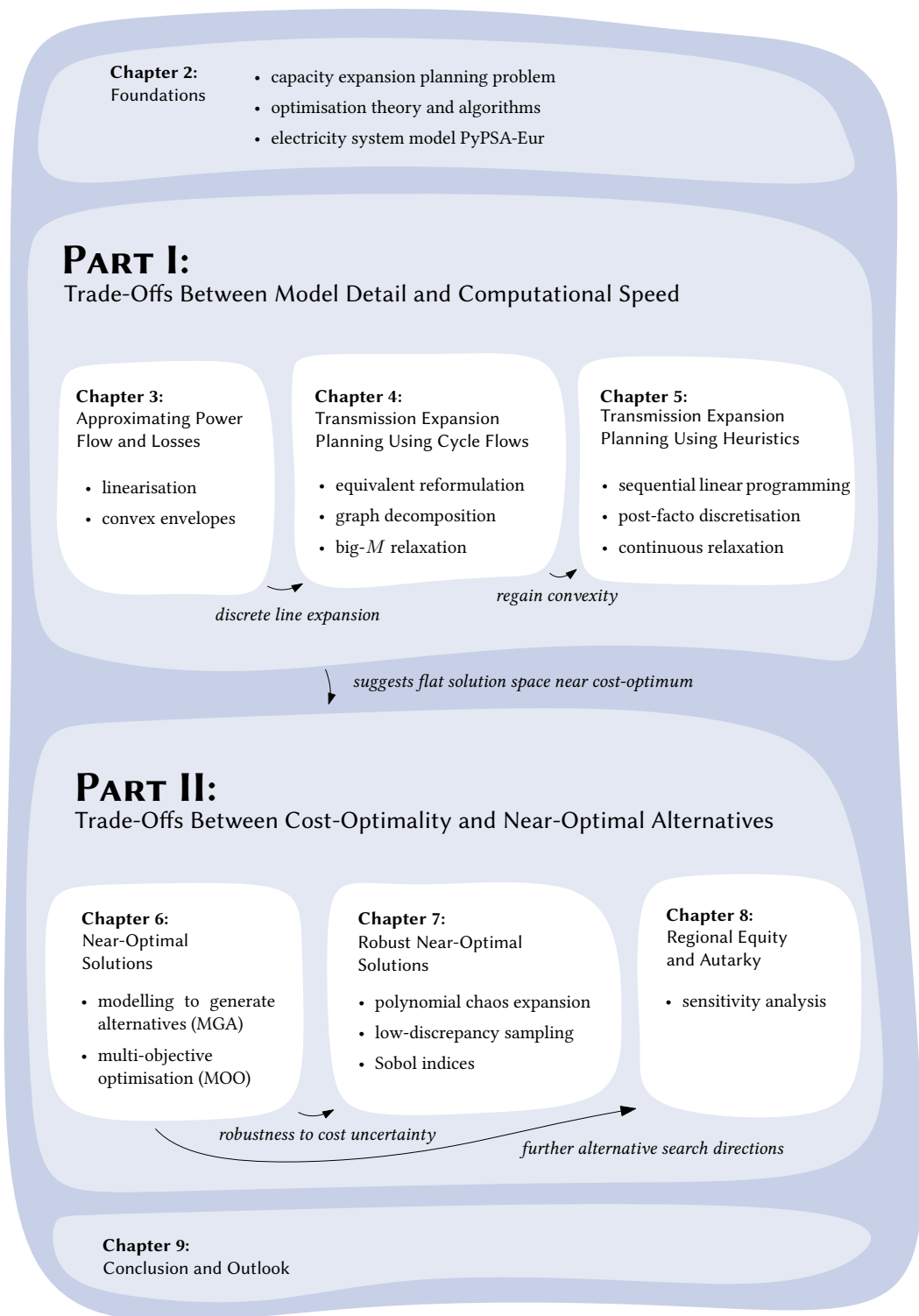


Figure 1.1: Graphical outline of the contents and methods of the present thesis.

Outline and Contributions

Overall, compromises and trade-offs are the common themes of the thesis. Following [Figure 1.1](#), the present thesis is organised in two major parts, enclosed by preliminary considerations and concluding remarks, each examining insightful compromises in power system planning from different angles.

[Part I](#) concerns trade-offs between model detail and computation times. From a modelling and algorithmic perspective, we seek to address the question of how we can circumvent some of the computationally challenging non-convexities induced by [AC](#) power flow equations, transmission losses, and discrete [TEP](#) while still capturing the core grid physics that many current energy system planning models ignore.

On the other hand, [Part II](#) examines trade-offs between cost-optimal solutions and marginally more expensive alternative system designs. Driven by the prospect of accommodating unmodelled social constraints, we explore broad ranges of robust investment configurations that are technologically diverse while taking account of technology cost uncertainty. In the context of this analysis, we apply various methods from sensitivity and uncertainty analysis, such as multi-objective optimisation ([MOO](#)), modelling-to-generate-alternatives ([MGA](#)), polynomial chaos expansion ([PCE](#)), the determination of Sobol indices, and low-discrepancy sampling techniques in a high-performance computing environment to cope with the incurred plethora of scenarios in parallel.

[Chapter 2 – Foundations](#) contains the preliminaries. Throughout the thesis, we consistently use variants of the same basic problem formulation, the same model of the European electricity system, and employ algorithms implemented in state-of-the-art software to solve the resulting optimisation problems. This chapter introduces these.

[Chapter 3 – Approximating Power Flow and Transmission Losses \[5\]](#)

starts [Part I](#) with approximations of two nonlinear phenomena in power systems, [AC](#) power flow and transmission losses, in the context of linear [CEMs](#). To navigate around the computationally burdensome non-convexities of the resulting nonlinear problems ([NLPs](#)), we apply convex envelopes in addition to further linearisation and relaxation techniques to derive various simplified power flow representations. We then compare and evaluate these based on differences in investment decisions, nodal prices, computational performance, and, foremost, by how much the modelled flows and losses deviate from simulated [AC](#) power flows. Particularly the latter is relevant to judge whether the model results from coarsely-resolved planning models are sufficiently physical to be used in more detailed feasibility studies.

Chapter 4 – Transmission Expansion Planning Using Cycle Flows [1]

turns to another prominent non-convexity in power system planning: the discrete expansion of transmission lines with changing effective line impedances when transmission corridors are reinforced. Using a big- M disjunctive relaxation, this involves a mixed-integer linear problem (MILP) that is commonly founded on a power flow linearisation that uses auxiliary optimisation variables for voltage angles to describe Kirchhoff's voltage law (KVL). As well as introducing a large number of auxiliary variables, this angle-based TEP formulation has the disadvantage that it is not well-suited to considering the connection of multiple disconnected networks. By exploiting insights from graph theory, we derive an equivalent reformulation of TEP based on a cycle decomposition of the network graph. This alternative formulation circumvents the auxiliary variables and reduces the required number of constraints by expressing KVL directly in terms of the power flows.

Since the new formulation neither compromises on model detail nor resolves the non-convexity of discrete line expansion options, unlike the other chapters of Part I, the focus is on computational benefits achieved solely by restating the solution space more efficiently. To demonstrate this, we compare the cycle-based reformulation to the standard angle-based TEP formulation and emphasise the combinatorics of the connection of multiple disconnected networks. This topic had not received attention in the literature but has particular relevance to regions like Europe, North America, China, Japan, and the Philippines with island systems or multiple synchronous zones.

Chapter 5 – Transmission Expansion Planning Using Heuristics [2]

continues the search for efficient ways to incorporate integer transmission expansion and changing line impedances. Since as a MILP even the cycle-based reformulation of TEP is hard to solve with state-of-the-art solvers in coordinating capacity expansion planning, in this chapter, we follow an alternative approach. To regain the lower computational effort of convex problems, we develop heuristics based on sequential linear programming (SLP), relaxation and retrospective discretisation, and benchmark their performance against the MILP formulation in terms of their computational gain, deviation from optimal system cost, and similarity of transmission line expansion.

Chapter 6 – Near-Optimal Solutions [3] initiates Part II, which centres on the perils of narrowly following optimisation results in electricity system planning. As previous modelling-focused chapters indicate degrees

of freedom in siting infrastructure without compromising the system's cost-effectiveness, in this chapter, we investigate the feasible space near the cost-optimum more systematically. We apply methods from multi-objective optimisation (MOO) and modelling-to-generate-alternatives (MGA) to delineate trade-offs between increasing system cost and minimising or maximising the build-out of individual technologies for generating, storing and transmitting electricity. Especially along dimensions that affect levels of public acceptance, e.g. the volume of transmission expansion or the scale of onshore wind capacities, we aim to establish a better understanding of which technologies are essential for given cost penalties. This allows us to communicate a broad spectrum of possibilities to policymakers.

Chapter 7 – Robust Near-Optimal Solutions [6] subsequently augments the robustness of outlined alternatives by additionally taking account of a wide range of different technology cost projections, given that rapidly evolving technologies like solar photovoltaics, batteries and hydrogen storage contain a high level of uncertainty. To manage the explosion of scenarios induced by searching for many near-optimal solutions alongside comprehensively sweeping the cost parameter space, we leverage multi-fidelity surrogate modelling techniques using sparse PCEs in combination with advanced low-discrepancy sampling techniques. Besides being able to allocate induced output uncertainty to individual cost parameters through Sobol indices, thereby, we can underpin alternative solutions beyond least-cost that have a high chance of involving a limited cost increase, just as we can identify regions of the solution space that are unlikely to be cost-efficient.

Chapter 8 – Regional Equity and Autarky [4] explores further dimensions of near-optimal alternatives besides the technology-oriented axes of **Chapters 6 to 7**. It is motivated by the observation that the purely resource-induced cost optimum can lead to very inhomogeneous regional distributions of capacities in a fully renewable European electricity system. This is deemed to be another factor that can be detrimental to levels of social acceptance. Within the context of sensitivity analysis, we explore the cost impact and changes in the system composition when the development of power supply infrastructure is more evenly shared among countries and smaller regions. Thereby, we also assess fully self-sufficient solutions without electricity transmission between individual regions.

Chapter 9 – Conclusions and Outlook offers a concluding review of the thesis, summarising the main insights from across all chapters and outlining worthwhile extensions to the conducted work for future investigations.

Foundations

The purpose of the present chapter is to set the foundations. It introduces the optimisation problem for joint long-term capacity expansion planning of electricity systems with high shares of renewables and energy storage (Section 2.1), and describes how can we solve such optimisation problems in general (Section 2.2) as well as how realistic inputs are obtained to model the European power system (Section 2.3).

2.1 Power System Planning Problem

For systems with high shares of renewables, the cost-optimal allocation of renewable generation capacities, energy storage and transmission reinforcements to supply electricity demands is strongly driven by operational aspects. Variable weather conditions largely determine generator dispatch, grid congestion, and storage needs. Because long-term investments in electricity system infrastructure and short-term system operation are so heavily intertwined, in this thesis, we assess different capacity expansion options based on an optimisation model that jointly optimises both, subject to physical, technical, and environmental constraints.

A successfully solved model returns both the optimised locations of generation, storage, and transmission capacities as well as the optimal dispatch of generators, storage units, and controllable network components like HVDC links throughout a reference year. It also shows levels of curtailment, network congestion, and nodal electricity prices.

In this partial equilibrium model [68], we seek to minimise the total annual system costs that comprises both investment costs and operational expenditures. To express both as annual costs, we use the annuity factor $(1 - (1 + \tau)^{-n})/\tau$ that, like a mortgage, converts the upfront investment of an asset to annual payments considering its lifetime n and cost of capital τ . Thus, the objective includes on one hand the annualised capital costs c_* for investments at locations i in generator capacity $G_{i,r} \in \mathbb{R}^+$ of technology r , storage power capacity $H_{i,s} \in \mathbb{R}^+$ of technology

s, and transmission line capacities $P_\ell \in \mathbb{R}^+$, as well as the variable operating costs $o_{i,r}$ for generator dispatch $g_{i,r,t} \in \mathbb{R}^+$ on the other:

$$\min_{G,H,F,g} \left[\sum_{i,r} c_{i,r} \cdot G_{i,r} + \sum_{i,s} c_{i,s} \cdot H_{i,s} + \sum_{\ell} c_{\ell} \cdot P_{\ell} + \sum_{i,r,t} w_t \cdot o_{i,r} \cdot g_{i,r,t} \right]. \quad (2.1)$$

Thereby, the representative time snapshots t are weighted by the time span w_t such that their total duration adds up to one year; $\sum_{t \in \mathcal{T}} w_t = 365 \cdot 24\text{h} = 8760\text{h}$. The subsequent problem description largely follows [5].

In addition to the cost-minimising objective function, we further impose a set of linear constraints that define limits on (i) the capacities of generation, storage and transmission infrastructure from geographical and technical potentials, (ii) the availability of variable renewable energy sources for each location and point in time, typically derived from historical weather data, (iii) the budget for greenhouse gas (GHG) emissions, (iv) storage consistency equations, and (v) a multi-period linear optimal power flow (LOPF) formulation. Overall, this results in a linear problem (LP).

The capacities of generation, storage and transmission infrastructure are constrained from above by their installable potentials and from below by any existing components:

$$\underline{G}_{i,r} \leq G_{i,r} \leq \overline{G}_{i,r} \quad \forall i, r \quad (2.2)$$

$$\underline{H}_{i,s} \leq H_{i,s} \leq \overline{H}_{i,s} \quad \forall i, s \quad (2.3)$$

$$\underline{P}_{\ell} \leq P_{\ell} \leq \overline{P}_{\ell} \quad \forall \ell \quad (2.4)$$

Moreover, the dispatch of a generator may not only be constrained by its rated capacity but also by the availability of variable renewable energy, which may be derived from reanalysis weather data. This can be expressed as a time- and location-dependent availability factor $\overline{g}_{i,r,t}$, given per unit of the generator's capacity:

$$0 \leq g_{i,r,t} \leq \overline{g}_{i,r,t} G_{i,r} \quad \forall i, r, t \quad (2.5)$$

The dispatch of storage units is split into two positive variables; one each for charging $h_{i,s,t}^+$ and discharging $h_{i,s,t}^-$. Both are limited by the power rating $H_{i,s}$ of the storage units.

$$0 \leq h_{i,s,t}^+ \leq H_{i,s} \quad \forall i, s, t \quad (2.6)$$

$$0 \leq h_{i,s,t}^- \leq H_{i,s} \quad \forall i, s, t \quad (2.7)$$

This formulation does not prevent simultaneous charging and discharging, in order to maintain the computational benefit of a convex feasible space but can be counter-

acted by adding a small marginal cost to the storage dispatch variables. The energy levels $e_{i,s,t}$ of all storage units have to be consistent with the dispatch in all hours

$$e_{i,s,t} = \eta_{i,s,0}^{w_t} \cdot e_{i,s,t-1} + w_t \cdot h_{i,s,t}^{\text{inflow}} - w_t \cdot h_{i,s,t}^{\text{spillage}} + \eta_{i,s,+} \cdot w_t \cdot h_{i,s,t}^+ - \eta_{i,s,-}^{-1} \cdot w_t \cdot h_{i,s,t}^- \quad \forall i, s, t \quad (2.8)$$

whereby storage units can have a standing loss $\eta_{i,s,0}$, a charging efficiency $\eta_{i,s,+}$, a discharging efficiency $\eta_{i,s,-}$, natural inflow $h_{i,s,t}^{\text{inflow}}$ and spillage $h_{i,s,t}^{\text{spillage}}$. Furthermore, the storage energy levels are assumed to be cyclic given that we consider a single reference year,

$$e_{i,s,0} = e_{i,s,|\mathcal{T}|} \quad \forall i, s, \quad (2.9)$$

and are constrained by their energy capacity

$$0 \leq e_{i,s,t} \leq \bar{T}_s \cdot H_{i,s} \quad \forall i, s, t. \quad (2.10)$$

To reduce the number of decision variables, here we link the energy storage volume to power ratings using a technology-specific parameter \bar{T}_s that describes the maximum duration a storage unit can discharge at full power rating.

Regarding the balance of supply and demand, Kirchhoff's current law (KCL) requires local generators and storage units as well as incoming or outgoing flows $f_{\ell,t}$ of incident transmission lines ℓ to balance the perfectly inelastic electricity demand $d_{i,t}$ at each location i and snapshot t

$$\sum_r g_{i,r,t} + \sum_s (h_{i,s,t}^- - h_{i,s,t}^+) + \sum_{\ell} K_{i\ell} f_{\ell,t} = d_{i,t} \quad \forall i, t, \quad (2.11)$$

where $K_{i\ell}$ is the incidence matrix of the network.

The power flows $p_{\ell,t}$ are limited by their nominal capacities P_{ℓ}

$$|p_{\ell,t}| \leq \bar{p}_{\ell} P_{\ell} \quad \forall \ell, t, \quad (2.12)$$

where \bar{p}_{ℓ} acts as an additional per-unit security margin on the line capacity to allow a buffer for the failure of single circuits ($N - 1$ criterion) and reactive power flows.

Kirchhoff's voltage law (KVL) imposes further constraints on the flow of AC lines and there are several ways to formulate KVL with large impacts on performance. As we will deal with the modelling implications of KVL in greater detail later in Chapter 3, for now we complete the problem description with one of them: Using linearised load flow assumptions, the voltage angle difference around every closed cycle in the network must add up to zero. Using a cycle basis C_{ℓ_c} of the

network graph where the independent cycles c are expressed as directed linear combinations of lines ℓ [69], we can write **KVL** as

$$\sum_{\ell} C_{\ell c} \cdot x_{\ell} \cdot f_{\ell,t} = 0 \quad \forall c, t \quad (2.13)$$

where x_{ℓ} is the series inductive reactance of line ℓ . Controllable **HVDC** links are not affected by this constraint.

We may further regard a constraint on the total annual CO_2 emissions Γ_{CO_2} to achieve sustainability goals. The emissions are determined from the time-weighted generator dispatch $w_t \cdot g_{i,r,t}$ using the specific emissions ρ_r of fuel r and the generator efficiencies $\eta_{i,r}$

$$\sum_{i,r,t} \rho_r \cdot \eta_{i,r}^{-1} \cdot w_t \cdot g_{i,r,t} \leq \Gamma_{\text{CO}_2}. \quad (2.14)$$

Note, that this formulation does not include pathway optimisation (i.e. no sequences of investments), but searches for a cost-optimal layout corresponding to a given **GHG** emission reduction level and assumes perfect foresight for the reference year based on which capacities are optimised. This optimisation problem is implemented in the open-source Python-based modelling framework **PyPSA** [12].

2.2 Optimisation Theory and Algorithms

Before we shortly proceed with populating the raw problem formulation with data, we first take a step back and briefly cover which algorithms we can use to solve optimisation problems in general.

Let us initially consider an objective function $f : \mathbb{R}^k \rightarrow \mathbb{R}$ which we seek to minimise

$$\min_x f(x). \quad (2.15)$$

In [Section 2.1](#) this corresponds to a mapping of investment and dispatch decisions $x \in \mathbb{R}^k$ to the total system cost. This decision space is further constrained by some equalities and inequalities

$$g_i(x) = c_i \quad \perp \quad \lambda_i \quad i = 1, \dots, n, \quad (2.16)$$

$$h_j(x) \leq d_j \quad \perp \quad \mu_j \quad j = 1, \dots, m. \quad (2.17)$$

For example, in [Section 2.1](#) we have inequality constraints for capacity limits and equality constraints to model the Kirchhoff laws. These constraints define the feasible space $X \subset \mathbb{R}^k$, which forms a convex polytope, a multi-dimensional polygon, if all constraints are affine. The dual variables λ_i and μ_j are the so-called Karush-Kuhn-Tucker (**KKT**) multipliers, which can be used to rewrite the constrained optimisation problem as a Lagrangian function

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_i \lambda_i (g_i(x) - c_i) + \sum_j \mu_j (h_j(x) - d_j). \quad (2.18)$$

Based on this Lagrangian function, we can apply the **KKT** conditions [\[70\]](#) formulate the necessary conditions for an optimal solution x^*, μ^*, λ^* , namely:

$$\text{stationarity: } \frac{\partial \mathcal{L}}{\partial x_\ell} = \frac{\partial f}{\partial x_\ell} + \sum_i \lambda_i^* \frac{\partial g_i}{\partial x_\ell} + \sum_j \mu_j^* \frac{\partial h_j}{\partial x_\ell} = 0 \quad (2.19)$$

$$\text{primal feasibility: } g_i(x^*) = c_i \quad (2.20)$$

$$h_j(x^*) \leq d_j \quad (2.21)$$

$$\text{dual feasibility: } \mu_j^* \geq 0 \quad (2.22)$$

$$\text{complementary slackness: } \mu_j^* (h_j(x^*) - d_j) = 0 \quad (2.23)$$

For optimisation problems with a convex objective function and exclusively affine constraints, the **KKT** conditions are even sufficient.

In the context of minimising the cost of future energy system designs, the **KKT** multipliers also have insightful economic interpretations. For instance, the **KKT** multipliers can translate a carbon cap to a carbon price, map a renewable generation target to a required subsidy, allocate a scarcity cost to limited potentials, or even convey the transmission congestion as nodal prices [\[71\]](#). As they measure the local sensitivity of the objective function to relaxing the associated constraint at the optimal point, they are frequently referred to as shadow prices.

In general, solving optimisation problems (and computing shadow prices on the side) is at worst **NP-hard**; i.e. there is no known all-encompassing algorithm that is guaranteed to solve in polynomial time. Non-convex objectives, constraints and discrete decision variables are the primary agitators of this level of complexity; and unfortunately we encounter quite a few of these in electricity system modelling. However, if we can force the investment planning problem into convexity, by formulating it as **LP** as in [Section 2.1](#), we can avail of fast polynomial-time algorithms [\[72\]](#).

Simplex Methods The simplex algorithm is one of the standard method to solve **LPs**. Starting at any vertex, it works by traversing the edges of the convex polytope that defines the feasible space. The method traces back to Dantzig's work in the late 1940s [\[73\]](#), and exploits the fact that the optimum occurs at (at least) one of

the polytope's vertices. But although the simplex method is demonstrated to solve in polynomial time on average, it requires exponential time in the worst case [74].

Interior-Point Methods On the other hand, interior-point methods (or barrier methods) are a more recent group of solving algorithms, which can even solve convex nonlinear problems (NLPs) in addition to LPs. Rather than examining the feasible space's vertices, interior-point methods approach the optimal solution from the inside by amending the objective with a barrier term that penalises solutions which come close to the feasible space boundary. In the 1980s, Karmarkar [75] developed an interior-point algorithm that completes in weakly polynomial time for LPs [76].

Branch and Bound Methods While the problem formulation in Section 2.1 did not feature any discrete decision variables so far, the expansion of transmission lines is often considered to be a discrete decision because circuits are only available in limited configurations. This results in a mixed-integer linear problem (MILP)¹, which complicates the search for the cost-optimal transmission reinforcement in some of the following chapters (4 and 5). One common practice for solving MILPs is the branch and bound method that originates from work in the 1960s by Land & Doig [77] and Dakin [78]. The branch and bound principle builds on a tree search rooted on an LP relaxation of the MILP which initially ignores all integrality constraints. Recursively the feasible space is then split on a chosen integer but fractionally optimised branching variable, whereby each new node entails a slightly more constrained LP relaxation until integer feasible solutions may be found. Based on continuously updated lower and upper bounds, parts of the tree known not to contain an optimal solution are then iteratively discarded. The lower bound climbs from the continuous relaxation's objective value, whereas the upper bound drops by leaps to the best known feasible solution. Throughout the tree search, the upper and lower bounds approach each other, and the algorithm terminates once they match up to a given tolerance.

Software For solving problem instances of both LPs and MILPs we draw on specialised software, which employs various additional speed-up techniques beyond the basic concepts sketched above. While there exist notable open-source solvers like Cbc [79] and GLPK [80], the immense computational burden incurred by resolving the European power system with hundreds of nodes and thousands of operating conditions in a single coordinating capacity expansion problem constrains us to significantly more performant commercial² solvers like Gurobi [54] or CPLEX [81] alongside utilising high-performance computing infrastructure.

¹Other common considerations that lead to MILPs in the context of power system expansion planning are so-called unit commitment constraints which impose constraints on the on-off status of individual generators, which is not regarded in this thesis' highly renewable scenarios.

²Both Gurobi and CPLEX provide free academic licenses.

2.3 Model of the European Power System

Knowing the mathematical problem formulation and how to solve such, this section outlines how we compile the data inputs for the planning problem of [Section 2.1](#) to prepare realistic case studies. For that purpose, we introduce [PyPSA-Eur](#) [13], the model of the European power system based on which the thesis' various algorithmic improvements, approximations and sensitivities are evaluated.

[PyPSA-Eur](#) is one of the only high-resolution models openly available. It is best described as a Python-based workflow that compiles and solves models of the European transmission system from publicly available data. [PyPSA-Eur](#) processes data on the topology of the power transmission network, historical time series of weather observations and electricity consumption, conventional power plants both fossil-fueled and hydro-electric, technical wind turbine and solar panel specifications, the geographic expansion potentials of wind and solar power, and cost and efficiency projections for generation, storage and transmission technologies to build a model that incorporates the complex spatio-temporal patterns that shape future investment decisions in renewable power systems with sufficient detail.

Thereby, the coming together of many different types of data presents a challenge to documentation and reproducibility. [PyPSA-Eur](#) handles this challenge by linking all intermediary processes in a dependency graph, where nodes represent scripts which are connected by edges that map outputs to inputs, using the *snakemake* workflow management tool [82].

Besides coherently compiling model instances, the [PyPSA-Eur](#) workflow is also immensely useful to conveniently manage numerous scenarios. This is because it is highly configurable. Among other settings, the spatio-temporal resolution and scope, the considered technologies, technology cost assumptions and aspired emission reduction levels can easily be varied. Additionally, *snakemake* provides user-friendly interfaces to high-performance computing infrastructure, such that many computationally burdensome scenarios can be run in parallel.

In the following, we outline the workflow's main steps and features. More details can be found in the source code repository at github.com/pypsa/pypsa-eur and the associated publication [13].

2.3.1 Transmission Network and Conventional Power Plants

The topology of the European transmission network is retrieved from the interactive [ENTSO-E](#) map [83] using a modified version of the [GridKit](#) [84] tool. As displayed in [Figure 2.1](#), the dataset includes [HVAC](#) lines at and above 220 kV across

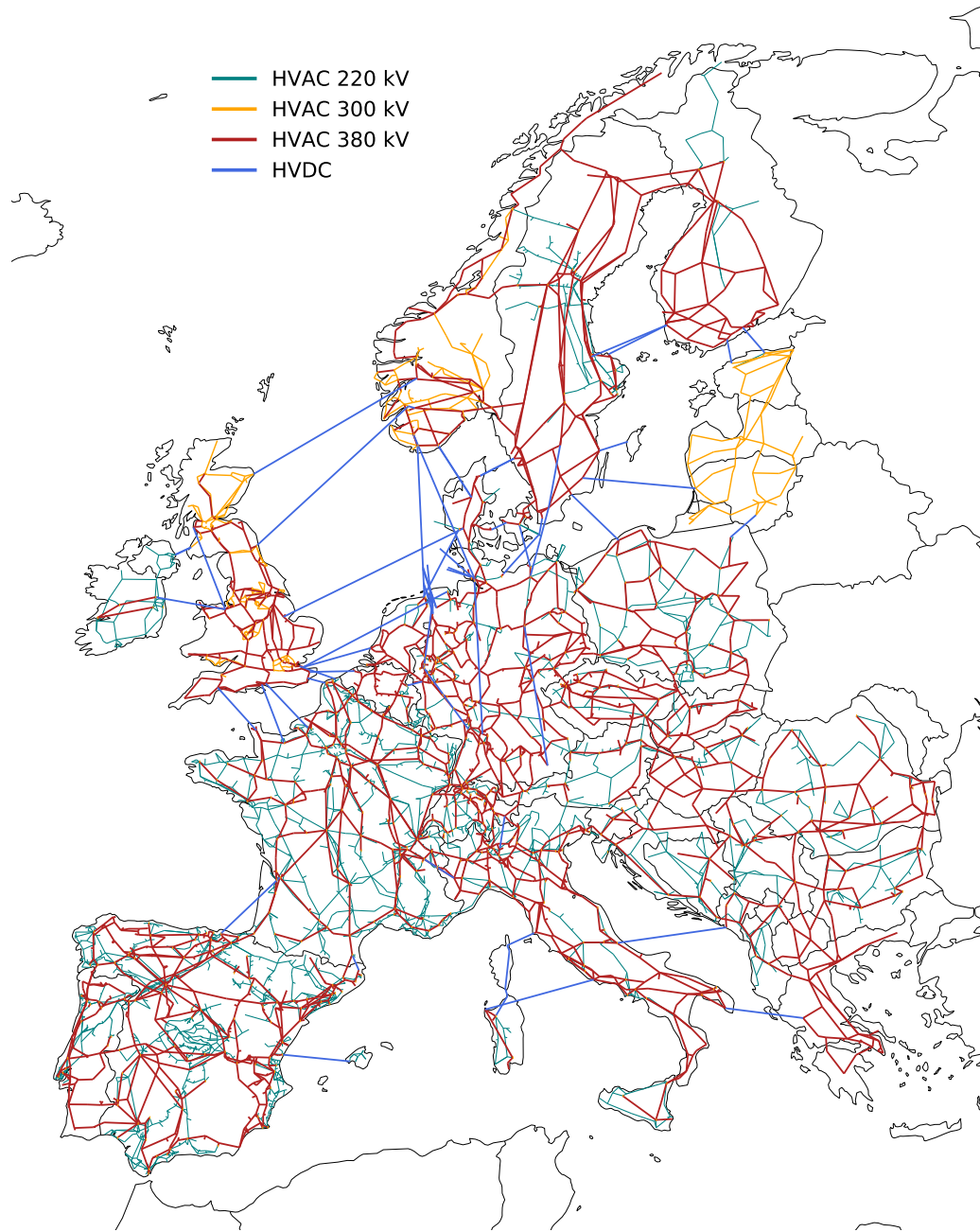
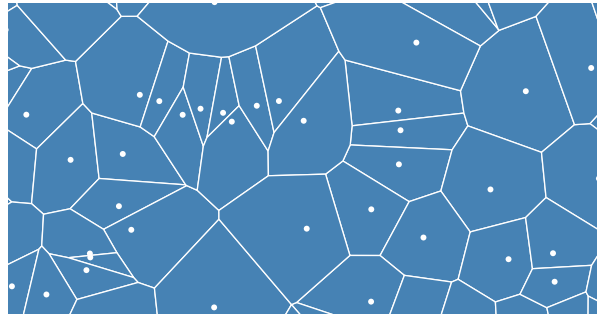


Figure 2.1: European transmission network topology at full resolution.

Figure 2.2: Exemplary Voronoi cells of the transmission network's substations.



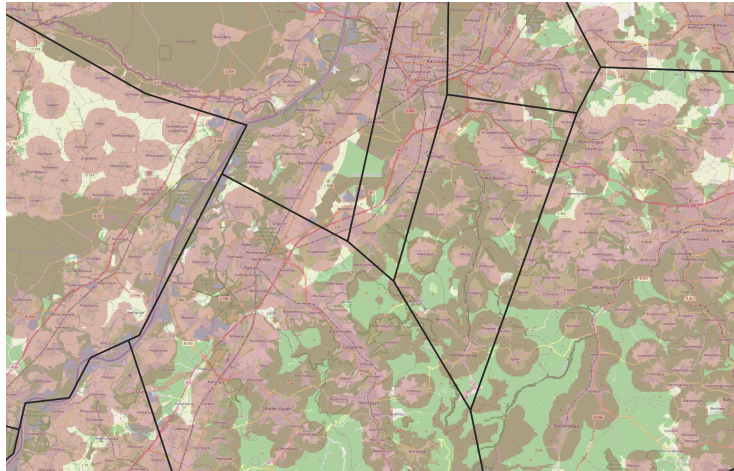
the multiple synchronous zones of the [ENTSO-E](#) area, but excludes Turkey and North-African countries which are also synchronised to the continental European grid, interconnections to Russia, Belarus and Ukraine, as well as small island networks with less than four nodes at transmission level, such as Cyprus, Crete and Malta. In total, the network encompasses around 3000 substations, 6600 [HVAC](#) lines, and around 70 [HVDC](#) links, some of which are planned projects from the Ten Year Network Development Plan ([TYNDP](#)) that are not yet in operation [85].

Capacities and electrical characteristics of transmission lines and substations, such as impedances and thermal ratings, are inferred from standard types for each voltage level. For each [HVAC](#) line, we further restrict line loading to 70% of the nominal rating to approximate $N - 1$ security, a criterion that mandates the system operation to continue in case any one line fails. This conservative security margin is commonly applied in the industry [37, 86].

Each substation has an associated Voronoi cell that describes the region that is closer to the substation than to any other substation except for country borders, which are kept to retain the integrity of country totals. Exemplary Voronoi cells are illustrated in [Figure 2.2](#). We use these as geographical catchment area for electricity loads, renewable resource potentials, and power plants, assuming that supply and demand always connect to the closest substation of the transmission network and disregarding any constraints on the low-voltage distribution level.

Besides the existing transmission network, [PyPSA-Eur](#) also includes the fleet of operational conventional power plants. This data is obtained from the *power-plantmatching* tool, which merges datasets from a variety of sources [87]. It provides data on the power plants about their location, technology and fuel type, age, and capacity, including hard coal, lignite, oil, open cycle gas turbine ([OCGT](#)), combined cycle gas turbine ([CCGT](#)), nuclear, and biomass generators. Furthermore, existing run-of-river, hydro-electric dams, pumped-hydro storage plants are also part of the dataset. In general, we suppose these to be non-extendable due to assumed geographical constraints. Because the continent-wide availability of data on the locations of wind and solar installations is fragmentary, we

Figure 2.3: Assumed eligible land for on-shore wind turbine development near Karlsruhe. Red zones are excluded.



disregard already existing wind and solar capacities and build them from zero at each node individually. The same applies to storage technologies for which the reasoning is more intuitive. They have not yet been deployed to a significant extent. We consider two extendable storage technologies: battery storage with a energy-to-power ratio of six hours representing short-term storage suited to balancing daily fluctuations and hydrogen storage which exemplifies long-term synoptic and seasonal. The latter constitutes a combination of electrolyzers to convert electricity to hydrogen, steel tanks or caverns for storage, and fuel cells for the reconversion to electricity subject to conversion losses.

2.3.2 Land Eligibility and Renewable Potentials

Eligible areas for developing renewable infrastructure are calculated per technology and the grid nodes' Voronoi cells using the *glaes* tool [88]. How much wind and solar capacity may be built in a particular region is constrained by eligible codes of the [CORINE](#) land use database and is further restricted by distance criteria and the natural protection areas specified in the [Natura 2000](#) dataset. For instance, wind turbines may not be built in forests or urban areas and must keep a minimum distance of 1000 metres to housing estates. Moreover, offshore wind farms may not be developed at sea depths exceeding 50 metres, as indicated by the [GEBCO](#) bathymetry dataset. This currently disregards the possibility of floating wind turbines. [Figure 2.3](#) showcases the exclusion procedure's level of detail by the example of available areas for onshore wind turbine development near Karlsruhe. To finally express the potential in terms of installable capacities, the available land is multiplied with an allowed deployment density, which we consider to be a fraction of the technology's technical deployment density to preempt public acceptance issues and conceivable regulation.

2.3.3 Time Series of Supply and Demand

The location-dependent renewables availability time series are generated based on two historical weather datasets. We retrieve wind speeds at 100 metres, surface roughness, and surface run-off from rainfall or melting snow from the ERA5 reanalysis dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) [53]. It provides hourly values for each of these parameters since 1950 on a $0.25^\circ \times 0.25^\circ$ grid. In Southern Germany, a weather cell expands approximately 20 km from east to west and 31 km from north to south. For the direct and diffuse solar irradiance, we use the satellite-aided SARAH-2 dataset, which assesses cloud cover in more detail than the ERA5 dataset. It features values from 1983 to 2015 at an even higher resolution with a $0.05^\circ \times 0.05^\circ$ grid and 30-minute intervals [89]. In general, the single reference weather year can be freely chosen for the optimisation, but throughout this thesis all analyses are based on the year 2013, which is regarded as a characteristic year for both wind and solar resources (e.g. [90]).

Models for wind turbines, solar panels, and the inflow into hydro basins convert the weather data to hourly capacity factors. Using power curves of typical wind turbines wind speeds scaled to the according hub height are mapped to power outputs. The solar photovoltaic panels' output is calculated based on the incidence angle of solar irradiation, the panel's tilt angle, and conversion efficiency. The tool *atlite* [9] provides functionality to perform such calculations efficiently. Finally, we aggregate the obtained time series to each grid node's Voronoi cell heuristically in proportion to the capacity factor.

In combination with the capacity potentials derived from the assumed land use restrictions, we can use the time-averaged capacity factors to determine the energy that could be produced from wind or solar energy in different regions of Europe. This is shown in Figures 2.4a to 2.4b, which exhibit locations that unite both highly productive sites and extensive land availability. For wind, the North Sea appears as the most attractive region, whereas for solar, the most promising locations are in the South of Spain and Italy. Figures 2.4c to 2.4d moreover present exemplary time series of wind and solar capacity factors, demonstrating the daily cycles of solar generation, the weekly patterns of wind, and also the smoothing effect of averaging over larger areas that is possible to realise with adequate transmission grids.

Like the renewable generation time series, data on electricity consumption in Europe is obtained from historical data. Because the load time series from the ENTSO-E statistics are only available on a country-level, the spatial detail is incomparably lower. Heuristically, they are distributed to each grid node to 40% by population density and to 60% by gross domestic product based on a regression performed for [13]. Figures 2.4e to 2.4f show a selection of the resulting spatio-temporal patterns, clearly outlining densely populated areas.

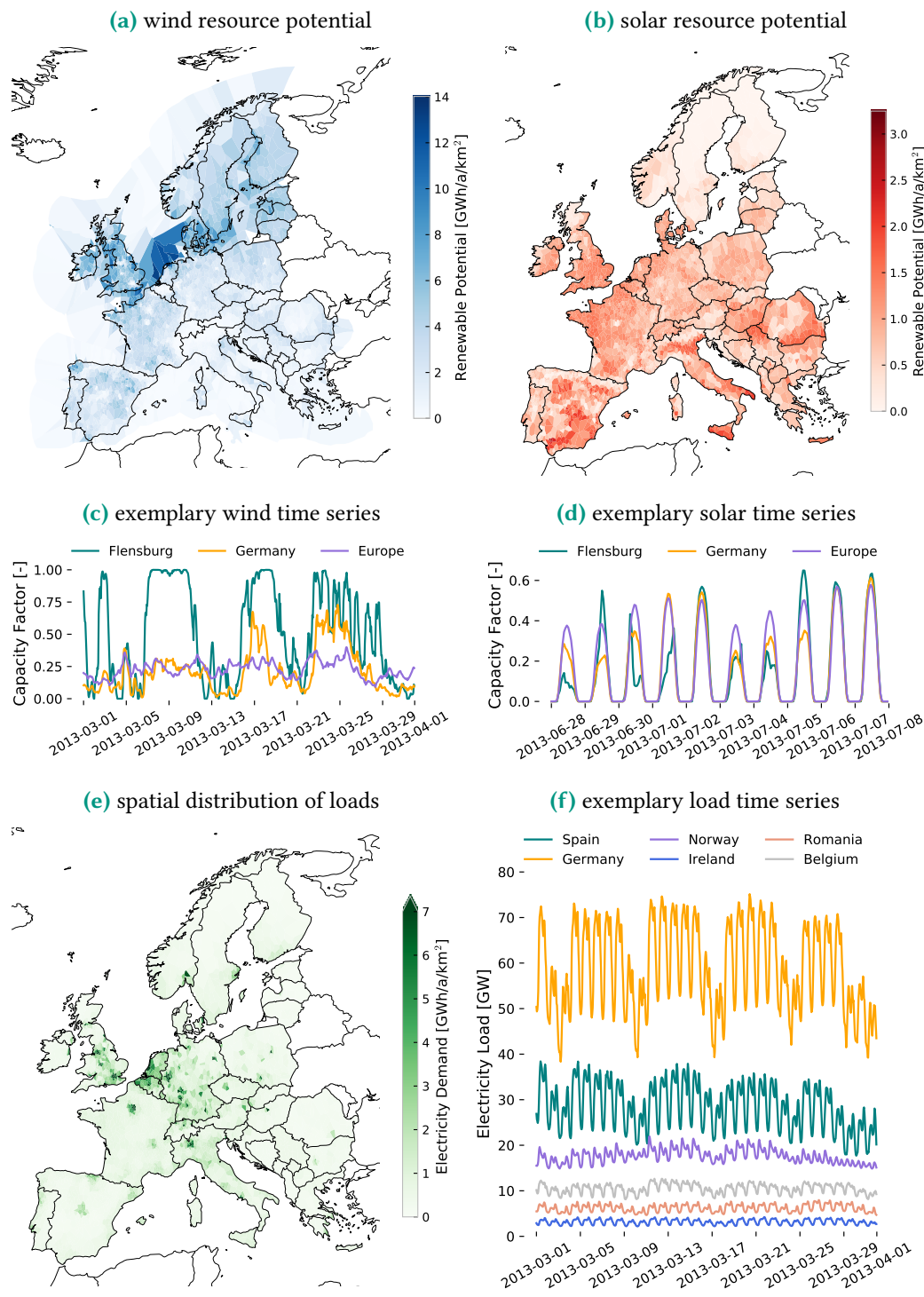


Figure 2.4: Spatial distribution of wind and solar resource potential (including land-use eligibility, deployment density and annual yield), patterns of electricity consumption in Europe, and exemplary hourly time series

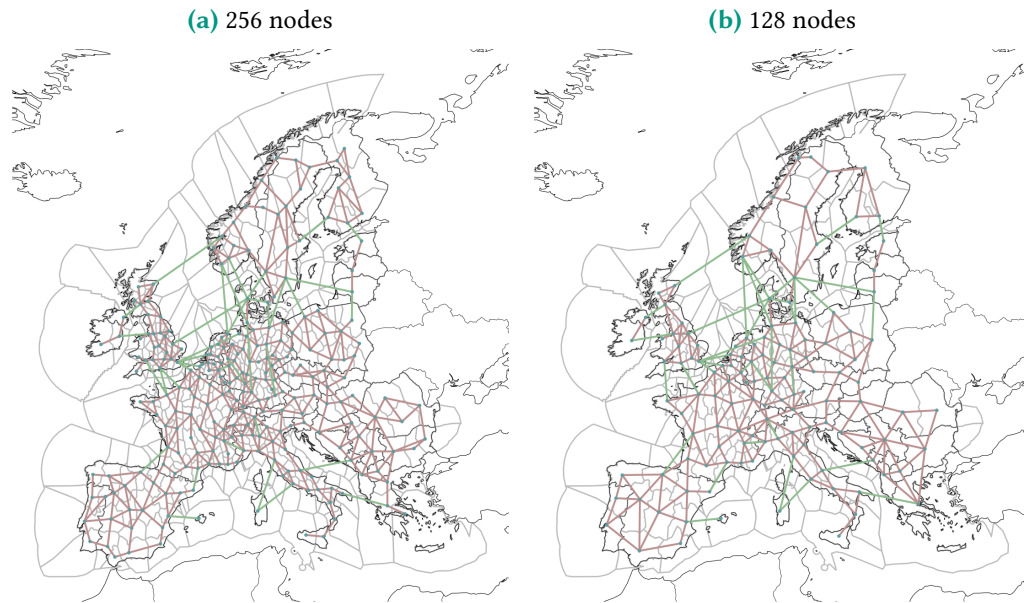


Figure 2.5: European transmission network clustered to varying spatial resolutions. Red lines indicate HVAC connections. Green lines indicate HVDC connections.

2.3.4 Network Simplification and Other Limitations

Modelling the European transmission system at full resolution is too complex to be solved in a reasonable time. Therefore, we must simplify the network topology by lowering the spatial and temporal resolution.

In terms of spatial detail, we initially remove the network’s radial paths, i.e. nodes with only one edge, by connecting remote resources to adjacent nodes and transform the network to a uniform voltage level. We also aggregate generators of the same kind that connect to the same substation. Based on these initial simplifications, the network resolution can be further reduced using a *k-means* network clustering algorithm, which by default uses regional electricity consumption as weights [28, 29]. The extent of the spatial aggregation can be flexibly chosen. It can be anything between the original resolution and one node per country (and synchronous zone if a country, like Denmark, is part of two synchronous areas). Figure 2.5 depicts two such levels of aggregation, 128 nodes and 256 nodes, that are within reach for solvers and exemplary of the network resolutions applied throughout this thesis.

To be able to evaluate energy storage needs properly, it is important to retain the sequence of time steps when adjusting the temporal resolution for computational gain. *PyPSA-Eur* allows two methods of temporal aggregation: simple downsampling from hourly to e.g. three-hourly snapshots and segmentation clustering

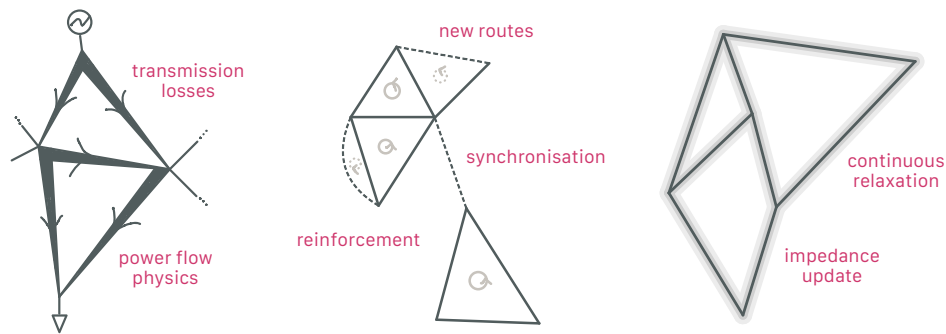
with the *tsam* package [91], which clusters only adjacent snapshots based on the similarity of the supply and demand time series.

Throughout the thesis, *PyPSA-Eur* will be a constant companion in various settings and scenarios, dynamically adapting the spatial and temporal resolution to reflect what is feasible to solve in a reasonable time for the respective problem types.

PART I

Model Detail versus Computational Speed

Overview Part I



To weigh the use of local renewable resources with storage against the costs of grid extension to sites with better resources, the siting of new wind and solar power plants requires close coordination with grid planning [24]. While co-optimisation of generation, storage and transmission capacity expansion is not always possible within the regulatory structures of some markets, the impetus may grow as public acceptance limits how much grid expansion is possible. However, co-optimisation demands high spatial detail to factor in grid bottlenecks and renewable potentials and temporal detail to capture the variability of renewables and loads [26, 28, 29]. Together, this leads to a high problem complexity.

Therefore, as a necessary reduction in accuracy, many capacity expansion models (CEMs) either employ a simple transport model for power transmission, which ignores grid physics, or assume a linearised power flow formulation but do not account correctly for how grid impedances change as lines are expanded nor include transmission losses. Chapter 3 distils the differences of established formulations in addition to contributing a convex loss approximation.

Furthermore, considerations of classical transmission expansion planning (TEP) complicate the problem because they typically entail modelling the available line types and the change of power flows in response to line expansion more realistically [30, 92]. Consequently, even if the power flow equations are linearised, the optimisation problem is still bilinear and mixed-integer due to the dependence of line expansion on line impedance and a discrete set of line expansion options [24, 93]. While Chapter 4 develops a more efficient equivalent reformulation of the problem using a cycle decomposition of the transmission network topology, Chapter 5 evaluates approximating LP-based heuristics that use sequential linear programming (SLP) and retrospective discretisation methods.

Recomposed from abstracts and introductions of papers forming this part's chapters [1, 2, 5].

Approximating Power Flow and Transmission Losses

Contents of this chapter based on

Neumann, F., Hagenmeyer, V. & Brown, T. Approximating Power Flow and Transmission Losses in Coordinated Capacity Expansion Problems. arXiv: 2008.11510 (2020). 

3.1 Introduction

To find credible capacity expansion plans as higher shares of renewables increase the frequency of transmission bottlenecks, more detailed grid modelling is needed that looks beyond import and export capacities but accounts for physical conditions such as loop flows, transmission losses, and curtailment due to otherwise overloaded lines [30, 94]. Especially for planning problems with both static investment and time-dependent dispatch variables spanning across thousands of operational conditions, a tractable yet sufficiently trustworthy representation of power flows is essential. Ideally, outputs are detailed enough to be used as inputs for more accurate analyses, bridging the granularity gap between coarsely-resolved planning models and more detailed engineering models. At the same time, the problem should still be solveable within reasonable time. Unfortunately, the first-choice AC power flow equations are nonlinear and non-convex, which makes the embedded AC optimal power flow problem NP-hard [58, 95, 96].

The transport model, that takes account only of power transfer capacities while ignoring impedances, and the linearised power flow model, which includes impedances to consider both Kirchhoff laws but no losses, are commonly used in capacity expansion models (CEMs). Among the models reviewed by Ringkjøb *et al.* [14] around four in five models use a transport model if flows are represented, whereas only one in five uses a linearised power flow model. Previous work has compared these two major variants [97–100], and some performed simulations of

AC power flow after optimisation [86, 100]. The comparisons indicate little discrepancy regarding total system cost and cross-border transmission, but also differences in nodal prices and overlooked line overloadings when checked against AC power flow calculations. However, the significance of existing comparisons is limited by the use of low spatial resolution models with fewer than 25 nodes. Furthermore, the consideration of losses is underrepresented in design studies, but alongside characteristic weather patterns shapes the trade-offs regarding the volume of transmitted energy because losses increase as more power is transported [101, 102].

In the present chapter, we offer a comprehensive comparison of linear representations of power flow and losses in theory and practice. We outline their characteristic benefits and shortcomings in the context of coordinated capacity expansion problems, where generation, transmission and storage infrastructure is jointly planned. Given the multitude of modelling uncertainties, we assess under which circumstances it is worth embedding more elaborate flow models than a simple transport model. We further extend beyond previous research by introducing a computationally inexpensive loss approximation that incorporates an efficient reformulation of the linearised power flow equations based on a cycle decomposition of the network graph. By using PyPSA-Eur [13], the open model of the European power system introduced in Section 2.3, which spans the whole continent with hundreds of nodes and hourly temporal coupling due to the consideration of storage units, we achieve advanced and reproducible comparisons in systems with high shares of renewables. While we take an investment planning perspective, the way that the transmission of power is represented is relevant beyond system planning. For instance, it plays a role in the design of future electricity markets with multiple bidding zones and flow-based market coupling [93, 99, 103].

We structured the present chapter as follows. We briefly review the physics of power flow in Section 3.2, before we continue with the different linear power flow representations and how they are embedded in the investment planning problem (see Section 2.1) in Section 3.3. In Section 3.4 we present the experimental setup, the results of which are discussed in Section 3.5 and critically appraised in Section 3.6. The chapter is concluded in Section 3.7 with a summary and recommendations.

3.2 Nonlinear Power Flow and Losses

The representation of power flows is one decisive constituent component of the full long-term power system planning problem we outlined in Section 2.1, but must be linearised to solve in reasonable time. Before we introduce different variants of such linearised power flow models, this section briefly revises the nonlinear AC power flow equations, important electrical parameters of transmission

lines, and how to calculate active power losses on a line. We do this to set the foundations for derivations of the covered flow models. For notational simplicity, the subsequent descriptions will omit the time index t .

3.2.1 Nonlinear Alternating Current Power Flow

The active power flow $p_{\ell(i,j)}$ of a line $\ell \equiv \ell(i,j)$ from bus i to bus j can be described in voltage-polar coordinates by

$$p_{\ell(i,j)} = g_{\ell} |V_i|^2 + |V_i| |V_j| [g_{\ell} \cos(\theta_i - \theta_j) - b_{\ell} \sin(\theta_i - \theta_j)] \quad (3.1)$$

and, analogously, the reactive power flow $q_{\ell(i,j)}$ is given by

$$q_{\ell(i,j)} = b_{\ell} |V_i|^2 + |V_i| |V_j| [g_{\ell} \sin(\theta_i - \theta_j) - b_{\ell} \cos(\theta_i - \theta_j)], \quad (3.2)$$

where $|V_i|$ is the per-unit bus voltage magnitude, θ_i is the bus voltage angle, g_{ℓ} is the line conductance, and b_{ℓ} is the line susceptance [104].

3.2.2 Nonlinear Active Power Losses

To derive an expression for the active power losses in a transmission line, we apply the convention that departing power flows are positive and arriving power flows are negative. Consequently, if power flows from bus i to j , $p_{\ell(i,j)} > 0$ and $p_{\ell(j,i)} < 0$. The losses ψ_{ℓ} are the difference between power sent and power received [104], therefore

$$\psi_{\ell} = p_{\ell(i,j)} + p_{\ell(j,i)}. \quad (3.3)$$

Substituting Equation (3.1) into Equation (3.3) yields

$$\psi_{\ell} = g_{\ell} |V_i|^2 + |V_i| |V_j| [g_{\ell} \cos(\theta_i - \theta_j) - b_{\ell} \sin(\theta_i - \theta_j)] \quad (3.4)$$

$$+ g_{\ell} |V_j|^2 + |V_j| |V_i| [g_{\ell} \cos(\theta_j - \theta_i) - b_{\ell} \sin(\theta_j - \theta_i)] \quad (3.5)$$

and using the trigonometric identities $\cos(-\alpha) = \cos(\alpha)$ and $\sin(-\alpha) = -\sin(\alpha)$ translates to

$$\psi_{\ell} = g_{\ell} |V_i|^2 + |V_i| |V_j| [g_{\ell} \cos(\theta_i - \theta_j) - b_{\ell} \sin(\theta_i - \theta_j)] \quad (3.6)$$

$$+ g_{\ell} |V_j|^2 + |V_j| |V_i| [g_{\ell} \cos(\theta_i - \theta_j) + b_{\ell} \sin(\theta_i - \theta_j)]. \quad (3.7)$$

We can further simplify this expression to the loss formula

$$\psi_{\ell} = g_{\ell} (|V_i|^2 + |V_j|^2) - 2 |V_i| |V_j| g_{\ell} \cos(\theta_i - \theta_j), \quad (3.8)$$

which we will use in [Section 3.3.4](#) to derive a linear approximation for losses.

3.2.3 Electrical Parameters of Transmission Lines

The electrical parameters of transmission lines play a significant role in determining the power flows and are therefore briefly reviewed in the following. The complex per-unit impedance $z_\ell = r_\ell + ix_\ell$ is composed of resistance r_ℓ and reactance x_ℓ . Likewise, the admittance $y_\ell = g_\ell + ib_\ell$ is composed of conductance g_ℓ and susceptance b_ℓ . Impedance and admittance are reciprocals ($y_\ell = z_\ell^{-1}$). Hence, if we assume a dominance of reactance over ohmic resistance ($r_\ell \ll x_\ell$), as applies for high voltage overhead transmission lines, we obtain the approximations

$$g_\ell \approx \frac{r_\ell}{x_\ell^2}, \quad (3.9)$$

$$b_\ell \approx \frac{1}{x_\ell}. \quad (3.10)$$

Following e.g. [105], the complex per-unit impedance $z_\ell = r_\ell + ix_\ell$ is composed of ohmic resistance r_ℓ and reactance x_ℓ . Likewise, the admittance $y_\ell = g_\ell + ib_\ell$ is composed of conductance g_ℓ and susceptance b_ℓ . Impedance and admittance are reciprocals ($y_\ell = z_\ell^{-1}$), hence we obtain the relations

$$g_\ell + ib_\ell = \frac{1}{r_\ell + ix_\ell}, \quad (3.11)$$

$$g_\ell + ib_\ell = \frac{r_\ell - ix_\ell}{(r_\ell + ix_\ell)(r_\ell - ix_\ell)}, \quad (3.12)$$

$$g_\ell + ib_\ell = \frac{r_\ell - ix_\ell}{r_\ell^2 + x_\ell^2}. \quad (3.13)$$

By splitting real and imaginary parts we can express conductance and susceptance in terms of impedance and reactance:

$$g_\ell = \operatorname{Re} \left[\frac{r_\ell - ix_\ell}{r_\ell^2 + x_\ell^2} \right] = \frac{r_\ell}{r_\ell^2 + x_\ell^2}, \quad (3.14)$$

$$b_\ell = \operatorname{Im} \left[\frac{r_\ell - ix_\ell}{r_\ell^2 + x_\ell^2} \right] = \frac{-x_\ell}{r_\ell^2 + x_\ell^2}. \quad (3.15)$$

We will use these relations in [Section 3.3.3](#) and [Section 3.3.4](#). In view of the approximation of line losses in later sections, note that although we suppose that resistance is dominated by reactance, we do not assume resistance to be zero (see [106, 107]).

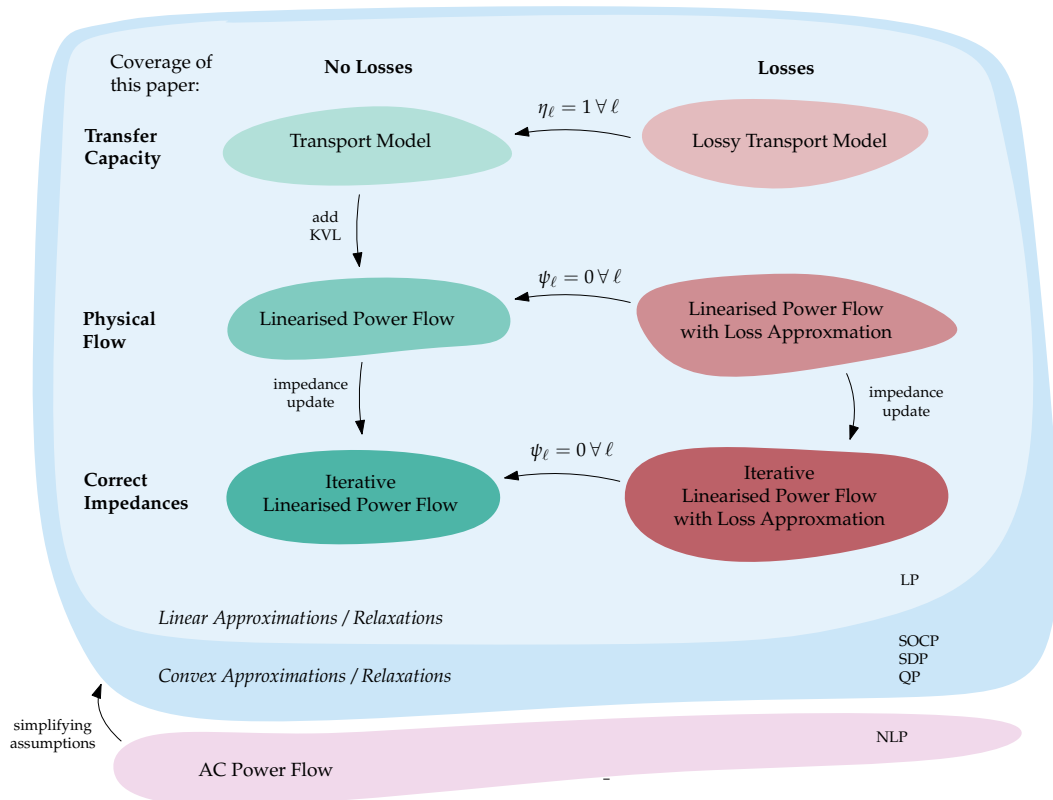


Figure 3.1: Illustration of the scope of the present chapter and its context. It shows the connections between the covered linear power flow models, their main features, and how they are related to other (convexified) nonlinear formulations.

3.3 Linearised Power Flow Models

The AC power flow Equations (3.1) to (3.2) are nonlinear and non-convex. This is challenging because multiple local minima exist due to the trigonometric expressions and when directly incorporated in the optimisation problem they would make the problem NP-hard [58, 95, 96]. To be able to run large optimisation problems of the continental power system at sufficient spatial and temporal resolution it is hence inevitable to retain a convex problem that can be solved in polynomial time and does not possess local minima.

In this section we describe and develop various linear representations of power flow. These are introduced in the order from least to most accurate, progressively increasing the complexity; namely (i) the common transport model, (ii) a lossy transport model, (iii) the lossless linearised power flow, and (iv) a lossy linearised power flow model. Figure 3.1 shows the relations between the formulations and Table 3.1 documents differences in the number of variables and constraints. The scope of the present chapter is deliberately constrained to:

Table 3.1: Comparison of the number of variables and equality/inequality constraints related to flow models per snapshot $t \in \mathcal{T}$. The constraint count excludes variable bounds. $|\mathcal{L}|$ is the number of lines, $|\mathcal{N}|$ is the number of nodes, and $2n$ represents the number of tangents used for the loss approximation.

Model		Variables	Eq. Constraints	Ineq. Constraints
Transport	Section 3.3.1	$ \mathcal{L} $	$ \mathcal{N} - 1$	0
Lossy Transport	Section 3.3.2	$2 \mathcal{L} $	$ \mathcal{N} - 1$	0
Linearised Power Flow	Section 3.3.3	$ \mathcal{L} $	$ \mathcal{L} $	0
Linearised Power Flow with Loss Approximation	Section 3.3.4	$2 \mathcal{L} $	$ \mathcal{L} $	$2n \mathcal{L} $

- *only linear problems:* To avail of powerful, scalable and fast interior-point solvers, and to guarantee an optimal solution, we only include formulations that entail linear problems. However, there exist promising second-order cone or semidefinite convex relaxations of the power flow equations. For excellent theoretical reviews of convex relaxations and approximations of power flow see Molzahn & Hiskens [108], Taylor [104], and Coffrin & Roald [109].
- *only active power:* We furthermore confine our analysis to formulations that do not capture reactive power flows or information on bus voltages. Nonetheless, linear problems that capture selected aspects of this are under active research; see e.g. Coffrin & Van Hentenryck [110].
- *only comparison of different feasible spaces:* We compare different linear flow models that define different feasible spaces. We do not compare equivalent reformulations of identical models, since this has been analysed in Hörsch *et al.* [69].
- *no copper plate model:* Although occasionally encountered in generation and storage CEMs, we do not include the copper plate relaxation in our comparisons because it does not capture information on power flows in transmission networks. The copper plate model removes all lines and aggregates all components to a single node. It is a relaxation because any transmission of power becomes unconstrained and incurs no cost. For the impact of spatial clustering – of which the copper plate model is the extreme – on optimal investments we refer to Hörsch & Brown [28] and Frysztacki *et al.* [29].

3.3.1 Transport Model

The transport model is also known as a network flow model, trade model, transportation model or net transfer capacity (NTC) model [104]. It ignores the effect of

impedances on flows (including losses) and, besides the capacity constraints of lines, only requires nodal power balance according to Kirchhoff's current law (KCL); i.e. the power injected at each bus must equal the power withdrawn by attached lines

$$p_i = \sum_{\ell} K_{i\ell} p_{\ell} \quad \forall i \in \mathcal{N}, \quad (3.16)$$

where p_i is the active power injected or consumed at node i and K is the incidence matrix of the network graph which has non-zero values $+1$ if line ℓ starts at bus i and -1 if line ℓ ends at bus i .

Because the columns of the incidence matrix each sum up to zero, KCL yields $|\mathcal{N}| - 1$ linearly independent constraints. These are not sufficient to uniquely determine the $|\mathcal{L}|$ unknown flows. The transport model allows arbitrary flows as long as flow conservation is fulfilled, also because it is free and lossless to transmit power. This makes the transport model degenerate, which can be detrimental to the solving speed. Also, of course, this does not adequately reflect the physical behaviour of power flows in the transmission network.

Despite its drawbacks, the transport model is very popular due to its simplicity. In the comprehensive review by Ringkjøb *et al.* [14], it is applied in a majority of models. This minimalistic representation of flows is useful to develop an understanding for the potential of increased transfer capacity between regions, rather than assessing specific transmission bottlenecks and reinforcement needs. It is often applied in investment models where the grid is highly aggregated to a few nodes (e.g. one node per country in Europe or federal state in the United States) or analyses of energy markets across multiple bidding zones. Its main advantages are ease of implementation and fast solving speed. For pure dispatch problems without investment decisions one can even utilise specialised network flow algorithms; for instance the minimum cost flow algorithm [111].

3.3.2 Transport Model with Loss Approximation

Part of the drawbacks and degeneracy of the transport model stems from the disregard of transmission losses. As partial remedy, we can amend the transport model with a simple loss approximation which assumes lines to have a constant transmission efficiency η_{ℓ} depending on their length. In this case, the power arriving at the receiving bus is lower than the power injected at the sending bus. To differentiate

between sending bus and receiving bus, we need to split the bidirectional power flow variable p_ℓ into forward flows p_ℓ^+ and backward flows p_ℓ^- with capacity limits

$$0 \leq p_\ell^+ \leq \bar{p}_\ell P_\ell \quad \forall \ell \in \mathcal{L} \quad (3.17)$$

$$0 \leq p_\ell^- \leq \bar{p}_\ell P_\ell \quad \forall \ell \in \mathcal{L} \quad (3.18)$$

which substitute the variables p_ℓ and their bounds given in Equation (2.12). Furthermore, we need to adjust the nodal balance constraints Equation (3.16) to reflect the transmission losses and separated power flow variables to

$$p_i = \sum_{\ell} K_{i\ell}^+ p_\ell^+ - \sum_{\ell} K_{i\ell}^- p_\ell^- \quad \forall i \in \mathcal{N}, \quad (3.19)$$

where K^+ is the lossy incidence matrix of the network graph regarding forward flows p_ℓ^+ which has non-zero values $+1$ if line ℓ starts at bus i and $-\eta_\ell$ if line ℓ ends at bus i . Analogously, K^- regards backward flows p_ℓ^- with non-zero values η_ℓ if line ℓ starts at bus i and -1 if line ℓ ends at bus i .

The transmission losses alleviate some degeneracy of the transport model since considering losses yields an incentive to minimise power flows rather than to distribute them arbitrarily. However, this is paid for with a doubling in the number of flow variables. Additionally, while the use of a constant transmission efficiency is an improvement from the plain transport model, it still ignores the quadratic relationship between power flow and losses [101]. Note, that if all lines have no losses ($\eta_\ell = 1$), the lossy transport model is equivalent to the regular transport model.

3.3.3 Linearised Power Flow

The linearised power flow model, which is also known as linearised load flow, DC power flow or B Θ model, extends the lossless transport model. In addition to the nodal power balance constraints Equation (3.16) from KCL and capacity limits Equation (2.12), linear constraints for Kirchhoff's voltage law (KVL) are included, which define how power flows split in parallel paths. We derive these by simplifying the nonlinear power flow Equations (3.1) to (3.2). Assuming

- all per-unit voltage magnitudes are close to one ($|V_i| \approx 1$),
- conductances g_ℓ are negligible relative to susceptances b_ℓ ($b_\ell \gg g_\ell$),
- voltage angle differences are small enough ($\sin(\theta_i - \theta_j) \approx \theta_i - \theta_j$ and $\cos(\theta_i - \theta_j) \approx 0$),
- reactive power flows q_ℓ are negligible compared to active power flows p_ℓ ($q_\ell \approx 0$),

leads to the constraint

$$p_\ell = b_\ell(\theta_i - \theta_j), \quad (3.20)$$

and when we further assume $r_\ell \ll x_\ell$, by substituting Equation (3.10) we obtain

$$p_\ell = \frac{\theta_i - \theta_j}{x_\ell}. \quad (3.21)$$

This angle-based formulation is the most common linear formulation of KVL [104]. But it is possible to avoid the auxiliary voltage angle variables and reduce the required number of constraints by using a cycle basis of the network graph [69]. Namely, KVL states that the sum of voltage angle differences across lines around all cycles in the network must sum up to zero. Considering a set of independent cycles c of the network forming a cycle basis, which are expressed as a directed linear combination of the lines ℓ in a cycle incidence matrix

$$C_{\ell c} = \begin{cases} 1 & \text{if edge } \ell \text{ is element of cycle } c, \\ -1 & \text{if reversed edge } \ell \text{ is element of cycle } c, \\ 0 & \text{otherwise,} \end{cases} \quad (3.22)$$

KVL is formulated by

$$\sum_{\ell} C_{\ell c}(\theta_i - \theta_j) = 0 \quad \forall c \in \mathcal{C}. \quad (3.23)$$

Using Equation (3.21), we can express KVL directly in terms of the power flows and circumvent the auxiliary voltage angle variables

$$\sum_{\ell} C_{\ell c} p_\ell x_\ell = 0 \quad \forall c \in \mathcal{C}. \quad (3.24)$$

Although less common, this cycle-based formulation Equation (3.24) has been shown to significantly outperform the angle-based formulation Equation (3.21) in operational problems generation capacity expansion planning [69]. In a subsequent chapter of this thesis (Chapter 4), we will further evaluate whether the computational advantage of the cycle-based formulation persists in the context of discrete transmission expansion planning (TEP). There are even further equivalent reformulations of the linearised power flow [69]; for example the power transfer distribution factor (PTDF) formulation, which directly relates nodal power injections to line flows. Because our focus lies on the comparison of different flow models, not their diverse reformulations, we only evaluate the computationally performant cycle-based formulation in the present chapter.

With the cycle-based formulation one can clearly see that the transport model is equivalent to the linearised power flow in radial networks; i.e. when the network has no cycles. Also, the absence of auxiliary voltage angle variables facilitates the insight that the transport model is a relaxation of the linearised power flow because the latter only adds constraints in the same variable space.

The linearised power flow model is claimed to be accurate when reactance dominates ($x_\ell \gg r_\ell$) and when parallel lines have similar ratios [112], but very long lines in highly aggregated networks can deteriorate the quality of the approximation (see Section 3.5.3). An advantage of this model over the transport model is that it captures some meaningful physical characteristics observed in the operation of electrical grids. Namely, it is capable of revealing loop flows in meshed networks; for instance recurring spillover effects between Germany and the Czech Republic. Nevertheless, it still disregards losses.

If we would consider that lines can be built between buses where there are currently none, another variant is the so-called hybrid model. This version formulates linearised power flow constraints for existing lines and employs a transport model for candidate lines.

3.3.4 Linearised Power Flow with Loss Approximation

Neglecting resistive losses is considered to be among the largest sources of error in the linearised power flow formulation, particularly in large networks [112]. The following extension of the lossless linearised power flow (Section 3.3.3) is a mixture of similar variants encountered in the literature with a focus on computational efficiency. We reference where we follow or deviate from previous work below. This or similar formulations have rarely been applied in the co-optimisation of transmission, storage and generation capacities, but rather in detailed operational optimal power flow (OPF) or TEP problems; see overview in [94].

We start by adding a loss variable $\psi_\ell \in \mathbb{R}^+$ for each line. Losses reduce the effective transmission capacity of a line

$$|p_\ell| \leq \bar{p}_\ell P_\ell - \psi_\ell \quad (3.25)$$

and must be accounted for in the nodal balance Equation (3.16)

$$p_i = \sum_\ell K_{i\ell} p_\ell + \frac{|K_{i\ell}|}{2} \psi_\ell \quad \forall i \in \mathcal{N}. \quad (3.26)$$

We split the losses ψ_ℓ equally between both buses (like in [113–115]) and do not allocate them at the sending bus exclusively (like in [107, 116]). The latter could

be modelled with an absolute value function in the linear problem. However, this would involve splitting flow and loss variables each into positive and negative segments. Because this adds many auxiliary decision variables, we decided in favor of distributing the losses evenly. This choice is paid for with the possibility of overestimating losses due to an extensive convex relaxation.

Assuming close to nominal per-unit voltage magnitudes $|V_i| \approx 1$ the loss formula given in equation [Equation \(3.8\)](#) becomes

$$\psi_\ell = 2g_\ell [1 - \cos(\theta_i - \theta_j)]. \quad (3.27)$$

This is the basis for the linearised loss formulation in [\[115\]](#). We can also express this in terms of active power flows p_ℓ by substituting [Equation \(3.21\)](#) into [Equation \(3.27\)](#)

$$\psi_\ell = 2g_\ell [1 - \cos(p_\ell x_\ell)]. \quad (3.28)$$

This makes the loss formulation independent from the voltage angle variables and we can therefore avail of the speed-up obtained by using the cycle-based formulation [Equation \(3.24\)](#).

Using the small-angle approximation $\cos(\alpha) \approx 1 - \alpha^2/2$, [Equation \(3.28\)](#) becomes quadratic

$$\psi_\ell = 2g_\ell \left[1 - \left(1 - \frac{(p_\ell x_\ell)^2}{2} \right) \right] = g_\ell (p_\ell x_\ell)^2. \quad (3.29)$$

By inserting [Equation \(3.9\)](#) we get

$$\psi_\ell = \frac{r_\ell}{x_\ell^2} (p_\ell x_\ell)^2 \quad (3.30)$$

or simply

$$\psi_\ell = r_\ell p_\ell^2. \quad (3.31)$$

This is the basis for the linearised loss formulation in [\[107\]](#). Because [Equation \(3.31\)](#) is a quadratic equality constraint, it is still non-convex. Other works have discussed or applied a piecewise linearisation of [Equation \(3.31\)](#) [\[108, 114, 115, 117, 118\]](#). But because the use of integer variables to define the segments would entail a non-convex mixed-integer linear problem (MILP), we choose not to pursue this approach. Instead, by building a convex envelope around this constraint from the upper and lower bounds for ψ_ℓ as well as a number of tangents as inequality constraints, we can incorporate transmission losses while retaining a linear optimisation problem. This is illustrated in [Figure 3.2](#). For setting the

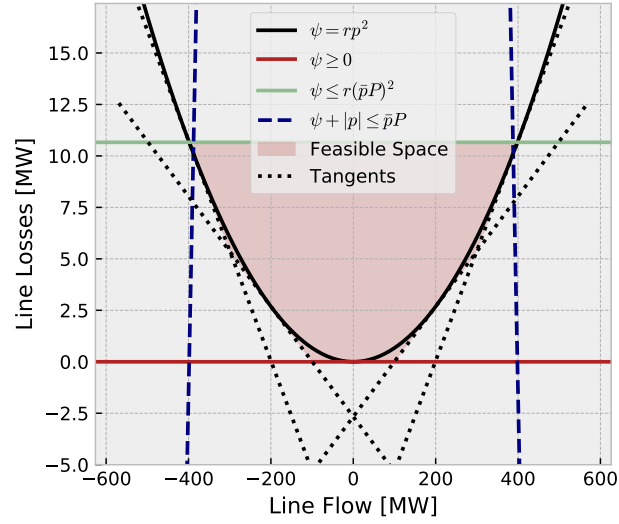


Figure 3.2: Illustration of the feasible space in the flow–loss ($p_\ell - \psi_\ell$) dimensions.

lower limit, by definition losses are positive

$$\psi_\ell \geq 0 \quad (3.32)$$

and by substituting maximal line flows

$$p_\ell \leq \bar{p}_\ell P_\ell \leq \bar{p}_\ell \bar{P}_\ell \quad (3.33)$$

into Equation (3.31) we obtain the upper limit

$$\psi_\ell \leq r_\ell (\bar{p}_\ell \bar{P}_\ell)^2. \quad (3.34)$$

Next, we derive $2n$ evenly spaced (like in [108]) mirrored tangents which approximate Equation (3.31) as inequalities from below. These have the form

$$\psi_\ell \geq m_k \cdot p_\ell + a_k \quad \forall k = 1, \dots, n \quad (3.35)$$

$$\psi_\ell \geq -m_k \cdot p_\ell + a_k \quad \forall k = 1, \dots, n \quad (3.36)$$

At segment k we calculate the losses

$$\psi_\ell(k) = r_\ell \left(\frac{k}{n} \cdot \bar{p}_\ell \bar{P}_\ell \right)^2 \quad (3.37)$$

and the corresponding slope

$$m_k = \frac{d\psi_\ell(k)}{dk} = 2r_\ell \left(\frac{k}{n} \cdot \bar{p}_\ell \bar{P}_\ell \right) \quad (3.38)$$

and the offset

$$a_k = \psi_\ell(k) - m_k \left(\frac{k}{n} \cdot \bar{p}_\ell \bar{P}_\ell \right). \quad (3.39)$$

Together, Equation (3.32) and Equations (3.34) to (3.36) form a convex envelope of Equation (3.31).

A recurring criticism of this extensive convex relaxation is the possibility of so-called fictitious or artificial losses [108, 119–122]. As illustrated by Figure 3.2, the model does indeed allow for overestimating losses. This can be economical if negative locational marginal prices occur. Overestimating losses is equivalent to dissipating power at a node. Another component in the problem formulation that already permits this behaviour are storage units (see Equations (2.6) to (2.8)). To avoid binary variables, storage units may charge and discharge at the same time. Power is then lost by cycling through the conversion efficiencies. We argue that fictitious losses are not problematic because (i) negative nodal prices are rare, (ii) such behaviour could be realised in operation by low-cost resistors and demand response, and (iii) the loss overestimation is bounded. We will substantiate this argument with experimental results in Section 3.5.2.

3.3.5 Iterative Linearised Power Flow (with Losses)

When using the linearised power flow, with and without losses (Section 3.3.3 and Section 3.3.4 respectively), the impedances of transmission lines affect the flows and losses. The relations of reactances x_ℓ determine the distribution of flows (see Equation (3.24)). The resistances r_ℓ set the losses (see Equation (3.31)). Thus, for reactances we are only interested in relative values, whereas for the resistances the absolute values are decisive.

Line impedances change as line capacities are increased ($x_\ell \propto 1/P_\ell$ and $r_\ell \propto 1/P_\ell$). Ignoring this dependency would result in distorted power flows. Expanded lines would experience less flow than they should. Losses may also be overestimated as the extension of parallel lines reduces the effective resistance.

Consequently, the representation of grid physics is improved by taking account of the relation between line capacities and impedances, yet also complicates the problem. If we considered discrete expansion options we would use a big- M disjunctive relaxation to resolve the nonlinearity [1]. We discuss this case in detail in the subsequent Chapter 4. But since we assume continuous line expansion in view of computational performance, we instead pursue an iterative heuristic approach, which we show later in Chapter 5 to be an acceptable approximation [2].

We sequentially solve the optimisation problem from [Section 2.1](#) and in each iteration update the line impedances according to their optimised capacities. We repeat this process until (i) line expansion choices do not change in subsequent iterations and convergence is reached, or (ii) a predetermined number of iterations are performed. In the latter case, the final iteration should be run with fixed line capacities such that impedances do not change anymore.

3.4 Simulation Setup

Having developed the individual power flow models in theory, this section outlines the setup we use to test them. First, in [Section 3.4.1](#) we list the settings we apply to the power system model we use as the case study. Second, we outline the methodology we use to validate the resulting approximated line flows in [Section 3.4.2](#).

3.4.1 Model Inputs

We evaluate the different flow models on the open power system dataset [PyPSA-Eur](#) we described in [Section 2.3](#), which covers the whole European transmission system [13]. We choose a spatial resolution of 250 nodes and a temporal resolution of 4380 snapshots, one for every two hours of a full year. This reflects the maximum for which all flow models presented in [Section 3.3](#) could be solved.

Targeting an emission reduction of 100% in the power sector, we only consider renewable resources [19]. Following the problem formulation from [Section 2.1](#), solar PV, onshore and offshore wind capacities are co-optimised with battery storage, hydrogen storage, and transmission infrastructure in a greenfield planning approach, subject to spatio-temporal capacity factors and geographic potentials. Exceptions to greenfield planning are existing transmission infrastructure, which can only be reinforced but not removed, and today's run-of-river and hydropower capacities including pumped hydroelectric energy storage, which are not extendable due to assumed geographical constraints. HVDC links are assumed to have losses of 3% per 1000 km [123] and can be expanded continuously up to 20 GW (each composed of several smaller parallel circuits). We only apply link losses to flow models which also account for losses in HVAC lines. HVAC line capacity can also be expanded continuously; by the maximum of doubled capacity or additional 5 GW. When using the lossy transport model, HVAC lines are assumed to have constant losses in the order of 5% per 1000 km [123].

Technically, the optimisation problem is implemented using the free Python modelling framework Python for Power System Analysis ([PyPSA](#)) (v0.17.0) working

with the Pyomo solver interface [12]. Both optimality and feasibility tolerances are set to a value of 0.1%, which is sufficient given the mentioned approximations made in the model. We use the cycle-based formulation of KVL for any model that accounts for it. The code to reproduce the experiments is openly available at github.com/fneum/power-flow-models.

In accordance with descriptions in Section 3.3, the following flow models are evaluated:

- lossless transport model as *Transport*,
- lossy transport model as *Lossy Transport*,
- lossless linear power flow with no iterations as *Lossless*,
- lossless linear power flow with 3 iterations as *Iterative Lossless*,
- lossy linear power flow with 6 tangents and no iterations as *Lossy*, and
- lossy linear power flow with 6 tangents and 3 iterations as *Iterative Lossy*.

3.4.2 Nonlinear Power Flow Simulation

All presented flow models approximate the AC power flow equations (Section 3.2). Thus, to identify possibly overlooked line overloading, and to demonstrate characteristic features of particular flow models, we use the AC power flow equations to assess the quality of the respective approximations.

We compare optimised flows to simulated flows which we obtain by solving the AC power flow equations ex-post based on the optimised dispatch of controllable system components. Specifically, we do not reoptimise dispatch decisions subject to the AC power flow model due to computational constraints given such large multi-period problems, but only check their feasibility. We use the Newton-Raphson method (see e.g. [105]) and distribute the total slack power across all buses in proportion to their total generation capacity [102, 124]:

Power Flow Calculation Without Distributed Slack Given nodal power imbalances S_n at any given snapshot for each bus n the AC power flow equations are given by

$$S_n = P_n + iQ_n = V_n I_n^* = V_n \left(\sum_m Y_{nm} V_m \right)^*, \quad (3.40)$$

where $V_n = |V_n| e^{i\theta_n}$ is the complex voltage, whose rotating angle is taken relative to the slack bus and Y_{nm} is the bus admittance matrix, based on the branch impedances and shunt admittances (including those attached to buses).

For the slack bus $n = 0$ it is assumed $|V_0|$ is given and that $\theta_0 = 0$; P and Q are to be found. For the PV buses, P and $|V|$ are given; Q and θ are to be found. For the PQ buses, P and Q are given; $|V|$ and θ are to be found.

Considering PV and PQ as sets of buses, then there are $|PV|+2|PQ|$ real-valued equations to solve:

$$\operatorname{Re} \left[V_n \left(\sum_m Y_{nm} V_m \right)^* \right] - P_n = 0 \quad \forall PV \cup PQ \quad (3.41)$$

$$\operatorname{Im} \left[V_n \left(\sum_m Y_{nm} V_m \right)^* \right] - Q_n = 0 \quad \forall PQ \quad (3.42)$$

We need to find θ_n for all PV and PQ buses and $|V_n|$ for all PQ buses.

These equations $f(x) = 0$ are solved using the Newton-Raphson method, with the Jacobian

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial P}{\partial \theta} & \frac{\partial P}{\partial |V|} \\ \frac{\partial Q}{\partial \theta} & \frac{\partial Q}{\partial |V|} \end{pmatrix} \quad (3.43)$$

and the initial guesses $\theta_n = 0$ and $|V_n| = 1$ for unknown quantities [105]. Commonly, the total active slack power, which balances remaining mismatches of power generation and demand resulting from the AC power flow equations, is fully allocated to the slack bus. This can be a crude assumption, particularly for large networks with a high penetration of renewables.

Power Flow Calculation With Distributed Slack A better alternative is to distribute the total active slack power across all generators in proportion to their capacities (or another distribution scheme) [124]. The active power flow equations are altered to

$$\operatorname{Re} \left[V_n \left(\sum_m Y_{nm} V_m \right)^* \right] - P_n - P_{\text{slack}} \gamma_n = 0 \quad \forall PV \cup PQ \cup \text{slack} \quad (3.44)$$

where P_{slack} is the total slack power and γ_n is the share of bus n of the total generation capacity, which is used as distribution key. We add an additional active power balance equation for the slack bus since it is now part of the distribution scheme.

The distributed slack approach extends the Jacobian by an additional row for the derivatives of the slack bus active power balance and by an additional column for the partial derivatives with respect to γ

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial P_0}{\partial \theta} & \frac{\partial P_0}{\partial |V|} & \frac{\partial P_0}{\partial \gamma} \\ \frac{\partial P}{\partial \theta} & \frac{\partial P}{\partial |V|} & \frac{\partial P}{\partial \gamma} \\ \frac{\partial Q}{\partial \theta} & \frac{\partial Q}{\partial |V|} & \frac{\partial Q}{\partial \gamma} \end{pmatrix}. \quad (3.45)$$

If $\gamma_n = 0$ for all buses but the slack bus, this is equivalent to a single slack bus model.

Moreover, we consider PV buses¹ at each node since the reactive power set points are unknown. Hence, we suppose that there is sufficient reactive power control infrastructure to maintain nominal voltages. We argue that in systems with high shares of renewables the PV bus assumption is justified in view of a growing number of distributed power generation units, each capable of contributing to voltage control by reactive power injection or consumption, and power electronic devices such as Flexible Alternating Current Transmission Systems (FACTS). While the linearised power flow approximations neglect the shunt capacitance of lines, these are taken into account in the subsequent AC power flow simulation according to the standard equivalent Π model [125]. Suitable short- to medium-length lines between 25km and 250km make up about 85% of all lines in the model. The remaining 15% of lines, which are longer than 250km, are modelled identically although more rigorous alternatives exist. These include partitioning long lines into multiple shorter sections to model series compensation [125], or using equations specifically for long lines that include fewer simplifying approximations of impedances than the Π model [126].

3.5 Results and Discussion

Having introduced the various model formulations and simulation setup, this section presents and discusses the results from the experiments as described in Section 3.4. As evaluation criteria we consider the total system costs and the optimal system composition (Section 3.5.1), the error of optimised losses (Section 3.5.2), the error of optimised flows compared to simulated flows (Section 3.5.3), as well as peak memory consumption and solving time (Section 3.5.4).

¹For PV buses, the nodal active power injections p_i and voltage magnitudes $|V_i|$ are known (we assume nominal voltage magnitudes $|V_i| = 1$). Bus voltage angles θ_i and reactive power feedin q_i are to be found. Conversely, for PQ buses the nodal active power injections p_i and reactive power injections q_i are known. Bus voltage magnitudes $|V_i|$ and angles θ_i are to be found.

Table 3.2: Various statistical indicators compared across covered flow models.

Indicator	Unit	Lossy			Iterative		
		Transport	Transport	Lossless	Lossless	Lossy	Lossy
System cost	bn€p.a.	220.2	226.2	224.9	225.7	243.8	238.5
	€/MWh	70.2	72.1	71.7	71.9	77.7	76.0
Energy transmitted	EWhkm	1.56	1.26	1.36	1.28	0.90	0.94
Network expansion	TWkm	216	214	206	206	160	170
Transmission losses	% of load	0	2.3	0	0	5.1	3.7
Curtailment	%	2.0	1.9	2.3	2.4	2.2	2.4
Share of $ \theta_i - \theta_j \geq 30^\circ$	%	5.1	3.7	4.6	3.9	1.4	1.5

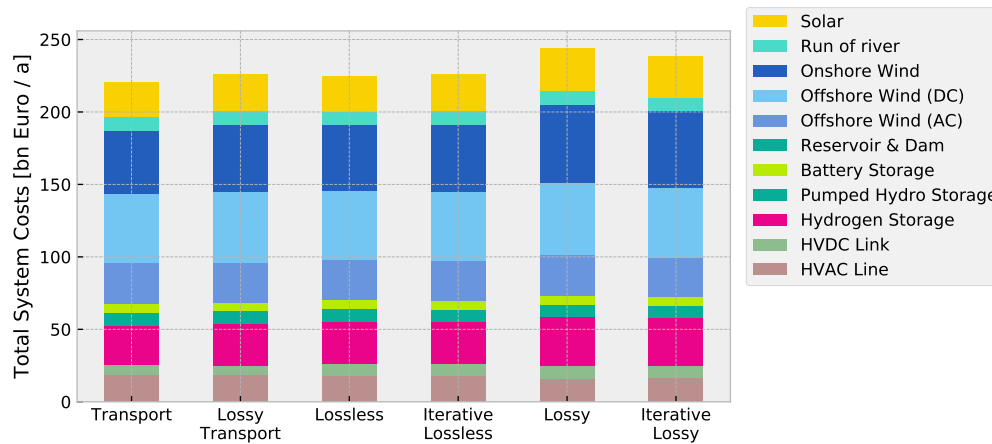


Figure 3.3: Comparison of total annual system costs split by system component for the covered flow models.

3.5.1 Investments, Nodal Prices and Total System Costs

Table 3.2 presents total transmission losses to sum up to around 4% of the total load when updated impedances according to line expansion are used. In comparison to the 1.2% transmission losses reported by the German Federal Network Agency for the year 2019 [127], this value is higher owing to the larger volume of power transmission across the whole continent in scenarios with high shares of renewables. Skipping the update of impedances overestimates losses (5.5%) because additional parallel lines reduce the total impedance. The lossy transport model underestimates losses (2.5%) since it neglects the quadratic relationship between power and losses. Table 3.2 further shows low curtailment at around 2% across all flow models due to generous line expansion allowances.

At first sight, the optimised technology mix appears relatively similar across all flow models, both in terms of cost composition in Figure 3.3 and the map of investments in Figure 3.6. This is further underlined by the high correlations of optimised capacities shown in Figure 3.4. Potentially due to some placement degeneracies, lowest correlations are found for battery and hydrogen storage. Further notable

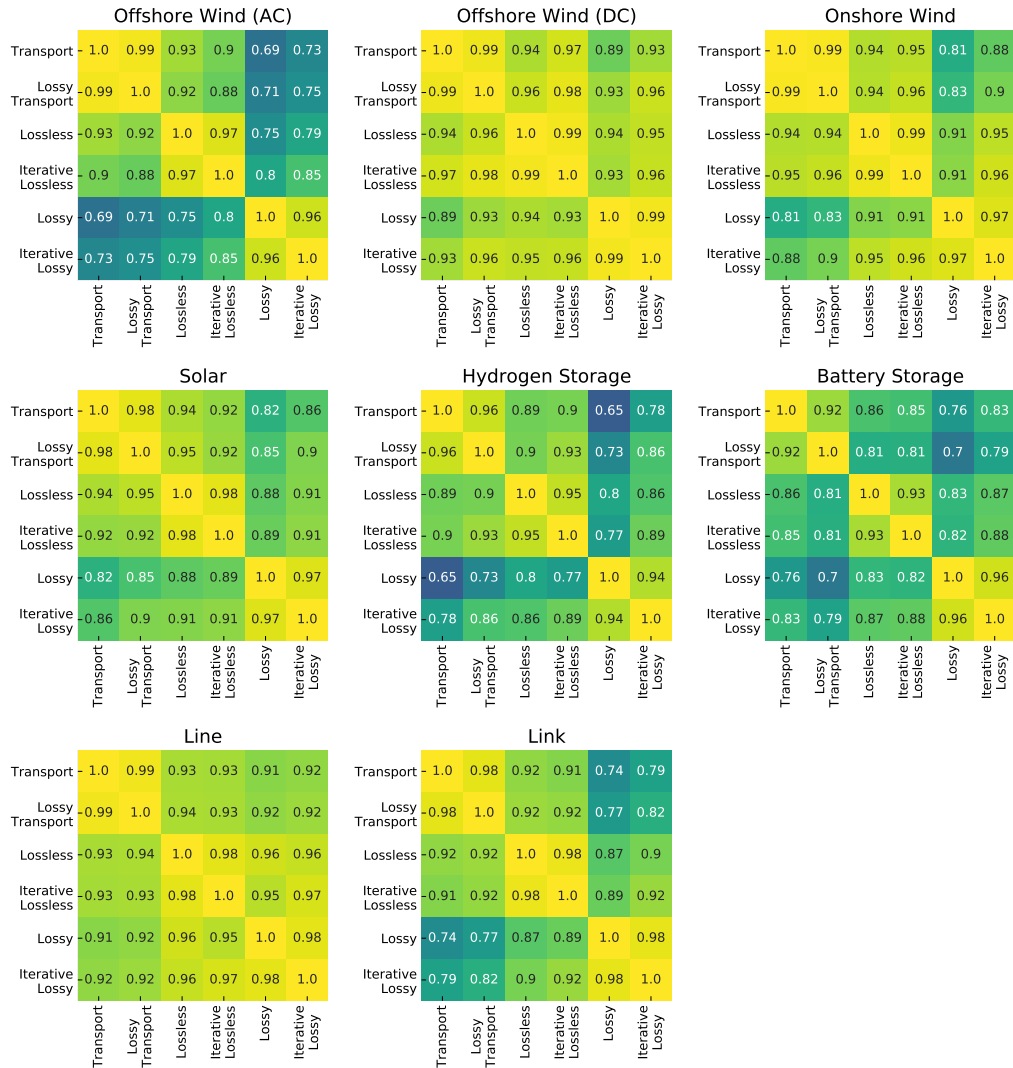


Figure 3.4: Capacity correlation of nodal investments distinguished by technology.

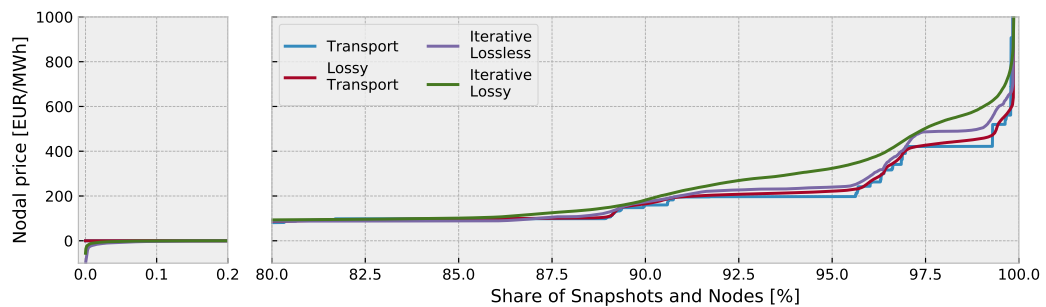


Figure 3.5: Nodal price duration curves (snapshots and nodes) for selected flow models. In the omitted section, prices rise steadily and similarly for all models. Some models allow for negative nodal prices with occurrence below 0.2%.

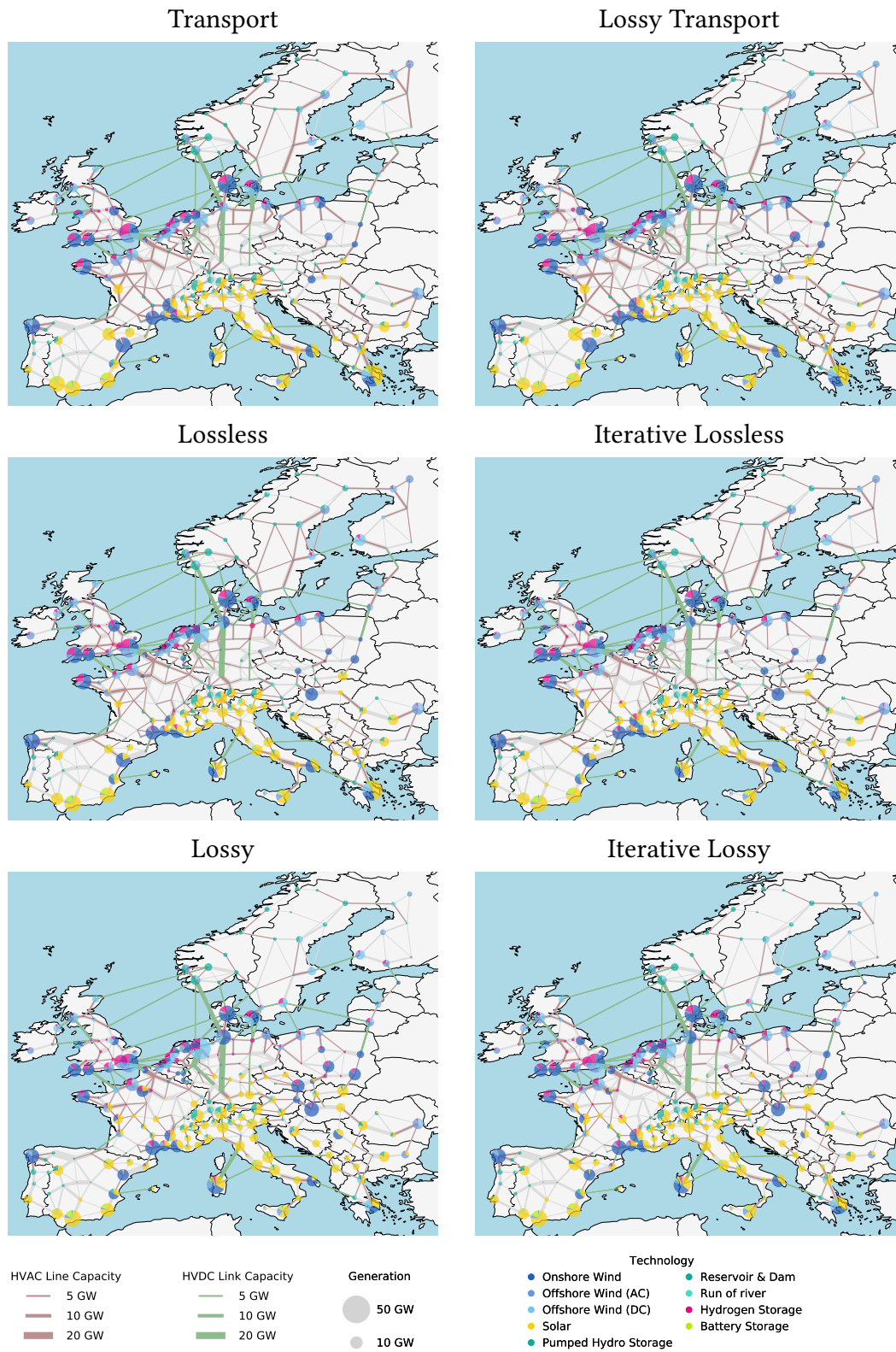


Figure 3.6: Maps of cost-optimal capacity expansion for the covered lossless flow models.

differences concern grid reinforcement. The lossless and lossy transport models feature many new transmission lines in France and Scandinavia, which disappear as more accurate flow models are applied. The difference adds up to 20% less network reinforcement. Likewise, the energy transmitted decreases as more constraints are imposed on power transmission. In order to avoid grid losses, models that consider transmission losses and KVL transmitted up to 66% less energy than the transport model. The reduced spatial transport of power is then compensated by a shift towards hydrogen storage and controllable HVDC links (e.g. in the West of Germany). Despite the involved conversion losses, balancing renewables in time through storage becomes more attractive. Additionally, to offset the energy lost by transmission but also the reduced amount of power transmission, lossy models feature more wind and solar generation capacity. This includes both more localised generation (e.g. more solar panels in Southern Germany and more onshore wind turbines in Eastern Europe) where previously there were few production sites, and more concentrated generation in the North Sea region to pair with the appended storage units. The added capacities raise the system cost. In total, the annual system costs increase by approximately 5.7% compared to iterative linearised power flow, or 8% relative to the transport model.

Besides investments, we also compare electricity prices in an idealised nodal market by using the dual variables of the nodal balance constraints. The price duration curves depicted in Figure 3.5 show that nodal prices are more evenly distributed in the lossless linear power flow compared to the transport model. The even distribution of prices was also found in Gunkel *et al.* [99]. The transport model and lossy transport model do not have the properties that would allow negative prices. Negative nodal prices are a consequence of KVL and occur when increasing demand at a bus relieves a transmission line, allowing power to be exported from somewhere cheap to somewhere expensive. This lowers the system cost and consequently results in a negative price at that bus. Other constraints that can generally entail negative prices are unit commitment constraints, but these are not considered in this contribution. We find that even for models with KVL and loss approximations, negative prices are rare ($\leq 0.2\%$). The major differences regarding nodal prices can be observed in the 10% of highest prices. The transport model features step-like price profiles, whereas the profiles of the other models are smoother. The iterative lossy linearised power flow model possesses the highest yet smoothest price duration curve.

3.5.2 Validation of Loss Approximation

Figure 3.7 relates optimised line flows $p_{\ell,t}$ to optimised losses $\psi_{\ell,t}$ for the lossy transport model and the iterative lossy linearised power flow model. The lossy

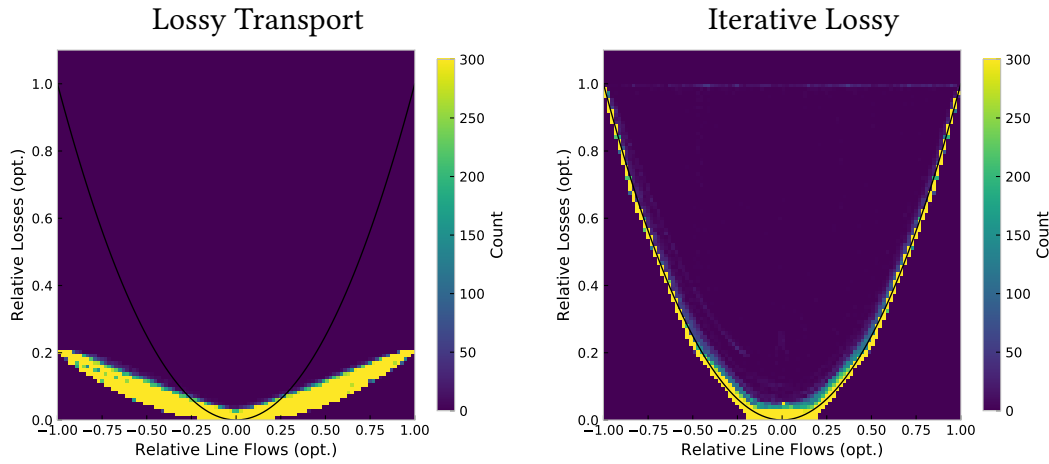


Figure 3.7: Examination of convex envelope relaxation around loss formula $\psi = rp^2$ given in Equation (3.31) for lossy transport model and the iterative lossy linearised power flow model in a two-dimensional histogram. The line flows are normalised by their nominal capacity including the $N - 1$ security margin ($p_{\ell,t}/\bar{p}_\ell P_\ell$) and maximum losses according to security-constrained line capacity respectively, such that lines with different electrical parameters can be mapped onto the same chart. The count refers to a tuple (ℓ, t) of line and snapshot. The black line depicts the normalised quadratic loss formula Equation (3.31).

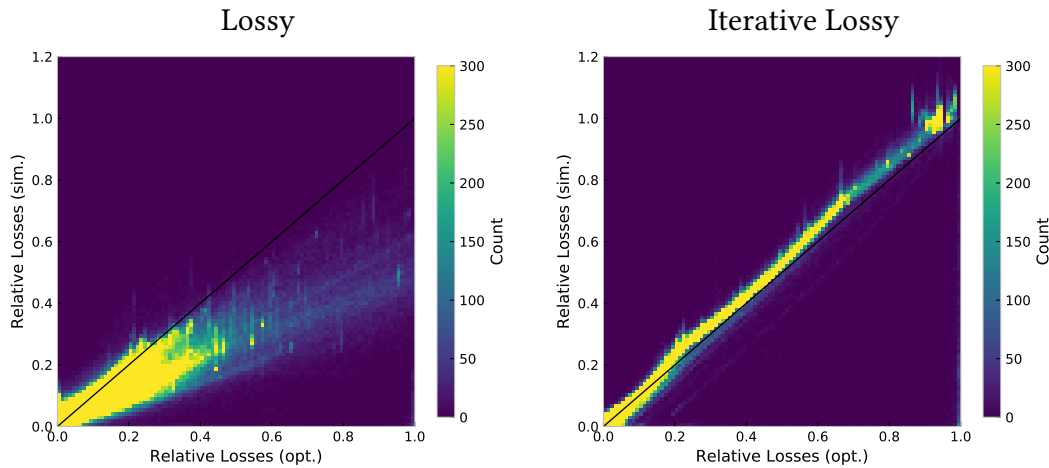


Figure 3.8: Comparison of simulated losses from AC power flow equations and optimised losses for iterative and non-iterative lossy linearised power flow in a two-dimensional histogram. Relative losses are shown as $\psi_{\ell,t}/\psi_\ell^{\max}$ according to security-constrained line capacity $\bar{p}_\ell P_\ell$. The count refers to a tuple (ℓ, t) of line and snapshot. The black line indicates perfect alignment of simulated and optimised losses.

Table 3.3: Flow errors compared across covered flow models.

Indicator	Unit	Lossy		Iterative Lossless		Iterative Lossy	
		Transport	Transport	Lossless	Lossless	Lossy	Lossy
Root Mean Squared (RMSE)	MW	1468	1059	790	679	298	60
Mean Absolute (MAE)	MW	775	707	269	207	194	35
Pearson Correlation (R)	-	0.91	0.94	0.97	0.98	0.99	0.998
Coef. of Determination (R ²)	-	0.83	0.89	0.94	0.95	0.98	0.996

transport model underestimates losses under high loading conditions depending on the assumed constant loss factor and fails to reflect the quadratic relationship between losses and flow. On the contrary, the results also confirm that approximating losses in linearised optimal power flow with a convex envelope does not degrade the obtained solutions. Although the envelope around the loss parabola Equation (3.31) (see Section 3.3.4, approximates cosine in Equation (3.28)) allows for losses to take values above the parabola, the cost associated with losses tends to push losses downwards. Substantial deviations from the parabola to above only occur when there is no cost (or even a benefit in the case of negative nodal prices) associated with higher losses; e.g. when energy is being curtailed, or when there is some extra consumption of interest to control power flows or some other problem degeneracy. As previously shown in Figure 3.5, negative nodal prices and consequently incentives for loss overestimation are rare ($\leq 0.2\%$). These circumstances cause the generous convex relaxation to function well. Underestimating losses is also possible, albeit to a much smaller extent, as a small fraction of the feasible space lies between the loss parabola and the tangents that form the convex envelope. Recall that the loss parabola Equation (3.31) is already an approximation of the cosine terms in Equation (3.28).

Figure 3.8 compares transmission losses retrieved from the optimisation problem to the simulated losses from AC power flow for the iterating and non-iterating loss approximation. Like in Figure 3.7, we note that the iterative lossy formulation manages to sufficiently represent losses observed in the respective AC power flow simulation. However, when the iteration is skipped and hence line impedances are not updated according to their optimised capacities, losses are overestimated.

3.5.3 Validation of Optimised Line Flows

Figure 3.9 compares line flows from optimisation to simulated line flows from AC power flow for each of the flow models in a two-dimensional histogram. Figure 3.10 displays the same information from a different perspective as duration curves of relative line loading for both simulated and optimised flows (inspired by Brown *et al.* [86]). Table 3.3 quantifies the alignment of optimised and simulated flows with

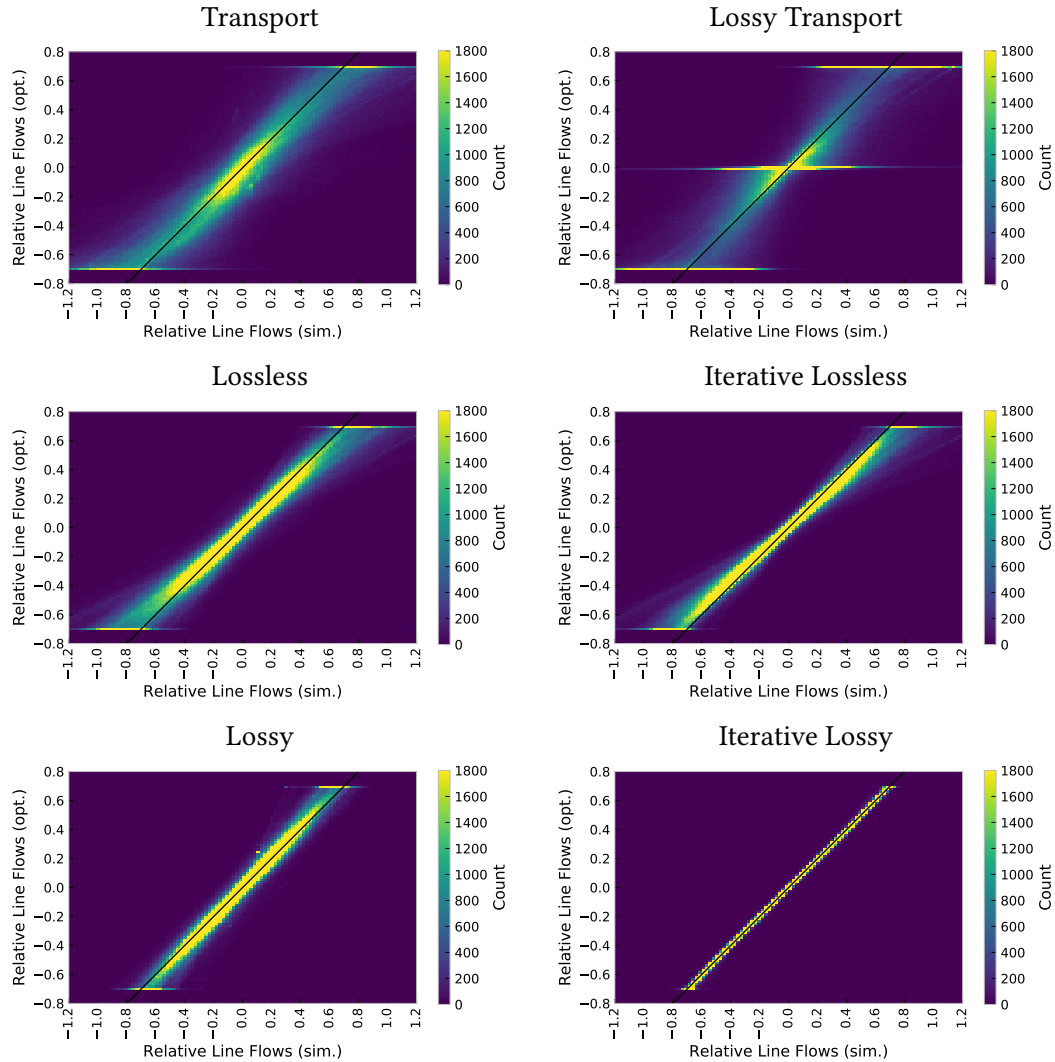


Figure 3.9: Two-dimensional histograms comparing simulated flows (AC power flow) and optimised flows of the indicated flow models. Relative line flows are shown as $p_{\ell,t}/P_{\ell}$. The count refers to a tuple (ℓ, t) of line and snapshot. The black line indicates perfect alignment of simulated and optimised flows.

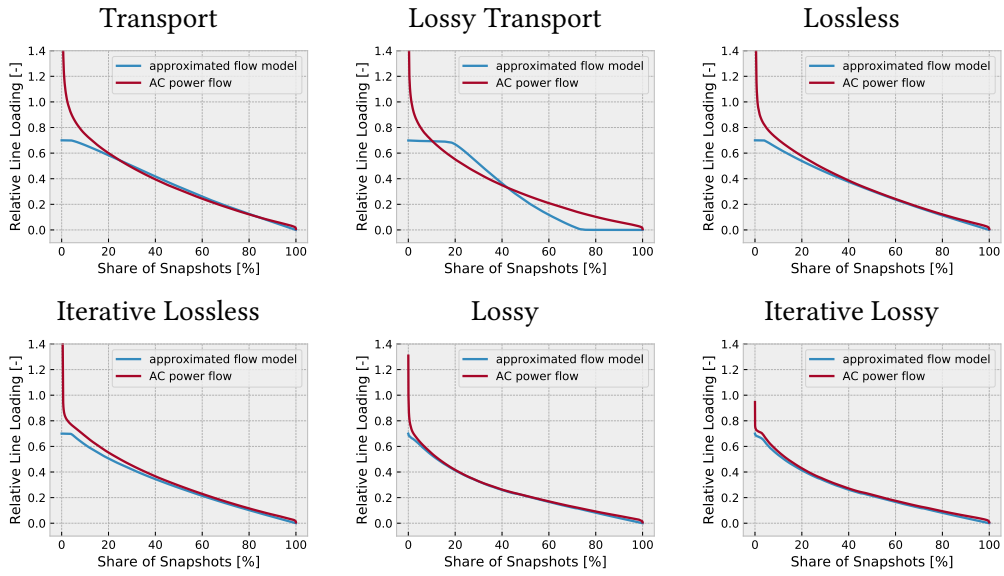


Figure 3.10: Flow duration curves of simulated flows (AC power flow) and optimised flows for the indicated flow models. Relative line loading is shown as $p_{\ell,t}/P_{\ell}$. The count refers to a tuple (ℓ, t) of line and snapshot.

Figure 3.11: Duration curves (lines and snapshots) of nodal price differences for lines experiencing no flow, congested lines, and all lines.

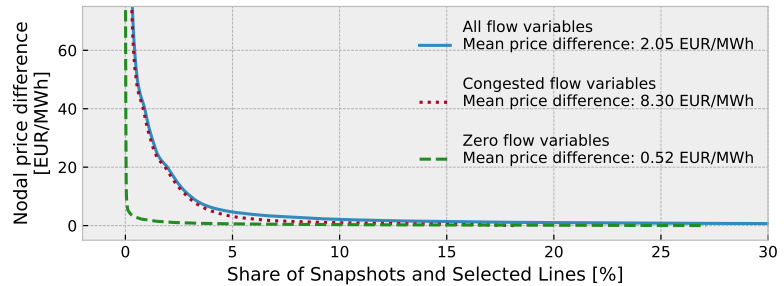
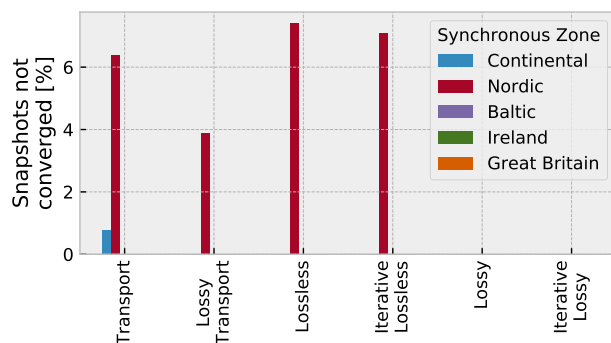


Figure 3.12: Share of snapshots where the Newton-Raphson algorithm for solving the AC power flow equations did not converge distinguished by colour-coded synchronous zone.



some standard absolute and relative measures of error that are frequently encountered in the literature (see [107]): root mean squared error (**RMSE**), mean absolute error (**MAE**), Pearson Correlation Coefficient (R), Coefficient of Determination (R^2).

First and foremost, the results reveal that the iterative lossy model matches simulated flows almost perfectly. Other formulations show deficiencies particularly under high loading conditions, but generally get the direction of flow right. The errors become significantly less pronounced and produce less undesired line overloading, the more physical characteristics of power flow are considered during optimisation. Limiting the utilisation of line capacities to 70% prevents abundant overloading. Remarkably, a high Pearson correlation coefficient of 0.91 is already achieved with the transport model, indicating that despite its simplicity the model can capture the dominant flow patterns we observe in the ex-post AC power flow simulation.

Lines with zero flow occur strikingly frequently in the lossy transport model, causing high deviations from the simulated flows. This can be explained with the aid of [Figure 3.11](#). There are many cases where prices are (almost) the same at two neighbouring buses. In such cases, there is no strict economic need to move power between them. With a lossless transport model there is no penalty for moving power between the two nodes, such that the optimisation yields a random value. However, for the lossy transport there is an incentive to set the flow to zero to avoid the losses, which is why exactly this phenomenon frequently occurs when there is no price difference. The physical flow constraints enforced by **KVL** make it complicated to realise zero flow on a line. This is the reason why we do not observe many lines with zero flow for models that enforce **KVL**. Conversely, [Figure 3.11](#) also shows that congested lines cause high nodal price differences.

In some cases the Newton-Raphson algorithm does not converge. Typical causes can be high voltage angle differences, voltage drops, and reactive power flows. The power flow simulation is run separately for each snapshot and each synchronous zone, so we can check individually what prevalent network characteristics, in combination with the underlying flow models, cause the failure to converge. The resulting share of snapshots not converged for each synchronous zone is presented in [Figure 3.12](#). Almost exclusively, difficulties are observed in the Nordic synchronous zone which possesses many long (aggregated) lines, which lead to high voltage angle differences. With regard to the whole European system, the number of snapshots where no convergence is reached is low. We observe better convergence rates for more detailed flow models and the issue is found to become less problematic as the spatial resolution of the transmission network is increased.

Given that high voltage angle differences diminish the accuracy of the linear power flow approximation, a maximum of up to $\pm 30^\circ$ is commonly tolerated in the literature [106, 128, 129]. This domain links to the range beyond which the

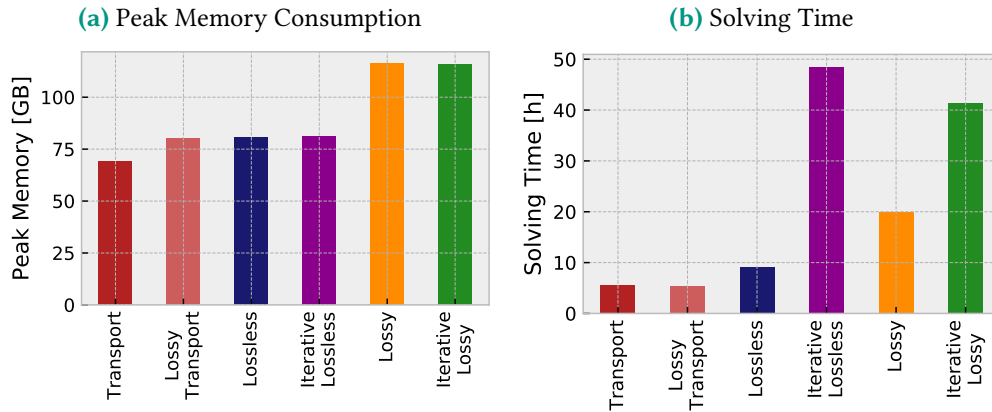


Figure 3.13: Comparison of computational performance in terms of peak memory consumption and solving time.

relative error of the small-angle approximation of the sine exceeds 5%. Since the cosine approximation is a second order Taylor series expansion, unlike the first order sine approximation, it does not reduce the acceptable range of angle differences further (see Table 3.3). We observe that across all flow models a majority of voltage angle differences lies within an uncritically low range where the sine approximation is quite precise (Table 3.2). The share of voltage angle differences outside $\pm 30^\circ$ reduces considerably with more physically accurate grid modelling (5% for transport model versus 1.5% for lossy model).

3.5.4 Computational Performance

The computational performance of the different flow models, both in terms of memory and computation time, is shown in Figure 3.13. More variables and constraints leads to higher peak memory consumption. The spectrum ranges from 70 GB to 130 GB (around factor 2). Particularly the loss approximation raises memory requirements significantly in relation to added KVL constraints or constant efficiencies, also depending on the number of tangents used for the convex envelope. Solving times range between 5 hours and 50 hours (factor 10). Lossy and lossless transport model are solved the fastest by far. The lossless linearised power flow model requires almost twice the time. Iteration has the biggest impact on solving times, multiplying with the number of iterations. Finally, we notice that the lossy formulations are more prone to numerical issues, which could be circumvented by increasing the numeric accuracy parameter of the solver at the cost of computational speed.

3.6 Critical Appraisal

The disregard of voltages and reactive power flows during optimisation ranks among the severest shortcomings of the presented flow models. The cost and required capacities for reactive power control are not assessed. The confinement to linear formulations may also be considered as a weakness in view of recent promising developments in convex second-order cone solvers.

Additionally, we consider the high-voltage transmission network only and do not assess the performance of flow models in low-voltage distribution grids. This is especially relevant in view of further closing the granularity gap. Furthermore, losses on the distribution level are not directly modelled but taken into account only through the electricity demand. Typically, the scale of losses is higher than at the transmission level, as for instance the German Federal Network Agency reports [127]. In 2019, losses at the transmission level amounted to 1.2%, whereas losses at the distribution level were as high as 3%. Moreover, the relations between ambient temperature, dynamic line rating and losses are not addressed. Higher ambient temperatures reduce the amount of power a transmission line can transmit safely but simultaneously increase the resistance, affecting the losses.

Although the clustered transmission system is of course also simplified due to computational constraints, we could observe consistent results for spatial aggregation to 100, 200 and 250 nodes in complementary simulations. However, the extent of network clustering also affects the length of modelled lines and we note that for very long lines with voltage angle differences beyond $\pm 30^\circ$ in highly aggregated grid models the standard equivalent Π model may not be suitable [125].

3.7 Conclusion

In the present chapter, we discussed best practices for incorporating two inherently nonlinear phenomena, power flow and transmission losses, into linear capacity expansion problems that co-optimize investments in generation, storage and transmission infrastructure.

High model fidelity comes at the cost of high computational burden. Given the cross-disciplinary nature of energy system modelling and differences in complexity, the selection of a suitable representation of power flows depends on the application, the user's availability of computational resources, and the level of spatial detail considered. A highly aggregated network will not benefit from detailed power

flow modelling, whereas modelling losses is critical in the presence of continent-spanning power transmission at sub-national detail. The present chapter provides a detailed comparison to facilitate this choice.

We find that already as little as three tangents are sufficient to accurately approximate the quadratic losses, which in turn approximate the trigonometric losses. We do not observe excessive fictitious losses despite the broad convex relaxation. However, we conclude that accounting for changing impedances as lines are expanded is essential. Otherwise, losses will be overestimated.

The literature predominantly employs the lossless transport model in design studies, which can already capture the main features of a cost-efficient system, but is too inexact for subsequent nonlinear power flow calculations. However, a representation of power flows that considers transmission losses as well as both Kirchhoff laws allows us to bridge between techno-economic models and more detailed electrotechnical models.

Transmission Expansion Planning Using Cycle Flows

Contents of this chapter based on

Neumann, F. & Brown, T. *Transmission Expansion Planning Using Cycle Flows in Proceedings of the Eleventh ACM International Conference on Future Energy Systems (ACM e-Energy 2020) (2020)*, 253–263. doi:[10/d3qk](https://doi.org/10/d3qk). arXiv:[2004.08702](https://arxiv.org/abs/2004.08702). ©2020 Neumann, F. and Brown T. Publication rights licensed to ACM. Reprinted with permission.

4.1 Introduction

Other than in the previous [Chapter 3](#), where the detailed coordination between generation, storage and transmission infrastructure coerced us to using an LP formulation with exclusively continuous variables, the classical transmission expansion planning (TEP) problem is often formulated as discrete optimisation problem, where line capacities cannot be continuously expanded because the choice of circuit types is limited. Nonetheless, large shares of weather-dependent renewables mean that also transmission grid reinforcements need to be optimised over many representative weather and load conditions, which together with the first-choice nonlinear AC power flow and the discrete line expansion choices drives up the computational burden of TEP [130].

A common approach to TEP in the literature is to linearise the power flow equations as a trade-off between computational complexity and accuracy. Using a big- M disjunctive relaxation, the linearisation allows TEP problems to be written as mixed-integer linear problem (MILP) and solved using decomposition methods and specialised commercial solvers [24, 131–133]. However, as we have seen in [Chapter 3](#), the common angle-based approach to power flow linearisation introduces many auxiliary variables for the voltage angles, which has two major drawbacks:

it introduces many new variables and constraints, which can lead to performance problems, and it is difficult to consider the connection of multiple disconnected networks. The latter difficulty is due to the fact that the voltage angles are only defined up to a constant in each connected network, and this redundancy must be managed with care when changing the connectivity. The connection of previously-disconnected networks is particularly relevant for the connection of island systems and regions with multiple synchronous zones, like Europe, North America, China, Japan, and the Philippines, where synchronisation via HVAC transmission lines competes with other technologies such as HVDC back-to-back coupling.

An alternative formulation of the linearised power flow equations has been used for operational linear optimal power flow (LOPF) problems without discrete TEP that uses constraints imposed directly on the power flows themselves, without the use of auxiliary variables, using a cycle decomposition of the flow patterns defined by Kirchhoff's voltage law (KVL) [69]. This cycle-based formulation was shown to reduce computation times by an order of magnitude compared to the angle-based formulation in LOPF problems with generation capacity expansion. Moreover, such an alternative formulation may converge in instances where the other does not.

The cycle-based formulation has also previously been applied to the optimal transmission switching (OTS) problem which is moderately related to the TEP problem [134]. OTS is an operational problem where the network topology can be changed by switching lines on and off. In many regards OTS could be viewed as reverse TEP. While TEP is about adding new transmission lines permanently, OTS is about deliberately deactivating transmission lines temporarily. However, using a cycle-based formulation in TEP has a distinct advantage over using it in OTS: while OTS needs to consider all simple cycles [134], TEP can be formulated by supplementing the initial cycle basis with new candidate cycles because existing lines are not removed.

In the present chapter, the cycle-based formulation is therefore extended to TEP problems. Unlike the previous Chapter 3 the angle of attack is not approximation of a non-convex feasible space, but a nimble reformulation of the identical feasible space. It is shown how to choose the big- M parameters necessary for the disjunctive relaxation, which is also present in the angle-based formulation. This is important because previous studies have reported a large impact of big- M coefficients on computation times and numerical stability [30]. For both formulations, it is shown how to formulate problems where multiple disconnected networks (also called synchronous zones) may be connected, which involves managing the choice of big- M parameters and, in the case of the angle-based formulation, the relaxation of the slack voltage angle constraints. The connection of networks is found to be both easier to formulate and faster to solve for the cycle-based formulation.

Realistic benchmark cases with varying spatial and temporal resolution are provided using the open model dataset `PyPSA-Eur` from [Section 2.3](#) [13, 28]. The model covers the European transmission system and includes regionally resolved time-series for renewable generator availability and is used to formulate a coordinated expansion planning problem of generation and transmission infrastructure. All formulations have been implemented for the Python-based power system analysis toolbox `PyPSA` [12].

The remainder of the chapter is structured as follows. First, [Section 4.2](#) guides through the foundations of angle-based and cycle-based linear power flow constraints, which are subsequently adapted to the `TEP` problem in [Section 4.3](#). The competing `TEP` formulations are benchmarked and assessed in [Section 4.4](#) and [Section 4.5](#), before conclusions are drawn in [Section 4.6](#).

4.2 Linear Optimal Power Flow

`LOPF` problems typically optimize the dispatch of generation assets in a network with the objective to minimise costs at the same time as enforcing the physical flow of power using the linear approximation of the power flow equations. More general problems consider multiple time periods, so that storage assets can be optimised as well as investments in assets taking into account representative load and weather situations.

The present chapter considers long-term investment planning problems like the one presented in [Section 2.1](#), that seek cost-effective solutions to reduce greenhouse gas emissions in the power system, of which `LOPF` is one principal building block. Among others, such problems constrain the absolute active power flows p_ℓ^0 in all existing lines $\ell \in \mathcal{L}^0$ to remain within their nominal capacities P_ℓ^0

$$|p_\ell^0| \leq P_\ell^0. \quad (4.1)$$

The label 0 indicates lines in the existing network. Additionally, Kirchhoff's current law (`KCL`) and `KVL` govern the flow p_ℓ^0 . A variety of mathematically equivalent `LOPF` formulations exists, many of which were compared and benchmarked in previous work [69]. In continuous linear capacity expansion problems without discrete transmission expansion planning the choice of the `LOPF` formulation was shown to impact computation times greatly.

In preparation for their extension to discrete transmission expansion planning in subsequent [Section 4.3](#), this section reviews two `LOPF` formulations used in this setting. These are (i) the common angle-based formulation using voltage angles as auxiliary variables ([Section 4.2.2](#)) and (ii) a more efficient cycle-based formulation

deduced from graph-theoretical considerations (Section 4.2.3). Both formulations share the constraints for representing KCL (Section 4.2.1), but differ in their formulation of KVL. While the former has previously been widely used in TEP studies [24, 131, 132], the application of the latter is a novel contribution of this chapter.

4.2.1 Kirchhoff's Current Law (KCL)

As known from Chapter 3, KCL requires the power injected at each bus to equal the power withdrawn by attached lines; i.e.

$$p_i = \sum_{\ell} K_{i\ell} p_{\ell}^0 \quad \forall i \in \mathcal{N} \quad (4.2)$$

where p_i is the active power injected or consumed at node $i \in \mathcal{N}$, p_{ℓ}^0 is the active power flow on line ℓ , and $K \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{L}^0|}$ is the incidence matrix of the network graph which has non-zero values +1 if line ℓ starts at bus i and -1 if line ℓ ends at bus i . The orientation of lines is arbitrary but fixed [135].

KCL provides $|\mathcal{N}|$ linear equations for the $|\mathcal{L}^0|$ unknown flows p_{ℓ}^0 , of which one is linearly dependent [69]. If the network is a tree with $|\mathcal{L}^0| = |\mathcal{N}| - 1$, Equation (4.2) is already sufficient to uniquely determine the flows p_{ℓ}^0 . However, in meshed networks $|\mathcal{L}^0| - |\mathcal{N}| + 1$ additional independent equations are required. These are provided by KVL.

4.2.2 Angle-based Kirchhoff's Voltage Law (KVL)

In textbooks and software toolboxes, KVL for the linearised power flow is commonly formulated in terms of the voltage phase angles $\{\theta_i\}_{i \in \mathcal{N}}$ [104, 105]. This angle-based formulation originates directly from applying the assumptions for linearised power flow to the nonlinear power flow equations in voltage-polar coordinates of lines $\ell \in \mathcal{L}^0$ (see Equations (3.1) to (3.2)). Assuming (i) all voltage magnitudes $|V_i|$ are close to one per unit, (ii) conductances g_{ℓ} are negligible relative to the susceptances b_{ℓ} , (iii) voltage angle differences are small enough such that $\sin(\theta_i - \theta_j) \approx \theta_i - \theta_j$, and (iv) reactive power flows q_{ℓ} are negligible compared to real power flows p_{ℓ} leads to

$$p_{\ell}^0 = \frac{\partial_{\ell}}{x_{\ell}^0} = \frac{1}{x_{\ell}^0} \sum_i K_{i\ell} \theta_i \quad \forall \ell \in \mathcal{L}^0 \quad (4.3)$$

where $x_\ell^0 = b_\ell^{-1}$ is the line reactance and $\vartheta_\ell = \theta_i - \theta_j$ is the voltage angle difference between nodes i and j which line ℓ connects [104]. Additionally, a reference voltage angle is commonly set at one bus for each synchronous zone

$$\theta_i = 0 \quad \forall i \in \mathcal{N}_0, \quad (4.4)$$

where \mathcal{N}_0 denotes the set of slack buses. This circumvents the rotational degeneracy¹ that originates from the invariance of the network flows to adding a constant to all voltage angles $\theta_i \rightarrow \theta_i + c$ [108]. Together with the KCL constraints, the angle-based formulation provides $|\mathcal{L}^0| + |\mathcal{N}|$ independent equality constraints to determine the $|\mathcal{L}^0|$ flows and $|\mathcal{N}|$ angles.

4.2.3 Cycle-based Kirchhoff's Voltage Law (KVL)

KVL states that the sum of voltage angle differences across lines around all cycles in the network must sum to zero. This allows a reformulation of the linearised power flow equations which circumvents the auxiliary voltage angle variables. The consistency of voltage angle summations within a connected network can alternatively be achieved by using a cycle basis of the network graph $\mathcal{G} = (\mathcal{N}, \mathcal{L}^0)$. A cycle basis is a subset of all simple cycles of \mathcal{G} such that any other cycle can be described by a linear combination of cycles in the cycle basis [136, 137]. It can be constructed from a minimum spanning tree \mathcal{T} of the network graph in $\tilde{O}(|\mathcal{N}| \cdot |\mathcal{L}^0|^2)$ [138]. The tree \mathcal{T} has $|\mathcal{N}| - 1$ edges [139]. Together with the path in \mathcal{T} connecting their nodes, each of the $|\mathcal{L}^0| - |\mathcal{N}| + 1$ remaining edges of \mathcal{G} creates a cycle of the cycle basis. These cycles are linearly independent because each cycle contains an edge that is not contained in the other cycles and consequently constitute a basis of the cycle space of \mathcal{G} [140]. These are sufficient to express KVL and uniquely determine the flows p_ℓ^0 [141]. The independent cycles $c \in \{1, \dots, |\mathcal{L}^0| - |\mathcal{N}| + 1\}$ are expressed as a directed linear combination of the lines ℓ in the cycle incidence matrix

$$C_{\ell c}^0 = \begin{cases} 1 & \text{if edge } \ell \text{ is element of cycle } c, \\ -1 & \text{if reversed edge } \ell \text{ is element of cycle } c, \\ 0 & \text{otherwise.} \end{cases} \quad (4.5)$$

Then, KVL can be written as

$$\sum_{\ell} C_{\ell c}^0 \vartheta_\ell = 0 \quad \forall c = 1, \dots, |\mathcal{L}^0| - |\mathcal{N}| + 1. \quad (4.6)$$

¹The term degeneracy is used to describe the condition where different values for optimisation variables yield same optimal objective value. Degeneracy is known to have a detrimental impact on the convergence of both simplex and interior-point methods.

where ϑ_ℓ is the voltage angle difference. By substituting Equation (4.3) into Equation (4.6), KVL can be expressed in terms of the power flows as

$$\sum_{\ell} C_{\ell c}^0 x_{\ell}^0 p_{\ell}^0 = 0 \quad \forall c = 1, \dots, |\mathcal{L}^0| - |\mathcal{N}| + 1. \quad (4.7)$$

Consequently, while the angle-based formulation defines KCL and KVL with $|\mathcal{L}^0| + |\mathcal{N}|$ variables and $|\mathcal{L}^0| + |\mathcal{N}|$ independent equality constraints, the equivalent cycle-based formulation requires only $|\mathcal{L}^0|$ variables and $|\mathcal{L}^0|$ independent equality constraints. Besides fewer variables and constraints, the cycle-based formulation moreover features sparser constraints than the angle-based formulation.

The computational appeal of this reformulation was evaluated in [69] for multi-period linear optimal power flow problems with generator capacity expansion and has been applied in other publications [134, 135, 139, 140, 142]. It has further been proven in [139] the cycle-based formulation also holds for multigraphs² which is particularly relevant for its extension to transmission expansion planning.

Retrospective Calculation of Voltage Angles

The cycle-based formulation does not include variables for the voltage angles. However, if needed for instance to assess the validity of linear power flow assumptions, they can be calculated after optimisation using the observed net nodal power injection or consumption p_i . By substituting Equation (4.3) into Equation (4.2) one obtains

$$p_i = \sum_{\ell \in \mathcal{L}^0} K_{i\ell} \frac{1}{x_{\ell}} \sum_{j \in \mathcal{N}} K_{j\ell} \theta_j \quad \forall i \in \mathcal{N}. \quad (4.8)$$

This can be rewritten with a weighted Laplacian $L = KBK^T$ where B is a diagonal matrix with elements $B_{\ell\ell} = b_{\ell} = x_{\ell}^{-1}$, leading to

$$p_i = \sum_j L_{ij} \theta_j \quad \forall i \in \mathcal{N}. \quad (4.9)$$

This can be solved for θ_i with

$$\theta_i = \sum_j (L^{-1})_{ij} p_j \quad \forall i \in \mathcal{N}. \quad (4.10)$$

However, L is not invertible as it has a zero eigenvalue with eigenvector $\mathbf{1}$. Since Equation (4.4) provides a reference voltage angle for one bus, the remaining voltage

²Multigraphs are graphs allowing parallel edges between the same two vertices.

angles $\{\theta_i\}_{i \in \mathcal{N} \setminus \mathcal{N}_0}$ can be found by inverting the submatrix $L' \in \mathbb{R}^{|\mathcal{N} \setminus \mathcal{N}_0| \times |\mathcal{N} \setminus \mathcal{N}_0|}$ of L which omits the row and column corresponding to the slack bus.

4.3 Transmission Expansion Planning

In [TEP](#) we consider the discrete reinforcement of transmission lines based on a set of candidate lines \mathcal{L}^1 . The label 1 indicates candidate lines. We extend the optimisation problem from [Section 4.2](#) by introducing a binary investment variable $i_\ell \in \mathbb{B}$ for each candidate line $\ell \in \mathcal{L}^1$ and then formulate constraints on the power flow p_ℓ^1 depending on the investment decision.

If the candidate line ℓ is not built, the power flow must be zero. Otherwise, the absolute power flow must not exceed the nominal capacity P_ℓ^1 of the candidate line. This is expressed by the constraint

$$|p_\ell^1| \leq i_\ell P_\ell^1 \quad \forall \ell \in \mathcal{L}^1. \quad (4.11)$$

Just like existing lines $\ell \in \mathcal{L}^0$ the capital cost of candidate lines $\ell \in \mathcal{L}^1$ are included in the objective function and nodal balance constraints in [Equation \(4.2\)](#) defining [KCL](#).

To define [KVL](#) for candidate lines, both angle-based and cycle-based [KVL](#) constraints, given in [Equation \(4.3\)](#) and [Equation \(4.7\)](#), need to be edited such that they can only be active if the associated candidate lines are built. To achieve this, both formulations make use of the big- M disjunctive relaxation. These modifications are subsequently developed in [Section 4.3.1](#) for the angle-based and [Section 4.3.2](#) for the cycle-based [KVL](#) constraints.

4.3.1 Angle-based Transmission Expansion Planning

The angle-based [KVL](#) constraint of the [TEP](#) problem is widely known from [[24](#), [104](#), [131](#), [132](#)]. It transforms the [KVL](#) equality constraint from [Equation \(4.3\)](#) into the two inequalities

$$\begin{aligned} p_\ell^1 - \frac{\vartheta_\ell}{x_\ell^1} &\geq -M_\ell^{\text{KVL}}(1 - i_\ell) \\ &\leq +M_\ell^{\text{KVL}}(1 - i_\ell) \quad \forall \ell \in \mathcal{L}^1, \end{aligned} \quad (4.12)$$

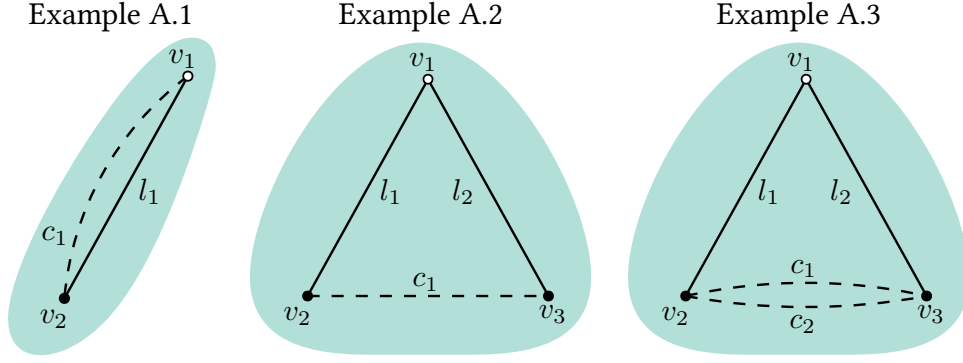


Figure 4.1: Example Group A. Candidate lines within a synchronous zone. Candidate lines denoted by c_i and existing lines by l_i .

where, as previously, $\vartheta_\ell = \theta_i - \theta_j$. If the big- M parameters M_ℓ^{KVL} are suitably chosen, the inequalities in Equation (4.12) are inactive if $i_\ell = 0$, but together form the original equality constraint if $i_\ell = 1$.

However, big- M parameters are known to easily incur numerical challenges [30, 131]. Therefore, M_ℓ^{KVL} are ideally chosen as large as necessary, to guarantee that the KVL constraint is inactive whenever the candidate line is not built, and as small as possible, to avoid a detrimental large value range in the optimisation problem's coefficient matrix.

For the derivation of appropriate values it is necessary to distinguish candidate lines which connect buses within the same synchronous zone ($\mathcal{L}_{\text{intra}}^1 \subseteq \mathcal{L}^1$, Section 4.3.1) and candidate lines which connect multiple synchronous zones ($\mathcal{L}_{\text{inter}}^1 \subseteq \mathcal{L}^1$, Section 4.3.1).

Big- M Parameters for KVL Constraints Within Synchronous Zone

The derivation of minimal values for M_ℓ^{KVL} for candidate lines $\ell \in \mathcal{L}_{\text{intra}}^1$ which connect buses of the same synchronous zone largely follows [131, 143], but is reproduced here to facilitate a comparison with the novel cycle-based formulation and to set the notation.

Theorem 1 *The value of the disjunctive constant M_ℓ^{KVL} for a candidate line ℓ that connects two buses i and j of the same synchronous zone can be chosen following*

$$M_\ell^{\text{KVL}} \geq \frac{|\mathcal{P}_{ij}^{\text{min}}|}{x_\ell^1} \quad (4.13)$$

where $|\mathcal{P}_{ij}^{\text{min}}|$ is the length of the shortest path between the buses i and j along edges k of the existing network graph $\mathcal{G} = (\mathcal{N}, \mathcal{L}^0)$ with weights $P_k^0 x_k^0$.

Proof 1 Let $\ell \in \mathcal{L}_{\text{intra}}^1$ be a particular candidate line for which Equation (4.11) and Equation (4.12) hold. In the case $i_\ell = 0$ it follows from Equation (4.11) that $p_\ell^1 = 0$ and from Equation (4.12) that

$$-M_\ell^{\text{KVL}} x_\ell^1 \leq \theta_i - \theta_j \leq M_\ell^{\text{KVL}} x_\ell^1. \quad (4.14)$$

Equation (4.14) represents a limit on the voltage angle difference and the value of M_ℓ^{KVL} must be chosen such that for as long as $i_\ell = 0$ this limit is never reached. Otherwise invalid limits on the angle difference are imposed. We must therefore derive valid big-M parameters from constraints on the voltage angle difference that are already enforced through the existing network.

If there exists a line $\ell \in \mathcal{L}^0$ in parallel to the candidate line (e.g. as in Example A.1 in Figure 4.1) we can obtain these by substituting Equation (4.3) into Equation (4.1), yielding the limits

$$-P_\ell^0 x_\ell^0 \leq \theta_i - \theta_j \leq P_\ell^0 x_\ell^0. \quad (4.15)$$

By combining Equation (4.14) and Equation (4.15) we can retrieve a minimum value for M_ℓ^{KVL} :

$$M_\ell^{\text{KVL}} \geq \frac{P_\ell^0 x_\ell^0}{x_\ell^1} \quad (4.16)$$

Now consider the slightly more complicated case where the candidate line ℓ is not a duplication of an existing line (e.g. as in Example A.2 in Figure 4.1). The theorem specifies that the buses i and j of ℓ are part of the same synchronous zone. Thus, there is at least one sequence $\mathcal{P}_{i,j} = \{k(i, b_1), k(b_1, b_2), \dots, k(b_n, j)\}$ of existing lines $k \in \mathcal{L}^0$ along buses $\{b_m\}_{m=1, \dots, n}$ which already connects these buses. Hence, just as with an existing parallel line there is an existing limit on the voltage angle difference, only that the limit is not given by just one existing line but by a set of existing lines:

$$-\sum_{k \in \mathcal{P}_{i,j}} P_k^0 x_k^0 \leq \theta_i - \theta_j \leq \sum_{k \in \mathcal{P}_{i,j}} P_k^0 x_k^0 \quad (4.17)$$

To find the tightest limit on $\theta_i - \theta_j$ we need to find the shortest path $\mathcal{P}_{i,j}^{\text{min}}$ among all possible paths $\mathcal{P}_{i,j}$ with weights $P_k^0 x_k^0$ using e.g. the Dijkstra algorithm, which then yields

$$M_\ell^{\text{KVL}} \geq \frac{|\mathcal{P}_{i,j}^{\text{min}}|}{x_\ell^1} = \frac{\sum_{k \in \mathcal{P}_{i,j}^{\text{min}}} P_k^0 x_k^0}{x_\ell^1} \quad (4.18)$$

as specified in the theorem.

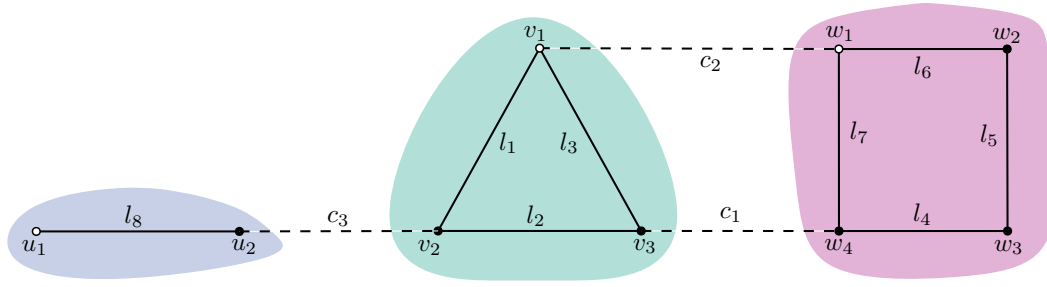


Figure 4.2: Example C. Candidate lines across synchronous zones. Candidate lines denoted by c_i and existing lines denoted by l_i .

Big- M Parameters for KVL Constraints Across Synchronous Zones

If the buses connected by candidate line ℓ are not part of the same synchronous zone and therefore no path exists to infer an existing limit on the voltage angle difference, it is possible to fall back to a significantly larger value

$$M_{\ell}^{\text{KVL}} \geq \frac{\sum_{k \in \mathcal{L}^0 \cup \mathcal{L}^1} P_k x_k}{x_{\ell}^1} \quad (4.19)$$

which has been proven in [144] to be a valid choice for any combination of line investment decisions, under the condition that a reference angle is defined for all synchronous zones such that $\theta_i = 0 \forall i \in \mathcal{N}_0$ if no new lines are built. Otherwise, due to the rotational degeneracy no relation could be established between the nodal voltage angles of disconnected networks.

Slack Constraints Across Synchronous Zones

If multiple synchronous zones may be connected by building new lines, the slack constraint in Equation (4.4) must also be modified, since it applies separately in each connected network. When two networks are connected, one of the slack constraints should be relaxed. The slack constraints cannot simply be dropped because the derivation of big- M parameters for the KVL constraints across synchronous zones (Section 4.3.1) depends on a calculable maximal voltage angle difference across synchronous zones even if they are not coupled. Available transmission expansion studies that alleviate rotational degeneracy of voltage angles with slack constraints have not dealt with this case. In this section a novel treatment of the connection of multiple synchronous zones is provided that handles the slack constraints by managing the combinatorics of possible relaxations that apply as networks are connected.

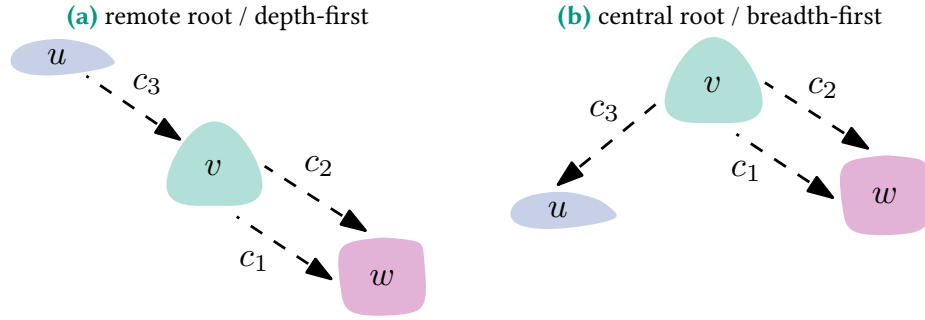


Figure 4.3: Example C shown as different directed rooted trees of the subnetwork graph. In the depth-first variant, u is the root subnetwork, c_3 relaxes the slack of v and c_1 or c_2 relax the slack of w . In the breadth-first variant, v is the root subnetwork, c_3 relaxes the slack of u and c_1 or c_2 relax the slack of w .

Initially, consider Example C in Figure 4.2 where only c_1 is a candidate line which, if built, would synchronize two synchronous zones v and w . If c_1 is built, one of the constraints in Equation (4.4) regarding the two slack buses v_1 and w_1 must be rendered ineffective. Otherwise, the nodal voltage angles would be fixed at two buses within the same synchronous zone, but the voltage angle difference between buses determines the flow. The solution would yield invalid or infeasible power flows. Therefore, we adjust the slack constraint of w to $|\theta_{w_1}| \leq i_{c_1} M_{c_1}^{\text{slack}}$, where $M_{c_1}^{\text{slack}}$ is a sufficiently large constant.

Now consider for Example C in Figure 4.2 the case where additionally c_2 is a candidate line which connects the same two synchronous zones as c_1 . In this case, we must agree on a single slack constraint relaxed by c_1 and c_2 as otherwise, if both are built, no slack constraint would remain to alleviate rotational degeneracy. Hence, the slack constraint of w is adjusted to $|\theta_{w_1}| \leq \sum_{\ell \in \{c_1, c_2\}} i_{\ell} M_{\ell}^{\text{slack}}$. The sum on the right-hand side acts as a logical OR expression such that each positive investment decision i_{ℓ} alone renders the constraint non-binding.

Next, consider the slightly more complicated Example C in Figure 4.2 where now three synchronous zones may be synchronised by candidates c_1 , c_2 and c_3 . In this case, it is essential to select a single root synchronous zone, the slack constraint of which is to be kept if all candidate lines are built. For instance, not all three candidate lines can relax the slack constraint of v as this would result in two remaining slack constraints in one synchronous zone.

Figure 4.3 sketches two possible relations between the candidate lines and the slack constraints they relax without the need to consider complementary investment decisions. It shows reduced graphs where the nodes \mathcal{S} represent all synchronous zones, and the directed edges represent the candidate lines in $\mathcal{L}_{\text{inter}}^1$ and point to the synchronous zone they affect. Since the connecting nodes are formally

different than in \mathcal{G} we label this edge set with \mathcal{L}_S^1 . In the following, we refer to this graph as the subnetwork graph $\mathcal{G}_S = (S, \mathcal{L}_S^1)$.

Generalising from the examples, we define sets of candidate lines $\mathcal{L}_v^1 \subseteq \mathcal{L}_{\text{inter}}^1$ which should turn the slack constraint of synchronous zone v non-binding. We can achieve a structure without complicating interdependencies of line investment variables if the graph of subnetworks \mathcal{G}_S is a forest of directed trees with a defined root (but allowing parallel edges). With an associated big- M constant M_ℓ^{slack} that is large enough regardless of all other investment decisions (see [Section 4.3.1](#)), we reformulate the slack constraints to

$$|\theta_v| \leq \sum_{\ell \in \mathcal{L}_v^1} i_\ell M_\ell^{\text{slack}} \quad (4.20)$$

which are correct for any combination of line investments.

If the subnetwork graph would not be a forest of directed rooted trees (with parallel edges), more interdependencies would arise due to the manifold of combinations of synchronisation scenarios. Consider Example D.1 in [Figure 4.6](#) where considering a dependency is inevitable. It is viable to encode one logical AND expression for two binary investment variables i_1 and i_2 in linear programming with an auxiliary variable i_{12} and the constraint

$$0 \leq i_1 + i_2 - 2i_{12} \leq 0 \quad (4.21)$$

[145]. But the rapidly growing number of additional binary auxiliary variables and constraints that would be required for only marginally more complicated cases, such as Example D.2, add to the appeal of reformulating the problem without voltage angle variables in cases where multiple synchronous zones may be joined.

Big- M Parameters for Slack Constraints

Having established that the subnetwork graph \mathcal{G}_S must be a forest of directed rooted trees in order to avoid considering interdependencies of investments, this section derives suitable big- M parameters for the modified slack constraints in [Equation \(4.20\)](#). It follows a similar logic as the preceding derivation for the [KVL constraints in Section 4.3.1](#).

For a start consider the simple case where there is only a single candidate line ℓ that would connect two asynchronous zones with reference buses v_1 and w_1 . Choose, without loss of generality, that ℓ relaxes the slack constraint of v_1 ($\ell \in \mathcal{L}_{v_1}^1$). Then if the candidate line built ($i_\ell = 1$),

$$\theta_{w_1} = 0 \quad \text{and} \quad |\theta_{v_1}| \leq M_\ell^{\text{slack}}, \quad (4.22)$$

where M_ℓ^{slack} is chosen such that the constraint is never binding. To determine M_ℓ^{slack} we need to find the maximum absolute voltage angle $|\theta_{v_1}|$ if the candidate line is built. This depends on the reference voltage angle θ_{w_1} . We can relate θ_{v_1} and θ_{w_1} by following a path $\mathcal{P}_{v_1, w_1}^\ell$ between the slack buses v_1 and w_1 through the graph $\mathcal{G}^\ell = (\mathcal{N}, \mathcal{L}^0 \cup \{\ell\})$ that consists of the existing network plus the candidate line ℓ via

$$\theta_{v_1} - \theta_{w_1} = \sum_{ij \in \mathcal{P}_{v_1, w_1}^\ell} \theta_i - \theta_j. \quad (4.23)$$

One can easily see this by following Example C in Figure 4.2.

$$(\theta_{v_1} - \theta_{v_3}) + (\theta_{v_3} - \theta_{w_4}) + (\theta_{w_4} - \theta_{w_1}) = \theta_{v_1} - \theta_{w_1}. \quad (4.24)$$

Knowing this we can calculate the maximum voltage angle difference between the two slack buses, as previously done in Equation (4.17) using the shortest path along lines in \mathcal{G}^ℓ with weights $P_\ell x_\ell$ to determine a lower bound for M_ℓ^{slack} :

$$M_\ell^{\text{slack}} \geq \sum_{k \in \mathcal{P}_{v_1, w_1}^{\ell, \min}} P_k x_k. \quad (4.25)$$

Now consider the slightly more complicated case of $|\mathcal{L}_{v_1}^1| \geq 2$ candidate lines $\ell \in \mathcal{L}_{v_1}^1$ where either line potentially synchronizes two separate power networks with reference buses v_1 and w_1 . We can repeat the preceding calculation of M_ℓ^{slack} for each candidate line $\ell \in \mathcal{L}_{v_1}^1$. However, the maximum voltage angle difference irrespective of all investment combinations is $\max \{M_\ell^{\text{slack}} \mid \ell \in \mathcal{L}_{v_1}^1\}$ and should therefore be chosen for both lines.

A hierarchical strategy based subnetwork graph \mathcal{G}_S is applied if multiple synchronous zones can be connected. We add the maximum big- M parameter of the upstream synchronous zone to all big- M parameters of the downstream synchronous zones, starting at the root. For instance, in Example C in Figure 4.3 using the remote root variant, the big- M constant for c_3 would be added to those of c_1 and c_2 . This approach does not yield minimal values, as it takes a detour via the slack bus of intermediate synchronous zones. However, it circumvents the need to consider investment dependencies to guarantee non-binding slack constraints. Due to this hierarchical approach, choosing a tree via breadth-first search from a central node of the subnetwork graph \mathcal{G}_S is advantageous as it generally results in lower big- M constants.

4.3.2 Cycle-based Transmission Expansion Planning

Investing in candidate lines in the transmission system can incur new cycles for which the **KVL** constraint in Equation (4.7) must hold if and only if all candidate lines which are part of a new cycle are built. In the following, these will be referred to as candidate cycles. Both existing and candidate lines can be involved in a candidate cycle. Given these candidate cycles as an incidence matrix $C_{\ell c}^1$ where $\ell \in \mathcal{L}^0 \cup \mathcal{L}^1$ we can formulate the **KVL** constraints analogously to the cycle-based load flow formulation from Equation (4.7) such that it is enforced only if all candidate lines of that cycle are built:

$$\sum_{\ell \in \mathcal{L}^0 \cup \mathcal{L}^1} C_{\ell c}^1 x_{\ell} p_{\ell} \geq -M_c^{\text{KVL}} \left(\sum_{\ell \in \mathcal{L}^1} C_{\ell c}^1 (1 - i_{\ell}) \right) \quad \forall c, \quad (4.26)$$

$$\sum_{\ell \in \mathcal{L}^0 \cup \mathcal{L}^1} C_{\ell c}^1 x_{\ell} p_{\ell} \leq +M_c^{\text{KVL}} \left(\sum_{\ell \in \mathcal{L}^1} C_{\ell c}^1 (1 - i_{\ell}) \right) \quad \forall c. \quad (4.27)$$

Like in the angle-based **TEP** formulation (cf. Section 4.3.1), the cycle-based **TEP** formulation relies on the big- M disjunctive relaxation with a sufficiently large parameter M_c^{KVL} for each candidate cycle c . The candidate cycle matrix $C_{\ell c}^1$ on the right-hand side acts as an indicator for whether candidate line ℓ is contained within the candidate cycle c . Only if all those $i_{\ell} = 1$, Equation (4.27) becomes binding.

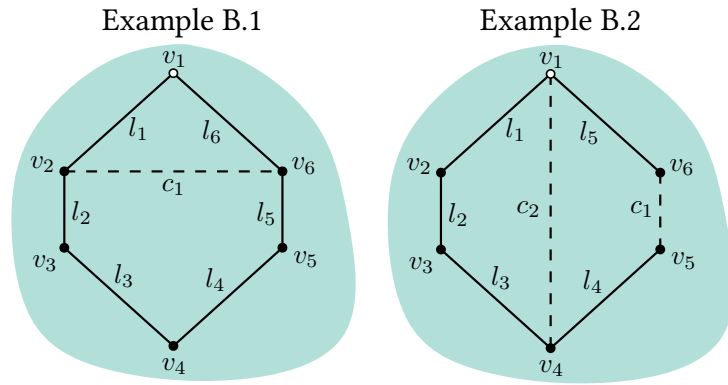
The cycle-based linear power flow equations have previously been applied to the related **OTS** [134]. However, using cycle-based power flow constraints in **TEP** has a distinct advantage over using it in **OTS**. Since usually in **TEP** problems existing transmission infrastructure cannot be removed, the **KVL** constraints from Equation (4.7) remain valid, regardless of the binary decision variables. Conversely, **OTS** needs to consider all simple cycles from the start because the initial network topology, and therefore the cycle basis, may not persist [134]. For **TEP** it is enough to append **KVL** constraints for supplemental candidate cycles according to Equation (4.27).

Candidate cycles can originate from (i) a candidate line parallel to an existing line, (ii) a candidate line connecting two buses which are already connected and are thereby part of the same synchronous zone, or (iii) multiple candidate lines connecting two or more synchronous zones which form cycles in the subnetwork graph \mathcal{G}_S .

Candidate Cycles Within Synchronous Zone

Finding candidate cycles within the same synchronous zone follows the subsequently described algorithm: For each candidate line $\ell \in \mathcal{L}^1$ connecting buses i and j find a shortest path \mathcal{P}_{ij}^{\min} through the network graph $\mathcal{G} = (\mathcal{N}, \mathcal{L}^0)$ with edge

Figure 4.4: Example Group B. Choice of candidate cycles within synchronous zone. Candidate lines denoted by c_i and existing lines denoted by l_i .



weights 1, which includes only the existing transmission infrastructure. The edges of the shortest path and the respective candidate line form a candidate cycle. The cycle incidence vector is formed according to equation Equation (4.5).

While any path through \mathcal{G} from i to j would yield a valid candidate cycle, it is computationally advantageous to minimise the size of the cycles to obtain sparser **KVL** constraints. For instance, in Example B.1 in Figure 4.4 the cycle for candidate line c_1 would consist of $\{c_1, l_1, l_6\}$ and not $\{c_1, l_5, l_4, l_3, l_2\}$. The potential **KVL** constraint would contain only three flow variables rather than five.

It is not required to add both cycles to the set of candidate cycles. If c_1 gets built, already one cycle in addition to the initial cycle basis ($\{l_1, l_2, l_3, l_4, l_5, l_6\}$) forms a cycle basis of the new network topology.

Furthermore, it is necessary to only consider existing lines and no other candidate lines for the shortest path search. Otherwise, a **KVL** constraint might be enforced only once a combination of candidate lines is built, although building one of the candidate lines alone would already introduce a new cycle. This is illustrated in Example B.2 in Figure 4.4. The cycles $\{c_1, l_5, c_2, l_4\}$ and $\{c_2, l_1, l_2, l_3\}$ would incur incorrect **KVL** constraints if c_1 is built but not c_2 . On the contrary, the longer cycles $\{c_2, l_1, l_2, l_3\}$ and $\{c_1, l_5, l_1, l_2, l_3, l_4\}$ obtained through the cycle search algorithm entail a correct modified cycle basis for either combination of investments.

Candidate Cycles Across Synchronous Zones

If two synchronous zones can only be synchronised by one particular candidate line (e.g. only c_1 in Example C in Figure 4.2), no new cycle has to be added. Then **KCL** alone already determines the power flow.

A new cycle must be introduced if two candidate lines connect to the same two synchronous zones. The cycle incidence vector is built from the two candidate lines and the existing lines on the shortest paths of \mathcal{G} through the synchronous

Figure 4.5: Example Group D. Choice of candidate cycles across synchronous zones and limits of the angle-based formulation. Candidate lines denoted by c_i and existing lines denoted by l_i .

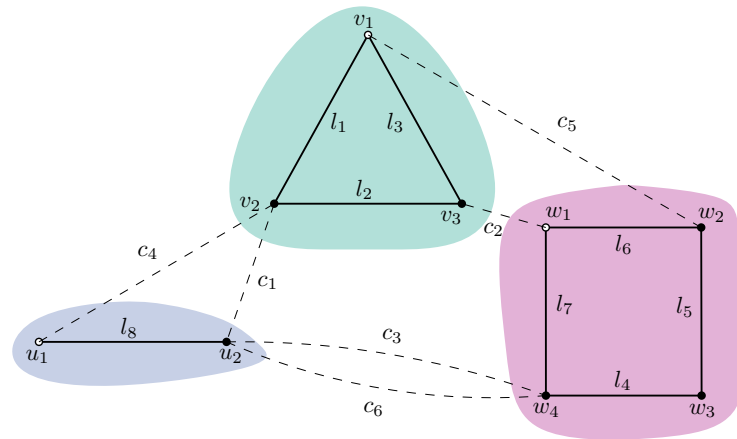
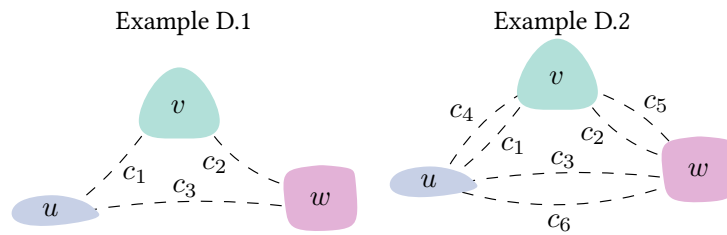


Figure 4.6: Example Group D as subnetwork graphs.



zones between the connection points, where edge weights are set to 1. In Example C in Figure 4.2, $\{c_1, l_3, c_2, l_7\}$ would form the according candidate cycle. Note, that also $\{c_1, l_2, l_3, c_2, l_6, l_5, l_4\}$ would be a correct candidate cycle, but the resulting conditional KVL constraint would be less sparse.

Additional cycles cannot only be incurred by the complementary investment of two candidate lines, but also from multiple candidate lines connecting three or more synchronous zones as depicted in Examples D.1 and D.2 in Figure 4.6. While Example D.1 has just one candidate cycle ($\{c_1, l_2, c_2, l_7, c_3\}$), Example D.2 with two candidate lines per pair of synchronous zones already has 11 candidate cycles to consider (3 cycles with two edges and 8 cycles with three edges). This is due to a growing number of interdependent combinations of investment decisions that would each demand conditional KVL constraints. Example D.2 creates a similar situation as in the OTS problem [134], where it becomes necessary to consider all simple cycles of the subnetwork graph \mathcal{G}_S (plus the corresponding shortest paths within the synchronous zones) a candidate cycle. Nonetheless, the network graph's initial cycle basis \mathcal{C} still remains intact.

Big-M Parameters for KVL Constraints

Having built the incidence matrix of the candidate cycles C_{lc}^1 , the subsequent step is to derive an appropriate big-M parameter M_c^{KVL} for each candidate cycle.

Theorem 2 *The value of the disjunctive constant M_c^{KVL} for a candidate cycle c can be chosen following*

$$M_c^{KVL} \geq \sum_{\ell \in \mathcal{L}^0 \cup \mathcal{L}^1} C_{\ell c}^1 x_\ell P_\ell \quad (4.28)$$

Proof 2 *Let c be a candidate cycle. If not all candidate lines of the candidate cycle are built, the corresponding cycle constraint must be inactive in all circumstances. In the case where n lines are not built Equation (4.27) becomes*

$$-nM_c^{KVL} \leq \sum_{\ell \in \mathcal{L}^0 \cup \mathcal{L}^1} C_{\ell c}^1 x_\ell p_\ell \leq nM_c^{KVL}. \quad (4.29)$$

Moreover, through Equation (4.1) the flow p_ℓ in lines $\ell \in \mathcal{L}^0 \cup \mathcal{L}^1$ is symmetrically limited by their nominal capacity P_ℓ . Hence,

$$\sum_{\ell \in \mathcal{L}^0 \cup \mathcal{L}^1} C_{\ell c}^1 x_\ell p_\ell \leq \sum_{\ell \in \mathcal{L}^0 \cup \mathcal{L}^1} C_{\ell c}^1 x_\ell P_\ell. \quad (4.30)$$

Constraint Equation (4.29) must be inactive even if an investment decision for only one candidate line is missing to close the candidate cycle. Therefore with $n = 1$ and the upper limit given in Equation (4.30), one obtains Equation (4.28).

Since there are no voltage angle variables and therefore no slack constraints in the cycle-based formulation, there is no need to calculate such big- M parameters. For calculating the voltage angles as outlined in Section 4.2.3, the slack buses can be chosen based on the resulting synchronous zones after the optimal investment decisions are known. This has the advantage over the angle-based formulation that matters of synchronisation do not have to be encoded into the optimisation problem.

4.4 Simulation Setup

We benchmark the presented transmission expansion planning formulations on multiple networks by again using the open European power transmission system model PyPSA-Eur [13] from Section 2.3 as a basis. The evaluation criteria are computational speed and peak memory consumption. The benchmark problems consider simultaneous generation and transmission capacity expansion each given a carbon budget of 40 Mt_{CO₂}, following the description of the long-term investment planning problem outlined in Section 2.1. Considered generation technologies include solar photovoltaics, onshore and offshore wind generators as well as open cycle gas turbines (OCGTs) and run-of-river power plants, but no storage units to maintain the independence of hourly snapshots and focus on transmission

expansion as balancing option for variable renewables. The necessary regionally resolved time series for electricity demand and renewable generator availability are included in the dataset. For candidate lines, we assume a standard line type for transmission lines at 380 kV with a capacity of approximately 1.7 GW [12].

To obtain a comprehensive sample of network topologies and operating conditions, we vary the number of clustered nodes in Europe {1000, 750, 500, 250}, the number of selected hours from a whole year {1, 5, 25, 50, 75}, the regional extract (see coloured areas in Figure 4.7), the tolerated MIP optimality gap {0.5%, 1%}, and the number of candidate lines per existing HVAC and HVDC corridor {1, 2}. In total, we evaluated 672 test problems. Synchronisation options are not considered in the benchmarks.

All formulations have been implemented for the power system analysis toolbox PyPSA [12]. Moreover, the optimisation problems are solved using the commercial solver Gurobi (version 9.0), given a time limit of 6 hours each. The code to reproduce the benchmarks is published at github.com/fneum/benchmark-teplop.

4.5 Results and Discussion

First, Figure 4.8 provides an initial insight into the problem sizes of the benchmark cases. The benchmark set covers a wide range of many smaller and some more complex problems. The largest involve up to 150,000 variables and 300,000 constraints and are the main target of speed improvements. The number of binary investment variables ranges from 34 to 612 candidate lines.

On average, using the cycle-based formulation reduces the total number of constraints to 95.3% compared to the angle-based formulation. Likewise, the average number of variables is reduced to 90.5%. As previously noted, this is due to the absence of the auxiliary voltage angle variables and fewer KVL constraints in the cycle-based formulation.

A share of 92% of all cycle-based problems and 82% of all angle-based problems were solved fastest using interior-point algorithms. Otherwise, dual simplex was quickest. To verify the formulations' objective values, while accounting for the fact that the MILPs only solve up to a predefined tolerance, we assert that the upper bound of one formulation is always larger than the other's lower bound. Across all instances, the total volume of transmission expansion ranges between 0% and 60% of the existing transmission network with up to 24 TWkm of additional network capacity. Due to the tolerances regarding the MIP gaps, both formulations can still yield slightly different transmission expansion plans.

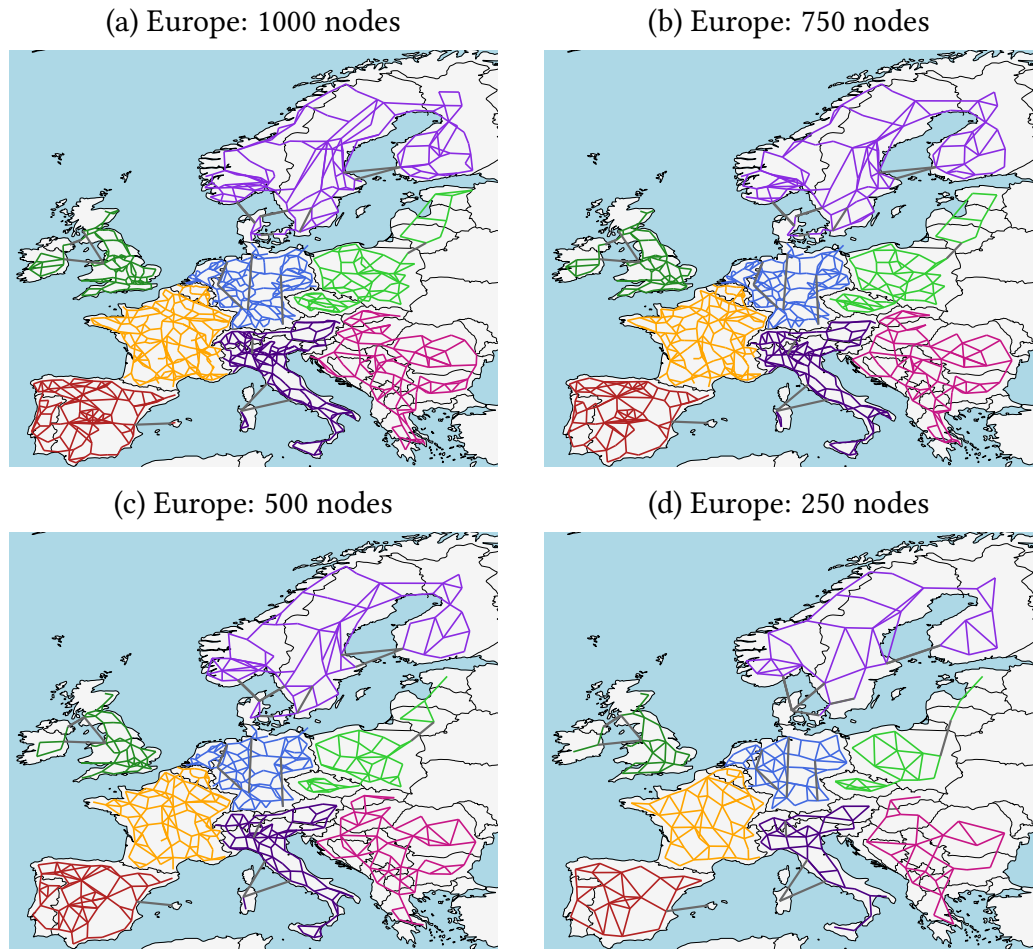


Figure 4.7: Clustered European transmission network models from which regional benchmark cases are formed. Each colour denotes a region for which an individual capacity planning problem is built. Coloured lines represent AC transmission lines at 380 kV, gray lines represent HVDC links.

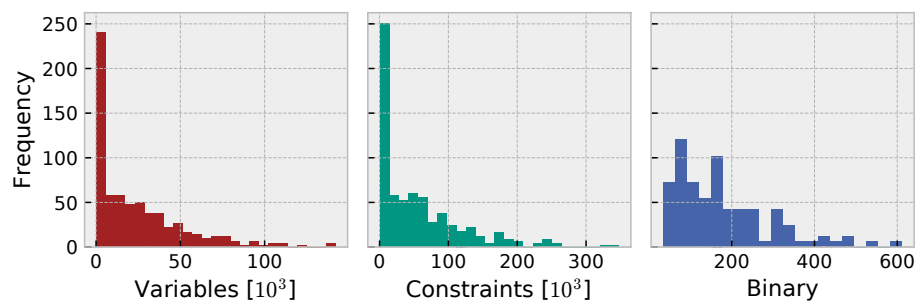


Figure 4.8: Histograms of the distribution of the number of variables, constraints and binary variables across the benchmark cases for angle-based planning problem.

Table 4.1: Numerical results of novel cycle-based and standard angle-based formulation. The speed-up factor is calculated by dividing the solving time of the angle-based formulation by the solving times of the cycle-based formulation.

	both ≤ 2 min	at least one > 2 min, excluding unsolved	one unsolved in walltime
instances	400	186	26
share faster	63.8%	87.6%	100%
speed-up factor (angle-based / cycle-based):			
– mean	1.28	3.94	3.12
– median	1.09	2.20	2.11
– maximum	9.40	31.25	14.70
– minimum	0.67	0.38	1.06

In terms of computation times, the competing formulations are contrasted in [Figure 4.9](#) and [Table 4.1](#). For individual benchmark cases, the relation between solving times is visible in [Figure 4.9](#). If a point is located on the identity line, both angle-based and cycle-based formulation took the same period of time to solve. If a point lies in the upper-left triangle, the cycle-based formulation solved faster, while a point in the lower-right triangle indicates that the angle-based formulation was quicker. Some instances have slightly exceeded the wall time of 6 hours due to system latency.

From [Figure 4.9](#) it becomes clear that the cycle-based formulation has a distinct advantage over the angle-based formulation in terms of computation times. In 60 out of the 672 instances both formulations did not satisfy the required MIP gap within the time limit. For the remaining instances, we distinguish the cases (i) both formulations solved in less than two minutes, (ii) at least one formulation took more than two minutes, but neither ran into the wall time, and (iii) exactly one formulation did not solve within the time limit. These categories are reflected in the summary of computational performance in [Table 4.1](#).

Instances of the most relevant group (ii), solve up to 31.25 times faster for particular cases, while averaging at a speed-up of factor 3.94 when using the cycle-based formulation instead of the angle-based variant. The median speed-up is 2.20. The angle-based formulation is outperformed in most (but not all) cases. Only in 12.4% of all cases, the angle based formulation was faster. For 26 instances, one formulation could not satisfy optimality tolerances within the time limit of 6 hours. In all such cases, the cycle-based formulation was solved, taking on average just 2 hours. The reduced computational advantage for small problems can partially be explained by the overhead that originates from determining the cycle basis and candidate cycles when building the problem.

The boxplots in [Figure 4.10](#) afford another interesting view on the solving times. They show the sensitivities of speed-up factors towards the spatial and temporal resolution of the network models. Besides many outliers, a trend towards a higher acceleration with larger networks is visible. Although acceleration tends

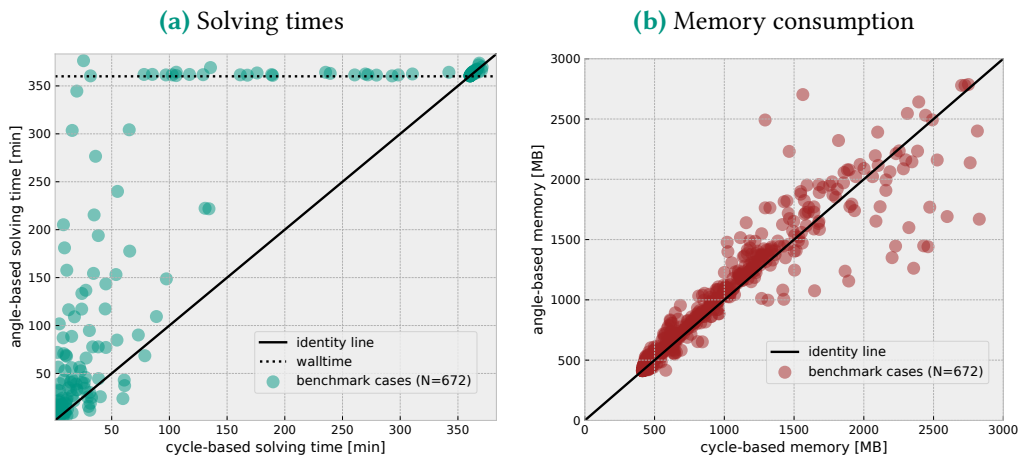


Figure 4.9: Solving times and memory consumption of cycle-based formulation versus solving times of angle-based formulation.

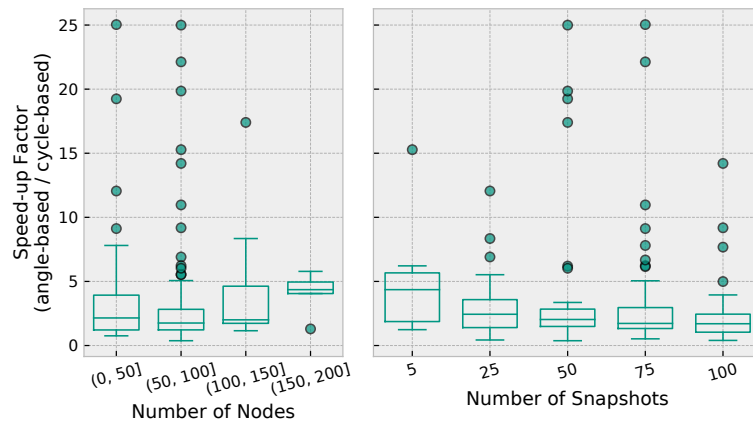


Figure 4.10: Sensitivities of speed-up distribution towards number of nodes and snapshots depicted as boxplots.

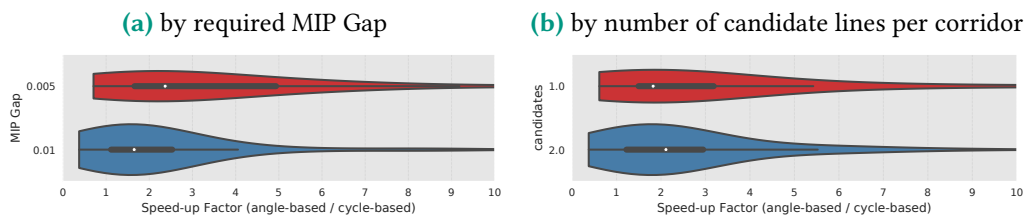


Figure 4.11: Sensitivities of speed-up factor distribution.

to decrease with a higher temporal resolution, the cycle-based formulation still outperforms the angle-based formulation considerably in most cases. [Figure 4.11](#) exhibits two further sensitivities. We found that a tighter MIP optimality gap further develops the advantage of the cycle-based formulation while considering slightly more candidate lines for each existing corridor tends to reduce its benefit. Contrary to computation times, as is shown in [Figure 4.9](#), there is no clear preference for either formulation in terms of peak memory consumption.

The fact that considerable speed-ups could be achieved already for problems with few snapshots makes the cycle-based reformulation suitable for combining it with Benders decomposition or related decomposition schemes. We did not apply any decomposition in this chapter because Benders decomposition's merits may be restricted to [TEP](#) problems where there are no complicating time-dependent constraints, e.g. from storage consistency equations or carbon budgets. Such intertemporal coupling would prohibit other essential acceleration techniques [[133](#)] but is pivotal to factor in the multitude of trade-offs in designing highly integrated renewable energy systems by co-optimisation.

4.6 Conclusion

The present chapter developed a novel cycle-based reformulation for the [TEP](#) problem with [LOPF](#) and compared it to the standard angle-based formulation. Instead of introducing many auxiliary voltage angle variables, the cycle-based formulation expresses Kirchhoff's voltage law directly in terms of the power flows, based on a cycle decomposition of the network graph. This results in fewer variables and sparser constraints. Moreover, the angle-based formulation has the disadvantage that it is not well-suited to considering the connection of multiple disconnected networks. The cycle-based formulation is shown to accommodate such synchronisation options conveniently. Since both formulations use the big- M disjunctive relaxation, helpful derivations for suitable big- M values were provided to avert numerical problems.

The competing formulations were benchmarked on 672 realistic generation and transmission expansion problems built from an open model of the European transmission system. For computationally challenging problems, the cycle-based formulation was shown to solve up to 31 times faster for particular cases, while averaging at a speed-up of factor 4. Hence, the cycle-based formulation is convincing not only because it can efficiently address synchronisation options, but also for its computational performance.

Transmission Expansion Planning Using Heuristics

Contents of this chapter based on

Neumann, F. & Brown, T. *Heuristics for Transmission Expansion Planning in Low-Carbon Energy System Models* in *16th International Conference on the European Energy Market* (2019), 1–8. doi:[10/d295](https://doi.org/10/d295). arXiv: [1907.10548](https://arxiv.org/abs/1907.10548). ©2019 IEEE. Reprinted, with permission.

5.1 Introduction

In [Chapter 4](#) we applied a big- M disjunctive relaxation to define the transmission expansion planning (TEP) problem as a mixed-integer linear problem (MILP) and demonstrated how one can nimbly reformulate this problem using insights from graph theory. And yet, with an increasing size of power networks, temporal resolution, and number of line expansion options, even solving this reformulation with state-of-the-art optimisation algorithms becomes prohibitively time-consuming.

In the present chapter, we therefore develop heuristics to incorporate integer transmission expansion and responsive line impedances into capacity expansion models, while retaining the lower computational effort of continuous linear problems (LPs) by applying sequential linear programming (SLP) techniques, relaxation, and post-facto discretisation. An LP, that jointly optimises generation and transmission capacities, is solved iteratively. In each iteration, the line impedances are adjusted to continuously expanded line capacities. Subsequently, the line extensions are fixed to their closest integral values followed by a final iteration of generation expansion. Such heuristics are of large avail when, for example, in a model with storage expansion the merits of frequently applied Benders decomposition cannot be leveraged because individual snapshots can no longer be decoupled [[133](#)].

SLP for joint optimisation of transmission and generation was introduced by Hagspiel *et al.* [[93](#)] but it was only tested against the exact solution for a 3-node

showcase example and did not feature benchmarks of different variants of the discretisation against the exact solution for realistic case studies.

In the present chapter, we remedy these gaps in the literature by evaluating the performance of different heuristic versions of SLP for a policy-relevant case study of the German transmission system. We compare the heuristics in terms of their speed-up in computation time, deviation from optimal total system cost, and similarity of line expansion. Furthermore, we show the error made by disregarding integrality constraints and responsive line impedances.

The remainder of the chapter is organised as follows: Section 5.2 guides through the mathematical formulation of the generation and transmission expansion planning problem and describes its standard solving algorithms, while Section 5.3 elaborates on heuristic algorithms for the problem at hand. In Section 5.4 we describe the case-study, the results of which are discussed in Section 5.5. The limitations are critically appraised in Section 5.6 before the chapter is concluded in Section 5.7.

5.2 Problem Description

Just as the previous thesis parts, the present chapter concerns a particular aspect of the long-term investment planning problem we first introduced in Section 2.1. One central component of the underlying linear optimal power flow (LOPF) constraints is Kirchhoff's voltage law (KVL)¹

$$p_\ell = b_\ell \vartheta_\ell, \quad (5.1)$$

which determines how power flows p_ℓ are distributed across the network's transmission lines given the line susceptance b_ℓ and the voltage angle difference ϑ_ℓ between the terminal buses of ℓ .²

It is a common assumption in capacity expansion models (CEMs) that line susceptances are invariant to line investments ($b_\ell = \tilde{b}_\ell$) and in some models line capacity is expanded continuously. On the other hand, in addition to acknowledging that the susceptance changes proportional to the capacity of a line ($x_\ell^{-1} = b_\ell \propto P_\ell$), TEP studies address the limited choice of circuit types. Instead of allowing the transmission capacities to expand continuously, they consider the number of added circuits

¹We illustrate the benefits of using heuristics for TEP with the angle-based KVL formulation known from Chapter 4. This choice is arbitrary and all considerations can straightforwardly be applied to the alternative cycle-based KVL formulation.

²For notational brevity we will omit the time index t in the following.

$\Phi_\ell \in \mathbb{N}$ to be within a set of candidate circuits $\mathcal{O}_\ell \subset \mathbb{N}_{\geq 0}$ for every extendable line $\ell \in \mathcal{L}_{\text{ext}}$.³ Then, the line capacity P_ℓ is linked to the choice of Φ_ℓ via

$$P_\ell = \left(1 + \frac{\Phi_\ell}{\tilde{\varphi}_\ell}\right) \tilde{P}_\ell, \quad (5.2)$$

where $\tilde{\varphi}_\ell \in \mathbb{N}$ denotes the initial number of circuits and \tilde{P}_ℓ is the initial line capacity. Amending Equation (5.1) with adaptive impedances and lumpy investment entails a set of nonlinear and non-convex constraints

$$p_\ell = \overbrace{\left(1 + \frac{\Phi_\ell}{\tilde{\varphi}_\ell}\right)}^{b_\ell} \tilde{b}_\ell \vartheta_\ell, \quad (5.3)$$

$$\Phi_\ell \in \mathcal{O}_\ell \subset \mathbb{N}_{\geq 0}, \quad (5.4)$$

where \tilde{b}_ℓ is the initial line susceptance.

With equations Equations (5.3) to (5.4) the optimisation problem classifies as a hard-to-solve mixed-integer nonlinear problem (MINLP) for which the present chapter seeks to find heuristic solutions.

5.2.1 Big- M Disjunctive Relaxation

However, as discussed in Chapter 4, Equations (5.3) to (5.4) can be reformulated as a MILP using the big- M disjunctive relaxation technique at the cost of numerous extra variables and constraints [24]. In shoret, the big- M reformulation uses one or multiple large constants in combination with binary variables to formulate disjunctive linear inequalities that redefine originally nonlinear constraints.

For the given TEP problem formulation, the integer variables $\Phi_\ell \in \mathcal{O}_\ell \subset \mathbb{N}_{\geq 0}$ are substituted with a binary variable $\Phi_{\ell,o} \in \mathbb{B}$ for each possible number of additional circuits $o \in \mathcal{O}_\ell$ and the constraint

$$\sum_{o \in \mathcal{O}_\ell} \Phi_{\ell,o} = 1 \quad (5.5)$$

³Note that this notation differs from the notation of Chapter 4. Here, the integer decision variable denotes the number of new circuits of identical type to be added to an existing line, whereas the previous chapter defined a binary decision variable for a specific line from a set of candidates.

is added to allow only one choice per line. We suppose that for all lines no capacity expansion is also a valid choice, i.e. $\forall \ell \in \mathcal{L}_{\text{ext}} : 0 \in \mathcal{O}_\ell$. The resulting disjunctive inequalities for all line investment options $o \in \mathcal{O}_\ell$

$$p_\ell - \left(1 + \frac{o}{\tilde{\varphi}_\ell}\right) \tilde{b}_\ell \vartheta_\ell \geq -\overline{M}_\ell^{\text{KVL}} (\Phi_{\ell,o} - 1) \quad (5.6)$$

$$\leq +\overline{M}_\ell^{\text{KVL}} (\Phi_{\ell,o} - 1) \quad (5.7)$$

are in effect equivalent to the equality constraint of Equation (5.3) if the big- M parameters $\overline{M}_\ell^{\text{KVL}}$ are suitably chosen. If $\Phi_{\ell,o}$ is 0, the inequalities formulated by Equations (5.6) to (5.7) are inactive, but if $\Phi_{\ell,o}$ is 1, the right-hand side is 0 and both merge to an equality constraint.

The choice of $\overline{M}_\ell^{\text{KVL}}$, however, is non-trivial. While the values must be chosen as large as necessary to guarantee that for a certain investment decision all other constraints linked to a different investment decision are inactive, they should also be as small as possible to avoid numerical instabilities. One approach for determining minimal values for $\overline{M}_\ell^{\text{KVL}}$ in the present formulation using shortest-path optimisation was outlined in the previous Chapter 4.

Adjusting Equation (5.2) to the binary nature of the line investment variables completes the reformulation:

$$P_\ell = \left(1 + \frac{\sum_{o \in \mathcal{O}_\ell} o \cdot \Phi_{\ell,o}}{\tilde{\varphi}_\ell}\right) \tilde{P}_\ell. \quad (5.8)$$

While applying the big- M formulation resolves the nonlinearities of Equation (5.3), the reformulation retains the integrality of line investment and therefore is still hard solve in reasonable time for large problem instances. More precisely, the reformulation considerably adds to the problem size in terms of the number of additional variables and constraints, where the cardinality of a set is denoted by $|\cdot|$:

$$\text{variables: } \sum_{\ell \in \mathcal{L}_{\text{ext}}} |\mathcal{O}_\ell| - |\mathcal{L}_{\text{ext}}| \quad (5.9)$$

$$\text{constraints: } |\mathcal{L}_{\text{ext}}| + |\mathcal{T}| \cdot \left(\sum_{\ell \in \mathcal{L}_{\text{ext}}} |\mathcal{O}_\ell| - |\mathcal{L}_{\text{ext}}| \right) \quad (5.10)$$

In this regard, the combinatorial difficulty of TEP has led researchers to focus only on promising line investments options. For instance, an automatic candidate selection scheme is discussed in [146].

Table 5.1: Solver settings for solving MILP in Gurobi.

Parameter	Value(s)
MIPGap	{0.015, 0.01, 0.005}
MIPFocus	2
Heuristics	0.5
Cuts	1
NumericFocus	3
TimeLimit	72 hours

5.2.2 Comments on Benders Decomposition

The computational challenges have led many researchers to apply decomposition techniques such as Benders decomposition (BD) to TEP [30, 131]. It is therefore worth briefly discussing its suitability for the problem at hand. BD divides the optimisation problem into an investment master problem with integer variables and an optimal power flow subproblem with exclusively continuous variables. By iteratively adding cuts (inequality constraints) to the master problem, derived from the duals of the subproblem given a specific set of investments, the operational expenses are approximated [131].

Despite its prevalence in the literature, there are concerns about the scalability of a straightforwardly applied BD algorithm for large-scale TEP problems that lead to investigations on acceleration techniques [147, 148]. Ramos & Lumberras [133] attained a speedup by around factor two by generating cuts based on relaxations and suboptimal solutions, removing inactive cuts, and, most significantly, by applying multicuts. The subproblem is split into individual snapshots and each of the resulting subproblems adds its own cut to the master problem. However, the merits of multicuts can only be leveraged if snapshots are independent of each other.

However, in the presence of storage consistency equations and renewable energy targets, there is a significant degree of intertemporal coupling. Additionally, capacity expansion planning that coordinate between generation, transmission and storage infrastructure investments, involve many more decision variables and, thereby, many more trade-offs than pure TEP studies. Hence, while the value of BD for detailed TEP studies remains unquestioned, for more general capacity expansion planning models working with heuristics and avoiding the NP-hardness of integer problems could prove more suitable.

In the present chapter, we therefore solve the big- M formulation without decomposition using a tuned parameter set (Table 5.1) for the Gurobi solver. In particular, the optimality gap (MIPGap) is a parameter which relates the lower and upper bounds of the objective value throughout the optimisation process and serves as termination condition.

Table 5.2: Experiment code glossary.

Code	Explanation
heur	Heuristic approach to solve MINLP
int	Integer line investment
iter	Iterating impedance updates until convergence
seqdisc	Sequential discretisation of impedances
postdisc	Post-facto discretisation with single threshold
postdisc-mult	Post-facto discretisation with multiple thresholds

5.3 Heuristic Approaches

In this section, we present the heuristics we propose to tackle the complexity of TEP problems, which comprise three central elements: relaxation of integer line investment (Section 5.3.1), iterative updates of line impedances (Section 5.3.2), and post-facto discretisation of line investment (Section 5.3.3). We label the different heuristic approaches presented above with abbreviations listed in Table 5.2.

5.3.1 Relaxation of Line Investment Variables

To avail of the computational merits of continuous LP, the integer investment decisions $\Phi_\ell \in \mathcal{O}_\ell \subset \mathbb{N}_{\geq 0}$ (tagged *heur-int*) may be relaxed to allow each line to be expanded continuously, i.e. $\Phi_\ell \in \mathbb{R}_{[0, \max\{\mathcal{O}_\ell\}]}$ (tagged *heur*).

5.3.2 Iterative Update of Line Impedances

To resolve the problem's nonlinearities, we draw on the concept of SLP. In SLP, constraints are iteratively linearised by a first-order Taylor series expansion around an operating point (e.g. the currently optimal line capacity) [149]. This way, instead of one complex nonlinear problem, multiple simpler LPs are solved consecutively. It can be shown that when SLP converges, the Karush-Kuhn-Tucker (KKT) conditions, the first-order necessary optimality conditions, are satisfied [150]. Additionally, move limits, which constrain the variable variation between two subsequent iterations to a linear trust region, are frequently supplemented but can be omitted if convergence is obtained without them [150].

For the given problem, Equation (5.3) is modified to

$$p_\ell = b_\ell^{(k)} \vartheta_\ell, \quad (5.11)$$

where in the first iteration the initial susceptances are used

$$b_\ell^{(1)} = \tilde{b}_\ell, \quad (5.12)$$

while for subsequent iterations $k+1$ the optimal line investment $\hat{\Phi}_\ell^{(k)}$ of the previous iteration k determines the physical line characteristics

$$b_\ell^{(k+1)} = \left(1 + \frac{\hat{\Phi}_\ell^{(k)}}{\tilde{\varphi}_\ell} \right) \tilde{b}_\ell \quad \forall k > 1. \quad (5.13)$$

Instead of adjusting the susceptances to values corresponding to fractional line capacities (tagged *iter*), another variant is to round any $\hat{\Phi}_\ell^{(k)} \in \mathbb{R}$ to their nearest integer value, referred to as sequential impedance discretisation (tagged *iter-seqdisc*).

The iteration loop terminates either if a pre-defined iteration limit is reached or if there is convergence; i.e. in two consecutive iterations there is no change in line investment or objective value and, therefore, also $b_\ell^{(k+1)} = b_\ell^{(k)}$. While Hagspiel *et al.* [93] only set a loose convergence tolerance for the objective function of 500,000 €/a was required, we tighten it to 1,000 €/a.

5.3.3 Post-facto Discretisation of Line Investment Variables

Since optimal line capacities are likely to be fractional when the continuous relaxation is applied, they do not represent a valid investment choice according to the original set of options \mathcal{O}_ℓ . Post-facto discretisation follows the iteration loop to produce a valid set of investment decisions.

The optimal capacities of continuously expanded lines are directly rounded to an integer value using a discretisation threshold ($z = 0.3$) on the fractional investment (tagged *postdisc*) [151]. The discretised line capacities are then fixed and line impedances are adapted accordingly for a final round of generation expansion only, to find the optimal mix of generator and HVDC link capacities given the discretised line capacities.

A more extensive post-facto discretisation procedure (tagged *postdisc_mult*) repeats the steps above for multiple discretisation thresholds ($z \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$) and chooses the configuration entailing the lowest total system costs. While this approach bears the chance of outcomes closer to the global optimum, it will naturally take longer to solve if thresholds are serially evaluated. However, this step lends itself to parallelisation to reduce computation times.

The post-facto discretisation step ensures that, ultimately, a feasible solution of the original problem (Section 5.2) is obtained.

5.4 Case Study of German Transmission System

The presented transmission expansion heuristics are evaluated on simplified models of the German transmission system. We choose Germany as our case study because it exhibits a distinct spatial mismatch of abundant wind resources in the North and high loads in the South, suggesting high future strains on the transmission network. Already in 2017, grid congestion causes curtailment of 5.52 TWh corresponding to 2.9% of renewable energy produced, and has been growing rapidly [152].

5.4.1 Model Inputs

Like in the previous chapter, we obtain the case study from the `PyPSA-Eur` model [13] of the European `ENTSO-E` transmission system we introduced in Section 2.3 and outline complementary considerations below. We allow simultaneous capacity expansion of transmission lines, `HVDC` links, and various types of generators: photovoltaic, onshore wind, and offshore wind generators with `AC` and `DC` grid connections, as well as open- and combined-cycle gas turbines. Existing run-of-river and biomass capacities are not extendable.

All transmission lines are assumed to be of the same type ‘Al/St 240/40 4-bundle 380.0’ with an Aluminium/Steel cross-section of 240/40 and a 4-bundle of wires per phase at 380 kV and allow for two additional parallel circuits fitted to the existing towers [153]. To avoid issues with social acceptance, a limit on the total additional volume of transmission capacities of 25% measured in TWkm is imposed. The corridors for `HVDC` links from the 2018 `TYNDP` may be expanded continuously with a capacity restriction of 8 GW each.

We neglect storage options (e.g. pumped-hydro, hydro dams, batteries, and power-to-hydrogen) to further simplify the case study and direct the focus to transmission expansion. Since balancing renewables in time through storage is an imperfect substitute for balancing renewables in space through transmission networks, ignoring storage options rather overestimates the cost-optimal grid reinforcements. A target share of renewable electricity of 70% is chosen since it is high enough to require transmission expansion and low enough to be viable without storage options.

To assess the scaling of the heuristic algorithms both in terms of spatial and temporal resolution, models with different levels of regional and time series aggregation are analyzed. The German power network is reduced to $n \in \{20, 40, 60, 80, 100\}$ nodes using the network clustering functionality of the power system analysis toolbox `PyPSA` [12], while the number of hourly snapshots is reduced to $t \in \{100, 200, 300, 400\}$ using the time series aggregation module `tsam` [26]. The cluster

weights represent the snapshot weightings and add up to 8760 hours to match the annuities of capital expenditures. Both reduced network size and temporal resolution are necessary limitations in accuracy to be able to solve the exact TEP formulations of Section 5.2 for up to 200 snapshots against which the heuristics of Section 5.3 are evaluated and may not accurately represent the necessary range of operating conditions to make qualified investment decisions.

5.4.2 Simulation Setup

The simulation setup comprises solving the case study with the ensemble of network sizes and snapshots outlined in Section 5.4.1 for the exact benchmark formulation that uses the big- M formulation described in Section 5.2 and the different variants of the proposed heuristic transmission expansion planning approaches named by combinations of the codes listed in Table 5.2 and described in Section 5.3. Note, that the approaches *heur* and *heur-iter* do not involve any discretisation and the former neither considers changing impedances. Therefore, they do not provide feasible solutions, but represent current practices in some CEMs. The solution to the big- M formulation comprises the lower and upper bounds that could be obtained for the lowest of the optimality gap tolerances of Table 5.1 within a walltime of 72 hours. It is further worth noting that the upper bound of the big- M formulation is the first feasible solution the solver could find that satisfies this tolerance.

5.5 Results and Discussion

We assess the heuristics in terms of (i) their deviation from optimal total annual system costs, (ii) the solving time, and (iii) the similarity of line expansion in relation to the solutions of the exact big- M reformulation. Figure 5.1 depicts these evaluation criteria for the network sizes outlined in Section 5.4.1 with 200 snapshots. The coloured markers represent feasible solutions to the exact MINLP, whereas black markers depict solutions of continuous relaxations of line investment and are generally not feasible MINLP solutions. The shaded area between the two vertical lines represents the range within which the optimal solution of the big- M formulation is contained, given the attained optimality gaps.

In terms of total system costs, all heuristics across all tested models are contained within a 1.5% relative cost increase compared to the lower bound of the big- M formulation obtained with the specified optimality gap tolerances. Since the optimal solution must be within the lower and upper bounds, the heuristics are therefore at most 1.5% more expensive than the optimal solution. Depending on

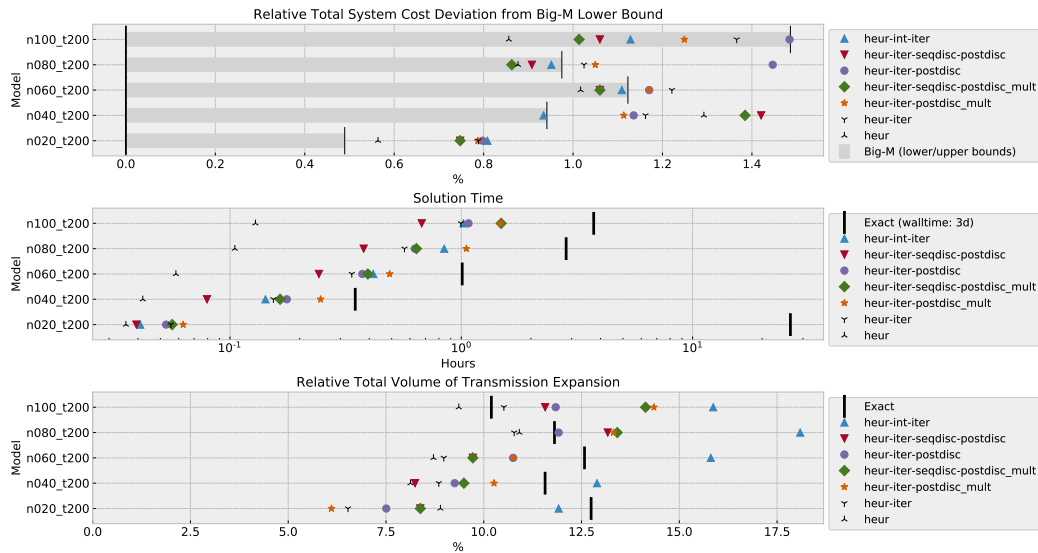


Figure 5.1: Comparison of (i) relative total system cost deviation from the lower objective bound of the big- M formulation, (ii) solving times, and (iii) relative volume of added transmission lines over the volume of original transmission lines (TWkm) across all presented heuristic approaches (as labelled in Table 5.2) and the upper objective bound of the big- M formulation (a feasible solution).

Table 5.3: Average number of iterations for heuristic approaches.

Algorithm	Average number of iterations
heur-int-iter	5.1
heur-iter-postdisc	10.0
heur-iter-postdisc-mult	10.0
heur-iter-seqdisc-postdisc(-mult)	4.4

the model, the spread among the heuristics lies between 0.1% and 0.5%. Another consistent observation is that the heuristics are close to the upper bound of the big- M formulation. In absolute numbers, total system costs roughly add up to around 32 billion €/a, of which around 2.5 billion €/a are attributed to transmission infrastructure. Thus, a relative cost deviation of around 1% corresponds to an absolute deviation of about 320 million €/a.

In terms of solving time, not iterating to update line impedances (*heur*) is naturally a fast but not MINLP-feasible approach. On the other hand, finding a solution to the big- M formulation consumes the most time. On average, the slowest heuristic is already faster by a factor of 2.2, but the approach *heur-iter-seqdisc-postdisc* is invariably the fastest among the MINLP-feasible heuristics with an average speed-up by factor 5.4. This is due to the sequential impedance discretisation (*iter-seqdisc*) which causes the iteration process to converge on average already after 4.4 iterations, while heuristics without sequential impedance discretisation consistently required at least the maximum number of 10 iterations (Table 5.3).

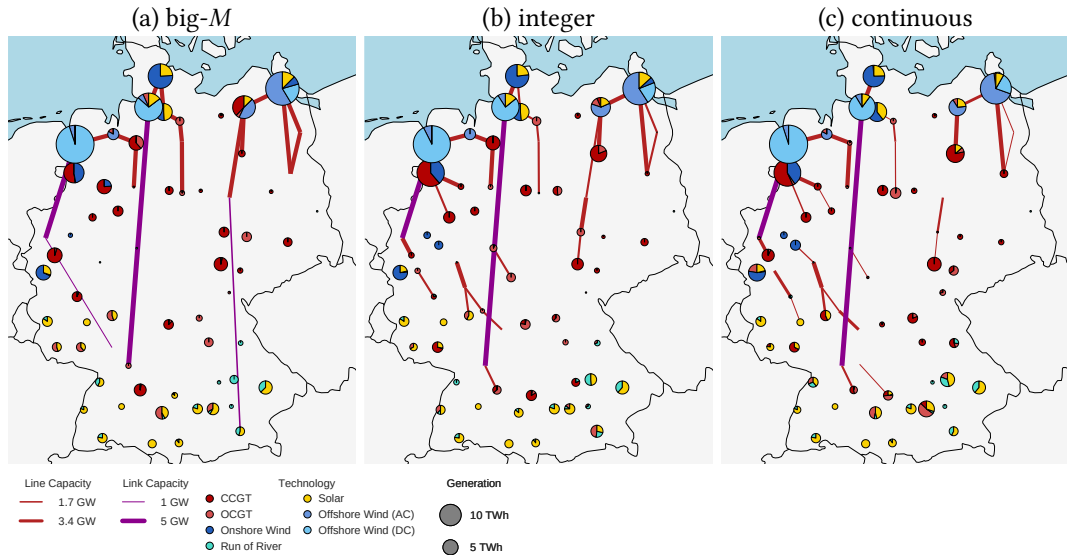


Figure 5.2: Map with transmission line expansion, HVDC link expansion and regional energy production for a model with 80 nodes and 200 snapshots obtained from (a) the exact big- M formulation with a 1% optimality gap, (b) an iterative heuristic approach with sequential impedance discretisation and post-facto line capacity discretisation, and (c) an iterative heuristic approach with continuous line expansion.

A wide spread among the approaches can be observed in terms of the total volume of transmission expansion, which we measure with a product of line capacity (MW) and length (km) and relate it to the original transmission network. The differences range between 5% and 7%. While *heur-int-iter* leads to the most line expansion for all models, it does not incur the highest total system costs. Approaches without discretisation (*heur* and *heur-iter*) can tailor line expansion more accurately to needs and, thus, entail less transmission expansion, but do not produce MINLP-feasible solutions.

Figure 5.2 offers a more visual approach to comparing the difference in transmission network expansion and power generation between the solutions of the big- M formulation, an iterative heuristic approach with sequential impedance discretisation and post-facto line capacity discretisation, and an iterative heuristic approach with continuous line expansion.

Across the three approaches, electricity production from renewables differs only marginally. Likewise, although the capacities and production from OCGT and CCGT power plants may interchange between neighbouring nodes, total volumes of gas-fired power plant capacities and production are very similar (Table 5.4). Thus, the regional generation mix is mostly insensitive to the applied heuristic.

Differences are more prevalent when looking at the expansion of HVDC links and HVAC transmission lines. While the common theme is a focus of line expansion

Table 5.4: Capacities and production of gas-fired powerplants for model and approaches of Figure 5.2.

		(i) big- M	(ii) integer	(iii) continuous
GW	CCGT	21.6	22.2	21.4
	OCGT	32.6	32.9	32.7
TWh	CCGT	96.9	96.1	96.8
	OCGT	42.1	42.9	42.2

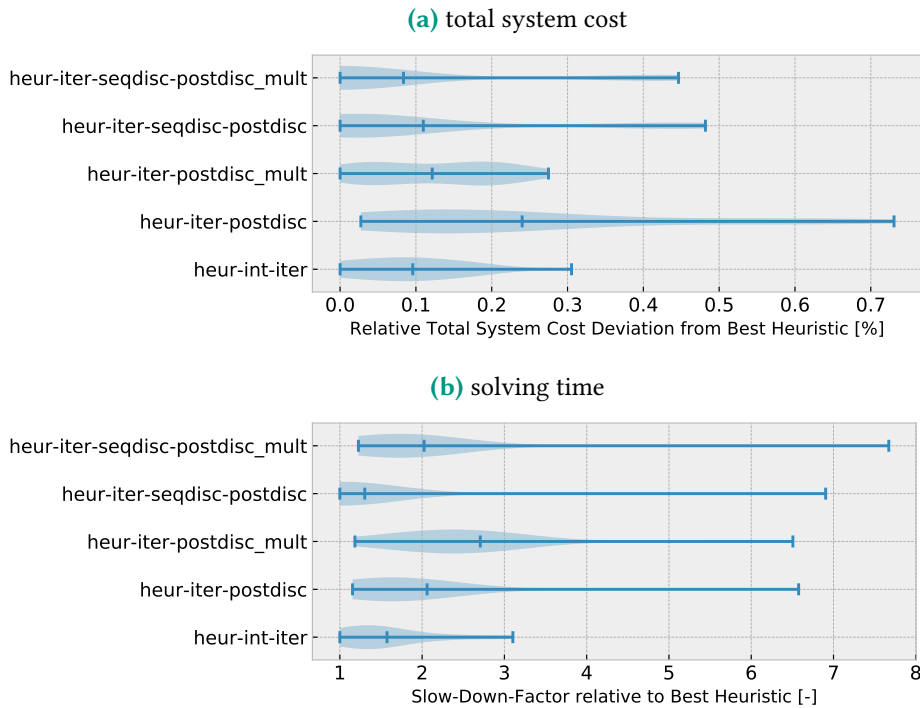


Figure 5.3: Comparison among heuristics in terms of relative total system cost and solving times. For each model, the objective values are related to the one heuristic approach with the lowest objective value or solving time respectively. The distribution of relative deviations from the best heuristic is depicted here.

in Northern Germany, the heuristics lead to an expansion of individual lines in Rhineland-Palatinate, Hesse and Saxony-Anhalt rather than building 2.5 GW on the route of the SuedOstLink (connecting the North-East of Germany with the South-East), which is the preferred choice of the big- M formulation. Both, big- M formulation and heuristics agree in utilising 8 GW HVDC links on the routes of SuedLink and A-North (connecting the North-South axis in the West of Germany).

Since the big- M formulation was only solved for up to 200 snapshots, Figure 5.3a and Figure 5.3b extend the comparison between heuristics to up to 400 snapshots and address the quality of approaches as well as their solving times. On average all heuristics perform similarly well and, neglecting rare outliers, differ only within a range of 0.3% in objective value. A glance at the solving times reiterates that

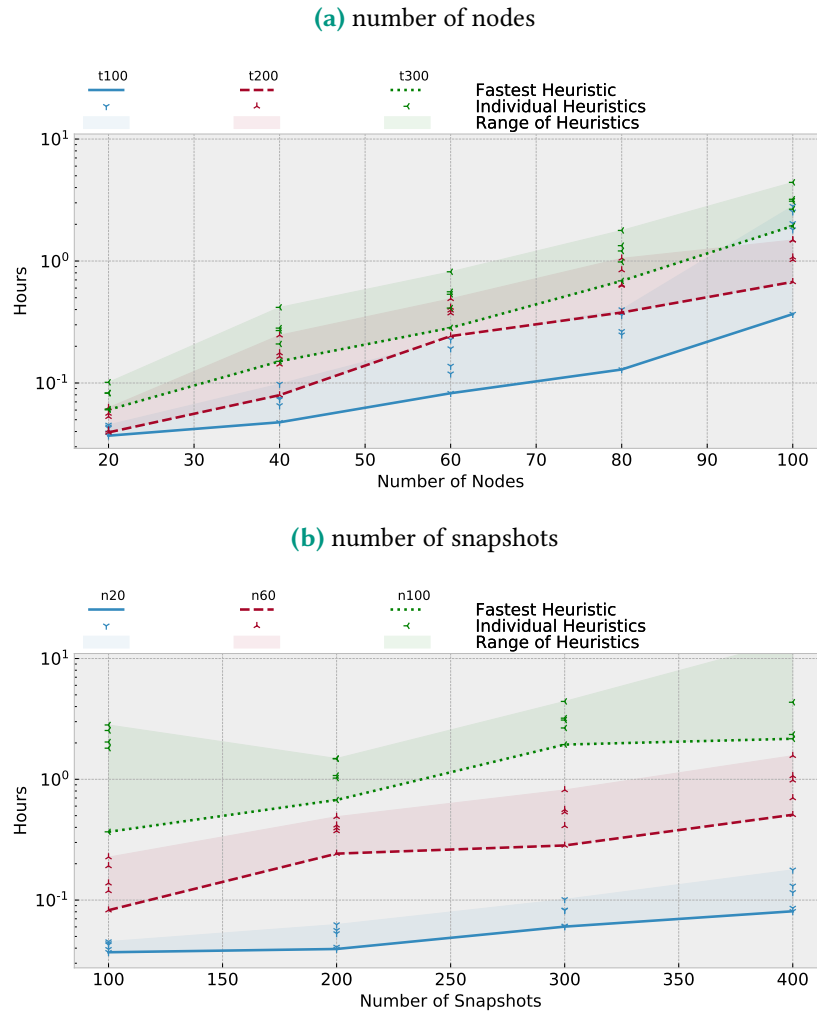


Figure 5.4: Scaling of solving time of heuristic approaches with with the number of nodes for models with 100, 200, and 300 snapshots and the number of snapshots for models with 20, 60, 100 nodes on a logarithmic scale.

sequential discretisation is an effective way of reducing the required number of iterations without a loss in solution quality. In contrast, evaluating multiple discretisation thresholds consumes additional time which does not benefit solution quality by much. Although the integer heuristic *heur-int-iter* produces good quality results in comparably little time, it is unlikely to scale well for much larger problems due to its integer, non-convex nature. Moreover, it is very sensitive to the choice of the optimality gap tolerance, which in this case was set to 0.5%. Additionally, solving times of integer problems can be very volatile. Such variance stems from randomness in the heuristics of most solvers (e.g. Gurobi). Instead, the heuristic *heur-iter-seqdisc-postdisc* seems to offer a sensible and well-scaling trade-off between solving time and finding low-cost solutions.

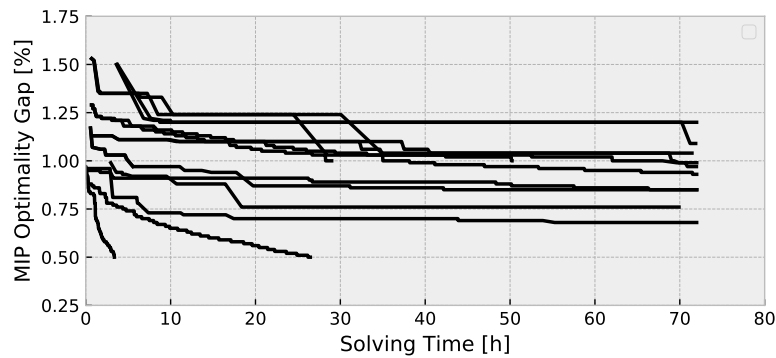


Figure 5.5: Progression of optimality gaps among all model optimisation runs which use the big- M reformulation. Lines start when the first feasible solution is found and end when either the optimality gap tolerance is satisfied or the walltime of the solver is reached.

To further investigate scaling, [Figure 5.4a](#) and [Figure 5.4b](#) depict how much longer the heuristic approaches take when the number of nodes or the number of snapshots is increased from two different perspectives. Both graphs show, that the sensitivity of solving times to model size is large and exponential, and that the heuristics seem to deal better with an increasing number of operating scenarios than with an increasing network size.

As previously noted, when solving the big- M formulation MILP, the upper bound is the best currently found feasible integer solution. [Figure 5.5](#) shows, that a feasible solution which lies within a 1.5% tolerance band can be found relatively quickly, but the progression from there becomes rapidly slower and may ultimately be terminated by the specified walltime. In consequence, the benefit of large runtimes is questionable. However, it is unclear whether feasible solutions with similarly low initial optimality gaps can also be found for a full European model and hourly time resolution for a full year. We could however observe, that the extent of allowed line expansion significantly influences solving time for the big- M formulation, and that the more ambitious the renewable target is chosen, the longer it takes to obtain an optimal solution.

5.6 Critical Appraisal

In the present chapter, the proximity of the objective functions for the heuristics are evaluated on relatively small networks and few snapshots, which in general would not provide a sufficient basis for reliable investment decisions. But, it allows a reasonable performance comparison of the heuristic approaches to exact formulations. Moreover, we disregarded storage options to avoid strong

temporal-coupling and furthermore ignored the currently existing fleet of fossil-fueled power plants, but limited those to gas-fired options. Neither do we consider sector-coupling, nor the transmission losses, unit commitment, or the provision of ancillary services. The case study also exclusively covers Germany, while questions of transmission expansion are in reality closely intertwined with considerations of other European countries.

Besides remedying these issues, another arising area of research regards the flatness of the solution space. The fact that the different heuristics provide solutions very close to the global optimum but differ in their individual investment decisions, indicates that many near-optimal solutions may exist which may be less susceptible to public or regulatory opposition. This constitutes the main research question of the following [Part II](#).

5.7 Conclusion

But first, a brief summary of the present chapter. Assuming a linear approximation of the power flow equations, we compared several heuristics for approximating the exact solution of joint transmission and generation expansion planning.

Particularly in light of the complexity already a limited choice of reinforcements entails, we conclude for models with high temporal and spatial resolution that using a continuous relaxation of line investment together with subsequent post-facto discretisation and applying sequential discretisation of line impedances to accelerate convergence is beneficial. This particular heuristic could reduce computation times of the joint optimisation by 82% with a maximal total system cost deviation of 1.5% for the models considered here.

Altogether, it was shown that the presented heuristics closely mirror optimal integer line investment of the more rigorous [MINLP](#) with considerable savings in solving time for policy-relevant low-carbon energy system optimisation models. With larger models that include storage and sector coupling it can be expected that only the heuristic methods will solve within reasonable time, and this contribution provides a reference to justify the application of such.

Summary Part I

In [Chapter 3](#), we demonstrated that, although the commonly used transport model can already identify key features of a cost-efficient renewable European electricity system while being computationally performant, deficiencies under high loading conditions arise due to the lack of a physical grid representation. Moreover, we showed that the disregard of transmission losses involves the danger of overestimating cost-optimal grid expansion by up to 20%.

Furthermore, we illustrated in the same [Chapter 3](#) that adding a convex relaxation of quadratic losses with no more than three tangents to the linearised power flow formulation while simultaneously accounting for changing line impedances as the network is reinforced suffices to represent power flows and losses adequately in design studies. We showed that the obtained investment and dispatch decisions are then sufficiently physical to be used in more detailed nonlinear simulations of AC power flow for evaluating their specific technical feasibility.

In the context of discrete transmission expansion planning (TEP), we showed in [Chapter 4](#) that a cycle-based reformulation could conveniently accommodate synchronisation options. Since both presented TEP formulations apply a big- M disjunctive relaxation, we moreover provided practical derivations for suitable big- M values to stabilise the numerics. Both competing formulations were also benchmarked on realistic generation and transmission expansion models derived from the European transmission system clustered to varying spatial and temporal resolutions, revealing that the proposed equivalent cycle-based formulation solved up to 31 times faster for particular cases than the typical angle-based formulation, while averaging at a speed-up of factor 4.

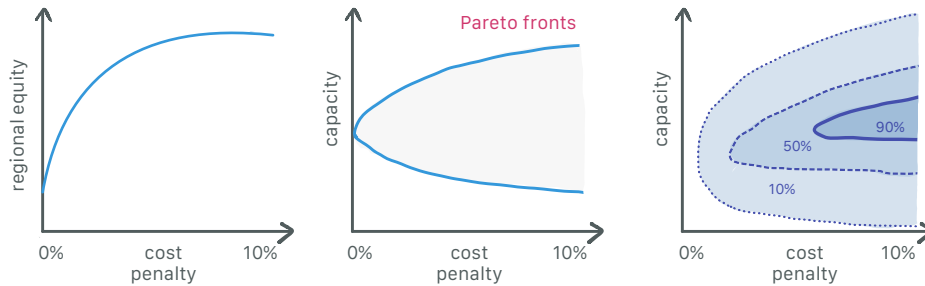
Finally, [Chapter 5](#) conveyed how using LP-based heuristics can reduce the computation times of joint generation and transmission capacity expansion by more than 80% while deviating at most by 1.5% from the original MILP formulation in terms of system costs. While the speed-up appears to be in the same order of magnitude as the cycle-based formulation of the MILP, aspiring to amplify model resolution and scope beyond the more constrained benchmark cases strongly shifts the preference towards the heuristic approaches because, with these, we can leverage weakly polynomial time algorithms. Overall, we found that the heuristics already closely mirror optimal integer line investment while exhibiting considerable time savings for studying policy-relevant scenarios of electricity systems with high shares of renewables.

Recomposed from abstracts of papers forming this part's chapters [1, 2, 5].

PART II

Cost-Optimality versus Near-Optimal Alternatives

Overview Part II



Frequently, energy system models studying the integration of renewables present just a single least-cost solution for a single set of cost assumptions. This constitutes a major weakness not only because it neglects uncertainties inherent to technology cost projections which capacity expansion models (CEMs) are susceptible to [154–156], but also because it hides a wide array of alternative solutions that are equally feasible and only marginally more expensive [3, 31, 157].

However, such trade-offs revealed by deviating from least-cost solutions would be extremely attractive for policymakers because they allow them to accommodate political and social dimensions that are otherwise hard to quantify without affecting the overall cost-effectiveness of the system [158]. This is particularly relevant considering rising public opposition towards the reinforcement of overhead transmission lines and the installation of onshore wind turbines in Europe.


Therefore, Chapter 6 systematically explores a selection of technology-oriented flat directions near the cost-optimum using a variant of modelling-to-generate-alternatives (MGA). The insights of this study are further strengthened in Chapter 7 by additionally taking account of uncertain technology cost projections. This analysis of near-optimal solutions for various cost scenarios is enabled by multi-fidelity surrogate modelling techniques with sparse polynomial chaos expansions (PCEs), low-discrepancy sampling, and heavy parallelisation on high-performance clusters.

Moreover, the spatial distribution of infrastructure plays a key role in shaping the levels of social acceptance, and concepts for nationally self-sufficient or even regionally autarkic distributed systems are recurrently debated. On the contrary, we observe that optimising for the least-cost renewable European electricity system leads to a very heterogeneous regional distribution of capacities. Chapter 8 is about investigating the cost sensitivity towards distributing capacities more evenly.

Recomposed from abstracts and introductions of papers forming this part's chapters [3, 4, 6].

Near-Optimal Solutions

Contents of this chapter based on

Neumann, F. & Brown, T. The Near-Optimal Feasible Space of a Renewable Power System Model. *Electric Power Systems Research* **190**, 106690. doi:[10/ghcpr2](https://doi.org/10/ghcpr2). arXiv: [1910.01891](https://arxiv.org/abs/1910.01891) (2021). Presented at *21st Power Systems Computation Conference 2020 (PSCC)*. 

6.1 Introduction

When planning cost-efficient future energy systems, providing just a singular optimal solution underplays an immense degree of freedom. It may well be that feasible but sub-optimal solutions are preferable for reasons not captured by model formulations [159], for instance because they rely less on building new overhead transmission lines or onshore wind turbines; two factors known to cause public opposition. Bypassing such issues to enable a swift decarbonisation of the energy system may justify a limited cost increase. Hence, presenting multiple alternative solutions and pointing out features that persist across many near-optimal solutions can remedy the lack of certainty in energy system models [55, 61]. Communicating these model results as a set of alternatives helps to identify *must-haves* (investment decisions common to all near-optimal solutions) and *must-avoids* (investment decisions not part of any near-optimal solution) [160]. In consequence, the resulting boundary conditions can then inform political debate and support consensus building.

A common technique for determining multiple near-optimal solutions is called modelling-to-generate-alternatives (*MGA*) which uses the optimal solution as an anchor point to explore the surrounding decision space for maximally different solutions [159]. Other methods, such as scenario analysis, global sensitivity analysis, Monte Carlo analysis and stochastic programming, that likewise address uncertainty in energy system modelling, concern *parametric* uncertainty, i.e. how investment choices change as cost assumptions are varied [61, 155, 161, 162]. Conversely, *MGA* explores investment flexibility for a single set of input parameters, by which it accounts for *structural* uncertainty. This can be simplifications of the model

equations, like linearising power flow and considering continuous transmission line expansion we previously examined in chapters [Chapter 3](#) and [Chapter 5](#), or degrees of freedom near the cost-optimum. In consequence, MGA is a complement rather than a substitute for methods sweeping across the parameter space, and we will engage in this aspect later in [Chapter 7](#).

Evidence from previous work suggests many technologically diverse solutions exist that result in similar total system costs for a sustainable European power system [37, 170]. These two studies research the sensitivity of cost input parameters or the relevance of transmission network expansion for low-cost power system layouts considering 30 regions.

Previous studies that applied MGA to long-term energy system planning problems or retrospective analyses are reviewed in [Table 6.1](#). The investigation presented in this chapter is the first to apply a variant of MGA to a European pan-continental electricity system model which includes an adequate number of regions and operating conditions to reflect the complex spatio-temporal patterns shaping cost-efficient investment strategies in a fully renewable system. Furthermore, the co-optimisation of generation, storage and transmission infrastructure subject to linear optimal power flow (LOPF) constraints is unique for MGA applications.

The goal of this chapter is to systematically explore the wide array of similarly costly but diverse technology mixes for the European power system, and derive a set of rules that must be satisfied to keep costs within pre-defined ranges. Additionally, we investigate how the extent of investment flexibility changes as we apply more ambitious greenhouse gas (GHG) emission reduction targets up to a complete decarbonisation and allow varying levels of relative cost increases.

The remainder of the chapter is structured as follows: [Section 6.2](#) guides through the problem formulation, the employed variant of MGA, sources of model input data, and the simulation setup. The results are presented and discussed from different perspectives in [Section 6.3](#) and critically appraised in [Section 6.4](#). The chapter is concluded in [Section 6.5](#).

6.2 Methods

6.2.1 Modelling-to-Generate-Alternatives (MGA)

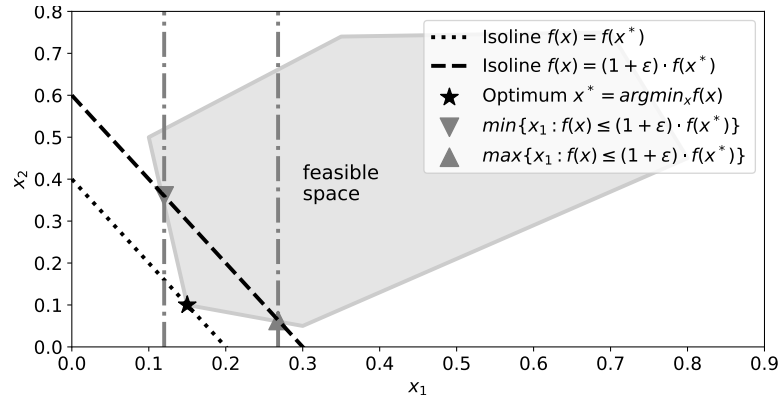
Solving the long-term power system planning problem from [Section 2.1](#) provides the anchor around which near-optimal alternatives are traversed. The optimisation seeks to minimise the total annual system costs comprising generation,

Table 6.1: Literature review of studies applying MGA to energy system models.*

Source	Main		Max. GHG				Near-optimal		
	Sector	Region	Nodes	Snapshots	Path	Reduction	Slack	Solutions	LOPF
Price & Keppo [163]	coupled	global	16	>1	yes	50%	<10%	30	no
DeCarolis [164]	electricity	USA	1	1	no	85%	<25%	9	no
DeCarolis <i>et al.</i> [159]	electricity	USA	1	1	yes	80%	<10%	28	no
Li & Trutnevyte [165]	electricity	UK	1	>1	yes	80%	<15%	800	no
Sasse & Trutnevyte [166]	electricity	Swiss	2,258	1	no	none	<20%	2,000	no
Trutnevyte [64]	electricity	UK	1	3	no	none	<23%	250,500	no
Berntsen & Trutnevyte [167]	electricity	Swiss	1	386	no	none	–	520	no
Nacken <i>et al.</i> [168]	coupled	Germany	1	8,760	no	95%	<10%	1,025	no
Hennen <i>et al.</i> [160]	urban	generic	1	>1	no	none	<10%	384	no
Pedersen <i>et al.</i> [169]	electricity	Europe	30	8,760	no	95%	<30%	–	no
Lombardi <i>et al.</i> [157]	electricity	Italy	20	8,760	no	100%	<20%	4,806	no
Sasse & Trutnevyte [31]	electricity	Central Europe	650/100	1/8,760	no	various	<20%	100	ex-post
The present chapter	electricity	Europe	100	4,380	no	100%	<10%	1,968	yes

* New works by Pedersen *et al.* [169], Sasse & Trutnevyte [31], Lombardi *et al.* [157] were amended retroactively.

Figure 6.1: Illustration of the near-optimal feasible space and the applied MGA method for a two-dimensional problem for the search-directions relating to dimension x_1 .



transmission and storage infrastructure in a fully renewable system subject to multi-period LOPF constraints, resulting in a linear problem (LP).

As illustrated in Figure 6.1, following a run of the original model, the objective function $f(G, H, F, g)$ is encoded as a constraint such that the original feasible space is limited by the optimal objective value $f(G^*, H^*, F^*, g^*)$ plus some acceptable relative cost increase ϵ ,

$$f(G, H, F, g) \leq (1 + \epsilon) \cdot f(G^*, H^*, F^*, g^*). \quad (6.1)$$

Other than preceding studies, the present chapter pursues a more structured approach to MGA to span the near-optimal feasible space, which resembles the ϵ -constraint method from multi-objective optimisation (MOO). The search directions are not determined by the hop-skip-jump (HSJ) algorithm that seeks to minimise the weighted sum of variables of previous solutions [155], but by pre-defined groups of investment variables. Consequently, the new objective function becomes to variously minimise and maximise sums of subsets of generation, storage and transmission capacity expansion variables given the ϵ -cost constraint. The groups can be formed by region and by technology. Examples for thought-provoking search directions are to minimise the sum of onshore wind capacity in Germany or the total volume of transmission expansion (see Section 6.2.3).

This process yields boundaries within which all near-optimal solutions are contained and can be interpreted as a set of rules that must be followed to become nearly cost-optimal. In fact, by arguments of convexity, it can be shown that near-optimal solutions exist for all values of a group's total capacity between their corresponding minima and maxima. The original problem is convex as it classifies as a linear program. Neither adding the linear ϵ -cost constraint nor introducing an auxiliary variable z that represents the sum of the group of variables alter this characteristic. Hence, for any total $z \in [z_{\min}, z_{\max}]$ a near-optimal solution exists, however not for any combination of its composites leading to this total.

6.2.2 Model Inputs

As has frequently been the case in previous parts of the thesis, the exploration of the near-optimal feasible space is executed for the open model dataset [PyPSA-Eur](#) [13] of the European power system which was introduced in [Section 2.3](#). Here, we consider a spatial resolution of 100 nodes and a temporal resolution of 4380 snapshots (two-hourly for a full year) [13]. The chosen levels of geographical and temporal aggregation reflect, at the upper end, the computational limits to calculate a large ensemble of near-optimal solutions and, at the lower end, the minimal requirements to expose transmission bottlenecks and account for spatially and temporally varying renewable potentials with passable detail [28, 29, 171].

Following a greenfield approach (with the exception of the transmission grid and hydro-electric installations), we allow simultaneous capacity expansion of transmission lines, [HVDC](#) links and multiple kinds of storage units and generators: solar photovoltaics, onshore wind turbines, offshore wind turbines with [AC](#) or [DC](#) grid connections, battery storage, hydrogen storage and, ultimately, open- and combined cycle gas turbines ([OCGT/CCGT](#)) as sole fossil-fueled plants. As in previous chapters, run-of-river and pumped-hydro capacities are not extendable due to assumed geographical constraints.

The corridors for new [HVDC](#) links from the 2018 [TYNDP](#) are limited to 30 GW [85]. Individual [AC](#) transmission line capacities may be expanded continuously up to four times their current capacity, but not reduced. Given the densely meshed and spatially aggregated transmission system, we do not add new expansion corridors but constrain options to reinforcement via parallel [AC](#) lines. The dependence of line capacity expansion on line impedance is addressed by the sequential linear programming ([SLP](#)) approach presented in [Chapter 5](#) [2].

6.2.3 Simulation Setup

The near-optimal analysis is run within a parallelised workflow for different deviations $\epsilon \in \{0.5\%, 1\%, 2\%, 3\%, 4\%, 5\%, 7.5\%, 10\%\}$ from the cost-optimal solution and for system-wide greenhouse-gas emission reduction targets of 80%, 95% and 100% compared to emission levels in 1990. This allows to follow the propagation of investment flexibility for increasing optimality tolerances and more ambitious climate protection plans. The alternative objectives are to variously minimise and maximise the generation capacity of all (i) wind turbines, (ii) onshore wind turbines, (iii) offshore wind turbines, (iv) solar panels, and (v) natural gas turbines. Moreover, we search for the minimal and maximal deployment of (vi) hydrogen storage, (vii) battery storage, and (viii) power transmission infrastructure. This setup yields

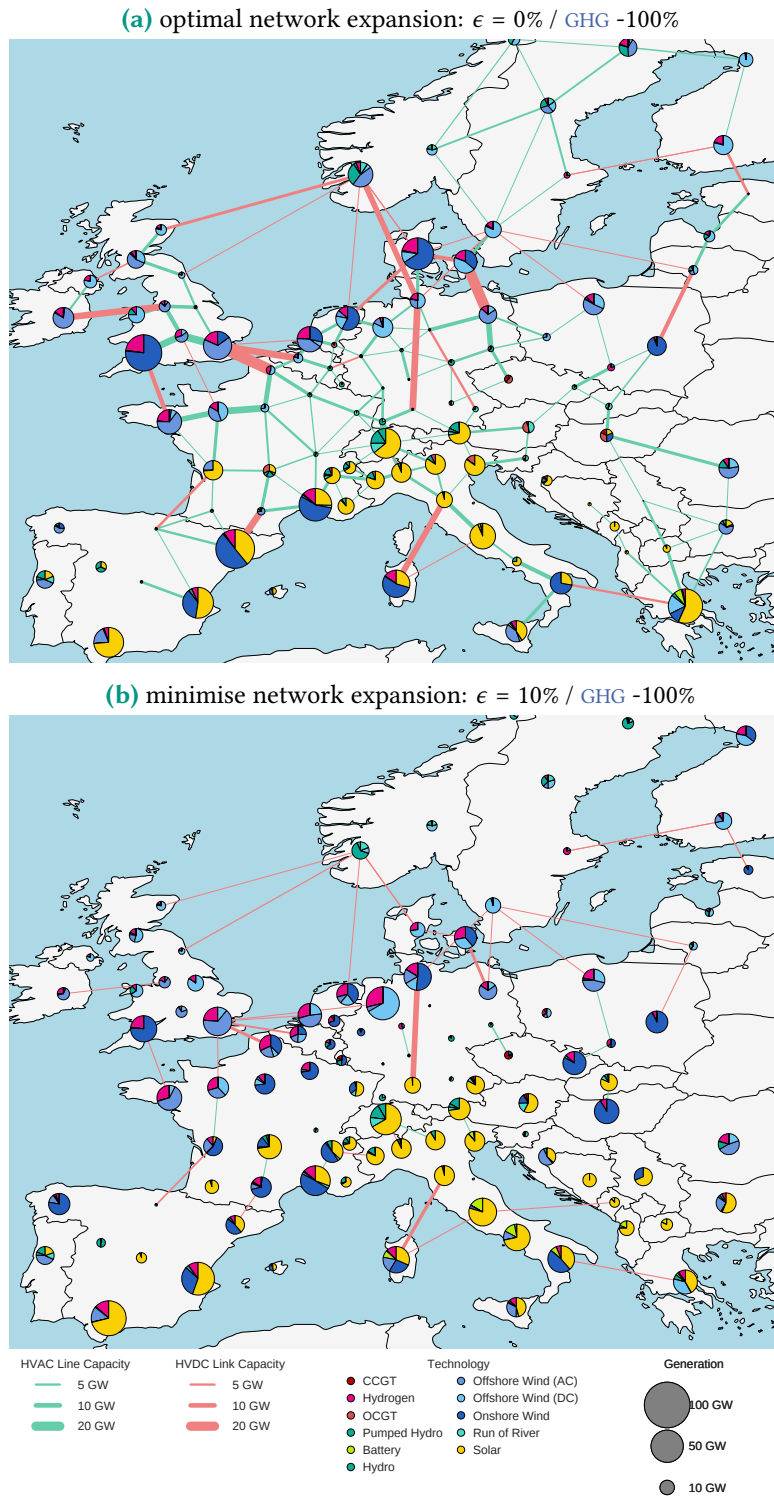


Figure 6.2: Maps of transmission line expansion and regional generator and storage capacities for a 100% renewable power system for the optimal solution and minimal transmission volume within a 10% cost increase.

Table 6.2: Statistics on optimal solutions for different GHG emission reduction levels.

GHG Emissions	-100%	-95%	-80%
Generation [TWh]			
– Onshore Wind	750 (24%)	421 (14%)	423 (15%)
– Offshore Wind (AC)	886 (28%)	568 (19%)	297 (10%)
– Offshore Wind (DC)	873 (27%)	1032 (34%)	769 (27%)
– Solar	502 (16%)	605 (20%)	381 (13%)
– Run of River	150 (5%)	153 (6%)	154 (5%)
– CCGT (gas)	0 (0%)	171 (6%)	761 (26%)
– OCGT (gas)	0 (0%)	41 (1%)	105 (4%)
Transmission [TWkm]	504 (+71%)	458 (+55%)	368 (+25%)
Load [TWh]	3,138	3,138	3,138
Total Cost [bn €/a]	246	207	165
Total Cost [€/MWh]	78.4	66.1	52.6

384 near-optimal solutions. The solutions delineate so-called Pareto frontiers where no criterion, neither reducing system cost nor extremising the capacity of a technology, can be improved without depressing the other. On average, each problem required 6.5 hours and 31 GB of memory to solve with the Gurobi solver.

For slacks $\epsilon \in \{1\%, 5\%, 10\%\}$ and a 95% emission reduction target a 3-hourly resolved model is run for country-wise minima and maxima of the investment groups above, resulting in additional 1584 near-optimal solutions. On average, each problem required 3.5 hours and 22 GB of memory to solve. The code used for the experiments is available at github.com/pypsa/pypsa-eur-mga.

6.3 Results and Discussion

6.3.1 Cost-Optimal Solutions

Before delving into near-optimal solutions, we first outline the characteristics of the optimal solutions for different emission reduction levels (see Table 6.2). A system optimised for a 100% emission reduction is strongly dominated by wind energy. More than half of the electricity is supplied by offshore wind installations. Onshore wind turbines provide another quarter. In contrast, photovoltaics account for only 16% of electricity generation. Strikingly, a system targeting a 95% reduction in greenhouse gases uses significantly less onshore wind generators but more solar energy in comparison to a completely decarbonised system, while keeping the share of offshore wind generation constant. Thus, for the last mile from 95% to 100% more onshore wind generation is preferred to phase out the last remaining natural-gas-fired power plants. The total system costs scale nonlinearly with more tight emission caps. Achieving an emission reduction

of 95% is roughly a quarter more expensive than a reduction by 80%, while a zero-emission system is almost 50% more expensive.

Also, the map in [Figure 6.2a](#) shows the optimal regional distribution of the capacities of power system components for a fully renewable European power system. Generation hubs tend to form along the coasts of North, Baltic and Mediterranean Sea, whereas inland regions produce little electricity. Expectedly, solar energy is the dominant carrier in the South, while wind energy prevails close to the coasts of the North Sea and the Baltic Sea. Most grid expansion can be found in Germany, France and the United Kingdom and individual HVDC links are built with capacities of up to 30 GW. The routes and capacities of HVDC links are well correlated with the placement of wind farms.

6.3.2 Nearly Cost-Optimal Solutions

In this section we extremise different groups of investments in generation, storage and transmission infrastructure. As an example, [Figure 6.2b](#) depicts a system that seeks to deviate from the optimum by minimising the volume of transmission network expansion up to a total cost increase of 10%, for instance as a concession to better social acceptance. With the results particularly the NordSued link connecting Northern and Southern Germany manifests as a no-regret investment decision up to a capacity of 15 GW in the context of full decarbonisation. It is one of the few persistent expansion routes. All other transmission expansion corridors are to a significant extent not compulsory. Missing transmission capacities can be compensated by adding storage capacity and more regionally dispersed power generation. Nevertheless, some transmission network reinforcement is indispensable to remain within the given cost bounds. These results are also broadly aligned with findings in [28, 37].

Beyond this example, the results offer insights about the structure of the near-optimal space. The intent is to portray a set of technology-specific rules that must be satisfied to keep costs within pre-defined ranges ϵ . Note, that the discontinuity created by ϵ restricts the accuracy of the solution space.

[Figure 6.3](#) reveals that wind generation, either onshore or offshore, is essential to set up a cost-efficient European power system for all three evaluated emission reduction levels. Whilst already a small cost increase of 0.5% yields investment flexibilities in the range of ± 100 GW, even a 10% more costly solution would still require more than 500 GW of wind generation capacity for a fully renewable system: two-thirds of the optimal system layout. However, even for a zero-emission system a cost increase of just 4% enables abstaining from onshore wind power, and a 7.5% more expensive alternative can function without offshore wind farms.

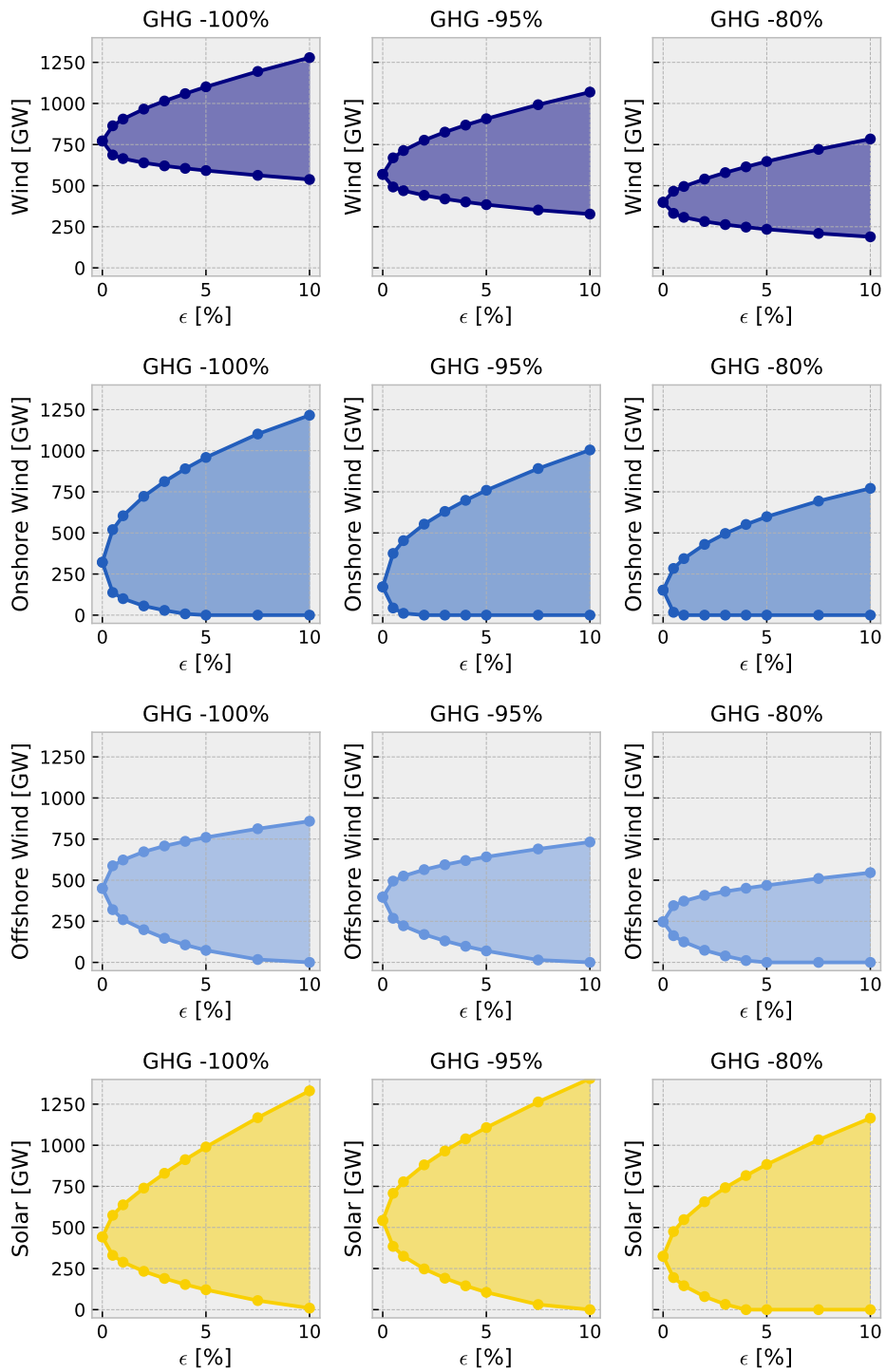


Figure 6.3: Solution space of renewable generation infrastructure by technology for different levels of slack ϵ and emission reduction targets.

The investment flexibility develops nonlinearly with increasing slack levels ϵ . Even a minor deviation from the cost optimum by 0.5% creates room for maneuver in the range of ± 200 GW for onshore and ± 150 GW offshore wind installations, which indicates a weak trade-off between onshore and offshore wind capacities very close to the optimum. Nonetheless, dispensing with both is not viable. Furthermore, 10% of total system costs must be spent to rule out solar panels, but already a slack of 1% allows to reduce the solar capacity by a third.

Price & Keppo [163] observed in their model that investment flexibility in generation infrastructure decreased as more tight caps on carbon-dioxide emissions were imposed. While it is true that more ambitious climate protection plans incur more must-haves (i.e. minimum requirements of capacity), for the case-study at hand the viable ranges of marginally inferior solutions increase as more total capacity is built.

As Figure 6.4 exhibits, even for a complete decarbonisation of the European power system building battery storage is not essential, although they are deployed in response to e.g. minimising network reinforcement. Conversely, once weather-independent dispatch flexibility from natural-gas-fired power plants is unavailable, it becomes imperative to counter-balance with hydrogen storage. The cutback of hydrogen infrastructure under these circumstances goes along with building excess generation capacities and multiplied amounts of curtailment.

The reinforcement of the transmission network becomes more pivotal the more the power system is based on renewables. Aiming for an emission reduction by 80% a 2% more expensive variant can get by without any grid reinforcement. Reducing emissions by 100% still requires some additional power transmission capacity at a 10% cost deviation. However, within this range, the transmission volume can deviate from almost double of today's network capacities to merely a marginal reinforcement (see Figure 6.2).

6.3.3 Correlations and Interactions

So far, the study of the near-optimal feasible space did not capture the interdependence between different system components apart from the envelopes the analysis provided for each technology. Shifting to the extremes of one technology will diminish the investment flexibility of other carriers. Figure 6.5 demonstrates how diversity in capacity mixes rises if more leeway is given in terms of system costs. The striking variety in capacity totals is largely attributable to the lower capacity factors of solar compared to wind energy.

Hennen *et al.* [160] suggested to present the intertwining of technologies through Pearson's correlation coefficient across all near-optimal solutions. Figure 6.6 con-

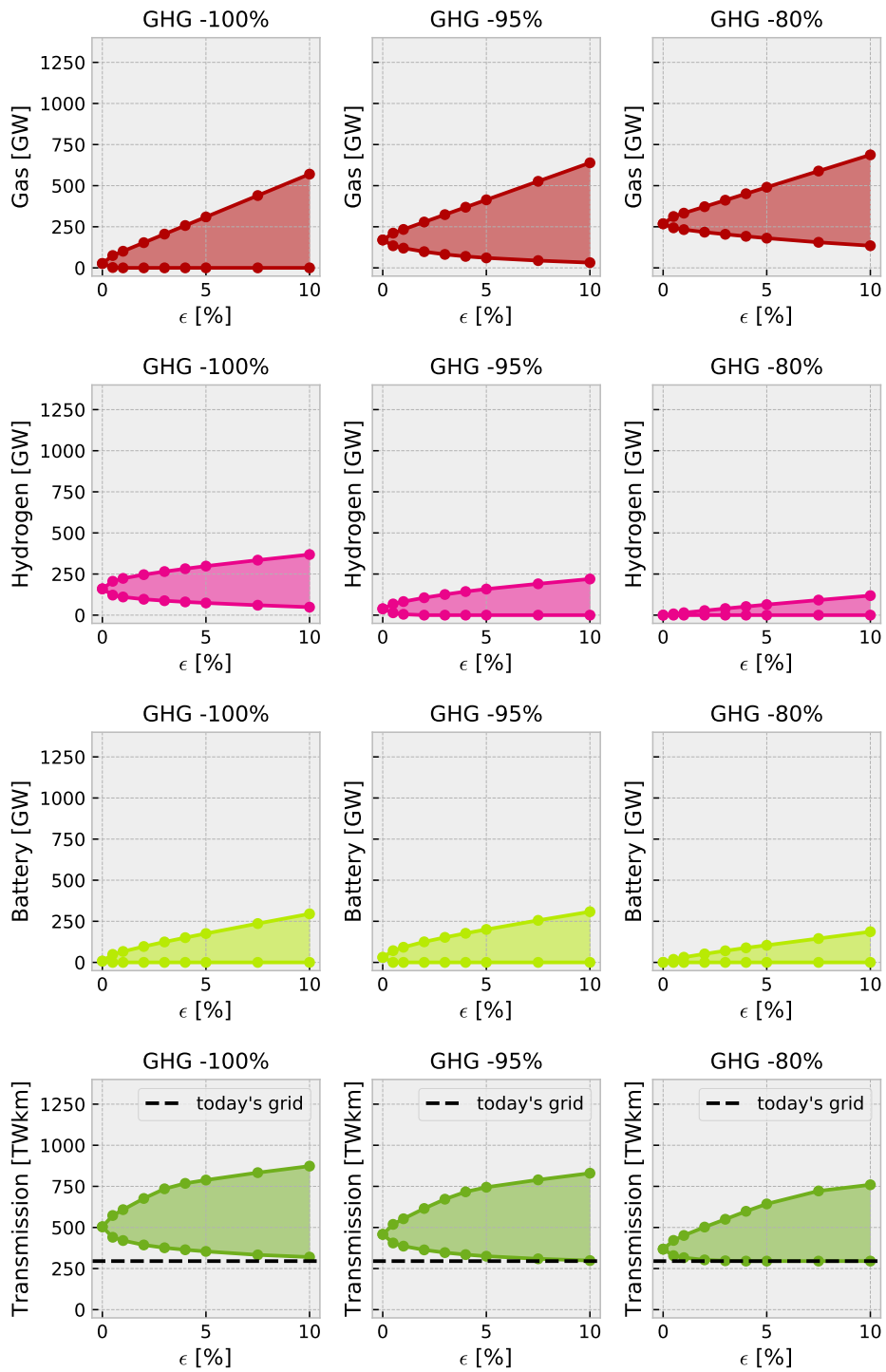


Figure 6.4: Solution space of storage, transmission and backup infrastructure by technology for different levels of slack ϵ and emission reduction targets.

Figure 6.5: Composition of generation and storage capacities for various near-optimal solutions with 100% renewables. Each subplot corresponds to a slack level ϵ and an optimisation sense. The labels of the bar charts indicate which group of investment variables is included in the objective.

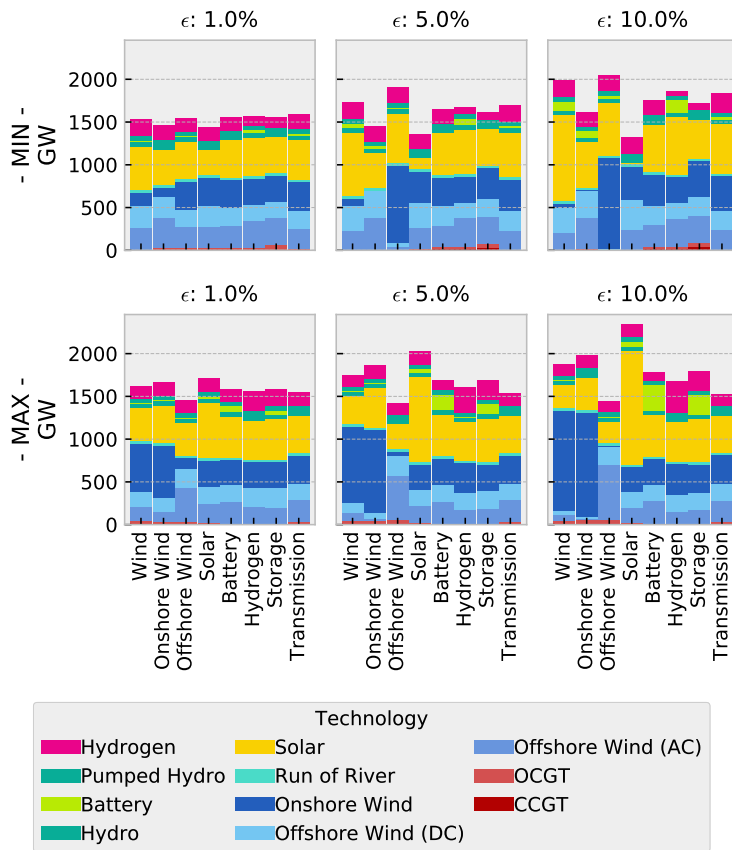
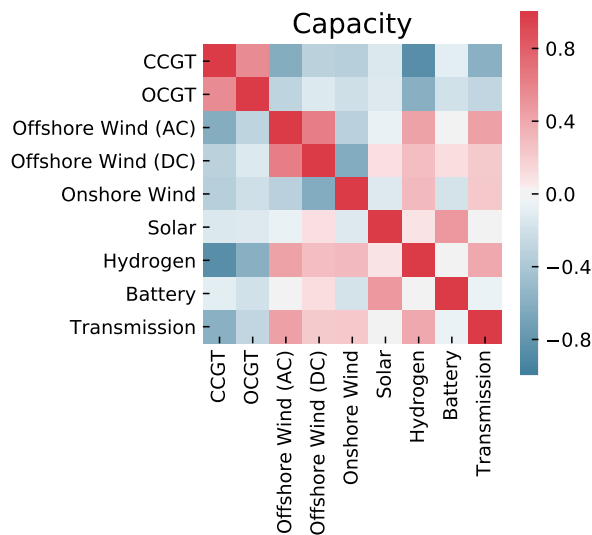


Figure 6.6: Correlations of capacities across all near-optimal solutions.



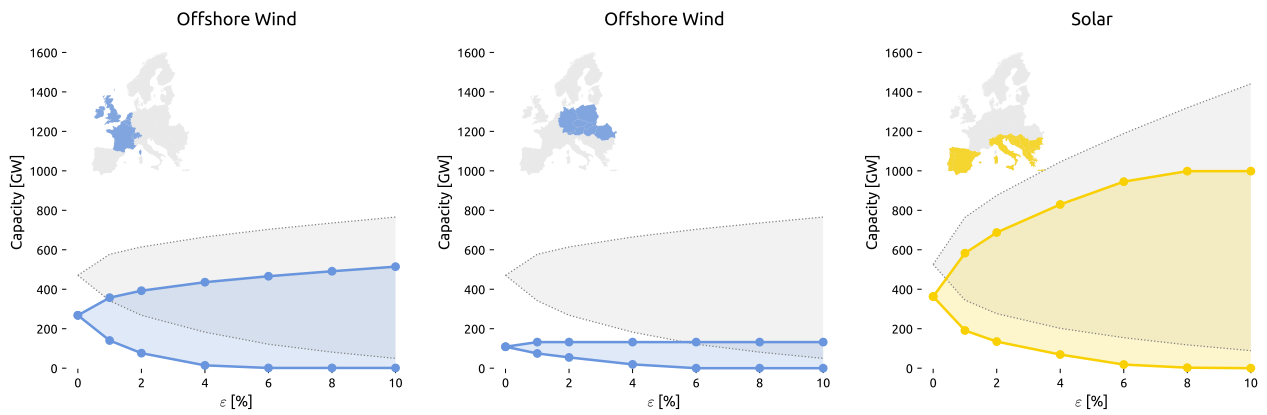


Figure 6.7: Near-optimal feasible space for specific regions of the European transmission system. It shows the three pairs of regions and technologies which have a minimum requirement beyond 1% of the cost-optimum. While the coloured areas depict the near-optimal space of the marked region, the grey shaded area portrays the system-wide near-optimal space.

firmly many of the previously noted connections. Hydrogen storage substitutes natural gas turbines and is positively correlated with onshore and offshore wind capacity, while battery deployment rather matches with solar installations. Likewise, transmission expansion occurs in unison with onshore and offshore wind deployment. Thereby, hydrogen storage and transmission become complements for high renewable energy scenarios. With caution should be noted that CCGT and OCGT as well as AC-connected and DC-connected offshore wind installations have high correlations because they are grouped in the MGA iterations.

6.3.4 Regional Must-Have Technologies

We also applied a near-optimal space analysis from a country-level perspective. Remarkably, we found that any one country could completely forego any one generation or storage technology and remain within 5% of the cost optimum when targeting a 95% reduction in greenhouse gas emissions. In this case, neighbouring countries must offset the absence of this technology.

Therefore, we aggregated the regional view even further in Figure 6.7, where we consider three regions for a fully renewable European system, to investigate whether there are any substantial constraints on the regional allocation of capacities. We find that even in this setting, only three pairs of regions and technologies featured must-haves, two of which regard offshore wind in Western and Eastern Europe. As one would intuitively expect, it moreover appears essential to build most solar capacities in Southern Europe.

6.4 Critical Appraisal

The present chapter covers the electricity sector only. Brown *et al.* [25] suggested that with an increasing coupling of energy sectors the benefit of the transmission system decreases. It is not far-fetched, that the near-optimal feasible space, in general, might look very different with a tightened sectoral integration. Within the computational constraints, it is moreover desirable to further enhance the spatial and temporal resolution to better reflect curtailment caused by transmission bottlenecks and factor in extreme weather events by taking account of multiple weather years [28, 29]. The present chapter further neglects parametric uncertainty, which we will address in the following [Chapter 7](#).

6.5 Conclusion


The present chapter shed light on the flatness of the near-optimal feasible decision space of a power system model with European scope for ambitious climate protection targets.

An understanding of alternatives beyond the least-cost solutions is indispensable to develop discussible policy guidelines. Therefore, we derived a set of technology-specific boundary conditions that must be satisfied to keep costs within pre-defined ranges using the [MGA](#) methodology. These rules permit well-informed discussions around social constraints to the exploitation of renewable resources or the extent to which the power transmission network can be reinforced in discussions.

Indeed, we observed high variance in the deployment of individual system components, even for a fully renewable system. Already a minor cost deviation of 0.5% offers a multitude of technologically diverse alternatives. It appears to be possible to dispense with onshore wind for a cost increase of 4%, and to forego solar for 10%. Nevertheless, either offshore or onshore wind energy plus at least some hydrogen storage and grid reinforcement seem to be essential to keep costs within 10% of the optimum.

Robust Near-Optimal Solutions

Contents of this chapter based on

Neumann, F. & Brown, T. Broad Ranges of Investment Configurations for Renewable Power Systems, Robust to Cost Uncertainty and Near-Optimality. eprint: <https://bit.ly/3fxGEAb> (2021). 

7.1 Introduction

Thus far, one of the major criticisms to be levelled against the analysis in [Chapter 6](#), just like most of the other studies that apply MGA-like methods, consists in the fact that it only uses a central cost projection for each considered technology. But recent decades have shown that many of these projections contain a high level of uncertainty, particularly for fast-moving technologies like solar, batteries and hydrogen storage [50–52]. This uncertainty propagates through the model to strongly affect the optimal and near-optimal system compositions, thus undermining any analysis of the trade-offs. Hence, it is crucial that apparent compromises are rigorously tested for robustness to technology cost uncertainty to raise confidence in conclusions about viable, cost-effective power system designs.

To thoroughly sweep the uncertainty space, we can fortunately avail of previous works on multi-dimensional global sensitivity analysis techniques in the context of least-cost optimisation [154, 172–175]. We expand their application to strengthening insights on the scope of near-optimal trade-offs, wherein the novelty of this contribution lies.

Here, we systematically explore robust trade-offs near the cost-optimum of a fully renewable European electricity system model, and investigate how they are affected by uncertain technology cost projections. Thereby, we evaluate both compromises between system cost and technology choices, as well as between pairs of technologies. We do so by solving numerous spatially and temporally explicit long-term investment planning problems that coordinate generation, transmission and

storage investments subject to multi-period linear optimal power flow constraints, while employing methods from global sensitivity analysis to account for a wide range of cost projections for wind, solar, battery and hydrogen storage technologies.

To handle the immense computational burden incurred by searching for near-optimal alternatives alongside evaluating many different cost parameter sets, we employ multi-fidelity surrogate modelling techniques, based on sparse polynomial chaos expansion (PCE) that allow us to merge results from one simpler and another more detailed model. This approach has been proven very effective in Tröndle *et al.* [154]. Heavy parallelisation with high-performance computing infrastructure allowed us to solve more than 50,000 resource-intensive optimisation problems which, in combination with surrogate modelling, admit spanning a probabilistic space of near-optimal solutions rather than putting single scenarios into the foreground.

Thereby, we are able to present alternative solutions beyond least-cost that have a high chance of involving a limited cost increase, just as we identify regions that are unlikely to be cost-efficient. We derive both ranges of options and technology-specific boundary conditions, that are not affected by cost uncertainty and must be met to keep the total system cost within a specified range. Our results show that indeed many such similarly costly but technologically diverse solutions exist regardless of how technology cost developments will unfold within the considered ranges.

7.2 Methods

In this section, we first outline how we obtain near-optimal solutions by deviating from the least-cost solution for a given cost parameter set. We then describe the model of the European power system and define the cost uncertainties. Finally, we explain how we make use of multi-fidelity surrogate modelling techniques based on PCEs and find an experimental design that efficiently covers the parameter space.

7.2.1 Finding Near-Optimal Alternatives

Using the least-cost solution of the thesis' common capacity expansion problem (Section 2.1) as an anchor, we use the ϵ -constraint method from multi-objective optimisation (MOO) to find near-optimal feasible solutions analogous to Chapter 6

[3, 176]. For notational brevity, let $c^\top x$ denote the linear objective function Equation (2.1) and $Ax \leq b$ the set of linear constraints Equations (2.2) to (2.13) in a space of continuous variables, such that the minimised system cost can be represented by

$$C = \min_x \{ c^\top x \mid Ax \leq b \}. \quad (7.1)$$

We then encode the original objective as a constraint such that the cost increase is limited to a given ϵ . Given this slack, we formulate alternative search directions in the objective. For instance, we can seek to minimise the sum of solar installations $x_s \subseteq x$ with

$$\bar{x}_s = \min_{x_s} \{ 1^\top x_s \mid Ax \leq b, \quad c^\top x \leq (1 + \epsilon) \cdot C \}. \quad (7.2)$$

To draw a full picture of the boundaries of the near-optimal feasible space, we systematically explore the extremes of various technologies: we both minimise and maximise the system-wide investments in solar, onshore wind, offshore wind, any wind, hydrogen storage, and battery storage capacities, as well as the total volume of transmission network expansion. Evaluating each of these technology groups for different cost deviations $\epsilon \in \{1\%, 2\%, 4\%, 6\%, 8\%\}$ allows us to observe how the degree of freedom regarding investment decisions rises as the optimality tolerance is increased, both at lower and upper ends. These boundaries yield Pareto frontiers where neither reducing system cost nor extremising the use of a particular technology, can be improved without diminishing the other.

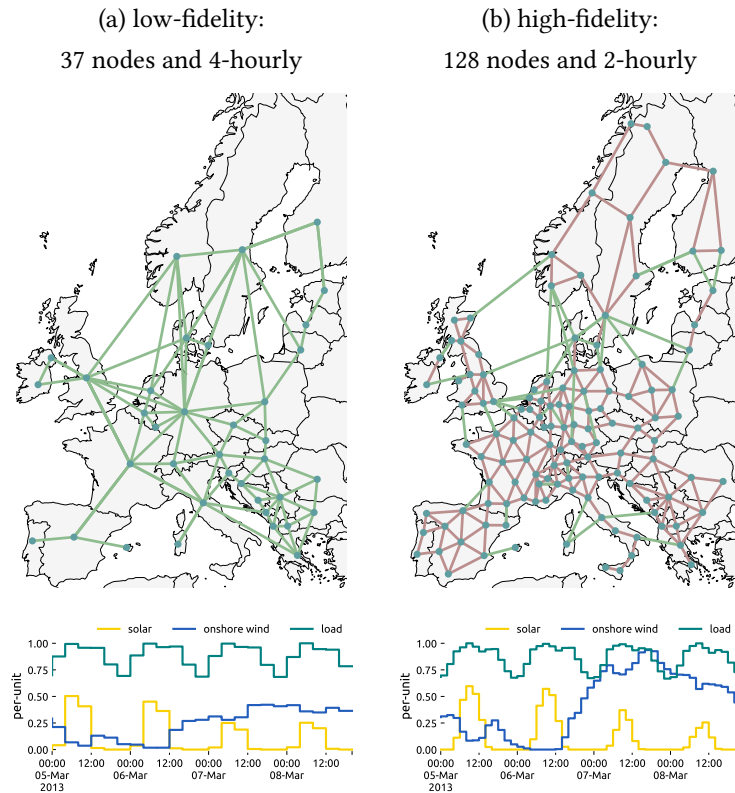
The near-optimal analysis above only explores the extremes of one technology at a time, i.e. one direction in the feasible space. But actually the space of attainable solutions within ϵ of the cost-optimum is multi-dimensional. To further investigate trade-offs between multiple technologies, in addition to the ϵ -constraint and the objective to extremise capacities of a particular technology, we formulate a constraint that fixes the capacity of another technology. For instance, we search for the minimum amount of wind capacity $x_w \subseteq x$ given that a certain amount of solar is built

$$\bar{x}_w = \min_{x_w} \{ 1^\top x_w \mid Ax \leq b, \quad c^\top x \leq (1 + \epsilon) \cdot C, \quad 1^\top x_s = \underline{x}_s + \alpha \cdot (\bar{x}_s - \underline{x}_s) \}. \quad (7.3)$$

The α denotes the relative position within the near-optimal range of solar capacities at given ϵ . For example, at $\alpha = 0\%$ we look for the least wind capacity given that minimal solar capacities are built. An alternative but more complex approach to spanning the space of near-optimal solutions in multiple dimensions at a time using a quick hull algorithm was presented by Pedersen *et al.* [169].

Due to computational constraints, we focus on technologies which are assumed to lend themselves to substitution and limit the corresponding analysis to a single cost

Figure 7.1: Spatial and temporal resolution of the low and high fidelity model. Green lines represent controllable HVDC lines. Red lines represent HVAC lines. The exemplary capacity factors for wind and solar are shown for four days in March at the northernmost node in Germany, alongside the normalised load profile.



increase level of $\epsilon = 6\%$. We consider the three pairs, (i) wind and solar, (ii) offshore and onshore wind, (iii) hydrogen and battery storage, by minimising and maximising the former while fixing the latter at positions $\alpha \in \{0\%, 25\%, 50\%, 75\%, 100\%\}$ within the respective near-optimal range.

7.2.2 Model Inputs

The instances of the coordinated capacity expansion problem (Section 2.1) are based on `PyPSA-Eur`, the open model of the European power transmission system introduced in Section 2.3 [13]. We target a fully renewable electricity system based on variable resources such as solar photovoltaics, onshore wind and offshore wind, that has not carbon emissions. We pursue a greenfield approach subject to a few notable exceptions. The existing hydro-electric infrastructure (run-of-river, hydro dams, pumped-storage) is included but not considered to be extendable due to assumed geographical constraints. Furthermore, the existing transmission infrastructure can only be reinforced continuously but may not be removed. In addition to balancing renewables in space with transmission networks, the model includes storage options at each node to balance renewables in time. We consider two extendable storage technologies: battery storage representing

Table 7.1: Technology cost uncertainty using optimistic and pessimistic assumptions from the Danish Energy Agency [177].

Technology	Lower Annuity	Upper Annuity	Unit
Onshore Wind	73	109	€/kW/a
Offshore Wind	178	245	€/kW/a
Solar	36	53	€/kW/a
Battery	30	125	€/kW/a
Hydrogen	111	259	€/kW/a

short-term storage suited to balancing daily fluctuations and hydrogen storage which exemplifies long-term synoptic and seasonal.

Since the spatial and temporal resolution strongly affects the size of the optimisation problem, running the model at full resolution is computationally infeasible. Throughout the chapter, we will therefore make use of two levels of aggregation, reflecting a compromise between the computational burden incurred by high-resolution models and the growing inaccuracies regarding transmission bottlenecks and resource distribution in low-resolution models. We consider a low-fidelity model with 37 nodes at a 4-hourly resolution for a full year that models power flow via a transport model (i.e. excluding KVL of Equation (2.13)) and a high-fidelity model with 128 nodes at a 2-hourly resolution that is subject to linearised load flow constraints (Figure 7.1).

7.2.3 Technology Cost Uncertainty

Uncertainty of technology cost projections is driven by two main factors: unknown learning rates (i.e. how quickly costs fall as more capacity is built) and unclear deployment rates (i.e. how much capacity will be built in the future) [178, 179]. As modelling technological learning endogeneously is computationally challenging due to the non-convexity it entails [180, 181], technology cost uncertainty is typically defined exogenously by an interval within which costs may vary and a distribution that specifies which segments are more probable.

Ranges of cost projections are best chosen as wide as possible to avoid excluding any plausible scenarios [63, 161]. When uncertainty has been considered in the literature, cost assumptions have commonly been modelled to vary between $\pm 20\%$ and $\pm 65\%$ depending on the technology's maturity [154, 161, 162, 173, 182]. In this study, we consider uncertainty regarding the annuities of onshore wind, offshore wind, solar PV, battery and hydrogen storage systems. The latter comprises the cost of electrolysis, cavern storage, and fuel cells. For solar PV we assume an even split between utility-scale PV and residential rooftop PV. Evaluating uncertainties based on annuities has a distinct advantage. They can be seen to simultaneously incorporate uncertainties about the overnight investments, fixed operation and

maintenance costs, their lifetime, and the discount rate, since multiple combinations lead to the same annuity. We built the uncertainty ranges presented in [Table 7.1](#) from the optimistic and pessimistic technology cost and lifetime projections for the year 2050 from the Danish Energy Agency, which correspond to 90% confidence intervals [177]. In cases where no uncertainty ranges were provided for the year 2050, such as for rooftop PV, projections for the year 2030 define the upper end of the uncertainty interval.

Distributions of cost projections have been assumed to follow normal [172] or triangular distributions [182]. But independent uniform distributions are the most prevalent assumption [154, 161, 162, 165, 183–186]. This approach is backed by the maximum entropy approach [154], which states that given the persistent lack of knowledge about the distribution the independent uniform distribution, that makes fewest assumptions, is most appropriate. Although the assumed independence may neglect synergies between technologies, for example, between offshore and onshore wind turbine development, we follow the literature by assuming that the cost are independent and uniformly distributed within the ranges specified in [Table 7.1](#).

7.2.4 Surrogate Modelling with Polynomial Chaos Expansion

Searching for least-cost solutions and many associated near-optimal alternatives of a highly resolved power system model on its own is already labour-intensive from a computational perspective. Repeating this search for a large variety of cost assumptions ([Section 7.2.3](#)), to be able to make statements about the robustness of investment flexibility near the optimum under uncertainty, adds another layer to the computational burden.

Surrogate models¹ offer a solution for such cases, where the outcome of the original model cannot be obtained easily. In contrast to the full model, they only imitate the input/output behaviour for a selection of aggregated outputs, but take much less time to compute [187]. Like other machine learning techniques, they generalise from a training dataset that comprises only a limited number of samples. As surrogate models interpolate gaps in the parameter space that are not contained in the sample set, which would otherwise be computationally expensive to fill, they are well suited to use cases such as parameter space exploration and sensitivity analysis.

Consequently, in this chapter we will make use of surrogate models that map the cost of onshore wind, offshore wind, solar, hydrogen, and battery storage ([Table 7.1](#)) onto a selection of eight system-level outputs. These are the total system cost and the installed onshore wind, offshore wind, solar, hydrogen, battery, and

¹Surrogate names are also known by names such as approximation models, response surface methods, metamodels and emulators.

transmission network capacities. We construct surrogate models for least-cost and near-optimal solutions separately for each system cost slack, search direction, fixed total capacity, and output variable. This results in a collection of 808 individual surrogate models based on 101 solved optimisation problems per set of cost assumptions. The method we choose from an abundance of alternatives is based on PCE [188–190]. We select this approach because the resulting approximations allow efficient analytical statistical evaluation [188] and can conveniently combine training data from variously detailed models [187].

The general idea of surrogate models based on PCE is to represent uncertain model outputs as a linear combination of orthogonal basis functions of the random input variables weighted by deterministic coefficients [191]. It is a Hilbert space technique that works in principle analogously to decomposing a periodic signal into its Fourier components [191]. Building the surrogate model consists of the following steps: (i) sampling a set of cost projections from the parameter space, (ii) solving the least-cost or near-optimal investment planning problem for each sample, (iii) selecting an expansion of orthogonal polynomials within the parameter space, (iv) performing a regression to calculate the polynomial coefficients, and ultimately (v) using the model approximation for statistical analysis. In the following, we will formalise this approach mathematically, which we implemented using the *chaospy* toolbox [192], and elaborate on individual aspects in more detail.

We start by defining the vector of random input variables as

$$\mathbf{x} = \{x_1, \dots, x_m\} \quad (7.4)$$

that represents the m uncertain cost projections. Further, we let

$$y = f(\mathbf{x}) \quad (7.5)$$

describe how the uncertainty of inputs \mathbf{x} propagates through the computationally intensive model f (i.e. the solving a large optimisation problem) to the outputs $y \in \mathbb{R}$.

We can represent the computational model f with its PCE

$$y = f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^m} r_\alpha \psi_\alpha(\mathbf{x}), \quad (7.6)$$

where ψ_α denotes multivariate orthogonal polynomials that form a Hilbertian basis and $r_\alpha \in \mathbb{R}$ are the corresponding polynomial coefficients [188]. The multiindex $\alpha = \{\alpha_1, \dots, \alpha_m\}$ denotes the degree of the polynomial ψ_α in each of the m random input variables x_i . As Equation (7.6) features an infinite number of

unknown coefficients, it is common practice to approximate by truncating the expansion to get a finite number of coefficients

$$f(\mathbf{x}) \approx f'(\mathbf{x}) = \sum_{\alpha \in \mathcal{A}^{m,p}} r_{\alpha} \psi_{\alpha}(\mathbf{x}). \quad (7.7)$$

In the standard truncation scheme [188, 190], all polynomials in m input variables where the total degree is less than p are selected. We can write this as a set of indices

$$\mathcal{A}^{m,p} = \{\alpha \in \mathbb{N}^m : |\alpha| \leq p\}, \quad (7.8)$$

where $|\alpha| = \sum_{i=1}^m \alpha_i$. Given the joint distribution of \mathbf{x} and a maximum degree, a suitable collection of orthogonal polynomials can be constructed using a three terms recurrence algorithm [192]. The cardinality of the truncated PCE,

$$q = \text{card } \mathcal{A}^{m,p} = \binom{m+p}{p} = \frac{(m+p)!}{m!p!}, \quad (7.9)$$

indicates the number of unknown polynomial coefficients.

We determine these coefficients by a regression based on a set of cost parameter samples and the corresponding outputs,

$$\mathcal{X} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\} \quad \text{and} \quad \mathcal{Y} = \{f(\mathbf{x}^{(1)}), \dots, f(\mathbf{x}^{(n)})\}. \quad (7.10)$$

Using this training dataset, we minimise the least-square residual of the polynomial approximation across all observations. We add an extra L_1 regularisation term, that induces a preference for fewer non-zero coefficients, and solve

$$\hat{\mathbf{r}} = \underset{\mathbf{r} \in \mathbb{R}^q}{\text{argmin}} \left[\frac{1}{n} \sum_{i=1}^n \left(f(\mathbf{x}^{(i)}) - \sum_{\alpha \in \mathcal{A}^{m,p}} r_{\alpha} \psi_{\alpha}(\mathbf{x}^{(i)}) \right)^2 + \lambda \|\mathbf{r}\|_1 \right], \quad (7.11)$$

where we set the regularisation penalty to $\lambda = 0.005$. This results in a sparse PCE that has proven to improve approximations in high-dimensional uncertainty spaces and to reduce the required number of samples for comparable approximation errors [190]. Knowing the optimised regression coefficients, we can now assemble the complete surrogate model

$$y = f(\mathbf{x}) \approx f'(\mathbf{x}) = \sum_{\alpha \in \mathcal{A}^{m,p}} \hat{r}_{\alpha} \psi_{\alpha}(\mathbf{x}). \quad (7.12)$$

7.2.5 Multifidelity Approach

To construct a sufficiently precise PCE-based surrogate model, it is desirable to base it on many samples from a high-fidelity model. However, this is likely prohibitively time-consuming. On the other hand, relying only on samples from a low-fidelity model may be too inaccurate [193]. For example, an investment model that features only a single node per country will underestimate transmission bottlenecks and regionally uneven resource or demand distribution. In Section 7.2.2 we already alluded to using two models with varying spatial and temporal resolution in the present chapter. We integrate both in a multi-fidelity approach [187, 193], and demonstrate how we can simultaneously avail of high coverage of the parameter space by sampling the simpler model many times, and the high spatio-temporal detail yielded by fewer more complex model runs.

The idea of the multi-fidelity approach is to build a corrective surrogate model $\Delta'(\mathbf{x})$ for the error of the low-fidelity model f_ℓ compared to the high-fidelity model f_h

$$\Delta(\mathbf{x}) = f_h(\mathbf{x}) - f_\ell(\mathbf{x}), \quad (7.13)$$

and add it to a surrogate model of the low-fidelity model to approximate the behaviour of the high-fidelity model

$$f'_h(\mathbf{x}) = f'_\ell(\mathbf{x}) + \Delta'(\mathbf{x}). \quad (7.14)$$

Typically, the corrective PCE rectifies only the lower order effects of the low-fidelity surrogate model [187]. The advantage is that this way the correction function can be determined based on fewer samples analogous to Section 7.2.4. To sample the errors, it is only required that the high-fidelity samples are a subset of the low-fidelity samples, e.g.

$$\mathcal{X}_h = \{ \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n_h)} \} \quad \text{and} \quad \mathcal{X}_\ell = \{ \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n_h)}, \dots, \mathbf{x}^{(n_\ell)} \}, \quad (7.15)$$

which we can easily guarantee by using deterministic low-discrepancy series in the experimental design (Section 7.2.6). With $p_c < p_\ell$ and consequently $\mathcal{A}_c \subset \mathcal{A}_\ell$, the multi-fidelity surrogate model can be written as a combination of low-fidelity and corrective polynomial coefficients

$$f'_h(\mathbf{x}) = \sum_{\alpha \in \mathcal{A}_\ell^{m,p_\ell} \cap \mathcal{A}_c^{m,p_c}} (r_{\ell,\alpha} + r_{c,\alpha}) \psi_\alpha(\mathbf{x}) + \sum_{\alpha \in \mathcal{A}_\ell^{m,p_\ell} \setminus \mathcal{A}_c^{m,p_c}} r_{\ell,\alpha} \psi_\alpha(\mathbf{x}). \quad (7.16)$$

In the present chapter, we apply a multi-fidelity surrogate model that considers effects up to order three observed in the low-fidelity model. These are then corrected with linear terms derived from insights from the high-fidelity model.

We justify this choice by experimentation in [Section 7.2.7](#), by testing against other typical choices between orders one to five [190]. Given the polynomial expansion order, the remaining question is how many samples are necessary to attain an acceptable approximation.

7.2.6 Experimental Design

The experimental design covers strategies to find sufficiently high coverage of the parameter space at low computational cost [175, 189]. It deals with how many samples are drawn and what sampling method is used.

Traditional Monte-Carlo sampling with pseudo-random numbers is known to possess slow convergence properties, especially in high-dimensional parameter spaces. So-called low-discrepancy series can greatly improve on random sampling. Because they are designed to avoid forming large gaps and clusters, these deterministic sequences efficiently sample from the parameter space [189]. Thus, we choose to draw our samples from a low-discrepancy Halton sequence.

For the question about how many samples should be drawn, we resort to the oversampling ratio (OSR) as a guideline. The OSR is defined as the ratio between the number of samples and the number of unknown coefficients [187]. The literature recommends values between two and three [187, 189, 190, 194]. In other words, for a sufficiently accurate approximation, there should be significantly more samples than unknown coefficients. If the OSR is lower, the regression is prone to the risk of overfitting. On the other hand, a high OSR may lead to a very coarse approximation [187].

According to [Equation \(7.9\)](#), targeting an OSR of two and considering the five uncertain technology cost parameters ([Table 7.1](#)), approximating linear effects would require at least 12 samples, whereas cubic relations would already need 112 samples. Even 504 samples would be necessary to model the dynamics of order 5. To investigate the quality of different PCE orders and retain a validation dataset, we draw 500 samples for the low-fidelity model. Due to the computational burden carried by the high-fidelity models, we settle on a linear correction in advance, such that 15 samples for the high-fidelity model are acceptable. In combination with 101 least-cost and near-optimal optimisation runs for each sample, this setup results in a total number of 50,500 runs of the low-fidelity model and 1,515 runs of the high-fidelity model. On average a single high-fidelity model run took 20 GB of memory and 5 hours to solve. Each low-fidelity model run on average consumed 3 GB of memory and completed within 5 minutes. This setup profits tremendously from parallelisation as it involves numerous independent optimisation runs. Moreover, it would have been infeasible to carry out without high-performance computing.

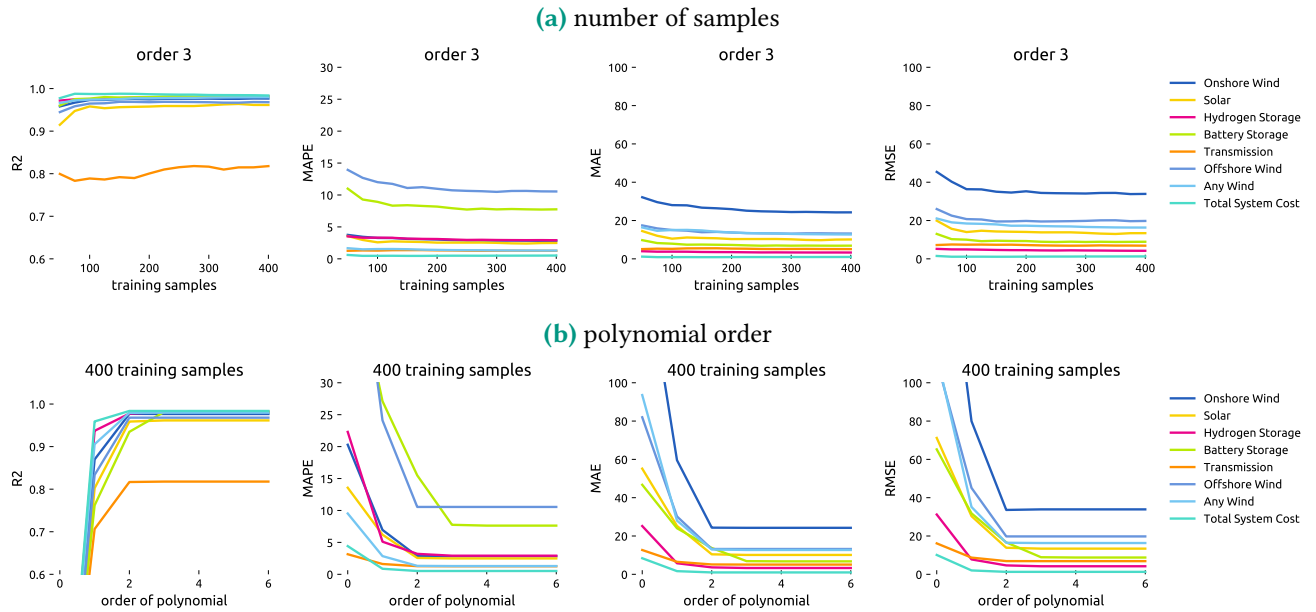


Figure 7.2: Cross-validation errors by output for varying sample sizes and polynomial orders of least-cost low-fidelity surrogate models.

7.2.7 Model Validation

We justify the use of surrogate modelling by cross-validation. Out of the 500 low-fidelity samples, 100 samples are not used in the regression. This validation dataset is unknown to the surrogate model and is consulted to assess the approximation’s quality. Because the high-fidelity sample size is limited and approximating near-optimal solutions is not assumed to fundamentally differ, we base the validation on low-fidelity least-cost solutions only. We experimentally evaluate the approximation errors between predicted and observed data for different combinations of polynomial order and sample size to decide on a suitable parameterisation. We present the coefficient of determination (R^2) for the variance captured, the mean absolute (percentage) errors (MAE/MAPE) for absolute and relative deviations, and the root mean squared error (RMSE).

Regarding the number of samples required, Figure 7.2a foremost illustrates that, given enough samples, we achieve average relative errors of less than 4% for most output variables. This is comparable to the cross-validation errors from Tröndle *et al.* [154] at rates below 5%. Only for offshore wind and battery storage, we observe larger errors. However, this can be explained by a distortion of the relative measure when these technologies are hardly built for some cost projections. On the contrary, the prediction of total system costs is remarkably accurate. Figure 7.2a also demonstrates that for a polynomial order of 3, we gain no significant improvement with more than 200 samples. In fact, thanks to the

regularisation term used in the regression, we already attain acceptable levels of accuracy with as few as 50 samples. Moreover, the high R^2 values underline that the surrogate model can explain most of the output variance.

Regarding the polynomial order, [Figure 7.2b](#) shows that an order of 2 and below may be too simple to capture the interaction between different parameters. On the other hand, an order of 4 and above yields no improvement and, were it not for the moderating regularisation term, would even result in a loss of generalisation properties due to overfitting. As higher-order approximations require significantly more samples, an order of 3 appears to be a suitable compromise to limit the computational burden.

7.3 Results and Discussion

In this section, we approach the uncertainty analysis to near-optimal solutions by reviewing the propagation of input uncertainties into least-cost solutions first and expanding gradually from there. This includes inspecting cost and capacity distributions induced by unknown future technology cost and conducting a global sensitivity analysis that identifies the most influential cost parameters for least-cost solutions. We then expand the uncertainty analysis to the space of nearly cost-optimal solutions, which yields us insights about the consistency of near-optimal alternatives across a variety of cost parameters.

7.3.1 Cost and Capacity Distribution of Least-Cost Solutions

Based on the uncertainty of cost inputs, the total annual system costs vary between 160 and 220 billion Euro per year, as displayed in [Figure 7.3](#). This means the most pessimistic cost projections entail about 40% higher cost than the most optimistic projections. All least-cost solutions build at least 350 GW solar and 600 GW wind, but no more than 1100 GW. While wind capacities tend towards higher values, solar capacities tend towards lower values. We observe that least-cost solutions clearly prefer onshore over offshore wind, yet onshore wind features the highest uncertainty range alongside battery storage. The cost optimum gravitates towards hydrogen storage rather than battery storage unless battery storage becomes very cheap. There are no least-cost solutions without hydrogen, only some without battery storage. Transmission network expansion is least affected by cost uncertainty and consistently doubled compared to today's capacities. The question arises, what we can conclude from these insights. The interpretation of the observed ranges may be limited because they are not robust when we look beyond the

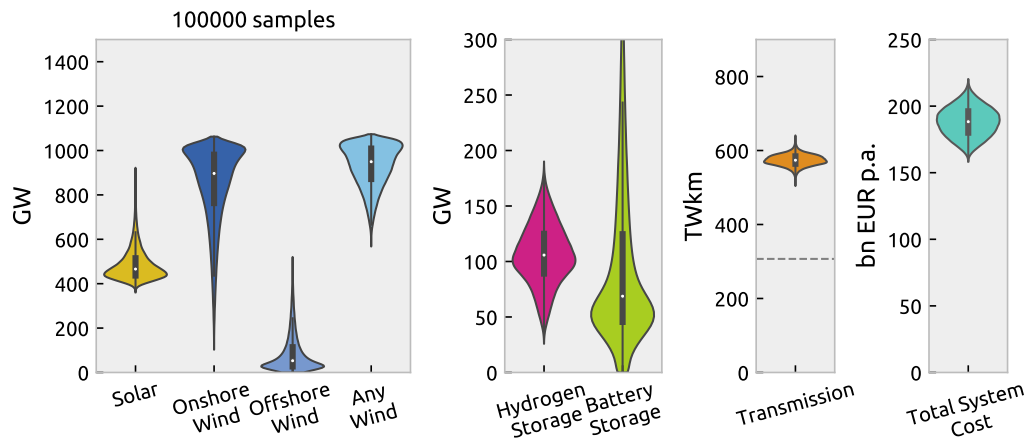


Figure 7.3: Distribution of total system cost, generation, storage, and transmission capacities for least-cost solutions.

least-cost solutions and acknowledge structural modelling uncertainties, such as social constraints. Moreover, the pure distribution of outputs does not yet convey information about how sensitive results are to particular cost assumptions. But knowing the technologies for which lowering overnight costs has a significant impact is important to promote technological learning in that direction.

7.3.2 Parameter Sweeps and Global Sensitivity Indices

Figure 7.4 addresses a selection of local self-sensitivities, i.e. how the cost of a technology influences its deployment while displaying the remaining uncertainty induced by other cost parameters. The overall tendency is easily explained: the cheaper a technology becomes, the more it is built. However, changes of slope and effects on the uncertainty range as one cost parameter is swept are insightful nonetheless. For instance, Figure 7.4 reveals that battery storage becomes significantly more attractive economically once its annuity falls below 75 €/kW/a (including 6h energy capacity at full power output) hydrogen storage features a steady slope. A low cost of onshore wind makes building much onshore wind capacity attractive with low uncertainty, whereas if onshore wind costs are high how much is built greatly depends on other cost parameters. The opposite behaviour is observed for offshore wind and solar. The cost of hydrogen storage mostly causes the limited uncertainty about cost-optimal levels of grid expansion. As the cost of hydrogen storage falls, less grid reinforcement is chosen. But since the presented self-sensitivities only exhibit a fraction of all sensitivities, in the next step we formalise how input uncertainties affect each outcome systematically

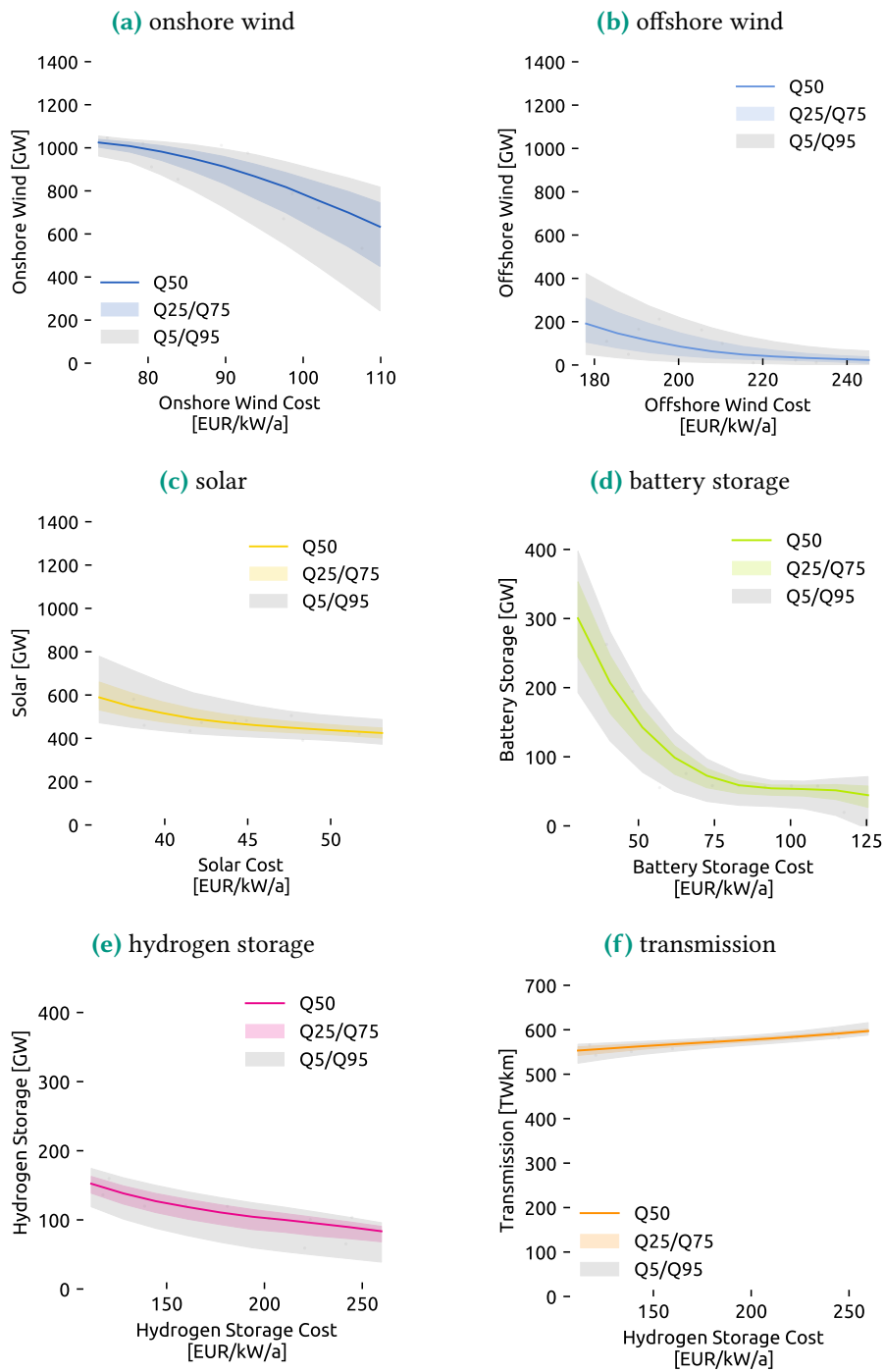


Figure 7.4: Sensitivity of capacities towards their own technology cost. The median (Q50) alongside the 5%, 25%, 75%, and 95% quantiles (Q5–Q95) display the sensitivity subject to the uncertainty induced by other cost parameters.

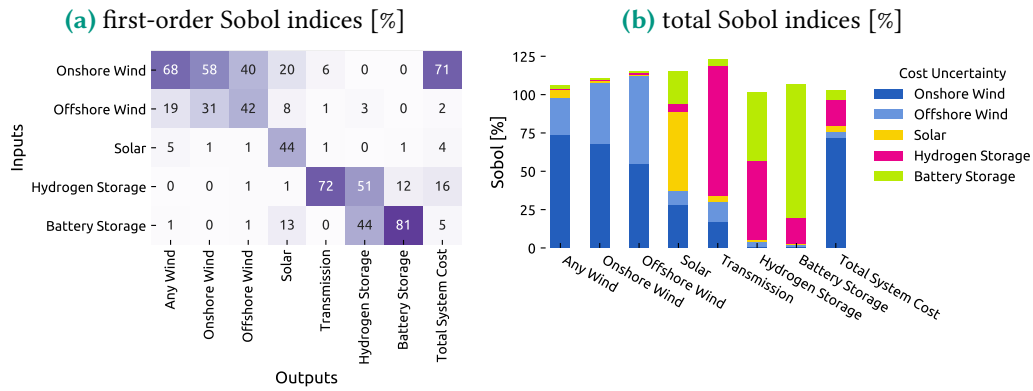


Figure 7.5: Sobol indices. These sensitivity indices attribute output variance to random input variables and reveal which inputs the outputs are most sensitive to. The first-order Sobol indices quantify the share of output variance due to variations in one input parameter alone. The total Sobol indices further include interactions with other input variables. Total Sobol indices can be greater than 100% if the contributions are not purely additive.

by applying variance-based global sensitivity analysis techniques, which have been applied in the context of energy systems, e.g. in [154, 172].

Sensitivity indices, or Sobol indices, attribute the observed output variance to each input and can be computed analytically from the PCE [188]. For our application, the Sobol indices can, for instance, tell us which technology cost contributes the most to total system cost or how much of a specific technology will be built. The first-order Sobol indices describe the share of output variance due to variations in one input alone averaged over variations in the other inputs. Total Sobol indices also consider higher-order interactions, which are greater than 100% if the relations are not purely additive.

The first-order and total Sobol indices for least-cost solutions in Figure 7.5 show that the total system cost is largely determined by how expensive it is to build onshore wind capacity, followed by the cost of hydrogen storage. The amount of wind in the system is almost exclusively governed by the cost of onshore and offshore wind parks. Other carriers yield a more varied picture. The cost-optimal solar capacities additionally depend on onshore wind and battery costs. The amount of hydrogen storage is influenced by battery and hydrogen storage cost alike. Although there are noticeable higher-order effects, which are most extensive for transmission, the first-order effects dominate. Strikingly, the volume of transmission network expansion strongly depends on the cost of hydrogen storage, which can be explained by the synoptic spatio-temporal scale of wind power variability across the European continent which both hydrogen storage and transmission networks seek to balance from different angles. While hydrogen storage typically balances multi-week variations in time, continent-spanning

transmission networks exploit the circumstance that as weather systems traverse the continent, it is likely always to be windy somewhere in Europe.

7.3.3 Fuzzy Near-Optimal Corridors with Rising Cost Slack

So far, we quantified the output uncertainty and analysed the sensitivity towards inputs at least-cost solutions only. Yet, it has been previously shown that even for a single cost parameter set a wide array of technologically diverse but similarly costly solutions exists [3]. In the next step, we examine how technology cost uncertainty affects the shape of the space of near-optimal solutions.

By identifying feasible alternatives common to all, few or no cost samples, we outline low-cost solutions common to most parameter sets (e.g. above 90% contour) as well as system layouts that do not meet low-cost criteria in any circumstances for varying ϵ in Figure 7.6. The wider the displayed contour lines are apart, the more uncertainty exists about the boundaries. The closer contour lines are together, the more specific the limits are. The height of the quantiles quantifies flexibility for a given level of certainty and slack; the angle presents information about the sensitivity towards cost slack.

From the fuzzy upper and lower Pareto fronts in Figure 7.6 we can see that it is highly likely that building 900 GW of wind capacity is possible within 3% of the optimum, and that conversely building less than 600 GW has a low chance of being near the cost optimum. Only a few solutions can forego onshore wind entirely and remain within 8% of the cost-optimum, whereas it is very likely possible to build a system without offshore wind at a cost penalty of 4% at most. On the other hand, more offshore wind generation is equally possible. Unlike for onshore wind, where it is more uncertain how little can be built, uncertainty regarding offshore wind deployment exists about how much can be built so that costs remain within a pre-specified range. For solar, the range of options within 8% of the cost optimum at 90% certainty is very wide. Anything between 100 GW and 1000 GW appears feasible. In comparison to onshore wind, the uncertainty about minimal solar requirements is smaller.

The level of required transmission expansion is least affected by the cost uncertainty. To remain within $\epsilon = 8\%$ it is just as likely possible to plan for moderate grid reinforcement by 30% as is initiating extensive remodelling of the grid by tripling the transmission volume compared to what is currently in operation. These results indicate that in any case some transmission reinforcement to balance renewable variations across the continent appears to be essential. Hydrogen storage, symbolising long-term storage, also gives the impression of a vital technology in many cases. Building 100 GW of hydrogen storage capacity is likely

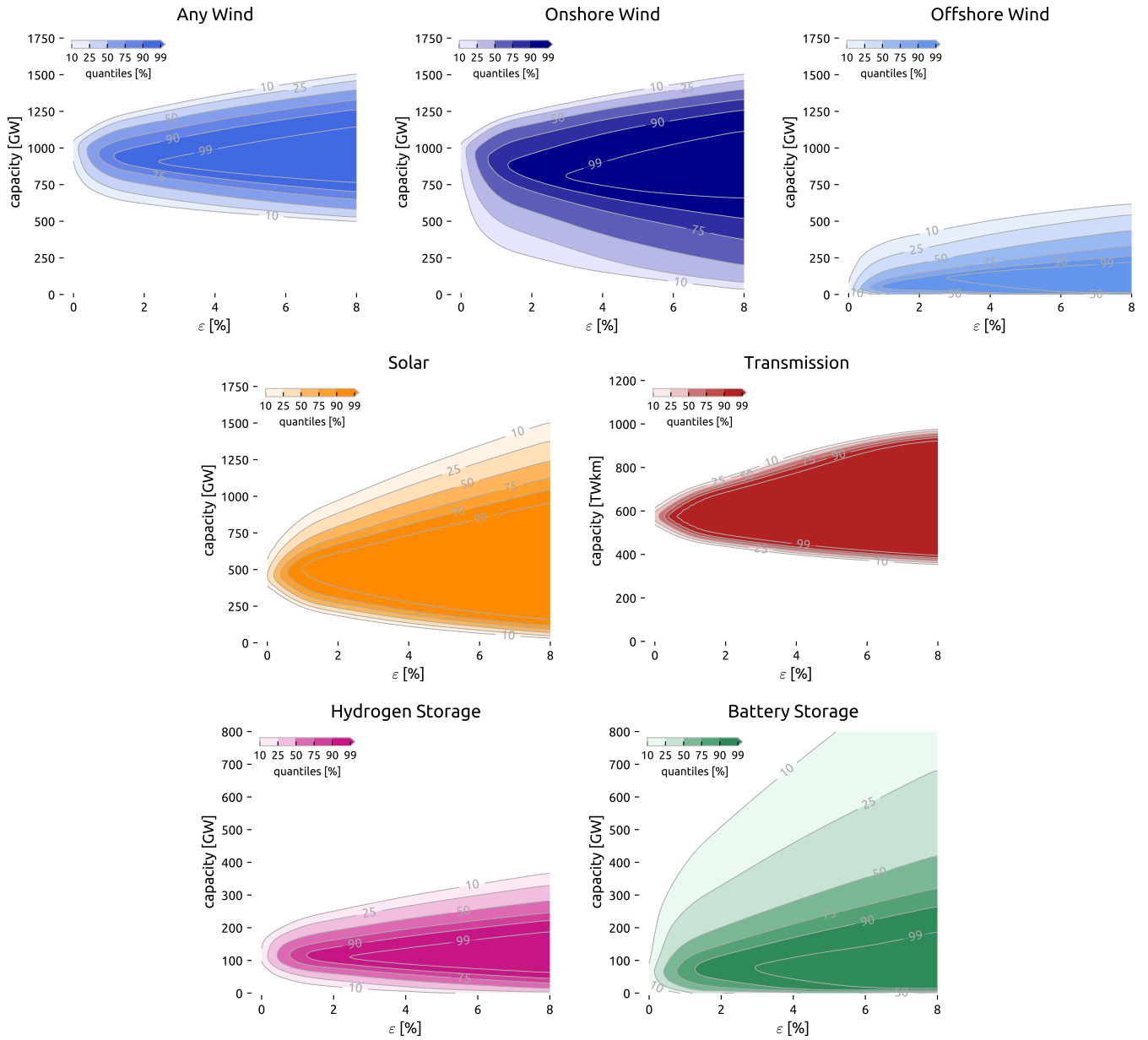


Figure 7.6: Space of near-optimal solutions by technology under cost uncertainty. For each technology and cost sample, the minimum and maximum capacities obtained for increasing cost penalties ϵ form a cone, starting from a common least-cost solution. By arguments of convexity, the capacity ranges contained by the cone can be near-optimal and feasible, given a degree of freedom in the other technologies. From optimisation theory, we know that the cones widen up for increased slacks. As we consider technology cost uncertainty, the cone will look slightly different for each sample. The contour lines represent the frequency a solution is inside the near-optimal cone over the whole parameter space. This is calculated from the overlap of many cones, each representing a set of cost assumptions. Due to discrete sampling points in the ϵ -dimension, the plots further apply quadratic interpolation and a Gaussian filter for smoothing.

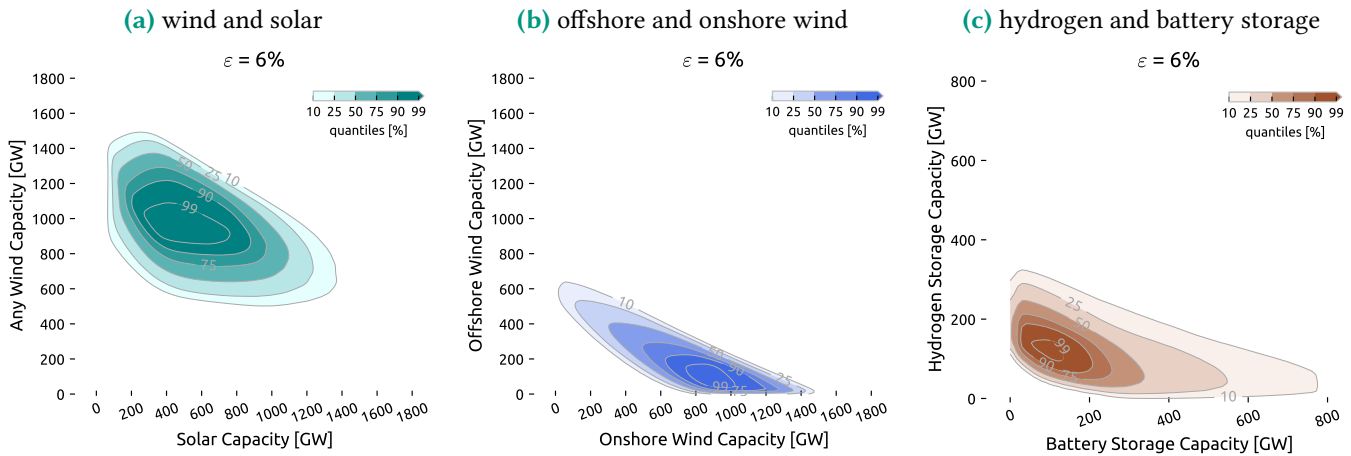


Figure 7.7: Space of near-optimal solutions by selected pairs of technologies under cost uncertainty. Just like in Figure 7.6, the contour lines depict the overlap of the space of near-optimal alternatives across the parameter space. It can be thought of as the cross-section of the probabilistic near-optimal feasible space for a given ϵ in two technology dimensions and highlights that the extremes of two technologies from Figure 7.6 cannot be achieved simultaneously.

viable within 2% of the cost optimum and, even at $\epsilon = 8\%$, only 25% of cost samples require no long-term storage; when battery costs are exceptionally low. Overall, 90% of cases appear to function without any short-term battery storage while the system cost rises by 4% at most. However, especially battery storage exhibits a large degree of freedom to build more.

7.3.4 Probabilistic Trade-Offs Between Two Technologies

The fuzzy cones from Figure 7.6 look at trade-offs between system cost and single technologies, assuming that the other technologies can be heavily optimised. But as there are dependencies between the technologies, in Figure 7.7 we furthermore evaluate trade-offs between technologies for three selected pairs at fixed system cost increase of $\epsilon = 6\%$, addressing which combinations of wind and solar capacity, offshore and onshore turbines, and hydrogen and battery storage are likely to be cost-efficient.

First, Figure 7.7a addresses constraints between wind and solar. The upper right boundary exists because building much of both wind and solar would be too expensive. The absence of solutions in the bottom left corner means that building too little of any wind or solar does not suffice to generate enough electricity. From the shape and contours, we see a high chance that building 1000 GW of wind and 400 GW of solar is within 6% of the cost-optimum. On the other hand, building

less than 200 GW of solar and 600 GW of wind is unlikely to yield a low-cost solution. In general, minimising the capacity of both primal energy sources will shift capacity installations to high-yield locations even if additional network expansion is necessary and boost the preference for highly efficient storage technologies. Overall, we can take away from this that, even considering combinations of wind and solar, a wide space of low-cost options exists with moderate to high likelihood, although the range of alternatives is shown to be more constrained.

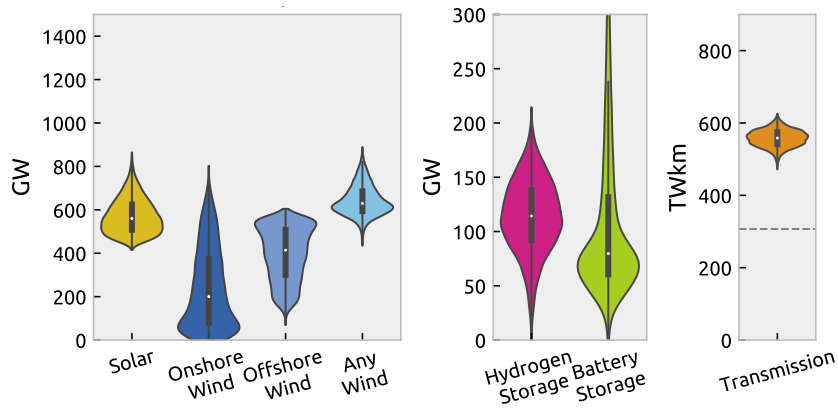
The trade-off between onshore wind and offshore wind is illustrated in [Figure 7.7b](#). Here, the most certain area is characterised by building more than 600 GW onshore wind, and less than 250 GW offshore wind capacity. However, there are some solutions with high substitutability between onshore and offshore wind, shown in the upper left bulge of the contour plot. Compared to wind and solar, the range of near-optimal solutions is even more constrained. The key role of energy storage in a fully renewable system is underlined in [Figure 7.7c](#). Around 50 GW of each is at least needed in any case, while highest likelihoods are attained when building 150 GW of each.

7.3.5 Selected Near-Optimal Capacity Distributions

The aforementioned contour plots [Figures 7.6 to 7.7](#) outline what is likely possible within specified cost ranges and subject to technology cost uncertainty, but do not expose the changes the overall system layout experiences when reaching for the extremes in one technology. Therefore, we show in [Figure 7.8](#) how the system-wide capacity distributions vary compared to the least-cost solutions ([Figure 7.3](#)) for two exemplary alternative objectives. For that, we chose minimising onshore wind capacity and transmission expansion because they are often linked to social acceptance issues.

[Figure 7.8a](#) illustrates that reducing onshore wind capacity is predominantly compensated by increased offshore wind generation but also added solar capacities. The increased focus on offshore wind also leads to a tendency towards more hydrogen storage, while transmission expansion levels are similarly distributed as for the least-cost solutions. From [Figure 7.8b](#) we can further extract that avoiding transmission expansion entails more hydrogen storage that compensates balancing in space with balancing in time, and more generation capacity overall, where resources are distributed to locations with high demand but weaker capacity factors and more heavily curtailed.

(a) minimal onshore wind with 8% system cost slack



(b) minimal transmission expansion with 8% system cost slack

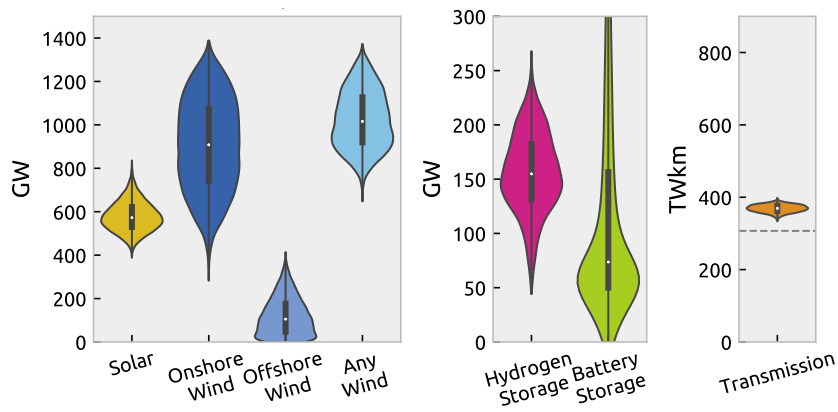


Figure 7.8: Distribution of total system cost, generation, storage, and transmission capacities for two near-optimal search directions with $\epsilon = 8\%$ system cost slack.

7.4 Critical Appraisal

The need to solve models for many cost projections and near-optimal search directions in reasonable time means that compromises had to be made in other modelling dimensions. For instance, the analysis would profit from a richer set of technologies and further uncertain input parameters, including efficiencies of fuel cells and electrolysis or the consideration of concentrating solar power, geothermal energy, biomass, and nuclear to name just a few. But as the number of considered technologies and parameters rises, so does the computational burden. Given the already considerable computational efforts involved in procuring our results, considering the full breadth of technologies and uncertainties would not have been feasible with the computational resources available. Moreover, limitations apply to the scope of the analysis which is limited to the electricity sector does not consider coupling to other energy sectors. However, accounting for interactions across sectors at high resolution in similarly set future studies is desirable and in development. Additionally, we assess no path dependencies via multi-period investments and endogenous learning, but optimise for an emission reduction in a particular target year based on annualised costs. We further disregard interannual variations of weather data by basing the analysis just on a single weather year for computational reasons. Lastly, aspects such as reserves, system adequacy and inertia have not been considered.

7.5 Conclusion

In the present chapter, we systematically explore a space of alternatives beyond least-cost solutions for society and politics to work with. We show how narrowly following cost-optimal results underplays an immense degree of freedom in designing future renewable power systems. To back our finding that there is no unique path to cost-efficiency, we account for the inherent uncertainties regarding technology cost projections, and draw robust conclusions about the range of options, boundary conditions and cost sensitivities:

Wide Range of Trade-Offs We find that there is a substantial range of options within 8% of the least-cost solution regardless of how cost developments will unfold. This holds across all technologies individually and even when considering dependencies between wind and solar, offshore and onshore wind, as well as hydrogen and battery storage.


Must-Avoid Boundary Conditions We also carve out a few boundary conditions which must be met to keep costs low and are not affected by the prevailing cost uncertainty. For a fully renewable power system, either offshore or onshore wind capacities in the order of 600 GW along with some long-term storage technology and transmission network reinforcement by more than 30% appears essential.

Technology Cost Sensitivities We identify onshore wind cost as the apparent main determinant of system cost, though it can often be substituted with offshore wind for a small additional cost. Moreover, the deployment of batteries is the most sensitive to its cost, whereas required levels of transmission expansion are least affected by cost uncertainty.

The robust investment flexibility in shaping a fully renewable power system we reveal opens the floor to discussions about social trade-offs and navigating around issues, such as public opposition towards wind turbines or transmission lines. Rather than modellers making normative choices about how the energy system should be optimised, we offer computational methods that present a wide spectrum of options and trade-offs that are feasible and within a reasonable cost range, to help society decide how to shape the future of the energy system.

Regional Equity and Autarky

Contents of this chapter based on

Neumann, F. Costs of Regional Equity and Autarky in a Renewable European Power System. *Energy Strategy Reviews* 35, 100652. doi:[10/gjr2fb](https://doi.org/10/gjr2fb). arXiv: [2007.08379](https://arxiv.org/abs/2007.08379) (2021). 

8.1 Introduction

Optimising for an unmitigated least-cost renewable power system entails very heterogeneous distributions of electricity generation in relation to demand when compared to national imbalances reported by [ENTSO-E](#) for 2018 ([Figure 8.1](#)) [195]. The system is dominated by many distinct net importers and exporters, whereas few supply just their own demand. This raises concerns about distributional equity, which in this chapter describes how evenly generation capacities are distributed relative to the regional demands.

Narrowly following the cost optimum risks inequitable outcomes and public headwind, bearing the potential of decelerating the energy transition. Particularly wind farms and transmission lines spark local opposition, which was found to be best counteracted by including the public in the planning process and by sharing profits [196]. Vice versa, also the absence of investments may have a detrimental impact on the economic prosperity of local communities.

Beyond the spatial distribution of generation capacities, numerous other equity principles exist [166]. Equity metrics can also relate to temporal, income, racial, labor and environmental aspects [197–199], and perceptions of fairness vary among stakeholders [200, 201]. Recent developments of pan-continental models with growing sub-national detail raise the need for recognising their regional implications

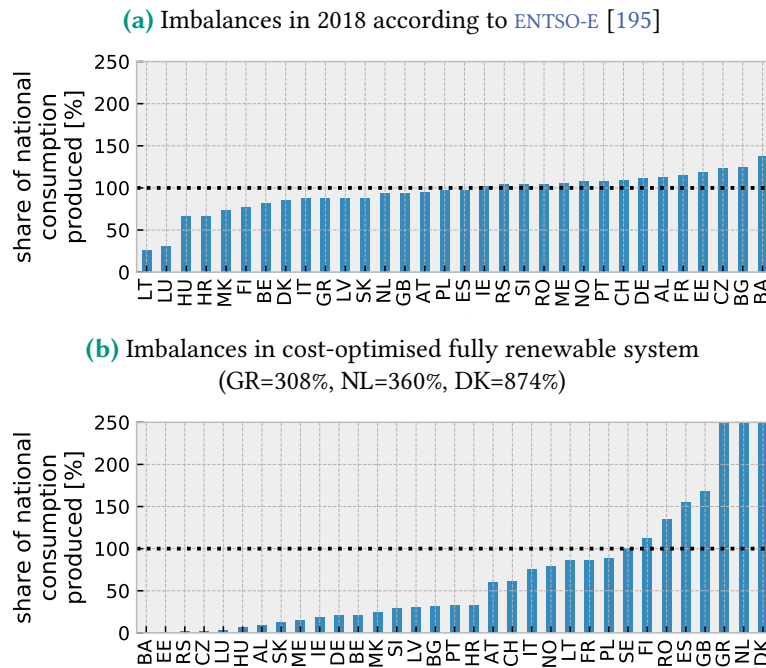


Figure 8.1: Imbalances observed in 2018 and in cost-optimised renewable system.

[199, 202]. However, such aspects are challenging to assess in endogenous modelling, and analyses have been limited to ex-post analysis [199]. Enhanced collaboration between social scientists and energy modellers has been encouraged [203].

Moreover, there is a trend towards discussing energy autarky, i.e. the ability to operate regions partially or completely independently [204, 205]. Positive associations with autonomy, control and independence drive aspirations for self-sufficiency for individuals and municipalities alike, and result in a higher willingness to pay and greater support for projects [205–208]. The debate also evolves around the resilience of more decentralised systems [209]. The primary resource-based feasibility of autarkic systems on different spatial levels was evaluated in Tröndle *et al.* [210]. High population density was found to sometimes be a limiting factor for small autarkic systems. Weinand *et al.* [211] found that about half of the 11,300 municipalities in Germany have sufficient potentials to become off-grid municipalities.

Further related work has assessed the benefit of transmission capacities between countries [37] and more heterogeneous distributions of generation assets [21]. It has moreover been evaluated what costs are incurred by reducing eligible potentials [212]. Previous work on distributional equity regarding power generation has however covered only a single country and neglected the variability of renewable generation and demand, as well as the interaction between storage and transmis-

sion infrastructure [166, 213]. A cost assessment of regional autarkic systems in Europe compared to the least-cost system does not appear to exist yet¹.

In the present chapter, we remedy the concerns about spatial scope and temporal resolution and explore at what cost more evenly distributed, or even autarkic, power supply could be achieved in Europe, regarding both countries and smaller regions.

8.2 Model Inputs and Simulation Setup

As previously, we use the open European transmission system model `PyPSA-Eur` presented in Section 2.3 with 200 nodes and 4380 snapshots, one for every two hours in a year [13]. We solve the long-term power system planning problem from Section 2.1 which seeks to minimise the total annual system costs comprising generation, transmission and storage infrastructure in a fully renewable system, subject to multi-period linear optimal power flow (LOPF) constraints.

We add constraints for each country or node to produce on average at least a given share of their annual consumption; i.e. we explore the sensitivity of increasing production equity requirements. The extreme cases are (i) every country or node produces as much as required for the cost-optimal system using the most productive locations (0%) and (ii) every country or node produces as much as they consume (100%). The experiments interpolate between the extremes in steps of 10%.

We further extend this setup by two experiments regarding absolute autarky: (i) one where there is no cross-border transmission of power between countries but which includes the intranational transmission grid, and (ii) one where each node fully supplies its own power demand at any time in isolation. The code and assumptions to reproduce all results is available at github.com/fneum/equity-and-autarky.

8.3 Results and Discussion

The discussion of results employs system costs, the technology mix, as well as the distribution of power system infrastructure capacity expansion as evaluation criteria, both regarding distributional equity and autarky considerations.

¹Meanwhile, two relevant studies, published after the preprint of this chapter's article, have looked at cost impacts of regionally autarkic supply and balancing [154] and regional impacts in the context of a European power system model with high spatio-temporal detail [31].

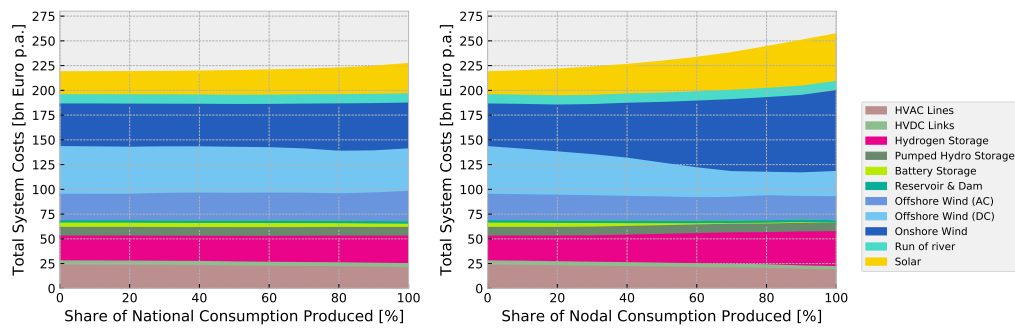


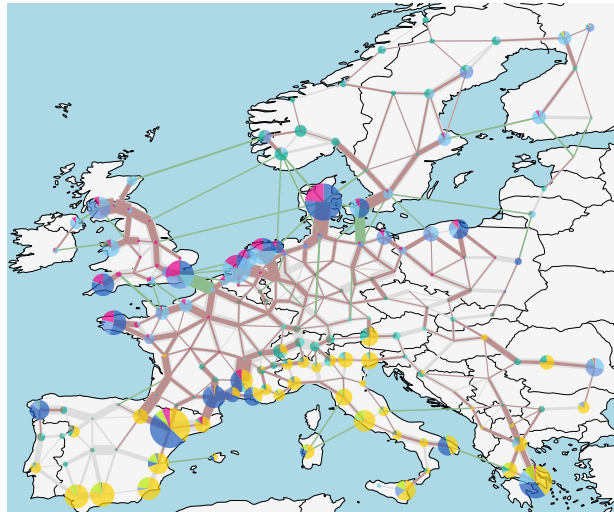
Figure 8.2: Sensitivity of system cost and composition to nodal and country-wide equity requirements.

Foremost, [Figure 8.2](#) displays the sensitivity of system costs towards nodal and country-wide equity requirements. Similar graphics were produced regarding the amount of cross-border transmission capacities by [Schlachtberger *et al.* \[37\]](#). National equity constraints cause a limited rise in the total system cost. The cost increase by less than 4% when every country produces as much as they consume; and by less than 2% when each produces at least 80%. They entail less grid reinforcement and some more solar installations. Conversely, the cost sensitivity is considerably higher for nodal equity constraints. When every node on average produces all they consume, costs inflate by 18%; and already at equity levels of 50% costs increase by 5%. Note, that the sensitivity is nonlinear. Nodal requirements shift expansion plans towards onshore wind, solar and hydrogen storage, while reducing network expansion and offshore wind capacities. This confirms but also extends on a finding by [Sasse & Trutnevyte \[166\]](#): indeed solar contributes to regional equity, but also onshore wind does. This is ambivalent since onshore wind is susceptible to local opposition.

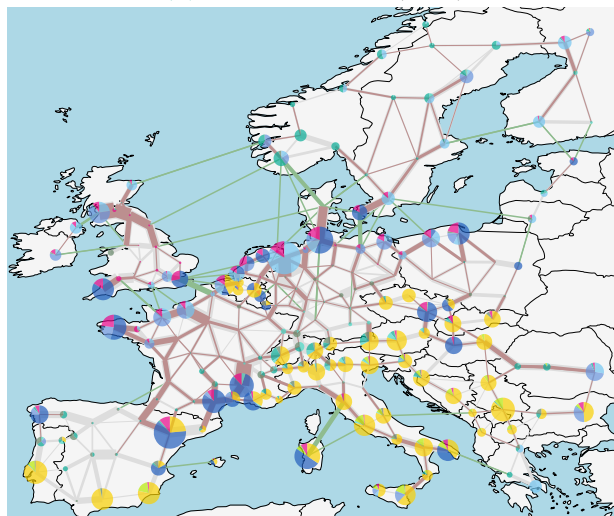
The maps of the optimised system capacities in [Figure 8.3](#) show less but still substantial amounts of transmission expansion in the case of nodal equity. Compared to the unrestricted least-cost solution, the deployment of solar panels progresses northbound and onshore wind capacities spread in Northern and Eastern Europe. Moreover, the storage infrastructure distributes more evenly.

[Figure 8.4](#) depicts Lorenz curves as equity measures for different equity constraints (see [\[166\]](#)). They relate the cumulative share of electricity generation to the cumulative share of demand in the 200 regions of the model. Like a load duration curve describes the share of time a particular level of electricity demand is exceeded, the Lorenz curve outlines the share of nodes where the ratio between total electricity generation and consumption exceeds a certain value. The Lorenz curve is on the identity line if annual sums of generation and load are equal at each node. While nodal equity requirements by definition lift the Lorenz curve, national requirements maintain an unequal distribution of infrastructure within each country.

(a) Least-Cost Solution (No Requirements)



(b) National Balance (100%)



(c) Nodal Balance (100%)

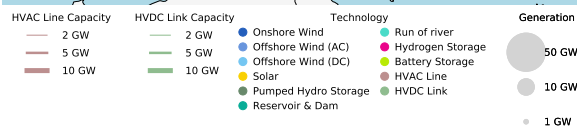
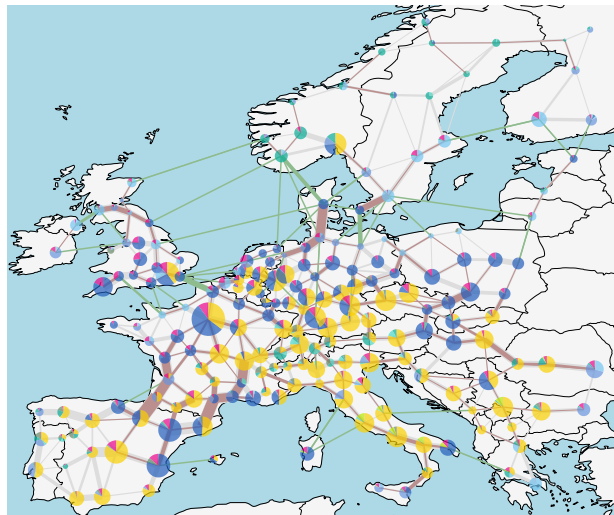


Figure 8.3: Maps of optimal system capacities for different equity requirements.

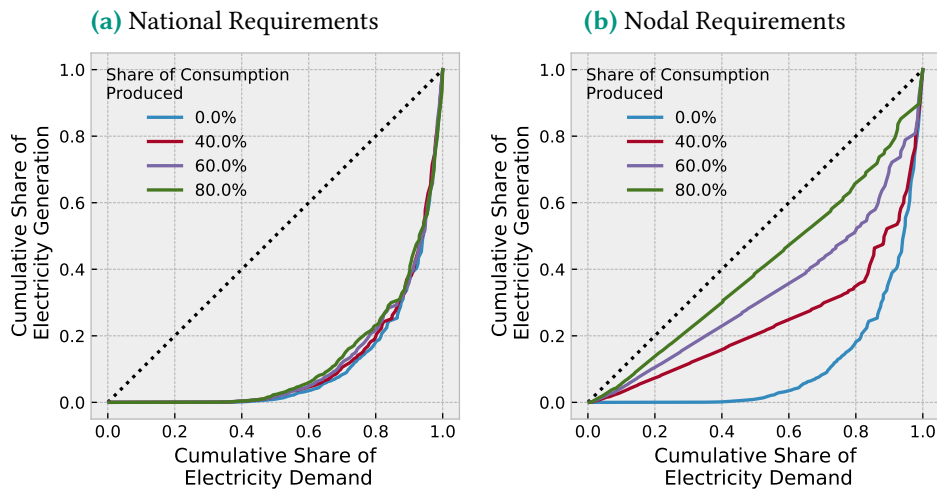


Figure 8.4: Lorenz curves for different equity requirements relating the cumulative share of electricity generation to the cumulative share of demand in the 200 regions of the European power system model.

Figure 8.5: National annual investment relative to annual demand when every country produces as much as they consume (100%).

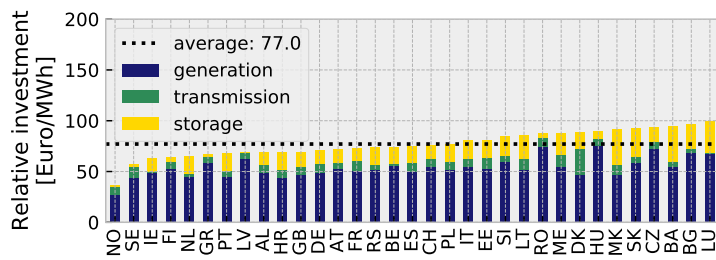
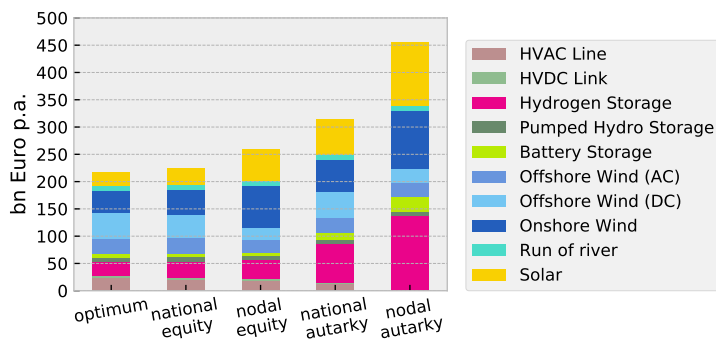


Figure 8.6: Total system cost impact of autarky on national and nodal levels compared to optimal solution and maximal equity constraints.



Results further show that, when every country balances generation and load on average, the national cost for capacity expansion relative to demand is more evenly distributed, ranging between 40 and 100 €/MWh (Figure 8.5). Generation infrastructure is the dominant component, followed by storage and transmission. Following conclusions from Li *et al.* [202], the shared burden of infrastructure build-out can be considered favourable for balanced regional economic development.

While even nodal production equity raises costs only to a limited extent below 20%, absolute autarky is significantly more costly already on a national level. Figure 8.6 shows that eliminating cross-border transmission capacities (i.e. no trade of power between countries) adds costs beyond 40%. Costs rise even more when each of the 200 regions is fully self-sufficient. With an additional 110%, costs more than doubled compared to minimising costs without equity requirements.

Figure 8.6 further shows that autarkic solutions compensate for the lack of power transmission options with extended deployment of hydrogen and also battery storage alongside additional onshore wind and solar capacities at locations with lower annual yields. Moreover, particularly for the nodal autarky scenario, the question arises whether there is sufficient renewable energy potential to cover demand. However, at least for the considered level of spatial aggregation, this concern can be dismissed as, assuming an even split between utility and rooftop PV, results reveal that no region uses more than 3% of its area for utility PV and only one in ten regions uses more than 1%. Nonetheless, landuse issues may become more critical if even smaller regions were modelled [210]. It further needs to be noted that hydrogen storage in salt caverns, as the cheaper alternative to steel tanks if the geological conditions admit it, was disregarded.

8.4 Critical Appraisal

In this chapter's evaluation, we focus on the mere cost perspective and do not assess whether these solutions would actually lead to higher social acceptance or whether they are preferable for any other reasons besides costs. One notable weakness of the simulation, stemming from computational constraints, is that they neglect the inherent uncertainty of technology cost projections, but for instance cheap battery storage and solar panels—technologies that are well suited to local use—may alleviate some of the cost penalty of self-sufficient solutions. In future work, the autarky analysis should be further expanded to a fully sector-coupled energy system which could include transport options for chemical energy carriers and opens the field to further flexibilities to locally balance supply and demand, possibly altering the cost impact of fully autarkic solutions.

8.5 Conclusion

Overall, it appears to be possible to strike a balance between cost-efficiency and balanced distribution of infrastructure at little additional expense. The analysis showed that aligning annual generation and consumption per country costs less than 4% more; per node, the costs increase by no more than 20%. National balancing, however, retains inhomogenous distributions within the countries, and even when each node produces as much as they consume, power is still extensively transmitted and regions are not self-sufficient. True autarky solutions without power transmission are substantially more expensive, nationally and even more so regionally, such that balancing via integrated power systems across regions and countries appears to remain vital component of low-cost future European power system designs.

Summary Part II

In [Chapter 6](#), we established that there exist many alternative solutions near the cost-optimum, where already a minor cost deviation below 1% opens a large range of possible technologically diverse investments to accommodate issues of social acceptance. However, wind energy, along with some hydrogen storage and transmission network reinforcement, appeared to be essential to keep costs within 10% of the optimum.

With [Chapter 7](#), we pursued the matter further by additionally taking account of uncertain technology cost assumptions and, thereby, ascertained that within the considered range of cost projections, the above-mentioned boundary conditions are robust. As the cost uncertainty propagates through the model to increase the vagueness of trade-offs between system cost and technology use, we could also quantify how they are affected by technology costs.

For instance, there is very little uncertainty about how much transmission expansion we can spare, whereas there is more ambiguity about how much onshore wind could be exchanged with offshore wind without significantly affecting the system's cost-effectiveness. Further insights regarded dependencies between technologies, i.a. between wind and solar generation, and how particularly battery storage comes into play once their cost falls below a certain threshold.

Finally, in [Chapter 8](#), we demonstrated that while purely cost-optimal solutions lead to very inhomogeneous distributions of capacities, more uniform expansion plans can be achieved nationally at little additional expense below 4%. However, completely autarkic solutions, without the use of electricity transmission, appear much more costly.

Recomposed from abstracts of papers forming this part's chapters [3, 4, 6].

Conclusions and Outlook

With rising shares of renewables and the need to properly assess trade-offs between transmission, storage and sectoral integration as balancing options, building a bridge between energy system models and detailed power flow studies becomes increasingly important, but is computationally challenging.

Trade-Offs Between Model Detail and Computational Speed

In the present thesis, we have investigated methods that allow us to increase the physical model detail of electricity transmission networks while remaining computationally performant. Considering a range of different linear approximations of power flow and transmission losses, we offered a decision aid for modellers who undertake capacity expansion studies. Thereby, we substantiated that by neglecting losses, cost-optimal levels of grid expansion are likely overestimated. We also emphasised that including linearised constraints for Kirchhoff's voltage law (KVL) and impedance changes is essential for capturing the basic physical conditions of transmission grids in design studies. This holds especially under high loading conditions.

In another chapter, we demonstrated performance improvements obtained without compromises in model detail but with a cycle decomposition of the network graph to efficiently reformulate the discrete transmission expansion planning (TEP) problem. This cycle-based formulation is also shown to integrate synchronisation options conveniently. With regards to commonly applied integer investment variables in TEP, we further established that continuous relaxation is justified as it removes the excessive computational burden of integer programming while yielding equally accurate solutions in light of typically tolerated optimality gaps in discrete optimisation and other, more decisive model condensations.

Trade-Offs Between Cost-Optimality and Near-Optimal Alternatives

Moreover, we addressed criticisms pertaining to a common focus of energy system design studies on single cost-optimal solutions. Along a number of different axes besides just economic efficiency, we explored alternative system layouts to build robust insights and intuitions about what actions are viable within given cost

ranges without being prescriptive. While taking account of the inherent uncertainties of technology cost projections, we identified a few boundary conditions that must be met for cost-efficiency. For instance, that some grid reinforcement and long-term storage alongside a significant amount of wind capacity is essential for building a fully renewable European electricity system. But, foremost, we reveal that near the cost-optimum a broad range of regionally and technologically diverse options exists, which provides policymakers with the building blocks to accommodate public acceptance issues.

Informatics Role and Contribution

Because computational performance limits what can be investigated, informatics becomes an increasingly important part of energy system modelling. It is methods from computer science that enable the high modelling detail required to improve our understanding of the drivers of cost-effective renewable energy system designs. Although this thesis is certainly a work between the disciplines, contributions to informatics manifest in several ways:

One part of the thesis had a modelling focus where we gauged various approximation errors of power flow, transmission losses, and grid reinforcement. In addition, we employed methods from graph theory, namely a cycle decomposition of the transmission network, in order to study how to formulate, more so than solve, the models at hand more efficiently. Thereby, we also addressed the combinatorics of connecting multiple synchronous zones.

The other part of the thesis followed a more applied direction, where the central question from an informatics perspective regarded how to efficiently compute many scenario runs. In combination with heavy parallelisation on high-performance clusters, we employed multi-fidelity surrogate modelling techniques with sparse polynomial chaos expansion (PCE) and efficient low-discrepancy sampling to comprehensively sweep the parameter space.

Ultimately, with all of the above analyses, we practised open-source modelling to provide accessible, transparent, and reproducible research results [61, 214], whereby the workflow management software *snakemake* was of great avail [82].

Enhancements and Outlook

Some ideas for further research have already been touched upon in the discussion of each chapter's limitations. However, bringing the thesis to a close, we briefly elaborate on the scope of future research in energy system modelling in a broader context and give a preview of what is to come next.

Computational Improvements with Heuristics and Decomposition

In light of the computational challenges of large-scale energy system planning and abundant expert knowledge about problem structures, further research on problem-specific heuristics, decomposition techniques, and algorithms should be undertaken to speed up the process of solving energy system optimisation problems.

Future studies could investigate the benefits of heuristics applied to energy system optimisation further, besides the heuristics for integer TEP in Chapter 5. For instance, further research might explore heuristics that quickly find good initial feasible solutions by leveraging problem-specific knowledge.

Also, more research could be conducted on decomposition techniques that exploit the typical sparsity and near block-angular structures of the capacity expansion model's coefficient matrix [56, 59, 215]. This could allow for much higher degrees of parallelisation and shorter computation times. Thereby, numerous closely related decomposition schemes, amongst others Benders decomposition (BD) and Dantzig-Wolfe decomposition (DWD), could be examined [216, 217]. Particularly, Lagrangean Decomposition (variable splitting) [218, 219] in combination with use of the alternating direction method of multipliers (ADMM) [220, 221] could be fruitful because capacity expansion models (CEMs) with high shares of renewables and storage feature intertemporal coupling both through investment *variables* and storage consistency *constraints*.

Moreover, for pure dispatch problems without investment decisions (e.g. coupled electricity markets based on net transfer capacities (NTCs) or subproblems in decomposition applications), future research may examine network flow algorithms as an alternative to interior-point and simplex methods; for instance, extensions of the minimum cost flow algorithm [111].

Sector Coupling

Among the most frequently mentioned drawbacks mentioned above is that all of the above analyses limited the scope to the European electricity sector and, thereby, neglected global aspects of the energy transition just as they disregarded the coupling with other energy sectors. However, to achieve ambitious sustainability goals, it is inevitable to decarbonise the other energy-consuming sectors, such as transportation, heating, industry, aviation and shipping, as well. Because many concepts for decarbonising those sectors involve electrification, an integrated system perspective gains in importance.

Thereby, new challenges arise from a modelling perspective. Besides the need to consider a much broader range of technologies, which further inflates the problem

size beyond the dimensions of spatial and temporal resolution, there are many more interdependencies to be captured, which require careful tracking of conversion paths between multiple energy carriers and carbon dioxide [25, 222]. Moreover, in light of growing demand in renewable energy generation, the impetus may grow to leverage international trade in zero-emission electricity or other fuels like hydrogen, ammonia or synthetic hydrocarbons.

All this reiterates that, especially when expanding the sectoral and spatial scope, we must continue to abstract from reality in numerous further domains to retain control of the computational complexity while keeping a tab on the approximation errors introduced. While the insights about simplifications and the way we formulate constraints of electricity transmission will prevail with increased sectoral integration, there is scope for transferring them to the modelling of other sectors. For example, as both power flow and gas flow are nonlinear phenomena, gas network modelling raises similar questions as electricity network modelling.

Pathway Optimisation

Moreover, while the thesis answered questions about *what* infrastructure could be build *where* under given policy goals, it left out *when* to build it. Which path to follow to achieve climate change mitigation targets across all energy sectors is a question of particular relevance. In view of many pending renewals of long-lived power plants and industrial production sites and the looming danger of lock-in effects and stranded assets it is strongly advisable to take account of pathways, i.e. considering the timing of investments across the upcoming decades.

Additionally, the opportunity to, thereby, also model the benefits gained by induced technological learning, i.e. exploiting how costs of a technology fall as its capacities are expanded, speaks in favour of adding this new dimension to the optimisation problem [180].

Because it also exhibits promising block-like structures, pathway optimisation also opens new windows for investigating the application of decomposition techniques to counteract the added computational burden.

Reliability

Besides economically achieving sustainability goals, it is another requirement to guarantee that energy services can be met whenever they are demanded. Modelling of reliability can further improve in several ways. Among other things, basing the optimisation on a single reference weather year may not result in an adequate

system design that can sustain extreme weather events or even a changing climate. The availability of decades of historical weather data lends itself to further strengthening the robustness towards interannual weather variations.

Another aspect of reliability where modelling can improve concerns the network operation. Because the consideration of the $N - 1$ security criterion (i.e. the condition that the failure of any one circuit may not cause any secondary outages) involves many what-if constraints, there is scope for further research on efficiently formulating such preventive measures by using algorithms from graph theory, heuristics, or some form of redundant constraint screening [223, 224].

Along these lines, it is also a worthwhile undertaking to evaluate the cost of novel curative measures like grid boosters, that compensate outages for brief transitional periods, compared to widespread preventative security margins on the transmission line capacities.

In addition to network security constraints, one may also evaluate the impact of generator failures more rigorously, e.g. due to the outage of a fossil plant or forecast deviations of renewable generators. Here, future research could look at reserve markets and the role of energy storage and demand-side management in it.

Further Trade-Offs

Finally, in [Part II](#) we focussed on trade-offs between system cost and the capacities of various technologies and their regional distribution. However, it should not be overlooked that there are many other trade-offs decision makers may have an interest in. These may evolve around the employment implications of expansion plans, the life cycle impact of different technologies, effects on air pollution, the sustainable use of biomass and nuclear energy, the ramifications on regional land-use, or even multiple such aspects at once and across different energy sectors. Consequently, future research may investigate further factors that resonate with the priorities of policymakers to inform their decision making by presenting an even wider menu of options to choose from.

As evermore complicated trade-offs will need to be assessed, which all require detailed modelling, computational aspects become a key success factor in energy system modelling. This thesis has leveraged methods from informatics to manage the computational challenges faced, with the ultimate goal to provide society with useful information about the choices available as we transform the energy system.

Bibliography

1. Neumann, F. & Brown, T. *Transmission Expansion Planning Using Cycle Flows in Proceedings of the Eleventh ACM International Conference on Future Energy Systems (ACM e-Energy 2020)* (2020), 253–263. doi:[10/d3qk](https://doi.org/10/d3qk). arXiv: [2004.08702](https://arxiv.org/abs/2004.08702) (cit. on pp. [ii](#), [vii](#), [7](#), [25](#), [39](#), [57](#), [95](#)).
2. Neumann, F. & Brown, T. *Heuristics for Transmission Expansion Planning in Low-Carbon Energy System Models in 16th International Conference on the European Energy Market* (2019), 1–8. doi:[10/d295](https://doi.org/10/d295). arXiv: [1907.10548](https://arxiv.org/abs/1907.10548) (cit. on pp. [ii](#), [vii](#), [7](#), [25](#), [39](#), [79](#), [95](#), [105](#)).
3. Neumann, F. & Brown, T. The Near-Optimal Feasible Space of a Renewable Power System Model. *Electric Power Systems Research* **190**, 106690. doi:[10/ghcpr2](https://doi.org/10/ghcpr2). arXiv: [1910.01891](https://arxiv.org/abs/1910.01891) (2021) (cit. on pp. [vii](#), [7](#), [99](#), [101](#), [117](#), [130](#), [145](#)).
4. Neumann, F. Costs of Regional Equity and Autarky in a Renewable European Power System. *Energy Strategy Reviews* **35**, 100652. doi:[10/gjr2fb](https://doi.org/10/gjr2fb). arXiv: [2007.08379](https://arxiv.org/abs/2007.08379) (2021) (cit. on pp. [vii](#), [8](#), [99](#), [137](#), [145](#)).
5. Neumann, F., Hagenmeyer, V. & Brown, T. Approximating Power Flow and Transmission Losses in Coordinated Capacity Expansion Problems. arXiv: [2008.11510](https://arxiv.org/abs/2008.11510) (2020) (cit. on pp. [vii](#), [6](#), [10](#), [25](#), [27](#), [95](#)).
6. Neumann, F. & Brown, T. Broad Ranges of Investment Configurations for Renewable Power Systems, Robust to Cost Uncertainty and Near-Optimality. eprint: <https://bit.ly/3fxGEAb> (2021) (cit. on pp. [vii](#), [8](#), [99](#), [115](#), [145](#)).
7. Rose, P. K. & Neumann, F. Hydrogen refueling station networks for heavy-duty vehicles in future power systems. *Transportation Research Part D: Transport and Environment* **83**, 102358. doi:[10/d7tn](https://doi.org/10/d7tn). arXiv: [1908.10119](https://arxiv.org/abs/1908.10119) (2020) (cit. on p. [vii](#)).
8. Sun, W., Neumann, F. & Harrison, G. P. Robust scheduling of Electric Vehicle Charging in LV Distribution Networks under Uncertainty. *IEEE Transactions on Industry Applications* **56**, 5785–5795. doi:[10/fccs](https://doi.org/10/fccs) (2020) (cit. on p. [viii](#)).
9. Hofmann, F., Hampp, J., Neumann, F., Brown, T. & Hörsch, J. atlite: A Lightweight Python Package for Calculating Renewable Power Potentials and Time Series. *Journal of Open Source Software* **6**, 3294. doi:[10/gn4b](https://doi.org/10/gn4b) (2021) (cit. on pp. [viii](#), [19](#)).

10. Parzen, M., Neumann, F., van der Weijde, A. H., Friedrich, D. & Kiprakis, A. Beyond cost reduction: Improving the value of energy storage in electricity systems. arXiv: [2101.10092](https://arxiv.org/abs/2101.10092) (2021) (cit. on p. [viii](#)).
11. Gazafroudi, A. S., Neumann, F. & Brown, T. Topology-based Approximations for $N - 1$ Contingency Constraints (2021) (cit. on p. [viii](#)).
12. Brown, T., Hörsch, J. & Schlachtberger, D. PyPSA : Python for Power System Analysis. *Journal of Open Research Software* **6**, 4. doi:[10/gfb7m9](https://doi.org/10/gfb7m9). arXiv: [1707.09913](https://arxiv.org/abs/1707.09913) (2018) (cit. on pp. [viii](#), [3](#), [12](#), [41](#), [59](#), [74](#), [86](#)).
13. Hörsch, J., Hofmann, F., Schlachtberger, D. & Brown, T. PyPSA-Eur: An open optimisation model of the European transmission system. *Energy Strategy Reviews* **22**, 207–215. doi:[10/d294](https://doi.org/10/d294). arXiv: [1806.01613](https://arxiv.org/abs/1806.01613) (2018) (cit. on pp. [viii](#), [3](#), [15](#), [19](#), [28](#), [40](#), [59](#), [73](#), [86](#), [105](#), [118](#), [139](#)).
14. Ringkjøb, H.-K., Haugan, P. M. & Solbrekke, I. M. A review of modelling tools for energy and electricity systems with large shares of variable renewables. *Renewable and Sustainable Energy Reviews* **96**, 440–459. doi:[10/gfgb5g](https://doi.org/10/gfgb5g) (2018) (cit. on pp. [1](#), [3](#), [27](#), [33](#)).
15. Jenkins, J. D., Luke, M. & Thernstrom, S. Getting to Zero Carbon Emissions in the Electric Power Sector. *Joule*, 1–12. doi:[10/gf6bh7](https://doi.org/10/gf6bh7) (2018) (cit. on p. [1](#)).
16. Groissböck, M. Are open source energy system optimization tools mature enough for serious use? *Renewable and Sustainable Energy Reviews* **102**, 234–248. doi:[10/gfs3rj](https://doi.org/10/gfs3rj) (2019) (cit. on p. [1](#)).
17. Oberle, S. & Elsland, R. Are open access models able to assess today’s energy scenarios? *Energy Strategy Reviews* **26**, 100396. doi:[10/gf59qp](https://doi.org/10/gf59qp) (2019) (cit. on p. [1](#)).
18. IPCC. *Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* tech. rep. (2018). eprint: <https://www.ipcc.ch/sr15/> (cit. on p. [1](#)).
19. Brown, T., Bischof-Niemz, T., Blok, K., Breyer, C., Lund, H. & Mathiesen, B. Response to ‘Burden of proof: A comprehensive review of the feasibility of 100% renewable-electricity systems’. *Renewable and Sustainable Energy Reviews* **92**, 834–847. doi:[10/gdwmvw](https://doi.org/10/gdwmvw). arXiv: [1709.05716](https://arxiv.org/abs/1709.05716) (2018) (cit. on pp. [1](#), [40](#)).
20. German Federal Network Agency (Bundesnetzagentur). *Monitoring Report 2020* https://www.bundesnetzagentur.de/SharedDocs/Mediathek/Berichte/2020/Monitoringbericht_Energie2020.pdf (cit. on p. [1](#)).

21. Eriksen, E. H., Schwenk-Nebbe, L. J., Tranberg, B., Brown, T. & Greiner, M. Optimal heterogeneity in a simplified highly renewable European electricity system. *Energy* **133**, 913–928. doi:[10/gbwppw](https://doi.org/10/gbwppw). arXiv: [1706.00463](https://arxiv.org/abs/1706.00463) (2017) (cit. on pp. [2](#), [138](#)).
22. Levi, P. J., Kurland, S. D., Carbajales-Dale, M., Weyant, J. P., Brandt, A. R. & Benson, S. M. Macro-Energy Systems: Toward a New Discipline. *Joule* **3**, 2282–2286. doi:[10/ghbkx3](https://doi.org/10/ghbkx3) (2019) (cit. on p. [2](#)).
23. Biggar, D. R. & Hesamzadeh, M. R. *The Economics of Electricity Markets* doi:[10/f3nmn9](https://doi.org/10/f3nmn9) (John Wiley & Sons, 2014) (cit. on p. [2](#)).
24. Krishnan, V., Ho, J., Hobbs, B. F., Liu, A. L., McCalley, J. D., Shahidehpour, M. & Zheng, Q. P. Co-optimization of electricity transmission and generation resources for planning and policy analysis: review of concepts and modeling approaches. *Energy Systems* **7**, 297–332. doi:[10/d3bd](https://doi.org/10/d3bd) (2016) (cit. on pp. [2](#), [25](#), [57](#), [60](#), [63](#), [81](#)).
25. Brown, T., Schlachtberger, D., Kies, A., Schramm, S. & Greiner, M. Synergies of sector coupling and transmission reinforcement in a cost-optimised, highly renewable European energy system. *Energy* **160**, 720–739. doi:[10/gfdstr](https://doi.org/10/gfdstr). arXiv: [1801.05290](https://arxiv.org/abs/1801.05290) (2018) (cit. on pp. [2](#), [114](#), [150](#)).
26. Kotzur, L., Markewitz, P., Robinius, M. & Stolten, D. Impact of different time series aggregation methods on optimal energy system design. *Renewable Energy* **117**, 474–487. doi:[10/d3bc](https://doi.org/10/d3bc). arXiv: [1708.00420](https://arxiv.org/abs/1708.00420) (2018) (cit. on pp. [3](#), [25](#), [86](#)).
27. Kotzur, L., Markewitz, P., Robinius, M. & Stolten, D. Time series aggregation for energy system design: Modeling seasonal storage. *Applied Energy* **213**, 123–135. doi:[10/gc5jn8](https://doi.org/10/gc5jn8). arXiv: [1710.07593](https://arxiv.org/abs/1710.07593) (2018) (cit. on p. [3](#)).
28. Hörsch, J. & Brown, T. The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios. *14th International Conference on the European Energy Market*, 1–8. arXiv: [1705.07617](https://arxiv.org/abs/1705.07617) (2017) (cit. on pp. [3](#), [21](#), [25](#), [32](#), [59](#), [105](#), [108](#), [114](#)).
29. Frysztacki, M. M., Hörsch, J., Hagenmeyer, V. & Brown, T. The strong effect of network resolution on electricity system models with high shares of wind and solar. arXiv: [2101.10859](https://arxiv.org/abs/2101.10859) (2021) (cit. on pp. [3](#), [21](#), [25](#), [32](#), [105](#), [114](#)).
30. Lumbreras, S., Ramos, A. & Banez Chicharro, F. Optimal transmission network expansion planning in real-sized power systems with high renewable penetration. *Electric Power Systems Research* **149**, 76–88. doi:[10/gbqp9m](https://doi.org/10/gbqp9m) (2017) (cit. on pp. [3](#), [25](#), [27](#), [58](#), [64](#), [83](#)).
31. Sasse, J.-P. & Trutnevyte, E. Regional impacts of electricity system transition in Central Europe until 2035. *Nature Communications* **11**, 4972. doi:[10/ghdwjs](https://doi.org/10/ghdwjs) (2020) (cit. on pp. [3](#), [99](#), [103](#), [139](#)).

32. Haller, M., Ludig, S. & Bauer, N. Decarbonization scenarios for the EU and MENA power system: Considering spatial distribution and short term dynamics of renewable generation. *Energy Policy* **47**, 282–290. doi:10/f35ntv (2012) (cit. on p. 3).
33. Nahmmacher, P., Schmid, E., Hirth, L. & Knopf, B. Carpe diem: A novel approach to select representative days for long-term power system modeling. *Energy* **112**, 430–442. doi:10/f88hnr (2016) (cit. on p. 3).
34. Reichenberg, L. & Hedenus, F. The Error induced by Using Representative Periods in Capacity Expansion Models – System Cost, Total Capacity Mix and Regional Capacity Mix. Preprints: <https://www.preprints.org/manuscript/202012.0229/v1> (2020) (cit. on p. 3).
35. Gils, H. C., Scholz, Y., Pregger, T., Luca de Tena, D. & Heide, D. Integrated modelling of variable renewable energy-based power supply in Europe. *Energy* **123**, 173–188. doi:10/f95f3x (2017) (cit. on p. 3).
36. Child, M., Kemfert, C., Bogdanov, D. & Breyer, C. Flexible electricity generation, grid exchange and storage for the transition to a 100% renewable energy system in Europe. *Renewable Energy* **139**, 80–101. doi:10/gf8fz3 (2019) (cit. on p. 3).
37. Schlachtberger, D. P., Brown, T., Schramm, S. & Greiner, M. The Benefits of Cooperation in a Highly Renewable European Electricity Network. *Energy* **134**, 469–481. doi:10/gbvvcf. arXiv: 1704.05492 (2017) (cit. on pp. 3, 17, 102, 108, 138, 140).
38. Zappa, W., Junginger, M. & van den Broek, M. Is a 100% renewable European power system feasible by 2050? *Applied Energy* **233-234**, 1027–1050. doi:10/gf8fz5 (2019) (cit. on p. 3).
39. Stockholm Environment Institute. *LEAP: The Low Emissions Analysis Platform* <https://leap.sei.org/default.asp> (cit. on p. 3).
40. IEA-ETSAP. *TIMES: The Integrated MARKAL-EFOM System* <https://iea-etsap.org/index.php/etsap-tools/model-generators/times> (cit. on p. 3).
41. Pfenninger, S. & Pickering, B. Calliope: A multi-scale energy systems modelling framework. *Journal of Open Source Software* **3**, 825. doi:10/f6tz (2018) (cit. on p. 3).
42. Hilpert, S., Kaldemeyer, C., Krien, U., Günther, S., Wingenbach, C. & Plessmann, G. The Open Energy Modelling Framework (oemof) - A new approach to facilitate open science in energy system modelling. *Energy Strategy Reviews* **22**, 16–25. doi:10/gjgxs3 (2018) (cit. on p. 3).
43. Howells, M., Rogner, H., Strachan, N., Heaps, C., Huntington, H., Kypreos, S., Hughes, A., Silveira, S., DeCarolis, J., Bazillian, M. & Roehrl, A. OSeMOSYS: The Open Source Energy Modeling System: An introduction to its ethos, structure and development. *Energy Policy* **39**, 5850–5870. doi:10/b8mqfv (2011) (cit. on p. 3).

44. Hunter, K., Sreepathi, S. & DeCarolis, J. F. Modeling for insight using Tools for Energy Model Optimization and Analysis (Temoa). *Energy Economics* **40**, 339–349. doi:[10/f5m5f4](https://doi.org/10/f5m5f4) (2013) (cit. on p. 3).
45. Thurner, L., Scheidler, A., Schäfer, F., Menke, J., Dollichon, J., Meier, F., Meinecke, S. & Braun, M. pandapower – An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems. *IEEE Transactions on Power Systems* **33**, 6510–6521. doi:[10/gfmh2m](https://doi.org/10/gfmh2m) (2018) (cit. on p. 3).
46. Lincoln, R. PYPOWER <https://github.com/rwl/PYPOWER> (cit. on p. 3).
47. Coffrin, C., Bent, R., Sundar, K., Ng, Y. & Lubin, M. *PowerModels.jl: An Open-Source Framework for Exploring Power Flow Formulations in 2018 Power Systems Computation Conference (PSCC)* (2018). doi:[10/f6t2](https://doi.org/10/f6t2) (cit. on p. 3).
48. Sørensen, B. Energy and Resources - A plan is outlined according to which solar and wind energy would supply Denmark's needs by the year 2050. *Science* **189**. doi:[10/c7ptsc](https://doi.org/10/c7ptsc) (1975) (cit. on p. 3).
49. Sørensen, B. On the fluctuating power generation of large wind energy converters, with and without storage facilities. *Solar Energy* **20**, 321–331. doi:[10/cw9jmr](https://doi.org/10/cw9jmr) (1978) (cit. on p. 3).
50. Jaxa-Rozen, M. & Trutnevyte, E. Sources of uncertainty in long-term global scenarios of solar photovoltaic technology. *Nature Climate Change* **11**, 266–273. doi:[10/fzt4](https://doi.org/10/fzt4) (2021) (cit. on pp. 3, 115).
51. Victoria, M., Haegel, N., Peters, I. M., Sinton, R., Jäger-Waldau, A., del Cañizo, C., Breyer, C., Stocks, M., Blakers, A., Kaizuka, I., Komoto, K. & Smets, A. Solar photovoltaics is ready to power a sustainable future. *Joule*, S2542435121001008. doi:[10/gjk5bx](https://doi.org/10/gjk5bx) (2021) (cit. on pp. 3, 115).
52. Xiao, M., Junne, T., Haas, J. & Klein, M. Plummeting costs of renewables - Are energy scenarios lagging? *Energy Strategy Reviews* **35**, 100636. doi:[10/gjg5sx](https://doi.org/10/gjg5sx) (2021) (cit. on pp. 3, 115).
53. Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. & Thépaut, J.-N. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**, 1999–2049. doi:[10/gg9wx7](https://doi.org/10/gg9wx7) (2020) (cit. on pp. 3, 19).
54. *Gurobi Optimization* <https://www.gurobi.com/> (cit. on pp. 3, 14).

55. Pfenninger, S., Hawkes, A. & Keirstead, J. Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews* **33**, 74–86. doi:[10/f526xf](https://doi.org/10/f526xf) (2014) (cit. on pp. 3, 101).
56. Cao, K.-K., von Krbeke, K., Wetzel, M., Cebulla, F. & Schreck, S. Classification and Evaluation of Concepts for Improving the Performance of Applied Energy System Optimization Models. *Energies* **12**, 4656. doi:[10/ggd8rf](https://doi.org/10/ggd8rf) (2019) (cit. on pp. 3, 149).
57. Ridha, E., Nolting, L. & Praktiknjo, A. Complexity profiles: A large-scale review of energy system models in terms of complexity. *Energy Strategy Reviews* **30**, 100515. doi:[10/gg3dmv](https://doi.org/10/gg3dmv) (2020) (cit. on p. 3).
58. Bienstock, D. & Verma, A. Strong NP-hardness of AC power flows feasibility. *Operations Research Letters* **47**, 494–501. doi:[10/d299](https://doi.org/10/d299). arXiv: [1512.07315v1](https://arxiv.org/abs/1512.07315v1) (2019) (cit. on pp. 4, 27, 31).
59. Kotzur, L., Nolting, L., Hoffmann, M., Groß, T., Smolenko, A., Priesmann, J., Büsing, H., Beer, R., Kullmann, F., Singh, B., Praktiknjo, A., Stolten, D. & Robinius, M. A modeler's guide to handle complexity in energy system optimization, 32. arXiv: [2009.07216](https://arxiv.org/abs/2009.07216) (2020) (cit. on pp. 4, 149).
60. Siala, K., Mier, M., Schmidt, L., Torralba-Díaz, L., Sheykkha, S. & Savvidis, G. Which model features matter? An experimental approach to evaluate power market modeling choices. arXiv: [2010.16142](https://arxiv.org/abs/2010.16142) (2020) (cit. on p. 4).
61. DeCarolis, J., Daly, H., Dodds, P., Keppo, I., Li, F., McDowall, W., Pye, S., Strachan, N., Trutnevyte, E., Usher, W., Winning, M., Yeh, S. & Zeyringer, M. Formalizing best practice for energy system optimization modelling. *Applied Energy* **194**, 184–198. doi:[10/f96rdx](https://doi.org/10/f96rdx) (2017) (cit. on pp. 4, 101, 148).
62. Pye, S., Broad, O., Bataille, C., Brockway, P., Daly, H. E., Freeman, R., Gambhir, A., Geden, O., Rogan, F., Sanghvi, S., Tomei, J., Vorushylo, I. & Watson, J. Modelling net-zero emissions energy systems requires a change in approach. *Climate Policy*, 1–10. doi:[10/ghdkp9](https://doi.org/10/ghdkp9) (2020) (cit. on p. 4).
63. McCollum, D. L., Gambhir, A., Rogelj, J. & Wilson, C. Energy modellers should explore extremes more systematically in scenarios. *Nature Energy* **5**, 104–107. doi:[10/ggk3hj](https://doi.org/10/ggk3hj) (2020) (cit. on pp. 4, 119).
64. Trutnevyte, E. Does cost optimization approximate the real-world energy transition? *Energy* **106**, 182–193. doi:[10/f8vrcz](https://doi.org/10/f8vrcz) (2016) (cit. on pp. 4, 103).
65. Trutnevyte, E., McDowall, W., Tomei, J. & Keppo, I. Energy scenario choices: Insights from a retrospective review of UK energy futures. *Renewable and Sustainable Energy Reviews* **55**, 326–337. doi:[10/gh276b](https://doi.org/10/gh276b) (2016) (cit. on p. 4).

66. Huntington, H. G., Weyant, J. P. & Sweeney, J. L. Modeling for insights, not numbers: the experiences of the energy modeling forum. *Omega* **10**, 449–462. doi:[10/dwm87g](https://doi.org/10/dwm87g) (1982) (cit. on p. 4).
67. Silvast, A., Laes, E., Abram, S. & Bombaerts, G. What do energy modellers know? An ethnography of epistemic values and knowledge models. *Energy Research & Social Science* **66**, 101495. doi:[10/gg8zmc](https://doi.org/10/gg8zmc) (2020) (cit. on p. 4).
68. Hanna, R. & Gross, R. How do energy systems model and scenario studies explicitly represent socio-economic, political and technological disruption and discontinuity? Implications for policy and practitioners. *Energy Policy* **149**, 111984. doi:[10/ghk7g4](https://doi.org/10/ghk7g4) (2021) (cit. on p. 9).
69. Hörsch, J., Ronellenfitsch, H., Witthaut, D. & Brown, T. Linear optimal power flow using cycle flows. *Electric Power Systems Research* **158**, 126–135. doi:[10/gdb8kx](https://doi.org/10/gdb8kx). arXiv: [1704.01881](https://arxiv.org/abs/1704.01881) (2018) (cit. on pp. 12, 32, 35, 58–60, 62).
70. Kuhn, H. & Tucker, A. *Nonlinear Programming in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (1951) (cit. on p. 13).
71. Brown, T. & Reichenberg, L. Decreasing market value of variable renewables is a result of policy, not variability. arXiv: [2002.05209](https://arxiv.org/abs/2002.05209) (2020) (cit. on p. 13).
72. Boyd, S. & Vandenberghe, L. *Convex Optimization* 1st ed. (Cambridge University Press, 2004) (cit. on p. 13).
73. Dantzig, G. B. Maximization of a linear function of variables subject to linear inequalities. *Activity analysis of production and allocation* **13**, 339–347 (1951) (cit. on p. 13).
74. Klee, V. & Minty, G. J. How good is the simplex algorithm. *Inequalities* **3**, 159–175 (1972) (cit. on p. 14).
75. Karmarkar, N. A new polynomial-time algorithm for linear programming. *Combinatorica* **4**, 373–395. doi:[10/czqmxn](https://doi.org/10/czqmxn) (1984) (cit. on p. 14).
76. Allamigeon, X., Benchimol, P., Gaubert, S. & Joswig, M. Log-Barrier Interior Point Methods Are Not Strongly Polynomial. *SIAM Journal on Applied Algebra and Geometry* **2**, 140–178. doi:[10/f5rw](https://doi.org/10/f5rw) (2018) (cit. on p. 14).
77. Land, A. H. & Doig, A. G. An Automatic Method of Solving Discrete Programming Problems. *Econometrica* **28**, 497. doi:[10/dh7zmm](https://doi.org/10/dh7zmm) (1960) (cit. on p. 14).
78. Dakin, R. J. A tree-search algorithm for mixed integer programming problems. *The Computer Journal* **8**, 250–255. doi:[10/fbpbk4f](https://doi.org/10/fbpbk4f) (1965) (cit. on p. 14).

79. Forrest, J., Ralphs, T., Vigerske, S., LouHafer, Kristjansson, B., Jpfasano, Edwin-Straver, Lubin, M., Santos, H. G., Rlougee & Saltzman, M. *Coin-Or/Cbc* doi:10.5281/zenodo.1317566 (cit. on p. 14).
80. *GNU Linear Programming Kit* <http://www.gnu.org/software/glpk/glpk.html> (cit. on p. 14).
81. *IBM ILOG CPLEX Optimization Studio* <https://www.ibm.com/products/ilog-cplex-optimization-studio> (cit. on p. 14).
82. Köster, J. & Rahmann, S. Snakemake – a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522. doi:10/gd2xzq (2012) (cit. on pp. 15, 148).
83. ENTSO-E. *ENTSO-E Transmission System Map* <https://www.entsoe.eu/data/map/> (Feb. 26, 2021) (cit. on p. 15).
84. Wiegmans, B. *GridKit* version v1.0. Mar. 2016. doi:10.5281/zenodo.47263 (cit. on p. 15).
85. ENTSO-E. *TYNDP: Ten-Year Network Development Plan* <https://tyndp.entsoe.eu/tyndp2018/> (Feb. 24, 2020) (cit. on pp. 17, 105).
86. Brown, T., Tröster, E., Schierhorn, P.-P. & Ackermann, T. Optimising the European transmission system for 77% renewable electricity by 2030. *IET Renewable Power Generation* **10**, 3–9. doi:10/f75tr3 (2016) (cit. on pp. 17, 28, 49).
87. Gotzens, F., Heinrichs, H., Hörsch, J. & Hofmann, F. Performing energy modelling exercises in a transparent way - The issue of data quality in power plant databases. *Energy Strategy Reviews* **23**, 1–12. doi:10/fzt3 (2019) (cit. on p. 17).
88. Ryberg, D., Robinius, M. & Stolten, D. Evaluating Land Eligibility Constraints of Renewable Energy Sources in Europe. *Energies* **11**, 1246. doi:10/gdtdrx (May 2018) (cit. on p. 18).
89. Pfeifroth, U., Kothe, S., Müller, R., Trentmann, J., Hollmann, R., Fuchs, P. & Werscheck, M. *Surface Radiation Data Set - Heliosat (SARAH) - Edition 2* Satellite Application Facility on Climate Monitoring, 2017. doi:10/f77h (cit. on p. 19).
90. Coker, P. J., Bloomfield, H. C., Drew, D. R. & Brayshaw, D. J. Interannual weather variability and the challenges for Great Britain's electricity market design. *Renewable Energy* **150**, 509–522. doi:10/ghfgtv (2020) (cit. on p. 19).
91. Hoffmann, M., Kotzur, L., Stolten, D. & Robinius, M. A Review on Time Series Aggregation Methods for Energy System Models. *Energies* **13**. doi:10/ggzkmd (2020) (cit. on p. 22).

92. Lumbreras, S., Ramos, A., Banez Chicharro, F., Olmos, L., Panciatici, P., Pache, C. & Maeght, J. Large-scale transmission expansion planning: from zonal results to a nodal expansion plan. *IET Generation, Transmission & Distribution* **11**, 2778–2786. doi:[10/gbxdsr](https://doi.org/10/gbxdsr) (2017) (cit. on p. 25).
93. Hagspiel, S., Jägemann, C., Lindenberger, D., Brown, T., Cherevatskiy, S. & Tröster, E. Cost-optimal power system extension under flow-based market coupling. *Energy* **66**, 654–666. doi:[10/f5zsq4](https://doi.org/10/f5zsq4) (2014) (cit. on pp. 25, 28, 79, 85).
94. Nolden, C., Schonfelder, M., Esser-Frey, A., Bertsch, V. & Fichtner, W. Network constraints in techno-economic energy system models: towards more accurate modeling of power flows in long-term energy system models. *Energy Systems* **4**, 267–287. doi:[10/ggk2h](https://doi.org/10/ggk2h) (2013) (cit. on pp. 27, 36).
95. Lehmann, K., Grastien, A. & Hentenryck, P. V. AC-feasibility on tree networks is NP-hard. *IEEE Transactions on Power Systems* **31**, 798–801. doi:[10/d3bb](https://doi.org/10/d3bb) (2016) (cit. on pp. 27, 31).
96. Sojoudi, S. & Lavaei, J. *Physics of power networks makes hard optimization problems easy to solve* in *IEEE Power and Energy Society General Meeting* (2012), 1–8. doi:[10/ggbgnc](https://doi.org/10/ggbgnc) (cit. on pp. 27, 31).
97. Nijs, W., Simoes, S., Sgobbi, A., Ruiz-Castello, P., Thiel, C., Giannakidis, G., Mantzaris, J., Tigas, K., Dimitroulas, D., Georgilakis, P. & Vournas, C. *Improved Representation of the European Power Grid in Long Term Energy System Models: Case Study of JRC-EU-TIMES in Informing Energy and Climate Policies Using Energy Systems Models* **30** (2015), 201–222. doi:[10/d297](https://doi.org/10/d297) (cit. on p. 27).
98. Haller, M., Ludig, S. & Bauer, N. Bridging the scales: A conceptual model for co-ordinated expansion of renewable power generation, transmission and storage. *Renewable and Sustainable Energy Reviews* **16**, 2687–2695. doi:[10/f335pp](https://doi.org/10/f335pp) (2012) (cit. on p. 27).
99. Gunkel, P. A., Koduvere, H., Kirkerud, J. G., Fausto, F. J. & Ravn, H. Modelling transmission systems in energy system analysis: A comparative study. *Journal of Environmental Management* **262**, 110289. doi:[10/ggq4mb](https://doi.org/10/ggq4mb) (2020) (cit. on pp. 27, 28, 47).
100. Schaber, K. *Integration of Variable Renewable Energies in the European power system: a model-based analysis of transmission grid extensions and energy sector coupling* PhD thesis (Technical University Munich, 2013). eprint: <https://mediatum.ub.tum.de/doc/1163646/document.pdf> (cit. on pp. 27, 28).

101. Fitiwi, D. Z., Olmos, L., Rivier, M., de Cuadra, F. & Pérez-Arriaga, I. J. Finding a representative network losses model for large-scale transmission expansion planning with renewable energy sources. *Energy* **101**, 343–358. doi:[10/f8k3ww](https://doi.org/10/f8k3ww) (2016) (cit. on pp. 28, 34).
102. Stott, B., Jardim, J. & Alsac, O. DC power flow revisited. *IEEE Transactions on Power Systems* **24**, 1290–1300. doi:[10/c8sx2b](https://doi.org/10/c8sx2b) (2009) (cit. on pp. 28, 41).
103. Hobbs, B., Drayton, G., Bartholomew Fisher, E. & Lise, W. Improved transmission representations in oligopolistic market models: quadratic losses, phase shifters, and DC lines. *IEEE Transactions on Power Systems* **23**, 1018–1029. doi:[10/b4df8d](https://doi.org/10/b4df8d) (2008) (cit. on p. 28).
104. Taylor, J. A. *Convex Optimization of Power Systems* doi:[10/fzts](https://doi.org/10/fzts) (Cambridge University Press, 2015) (cit. on pp. 29, 32, 35, 60, 61, 63).
105. Grainer, J. J. & Stevenson, W. *Power System Analysis* (McGraw-Hill Education, 1994) (cit. on pp. 30, 41, 42, 60).
106. Dörfler, F. & Bullo, F. *Novel insights into lossless AC and DC power flow* in 2013 IEEE Power Energy Society General Meeting (2013). doi:[10/d6z4](https://doi.org/10/d6z4) (cit. on pp. 30, 52).
107. Coffrin, C., Hentenryck, P. V. & Bent, R. *Approximating line losses and apparent power in AC power flow linearizations* in 2012 IEEE Power and Energy Society General Meeting (2012). doi:[10/ggfzwr](https://doi.org/10/ggfzwr) (cit. on pp. 30, 36, 37, 52).
108. Molzahn, D. K. & Hiskens, I. A. A Survey of Relaxations and Approximations of the Power Flow Equations. *Foundations and Trends in Electric Energy Systems* **4**, 1–221. doi:[10/d298](https://doi.org/10/d298) (2019) (cit. on pp. 32, 37–39, 61).
109. Coffrin, C. & Roald, L. Convex Relaxations in Power System Optimization: A Brief Introduction. arXiv: [1807.07227](https://arxiv.org/abs/1807.07227) (2018) (cit. on p. 32).
110. Coffrin, C. & Van Hentenryck, P. A linear-programming approximation of AC power flows. *INFORMS Journal on Computing* **26**, 718–734. doi:[10/f6km6h](https://doi.org/10/f6km6h) (2014) (cit. on p. 32).
111. Ahuja, R. K., Magnanti, T. L. & Orlin, J. B. *Network Flows: Theory, Algorithms, and Applications* (Prentice Hall, 1993) (cit. on pp. 33, 149).
112. Simpson-Porco, J. W. Lossy DC Power Flow. *IEEE Transactions on Power Systems* **33**, 2477–2485. doi:[10/gdgqrq](https://doi.org/10/gdgqrq). arXiv: [1611.05953](https://arxiv.org/abs/1611.05953) (2018) (cit. on p. 36).
113. Zhang, H., Vittal, V., Heydt, G. T. & Quintero, J. A relaxed AC optimal power flow model based on a Taylor series. *IEEE Innovative Smart Grid Technologies Asia*, 5–9. doi:[10/d296](https://doi.org/10/d296) (2013) (cit. on p. 36).

114. Zhang, H., Heydt, G. T., Vittal, V. & Quintero, J. An improved network model for transmission expansion planning considering reactive power and network losses. *IEEE Transactions on Power Systems* **28**, 3471–3479. doi:10/f472bm (2013) (cit. on pp. 36, 37).
115. Sanchez-Martin, P. & Ramos, A. Modeling Transmission Ohmic Losses in a Stochastic Bulk Production Cost Model. eprint: <https://pdfs.semanticscholar.org/d981/75cbc4e830fb68612b0f319eb5b19fed4a84.pdf> (2006) (cit. on pp. 36, 37).
116. Farahmand, H., Huertas-Hernando, D., Warland, L., Korpas, M. & Svendsen, H. G. *Impact of system power losses on the value of an offshore grid for North Sea offshore wind* in *2011 IEEE Trondheim PowerTech* (2011), 1–7. doi:10/cnxm9d (cit. on p. 36).
117. Vibrant Clean Energy & Electric Power Research Institute. *Program on Technology Innovation: Coordinated Expansion Planning: Status and Research Challenges* tech. rep. 3012016661 (2019). eprint: https://www.vibrantcleanenergy.com/wp-content/uploads/2019/12/EPRI_Report_Dec2019.pdf (cit. on p. 37).
118. Dos Santos, T. N. & Diniz, A. L. A dynamic piecewise linear model for DC transmission losses in optimal scheduling problems. *IEEE Transactions on Power Systems* **26**, 508–519. doi:10/b5rk9g (2011) (cit. on p. 37).
119. Fortenbacher, P. & Demiray, T. Linear/quadratic programming-based optimal power flow using linear power flow and absolute loss approximations. doi:10/gd6svt. arXiv:1711.00317 (2017) (cit. on p. 39).
120. Palma-Benhke, R., Philpott, A., Jofré, A. & Cortés-Carmona, M. Modelling network constrained economic dispatch problems. *Optimization and Engineering* **14**, 417–430. doi:10/ggq5gt (2013) (cit. on p. 39).
121. Yang, Z., Zhong, H., Bose, A., Zheng, T., Xia, Q. & Kang, C. A linearized OPF model with reactive power and voltage magnitude: a pathway to improve the MW-Only DC OPF. *IEEE Transactions on Power Systems* **33**, 1734–1745. doi:10/gc5jrd (2018) (cit. on p. 39).
122. Zhong, H., Xia, Q., Wang, Y. & Kang, C. Dynamic economic dispatch considering transmission losses using quadratically constrained quadratic program method. *IEEE Transactions on Power Systems* **28**, 2232–2241. doi:10/gfzp35 (2013) (cit. on p. 39).
123. Siemens. *HVDC Classic – powerful and economical: High-performance power transmission* <https://tinyurl.com/yy3ouxwm> (cit. on p. 40).
124. Milano, F. *PSAT documentation: version 2.0.0* <https://www.eecs.wsu.edu/~ee521/Material/20120927/psat-20080214.pdf> (cit. on pp. 41, 42).

125. Kundur, P. *Power System Stability and Control* (McGraw-Hill, 1994) (cit. on pp. 43, 54).
126. Machowski, J., Bialek, J. & Bumby, J. *Power System Dynamics: Stability and Control* 2nd ed. (Wiley-Blackwell, 2008) (cit. on p. 43).
127. German Federal Network Agency (Bundesnetzagentur). *Monitoring Report 2019* <https://www.bundesnetzagentur.de/SharedDocs/Downloads/EN/Areas/ElectricityGas/CollectionCompanySpecificData/Monitoring/MonitoringReport2019.pdf> (cit. on pp. 44, 54).
128. Dvijotham, K. & Molzahn, D. K. *Error bounds on the DC power flow approximation: A convex relaxation approach* in *2016 IEEE 55th Conference on Decision and Control (CDC)* (2016), 2411–2418. doi:10/d64w (cit. on p. 52).
129. Purchala, K., Meeus, L., Van Dommelen, D. & Belmans, R. *Usefulness of DC power flow for active power flow analysis* in *IEEE Power Engineering Society General Meeting, 2005* (2005), 454–459. doi:10/d4ww62 (cit. on p. 52).
130. Lumbreras, S. & Ramos, A. The new challenges to transmission expansion planning. Survey of recent practice and literature review. *Electric Power Systems Research* **134**, 19–29. doi:10/ghdmz6 (2016) (cit. on p. 57).
131. Binato, S., Pereira, M. & Granville, S. A new Benders decomposition approach to solve power transmission network design problems. *IEEE Transactions on Power Systems* **16**, 235–240. doi:10/fdb26t (2001) (cit. on pp. 57, 60, 63, 64, 83).
132. Romero, R., Monticelli, A., Garcia, A. V. & Haffner, S. Test systems and mathematical models for transmission network expansion planning. *IEE Proceedings - Generation, Transmission and Distribution* **149**, 27. doi:10/b7jzbs (2002) (cit. on pp. 57, 60, 63).
133. Ramos, A. & Lumbreras, S. How to solve the transmission expansion planning problem faster: acceleration techniques applied to Benders' decomposition. *IET Generation, Transmission & Distribution* **10**, 2351–2359. doi:10/f8w2bd (2016) (cit. on pp. 57, 78, 79, 83).
134. Kocuk, B., Jeon, H., Dey, S. S., Linderoth, J., Luedtke, J. & Sun, A. A Cycle-Based Formulation and Valid Inequalities for DC Power Transmission Problems with Switching. doi:10/f8zkv3. arXiv: 1412.6245 (2014) (cit. on pp. 58, 62, 70, 72).
135. Ronellenfitsch, H., Manik, D., Hörsch, J., Brown, T. & Witthaut, D. Dual Theory of Transmission Line Outages. *IEEE Transactions on Power Systems* **32**, 4060–4068. doi:10/gbtpn2. arXiv: 1606.07276v2 (2017) (cit. on pp. 60, 62).
136. Biggs, N. *Algebraic Graph Theory* doi:10/fztq (Cambridge University Press, 1974) (cit. on p. 61).

137. Kavitha, T., Liebchen, C., Mehlhorn, K., Michail, D., Rizzi, R., Ueckerdt, T. & Zweig, K. A. Cycle bases in graphs characterization, algorithms, complexity, and applications. *Computer Science Review* **3**, 199–243. doi:[10/fc8qzp](https://doi.org/10.1016/j.csr.2009.05.001) (2009) (cit. on p. 61).
138. Kavitha, T., Mehlhorn, K., Michail, D. & Paluch, K. E. An $O(m^2n)$ Algorithm for Minimum Cycle Basis of Graphs. *Algorithmica* **52**, 333–349. doi:[10/c4xt4n](https://doi.org/10.1007/s00453-008-9141-1) (2008) (cit. on p. 61).
139. Bollobás, B. *Modern graph theory* corrected. *Graduate texts in mathematics* **184**. doi:[10/d57dnr](https://doi.org/10.1007/978-1-4419-9964-0) (Springer, 1998) (cit. on pp. 61, 62).
140. Ronellenfitsch, H., Timme, M. & Witthaut, D. A Dual Method for Computing Power Transfer Distribution Factors. *IEEE Transactions on Power Systems* **32**, 1007–1015. doi:[10/gf5p7j](https://doi.org/10.1109/TPWRS.2017.2705171) (2017) (cit. on pp. 61, 62).
141. Manik, D., Timme, M. & Witthaut, D. Cycle flows and multistability in oscillatory networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **27**, 083123. doi:[10/gbvj97](https://doi.org/10.1063/1.497997). arXiv: [1611.09825v2](https://arxiv.org/abs/1611.09825v2) (2017) (cit. on p. 61).
142. Carvalho, M., Soares, S. & Ohishi, T. Optimal active power dispatch by network flow approach. *IEEE Transactions on Power Systems* **3**, 1640–1647. doi:[10/c83ffp](https://doi.org/10.1109/TPWRS.1988.1039100) (1988) (cit. on p. 62).
143. Binato, S. *Expansão Ótima de Sistemas de Transmissão Através de Decomposição de Binders e Técnicas de Planos Cortantes* Portuguese. PhD thesis (Federal University of Rio de Janeiro, 2000). eprint: <https://www.cos.ufrj.br/uploadfile/publicacao/831.pdf> (cit. on p. 64).
144. Tsamasphyrou, P., Renaud, A. & Carpentier, P. Transmission network planning under uncertainty with Benders decomposition in *Optimization* **481** (2000), 457–472. doi:[10/dstwkf](https://doi.org/10.1080/01674940008839200) (cit. on p. 66).
145. Stevens, S. P. & Palocsay, S. W. Teaching Use of Binary Variables in Integer Linear Programs: Formulating Logical Conditions. *INFORMS Transactions on Education* **18**, 28–36. doi:[10/ggjkvf](https://doi.org/10.1287/ited.2017.18.1.28) (2017) (cit. on p. 68).
146. Lumbreras, S., Ramos, A. & Sánchez, P. Automatic selection of candidate investments for Transmission Expansion Planning. *International Journal of Electrical Power and Energy Systems* **59**, 130–140. doi:[10/f52p6f](https://doi.org/10.1016/j.ijepes.2014.05.001) (2014) (cit. on p. 82).
147. Majidi-Qadikolai, M. & Baldick, R. A Generalized Decomposition Framework for Large-Scale Transmission Expansion Planning. *IEEE Transactions on Power Systems* **33**, 1635–1649. doi:[10/gc5kf2](https://doi.org/10.1109/TPWRS.2018.2819142) (2018) (cit. on p. 83).

148. Munoz, F. D., Hobbs, B. F. & Watson, J. New bounding and decomposition approaches for MILP investment problems: Multi-area transmission and generation planning under policy constraints. *European Journal of Operational Research* **248**, 888–898. doi:10/fztw (2016) (cit. on p. 83).
149. Griffith, R. E. & Stewart, R. A. A Nonlinear Programming Technique for the Optimization of Continuous Processing Systems. *Management Science* **7**, 379–392. doi:10/fhttqf (1961) (cit. on p. 84).
150. Bazaraa, M. S., Sherali, H. D. & Shetty, C. *Nonlinear Programming: Theory and Algorithms* doi:10/cwps3z (Wiley-Interscience, 2006) (cit. on p. 84).
151. Chekuri, C. & Tseng, L. *Approximation Algorithms: Vertex Cover and Set Cover via LP* https://courses.engr.illinois.edu/cs598csc/sp2011/Lectures/lecture_4.pdf (cit. on p. 85).
152. German Federal Network Agency (Bundesnetzagentur). *Monitoring Report 2018* https://www.bundesnetzagentur.de/EN/Areas/Energy/Companies/DataCollection/Monitoring/MonitoringBenchmarkReport/Monitoring_Benchmark_Report_node.html (cit. on p. 86).
153. Oeding, D. & Oswald, B. R. *Elektrische Kraftwerke und Netze* 7th ed. doi:10/c2zwbm (Springer, 2011) (cit. on p. 86).
154. Tröndle, T., Lilliestam, J., Marelli, S. & Pfenninger, S. Trade-Offs between Geographic Scale, Cost, and Infrastructure Requirements for Fully Renewable Electricity in Europe. *Joule*, S2542435120303366. doi:10/gg8zk2 (2020) (cit. on pp. 99, 115, 116, 119, 120, 125, 129, 139).
155. Yue, X., Pye, S., DeCarolis, J., Li, F. G., Rogan, F. & Gallachoir, B. O. A review of approaches to uncertainty assessment in energy system optimization models. *Energy Strategy Reviews* **21**, 204–217. doi:10/gf75x4 (2018) (cit. on pp. 99, 101, 104).
156. Pye, S., Li, F. G., Petersen, A., Broad, O., McDowall, W., Price, J. & Usher, W. Assessing qualitative and quantitative dimensions of uncertainty in energy modelling for policy support in the United Kingdom. *Energy Research & Social Science* **46**, 332–344. doi:10/gf8nvv (2018) (cit. on p. 99).
157. Lombardi, F., Pickering, B., Colombo, E. & Pfenninger, S. Policy Decision Support for Renewables Deployment through Spatially Explicit Practically Optimal Alternatives. *Joule*, S2542435120303482. doi:10/gg8z6v (2020) (cit. on pp. 99, 103).
158. Strachan, N., Fais, B. & Daly, H. Reinventing the energy modelling–policy interface. *Nature Energy* **1**, 16012. doi:10/gh626j (2016) (cit. on p. 99).

159. DeCarolis, J. F., Babaei, S., Li, B. & Kanungo, S. Modelling to generate alternatives with an energy system optimization model. *Environmental Modelling and Software* **79**, 300–310. doi:[10/f8n923](https://doi.org/10.1016/j.envsoft.2016.08.013) (2016) (cit. on pp. [101](#), [103](#)).
160. Hennen, M., Lampe, M., Voll, P. & Bardow, A. SPREAD – Exploring the decision space in energy systems synthesis. *Computers and Chemical Engineering* **106**, 297–308. doi:[10/gb3czd](https://doi.org/10.1016/j.compchemeng.2017.07.011) (2017) (cit. on pp. [101](#), [103](#), [110](#)).
161. Moret, S., Codina Gironès, V., Bierlaire, M. & Maréchal, F. Characterization of input uncertainties in strategic energy planning models. *Applied Energy* **202**, 597–617. doi:[10/gbsxmz](https://doi.org/10.1016/j.apenergy.2017.07.071) (2017) (cit. on pp. [101](#), [119](#), [120](#)).
162. Shirizadeh, B., Perrier, Q. & Quirion, P. How sensitive are optimal fully renewable power systems to technology cost uncertainty? eprint: <https://EconPapers.repec.org/RePEc:fae:paper:2019.04> (2019) (cit. on pp. [101](#), [119](#), [120](#)).
163. Price, J. & Keppo, I. Modelling to generate alternatives: A technique to explore uncertainty in energy-environment-economy models. *Applied Energy* **195**, 356–369. doi:[10/ggfzqq](https://doi.org/10.1016/j.apenergy.2017.07.071) (2017) (cit. on pp. [103](#), [110](#)).
164. DeCarolis, J. F. Using modeling to generate alternatives (MGA) to expand our thinking on energy futures. *Energy Economics* **33**, 145–152. doi:[10/fm2vdq](https://doi.org/10.1016/j.eneco.2011.05.001) (2011) (cit. on p. [103](#)).
165. Li, F. G. & Trutnevyte, E. Investment appraisal of cost-optimal and near-optimal pathways for the UK electricity sector transition to 2050. *Applied Energy* **189**, 89–109. doi:[10/f9qtrf](https://doi.org/10.1016/j.apenergy.2017.07.071) (2017) (cit. on pp. [103](#), [120](#)).
166. Sasse, J.-P. & Trutnevyte, E. Distributional trade-offs between regionally equitable and cost-efficient allocation of renewable electricity generation. *Applied Energy* **254**, 113724. doi:[10/gf69s8](https://doi.org/10.1016/j.apenergy.2019.113724) (2019) (cit. on pp. [103](#), [137](#), [139](#), [140](#)).
167. Berntsen, P. B. & Trutnevyte, E. Ensuring diversity of national energy scenarios: Bottom-up energy system model with Modeling to Generate Alternatives. *Energy* **126**, 886–898. doi:[10/gbmm36](https://doi.org/10.1016/j.energy.2017.07.071) (2017) (cit. on p. [103](#)).
168. Nacken, L., Krebs, F., Fischer, T. & Hoffmann, C. Integrated renewable energy systems for Germany – A model-based exploration of the decision space. *16th International Conference on the European Energy Market*, 8. doi:[10/fzt8](https://doi.org/10.1016/j.fzt.2019.08.001) (2019) (cit. on p. [103](#)).
169. Pedersen, T. T., Victoria, M., Rasmussen, M. G. & Andresen, G. B. Modeling all alternative solutions for highly renewable energy systems. arXiv: [2010.00836](https://arxiv.org/abs/2010.00836) (2020) (cit. on pp. [103](#), [117](#)).

170. Schlachtberger, D. P., Brown, T., Schäfer, M., Schramm, S. & Greiner, M. Cost optimal scenarios of a future highly renewable European electricity system: Exploring the influence of weather data, cost parameters and policy constraints. *Energy* **163**, 100–114. doi:10/gfk5cj. arXiv: 1803.09711 (2018) (cit. on p. 102).
171. Pfenninger, S. Dealing with multiple decades of hourly wind and PV time series in energy models: A comparison of methods to reduce time resolution and the planning implications of inter-annual variability. *Applied Energy* **197**, 1–13. doi:10/gft8fx (2017) (cit. on p. 105).
172. Mavromatidis, G., Orehounig, K. & Carmeliet, J. Uncertainty and global sensitivity analysis for the optimal design of distributed energy systems. *Applied Energy* **214**, 219–238. doi:10/gc6n48 (2018) (cit. on pp. 115, 120, 129).
173. Pizarro-Alonso, A., Ravn, H. & Münster, M. Uncertainties towards a fossil-free system with high integration of wind energy in long-term planning. *Applied Energy* **253**, 113528. doi:10/ghfgtx (2019) (cit. on pp. 115, 119).
174. Fais, B., Keppo, I., Zeyringer, M., Usher, W. & Daly, H. Impact of technology uncertainty on future low-carbon pathways in the UK. *Energy Strategy Reviews* **13–14**, 154–168. doi:10/f9gmpd (2016) (cit. on p. 115).
175. Usher, W. *The Value of Global Sensitivity Analysis for Energy System Modelling in International Energy Workshop* (2015), 29. eprint: https://www.researchgate.net/publication/322721974_The_Value_of_Global_Sensitivity_Analysis_for_Energy_System_Modelling (cit. on pp. 115, 124).
176. Mavrotas, G. Effective implementation of the ϵ -constraint method in Multi-Objective Mathematical Programming problems. *Applied Mathematics and Computation* **213**, 455–465. doi:10/bwkzrt (2009) (cit. on p. 117).
177. Danish Energy Agency. *Technology Data* <https://ens.dk/en/our-services/projections-and-models/technology-data> (Oct. 1, 2020) (cit. on pp. 119, 120).
178. Gritsevskiy, A. & Nakićenovi, N. Modeling uncertainty of induced technological change. *Energy Policy* **28**, 907–921. doi:10/d4rkhb (2000) (cit. on p. 119).
179. Yeh, S. & Rubin, E. S. A review of uncertainties in technology experience curves. *Energy Economics* **34**, 762–771. doi:10/dzc2b4 (2012) (cit. on p. 119).
180. Heuberger, C. F., Rubin, E. S., Staffell, I., Shah, N. & Mac Dowell, N. Power capacity expansion planning considering endogenous technology cost learning. *Applied Energy* **204**, 831–845. doi:10/gcgs2v (2017) (cit. on pp. 119, 150).
181. Mattsson, N. *Learning by modeling energy systems* PhD thesis (Chalmers University of Technology, 2019). eprint: <https://research.chalmers.se/publication/514513> (cit. on p. 119).

182. Li, P.-H., Pye, S. & Keppo, I. Using clustering algorithms to characterise uncertain long-term decarbonisation pathways. *Applied Energy* **268**, 114947. doi:[10/ghchgg](https://doi.org/10/ghchgg) (2020) (cit. on pp. [119](#), [120](#)).
183. Moret, S., Bierlaire, M. & Maréchal, F. Robust Optimization for Strategic Energy Planning. *Informatica* **27**, 625–648. doi:[10/f87mh8](https://doi.org/10/f87mh8) (2016) (cit. on p. [120](#)).
184. Pilpola, S. & Lund, P. D. Analyzing the effects of uncertainties on the modelling of low-carbon energy system pathways. *Energy* **201**, 117652. doi:[10/ggtk2r](https://doi.org/10/ggtk2r) (2020) (cit. on p. [120](#)).
185. Trutnevyte, E. EXPANSE methodology for evaluating the economic potential of renewable energy from an energy mix perspective. *Applied Energy* **111**, 593–601. doi:[10/ftdb](https://doi.org/10/ftdb) (2013) (cit. on p. [120](#)).
186. Lopion, Markewitz, Stolten & Robinius. Cost Uncertainties in Energy System Optimization Models: A Quadratic Programming Approach for Avoiding Penny Switching Effects. *Energies* **12**, 4006. doi:[10/ghchff](https://doi.org/10/ghchff) (2019) (cit. on p. [120](#)).
187. Palar, P. S., Tsuchiya, T. & Parks, G. T. Multi-fidelity non-intrusive polynomial chaos based on regression. *Computer Methods in Applied Mechanics and Engineering* **305**, 579–606. doi:[10/f8jsb6](https://doi.org/10/f8jsb6) (2016) (cit. on pp. [120](#), [121](#), [123](#), [124](#)).
188. Sudret, B. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety* **93**, 964–979. doi:[10/bjpp5r](https://doi.org/10/bjpp5r) (2008) (cit. on pp. [121](#), [122](#), [129](#)).
189. Fajraoui, N., Marelli, S. & Sudret, B. On optimal experimental designs for Sparse Polynomial Chaos Expansions. arXiv: [1703.05312](https://arxiv.org/abs/1703.05312) (2017) (cit. on pp. [121](#), [124](#)).
190. Gratiet, L. L., Marelli, S. & Sudret, B. Metamodel-based sensitivity analysis: Polynomial chaos expansions and Gaussian processes, 1–37. doi:[10/ggxtbp](https://doi.org/10/ggxtbp). arXiv: [1606.04273](https://arxiv.org/abs/1606.04273) (2015) (cit. on pp. [121](#), [122](#), [124](#)).
191. Mühlpfordt, T. *Uncertainty Quantification via Polynomial Chaos Expansion – Methods and Applications for Optimization of Power Systems* PhD thesis (Karlsruhe Institute of Technology, 2020). doi:[10/ggxtbr](https://doi.org/10/ggxtbr) (cit. on p. [121](#)).
192. Feinberg, J. & Langtangen, H. P. Chaospy: An open source tool for designing methods of uncertainty quantification. *Journal of Computational Science* **11**, 46–57. doi:[10/gddwzk](https://doi.org/10/gddwzk) (2015) (cit. on pp. [121](#), [122](#)).
193. Ng, L. W.-T. & Eldred, M. *Multifidelity Uncertainty Quantification Using Non-Intrusive Polynomial Chaos and Stochastic Collocation* in *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference* (2012). doi:[10/ghknms](https://doi.org/10/ghknms) (cit. on p. [123](#)).

194. Hosder, S., Walters, R. & Balch, M. *Efficient Sampling for Non-Intrusive Polynomial Chaos Applications with Multiple Uncertain Input Variables* in *48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference* (2007). doi:10/ftdc (cit. on p. 124).
195. *Statistical Factsheet 2018* <https://www.entsoe.eu/publications/statistics-and-data/#statistical-factsheet> (cit. on pp. 137, 138).
196. Cohen, J. J., Reichl, J. & Schmidthaler, M. Re-focussing research efforts on the public acceptance of energy infrastructure: A critical review. *Energy* **76**, 4–9. doi:10/f6pw36 (2014) (cit. on p. 137).
197. Jenkins, K., McCauley, D., Heffron, R., Stephan, H. & Rehner, R. Energy justice: A conceptual review. *Energy Research & Social Science* **11**, 174–182. doi:10/gftt5d (2016) (cit. on p. 137).
198. Mayfield, E. N., Cohon, J. L., Muller, N. Z., Azevedo, I. M. L. & Robinson, A. L. Quantifying the social equity state of an energy system: environmental and labor market equity of the shale gas boom in Appalachia. *Environmental Research Letters* **14**, 124072. doi:10/gghf65 (2019) (cit. on p. 137).
199. Fell, M. J., Pye, S. & Hamilton, I. Capturing the distributional impacts of long-term low-carbon transitions. *Environmental Innovation and Societal Transitions* **35**, 346–356. doi:10/ggzhnk (2020) (cit. on pp. 137, 138).
200. Batel, S., Devine-Wright, P. & Tangeland, T. Social acceptance of low carbon energy and associated infrastructures: A critical discussion. *Energy Policy* **58**, 1–5. doi:10/f4zsr3 (2013) (cit. on p. 137).
201. Lehmann, P., Ammermann, K., Gawel, E., Geiger, C., Hauck, J., Heilmann, J., Meier, J.-N., Ponitka, J., Schicketanz, S., Stemmer, B., Tafarte, P., Thrän, D. & Wolfram, E. Managing spatial sustainability trade-offs: The case of wind power. *UFZ Discussion Papers*. eprint: <https://EconPapers.repec.org/RePEc:zbw:ufzdps:42020> (2020) (cit. on p. 137).
202. Li, F. G., Pye, S. & Strachan, N. Regional winners and losers in future UK energy system transitions. *Energy Strategy Reviews* **13-14**, 11–31. doi:10/gg3dpx (2016) (cit. on pp. 138, 143).
203. Trutnevyte, E., Hirt, L. F., Bauer, N., Cherp, A., Hawkes, A., Edelenbosch, O. Y., Pedde, S. & van Vuuren, D. P. Societal Transformations in Models for Energy and Climate Policy: The Ambitious Next Step. *One Earth* **1**, 423–433. doi:10/ggfvzr (2019) (cit. on p. 138).

204. Schmidt, J., Schönhart, M., Biberacher, M., Guggenberger, T., Hausl, S., Kalt, G., Leduc, S., Schardinger, I. & Schmid, E. Regional energy autarky: Potentials, costs and consequences for an Austrian region. *Energy Policy* **47**, 211–221. doi:[10/f35m37](https://doi.org/10/f35m37) (2012) (cit. on p. 138).
205. Engelken, M., Römer, B., Drescher, M. & Welp, I. Transforming the energy system: Why municipalities strive for energy self-sufficiency. *Energy Policy* **98**, 365–377. doi:[10/f9bmw3](https://doi.org/10/f9bmw3) (2016) (cit. on p. 138).
206. Müller, M. O., Stämpfli, A., Dold, U. & Hammer, T. Energy autarky: A conceptual framework for sustainable regional development. *Energy Policy* **39**, 5800–5810. doi:[10/d5f6zz](https://doi.org/10/d5f6zz) (2011) (cit. on p. 138).
207. Ecker, F., Hahnel, U. J. J. & Spada, H. Promoting Decentralized Sustainable Energy Systems in Different Supply Scenarios: The Role of Autarky Aspiration. *Frontiers in Energy Research* **5**, 14. doi:[10/gg3m9w](https://doi.org/10/gg3m9w) (2017) (cit. on p. 138).
208. Rae, C. & Bradley, F. Energy autonomy in sustainable communities—A review of key issues. *Renewable and Sustainable Energy Reviews* **16**, 6497–6506. doi:[10/f4jwnq](https://doi.org/10/f4jwnq) (2012) (cit. on p. 138).
209. Panteli, M. & Mancarella, P. The Grid: Stronger, Bigger, Smarter?: Presenting a Conceptual Framework of Power System Resilience. *IEEE Power and Energy Magazine* **13**, 58–66. doi:[10/gfvdfn](https://doi.org/10/gfvdfn) (2015) (cit. on p. 138).
210. Tröndle, T., Pfenninger, S. & Lilliestam, J. Home-made or imported: On the possibility for renewable electricity autarky on all scales in Europe. *Energy Strategy Reviews* **26**, 100388. doi:[10/gf69s7](https://doi.org/10/gf69s7) (2019) (cit. on pp. 138, 143).
211. Weinand, J. M., Ried, S., Kleinebrahm, M., McKenna, R. & Fichtner, W. Identification of potential off-grid municipalities with 100% renewable energy supply. *Working Paper Series in Production and Energy* **40**. doi:[10/d3xm](https://doi.org/10/d3xm) (2020) (cit. on p. 138).
212. Bolwig, S., Bolkesjø, T. F., Klitkou, A., Lund, P. D., Bergaentzlé, C., Borch, K., Olsen, O. J., Kirkerud, J. G., Chen, Y.-k., Gunkel, P. A. & Skytte, K. Climate-friendly but socially rejected energy-transition pathways: The integration of techno-economic and socio-technical approaches in the Nordic-Baltic region. *Energy Research & Social Science* **67**, 101559. doi:[10/ggvc9v](https://doi.org/10/ggvc9v) (2020) (cit. on p. 138).
213. Drechsler, M., Egerer, J., Lange, M., Masurowski, F., Meyerhoff, J. & Oehlmann, M. Efficient and equitable spatial allocation of renewable power plants at the country scale. *Nature Energy* **2**, 17124. doi:[10/gf75zb](https://doi.org/10/gf75zb) (2017) (cit. on p. 139).

214. DeCarolis, J. F., Jaramillo, P., Johnson, J. X., McCollum, D. L., Trutnevyte, E., Daniels, D. C., Akin-Olçum, G., Bergerson, J., Cho, S., Choi, J.-H., Craig, M. T., de Queiroz, A. R., Eshraghi, H., Galik, C. S., Gutowski, T. G., Haapala, K. R., Hodge, B.-M., Hoque, S., Jenkins, J. D., Jenn, A., Johansson, D. J., Kaufman, N., Kiviluoma, J., Lin, Z., MacLean, H. L., Masanet, E., Masnadi, M. S., McMillan, C. A., Nock, D. S., Patankar, N., Patino-Echeverri, D., Schivley, G., Siddiqui, S., Smith, A. D., Venkatesh, A., Wagner, G., Yeh, S. & Zhou, Y. Leveraging Open-Source Tools for Collaborative Macro-energy System Modeling Efforts. *Joule* **4**, 2523–2526. doi:[10/gkh7p4](https://doi.org/10/gkh7p4) (2020) (cit. on p. 148).
215. Conejo, A., Castillo, E., Minguéz, R. & Garcia-Bertrand, R. *Decomposition techniques in mathematical programming: engineering and science applications* **04**. doi:[10/ggkjgn](https://doi.org/10/ggkjgn) (Springer, 2006) (cit. on p. 149).
216. Lim, C. in *Wiley Encyclopedia of Operations Research and Management Science* (eds Cochran, J. J., Cox, L. A., Keskinocak, P., Kharoufeh, J. P. & Smith, J. C.) chap. Relationship among Benders, Dantzig-Wolfe, and Lagrangian Optimization (John Wiley & Sons, Inc., 2011). doi:[10/fcgdfc](https://doi.org/10/fcgdfc) (cit. on p. 149).
217. Létocart, L., Nagih, A. & Moun gla, N. T. Dantzig-Wolfe and Lagrangian decompositions in integer linear programming. *International Journal of Mathematics in Operational Research* **4**, 247. doi:[10/ggkktn](https://doi.org/10/ggkktn) (2012) (cit. on p. 149).
218. Guignard, M. & Kim, S. Lagrangean decomposition: A model yielding stronger lagrangean bounds. *Mathematical Programming* **39**, 215–228. doi:[10/cg48vx](https://doi.org/10/cg48vx) (1987) (cit. on p. 149).
219. Goldfarb, D. & Ma, S. Fast Multiple-Splitting Algorithms for Convex Optimization. *SIAM Journal on Optimization* **22**, 533–556. doi:[10/f345xq](https://doi.org/10/f345xq) (2012) (cit. on p. 149).
220. Glowinski, R. & Marroco, A. Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité d’une classe de problèmes de Dirichlet non linéaires. fr. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique* **9**, 41–76 (1975) (cit. on p. 149).
221. Gabay, D. & Mercier, B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications* **2**, 17–40. doi:[10/fp8fvz](https://doi.org/10/fp8fvz) (1976) (cit. on p. 149).
222. Victoria, M., Zhu, K., Brown, T., Andresen, G. B. & Greiner, M. Early decarbonisation of the European energy system pays off. *Nature Communications* **11**, 6223. doi:[10/ghpp93](https://doi.org/10/ghpp93). arXiv: [2004.11009](https://arxiv.org/abs/2004.11009) (2020) (cit. on p. 150).
223. Ardakani, A. J. & Bouffard, F. Identification of Umbrella Constraints in DC-Based Security-Constrained Optimal Power Flow. *IEEE Transactions on Power Systems* **28**, 3924–3934. doi:[10/f5gf9h](https://doi.org/10/f5gf9h) (2013) (cit. on p. 151).

224. Weinhold, R. & Mieth, R. Fast Security-Constrained Optimal Power Flow Through Low-Impact and Redundancy Screening. *IEEE Transactions on Power Systems* **35**, 4574–4584. doi:[10/ghhv3x](https://doi.org/10.1109/TPWRS.2020.3000000) (2020) (cit. on p. 151).

List of Abbreviations

AC	alternating current
ADMM	alternating direction method of multipliers
BD	Benders decomposition
CCGT	combined cycle gas turbine
CEM	capacity expansion model
CORINE	Coordinate Information on the Environment
DC	direct current
DWD	Dantzig-Wolfe decomposition
ECMWF	European Centre for Medium-Range Weather Forecasts
ENTSO-E	European Network of Transmission System Operators for Electricity
ERA5	ECMWF reanalysis dataset
FACTS	Flexible Alternating Current Transmission Systems
GEBCO	General Bathymetric Chart of the Oceans
GHG	greenhouse gas
HSJ	hop-skip-jump
HVAC	high-voltage alternating current
HVDC	high-voltage direct current
KCL	Kirchhoff's current law
KKT	Karush-Kuhn-Tucker
KVL	Kirchhoff's voltage law
LOPF	linear optimal power flow
LP	linear problem
MAE	mean absolute error
MAPE	mean absolute percentage error
MGA	modelling-to-generate-alternatives
MILP	mixed-integer linear problem
MINLP	mixed-integer nonlinear problem
MOO	multi-objective optimisation
Natura 2000	Network of Nature Protection Areas
NLP	nonlinear problem
NP	nondeterministic polynomial time
NTC	net transfer capacity
OCGT	open cycle gas turbine

OPF	optimal power flow
OSR	oversampling ratio
OTS	optimal transmission switching
PCE	polynomial chaos expansion
PTDF	power transfer distribution factor
PyPSA	Python for Power System Analysis
PyPSA-Eur	European Transmission System Model with PyPSA
RMSE	root mean squared error
SARAH-2	Surface Solar Radiation Data Set - Heliosat
SLP	sequential linear programming
TEP	transmission expansion planning
TYNDP	Ten Year Network Development Plan

List of Figures

1.1	Graphical outline of the contents and methods of the present thesis.	5
2.1	European transmission network topology at full resolution.	16
2.2	Exemplary Voronoi cells of the transmission network's substations.	17
2.3	Assumed eligible land for onshore wind turbine development near Karlsruhe. Red zones are excluded.	18
2.4	Patterns of electricity consumption, wind, and solar resource potential.	20
2.5	European transmission network clustered to varying spatial resolutions.	21
3.1	Illustration of the scope of the present chapter and its context.	31
3.2	Illustration of the feasible space in the flow-loss ($p_\ell - \psi_\ell$) dimensions.	38
3.3	Annual system costs split by system component for covered flow models.	44
3.4	Capacity correlation of nodal investments distinguished by technology.	45
3.5	Nodal price duration curves (snapshots and nodes) for selected flow models.	45
3.6	Maps of cost-optimal capacity expansion for the covered lossless flow models.	46
3.7	Convex envelope relaxation around loss formula $\psi = rp^2$ for lossy transport model and the iterative lossy linearised power flow model.	48
3.8	Comparison of simulated losses from AC power flow equations and optimised losses for iterative and non-iterative lossy linearised power flow.	48
3.9	Two-dimensional histograms comparing simulated flows (AC power flow) and optimised flows of the indicated flow models.	50
3.10	Flow duration curves of simulated flows and optimised flows.	51
3.11	Duration curves (lines and snapshots) of nodal price differences for lines experiencing no flow, congested lines, and all lines.	51
3.12	Convergence of Newton-Raphson algorithm.	51
3.13	Performance comparison by peak memory consumption and solving time.	53
4.1	Candidate lines within a synchronous zone.	64
4.2	Candidate lines across synchronous zones.	66
4.3	Example C shown as different directed rooted trees of the subnetwork graph.	67
4.4	Choice of candidate cycles within synchronous zone.	71
4.5	Choice of candidate cycles across synchronous zones.	72
4.6	Example Group D as subnetwork graphs.	72
4.7	Clustered European transmission network models.	75
4.8	Histograms of the distribution of the number of variables, constraints and binary variables across the benchmark cases for angle-based planning problem.	75

4.9	Solving times and memory consumption of cycle- and angle-based formulation.	77
4.10	Sensitivities of speed-up towards number of nodes and snapshots.	77
4.11	Sensitivities of speed-up factor distribution.	77
5.1	Comparison across all presented heuristic approaches.	88
5.2	Transmission line expansions on a map for different approaches.	89
5.3	Comparison of heuristics by total system cost and solving times.	90
5.4	Scaling of heuristics with with the number of nodes and snapshots.	91
5.5	Progression of optimality gaps.	92
6.1	Illustration of the near-optimal feasible space and the applied MGA method.	104
6.2	Cost-optimal and a near-optimal system layouts on a map.	106
6.3	Solution space of renewable generation infrastructure by technology for different levels of slack ϵ and emission reduction targets.	109
6.4	Solution space of storage, transmission and backup infrastructure by technology for different levels of slack ϵ and emission reduction targets.	111
6.5	Composition of generation and storage capacities for near-optimal solutions.	112
6.6	Correlations of capacities across all near-optimal solutions.	112
6.7	Near-optimal feasible space for specific regions.	113
7.1	Spatial and temporal resolution of the low and high fidelity model	118
7.2	Cross-validation errors for varying sample sizes and polynomial orders.	125
7.3	Distribution of total system cost, generation, storage, and transmission capacities for least-cost solutions.	127
7.4	Sensitivity of capacities towards their own technology cost.	128
7.5	First-order and total Sobol indices.	129
7.6	Space of near-optimal solutions by technology under cost uncertainty.	131
7.7	Space of near-optimal solutions by selected pairs of technologies under cost uncertainty.	132
7.8	Distribution of total system cost, generation, storage, and transmission capacities for two near-optimal search directions with $\epsilon = 8\%$ system cost slack.	134
8.1	Imbalances observed in 2018 and in cost-optimised renewable system.	138
8.2	Sensitivity of system cost and composition to nodal and country-wide equity requirements.	140
8.3	Maps of optimal system capacities for different equity requirements.	141
8.4	Lorenz curves for different equity requirements.	142
8.5	National annual investment relative to annual demand when every country produces as much as they consume (100%).	142
8.6	Total system cost impact of autarky on national and nodal levels.	142

List of Tables

3.1 Comparison of the number of variables and equality/inequality constraints.	32
3.2 Various statistical indicators compared accross covered flow models.	44
3.3 Flow errors compared accross covered flow models.	49
4.1 Numerical results of cycle-based and angle-based formulation.	76
5.1 Solver settings for solving MILP in Gurobi.	83
5.2 Experiment code glossary.	84
5.3 Average number of iterations for heuristic approaches.	88
5.4 Capacities and production of gas-fired powerplants.	90
6.1 Literature review of studies applying MGA to energy system models.	103
6.2 Statistics on optimal solutions for different GHG emission reduction levels.	107
7.1 Technology cost uncertainty.	119

Funding

This work was funded by the Helmholtz Association under grant no. VH-NG-1352.

Colophon

This thesis was typeset with $\text{\LaTeX} 2_{\epsilon}$. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at <http://cleanthesis.der-ric.de/>.