Spatial scaling in renewable energy networks

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich Physik

der Johann Wolfgang Goethe-Universität

in Frankfurt am Main

von

Jonas Hörsch

aus Bamberg

Frankfurt 2018

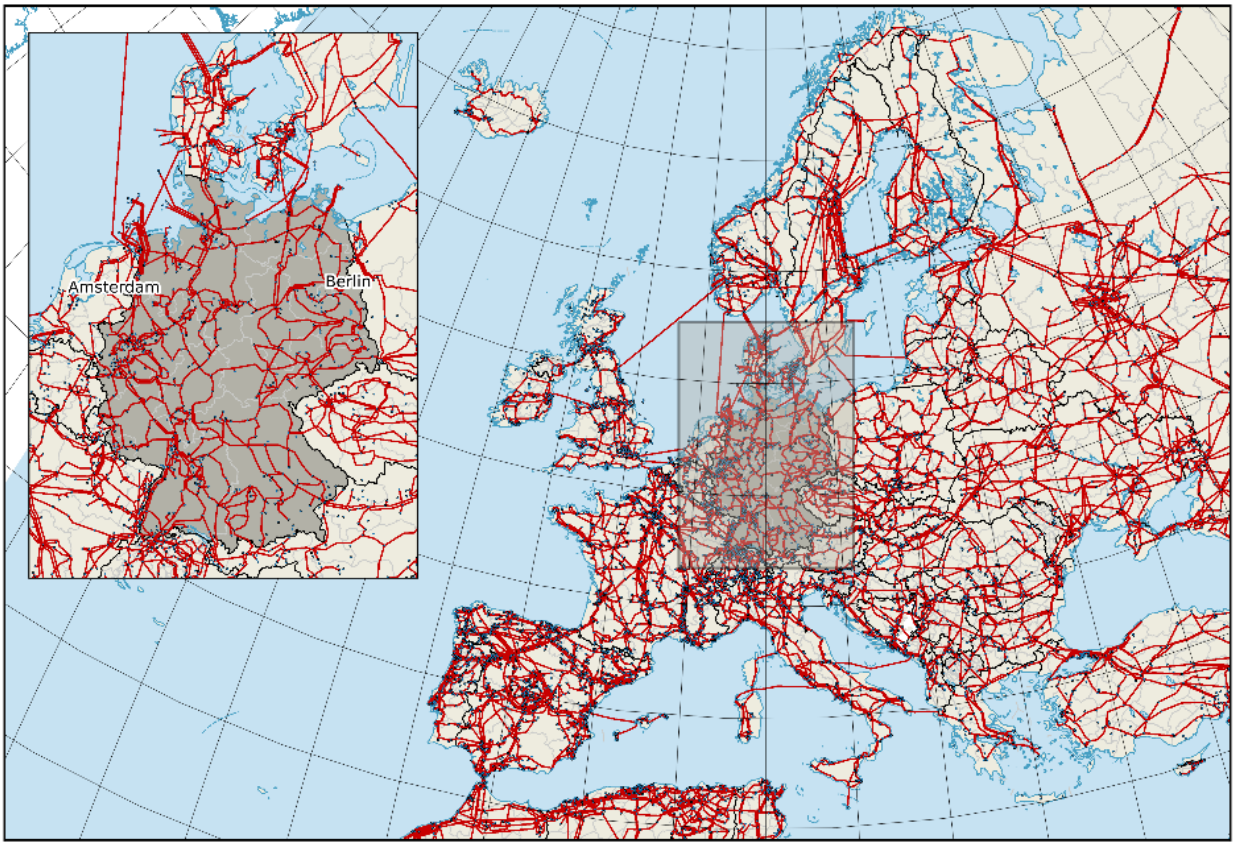
D 30

vom Fachbereich Physik der
Johann Wolfgang Goethe-Universität
als Dissertation angenommen.

Dekan: Owe Philipsen
Gutachter: Stefan Schramm, Martin Greiner, Holger Podlech
Datum der Disputation: 16.10.2018

SPATIAL SCALING IN RENEWABLE ENERGY NETWORKS

JONAS HÖRSCH



Spatially resolved co-optimization of generation, storage and transmission capacities at high shares of renewable energy

Mai 2018

ABSTRACT

Defossiliation of the energy system is crucial in the face of the impending risks of climate change. Electricity generation by burning fossil fuels is being displaced by renewable energy sources like hydro, wind and solar, driven by support schemes and falling costs from technological advances as well as manufacturing scale effects. The unavoidable shift from flexibly dispatchable generation to weather-dependent spatio-temporally varying generators transforms the generation and distribution of electricity into highly interdependent complex systems in multiple dimensions and disciplines:

In *time*, different scales, stretching from intra-day, diurnal, synoptic to seasonal oscillations of the weather interact with years and decades of planning and construction of capacity. In *space*, long-range correlations and local variations of weather systems as well as local bottlenecks in transmission networks affect solutions. The investment decisions about technological mix and spatial distribution of capacity follow *economic* principles, within restrictions which adapt in *social* feedback loops to public opinion and lobbyist influences.

In this work, a family of self-consistent models is developed which map physical steady-state operation, capacity investments and exogeneous restrictions of a European electricity system, in higher simultaneous spatial and temporal detail as well as scope than has previously been computationally tractable. Increasing the spatial detail of the renewable resources and co-optimizing the expansion of only a few transmission lines, reveals solutions to serve the European electricity demand at about today's electricity cost with only 5% of its carbon-dioxide emissions; and importantly their electricity mix differs from the findings at low spatial resolution.

As important intermediate steps,

- new algorithms for the convex optimization of electricity system infrastructure are derived from graph-theoretic decompositions of network flows. Only these enable the investigation of model detail beyond previous computational limitations.

- a comprehensive European electricity network model down to individual substations at the transmission voltage levels is built by combining and completing data from freely available sources.

- a network reduction technique is developed to approximate the detailed model at a sequence of spatial resolutions to investigate the role of spatial scale, and identify a level of spatial resolution which captures all relevant detail, but is still computationally tractable.

- a method to trace the flow of power through the network, which is related to a vector diffusion process on a directed flow graph embedded in a network, is used to analyse the resulting technology mix and its interactions with the power network.

The open-source nature of the model and restriction to freely available data encourages an accessible and transparent discussion about the future European electricity system, primarily based on renewable wind and solar resources.

PUBLICATIONS

Peer-reviewed articles

[1] Jonas Hörsch, Henrik Ronellenfitsch, Dirk Witthaut, and Tom Brown. "Linear optimal power flow using cycle flows". In: *Electric Power Systems Research* 158 (2018), pp. 126–135. ISSN: 0378-7796. DOI: 10.1016/j.epsr.2017.12.034. arXiv: 1704.01881.

[2] Tom Brown, Jonas Hörsch, and David Schlachtberger. "PyPSA: Python for Power System Analysis". In: *Journal of Open Research Software* 6.4 (1 2018). DOI: 10.5334/jors.188. arXiv: 1707.09913.

[3] Jonas Hörsch, Mirko Schäfer, Sarah Becker, Stefan Schramm, and Martin Greiner. "Flow tracing as a tool set for the analysis of networked large-scale renewable electricity systems". In: *International Journal of Electrical Power & Energy Systems* 96 (2018), pp. 390–397. ISSN: 0142-0615. DOI: 10.1016/j.ijepes.2017.10.024. arXiv: 1609.02977.

[4] Henrik Ronellenfitsch, Debsankha Manik, Jonas Hörsch, Tom Brown, and Dirk Witthaut. "Dual theory of transmission line outages". In: *IEEE Transactions on Power Systems* PP.99 (2017). ISSN: 0885-8950. DOI: 10.1109/TPWRS.2017.2658022. arXiv: 1606.07276.

[5] Jonas Hörsch and Tom Brown. "The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios". In: *14th International Conference on the European Energy Market*. Accepted after peer-review. June 2017, pp. 1–7. DOI: 10.1109/EEM.2017.7982024. arXiv: 1705.07617.

[6] Francesco Gardumi et al. "From the development of an open source energy modelling tool to its application and the creation of communities of practice: the example of OSeMOSYS". In: *Energy Strategy Reviews* 20 (2018), pp. 209–228. DOI: 10.1016/j.esr.2018.03.005.

[7] Bo Tranberg, Leon Schwenk-Nebbe, Mirko Schäfer, Jonas Hörsch, and Martin Greiner. "Flow-based nodal cost allocation in a heterogeneous highly renewable European electricity network". In: *Energy* 150 (2018), pp. 122–133. DOI: 10.1016/j.energy.2018.02.129. arXiv: 1801.10451.

[8] Jonas Hörsch, Fabian Hofmann, David Schlachtberger, and Tom Brown. "PyPSA-Eur: An open optimisation model of the European transmissions system". In: *Energy Strategy Reviews* 22 (2018), pp. 207–215. ISSN: 2211-467X. DOI: 10.1016/j.esr.2018.08.012. arXiv: 1806.01613.

[9] Fabian Gotzens, Heidi Heinrichs, Jonas Hörsch, and Fabian Hofmann. "Performing energy modelling exercises in a transparent way - the issue of data quality in power plant databases". In: *Energy Strategy Reviews* 23 (2019), pp. 1–12. DOI: 10.1016/j.esr.2018.11.004. arXiv: 1809.00974.

[10] Juliane Weber, Heidi Ursula Heinrichs, Bastian Gillesen, Diana Maybe, Jonas Hörsch, Tom Brown, and Dirk Witthaut. "Counter-intuitive behaviour of energy system models under CO_2 caps and prices". In: *Energy* (2018). DOI: 10.1016/j.energy.2018.12.052. arXiv: 1809.03157.

[11] Fabian Hofmann, Mirko Schäfer, Tom Brown, Jonas Hörsch, Stefan Schramm, and Martin Greiner. "Principal flow patterns across renewable electricity networks". In: *Europhysics Letters* 124.1 (2018), p. 18005. DOI: 10.1209/0295-5075/124/18005. arXiv: 1807.07771.

Conference papers and other publications

[12] Bo Tranberg, Mirko Schäfer, Tom Brown, Jonas Hörsch, and Martin Greiner. "Flow-based analysis of storage usage in a low-carbon European electricity scenario". In: *15th International Conference on the European Energy Market*. 2018. DOI: 10.1109/EEM.2018.8469951. arXiv: 1806.02549.

[13] Jonas Hörsch and Joanne Calitz. "PyPSA-ZA: Investment and operation co-optimization of integrating wind and solar in South Africa at high spatial and temporal detail". In: *WindAc Africa Conference, Cape Town*. Nov. 2017. arXiv: 1710.11199.

[14] Jonas Hörsch and Carla Mendes. "Integrating balancing reserves and congestion management to re-balance the German system". In: *14th International Conference on the European Energy Market*. June 2017, pp. 1–6. DOI: 10.1109/EEM.2017.7982014.

[15] Mirko Schäfer, Leon Schwenk-Nebbe, Jonas Hörsch, Bo Tranberg, and Martin Greiner. "Allocation of nodal costs in heterogeneous highly renewable European electricity networks". In: *14th International Conference on the European Energy Market*. June 2017, pp. 1–6. DOI: 10.1109/EEM.2017.7981964.

[16] Jonas Hörsch, Tom Brown, and Stefan Schramm. "Spatial Scale Dependence in Joint Optimisation of Generation and Transmission for Highly Renewable Scenarios". In: *15th International Workshop on Large-Scale Integration of Wind Power into Power Systems*. Paper: WIW16-124. Nov. 2016.

[17] Mirko Schäfer, Sabrina Hempel, Jonas Hörsch, Tranberg Bo, Stefan Schramm, and Martin Greiner. "Power flow tracing in complex networks". In: *New Horizons in Fundamental Physics*. Ed. by Stefan Schramm and Mirko Schäfer. FIAS Interdisciplinary Science Series. Nov. 2016, pp. 357–373. DOI: 10.1007/978-3-319-44165-8_26.

[18] Jonas Hörsch, Mirko Schäfer, Sarah Becker, Stefan Schramm, and Martin Greiner. "Power flow tracing across Germany's transmission grid". In: *10th Conference on sustainable development of energy, water and environment systems (SDEWES)*. Paper: SDEWES2015.0410. Sept. 2015.

CONTENTS

ACRONYMS

CO_2	carbon dioxide
NO	nitrous oxide
CH_4	methane
GHG	greenhouse gases
RCM	radiative-convective model
GCM	global circulation model
TSO	transmission system operator
UHV	ulta-high voltage
EHV	extra-high voltage
HVDC	high-voltage direct current
HVAC	high-voltage alternating current
IPCC AR5	Fifth assessment report by the International Panel on Climate Change
UNFCCC	United Nations Framework Convention on Climate Change
IAM	integrated assessment model
RCP	representative concentration pathway
NET	negative emission technologies
BECCS	bioenergy with carbon capture and storage
DSM	demand-side management
PV	photovoltaic
OPF	optimal power flow
LOPF	linear OPF
PTDF	power transfer distribution factor
AVR	active voltage regulator
DC	direct current
AC	alternating current
KCL	Kirchhoff's current law

KVL	Kirchhoff's voltage law
PyPSA	Python for Power System Analysis
PPM	powerplantmatching
ENTSO-E	European Network of Transmission System Operators for Electricity
TYNDP	Ten Year Network Development Plan
SO&AF	ENTSO-E Scenario Outlook and Adequacy Forecast
OPSD	Open Power System Data
ESE	DOE Energy Storage Exchange
GEO	Global Energy Observatory
CARMA	Carbon Monitoring for Action
BNetzA	Bundesnetzagentur
EIA	U.S. Energy Information Administration
NREL	National Renewable Energy Laboratory
OCGT	open-cycle gas turbine
CCGT	closed-cycle gas turbine
GDP	gross domestic product
NTC	net transfer capacities
ATC	available transfer capacities
FBMC	flow-based market coupling
LMP	locational marginal price
LCOE	levelized cost of electricity
FOM	fixed operation and maintenance
NUTS3	Nomenclature of Territorial Units for Statistics level 3

1 INTRODUCTION

1.1 CLIMATE CHANGE

Physics

The basic understanding that the temperature equilibrium at the Earth's surface depends crucially on the asymmetric absorption of thermal radiation by the atmosphere was established in the 19th century thanks to largely Fourier, Tyndall and Arrhenius. While the atmosphere is nearly transparent to solar radiation at visible frequencies[1], it is mostly opaque to thermal radiation from the Earth's surface at infrared wavelengths due to absorption by greenhouse gases (GHG), i. e. water vapour, carbon dioxide (CO_2) and other trace gases[2]. Instead, the energy flux through the lower atmosphere is dominated by heat convection and short-ranged radiative transformations (repetitive absorption and re-emission by GHG), as was only determined by numerical radiative-convective models (RCMs) in the 1960s. The first realistic RCMs described a single vertical column of the troposphere (the lower 8 to 16 km of the atmosphere) and found that doubling of the CO_2 concentration from 300 to 600 ppm would increase the equilibrium temperature of the Earth's surface by 1.95 to 2.36 K [1, 2], in agreement with the current estimate of 2.2 K and a 5 to 95% model range of 1.4 to 3.1 K [3]. Qualitatively, the additional CO_2 plays a critical role by raising the temperature in the dry – since cold – layers of the upper troposphere and *consequently* the concentration of water vapour. Only by pushing the model complexity to 3-dimensional global circulation models (GCMs), did it become possible to quantify the other predominant feedback loops like the surface albedo (reflectivity of the ground) and altitude, depth and cover of clouds [4, 5]. How the stable temperature structure to sustain atmospheric water vapour and clouds is provided by non-condensing GHG was demonstrated by Lacis et al. [6]. They removed the concentrations of non-condensing GHG in a current GCM and followed it plunge, within five decades, into an ice age with mean surface temperature at $-21\,^{\circ}C$.

Projections

Research in the last decades quantified and reduced uncertainties on all known feedback loops and compiled global and regionalized projections of the impending risks: Rising sea levels from expanding sea water and melting glaciers paired with storm surges will disrupt livelihoods on several islands and coastal

1 Black body radiation is emitted from the Sun's surface at about 5000 K with a spectrum peaking at approximately 500 nm.

2 Methane (CH_4), nitrous oxide (NO) and halo-carbons are the next important contenders.

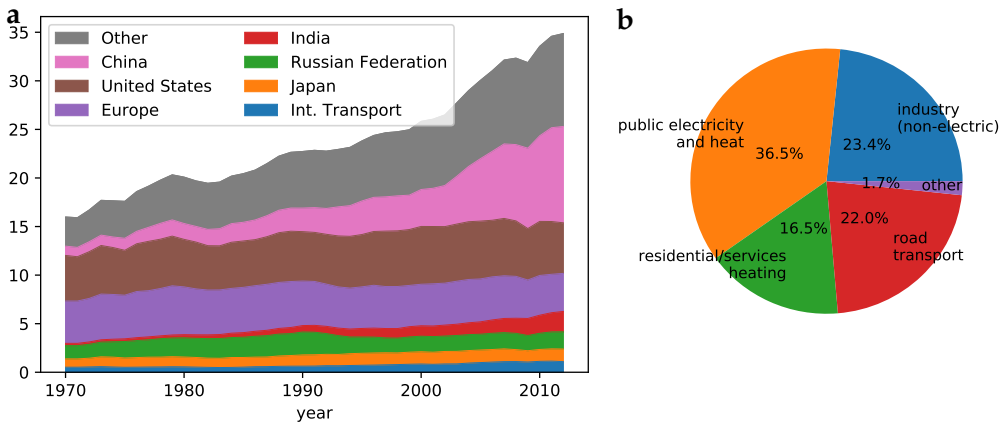


Figure 1.1: CO_2 emissions as compiled by [10]. (a) Annual time-series 1970-2012 for the largest seven emitting "regions" and the aggregated rest. (b) European emissions 2012 split into sectors.

regions. Extreme weather events such as heat waves and heavy precipitation are increasing in frequency and lead to water and food insecurity. They are unevenly distributed: risks are in general greater for disadvantaged people and communities everywhere. Climate change has already been positively linked to increasing human displacement and the outbreak of violent conflict [3, 7]. In view of the risks, the *Paris agreement* was adopted by the representatives of 196 member states of the United Nations Framework Convention on Climate Change (UNFCCC) in December 2015. It aims to keep the mean global temperature rise relative from pre-industrial levels well below 2 °C; at least during the current century [8].

Multi-model results show that upholding this limit with a probability $> 66\%$ requires limiting cumulative anthropogenic CO_2 emissions *from 2011* to about 1000 $GtCO_2$, when confining non-CO_2 forcing to the RCP2.6 scenario[3] (Anthropogenic emissions between 1870 and 2010 add up to 1800 $GtCO_2$). Equivalent scenario runs by integrated assessment models (IAMs) from the database of the Fifth assessment report by the International Panel on Climate Change (IPCC AR5) which realize the temperature limit have emissions in the range of 750 to 1400 $GtCO_2$ [3]. The majority of these scenarios depend crucially on negative emission technologies (NET) like bioenergy with carbon capture and storage (BECCS), forestry policies and direct air capture, which are not demonstrated or invested in at scale. Holding to the cumulative carbon budget means reducing annual CO_2 emissions from current 35 $GtCO_2/a$ progressively down by 57-83% until 2050 or even by 67-95% to not rely on NET [9].

Carbon dioxide emissions

World-wide annual CO_2 emissions shown in 1.1(a) have doubled over the last four decades largely driven by the strong economic development in China and

3 RCP2.6 is a selected representative for a stringent mitigation scenario. It fixes GHG emissions, atmospheric concentrations, air pollutant emissions and landuse development in agreement with keeping global warming below 2 °C above pre-industrial temperature levels [3].

India, but also by the increasing international trade and mobility (emissions by international aviation and navigation/nautical shipping). Emissions from the other large emitters, the United States, Russia, Japan and Europe, has approximately plateaued. European emissions in 2012 make up 3.9 $GtCO_2$ or 11% of global emissions and are attributed mainly to electricity and heat generation, followed by road transport as presented in 1.1(b). Electricity generation by power plants and district heating are only reported jointly as *public electricity and heat*. They comprise as much as 36% of European CO_2 emissions. This thesis implements technologically detailed optimizations demonstrating the economic viability to supply the European electricity demand for emissions as low as 77.5 $MtCO_2$ (a 95% reduction from the 1990 level) based on renewable generation.

1.2 RENEWABLE ENERGY

In 2016, 34% of the electricity generation in Europe was from renewable sources. Up to 2016, the well-established hydrological electricity generation dominated the share Europe-wide from the key countries Norway and Switzerland. It provides a steady amount of about 17% of the electricity demand. But its potential is largely saturated [11], and it is being displaced by the two fastest growing generation sources today which are wind (9%) and photovoltaic generation (3%) [12, 13].

Their technologies are well-understood and built in scale:

HORIZONTAL WIND TURBINES extract up to close to 60% of the wind energy flux through the surface spanned by the rotor blades. The energy scales by the third power in the wind speed, up until plateauing at the generator capacity using pitch control. In 2016 and 2017, about 50 GW have been added each year worldwide, about 14 GW in Europe [14].

PHOTOVOLTAIC PANELS (PV) consist of several layers of semiconductors producing an electrical current by separating photoelectron-hole-pairs with the static electric field of a p-n-junction. They are thus able to convert up-to a fourth of the solar irradiation into electrical power[4]. In 2016 and 2017, 72 and 94 GW of capacity have been built, half of it in China [14].

CONCENTRATED SOLAR POWER (CSP) collects and focuses solar irradiance to heat a thermal liquid and drive a steam engine. It has higher levelized costs and can only make use of the direct component of sunlight, but allows on the other hand to decouple the volatile irradiation from the generation of electrical power with heat tanks.

With the advances in renewable energy technologies in the last decade, the discussion about the decarbonization of electricity generation and further energy sectors has shifted from arguments about the technical feasibility of renewable energy-based supply to the conditions of its economic viability [16],

4 In 2018, the efficiencies of average commercial modules were reported as 17% for silicon wafer based panels and 15% for thinfilm panels, while laboratory records reached 27% and 21% respectively [15]

as well as how to efficiently address the new operational and organizational challenges raised by fluctuating, weather-dependent availability and vanishing marginal costs for the planning of renewable energy networks.

1.3 DIMENSIONS AND SCALES

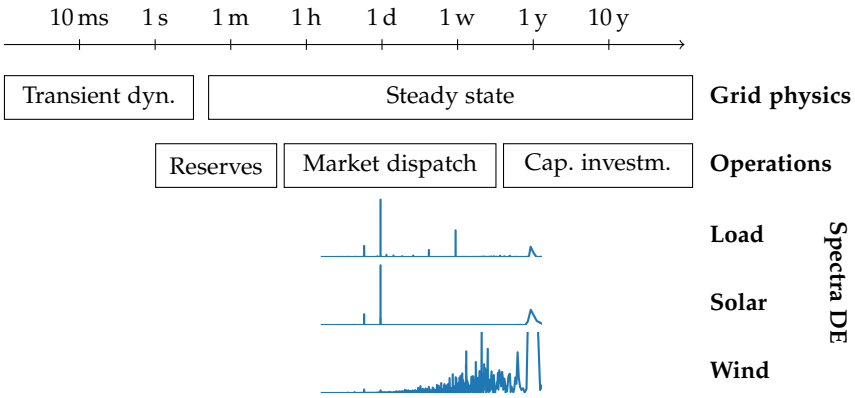


Figure 1.2: Time-scales in power system analysis, due to physics of electrically coupled devices, stability and economically-determined operation and capacity investments according to profit expectations are compared to spectra of the historically required demand and availability of solar and wind (DE 2011–2017 [17]).

An electrical energy network is a system of electrically coupled devices whose purpose is to deliver power from generators to loads (consumers). Generators as well as loads have technical and economic limitations in their availability in time and space like limited ramping capabilities or the weather-dependent availability of wind or solar generation. Hence, the time-scales in power system analysis vary from partial cycles of the 50 Hz grid frequency to the years required for planning and constructing new generation or transmission capacity as arranged in Fig. 1.2 and put into context with the dominant periods in historical German time-series of demand and generation availability of solar and wind.

Solar generation is strongly dominated by the *diurnal* cycle of the Earth's rotation, but also shows *seasonal* variations due to the eccentricity and inclination of the Earth's solar orbit.

Wind speed variations in space and time are governed by the migration of high and low pressure systems in the lower troposphere. On the leading *synoptic* scale, their dynamic behaviour is approximated by the vorticity equation with coriolis force and pressure gradient terms [18]. Empirically, the synoptic scale is quantified by the wind correlation length of 400 to 600 km [19] and is closely related to a time-scale of several weeks due to the slow migration speed of these pressure systems, appearing as a broad band in the fourier spectrum of the wind generation availability at the corresponding frequencies.

Before the advent of renewable generation with time-dependent availability and vanishing marginal cost, the planning of electricity systems naturally "di-

agonalized" into generation investment planning and generation and transmission adequacy evaluations: the generation capacity investment planning used integrated assessment tools (TIMES, MARKAL, MESSAGE) to evaluate the expected profits and rough capacity requirements. The tools approximate the operation of the electricity system by *load factors* (average per-unit generation) and system stability by a *reserve margin* (relative capacity above the need to meet peak demand). The load factors and required reserves were determined in separate generation adequacy forecasts carried out by the transmission system operators (TSOs), which went through the extra complexity to evaluate the operation of the electricity system with higher temporal resolutions. For flexibly dispatchable generating units alone, the optimal generation capacity to serve a particular load duration curve could even be determined analytically with so-called screening curves. As an orthogonal planning step of the TSO the required transmission system extensions were determined from the forecasted network flows [20–24].

Varying zero-marginal cost generation disrupts the use of load factors and the planning according to a fixed reserve margin, since the factors of all technologies become functions of the renewable energy feed-in, which depends on the amount of renewable generation capacities. Furthermore, many studies evaluating high shares of renewables have identified a growing need for flexibility which can be provided aside from conventional thermal peaking units by short- and long-term storage units [25–27], additional transmission capacities [28–31] or the incorporation and coordination of flexible demands from other energy sectors like heating and transport [32, 33]; all of which with their own trade-offs, but most of all requirements on model detail.

In regards to the *temporal scale* there is a consensus that the dimensioning of photovoltaic capacities and short-term storage units requires resolving the intra-diurnal load and solar feed-in on the lower end, while for long-term storage the full seasonal variation of load, wind and solar should be represented by modelling one full year or at least representative weeks [34, 35][5].

The required *spatial scales* are less well determined:

Previous investigations for Europe and the surrounding regions show that strengthened interconnectors between countries are used to smooth the wind feed-in over distances larger than the correlation length [28, 31, 36–40]. For instance, Schlachtberger et al. demonstrate that a wind-dominated European electricity system is about 33% cheaper to realize with strong tie-lines than as isolated countries dependant on storage. Thus, a continent-wide scope should be assumed as the upper bound.

On the lower end, several studies have investigated the synergies of generation and transmission capacity expansions in single countries using the areas around the ulta-high voltage (UHV) and extra-high voltage (EHV) buses as natural scale with a specific length[6] between 13 to 50 km [41]. For an investigation on European scope, the equivalent amount of buses ($\approx$ 6000) is not computationally tractable at the temporal scales and with the relevant technologies iden-

5 The investigation of a build-out pathway consequently requires several representative years at 5 to 10 year-steps in hourly resolution.

6 Determined as 25% and 75% quantiles of the transmission line length distribution.

tified above. The highly resolved ELMOD by Egerer, Gerbaulet, and Lorenz [42] was used to examine transmission expansion for the full European grid, but their limited temporal scope to just 18 separate hours forced them to assume fixed generation fleets and prevented them from evaluating storage. The transmission benefits for a European renewable energy network were analysed at an intermediate spatial scale of 88 buses by [30, 43] and 200 buses by [44], still with fixed generation and storage capacities, though. Hagspiel et al. [45] used an iterative scheme to jointly optimize generation, storage and transmission of the European electricity system at the same scale of 200 buses, but limited the temporal dimension again to only eight typical days.

In summary, electricity system planning with varying renewable generation must decide about generation, transmission and storage capacities accounting for the intra-diurnal (usually hourly) up to the seasonal time-scale to represent all the principal frequencies (see Fig. 1.2) and the spatial scales to encompass a continental electricity system down to an uncharted spatial resolution represented by up to 6000 buses.

1.4 APPROACH AND ORGANIZATION OF THIS WORK

The decision about the spatial distribution of the capacities in our political and societal framework is made by the combined actions of all investors according to their profit expectations. In an idealized market, with enough well-endowed independent players gifted with perfect foresight, the joint investment and operation would maximize the system's welfare according to basic market theory. Since the demand in electricity systems is typically very inelastic – even though demand-side management (DSM) aims to use it as an additional flexibility provider – a techno-economic optimization which minimizes the total system cost is able to (1) show the economic viability of an electricity system configuration, as well as (2) approximate the simulation of its evolution. Political goals and instruments like the European emission trading system enter the optimization either as cost adjustments or additional constraints like a CO_2 cap [24, 46].

This contribution starts from the assumption that a decarbonized renewable electricity system based on the temporally and spatially varying energy sources wind and solar introduces complex interdependencies with the other system components needed to provide flexibility. As long as these interdependencies are not (yet) fully understood, they have to be evaluated from a system's perspective – which can approximately be assumed with a least-cost optimization of the spatially and temporally resolved investment and operation of the electricity system. The prominent flexibility options are continental transmission, dispatchable hydro-electric generation, long- as well as short-term storage and flexibly dispatchable fossil generation at low load factors. Even in its simplest realization as a minimization of a linear cost objective under a set of linear equality and inequality constraints, finding the solution exceeds today's computational resources. Compromises about the resolution and scope must be made.

In this work, we attack the the dilemma head-on. In Chapter 2, we present new formulations of the linear optimization problem and benchmark them against the ones in use in the literature: A formulation that draws equally from the network graph as well as from its dual to model the Kirchhoff circuit laws is able to minimize the number of explicit variables and constraints, while preserving sparsity. The solution time of the electricity system optimization is reduced by up to two orders of magnitude.

Chapter 3 builds a comprehensive European electricity model from freely available sources: The topology of the European UHV/EHV network layers is extracted from online data and combined with standard electrical parameters from literature. Several incomplete public power plant databases are funnelled into a consistent dataset. Weather-dependent availability and expansion potential of on- and offshore wind-turbines as well as PV panels in the area of each substation is estimated from reanalysis weather data and information about land-use. Basic validation against published aggregate values is carried out, albeit tentatively.

Chapter 4 forms the core of this work. It applies a spatial clustering methodology on the detailed European model to parametrically interpolate from one-node-per-country to many-node-per-country electricity system models. The scaling of the cost components of the optimal system without and with transmission expansion is examined. Furthermore, since every new transmission line must be weighed against the possibility of local resistance due to landscape impacts, the effects from constraining the allowed expansion of the total transmission line volume are determined.

Chapter 5 generalizes a methodology to follow the energy flows from individual generators through the network for analysing the aggregate flows from generation technologies or from and to regions. A transmission usage measure is introduced for estimating the relative transmission volume used by wind generation, solar generation and thermal generation and is demonstrated on a 118-bus test electricity model.

The final Chapter 6 gives a conclusive summary of the technical achievements and concrete modeling results. It closes by highlighting next steps for addressing the remaining limitations.

2 LINEAR OPTIMAL POWER FLOW USING CYCLE FLOWS[1]

2.1 INTRODUCTION

Optimal power flow (OPF) problems can be constructed to find the welfare-maximizing generation and consumption levels in a network given the physical load flow equations, branch loading limits and generator cost functions. The full load flow equations are non-linear and the resulting optimization problem is non-convex, which makes it both challenging and computationally expensive to find a global optimum. In transmission networks with sufficient reactive power compensation, linearizing the load flow equations introduces only small errors [48, 49], with the benefit that the linear OPF (LOPF) can be expressed as a linear problem, whose convexity guarantees that a local optimum is a global optimum.

LOPF algorithms are used to clear markets with nodal pricing [50], to determine redispatch measures in markets with zonal pricing [51], to optimize dispatch taking account of contingencies (security constrained LOPF (SCLOPF)) [52] and in the long-term optimization of investment in generation and transmission assets [53]. LOPF is becoming more important with the growth of renewable energy, since the fluctuating feed-in has led to more frequent situations where the network is highly loaded [54]. When large networks are optimized over multiple representative feed-in situations, especially with discrete constraints on generation dispatch, the LOPF problems can take a significant time to solve, despite the linearization of the problem. Approaches in the literature to reducing the computational times include decomposition [55–57] and a parallelisable algorithm using the primal-dual interior point method [58].

In textbooks [52, 59] and major software packages such as MATPOWER [60], PYPOWER [61] and DIgSILENT PowerFactory [62], the linearization of the relations between power flows in the network and power injection at the buses is expressed indirectly through auxiliary variables that represent the voltage angles at the buses. Here we introduce two new formulations of the linear equations that use the power flows directly, decomposed using graph theoretic techniques into flows on a spanning tree and flows around closed cycles in the network. We evaluate the computational performance of the various methods for the LOPF problem, showing that the new formulations can solve significantly faster than the traditional angle-based formulation.

Cycle-flow techniques have already been used in [63] to improve the calculation times of power transfer distribution factors (PTDFs) and to gain a new understanding of the propagation of line outages in networks [64]. While preparing this manuscript, another paper [65] using cycle flows for LOPF in the context

1 This chapter is published as "Linear optimal power flow using cycle flows" by Hörsch et al. [47]. Sec. 2.2 was added to provide a more detailed introduction into electrical power system models to the uninitiated reader.

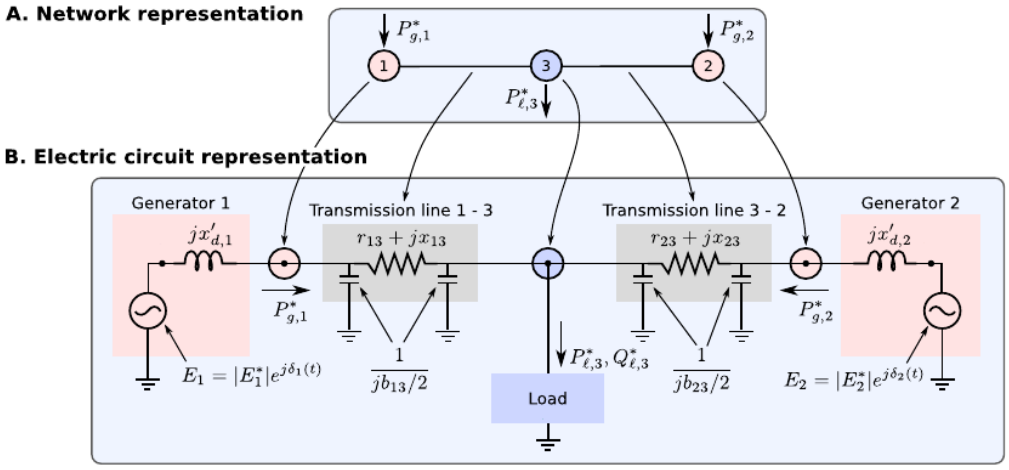


Figure 2.1: Representations of a simple 3-bus electricity network as a (A) topological network, and a (B) circuit diagram. Extracted from [66].

of optimal transmission switching was published; in contrast to that paper, here we provide an additional new formulation and benchmark both formulations against established formulations.

In Section 2.2, we give a short introduction into the representation of electricity networks, before in Section 2.3 the different formulations of the linear load flow are reviewed to prepare for the introduction of the optimization in Section 2.4. Extensions beyond the basic LOPF problem are described in Section 2.5 and the results of the performance analysis are presented in Section 2.6.

2.2 ELECTRICAL MODEL

Generators and loads are attached at buses and connected by branches, either transmission lines or transformers. Buses and branches as nodes and edges define the topology of a spatial graph, as illustrated in Fig. 2.1 (A) for a simple 3-bus network with two generators (1 and 3) and one load (2) connected by two transmission lines. The electrical representation of one of the three symmetric phases is shown in Fig. 2.1 (B): As is common in power system analysis, the continuous impedance and capacitance of a transmission line is lumped in the π-model to an impedance $z_\ell = (y_\ell)^{-1}$ surrounded by two symmetric ground-connected capacitors specified by the shunt admittance $y_\ell^{sh} = ib_\ell/2$ (refer also to App. A). Similarly, a transformer is described as maintaining a constant complex-valued ratio (magnitude and phase-shift) between the voltages of its two terminal buses.

Power flow equations

The voltage differences from a bus to the ground and to neighbouring buses lead to the branch current leaving bus i to bus j on branch ℓ,

$$I_{i\to j} = y_\ell^{sh} V_i + y_\ell (V_i - V_j) \; . \tag{2.1}$$

Variable	Definition
$i, j \in \{1, \dots N\}$	Bus labels
$s \in \{1, \dots G\}$	Generation source labels (wind, solar, gas, etc.)
$k, \ell \in \{1, \dots L\}$	Branch labels
$c, d \in \{1, \dots L - N + 1\}$	Cycle labels
$t \in \{1, \dots T\}$	Snapshot / time point labels
I_i	Complex nodal current through node i
V_i	Complex voltage at node i
$d_{i,s}$	Dispatch of generator at node i with source s
$D_{i,s}$	Available power of generator i, s
l_i	Electrical load at node i
θ_i	Voltage angle at node i
s_i	Complex power injection/withdrawal at node i
p_i	Total active power injection
q_i	Total reactive power injection
θ_ℓ	Voltage angle across a branch
f_ℓ	Branch active power flow
g_ℓ	Flow on spanning tree (zero if ℓ not in tree)
h_c	Flow around cycle c
F_ℓ	Branch active power rating
x_ℓ	Branch series reactance
r_ℓ	Branch series resistance
$K_{i\ell}$	$N \times L$ incidence matrix
$C_{\ell c}$	$L \times (L - N + 1)$ cycle matrix
$T_{\ell i}$	$L \times N$ tree matrix
$B_{\ell k}$	Diagonal $L \times L$ matrix of branch susceptances
Λ	$N \times N$ weighted Laplacian matrix $\Lambda = KBK^T$

Table 2.1: Variable definitions

Typically the branch currents are combined to nodal currents

$$I_i = \sum_j I_{i \to j} = \sum_j Y_{i,j} V_j \tag{2.2}$$

by defining the nodal admittance matrix Y. Its off-diagonal elements $Y_{i,j}$ are given by the negative admittance between the adjacent buses i and j and its diagonal elements $Y_{i,i}$ hold the sum of all admittances connected to bus i including the shunt admittance y_ℓ^{sh} [59]. The instantaneous power flows through

the network connecting power injections and withdrawals s_i are described by the *power flow equations*

$$s_i = \sum_j s_{i\to j} = \sum_j V_i I^*_{i\to j} = V_i I^*_i = \sum_j V_i Y^*_{i,j} V^*_j \ . \tag{2.3}$$

This set of $2N$ non-linear equations can be solved for all $4N$ quantities of the power flow state, if two real quantities per bus are given as parameters. Usually, one specifies (1) at buses with generators, the voltage magnitude $|V_i|$ and active power injection set-point of the generator $p_i = Re(s_i)$, while (2) at load only-buses the full complex-valued power withdrawal $s_i = p_i + iq_i$ must be given as data. The distinction arises from the two simple facts that the feed-in power of a generator is an economically determined parameter and (at least synchronous) generators actively maintain their voltage magnitude using active voltage regulators (AVRs).

The equations are solved with Newton-Raphson[2] or approximated by decoupling active and reactive power flow or a linearization as shown below. They are multi-stable and the number of solutions is closely linked to the number of cycles of the network topology[3] [67].

Equation 2.3 is valid at each point in time, but does not describe the evolution of the energy network, although the parameters s_i and p_i vary considerably in time and space. In the case of large-scale disturbances like transmission line or generator contingencies, feedback mechanisms of the generators and loads lead to complex transient network dynamics investigated with the swing equation (see also App. B) and other higher dimensional differential equations [66, 68, 69].

Linearization

For investigating the economic viability of a decarbonized electricity system, we are interested in time-scales several orders above the typical relaxation time in an already well-compensated high voltage network. There, the following assumptions are typically made [48, 49, 70]:

- All voltage magnitudes are set to $|V_n| = 1$, meaning that there are no voltage drops.
- Reactive power is neglected, i. e. assumed to be sufficiently provided.
- Losses are neglected, and line series reactance is much larger than the resistance $x_\ell \gg r_\ell \approx 0$, as well as the shunt admittance $y^{sh} \approx 0$, s. t. the admittance matrix reduces to the weighted Laplacian Λ

$$Y_{i,j} \approx \sum_\ell K_{i,\ell} (i x_\ell)^{-1} K^T_{\ell,j} = (K B K^T)_{i,j} = \Lambda_{i,j} \ , \tag{2.4}$$

2 Iterative jumps by $-J^{-1}r$ with the Jacobian J and residual r, until the latter is small enough.

3 Arbitrary closed loop currents are allowed. They are highly undesired, since a continuous inflow of power must be fed into the damping.

where the diagonal branch susceptance matrix is defined with $B_{\ell,\ell} = (ix_\ell)^{-1}$.

- Voltage angle differences are small, such that $e^{i\theta_\ell} \approx 1 + i\theta_\ell$.

They hold under normal working conditions and linearize the power flow equations in (2.3) to

$$p_i = Re(\sum_j V_i Y^*_{i,j} V^*_j) \approx \sum_j \Lambda_{i,j}\theta_j \,, \tag{2.5}$$

$$f_\ell = Re\left(V_i \left(y^{sh}_\ell V_i + y_\ell(V_i - V_j)\right)^*\right) \approx \frac{\theta_\ell}{x_\ell} \,. \tag{2.6}$$

The linearized active power flow in Eqs. (2.5)-(2.6) is form invariant with the real-valued current-voltage relations in a direct current (DC) circuit. Since these follow from simple Kirchhoffs node and loop circuit rules, we find by back-substitution that the linearized power flow equations are equivalent to a Kirchhoff's current law (KCL) for active power p_i and a Kirchhoff's voltage law (KVL) for voltage angle θ_i.

In the next section, which is the starting point of the published paper [47], we capitalized on this equivalency to reformulate the linearized power flow using different sets of equivalent variables to speed up the solution time of the linear optimal power flow, which will be introduced in Sec. 2.4.

2.3 LINEAR LOAD FLOW FORMULATIONS

The aim of the linear load flow calculation is to calculate the active power flow f_ℓ on each of the branches $\ell = 1, \ldots, L$ in terms of the active power p_i injected or consumed at each of the nodes $i = 1, \ldots, N$. In this section four methods are presented for solving the linear load flow, which lead to different formulations of the LOPF problem discussed below.

The linear approximation is valid if all branch resistances r_ℓ are negligible compared to the branch reactances x_ℓ, $r_\ell \ll |x_\ell|$, reactive power flows may be neglected, all voltage magnitudes are kept at nominal value and if all voltage angle differences across branches θ_ℓ are small enough that we can approximate $\sin\theta_\ell \sim \theta_\ell$ [48, 49]. Then the real power over a transmission line ℓ is given by

$$f_\ell = \frac{\theta_\ell}{x_\ell}, \tag{2.7}$$

where θ_ℓ is the voltage angle difference between the terminal buses of line ℓ.

The flows f_ℓ are constrained to be physical by the two Kirchhoff circuit laws for the current and voltage. KCL states that the current injected at each bus must equal the current withdrawn by the branches attached to the bus. This law can be expressed using the incidence matrix $K_{i\ell}$, which has non-zero values $+1$ if branch ℓ starts on node i and -1 if branch ℓ ends on node i. KCL then reads

$$p_i = \sum_\ell K_{i\ell} f_\ell \qquad \forall i = 1, \ldots, N. \tag{2.8}$$

KCL directly implies power conservation $\sum_i p_i = 0$ because $\sum_i K_{i\ell} = 0$ for all lines ℓ. KCL provides N linear equations for the L unknown flows f_ℓ, of which one is linearly dependent. This is not sufficient to uniquely determine the flows unless the network is a tree. Hence, $L - N + 1$ additional independent equations are needed.

The necessary equations and physicality are provided by the KVL, which states that the sum of potential differences across branches around all cycles in the network must sum to zero. It follows from graph theory that there are $L - N + 1$ independent cycles for a connected graph [71], which provides enough equations to constrain the f_ℓ completely. The independent cycles $c \in \{1, \dots L - N + 1\}$ are expressed as a directed linear combination of the branches ℓ in the cycle incidence matrix

$$C_{\ell,c} = \begin{cases} 1 & \text{if edge } \ell \text{ is element of cycle } c, \\ -1 & \text{if reversed edge } \ell \text{ is element of cycle } c, \\ 0 & \text{otherwise.} \end{cases} \tag{2.9}$$

Then the KVL becomes

$$\sum_\ell C_{\ell c}\theta_\ell = 0 \qquad \forall c = 1, \dots, L - N + 1. \tag{2.10}$$

where $\theta_\ell = \theta_i - \theta_j$ is the angle difference between the two nodes i, j which branch ℓ connects. Using equation (2.7), KVL can be expressed in terms of the power flows as

$$\sum_\ell C_{\ell c} x_\ell f_\ell = 0 \qquad \forall c = 1, \dots, L - N + 1. \tag{2.11}$$

Angle formulation

Commonly, the linear load flow problem is formulated in terms of the voltage phase angles $\theta_i, i \in \{1, \dots, N\}$. Using the incidence matrix the power flows are expressed as

$$f_\ell = \frac{1}{x_\ell} \sum_i K_{i\ell}\theta_i \qquad \forall \ell = 1, \dots, L \tag{2.12}$$

If the $L \times L$ diagonal matrix B is defined with $B_{\ell\ell} = \frac{1}{x_\ell}$ then the KCL equation (2.8) becomes

$$\begin{aligned} p_i &= \sum_{\ell,k,j} K_{i\ell} B_{\ell k} K_{jk} \theta_j \\ &= \sum_j \Lambda_{ij}\theta_j, \qquad \forall i = 1, \dots, N, \end{aligned} \tag{2.13}$$

using the nodal susceptance matrix $\Lambda = KBK^T$. In mathematical terms, Λ is a weighted network Laplacian [72].

The Angle formulation thus consists of two consecutive steps to calculate the flows f_ℓ. First, equation (2.13) is solved to obtain the N voltage angles θ_i. The equation provides only $N-1$ independent conditions such that we typically fix the voltage angle at a slack bus as $\theta_0 = 0$. Second, the flows are calculated via Equation (2.12). KVL is automatically satisfied as all closed cycles are in the kernel of the incidence matrix such that

$$\sum_\ell K_{i\ell} C_{\ell c} = 0 \qquad \forall c = 1, \ldots, L - N + 1. \tag{2.14}$$

PTDF formulation

For the power transfer distribution factor (PTDF) formulation [52] the matrix defining equation (2.13) is explicitly inverted to get the angles in terms of the power injections, and the resulting expression for the angles inserted into (2.12) to get a direct linear relation:

$$f_\ell = \sum_i PTDF_{\ell i}\, p_i \qquad \forall \ell = 1, \ldots, L, \tag{2.15}$$

where the PTDF matrix is given by $PTDF = BK^T\Lambda^*$. The pseudo-inverse Λ^* is used because Λ contains a zero eigenvalue for a connected network. Because KCL is no longer explicitly enforced, power conservation $\sum_i p_i = 0$ must be added as an explicit constraint for each connected network. The need to calculate the explicit pseudo-inverse of Λ makes this slow compared to the Angle formulation for single calculations, but once the PTDF has been computed, repeated application involves only matrix multiplication and no equation-solving. However, the PTDF matrix is typically dense, while Λ and K are sparse.

Kirchhoff formulation

In what we call the 'Kirchhoff formulation', the linear load flow is expressed as explicit linear constraints on the flows themselves. To the $N-1$ independent equations of the KCL equation from (2.8) we add the $L-N+1$ constraints of the KVL from (2.11). Together, this provides a system of L independent equations for the L variables f_ℓ and can therefore be solved.

Cycle formulation

In what we call the 'Cycle formulation' the flows f_ℓ are decomposed into flows g_ℓ on a spanning tree of the network, which ensure KCL is satisfied, and into cycle flows h_c that flow around each independent cycle c in the network without altering the power balance at any node [63]. We thus have:

$$f_\ell = g_\ell + \sum_c C_{\ell c} h_c. \tag{2.16}$$

The g_ℓ are only non-zero on the $N-1$ edges of a chosen spanning tree of the connected network. They are uniquely determined from the power imbalances by a matrix T

$$g_\ell = \sum_i T_{\ell i} p_i. \tag{2.17}$$

T is determined by fixing a slack bus and giving $T_{\ell i}$ value $+1$ if branch ℓ is in the directed path in the spanning tree from i to the slack bus or -1 if it is in the directed path but with reversed orientation [63]. This guarantees that KCL is satisfied at every node given that the power is balanced, $\sum_i p_i = 0$. Note that T only has to be calculated once for a network and is independent of the p_i. There is freedom both in the choice of spanning tree and in the choice of the slack bus used to determine the matrix T. The remaining $L-N+1$ degrees of freedom for the cycle flows h_c are fixed by the $L-N+1$ additional constraints from KVL (2.11)

$$\sum_\ell C_{\ell c} x_\ell \left(g_\ell + \sum_d C_{\ell d} h_d \right) = 0 \qquad \forall c \tag{2.18}$$

Solving this equation for the h_c involves solving $L-N+1$ linear equations. Power networks are not so heavily meshed, typically $L-N+1 < N-1$, such that this method can be significantly faster than the Angle formulation [63, 64].

2.4 LINEAR OPTIMAL POWER FLOW FORMULATIONS

In this section the linear load flow methods from Section 2.3 are transposed to the linear OPF (LOPF). In optimal power flow, power plant dispatch is optimized to minimize dispatch costs, assuming that no branch flows f_ℓ exceed their loading limits F_ℓ, i. e. $|f_\ell| \leq F_\ell$ [52].

The factors which control the speed of the solution to the LOPF problem are now more subtle. They include: i) the number of optimization variables; ii) the number of constraints; iii) the sparsity or density of the constraint matrix; iv) the shape of the feasible space near the optimal point; v) the method used to solve the linear problem. The first three factors are summarized for each of the formulations in Table 2.2.

The objective function for the LOPF has the generic form

$$\min_{\{d_{i,s}\},\{z_a\}} \left[\sum_{i,s} c_{i,s} d_{i,s} \right] \tag{2.19}$$

where $d_{i,s}$ is the dispatch of generator s at node i and $c_{i,s}$ is its operating cost. The z_a are auxiliary variables which implement the network constraints. One can also include the line flows f_ℓ as explicit optimization variables. The generic optimization problem then reads

$$\min_{\{d_{i,s}\},\{z_a\},\{f_\ell\}} \left[\sum_{i,s} c_{i,s} d_{i,s} \right] \tag{2.20}$$

Formulation	Variables	# Variables	# Equ. cons.	# Inequ. c.	Matrices
Pure Angle	$d_{i,s}, \theta_i$	$G+N$	$N+1$	$G+2L$	sparse
Angle+Flow	$d_{i,s}, f_\ell, \theta_i$	$G+L+N$	$L+N+1$	$G+2L$	sparse
Pure PTDF	$d_{i,s}$	G	1	$G+2L$	dense
PTDF+Flow	$d_{i,s}, f_\ell$	$G+L$	$L+1$	$G+2L$	dense
Kirchhoff	$d_{i,s}, f_\ell$	$G+L$	$L+1$	$G+2L$	sparse
Pure Cycle	$d_{i,s}, h_c$	$G+L-N+1$	$L-N+2$	$G+2L$	semi-sparse
Cycle+Flow	$d_{i,s}, h_c, f_\ell$	$G+2L-N+1$	$2L-N+2$	$G+2L$	semi-sparse

Table 2.2: Overview over the different formulations of the LOPF problem (N: number of buses. L : number of transmission lines, G : number of dispatchable generators)

All variables and their definitions are listed in Table 2.1.

The optimization must respect several constraints. First, the load l_i at each bus (which is assumed to be inelastic) must always be met. The bus power balance is the difference between generation and the electrical load l_i at the bus

$$p_i = \sum_s d_{is} - l_i \,. \tag{2.21}$$

If $p_i > 0$ then the node is a net exporter of power; if $p_i < 0$ then the node is a net importer of power. Note that p_i is only used to organize the presentation of the equations and is not an explicit optimization variable. Second, no generator may dispatch above its available power

$$d_{i,s} \leq D_{i,s} \qquad \forall \text{ generators} \,. \tag{2.22}$$

Third, the real power flows must remain within the loading limits of the lines

$$|f_\ell| \leq F_\ell \qquad \forall\ \ell = 1, \ldots, L. \tag{2.23}$$

It is sometimes desirable to limit the magnitude of the voltage angle differences θ_ℓ across the branches, to maintain the $\sin\theta_\ell \sim \theta_\ell$ approximation and avoid voltage stability problems [73]. Since $\theta_\ell = x_\ell f_\ell$, this constraint has the same form as the loading limit constraint (2.23), so we do not consider it further. Note that the load at each node l_i, specific costs $c_{i,s}$, generation upper limits $D_{i,s}$, branch loading limits F_ℓ and branch reactances x_ℓ are all exogenous data inputs and not subject to optimization in the considerations here. In all cases here only a single time point is considered and the network is assumed to be connected.

Finally active power flows on each branch f_ℓ are determined by the p_i and the auxiliary variables z_a through the constraints

$$f_\ell \equiv f_\ell(p_i, z_a) \tag{2.24}$$

The different formulations of the network equations presented in Section 2.3 give rise to different formulations of the linear optimal power flow. Whether we include the flows f_ℓ and additional auxiliary variables z_a as optimization variables has a significant impact on the computational resources needed to solve the optimization task. In the following we specify the different formulations of the LOPF in detail; their properties are summarized in Table 2.2. Note that for a uniquely-defined problem, all the formulations deliver the same optimum.

Pure Angle formulation

In the Pure Angle formulation the optimization problem (2.19) is solved with the voltage angles as auxiliary variables $\{z_a\} = \{\theta_i\}$ subject to the constraints (2.22) and

$$\begin{aligned} &\Big|\sum_i (BK^T)_{\ell i}\theta_i\Big| \leq F_\ell && \forall\, \ell = 1,\ldots,L, \\ &p_i = \sum_j \Lambda_j \theta_j && \forall\, i = 1,\ldots,N, \\ &\theta_0 = 0. \end{aligned} \tag{2.25}$$

The first equation ensures no branch overloading (note that it is sparse, inheriting the sparsity of K), the second equation is KCL and in the final equation the phase angle is fixed at the reference bus, which removes an unnecessary degree of freedom. Here and in the following the p_i are used as a short-hand notation according to equation (2.21).

The Pure Angle formulation is used in the free software tools MATPOWER [60] and PYPOWER [61]; it is therefore used as the benchmark implementation against which we compare all other formulations in Section 2.6.

Angle+Flow formulation

For the Angle+Flow formulation of the LOPF the flows f_ℓ are introduced as explicit optimization variables and the voltage angles are retained as auxiliary variables. Hence we have to solve the optimization problem (2.20) with N auxiliary variables, $\{z_a\} = \{\theta_i\}$ subject to the constraints (2.23) and (2.22) and the network equations

$$\begin{aligned} &f_\ell = \sum_i (BK^T)_{\ell i}\theta_i && \forall\, \ell = 1,\ldots,L, \\ &p_i = \sum_\ell K_{i\ell} f_\ell && \forall\, i = 1,\ldots,N, \\ &\theta_0 = 0. \end{aligned} \tag{2.26}$$

The introduction of additional optimization variables f_ℓ might appear to be redundant and unnecessary, but it will be shown to cause a significant speed-up in some cases. This is because modern solvers have sophisticated algorithms to 'pre-solve' solutions and remove redundancy that may not be obvious.

Pure PTDF formulation

In the Pure PTDF formulation no auxiliary variables are used such that the optimization problem is given by (2.19) subject to the constraints (2.22) and

$$\begin{aligned} \Big|\sum_i PTDF_{\ell,i} p_i\Big| &\leq F_\ell \qquad \forall\, \ell = 1,\dots,L. \\ \sum p_i &= 0. \end{aligned} \tag{2.27}$$

This formulation minimizes the number of optimization variables, but suffers from the fact that the matrix $PTDF$ is dense. This generates a large number of dense inequalities, which may make the feasible space complicated by introducing lots of interdependencies between the variables. This formulation has been used in the literature in, for example, [45].

PTDF+Flow formulation

The PTDF+Flow formulation does not use any auxiliary variables, but keeps the flows as explicit optimization variables. Hence we have to solve the optimization problem (2.20) subject to the constraints (2.23) and (2.22) and the network equations

$$\begin{aligned} f_\ell &= \sum_i PTDF_{\ell i} p_i \qquad \forall\, \ell = 1,\dots,L, \\ \sum_i p_i &= 0. \end{aligned} \tag{2.28}$$

Kirchhoff formulation

The Kirchhoff formulation is a new formulation of the LOPF which only requires the flow variables f_ℓ and introduces no additional auxiliary variables. The optimization problem is given by (2.20) subject to the constraints (2.23) and (2.22) and the network equations

$$\begin{aligned} \sum_\ell K_{i\ell} f_\ell &= p_i && \forall\, \ell = 1,\dots,L, \\ \sum_\ell C_{\ell c} x_\ell f_\ell &= 0 && \forall\, c = 1,\dots,L-N+1. \end{aligned} \tag{2.29}$$

This method implements the Kirchhoff circuit laws directly on the flow variables. It has both a small number of variables and extremely sparse constraints. This formulation was also introduced in the recent paper [65], which appeared while this manuscript was being prepared.

Pure Cycle formulation

The Cycle formulations of the linear load flow problem introduced in Section (2.3) leads to new formulations of the LOPF. In the Pure Cycle formulation we

solve the optimization problem (2.19) by adding $L - N + 1$ auxiliary variables $\{z_a\} = \{h_c\}$ subject to the constraints (2.22) and

$$\begin{aligned}
&\left| \sum_i T_{\ell i} p_i + \sum_c C_{\ell c} h_c \right| \leq F_\ell \qquad \forall\, \ell = 1, \ldots, L, \\
&\sum_\ell C_{\ell c} x_\ell \left[\sum_i T_{\ell i} p_i + \sum_{c'} C_{\ell c'} h_{c'} \right] = 0 \\
&\qquad\qquad \forall\, c = 1, \ldots, L - N + 1, \\
&\sum_i p_i = 0.
\end{aligned} \tag{2.30}$$

This involves fewer constraints than the Pure Angle formulation if $L < 2N$, which is typically true for power networks.

Cycle+Flow formulation

In the Cycle+Flow formulation we add auxiliary variables $\{z_a\} = \{h_c\}$ and include the flow variables f_ℓ as explicit optimization variables. The optimization problem is then given by (2.20) subject to the constraints (2.23) and (2.22) and the network equations

$$\begin{aligned}
&f_\ell = \sum_i T_{\ell i} p_i + \sum_c C_{\ell c} h_c && \forall\, \ell = 1, \ldots, L, \\
&\sum_\ell C_{\ell c} x_\ell f_\ell = 0 && \forall\, c = 1, \ldots, L - N + 1, \\
&\sum_i p_i = 0.
\end{aligned} \tag{2.31}$$

2.5 EXTENSIONS TO LOPF

In this section we briefly sketch some extensions of the LOPF problem to related problems for which the methodology also applies.

Multi-period and stochastic optimization

Inter-temporal aspects of optimal power flow, such as the operation of storage units or power plant unit commitment, can be considered using multi-period OPF [46, 52]. For periods labeled t with weighting π_t the objective function becomes

$$\min_{\{d_{i,s,t}\},\{z_{a,t}\},\{f_{\ell,t}\}} \left[\sum_{i,s,t} \pi_t c_{i,s} d_{i,s,t} \right]. \tag{2.32}$$

The network flow constraints repeat for each period t.

Storage introduces inter-temporal constraints that ensure that the storage state of charge $soc_{i,s,t}$ stays below the maximum energy storage capacity $SOC_{i,s}$:

$$\begin{aligned} soc_{i,s,t} &= soc_{i,s,t-1} + \eta_1 d_{i,s,t,charge} - \eta_2^{-1} d_{i,s,t,discharge} \\ 0 \leq soc_{i,s,t} &\leq SOC_{i,s} \qquad \forall\, i,s,t \end{aligned} \tag{2.33}$$

The efficiencies η_1, η_2 determine the losses during charging and discharging, respectively.

For stochastic optimization the periods t can represent different scenarios with probability π_t [46, 74].

Generation investment optimization

For generation investment optimization, the power plant capacities $D_{i,s}$ are promoted from exogenous parameters to optimization variables with capital costs $C_{i,s}$ [46]. The objective function becomes

$$\min_{\{D_{i,s}\},\{d_{i,s,t}\},\{z_{a,t}\},\{f_{\ell,t}\}} \left[\sum_{i,s} C_{i,s} D_{i,s} + \sum_{i,s,t} \pi_t c_{i,s} d_{i,s,t} \right] .$$

2.6 RESULTS

In this section we compare the computational performance of the different formulations of the LOPF problem introduced in Section 2.4 for various different test grids. All LOPF formulations are implemented in *Python for Power System Analysis* (PyPSA) [75], a free software tool developed at the Frankfurt Institute for Advanced Studies (FIAS). The formulation can be changed simply by passing a different argument 'formulation' to the LOPF function. Python for Power System Analysis (PyPSA) is used to generate linear program files (in CPLEX's .lp format), which are then passed to a linear solver (here we use the commercial software Gurobi [76]). The solver is then run using different algorithms for the linear program (primal and dual simplex, interior point) and the total solving time averaged over multiple runs is compared. The total solving time includes reading in the .lp file, pre-solving the matrix system and the solution algorithm. A computer system with 20 Intel Xeon E5-2650 cores @ 2.30GHz each and 128 GB RAM was used for each benchmark.

Problem preparation

Six standard network topologies are considered. case1354pegase, case1951rte, case2383wp, case2869pegase, case118 and case300 are standard IEEE cases taken from the MATPOWER software package [60]. In addition the open data SciGRID model of Germany's transmission network [77] is also tested, which has 585 nodes and 948 branches.

		Mean solution time	Speed-up compared to Pure Angle					
mode	case	Pure Angle [s]	Angle+ Flow	Pure PTDF	PTDF+ Flow	Kirch-hoff	Pure Cycle	Cycle+ Flow
p	case118	0.20	1.13	0.24	0.53	1.27	0.76	0.98
	case300	0.45	1.00	0.27	0.59	1.12	0.60	0.67
	case1354pegase	1.92	1.07	0.10	0.17	0.99	0.23	0.43
	case1951rte	3.21	0.22	0.14	0.27	1.30	0.32	0.55
	case2383wp	9.17	0.75	0.27	0.44	1.43	0.42	0.35
	case2869pegase	14.94	2.19	0.30	0.52	2.15	0.41	0.85
	scigrid	2.01	1.44	0.10	0.19	1.60	0.57	1.08
r	case118	0.25	0.99	0.12	0.23	1.22	0.58	0.88
	case300	0.77	1.12	0.11	0.20	1.37	0.54	0.73
	case1354pegase	7.58	1.38	0.06	0.10	2.55	0.42	0.87
	case1951rte	11.96	0.57	0.05	0.09	2.70	0.46	0.93
	case2383wp	65.17	3.40	0.13	0.24	4.31	1.13	1.55
	case2869pegase	51.83	0.83	0.06	0.10	3.60	0.43	1.18
	scigrid	3.60	1.62	0.06	0.12	2.44	0.75	1.14
rs	case118	0.26	0.99	0.13	0.23	1.24	0.61	0.90
	case300	0.77	1.11	0.11	0.19	1.38	0.55	0.73
	case1354pegase	7.45	1.35	0.06	0.10	2.42	0.42	0.89
	case1951rte	11.91	0.58	0.05	0.09	2.62	0.46	0.90
	case2383wp	60.73	3.22	0.14	0.25	4.12	1.10	1.44
	case2869pegase	52.88	0.85	0.07	0.11	3.61	0.45	1.20
	scigrid	7.26	2.70	0.12	0.25	4.14	1.33	2.03

Table 2.3: Speed-up compared to the Pure Angle formulation (> 1 means faster), best formulation marked green, worst marked red

To ensure that the timings were really testing the solution speed, the optimization problems were all made large enough so that no small one-off delays could significantly affect the timing. This was done by only considering large networks and only considering multi-period optimization with 24 hours represented in each problem. Large problems also represent the main target of efforts to improve computational speed.

Each test grid only has a single snapshot of the load. This was extended to 24 hours by subtracting a small fraction of normally distributed random noise $\varepsilon \sim \mathcal{N}(0, 0.2)$

$$d_{i,s,t} = d_{i,s} - d_{i,s} \left|\varepsilon_{i,s,t}\right|, \tag{2.34}$$

to ensure that the problem remained feasible and the solver was unable to reduce the problem from 24 identical problems to a single one.

The configuration of the generation was varied in three different 'modes':

- p: (plain): Only the conventional generators of the model are available. There is no inter-temporal linkage between the snapshots.

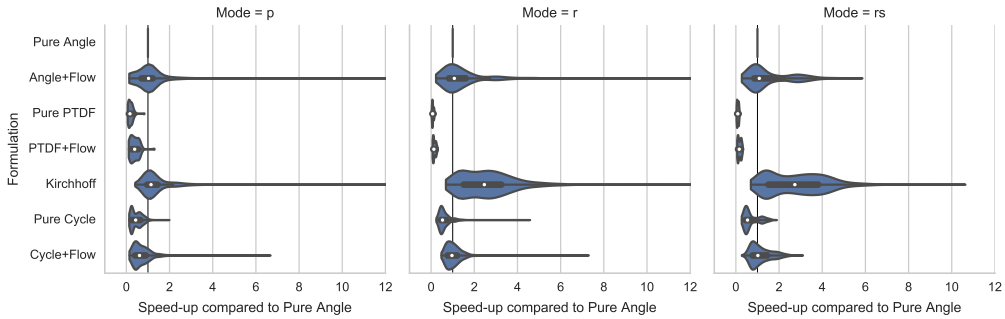


Figure 2.2: Speed-up compared to Pure Angle for total time (read + pre-solve + solve). The violin plots give the distribution of speed-ups, while the box plots mark the 25% and 75% quantiles and the dot marks the median.

- r: Compared to p, variable renewable generators are added to every single node to represent decentralized generation. The time series of the power availability of the renewable generators are taken at random from wind and time series for Germany for the year 2011 generated using the Aarhus Renewable Energy Atlas [78]. The renewable generators may be curtailed such that they correspond to dispatchable generators with no variable costs. There is no inter-temporal linkage between the snapshots.
- rs: Compared to r, storage units with a power capacity of a third of the nodal mean load are added to the fifteen buses with the highest average load. They provide an energy capacity of 6 hours at full power capacity and link the snapshots. More than 15 storage units made the computation times intractable.

For each network, mode and formulation, Gurobi was run in parallel using the primal simplex, the dual simplex and the interior point algorithms on at most four cores in parallel. The fastest solution was always taken. For each case and mode combination, 100 instances (i. e. different randomizations of the load and selections of the renewable time series) were generated and timed for all formulations except for the Pure PTDF and the PTDF+Flow formulations. For these only 10 instances were investigated, since the generation of a single of their lp files took up to 6 hours. The code for running the simulations with Snakemake [79] will be linked from the PyPSA website [75]. It was checked that all formulations gave identical results for the same problem.

Comparing average speed-up of the different formulations

In Table 2.3 the speed-up for the different formulations for the different problems (averaged over 100 instances) are shown, compared to the standard Pure Angle formulation.

The Kirchhoff formulation is the fastest in all cases where decentralized renewables are present in the network and the fastest in all but two cases for the 'plain' mode, where the Angle+Flow formulation is faster by a small margin. For the Kirchhoff formulation the speed-up factor averages 1.4 in mode 'p', 2.6

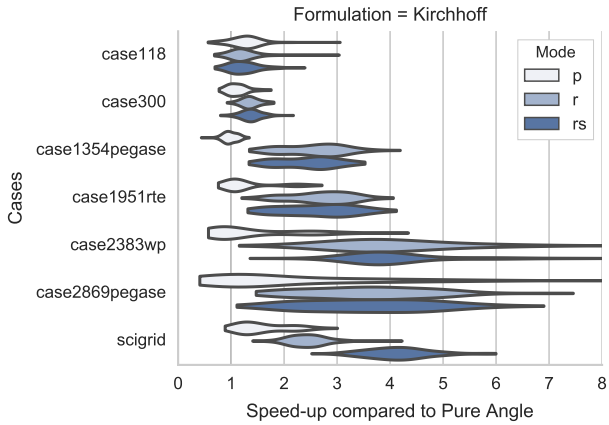


Figure 2.3: Kirchhoff speed-up compared to Pure Angle for total time (read + presolve + solve) per network case.

in mode 'r' and 2.8 for mode 'rs'. One reason the speed-up is high with renewables is that the optimization has to weigh up the dispatch at every single node and their effects on the flows. A sparser, less interdependent constraint set is a bigger advantage than in mode 'p', where only a few nodes have controllable generators. Inter-temporal storage introduces even more interdependences between variables, which again favours the sparse formulations.

The Angle+Flow formulation is the next fastest, averaging a speed improvement of 1.11 in mode 'p', 1.42 in mode 'r' and 1.54 in mode 'rs', despite the fact that there are more variables than the Pure Angle formulation.

The Cycle+Flow formulation is a factor 0.7 slower than the Pure Angle formulation in mode 'p', but faster by factor 1.04 in mode 'r' and 1.16 in mode 'rs'. The Pure Cycle formulation is on average slower in all modes. In particular cases the Pure Cycle formulation is faster than Pure Angle, but in each of those cases Cycle+Flow is faster.

The PTDF methods are slowest of all, with the Pure PTDF being the slowest. This is primarily driven by the size of the linear programming problem file, which takes a long time to read in by the solver. The size of the file is driven by the dense constraints coming from the dense PTDF matrix. Once the .lp problem is read in and pre-solved, the solving time is in some cases faster than the other methods.

Comparing specific speed-ups of the different formulations

The average speed-ups of the different formulations in the different modes masks considerable variations, both between the different network cases considered and within the instances for each case. Figure 2.2 shows violin plots of all the instances and all the cases for each mode and formulation combination, while in Figure 2.3 the different cases can be seen more clearly for the Kirchhoff formulation.

Take the speed-up of the Kirchhoff formulation in mode 'r' as an example. The average speed-up is 2.6, but this masks speed-ups for particular instances

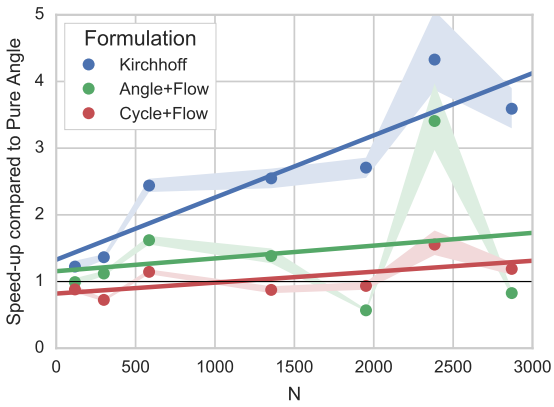


Figure 2.4: Speed-up compared to Pure Angle per nodes, shown are the mean values with 99% confidence interval and the result of a linear regression of all values for the three fastest formulations in mode 'r'.

that range from a factor 0.7 (i. e. a 30% slow-down, for an instance of case118) to factor 20 (for an instance of case2383wp). Even within a particular case there is significant variation for particular instances, ranging for case2383wp from 1.2 up to 20, although with a strong clustering around the mean of 4.3.

In 3% of the instances the Kirchhoff formulation in mode 'r' is in fact slower than the Pure Angle formulation, and all these instances are for the cases with a smaller number of nodes, case118 and case300. Figure 2.4 reveals that this is part of a bigger trend: In the 'r' mode, the Kirchhoff formulation speed-up grows with the size of the network, measured in terms of the number of nodes. The increase in speed-up with network size also holds true for the 'rs' mode.

Of all the cases, instances and modes, the Kirchhoff formulation was fastest in 79.3% of the problems, while the Angle+Flow was fastest in 12.5%, Angle in 7.5% and Cycle+Flow in 0.7%. If we restrict to the modes 'r' and 'rs', then the Kirchhoff is fastest in 91.6% of the problems, Angle+Flow in 5.9%, Angle in 2.1% and Cycle+Flow in 0.4%.

The high level of variation of the speed-up for different cases and instances (reflecting different load and renewable profiles) means that in practice it may be advisable, given a particular problem, to run several formulations in parallel and take the solution from whichever solves first, much as linear program solvers like Gurobi can be configured to run multiple solution algorithms in parallel, given the difficulty in predicting the runtime in advance.

2.7 CONCLUSION

In this chapter two new formulations of the linear OPF (LOPF) problem, the Kirchhoff formulation and the Cycle formulation, have been presented and a comprehensive study of the numerical performance has been provided. The new formulations both use a graph-theoretic decomposition of the network into a spanning tree and closed cycles.

In one formulation, the Kirchhoff formulation, which implements the two Kirchhoff circuit laws directly on the flow variables, the LOPF is shown to perform considerably faster than the standard Angle formulation used in today's power system tools. It shows the greatest speed-up in very large networks with decentralized generation, which are exactly the kinds of problems that are becoming increasingly important with the rise of distributed renewable energy. In the Kirchhoff formulation the LOPF can solve up to 20 times faster for particular cases, while averaging a speed-up of approx. 3 for the networks considered in this chapter. In 92% of the problems with distributed generation, the Kirchhoff formulation was the fastest formulation. In a very small number of specific cases the Cycles formulation was the fastest.

Future further applications could include the transmission expansion problem and the application of graph decomposition to the full non-linear optimal power flow problem.

3 MODEL DATA

3.1 INTRODUCTION

The energy system in Europe is undergoing a far-reaching transformation on multiple fronts: generation from variable renewable energy sources, such as wind and solar power, is growing due to the imperative of tackling climate change; electricity provision has been unbundled and liberalised, raising complex challenges for the efficient design and regulation of electricity markets; the need to decarbonise heating and transport is driving electrification of these sectors; and finally energy markets are being integrated across the continent [80].

To study this transformation, accurate modelling of the transmission grid is required. The need to take account of international electricity trading and the possibility of smoothing variable renewable feed-in over large distances (wind generation has a typical correlation length of around 600 km [19]) mean that models should have a continental scope. At the same time, high spatial detail is required, since national grid bottlenecks are already hindering the uptake of renewable energy today [81], and given persistent public acceptance problems facing new transmission projects [82], severe grid bottlenecks will remain a feature of the energy system for decades to come.

Currently there is no openly-available model of the full European transmission network with which researchers can investigate and compare different approaches to the energy transformation. The transmission grid dataset provided by the European Network of Transmission System Operators for Electricity (ENTSO-E) for the 2016 Ten Year Network Development Plan (TYNDP) [83] is rendered unusable by restrictive licensing, the exclusion of Finland, Norway and Sweden, and a lack of geographical localisation of the represented substations. The lack of geo-data means that the crucial weather system correlations and dynamics cannot be mapped onto the network. In 2005 in [84] an openly-available model of the continental European transmission network (i. e. excluding the UK, Ireland, Scandinavia and the Baltic states) was presented using a manual matching of buses and lines to the raster graphic of the ENTSO-E map, along with an open power plant database based on the Global Energy Observatory [85]; this model was updated to the network of 2009 in [86]. Apart from not covering the full ENTSO-E area, this dataset has the problem that much of the data was extracted manually, which is potentially error-prone and hard to repeat as new data becomes available, the buses are missing geo-coordinates and the power plant database is incomplete. In [87, 88] geo-coordinates and data for wind and solar plants were added to the dataset. Open datasets based on OpenStreetMap [89], such as the SciGRID network [77] and the osmTGmod [90] network, are of high quality in Germany, where data is well organised, but are not yet accurate for the rest of Europe. Similarly the open electricity model pro-

vided by the German Institute for Economic Research (DIW), ELMOD-DE [91], only covers Germany.

In this chapter we present a model of the European power system at the transmission network level which remedies these many deficiencies: it is not only open but also contains a high level of detail for the full ENTSO-E area. Grid data is provided by an automatic extraction of the ENTSO-E grid map; a power plant database is presented using a sophisticated algorithm that matches records from a wide range of available sources and includes geo-data; other data, such as time series for electrical load and wind, solar and hydro-electric availability, and geographic potentials for the expansion of wind and solar power, are also described. A new technique for comparing network datasets is presented and used to validate the grid data. The dataset and all code used to generate it from the raw data are available online [92], as a model for the PyPSA framework version 0.9 [93].

In Section 3.2 the data sources and processing methods are presented; the data is validated in Section 3.3; limitations of the dataset are discussed in Section 3.4; conclusions are drawn in Section 3.5.

3.2 DATA SOURCES AND METHODS

Network topology

The network topology and geography of substations and transmission lines have been extracted from the geographical vector data of the online ENTSO-E Interactive Map [94] by the GridKit toolkit [95] and published at [96]. The extract was corrected in several steps:

1. 29 alternating current (AC) lines were removed, which were identified as inadvertent duplicates by manual comparison to the online map.
2. Three converters at the end of dangling high-voltage direct current (HVDC) lines and at the border between Poland and Lithuania were introduced.
3. 64 transformers and 12 lines were added between buses with a distance of less than 1 km.
4. 60 AC lines carrying circuits of two different voltage levels were identified by inspecting the descriptive text tag and are split into several lines.

The electrical parameters are derived by assuming the standard AC line types in Table 3.1 for the length and number of circuits. The DC line capacities are assigned from the table in [97]. No transformer information is contained in the map, so a single transformer of capacity 2 GW (i. e. equivalent to four 500 MW transformers) is placed between buses of different voltage levels at the same location, with a reactance of 0.1 per unit. The transformer capacity assumption is on the high side to avoid introducing constraints where none exist in reality.

The restriction to buses and transmission lines of the voltage levels 220 kV, 300 kV and 380 kV in the landmass or exclusive economic zones of the European countries and the removal of 41 disconnected stub sub-networks (of less

Volt. level (kV)	Wires	Series resist. (Ω/km)	Series ind. reactance (Ω/km)	Shunt capacit. (nF/km)	Current therm. limit (A)	App. power therm. limit (MVA)
220	2	0.06	0.301	12.5	1290	492
300	3	0.04	0.265	13.2	1935	1005
380	4	0.03	0.246	13.8	2580	1698

Table 3.1: Standard line types for overhead AC lines [98]

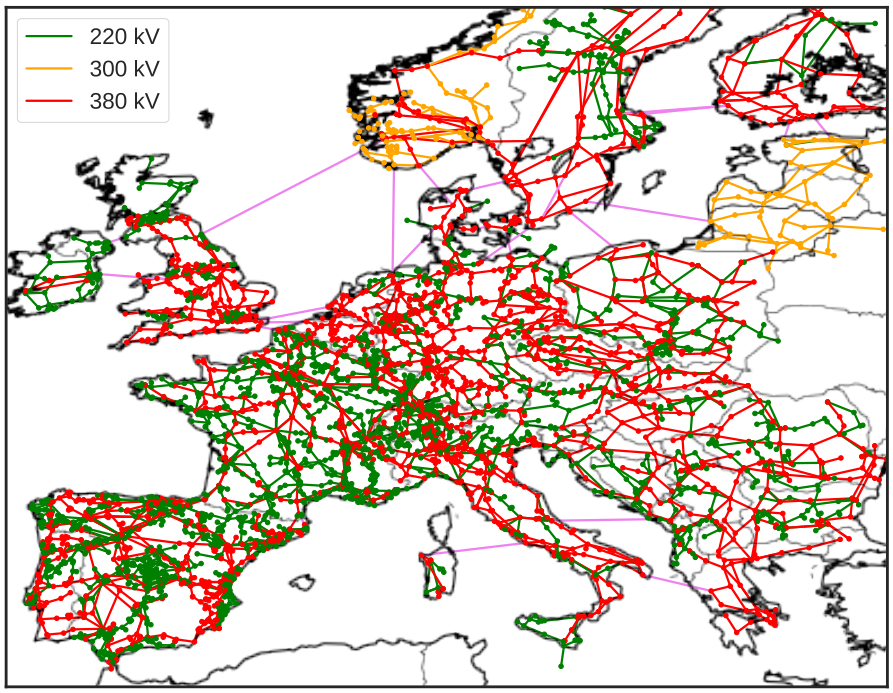


Figure 3.1: Transmission network model

than 10 buses) produces the transmission network in Figure 3.1 of all current transmission lines plus several ones which are already under or close to construction (these are marked in the dataset). In total the model contains 5586 high-voltage alternating current (HVAC) lines with a volume of 241.3 TW km (of which 11.4 TW km are still under construction), 26 HVDC lines with a volume of 3.4 TW km (of which 0.5 TW km are still under construction) and 4653 substations.

The countries are partitioned into Voronoi cells as catchment areas, each of which is assumed to be connected to the substation by lower voltage network layers. These Voronoi cells are used to link power plant capacities and determine feed-in by potential renewable energy generation, as well as the share of demand drawn at the substation.

Conventional power plants

Official sources often only report on country-wide capacity totals keyed by fuel-type and year like the Eurostat nrg_113a database [99], the ENTSO-E net generation capacity [100] or the ENTSO-E Scenario Outlook and Adequacy Forecast (SO&AF) [101, 102], while only seven countries[1] have official power plant lists collected and standardised by the Open Power System Data (OPSD) project [103]. This gap has been gradually closing since ENTSO-E started maintaining a power plant list (ENTSO-E PPL) on their Transparency Platform [104]. Unfortunately, it is still far from complete, for instance even after excluding solar and wind generators, the total capacity represented in Germany amounts only to about 54.5 GW, while the SO&AF reports 111 GW, 109 GW of which are also covered in the German Bundesnetzagentur (BNetzA) Kraftwerksliste [105] excluding power plants that have been permanently shut down.

The powerplantmatching (PPM) tool and database [106] we present in this section achieves good coverage by (1) standardising the records of several freely available databases, (2) linking them using a deduplication and record linkage application and (3) reducing the connected claims about fuel type, technology, capacity and location to the most likely ones.

PPM incorporates several power plant databases that are either published under free licenses allowing redistribution and reuse or are at least freely accessible. In the order of approximate reliability, there are OPSD [103], ENTSO-E PPL [104], DOE Energy Storage Exchange (ESE) [107], Global Energy Observatory (GEO) [85], Carbon Monitoring for Action (CARMA) from 2009 [108, 109] and the WRI Powerwatch project [110]. All of them are brought into the standardised tabular structure outlined in Table 3.2 by explicit maps between the various naming schemes and additional heuristics identifying common fuel-type or technology keywords like *lignite* or *CHP* in the *Name* column. Furthermore, the *Name* column is cleaned by removing frequently occurring tokens, *power plant* or block numbers, for instance.

Since OPSD, ENTSO-E PPL and ESE report individual power plant units for at least some power plants, in a first step we use the deduplication mode of the java application Duke to determine units of the same power plant. Duke [111] is a free software extension of the search engine library Lucene that determines probabilities whether pairs of records (of the same or different tables) refer to the same entity. It computes conditional probabilities $p_c^{i,j} := P(M^{i,j}|x_c^{i,j})$ for the event $M_{i,j}$:= "records i and j match" given the data $x_c^{i,j}$ in column c of these records from mostly character-based similarity metrics like the *Jaro distance* or the *Q-gram distance* [112] skewed into a configurable interval and combines them into an overall matching probability as

$$p_{i,j} := P(M^{i,j}|\cap_c x_c^{i,j}) = \frac{\prod_c p_c^{i,j}}{\prod_c p_c^{i,j} + \prod_c (1 - p_c^{i,j})} \,. \tag{3.1}$$

1 BE, DE, FR, HU, IE, IT, LT as listed by the Open Power System Data project at http://open-power-system-data.org/data-sources#23_National_sources

Column	Argument
Name	Power plant name
Fueltype	{Bioenergy, Geothermal, Hard Coal, Hydro, Lignite, Nuclear, Natural Gas, Oil, Solar, Wind, Other }
Technology	{CCGT, OCGT, Steam Turbine, Combustion Engine, Run-Of-River, Pumped Storage, Reservoir }
Set	{PP, CHP}
Capacity	Generation capacity in MW
lat/lon	Latitude and Longitude
Country	{EU-27 + CH + NO (+ UK) minus Cyprus and Malta}
YearCommissioned	Commissioning year
File	Source file of the data record
projectID	Identifier of the power plant in the original source file

Table 3.2: Standardised data structure for the power plant databases.

This formula, a simplified variant of Naive Bayesian Classification, can be derived from the Bayes Theorem under the assumptions of pairwise conditional independence of the $x_c^{i,j}$ and unbiased prior probabilities whether two records match or do not, i. e. $P(M^{i,j}) = P((M^{i,j})^C) = 0.5$. The former assumption underlies all naive bayesian classifiers and ignores for instance the correlation between technology and capacity (run-of-the-river turbines are typical small (20 MW), whereas nuclear power plants are typically large, with a median capacity of 2 GW). The latter assumption means literally that any two power plant entries from two different datasets have a prior probability of 50% to refer to the same power plant, while the real probability is less than $\frac{N}{N^2} = \frac{1}{N}$, seen from the comparison of two identical datasets of length N. To include such a more realistic prior assumption into the matching process would require changing several internals of the Duke library and is out of the scope of this work. Nevertheless, the model has already been successfully applied in practice [113, 114].

For the aggregation of power plant units, Duke is configured to use the metrics and intervals described in Table 3.3 to return the probabilities $p_{i,j} > 0.985$ between likely pairs i and j. In the power plant matching tool of the authors, these are used as edges in a directed graph of records and the cliques of this graph[2] are aggregated as power plants. Note the low end of the interval for measuring the similarity of the fuel-type chosen to prevent merging units with different fuel-types into the same power plant.

For linking the six databases, PPM runs Duke in Record linkage mode on every pair of databases and determines the most likely links above the threshold of 0.985. These links are joined to chains by collecting the records across all databases that match to the same plant in any database. The chains are reduced by keeping only the longest chains, until they are consistent, i. e. each power

2 A clique in a directed graph is a subset of the nodes such that every two distinct nodes are adjacent.

Column	Deduplication			Record linkage		
	Comparator	low	high	Comparator	low	high
Name	JaroToken	0.09	0.99	JaroToken	0.09	0.99
Fueltype	QGram	0.09	0.65	QGram	0.09	0.7
Country	QGram	0.01	0.51	QGram	0.0	0.53
Capacity	Numeric	0.49	0.51	Numeric	0.1	0.75
Geoposition	Geo	0.05	0.55	Geo	0.1	0.8

Table 3.3: Duke comparison metrics and intervals for aggregation of power plant units (deduplication) and linking different power plant tables (record linkage). JaroToken breaks the full string into several tokens, evaluates the Jaro Winkler distance metric for each and returns the compound Jaccard index [112]. These parameters have been chosen by hand and plausibility, while instead they should be tuned for a representative subset to an ideal match by Duke's Genetic algorithm.

plant appears only in at most one chain. This could likely be improved by joining chains recursively, while keeping track of the chain probability based on a variant of Eq. (3.1) at the expense of not being able to rely on the fast pandas routines any more.

For the remaining chains the power plant information is aggregated by taking the most frequent *Fueltype*, a comma separated list of the *Technology*(-ies), the mean *lat/lon* and the median *Capacity*. The latter ensures that the shutdown or addition of a block of a power plant which is not yet reflected in a minority of databases does not distort the final capacity.

The compound dataset, at the time of writing, contains 3465 power plants with a total capacity of 705 GW. Less than a third of these are represented in 3 or more sources, but still account for about two third of the capacity. 2494 small power plants with an average capacity of about 93 MW appear in only two databases. There are a further 841 power plants with 21.7 GW capacity in the OPSD dataset unmatched by the other free datasets and exclusively compiled from official sources. After including these power plants the mean absolute error from the SO&AF country-wise capacity is at 9% of the average capacity and below a 33% deviation in each single country except for Bulgaria and Lithuania. Refer to the companion paper for a more detailed comparison of the free dataset with the proprietary World Electric Power Plants dataset [115].

Hydro-electric generation

Existing hydroelectric capacities ensue from the same matching process as the conventional power plants, particularly based on the sources ESE and ENTSO-E. The capacities are categorised into run-of-river, reservoir and pumped storage. Reservoir and pumped storage have energy storage capacities that are estimated by distributing the country-aggregated energy storage capacities reported by [116, 117] in proportion to power capacity. Run-of-river as well as

reservoir hydro capacities receive an hourly-resolved in-flow of energy. Extensions to the current hydro capacities are not considered.

Renewable generation time series like hydro-electric in-flow, wind and solar are derived from the re-analysis weather dataset CFSv2 by the US National Oceanic Atmospheric Administration [118]. It provides wind speeds, irradiation, surface-roughness, temperature and run-off on a $0.2° \times 0.2°$ spatial raster ($x \in \mathcal{X}$) in hourly resolution since 2011.

The simplified in-flow time series is generated as in [40, 116] by aggregating the total potential energy at height h_x relative to ocean level of the CFSv2 run-off data $\mathcal{R}_x$ in each country c by

$$G_c^H(t) = \mathcal{N} \sum_{x \in \mathcal{X}(c)} h_x \mathcal{R}_x(t) \tag{3.2}$$

where $\mathcal{N}$ is chosen so that $\int_t G_c^H(t)\,\mathrm{d}t$ matches the U.S. Energy Information Administration (EIA) annual hydroelectricity generation [119]. The in-flow is distributed to all run-of-river and reservoir capacities in proportion to their power capacity.

Wind generation

Following the methodology in [78], the wind speeds at 10 m above ground $u_x^{10\,\mathrm{m}}(t)$ are extrapolated to turbine hub-height h using the surface roughness z_x^0 with the logarithmic law

$$u_x^h(t) = u_x^{10\,\mathrm{m}}(t) \frac{\ln\left(h/z_x^0\right)}{\ln\left(10\,\mathrm{m}/z_x^0\right)}\,. \tag{3.3}$$

The capacity factor of each raster cell x for a wind turbine with powercurve $P_w(u)$ and generator capacity P_w^{max} is determined as

$$c_{x,w} = \frac{\left\langle P_w(u_x^h(t)) \right\rangle_t}{P_w^{max}} \tag{3.4}$$

and together with the usable area $A_{x,w}$ the maximally installable wind generation capacity $G_{x,w}^{max} = 0.3 \cdot 10\,\mathrm{MW/km^2} \cdot A_{x,w}$ is calculated, where $10\,\mathrm{MW/km^2}$ is the technical potential density [28] and 0.3 arises out of considering competing land use and issues of public acceptance.

The usable area is restricted by the following constraints: Onshore wind can only be built in land use types of the CORINE Land Cover database [120] associated to *Agricultural areas* and *Forest and semi natural areas* and furthermore a minimum distance of 1000 m from *Urban fabric* and *Industrial, commercial and transport units* must be respected. Offshore wind can only be constructed in water depths up to 50 m. Additionally, all nature reserves and restricted areas listed in the Natura2000 database [121] are excluded. The wind generation potential in Germany is shown in Figure 3.2.

Each Voronoi cell V of a substation covers multiple cells of the re-analysis weather grid, as described by the indicatormatrix $\mathcal{I}_{V,x} = \mathrm{area}(V \cap x)/\mathrm{area}\, x$,

and we distribute the wind turbine capacity according to a normed capacity layout

$$G^{p.u.}_{V,x,w} = \frac{\mathcal{I}_{V,x} c_{x,w} G^{max}_{x,w}}{\sum_x (\cdot)} \tag{3.5}$$

which prefers cells x with high capacity factor $c_{x,w}$ and high maximally installable capacity $G^{max}_{x,w}$. The wind generation availability time-series at a substation with Voronoi cell V is, thus,

$$\bar{g}_{V,w}(t) = G_{V,w} \frac{\sum_x G^{p.u.}_{V,x,w} P_w(u^h_x(t))}{P^{max}_w} \tag{3.6}$$

for an installed capacity $G_{V,w}$. This capacity is expandable until reaching $G^{max}_{x,w}$ in any grid cell up to

$$G^{max}_{V,w} = \min_{\{x|\mathcal{I}_{V,x}>0\}} \frac{\mathcal{I}_{V,x} G^{max}_{x,w}}{G^{p.u.}_{V,x,w}} \,. \tag{3.7}$$

The power curve of the turbine Vestas V112 with a turbine capacity of 3 MW and a hub height 80 m is used to generate the onshore wind time-series and the National Renewable Energy Laboratory (NREL) Reference Turbine with 5 MW at 90 m is used for the offshore wind time-series. The accuracy of the wind generation time-series are improved to account for effects of spatial wind speed variations within a grid cell by smoothing the power curves with a Gaussian kernel as

$$P_w(u) = \eta \int_0^\infty P_0(u'') \frac{1}{\sqrt{2\pi\sigma_0^2}} \mathrm{e}^{-\frac{(u-u'+\Delta u)^2}{2\sigma_0^2}} \mathrm{d}u' \,, \tag{3.8}$$

where $\eta = 0.95$, $\Delta u = 1.27\,\mathrm{m/s}$ and $\sigma_0 = 2.29\,\mathrm{m/s}$ are the optimal parameters minimising the error between the re-analysis-based time-series and a year of Danish wind feed-in [78]. A study comparing the wind generation time-series based on the re-analysis MERRA-2 dataset for a 20 year period to the per-country wind feed-in and several wind park generation measurements found non-negligible discrepancies of the optimal bias correction parameters between different countries [122]. They will be incorporated in a future version of the presented model.

Photovoltaic generation

The solar availability time-series and maximally installable capacity per substation are like the wind generation in the previous section based on the re-analysis weather dataset CFSv2 and we will focus on the differences.

The photovoltaic generation $P_{x,s}(t)$ for a panel of nominal capacity P^{max}_s of a point in time t and space resp. grid cell x is calculated from the re-analysis short-wave radiation. A direct, a diffuse and a ground-reflected irradiation component are derived from the clearsky model of Reindl [123] and geomet-

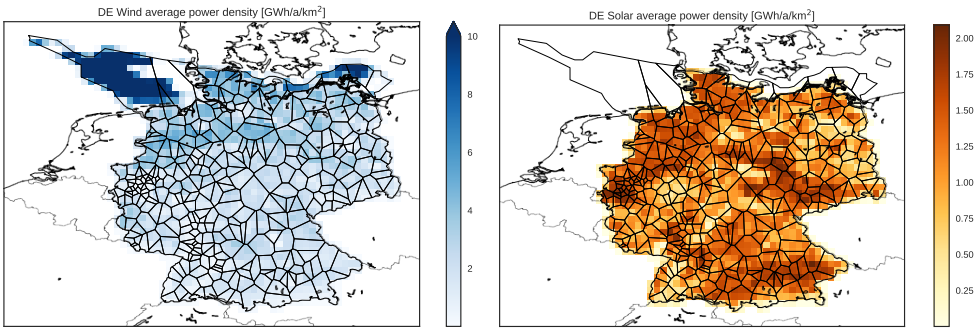


Figure 3.2: Wind (l.) and solar (r.) potential power generation after landuse restrictions for weather grid cells in Germany. The generation of all grid cells in a Voronoi cell (also shown in black) is fed into the central substation.

ric relations of the trajectory of the sun and the tilted panel surface [124, 125]. An effective electric model by Bofinger et al. [126] determines the active power output from the total irradiation and the ambient temperature. Implementation details are found in the pv sub-package of the atlite package [127] are based on the Renewable Energy Atlas developed at Aarhus University [78]. An inverter inefficiency reducing the solar generation by 21% is assumed to match on average solar capacity factors reported by [128].

For each raster cell $x \in \mathcal{X}$ the capacity factor $c_{x,s}$ and the maximally installable capacity $G^{max}_{x,s} = 0.01 \cdot 145\,\mathrm{MW/km^2} \cdot A_{x,s}$ is determined as for wind, with the difference that the high technical potential of $145\,\mathrm{MW/km^2}$ corresponds to an unrealistic full surface of solar cells, which is offset by allowing only up to 1%. The permitted CORINE land use types are *Artificial surfaces*, most *Agricultural areas* except for those with forests and then including only few sub-categories of *Forest and semi natural areas*: *Scrub and/or herbaceous vegetation associations*, *Bare rocks* and *Sparsely vegetated areas*. Figure 3.2 shows the solar generation potentials.

Equations (3.5)-(3.7) are applied analogously to generate the solar availability time-series $\bar{g}_{V,s}(t)$ and to find the solar expansion potential $G^{max}_{V,s}$. The reference solar panel is the U-EA120 type thin-film silicon panel by Kaneka. Note that photovoltaic generation models based on satellite imaging as provided by METEOSAT have been found to recover measured feed-in time-series and capacity factors with a higher accuracy [129, 130].

Demand

There are two classes of approaches in the literature: Most employ a top-down approach by distributing the historical demand curves of each country to the substations in the country according to some geographical key. [87] uses population, ELMOD uses a weighted convex combination of population and gross domestic product [91], the REMIX model relies on the area of artificial land-surface from land-cover data [28]. Hülk* et al. [131] extend this approach in two aspects: firstly, they use a rule-based partitioning of the geographical surface combining Voronoi cells with administrative boundaries; secondly, the electric-

Circuit length	DE			EU		
in 1000 km	220 kV	300 kV	380 kV	220 kV	300 kV	380 kV
ENTSO-E	13.70	0.0	20.92	117.25	9.96	146.82
PyPSA-Eur	11.04	0.0	23.76	115.63	10.00	152.55

Table 3.4: AC lines circuit lengths of the whole of Europe and Germany as an example

ity consumption for each load area is derived for each sector (residential, industrial, retail and agricultural) separately based on OpenStreetMap land use and industrial infrastructure data as well as population density. Alternatively, demand time-series can be compiled in a more involved bottom-up approach by combining industrial and domestic reference profiles according to regionalized sectoral statistics and finally overlaid by country load profiles, as recently attempted at the KIT [132]; unfortunately detailed documentation of the results has not been published.

For PyPSA-Eur, the hourly electricity demand profiles of each country from 2011 to 2016 are taken from the European Network of Transmission System Operators for Electricity (ENTSO-E) website [133]. The load time-series is distributed to the substations in each country by 60% according to the gross domestic product (GDP) as a proxy for industrial demand and by 40% as residential demand according to population in a Voronoi cell. The 60-40% split is based on a linear regression analysis of the per-country data and agrees with values used in [91]. The two statistics are mapped from the Eurostat Regional Economic Accounts database (nama_10-reg) for NUTS3 regions to the Voronoi cells in proportion to their geographic overlap.

3.3 VALIDATION

Network total line lengths

In this subsection, total line circuit lengths at different voltage levels in the model are compared with official statistics from ENTSO-E. The lengths of AC circuits [134] per voltage level and country are compared to aggregations of line lengths times circuits from PyPSA-Eur, so that cross-border lines are equally attributed to both adjacent countries. In Table 3.4 the total line lengths for the whole of Europe and Germany are presented as examples. Considering the data for all countries, the lines in the PyPSA-Eur dataset deviate from the ENTSO-E lengths of circuits by a mean absolute error of 14% for 220 kV, 11% for 300 kV and 9% for 380 kV lines. These deviations are accounted for by the fact that the ENTSO-E map [94] from which the PyPSA-Eur network is derived is only an artistic representation and does not follow the exact contours of each transmission line. Some differences may also be due to incorrect classification of 220 kV lines as 380 kV lines, or due to the fact that the ENTSO-E map on which PyPSA-Eur is based is more up-to-date with regard to recent upgrades to the transmission network.

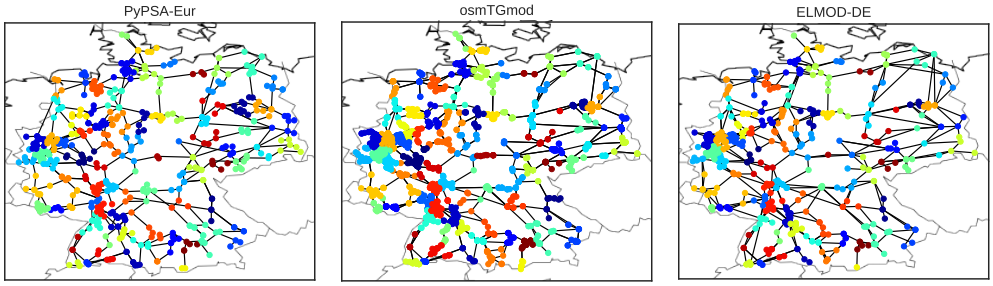


Figure 3.3: 80 clusters jointly identified by colour in the network topologies of the models PyPSA-Eur, osmTGmod and ELMOD-DE.

Network topology

While the total circuit lengths might agree, it does not necessarily mean that the lines are in the right places with the right topology. The *ENTSO-E Interconnected network map* [94] is the source of the network topology in PyPSA-Eur, so it naturally agrees well in visual examination. A comparison with the network topology published with the TYNDP [83] is hindered by shortened substation names and missing geo-locations in that dataset. Instead, in this section the network topology of PyPSA-Eur is compared to the open network datasets available for Germany, which are derived using a different methodology. New, experimental algorithms are presented to compare network topologies, since few appropriate algorithms exist in the literature. This is a difficult problem because neither the locations nor the number of the buses and lines in the different models necessarily agree. Our methodology works by first establishing a common set of aggregated buses for the different networks, then comparing the networks once all lines and other elements have been reattached to the aggregated buses.

More precisely, we present a new technique of applying k-means clustering to measure the similarity of several geo-located network models, specifically the German 220 kV and 380 kV voltage layers represented in PyPSA-Eur, osmTGmod and ELMOD-DE. The buses of all networks are jointly clustered together into k clusters by minimizing the distances in each cluster π_j

$$\mathcal{D}(\pi_j) = \sum_{j=1}^{k} \sum_{a \in \pi_j} w(a)\, ||a - m_j||^2 \tag{3.9}$$

from its buses a to the center $m_j = \sum_{b \in \pi_j} w(b)\, b / \sum_{b \in \pi_j} w(b)$ with chosen bus-weights $w(b)$. Since Eq. (3.9) expands to a polynomial in the scalar products $\langle a, b \rangle$ between buses, kernel k-means allows the use of general scalar products by evaluating them on every pair of buses [135]. In the kernel of the scalar product we propose

$$K_{a,b} = e^{-||a-b||_2^2/N} + \nu B^{\dagger}_{a,b} \tag{3.10}$$

spatial cohesion comes from the first term, a radial basis function based on the Euclidean distance $||\cdot||_2$ over the number of buses N, which favours short ge-

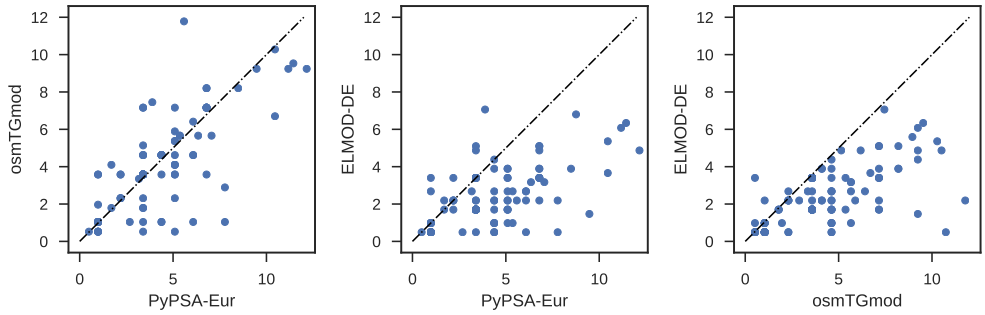


Figure 3.4: Capacity connecting the same clusters in GW

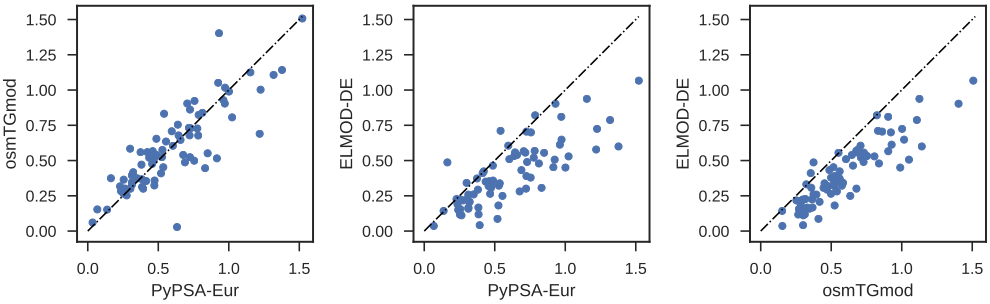


Figure 3.5: Line volume at buses in the same cluster in TWkm

ometric distances and connectedness over spatial convexity, while the pseudo inverse of the admittance matrix B in each network induces an electrical reactance distance, shown to lead to electrically cohesive clusters [136]. With the weights $w(a \in \text{PyPSA-Eur/ELMOD-DE}) = 5$ and $w(a \in \text{osmTGmod}) = 1$ balancing the five times as many buses in osmTGmod and the relative weight $\nu = 200$, the clustering algorithm is able to distribute 80 clusters across the three networks, by starting from the labels found by regular k-means and by picking the best result from 20 runs. The networks and the associated buses are shown in Figure 3.3.

As the clusters are the same in each network, the aggregate capacity between every two clusters is now comparable for different networks. Unfortunately, Figure 3.4 reveals bad agreement for the capacity between pairs of clusters, due to its high sensitivity to errors arising from clustering topologically distinct buses; i. e. buses lying on distinct lines are inadvertently joined together. Increasing the weight of the electrical distance ν dampens the appearance of these associations, but worsens convergence and increasingly finds solutions in which clusters in the electrically well connected areas as the Ruhrpott detach from one of the three networks. With a lower number of clusters between 20 to 40, ν can be increased by an order of magnitude and the topological errors are less important, then the aggregate capacities between large network zones can be compared and deviations identified.

Another approach is simply to aggregate line volumes within and attached to each cluster of buses. These are compared in Figure 3.5 and turn out to be quite robust against topologically problematic associations and show a high

	PyPSA-Eur	osmTGmod	ELMOD-DE
PyPSA-Eur	1.000	0.846	0.744
osmTGmod		1.000	0.868
ELMOD-DE			1.000

Table 3.5: Pearson correlation coefficients between line volume at different buses

correlation across networks in Table 3.5. ELMOD-DE has proportionally less line volume than osmTGmod and PyPSA-Eur, but with approximately the same spatial distribution. osmTGmod and PyPSA-Eur agree well.

Potentials for expansion of renewables

Geographic potentials for the expansion of wind and solar power depend strongly on technical, environmental, social and political constraints. Different organisations offer different assessments of acceptable potentials, which involve a complex balance between land availability, landscape impact and species protection. In this section we compare aggregated total potentials for Germany in the PyPSA-Eur model derived using the methodologies described in Sections 3.2 and 3.2 with other studies.

For onshore wind, there is an installable potential of 447 GW in Germany in the model. Assessments in the literature range from 198 GW [137] (based on a 'realistic' restriction to 2% of total land area, although 8% is available when excluding forests and protected areas) up to 1190 GW [138] (using 13.8% of the total land area, ignoring species protection and whether locations are economically exploitable).

For offshore wind, 86 GW of fixed-foundation capacity is available for installation in PyPSA-Eur in Germany. Estimates in the literature range from 38 to 85 GW [139, 140].

345 GW of solar photovoltaics is installable in Germany in PyPSA-Eur. The potential depends strongly on what land areas are permitted, but typical values range from 360 GW [139] to 400 GW [140] (including roofs, facades and railway/motorway sidings, but excluding free space).

Model validation: Linear optimal power flow

As a validation of the model at large, it is formulated within the PyPSA framework [93, 141] and the linear optimal power flow of the European peak-load hour is considered to check the feasibility of the combined network, generation and demand data for supplying the most extreme demand. Solar- and wind feed-in are not allowed to reduce the load, while hydro-electric installations may be discharged at their full power capacity. European overall peak-load of 0.54 TW in the dataset happens at 17:00 on 08-02-2012 and leads to 25 GW of load shedding in the vicinity of large agglomerations, primarily in Paris (6 GW) and London (5.4 GW). Since there is sufficient generation capacity to cover the

peak load, this load-shedding is due to grid bottlenecks which appear in the model (but not in reality, since grid bottlenecks do not cause load-shedding in today's European network). The amount of shedding decreases considerably by lifting capacity constraints on short lines, for example if lines shorter than 25 km are not limited in power capacity, only about 10 GW, 2%, of load has to be shed. Similarly easing local restrictions by clustering the network using a k-means algorithm as detailed in [142] to 1500 buses reduces shedding to 1.3% of peak load, while clustering to 356 buses allows the whole demand to be supplied. This is indicative of local assignment errors of load and supply, when the Voronoi cells used to assign load and generators to transmission substations do not represent the true distribution grid topology at each transmission substation, and/or an underrepresentation of inner-city underground cabling, which is not always shown on the map. Clustering the network, so that each bus represents a larger area, smooths out local assignment errors. Since there are several heuristic remedies from expanding the loaded lines over rearranging the load to using clustered topologies as done manually by [84, 86], we decided not to perform any corrections, but publish the dataset as is.

3.4 LIMITATIONS

While the benefit of an openly available, functional and partially validated model of the European transmission system is high, many approximations have been made due to missing data. In this section we summarise the limitations of the dataset, both as a warning to the user and as an encouragement to assist in improving the approximations.

The grid data is based on a map of the ENTSO-E area [94] that is known to contain small distortions to improve readability. Since the exact impedances of the lines are unknown, approximations based on line lengths and standard line parameters were made that ignore specific conductoring choices for particular lines. There is no openly available data on busbar configurations, switch locations, transformers or reactive power compensation assets.

Using Voronoi cells to aggregate load and generator data to transmission network substations ignores the topology of the underlying distribution network, meaning that assets may be connected to the wrong substation. Assumptions have been made about the distribution of load in each country proportional to population and GDP that may not reflect local circumstances. Openly available data on load time series may not correspond to the true vertical load [143] and is not spatially disaggregated; assuming, as we have done, that the load time series shape is the same at each node within each country ignores local differences.

Information on existing wind, solar and small hydro, geothermal, marine and biomass power plants are excluded from the dataset because of a lack of data availability in many countries. Approximate distributions of wind and solar plants in each country can be generated that are proportional to the capacity factor at each location.

The database of hydro-electric power plants does not include plant-specific energy storage information, so that blanket values based on country storage totals have been used. Inflow time series are based on country-wide approximations, ignoring local topography and basin drainage; in principle a full hydrological model should be used. Border connections and power flows to Russia, Belarus, Ukraine, Turkey and Morocco have not been taken into account; islands which are not connected to the main European system, such as Malta, Crete and Cyprus, are also excluded from the model.

3.5 CONCLUSIONS

In this chapter a dataset PyPSA-Eur has been presented of the full European transmission system, including a high resolution grid model, load data, a new geo-referenced database of conventional power plants, potentials for the expansion of wind and solar, and time series for the load and variable renewable power availability. The model is only based on publicly available and open datasets, and all code and data has been made available, making it the first open model of the full European system at such high spatial resolution.

Validation results, including a new technique for validating grid data, have been presented which demonstrate that the model is a plausible approximation of the European power system. Further validation is desirable.

The dataset has been designed primarily for the optimisation of future investment in generation and transmission, but can also be adapted to studies of the operation of the current power system. We hope that it will contribute towards a transparent discussion of the future needs of the European energy system.

4 SPATIAL SCALE IN OPTIMAL RENEWABLE ENERGY NETWORKS[1]

4.1 INTRODUCTION

Optimising investment in the electricity system to reduce greenhouse gas emissions is computationally intensive. Transmission investment should be jointly optimised with generation investment, so that the benefits of exploiting the sites with the best renewable resources can be balanced against the network expansion costs; continental-scale areas should be considered, so that synoptic-scale weather variations ($\sim 600 - 1000$ km), which particularly affect wind generation, can be balanced; at the same time, high spatial detail is required to capture both variations in renewable resources and existing transmission bottlenecks.

Previous studies have typically sacrificed at least one of these goals. In some studies, only a single node per country or group of countries has been considered [28, 36, 38, 40], ignoring national transmission networks and local differences in weather conditions. Other studies consider the transmission network in detail, but only for single countries [41, 144], neglecting the benefits of international cooperation. Other studies maintain both a pan-continental scope and transmission network detail, but fix the generation fleet and only optimise the transmission network [42, 44].

In this chapter it is attempted to bring a more systematic approach to the question of spatial resolution in electricity system optimisations. A clustering methodology called '*k*-means' is used to successively reduce the number of nodes in the European transmission network from its full level of spatial detail down to a level where there is only one node per country. The effects on the results of the optimal investments in generation and transmission are then studied as the spatial resolution is changed. A high spatial resolution reveals transmission bottlenecks that might either restrict welfare-enhancing transfers or force transmission upgrades; ignoring these effects in a low resolution model leads to an underestimate of the total costs. In a low resolution model one is also forced to average the renewable resources over a larger area, which lowers the average capacity factors, even with a weighting towards better sites; at high resolution the sites with the highest capacity factors can be fully exploited, particularly for wind.

In recent decades as large-scale optimisation has gained in importance, many methods have been suggested in the literature to reduce the whole network to a number of clusters, rather than focussing on a binary exterior-interior division like the Ward or Radial-Equivalent-Independent methods [145]. Standard clustering algorithms from complex network theory [146] have been applied

1 This chapter is published as "The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios" by Hörsch and Brown [142]. It was only slightly modified for this work.

on the network structure, including k-means clustering [147] on electrical distance between buses [136, 148, 149] and spectral partitioning of the Laplacian matrix [150]. Equivalents based on zonal PTDFs were considered in [151], while a methodology based on available transfer capacities (ATC) was developed in [152]. A more economic focus was taken in [153], where buses were clustered based on similar average locational marginal price (LMP). The final report of the recent e-Highways 2050 project [154] that considered network expansion needs in Europe contains both a summary of network clustering methods and suggestions for mixed metrics combining several characteristics to define nodal similarity.

In the following sections we review the clustering methodology, the investment model, the data input for the European electricity system and the results for different aggregation scales and levels of grid expansion.

4.2 METHODOLOGY FOR NETWORK REDUCTION

We first describe the method to derive a clustered equivalent network, with fewer buses and lines, from a more detailed network. First, the network buses are partitioned into clusters, then an equivalent network is constructed with one bus per cluster and aggregated lines between the new buses.

The network reduction method used here for an equivalent network of k buses consists of the following steps (steps 3-6 are schematically represented in Fig. 4.1)

1. Univalent buses, i. e. network stubs or 'dead-ends', are aggregated to their neighbours in an iterative process until all buses are multi-valent, since such stubs are typically short lines, connecting single generators to the main network.

2. The remaining buses labelled by n are assigned a weight w_n proportional to the load and today's conventional generation capacities at the bus and coordinates x_n based on their geographical location.

3. The k-means algorithm is used to find the geographical positions of k centroids $\{x_c\}$ for $c = 1, \dots k$ by minimising the weighted sum of squared distances from each centroid to its clustered members N_c:

$$\min_{\{x_c\}} \sum_{c=1}^{k} \sum_{n \in N_c} w_n ||x_c - x_n||^2 \tag{4.1}$$

   To lessen the risk of finding local minima the k-means algorithm is run on 10 different starting conditions and all but the best found centroid configuration are discarded. Further, the clustering is constrained so that for each country and synchronous zone a number of clusters proportional to its overall mean load is chosen.

4. A new bus c is created at each centroid x_c to represent the set of clustered nodes.

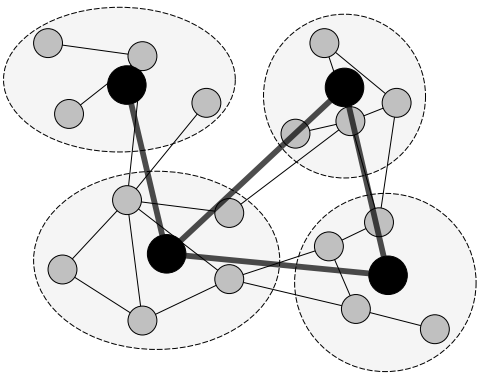


Figure 4.1: Schematic clustering representation of steps 4-6. Detailed buses (gray) are grouped to clusters, so that internal distances are minimized. The cluster is represented by its centroid as new bus (black), to which generators and storage units are aggregated. Lines between clusters are replaced by a single equivalent line.

5. All generators, storage units and loads that were connected to the original buses in N_c are then aggregated by technology type at the equivalent bus c. The maximum expansion potential of generators of the same technology type are added for the new aggregated generator, while the weather-dependent availability time series for renewable generators are averaged with a weighting.

6. Lines between the clusters are replaced by a single equivalent line with a length of 1.25 times the crow-flies-distance, whose capacity is given by the sum of the capacity of the replaced lines, and whose impedance is given by the equivalent impedance of the parallel lines.

Note that by focussing on the geographical distribution of the load and conventional generation, this method ignores both the electrical distance between the buses and the grid topology. Electrical distance is ignored because the network clustering should be independent of existing grid capacities, given that these capacities will be optimised later; for the optimisation, geographical distance is more important because it determines the cost of the grid expansion. The topology is ignored because it is expected that the grid topology was designed to connect major load and conventional generation centres, so that focussing on the load and generation is sufficient to capture the important conglomerations and the transmission corridors between them.

N-1 security is modelled by scaling down the available transmission capacity to 0.7 times nominal capacity for high resolution networks ≥ 200 clusters and linearly shrinking it down until reaching half the nominal values at 37 clusters to account for cluster-internal bottlenecks.

The network model is clustered down to $k = 37, 45, 64, 90, 128, 181, 256, 362$ buses (see Figure 4.2 for the clusterings with 64 and 362 buses) and the different results of the system optimisation are examined for each level of clustering in several grid expansion scenarios.

The network reduction algorithms are implemented in the free software PyPSA Version 0.8.0 [75, 141] which is developed at the Frankfurt Institute for Advanced Studies (FIAS). PyPSA uses the scikit-learn Python package [155] for the k-means clustering.

4.3 MODEL FOR INVESTMENT OPTIMISATION

The model minimises total annual system costs, which include the variable and fixed costs of generation, storage and transmission, given technical and physical constraints.

To obtain a representative selection of weather and demand conditions while keeping computation times reasonable, the model is run over every third hour of a full historical year of weather and demand data assuming perfect foresight, with 2012 chosen as the representative year. Each time point t is weighted by $w_t = 3$ in the objective function and storage constraints, to account for the fact that it represents three hours.

The optimisation minimises total annual system costs, with objective function

$$\min_{\substack{G_{n,s},F_\ell,\\ g_{n,s,t},f_{\ell,t}}} \left[\sum_{n,s} c_{n,s} G_{n,s} + \sum_{\ell} c_\ell F_\ell + \sum_{n,s,t} w_t o_{n,s} g_{n,s,t} \right] \tag{4.2}$$

consists of the capacities $G_{n,s}$ at each bus n for generation and storage technologies s and their associated annualised fixed costs $c_{n,s}$, the dispatch $g_{n,s,t}$ of the unit in time t and the associated variable costs $o_{n,s}$, and the line capacities F_ℓ for each line ℓ (including both high voltage HVAC and HVDC lines) and their annualised fixed costs c_ℓ.

The dispatch of conventional generators $g_{n,s,t}$ is constrained by their capacity $G_{n,s}$

$$0 \leq g_{n,s,t} \leq G_{n,s} \qquad \forall\, n, s, t \tag{4.3}$$

The maximum producible power of renewable generators depends on the weather conditions, which is expressed as an availability $\bar{g}_{n,s,t}$ per unit of its capacity:

$$0 \leq g_{n,s,t} \leq \bar{g}_{n,s,t} G_{n,s} \qquad \forall\, n, s, t \tag{4.4}$$

The energy levels $e_{n,s,t}$ of all storage units have to be consistent between all hours and are limited by the storage energy capacity $E_{n,s}$

$$\begin{aligned} e_{n,s,t} = \eta_0^{w_t} e_{n,s,t-1} - \eta_1 w_t \left[g_{n,s,t}\right]^- + \eta_2^{-1} w_t \left[g_{n,s,t}\right]^+ \\ + w_t g_{n,s,t,\text{inflow}} - w_t g_{n,s,t,\text{spillage}} \\ 0 \leq e_{n,s,t} \leq E_{n,s} \qquad \forall\, n, s, t \end{aligned} \tag{4.5}$$

Positive and negative parts of a value are denoted as $[\cdot]^+ = \max(\cdot, 0)$, $[\cdot]^- = -\min(\cdot, 0)$. The storage units can have a standing loss η_0, a charging efficiency η_1, a discharging efficiency η_2, inflow (e. g. river inflow in a reservoir) and spillage. The energy level is assumed to be cyclic, i. e. $e_{n,s,t=0} = e_{n,s,t=T}$.

CO_2 emissions are limited by a cap CAP_{CO2}, implemented using the specific emissions e_s in CO_2-tonne-per-MWh of the fuel s and the efficiency $\eta_{n,s}$ of the generator:

$$\sum_{n,s,t} \frac{1}{\eta_{n,s}} w_t g_{n,s,t} \cdot e_s \leq \mathrm{CAP}_{CO2} \quad \leftrightarrow \quad \mu_{CO2} \tag{4.6}$$

In all simulations this cap was set at a reduction of 95% of the electricity sector emissions from 1990.

The (inelastic) electricity demand $d_{n,t}$ at each bus n must be met at each time t by either local generators and storage or by the flow $f_{\ell,t}$ from a transmission line ℓ

$$\sum_s g_{n,s,t} - d_{n,t} = \sum_\ell K_{n\ell} f_{\ell,t} \qquad \forall\, n, t \tag{4.7}$$

where $K_{n\ell}$ is the incidence matrix of the network. This equation is essentially KCL.

In this chapter it is assumed that the linear load flow is a good approximation for a well-compensated transmission network [44, 49]. To guarantee the physicality of the network flows, in addition to KCL, KVL must be enforced in each connected network. KVL states that the voltage differences around any closed cycle in the network must sum to zero. If each independent cycle c is expressed as a directed combination of lines ℓ by a matrix $C_{\ell c}$ then KVL becomes the constraint

$$\sum_\ell C_{\ell c} x_\ell f_{\ell,t} = 0 \qquad \forall c, t \tag{4.8}$$

where x_ℓ is the series inductive reactance of line ℓ. Note that point-to-point HVDC lines have no cycles, so there is no constraint on their flow beyond KCL.

The flows are also constrained by the line capacities F_ℓ

$$|f_{\ell,t}| \leq F_\ell \qquad \forall\, \ell, t \tag{4.9}$$

Although the capacities F_ℓ are subject to optimisation, no new grid topologies are considered.

Since line capacities F_ℓ can be continuously expanded to represent the addition of new circuits, the impedances x_ℓ of the lines would also decrease. In principle this would introduce a bilinear coupling in equation (4.8) between the x_ℓ and the $f_{\ell,t}$. To keep the optimisation problem linear and therefore computationally fast, x_ℓ is left fixed in each optimisation problem, updated and then the optimisation problem is rerun in up to 4 iterations to ensure convergence, following the methodology of [45].

In order to investigate the interactions of spatial scale with transmission expansion, the sum of all transmission line capacities (HVAC and HVDC) multi-

Quantity	Overnight Cost [€]	Unit	FOM [%/a]	Lifetime [a]
Wind onshore	1182	kW_{el}	3	20
Wind offshore	2506	kW_{el}	3	20
Solar PV	600	kW_{el}	4	20
Gas	400	kW_{el}	4	30
Battery storage	1275	kW_{el}	3	20
Hydrogen storage	2070	kW_{el}	1.7	20
Transmission line	400	MWkm	2	40

Table 4.1: Investment costs

plied by their lengths l_ℓ is restricted by a line volume cap $\mathrm{CAP}_{\mathrm{trans}}$, which is then varied in different simulations:

$$\sum_\ell l_\ell \cdot F_\ell \leq \mathrm{CAP}_{\mathrm{trans}} \quad \leftrightarrow \quad \mu_{\mathrm{trans}} \tag{4.10}$$

The caps are defined in relation to today's line capacities F_ℓ^{today}, i. e.

$$\mathrm{CAP}_{\mathrm{trans}} = x \cdot \mathrm{CAP}_{\mathrm{trans}}^{\mathrm{today}} = x \cdot \sum_\ell l_\ell \cdot F_\ell^{\mathrm{today}} \,. \tag{4.11}$$

The discussion in Section 4.5 starts off with the no expansion scenario, where $\mathrm{CAP}_{\mathrm{trans}} = \mathrm{CAP}_{\mathrm{trans}}^{\mathrm{today}}$ impedes any network expansion beyond today's line capacities. In this scenario transmission bottlenecks restrict the exploitation of the best renewable sites and the smoothing effects across the continent; generation is forced to be more localised and renewable variability may have to be balanced by storage. Then, five expansion scenarios are studied by gradually easing the cap $\mathrm{CAP}_{\mathrm{trans}} = x \cdot \mathrm{CAP}_{\mathrm{trans}}^{\mathrm{today}}$ with $x = 1.125, 1.25, 1.5, 2, 3$ until reaching three times today's transmission volume, which is already above the optimal value for overhead lines at high numbers of clusters, as we will discuss in Section 4.5.

The optimisation model was also implemented in PyPSA.

4.4 DATA INPUTS

The network reduction and subsequent investment optimisation were run on a full model of the European electricity transmission system.

The existing network capacities and topology for the ENTSO-E area (including continental Europe, Scandinavia, the Baltic countries, Great Britain and Ireland) were taken from the GridKit extract [96] of the online ENTSO-E Interactive Transmission Map [94]. The model includes all transmission lines with voltages above 220 kV and all HVDC lines in the ENTSO-E area (see Figure 3.1).

In total the model contains 5586 HVAC lines with a volume of 241.3 TWkm (of which 11.4 TWkm are still under construction), 26 HVDC lines with a volume of 3.4 TWkm (of which 0.5 TWkm are still under construction) and 4653 substations.

The hourly electricity demand profiles for each country in 2012 are taken from the European Network of Transmission System Operators for Electricity (ENTSO-E) website [133]. The geographical distribution of load in each country is based on GDP and population statistics for the Nomenclature of Territorial Units for Statistics level 3 ($NUTS_3$) regions.

Electricity generation in the model is allowed from the following technologies: hydroelectricity, natural gas, solar PV, onshore wind and offshore wind. Gas, solar and wind capacities may be expanded within the model constraints.

Existing hydroelectricity capacities (including run-of-river, reservoirs and pumped storage) were compiled by matching databases CARMA [109], GEO [85], DOE Global Energy Storage Database [107] and the PowerWatch project coordinated by the World Resources Institute [110]; no expansion of existing hydro capacities is considered in the model. The hydro energy storage capacities are based on country-aggregated data reported by [116, 117] and the inflow time series are provided by [116].

The only fossil fuel generators in the model are open cycle gas turbines, whose efficiency is assumed to be 39%. Their usage is limited by the CO_2 cap in equation (4.6).

The potential generation time series for wind and solar generators are computed with the Aarhus renewable energy atlas [78] from hourly historical weather data from 2012 with a spatial resolution of $40 \times 40\text{km}^2$ provided by the US National Oceanic and Atmospheric Administration [118].

The distribution of these generators is proportional to the quality of each site given by the local capacity factor multiplied with the maximum installable capacity of the site. However, protected sites as listed in Natura2000 [121] are excluded, as well as areas with certain land use types from the Corine Land Cover database [120], as specified by [28], to avoid building, e. g., wind turbines in urban areas. The maximum water depths for offshore wind turbines is assumed to be 50 m. The maximum installable capacity per bus and generator type is then determined by scaling these layouts until the first site reaches a maximum installation density of 2 and 1.7 MW/km^2 for wind and solar, respectively. These maximum densities are chosen conservatively to take account of competing land use and minimum-distance regulations for onshore wind turbines.

The model contains two extendable types of storage units: batteries and hydrogen storage. Their charging and discharging efficiencies, as well as cost assumptions for their power and energy storage capacities are taken from [156]. It is assumed that the charging and discharging power capacities of a unit are equal, and the energy capacity $E_{n,s} = h_{max,s} * G_{n,s}$ is proportional to this power capacity. The factor $h_{max,s}$ determines the time for charging or discharging the storage completely at maximum power, and is set to $h_{max} = 6\,\text{h}$ for batteries and to 168 h for H2 storage.

Investment and fixed operation and maintenance (FOM) costs for all assets are listed in Table 4.1. The costs for generating assets are based on predictions for 2030 from DIW [157]; the costs for battery and hydrogen electricity storage power capacity and energy storage capacity come from [156]. Although the costs of lines c_ℓ are set to zero, as they are dual to the line volume cap, these costs are added in afterwards in the results. For the annualisation of overnight costs a discount rate of 7% is used. Gas variable costs add up to 21.6 €/MWh$_{th}$[157].

4.5 RESULTS

The original European grid model has already been shown in the previous Chapter 3 in Figure 3.1 and can be compared to two clustered networks in Figure 4.2; the total annual system costs in the three scenarios as a function of the number of clusters is found in Figure 4.3a and these costs are broken down into components in Figure 4.3(b)-(d); the expansion of the transmission network is shown in further detail in Figure 4.4; the system costs as a function of the transmission cap are plotted in Figure 4.5; finally the shadow price of the transmission cap can be studied in Figure 4.6. The results of the scenarios are now discussed in detail.

Spatial scale dependence

Without any expansion of the transmission network ($x = 1.0$ in Figure 4.3(b)), the total annual sytem cost remains approximately steady as the number of clusters increases at 260 billion euros (an average of around € 82/MWh), due to a coincidental balance of the two driving effects: (1) The sites with high capacity factors are more finely resolved with a higher number of clusters, allowing the model to put more capacity at the best sites. With smaller numbers of clusters, the capacity factors are averaged with a weighting over a larger area, bringing the capacity factors down. For example, the best cluster for onshore wind in Germany with 362 clusters has a capacity factor of about 40 %, whereas with one node for the whole of Germany, the weighted average capacity factor is only 26 %. (2) As the number of clusters increases, the bottlenecks inside each country's network become constraining and prevent the wind generated at high capacity factors, localised on the coastlines and offshore, to be transported to load centres.

In panel (b) in Figure 4.3, the two effects, the increasing effective capacity factor of onshore wind combined with intra-country bottlenecks becoming more important, lead to the considerable decrease in the built offshore wind capacity, since better sited onshore wind and solar installations produce more energy closer to the load. The increasing solar generation drives an increase in battery capacities to smooth short-term diurnal variability. Hydrogen storage, which balances longer-term synoptic and seasonal variability, decreases gently with the number of clusters at a higher level than the other two scenarios. Gas generation is fixed because of the CO_2 constraint. The grid costs increase

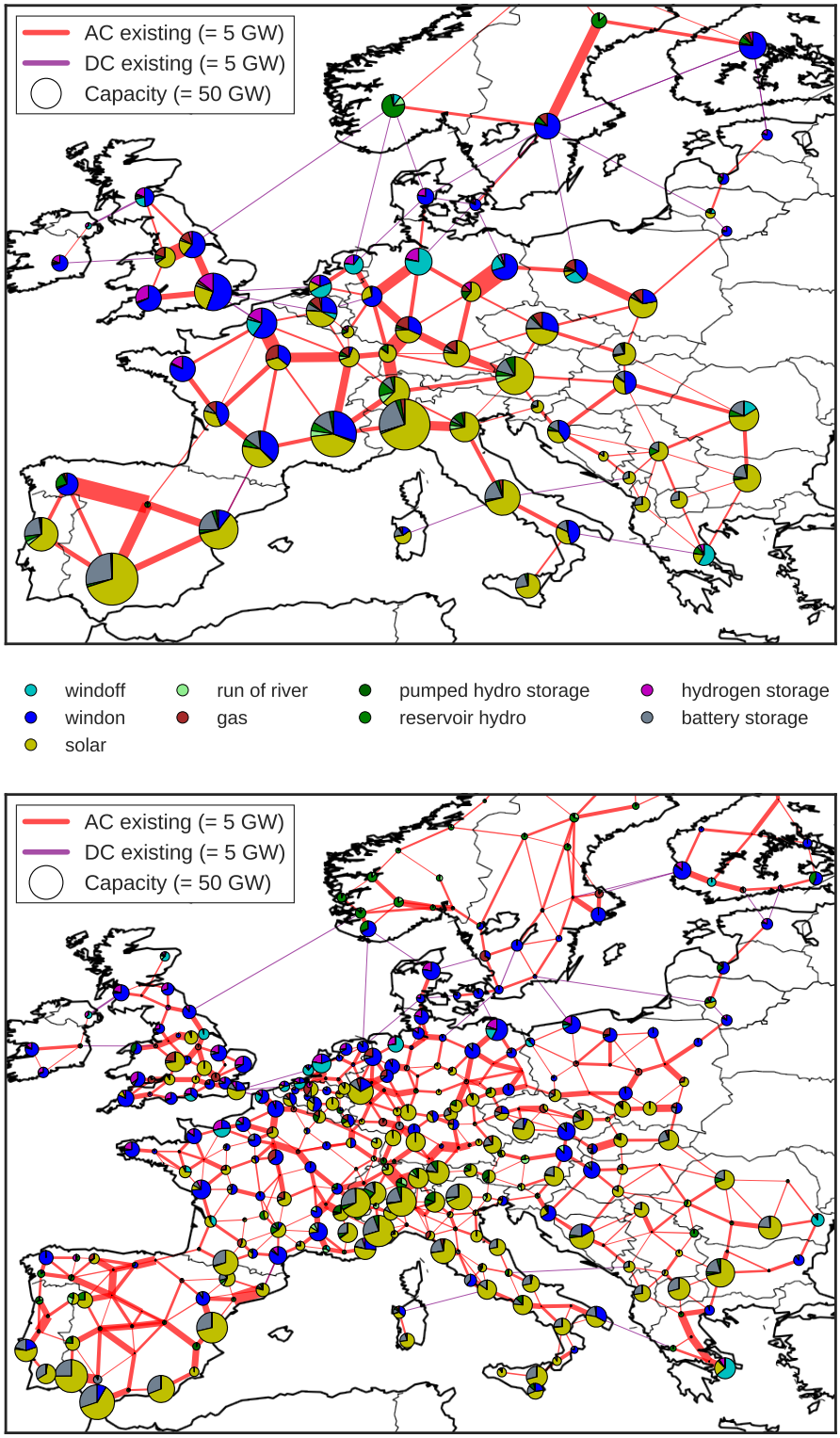


Figure 4.2: The clusterings with 64 buses (above) and 362 buses (below). Results for the distribution of generation capacities at each node are shown as pie charts for the no expansion scenario (existing and planned projects only).

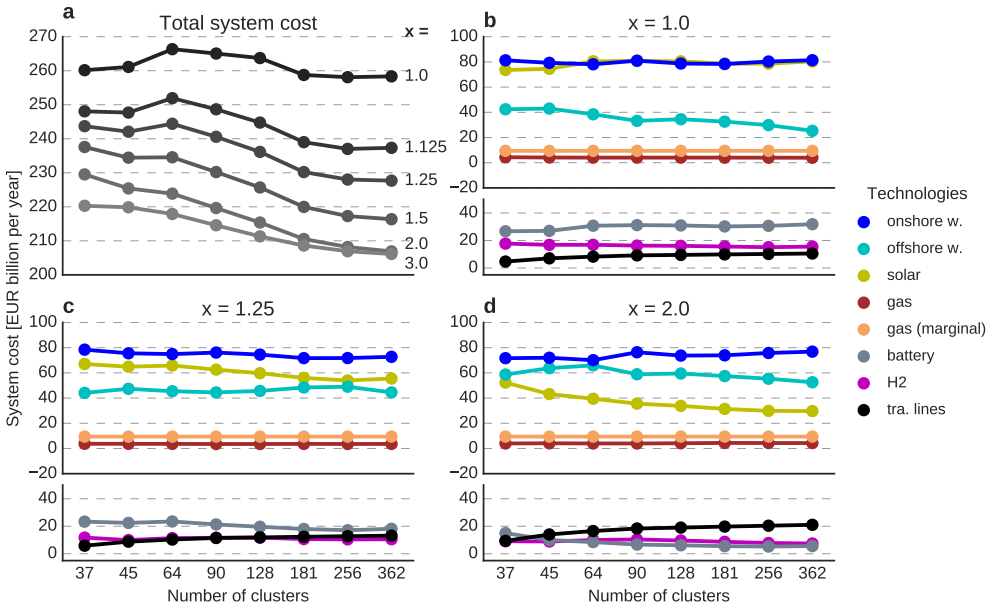


Figure 4.3: System costs. (a) Total annual system costs as a function of the number of clusters for the six scenarios including the costs for overhead transmission lines. (b)-(d) Breakdown of the annual system costs for generation (top) and flexibility options (bottom) as a function of the number of clusters for the no expansion scenario and the expansion scenarios with $x = 1.25$ and $x = 2$.

monotonically as more line capacity and line constraints are seen by the model, but flatten out with the exponentially increasing number of clusters. This is a good indication that the clustering is capturing the major transmission corridors even with smaller numbers of clusters.

Turning back to Figure 4.3(a), the expansion of the network lifts transmission bottlenecks and the first effect wins out, better exploitation of good sites with higher numbers of clusters decrease the system cost. As the grid is gradually expanded the system cost decreases in a very non-linear manner: The expansion by 25% reduces the total system cost already by 30 billion euros of the 50 billion euros in cost reduction available down to 210 billion euros (an average of € 66/MWh). Nevertheless the overall cost reduction possible by expanding the network is a moderate 20%.

In the technology break-down in the lower panels (c) and (d) of Figure 4.3, with the additional line volume the joint solar and battery capacities are replaced by offshore wind turbines. Solar is favoured with limited transmission capacity because it can be built close to demand everywhere and reasonably balanced during its principal short-term diurnal variation using battery storage, whereas the good wind sites are concentrated in Northern Europe and their energy cannot be transported to loads in large quantities without an expansion of the transmission grid. Wind generation additionally benefits from expanding the transmission capacities so that the spatial variation on the continental scale is used for smoothing the temporal fluctuations on the synoptic scale to relieve expensive hydrogen storage. The extra transmission capacity does not offset the low significance of the transmission network cost.

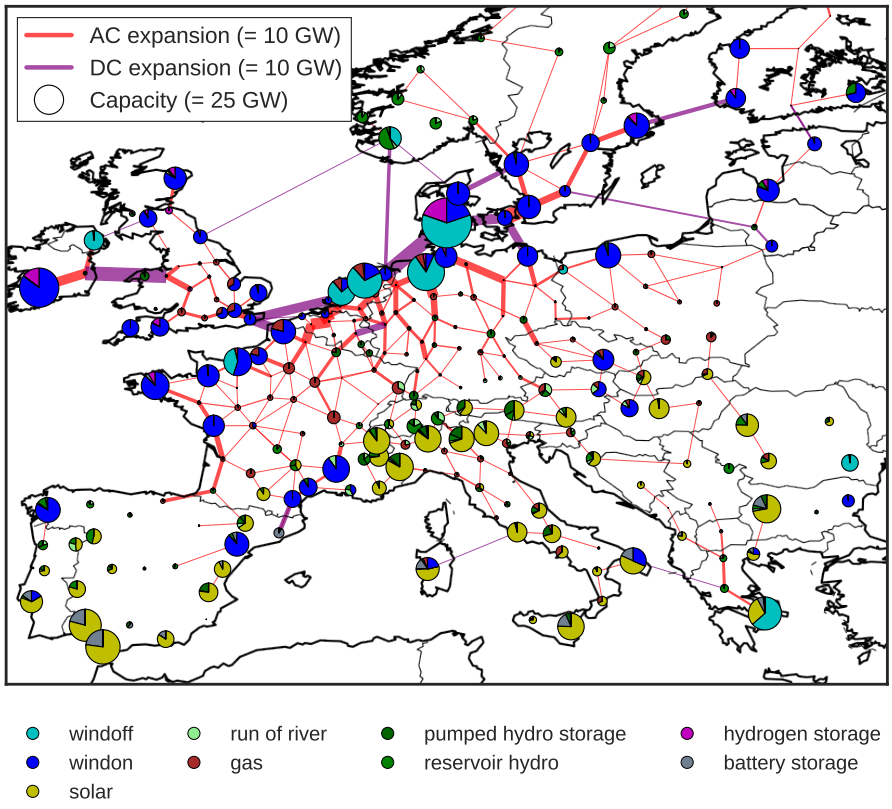


Figure 4.4: Optimal generation capacities and transmission line expansion for 256 buses in the expansion scenario with the transmission cap at $x = 2$.

These trends are all pronounced if the results for Germany are considered in isolation. Transmission bottlenecks within Germany complicate transporting offshore wind energy away from the coast with higher numbers of clusters, forcing a dramatic substitution by solar instead, i. e. the German offshore wind capacity falls from 40 GW to 12 GW from 37 to 362 clusters, while solar peak capacities increase from 46 GW to 100 GW and onshore wind remains largely unaffected despite an intermediate decrease.

The effects disappear for about 200 clusters and above, a level of resolution above which all the results are more-or-less steady.

Transmission volume cap

After ensuring that the solutions have already stabilized at 200 clusters and are thus, likely, a good proxy for the relations on the full network, we want to focus in more detail on the solutions for 256 clusters while varying the allowed overall transmission volume to find the most important lines for expansion and estimate the benefits of a partial expansion deviating from the optimal solution, which might be preferable vis-à-vis problems of public acceptance.

Figure 4.4 shows the optimal generation capacities and transmission expansion for a challenging doubling of the existing transmission capacity ($x = 2$), which was also the subject of the last panel (d) in Figure 4.3. Transmission is

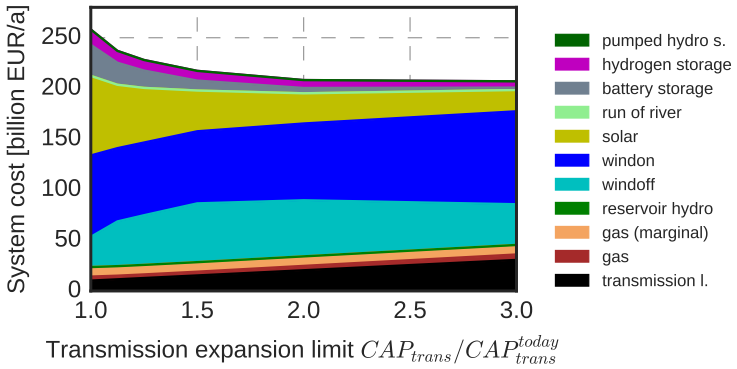


Figure 4.5: Annual total system cost at 256 clusters for different values of the transmission cap $\mathrm{CAP}_{\mathrm{trans}}$.

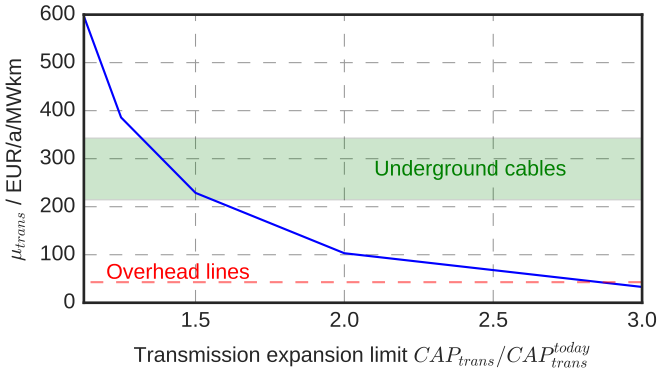


Figure 4.6: Shadow price of the line volume constraint μ_{trans} for different values of the transmission cap $\mathrm{CAP}_{\mathrm{trans}}$ for 256 clusters.

foremost expanded in the proximity of wind capacity installations forming a wide band along the shore of the north and east sea with branches leading inland. This band allows synoptic-scale balancing as weather systems pass from west to east over Europe. It provides the flexibility for the energy of large-scale wind installations to replace a significant amount of solar capacity in Southern Europe and Italy in particular, which also lessens the need for short-term battery storage.

The total system cost in respect to allowed transmission volume in Figure 4.5 decreases non-linearly as has already been observed in the detailed study in the one-node-per-country setting by Schlachtberger et. al [40]. More than half of the overall benefit of transmission of 50 billion EUR per year is already locked in at an expansion by a fourth to $1.25 \cdot \mathrm{CAP}_{\mathrm{trans}}^{\mathrm{today}}$ and after reaching two times today's line volume ($x = 2$) does not increase significantly anymore (also compare the vertical slice at 256 clusters in Figure 4.3). From a system constrained to today's transmission capacities to the optimal solution, the cost composition reduces the component spent on solar and battery in favour of offshore wind and, then, also onshore wind.

Finally, the shadow price of the transmission cap, μ_{trans} introduced in Equation (4.10) is shown in Figure 4.6. It indicates the marginal value of an increase in line volume at each level of network expansion; it can also be interpreted

as the transmission line cost per MWkm necessary for the optimal solution to have the transmission line volume CAP_{trans}. For the assumed costs for overland transmission lines of 400 €/MWkm the model finds the optimal grid volume at slightly below $3 \cdot CAP_{trans}^{today}$. If the expansion instead were to be carried out with underground cabling at a 4 to 8 times higher cost, the economically optimal solution would still be to expand the line volume to between 1.25 and 1.5 times the existing volume.

4.6 CRITICAL APPRAISAL

Although the clustering algorithm presented in Section 4.2 captures the major transmission corridors well, it would be interesting to benchmark the different clustering algorithms mentioned in the introduction based on comparable criteria, such as their ability to capture power flows in the original unclustered network. Results with a higher number of nodes would also be desirable, if this is computationally possible.

Additional aspects, such as distribution grid costs, reserve power, stability and sector-coupling, have not been considered here.

4.7 DISCUSSION AND CONCLUSIONS

The results of this chapter are two fold: Firstly, a network clustering method has been demonstrated that can reduce the number of buses in a given electricity network while maintaining the major transmission corridors for network analysis. With this network reduction method the effects of spatial resolution, i. e. the number of clusters, on the joint optimisation of transmission and generation investment for highly renewable systems in Europe have been investigated. Secondly, the techno-economic European model was optimised at a sufficient level of resolution to determine the hotspots and benefits of transmission expansion.

The systems optimised to reduce CO_2 emissions by 95% with no grid expansion are consistently only around 20% more expensive than systems with grid expansion and half of that cost benefit can already be locked in with an expansion of the line volume by a fourth, which may be a price worth paying given public acceptance problems for new transmission lines.

One must note, though, that in the time horizon until 2050 in which the studied reduction of emissions is to be implemented a significant amount of the current conventional generation park will not yet have passed their lifetime and an important next step is to confirm our greenfield results accounting for this inertia. Further, one should be clear that the feasibility of these solutions is based on a fully integrated European market with nodal prices, high CO_2 price and optimally real-time prices for distributed generation and storage.

5 FLOW TRACING IN RENEWABLE ENERGY NETWORKS[1]

NOMENCLATURE

Indices and Labels

n, m, k	Index of buses.
l, l'	Index of lines.
α, β, τ	Labels of regions and technologies for grouping the power injection and flows.

Constants, Variables and Functions

$P_n(t)$	Net power injection at bus n (MW).
$G_n^\tau(t)$	Power generation by technology τ at bus n (MW).
$L_n(t)$	Load at bus n (MW).
$F_{n\to m}^{out}(t)$	Power outflow from bus n in direction of bus m (MW).
$F_{n\to m}^{in}(t)$	Power inflow to bus m from bus n (MW).
$F_l(t)$	Absolute value of the power flow on line l.
$\chi_{n\to m}(t)$	Loss in the transmission line between bus n and m (MW).
$q_{n,\alpha}^{in}(t)$	In-partition, the share of the injected power at bus n attributed to component α.
$q_{n,\alpha}^{out}(t)$	Out-partition, the share of the consumed power at bus n attributed to component α.
$q_{l,\alpha}(t)$	Line-flow partition, the share of the power flow through line l attributed to component α.
$p_l(F_l)$	Probability for a flow F_l on line l.
$p_l(q_{l,\alpha}\|F_l)$	Conditional probability for a share $q_{l,\alpha}$ of component α in case of a flow F_l.
$h_{l,\alpha}(F_l)$	Average share of owner α on the link l for a flow F_l.
$w_{l,\alpha}(\mathcal{K})$	Weight for the usage of the capacity increment between $\mathcal{K}$ and $d\mathcal{K}$ attributed to owner α on the link l.
$\mathcal{K}^T$	Transmission capacity of the network (MW).
$\mathcal{K}_l^T$	Transmission capacity of line l (MW).
$\bar{\mathcal{K}}^T$	Transmission capacity of the network including length (MW km).
$\bar{L}_l$	Length of transmission line l (km).
$\bar{D}_n$	Average graph distance of bus n (km).
$\mathcal{M}_{\alpha,\tau}^{(1\ldots4)}$	Transmission network usage measures (MW km).

1 This chapter is published as "Flow tracing as a tool set for the analysis of networked large-scale renewable electricity systems" by Hörsch et al. [158]. It was only slightly modified for this work.

5.1 INTRODUCTION

The electricity system is built up of a complex interwoven network of technologies, which provides the backbone for our modern society. In the past, this network was characterized by power flows from large central power plants downstream through the grid to the consumers, with only very limited interactions between different geographical regions. Today, the rising share of decentralized, fluctuating renewable generation and the increasing inter-dependence of international electricity markets has led to a more dynamical system: the power grid has become the underlying infrastructure for a complex pattern of long-range power flows between a heterogeneous distribution of power generation to consumers, integrating not only dispatchable conventional generation, but also electricity from offshore wind farms, wind and solar parks and roof-top solar panels. In this context, a deeper understanding of the emerging power flow patterns is of paramount importance on different levels: For instance, internationally integrated electricity markets need to incorporate possible network congestion into their market design [159], whereas network expansion plans attempt to minimize this congestion in the long run [45, 160]. Also the delevopment of fair and transparent grid usage fee systems, or public discussions concerning the benefit of new infrastructure projects rely strongly on insights concerning the composition and dynamics of the flow pattern in the network [161, 162]. In this article we present a reformulation of a well-known method of flow allocation, denoted as average participation or flow tracing, that is well adapted to the challenges of the system analysis of complex modern electricity systems. Different approaches to the problem of flow allocation in power grids are often derived from circuit theory [163, 164] or are based on approximations of the complex power flow equations for AC electrical networks [165, 166]. For the application of such methods to the problem of flow allocation in large-scale models of electricity systems, one has to factor in the potentially coarse-grained nature of such models. Both the network buses and transmission lines might be aggregated representations of lower level infrastructures, which cannot be included in detail in the model due to computational limitations or lack of data [20, 22, 167]. The method of flow tracing can be applied directly to the overall power flow pattern in the system, and thus does not explicitly have to take into account the underlying modeling details. By tracing what we term in-partitions, we show how the known composition of network-injected power generation can be followed through the grid and thus be transferred to the power flows and composition of net consumption at the sink nodes. In this way the location of generation of power flow can be connected to its location of consumption, thus disentangling the complex spatio-temporal patterns of imports and exports inherent to interconnected electricity systems with a high share of renewable generation. We showcase the potential of this methodological tool set by application to the Scenario 2023B of the IEEE 118-bus model adapted by Barrios et al. at RWTH Aachen with renewable generation capacities and hourly availability for a model year as a benchmark for transmission expansion algorithms [168].

After a short review of flow tracing, Sec. 5.2 introduces the reformulated flow tracing technique and a measure of network usage. The subsequent Sec. 5.3 showcases two exemplary applications: Firstly the tracing of power flow of different generation types between several regions across a network model based on the IEEE 118 bus case, and secondly a comparison of a statistical transmission capacity usage measure with several alternative allocation mechanisms. Sec. 5.4 throws a quick glance at the usage of storage in the decarbonized scenarios of the European electricity system from the previous chapter. Section 5.5 concludes the chapter.

5.2 METHODOLOGY

Flow tracing was introduced as a loss-allocation scheme by Bialek et al. based on solving linear equations [169] and in parallel by Kirschen et al. as an analytical tool using a graph-based, iterative approach [170].

It was soon after proposed as a transmission-usage allocation scheme [171–174]. Subsequently, the method was discussed to cover concrete supplementary charge schemes for cross-border trades [175, 176], in view of the discussion about the mechanism of inter-transmission system operator compensation in Europe [161, 177, 178].

Of the other network-cost allocation methods – reviewed in [179] or [180], for instance – we only want to highlight marginal participation [181] and the related decomposition method [182], which attribute transmission capacity according to linear sensitivities of network flows to differential bus injections as captured by the power transfer distribution factors (PTDF) [59]. Due to its influence on the PTDF, for this method the choice of the slack bus has to be taken into account explicitly [183], whereas for the flow tracing technique this choice only affects the total power flow but not the allocation mechanism.

Power flow

The active power flow in an electricity system satisfies Kirchhoff's current law. If the net power injection at bus n from generators and loads is given by P_n, and $F^{in/out}_{n\to m}$ are the power in- and outflows from bus n to m, then the power flow through node n is conserved as

$$P^{in}_n + \sum_m F^{in}_{m\to n} = P^{out}_n + \sum_m F^{out}_{n\to m} \,. \tag{5.1}$$

Here we use the positive and negative injections P^{in}_n and P^{out}_n at node n and invoke the convention that all $F^{out}_{m\to n}$ and $F^{in}_{m\to n}$ are positive or zero.

Table 5.1 introduces a particular snapshot in a simple network with four buses with generation G_n, load L_n and im-/exports I_n/X_n with other buses not represented explicitly. In this example, we take the positive injection as the net

n	G_n	of which in %			L_n	$G_n - L_n$	I_n	X_n
		wind	solar	other				
1	76.0	16	19	65	65.5	10.5	0.9	5.6
2	20.5	8	0	92	21.1	-0.6	0.9	0.6
3	8.5	2	13	85	8.0	0.5	0.0	1.8
4	7.3	12	6	82	7.5	-0.3	0.0	2.5

Table 5.1: Power generation and consumption of a simple four bus network with im-/exports with external buses in GW.

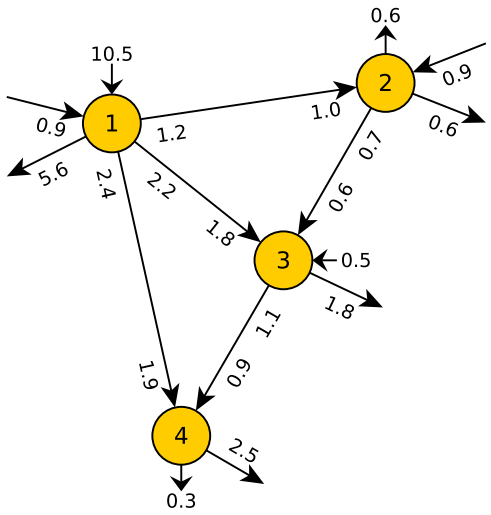


Figure 5.1: Power flows and injections in the simple four bus network introduced in Table 5.1 in units of GW.

surplus between generation G_n and demand L_n plus the imports I_n, while the negative injection follows from the deficit and exports X_n, as

$$P_n^{in} = \max\{(G_n - L_n), 0\} + I_n \,, \qquad P_n^{out} = \max\{-(G_n - L_n), 0\} + X_n \,. \tag{5.2}$$

The flows and line-losses are illustrated in Fig. 5.1. The convention means that the line from bus 1 to bus 3 is described by $F^{out}_{1\to3} = 2.2\,\text{GW}$, $F^{in}_{1\to3} = 1.8\,\text{GW}$, $F^{in}_{3\to1} = 0$ and $F^{out}_{3\to1} = 0$.

Here and in general the outflow from bus n to m, $F^{out}_{n\to m}$, is larger than the inflow to m, $F^{in}_{n\to m}$ due to losses in the transmission line $n \to m$. We denote them by $\chi_{n\to m} = F^{out}_{n\to m} - F^{in}_{n\to m}$.

Flow tracing

The flow tracing method by Bialek and Kirschen [169, 170] follows the power flow from individual buses through the network and decomposes the flow on the power lines into contributions associated to each bus. Since for large-scale electricity systems, the injection P_n^{in}, in general, already contains several constituents, we introduce an in-partition $q^{in}_{n,\alpha}$ associating the power injection

at each bus n to a set of components α. For the power flows of the four bus example, we will use the components $\{1, 2, 3, 4, I\}$ with the in-partition

$$q^{in}_{n,\alpha} = \begin{cases} \dfrac{\max\{G_n - L_n, 0\}}{P^{in}_n} & \text{for } \alpha = n \wedge P^{in}_n > 0\,, \\ \dfrac{I_n}{P^{in}_n} & \text{for } \alpha = I \wedge P^{in}_n > 0\,, \\ 0 & \text{else.} \end{cases} = \begin{pmatrix} \frac{10.5}{11.4} & 0 & 0 & 0 & \frac{0.9}{11.4} \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{5.3}$$

to differentiate the imports I_n entering at each bus from the power generated there. Note that the component I is associated with injected power throughout the network. Similarly, another in-partition for components {wind, solar, other, imports} is able to encode the relative shares of wind, solar and other energy generation sources from Table 5.1.

Flow tracing follows the diffusion of the different components α by assuming conservation of the partial power flows at bus n in analogy to (5.1)

$$q^{in}_{n,\alpha} P^{in}_n + \sum_m q^{(m\to n)}_{m,\alpha} F^{in}_{m\to n} = q^{(out)}_{n,\alpha} P^{out}_n + \sum_m q^{(n\to m)}_{n,\alpha} F^{out}_{n\to m}\,. \tag{5.4}$$

In general there is a degree of freedom in relating $q^{(out)}_{n,\alpha}$ and $q^{(n\to m)}_{n,\alpha}$ under the boundary condition of assuring conservation of partial flows. It is nevertheless intuitive to assume that the power contributions mix perfectly in each bus and the partitions of the flows leaving a bus are all identical $q_{n,\alpha} = q^{(out)}_{n,\alpha} = q^{(n\to m)}_{n,\alpha}$. This assumption equally underlies the tracing by Bialek and Kirschen and is known as proportional sharing. Bialek et. al. were able to show the proportional sharing principle to coincide with the Shapley value of a stylized game of loss attribution played by two generators feeding into the same line [184], but it can rightfully be contested for practical purposes. For the purpose of flow allocation in large-scale electricity models, we suggest this realization in particular due to its intuitiveness and lack of additional parameters. With the proportional sharing assumption Eq. (5.4) reduces to a system of $N \times A$ equations for $N \times A$ unknowns $q_{n,\alpha}$

$$q^{in}_{n,\alpha} P^{in}_n + \sum_m q_{m,\alpha} F^{in}_{m\to n} = q_{n,\alpha} \left(P^{out}_n + \sum_m F^{out}_{n\to m} \right)\,, \tag{5.5}$$

with A denoting the number of components α, as we had already reported in [185]. If we eliminate inert buses without any flows from the network (without any loss of generality) and abbreviate the power leaving a bus as nodal flow $F_n := P^{out}_n + \sum_k F^{out}_{n\to k}$, we can rearrange Eq. (5.5) to

$$q^{in}_{n,\alpha} P^{in}_n = F_n \sum_m \left[\delta_{n,m} - \frac{F^{in}_{m\to n}}{F_n} \right] q_{m,\alpha} \tag{5.6}$$

with the Kronecker delta $\delta_{n,m}$. Finally, with the definition of the matrix

$$D_{n,m} = \frac{F^{in}_{m\to n}}{F_n} \tag{5.7}$$

capturing the share that the power from bus m contributes to the nodal flow through bus n, Eq. (5.6) can be rendered in matrix notation as

$$\mathrm{diag}(P^{in})q^{in} = \mathrm{diag}(F)(1-D)q\ . \tag{5.8}$$

For a lossless power flow, $F^{out}_{n\to m} = F^{in}_{n\to m}$, $1-D$ is the transpose of the downstream distribution matrix A_d in Bialek's formulation and together with $q^{in} = q = 1$ the proposed method reduces to Bialek's flow tracing. The steps from Eq. (5.5) to Eq. (5.8) illustrate the equivalence of the two formulations of flow tracing discussed in the literature as a linear algebra problem [169] and a graph-based algorithm [170].

Eq. (5.8) is solved formally as

$$q = (1-D)^{-1}\mathrm{diag}(P^{in}/F)q^{in}\ , \tag{5.9}$$

where the inverse of $1-D$ can be shown to exist as Neumann series $(1-D)^{-1} = \sum_{k=0}^{\infty} D^k$, since the absolute value of each eigenvalue of D is smaller than 1, if there is at least one bus with a positive power injection in each connected component, similarly to [186]. While, therefore, the method is formally applicable also in the presence of loop flows, the interpretation of the resulting flow attribution still remains to be investigated.

If all power injections are attributed to a component, i. e. $\sum_\alpha q^{in}_{n,\alpha} = 1$, the partial flows must add up to the total power flow, as we show in the next few sentences: If we sum over all components α in Eq. (5.5), we obtain

$$P^{in}_n + \sum_m Q_m F_{m\to n} = Q_n\left(P^{in}_n + \sum_m F_{m\to n}\right)\ , \tag{5.10}$$

where we have used the abbreviation $Q_m = \sum_\alpha q_{m,\alpha}$. This equation is solved by $Q_n = 1$ $(\forall n)$ due to flow conservation from Eq. (5.1). The solution is also unique, since by following the steps between Eqs. (5.5)-(5.9), one arrives at

$$Q = (1-D)^{-1}\mathrm{diag}(F^{-1})P^{in}\ . \tag{5.11}$$

Thus, $\sum_\alpha q_{n,\alpha} = 1$ at all buses n.

To apply flow tracing to the 4-bus example shown in Fig. 5.1 one calculates the in-partition q^{in} (Eq. (5.3)), nodal flows F, power injections P^{in} and matrix $1 - D$

$$F = \begin{pmatrix} 11.4 & 1.9 & 2.9 & 2.8 \end{pmatrix} , \tag{5.12}$$

$$P^{in} = \begin{pmatrix} 11.4 & 0.9 & 0.5 & 0 \end{pmatrix} , \tag{5.13}$$

$$1 - D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1/1.9 & 1 & 0 & 0 \\ -1.8/2.9 & -0.6/2.9 & 1 & 0 \\ -1.9/2.8 & 0 & -0.9/2.8 & 1 \end{pmatrix} . \tag{5.14}$$

By evaluation of Eq. (5.9), one then finds

$$q \approx \begin{pmatrix} 0.921 & 0.0 & 0.000 & 0.0 & 0.079 \\ 0.485 & 0.0 & 0.000 & 0.0 & 0.515 \\ 0.672 & 0.0 & 0.172 & 0.0 & 0.156 \\ 0.841 & 0.0 & 0.055 & 0.0 & 0.104 \end{pmatrix} , \tag{5.15}$$

where each column corresponds to the share of each nodal flow associated with a component, and consequently also to the share on the out-going lines. Since buses 2 and 4 do not feed any power into the network, they do not contribute to any flows. The shares of bus 1, which feeds all buses directly, are strongest at bus 1 and 4, while at bus 2 a strong in-flow by imports dilutes the share of bus 1. Power entering the network as imports is present at every bus, making up 15.6% of the nodal flow through bus 3 and 10.4% of the nodal flow at bus 4. The generation of bus 3 only appears at buses 3 and 4.

Since the power loss happens on the links it is natural to attribute a loss $q_{n,\alpha}\chi_{n\to m}$ to entity α, f.ex. imported power leads to a loss of $q_{2,I}(F^{out}_{2\to 3} - F^{in}_{2\to 3}) = 0.515 \cdot 0.1\,\mathrm{GW}$ in line $2 \to 3$ according to flow tracing. Substituting $F^{out}_{m\to n} = F^{in}_{m\to n} + \chi_{m\to n}$ in Eq. (5.5)

$$q^{in}_{n,\alpha} P^{+}_{n} + \sum_m q_{m,\alpha} F^{in}_{m\to n} = q_{n,\alpha}\left(P^{-}_{n} + \sum_m \chi_{n\to m}\right) + q_{n,\alpha}\sum_m F^{in}_{n\to m} , \tag{5.16}$$

reveals by comparing the structure again to Eq. (5.5) that this loss allocation scheme is equivalent to treating a loss on line $n \to m$ as an additional load at the out-flowing bus n combined with flow tracing on the inflows indiscriminately, the procedure Bialek introduced as net flows [169].

In summary, for a given flow pattern $F_{n\to m}$ and a fixed attribution of the generated power $q^{in}_{n,\alpha}P^{in}_{n}$ to a set of components α the flow tracing algorithm yields the attribution of all flows along the links $q_{n,\alpha}F_{n\to m}$ and the attribution of the power flowing into the consuming nodes $q_{n,\alpha}P^{out}_{n}$.

Note that we are able to invert the injection pattern and flow graph consistently by switching the in- and outputs $P^{in/out}_{n} \to P^{out/in}_{n}$ and the flows $F^{in/out}_{n\to m} \to F^{out/in}_{m\to n}$. This procedure allows a given out-partition to be considered

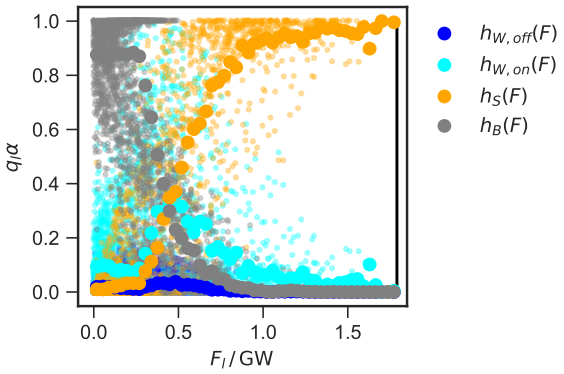


Figure 5.2: Usage shares of generation types in a chosen transmission line in a realistic examplary model with a high share of renewable generation (discussed in detail in Section 5.3).

as the input for the flow tracing algorithm, which then assigns shares of the power flow and injected power at the source node according to this partition.

A measure of transmission line usage

The flow tracing method as displayed in the last section refers to the application to a single flow pattern. However, for the analysis of complex modern electricity systems, one rather has to consider whole time series of fluctuating injection and flow patterns taking place on the underlying power grid. The application of the flow tracing method then yields a time series $(F_l(t), \{q_{l,\alpha}(t)\})$, containing the power flows $F_l(t)$ and the respective shares $q_{l,\alpha}(t)$ assigned to the components α for each link l. In order to derive the respective grid usage over the whole time series, this information has to be integrated into a suitable transmission capacity usage measure. Fig. 5.2 illustrates the need for such a non-trivial measure in a realistic example: Consider, for instance, in cyan the shares associated with onshore wind on a specific line; while the shares shown as small dots vary strongly over time, their conditional averages,

$$h_{l,\alpha}(F) = \langle\, q_{l,\alpha}(t) \,\rangle_{\{t|F_l(t)=F\}} \ , \tag{5.17}$$

depend smoothly on the absolute line flow at which the average is taken. In the presented line the power transmitted at a low line-loading is traced back to conventional generators, while in hours with a high line-loading on- and offshore wind contribute nearly all the power. Such a striking difference should be accounted for as relevant information by an adequate usage measure. In the following we briefly review such a capacity usage measure introduced in [187].

Their central idea is that the transmission line capacity of a small increment between $\mathcal{K}$ and $\mathcal{K} + d\mathcal{K}$ is only used by flows $F(t) > \mathcal{K}$ and, thus, the usage share of a component for this capacity increment is determined only from those as

$$w_{l,\alpha}(\mathcal{K}) = \langle\, q_{l,\alpha}(t) \,\rangle_{\{t|F_l(t)>\mathcal{K}\}} \ . \tag{5.18}$$

Mind ">" in the subscript. The capacity of the whole line $\mathcal{K}_{l,\alpha}^{T}$ can then be split for the individual components α by summing all increments to

$$\mathcal{K}_{l,\alpha}^{T} = \frac{\mathcal{K}_{l}^{T}}{\max_t F_l(t)} \int_0^{\max_t F_l(t)} \langle\, q_{l,\alpha}(t) \,\rangle_{\{t|F_l(t)>\mathcal{K}\}} \, \mathrm{d}\mathcal{K} \,. \tag{5.19}$$

The proportional factor in front of the integral accounts for the fact that $w_{l,\alpha}(\mathcal{K})$ vanishes at the maximum flow by sharing the remaining security margin $\mathcal{K}_{l,\alpha}^{T} - \max_t F_l(t)$ proportionally, since it is important to all users of the capacity in an analogous way to the actually used capacity. Nevertheless, depending on the details of the system under investigation other schemes are possible.

5.3 FLOW TRACING APPLIED TO A 118-BUS ELECTRICITY NETWORK MODEL

For demonstrating the application of the reformulated flow tracing methodology, we briefly introduce an electricity system model that has been developed as a benchmark for transmission expansion methods. The IEEE 118 bus network model has been geographically embedded and augmented by attaching specific loads and conventional as well as renewable generators to the system by Barrios et al. at RWTH Aachen [168]. The load curves and the renewable generation availability span all hours in a model year. The geographic regions have weather characteristics in line with the artificial TRY-Regions of the so-called TRY reference data set of the German weather service (DWD), which feature a higher solar capacity factor in the North-West and a higher wind capacity factor in the East, where, in addition, an offshore-wind region is located. The network topology and generation capacities are shown in Fig. 5.3, while the average generation and consumption of each region are included in Fig. 5.4.

We use our electricity system modeling framework PyPSA [141] to determine the linear optimal power flow (LOPF), i. e. the dispatch of the generators is solved by a convex linear optimization minimizing the total cost based on the marginal costs of the conventional generators and the spatially and temporally fluctuating availability of renewable generation subject to meeting the load curve and the transmission constraints in all hours. Once the generator dispatch has been determined, the non-linear power flow is found by a standard Newton-Raphson iteration. Several key figures of the optimization are summarized in Table 5.2.

Analyzing mean flow patterns

To adopt the flow tracing method one initially distinguishes the injections per region by choosing an in-partition

$$q_{\alpha,n}^{in} = \delta_{\alpha,n} := \begin{cases} 1 & \text{for node } n \text{ in region } \alpha, \\ 0 & \text{else} \end{cases} . \tag{5.20}$$

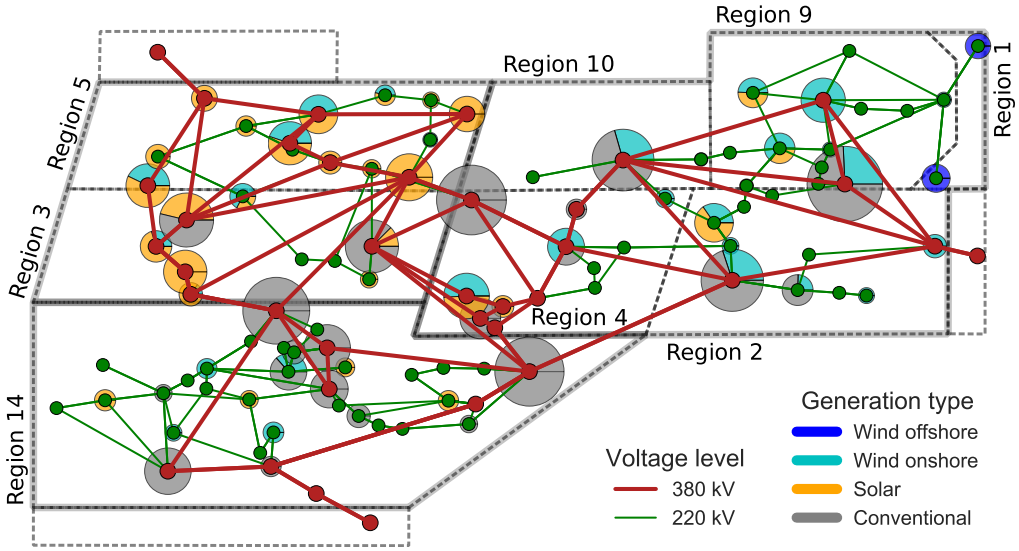


Figure 5.3: Scenario 2023B of the 118-bus transmission expansion benchmark case with renewable generation capacities from [168]. In the background the relative composition of the generation capacities of each region are indicated.

	Generation				Load	Loss
	wind off.	wind on.	solar	conv.		
capacities	1.8	21.2	22.3	27.9	-	-
mean	0.8	4.5	2.6	10.9	18.6	0.2
std	0.7	4.3	4.1	5.9	3.2	0.16
min	0	0	0	0.05	9.4	0.03
max	1.8	21.0	19.7	25.9	26.3	0.9

Table 5.2: Characteristic figures of the LOPF solution in units of GW

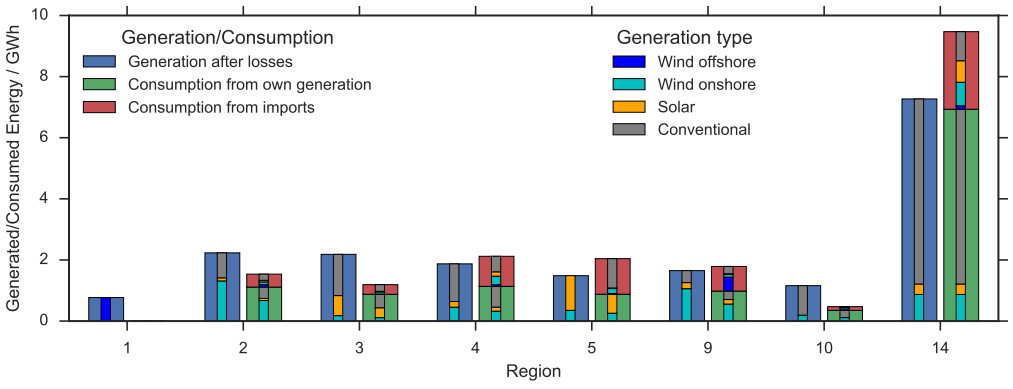


Figure 5.4: Comparison of the generated energy (blue) in a region with its consumption from own generation (green) and from imports (red). The inset decomposes the energy by generation type.

Using the flow tracing solution Eq. (5.9) we find the partition $q^{out}_{\alpha,n}(t)$ as the share of the energy consumed in bus n that is generated in region α, i.e. $q_{\alpha,n}(t)P^{out}_n(t)$. For the total amount of energy from a region α, we only need to correct for the energy generated and consumed directly at bus n given by $L_n - P^{out}_n$, if bus n also belongs to region a. The average inter-region flow from region α to region β then adds up to

$$E_{\alpha,\beta} = \sum_{n \text{ in region } \beta} \left\langle q^{out}_{\alpha,n} \cdot P^{out}_n + \delta_{\alpha,n} \cdot (L_n - P^{out}_n) \right\rangle_t \ . \tag{5.21}$$

These flows are illustrated in Fig. 5.4 and 5.5.

If you ignore the inset decomposition about renewables and focus on the outer blocks in blue, red and green for now, Fig. 5.4 compares the net generated energy $\sum_b E_{\alpha,\beta}$ to the consumed energy $\sum_\alpha E_{\alpha,\beta}$ in each region. The consumption has been decomposed into two parts which are covered by local production and by imports. In contrast to summing up the generation independent of flow tracing as $\sum_{n \text{ in region } \alpha} G_n$, the small losses of about 5% for the energy generated in the offshore Region 1 and about 2% for the other regions have automatically been netted away by considering directly the consumed energy.

The full benefit of using flow tracing for the average flow statistics becomes only clear once we distinguish also between different generation types. We use the components

$$\{(\alpha,\tau) | \alpha \in \text{regions}, \tau \in \{\text{wind}_{\text{on}}, \text{wind}_{\text{off}}, \text{solar}, \text{other}\}\} \tag{5.22}$$

and extend the in-partition from Eq. (5.20) to

$$q^{in}_{(\alpha,\tau),n}(t) = \delta_{\alpha,n} G^\tau_n(t) / \sum_{\tau'} G^{\tau'}_n \ , \tag{5.23}$$

while Eq. (5.21) is adapted by substituting $\alpha \to (\alpha,\tau)$. The resulting measure $E_{(\alpha,\tau),\beta}$ yields the decomposition in generation types shown as inner bars in Fig. 5.4.

Regions with only one or two types of generation capacities in the studied network model usually import a generation mix that is far more balanced. This can be observed, for instance, in Region 5, which only generates solar and wind energy, but consumes nevertheless more than a third of conventionally generated energy and Region 14 with mostly conventional generation capacities importing also a significant amount of energy from renewable generation. It is also found that the energy generated by offshore wind in Region 1 is mainly consumed (to 56%) in the adjacent Region 9 and only a tiny amount of 5% reaches the remote Region 14.

To study the spatial pattern on imports and exports in more detail, we decompose the imports of each region further into the partial flows originating from each of the other regions in Fig. 5.5. The matrix shows relative imports $E_{\alpha,\beta} / \sum_{\alpha \neq \beta} E_{\alpha,\beta}$. For example, the value 0.6 between Region 1 and 9 means that 60% of the consumption from imports is covered by energy generated in Region 1. The order of the regions is chosen from the North-East to the South-West highlighting two local clusters between regions 1, 9, 2 and 10 and between

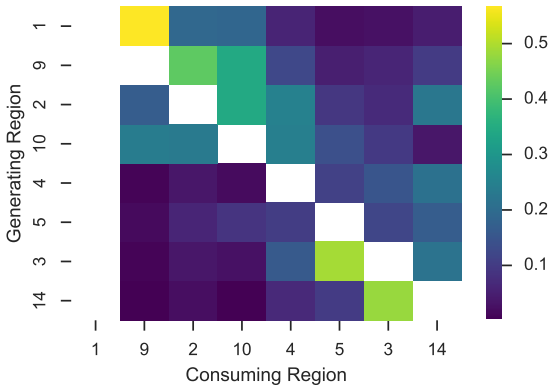


Figure 5.5: Relative imports between regions ($E_{\alpha,\beta} / \sum_{\alpha \neq \beta} E_{\alpha,\beta}$).

regions 5, 3 and 14. Region 4 has a status of its own, since it receives most of its imports from the north-eastern cluster, while it exports to the south-western cluster. The high-load Region 14 satisfies also about a fourth of its imports from regions 2 and 9 outside of its own cluster. This indicates a net flow from the North-East to the South-West not unlike the German situation of wind energy surpluses in the North-East flowing to the load-intensive South and West.

Note that while, for simplicity, we studied the average energy flows, all the partial flows are available as time-series retaining correlations to important network characteristics. The following section uses the correlations to line-loading for attributing transmission line capacity.

Attributing transmission capacity

In this section we will demonstrate the application of the line usage measure reviewed in Sec. 5.2 to determine the transmission capacity that is attributed to the four generation types on each link. Extending the investigations in [187], we will then compare the results to several other allocation measures. In contrast to specific cost allocation models as f.ex. Soares et al. [188] propose for pricing distribution grid capacities, our focus lies on improving the underlying usage measure, in particular by incorporating correlations to the absolute value of the power flows as detailed in Sec. 5.2.

Usage shares of the transmission lines for generation types τ are captured by the line-flow partition $\{q_{l,\tau}(t)\}$ which results from flow tracing on an in-partition

$$q^{in}_{n,\tau}(t) = G^{\tau}_{n}(t) / \sum_{\tau''} \left(G^{\tau'}_{n}(t) \right) \ , \tag{5.24}$$

based on the hourly energy generation mix $G^{\tau}_{n}(t)$.

These shares vary significantly with the flow on a power line, as we have already seen in Fig. 5.2. The line that was shown there carried high shares of conventionally generated power only in hours with low amounts of flow. In hours with a high line-loading this line is mainly occupied by energy traced back to wind turbines. On another line further in the West, the usage shares presented

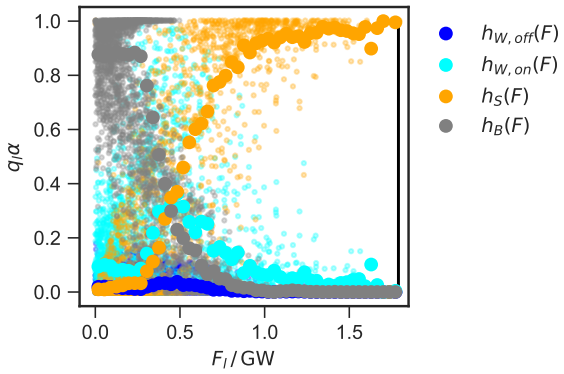


Figure 5.6: Usage shares of the power line connecting Region 3 and Region 14, the western line highlighted in Fig. 5.7.

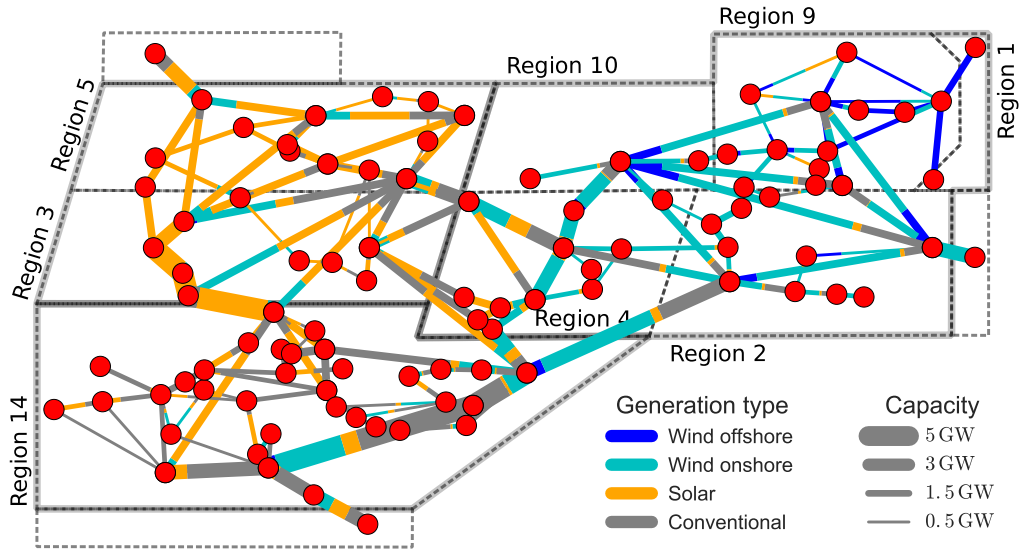


Figure 5.7: Line capacities $\bar{\mathcal{K}}^T_{l,\alpha}$ attributed to the four generation types α for each link l throughout the network of the benchmark case. The usage shares of individual snapshots for the links highlighted by a black frame have been resolved in Figures 5.2 and 5.6.

in Fig. 5.6 exhibit similar characteristics of high shares of conventional power at low line-loadings and only that solar power dominates at higher flows. Both lines are highlighted in Fig. 5.7, the former connects Regions 2 and 14 in the East, the latter is between Regions 3 and 14 in the West.

If one neglected this correlation for assessing the usage of the eastern power line, one would find that the total amount of conventional power is about a third higher than the amount of wind power flowing through it and, thus, would conclude that the costs of the power line should be split in the same proportion. Instead, the reviewed usage measure from Sec. 5.2 gives a higher weight to the shares with high line loads. To compare attributed transmission

capacities in line with transmission cost drivers length and capacity, we multiply it by length $\bar{L}_l$ as

$$\bar{\mathcal{K}}^T_{l,\alpha} = \frac{\bar{L}_l \mathcal{K}^T_l}{\max_t F_l(t)} \int_0^{\max_t F_l(t)} \langle\, q_{l,\alpha}(t) \,\rangle_{\{t|F_l(t)>\mathcal{K}\}} \, \mathrm{d}\mathcal{K} \,. \tag{5.25}$$

Similarly we understand the total transmission capacity of the network to be given by $\bar{\mathcal{K}}^T = \sum_l \mathcal{K}^T_l \bar{L}_l$ in units of MW km. The evaluation of the measure for the eastern line in Fig. 5.6 attributes 54% to onshore wind and only 34% to conventionally generated power.

The attributed capacities of all the transmission lines in the network are shown schematically in Fig. 5.7. Power from onshore wind turbines takes up most of the capacity in the East, while power generated by solar panels is attributed the transmission capacity in the West. This separation mirrors the distribution of the generation capacities (cf. Fig. 5.3). Renewable generation is attributed a share of the transmission capacity that is disproportionately high compared to the average energy generation mix, given in Table 5.2. In Region 14, for instance, where only few renewable generation capacities are located, significant amounts of transmission capacity are attributed to solar and wind generation.

We finally compare the flow tracing based usage measure in Eq. (5.25) with several alternative allocation mechanisms for transmission capacity:

Average power injection splits the transmission capacity of the network $\bar{\mathcal{K}}^T$ in proportion to the amount of injected power of each generator (n, τ), i. e.

$$\mathcal{M}^{(1)}_{\alpha,\tau} = \left(\frac{\sum_{n \text{ in region } \alpha} \langle\, P^{in}_{n,\tau} \,\rangle_t}{\sum_m \langle\, P^{in}_{m,\tau} \,\rangle_t} \right) \bar{\mathcal{K}}^T \,, \tag{5.26}$$

where $P^{in}_{n,\tau}(t)$ is the power injected by a generation type τ at bus n. This scheme corresponds to the widely used postage stamp pricing mechanism.

Average power injection with topological correction adjusts $\mathcal{M}^{(1)}_{\alpha,\tau}$ with an additional factor penalizing remote locations, where the generators on average have to send their energy farther through the network than from a central bus.

$$\mathcal{M}^{(2)}_{\alpha,\tau} = \left(\frac{\sum_{n \text{ in region } \alpha} \langle\, P^{in}_{n,\tau} \,\rangle_t \bar{D}_n}{\sum_m \langle\, P^{in}_{m,\tau} \,\rangle_t \bar{D}_m} \right) \bar{\mathcal{K}}^T \,. \tag{5.27}$$

$\bar{D}_n$ is the average graph distance of the bus n, which is the mean distance to the other buses [72].

Flow tracing mean usage weights the attributions with a distribution determined from the average line loading of each generation type and region.

$$\mathcal{M}^{(3)}_{\alpha,\tau} = \left(\frac{\sum_l \langle\, q_{l,\alpha,\tau} F_l \,\rangle_t \bar{L}_l}{\sum_{l'} \langle\, F_{l'} \,\rangle_t \bar{L}_{l'}} \right) \bar{\mathcal{K}}^T \tag{5.28}$$

It is similar to previously proposed pricing schemes based on flow tracing [171, 172].

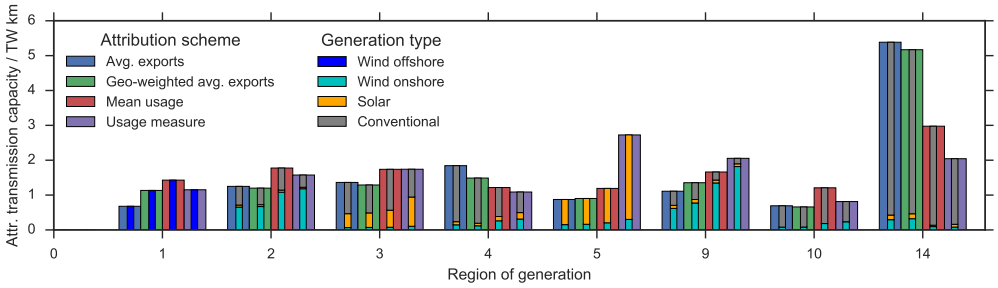


Figure 5.8: Line capacities of the overall electricity system attributed to the generators of each Region using four different assignment schemes.

Finally, *Flow tracing usage measure* distributes the capacity of each line by the usage measure from Eq. (5.25).

$$\mathcal{M}^{(4)}_{\alpha,\tau} = \sum_l \bar{K}^T_{l,\alpha,\tau} \tag{5.29}$$

The four measures $\mathcal{M}^{(1)}_{\alpha,\tau}$ to $\mathcal{M}^{(4)}_{\alpha,\tau}$ for the 118-bus network model are illustrated in Figure 5.8.

The geometry factor which distinguishes $\mathcal{M}^{(2)}_{\alpha,\tau}$ from $\mathcal{M}^{(1)}_{\alpha,\tau}$ only has a marginal effect on the allocation, nevertheless it is still worth noting that the modification is mostly in direction of the results of more elaborate measures.

For most regions and generation types the simple measures $\mathcal{M}^{(1/2)}_{\alpha,\tau}$ agree quite well with the flow tracing based measures $\mathcal{M}^{(3/4)}_{\alpha,\tau}$. Incorporating the actual shares of the line loading by flow tracing turns out to have the largest effect for Region 14 which has a large consumption and exclusively conventional generation. Most of the power that is generated in Region 14 is consumed within few line kilometers so that actual network transmission is kept to a minimum, although the total power injected into the network is very high. This indicates that an average distance to the load centers instead of to all the buses in equal weights might be a better measure.

The capacity attributed to Region 14 is further reduced by taking the line-loading correlations into account, since its conventional generation is mainly dispatched in times with low renewable generation and, thus, also small overall flows. But the most striking adjustment from $\mathcal{M}^{(3)}_{\alpha,\tau}$ to $\mathcal{M}^{(4)}_{\alpha,\tau}$ is that the capacity attributed to the solar generation in Region 5 doubles, which is due to the strong correlation between line-loading and solar flows already visible in the single line usage share of Fig. 5.6. The same effect can also be seen in the usage measure component for solar power generated in Region 3, only that the overall capacity for Region 4 balances out thanks to its conventionally dominated generation mix.

Overall, we find that wind generators are strongly affected by switching from a postage stamp pricing mechanism to a flow tracing based one, since the volume of wind energy in the network is often disproportionately high as a study by Brown et al., based on marginal participation, has already pointed out [166]. Additionally taking the correlation between line-loading and usage

shares into account strongly impacts the capacities attributed to solar generation. The choice of a suitable capacity allocation measure thus depends on the range of system properties which should be represented. Whereas a simple postage stamp method might cover average imports and exports of the system participants, only more elaborate techniques based on flow allocation are able to incorporate the correlations and patterns emerging from the fluctuating imports and exports in large-scale electricity systems with a high share of renewable generation.

5.4 STORAGE USAGE IN A LOW-CARBON EUROPEAN ELECTRICITY SCENARIO[2]

In the last chapter, we noted already a strong spatial relationship of hydrogen storage with wind turbines and batteries with PV panels emerging from optimization of low-carbon solutions to the European electricity system, particularly visible in Fig. 4.4.

The technology mix of the power flowing through any bus in the network can be determined by flow tracing using the in-partition defined in Eq. (5.24). The average mix of the power $[g_{n,s}(t)]^-$ charging a particular storage (n,s) at time t follows then from correcting the out-partition by the bus-internal generation similar to Eq. (5.21) as

$$\overline{[g_{n,s}]^-}_\tau = \left\langle [g_{n,s}]^- \cdot \frac{[g_{n,\tau}]^+ + q^{out}_{\tau,n} P^{out}_n}{\sum_{\tau'} [g_{n,\tau'}]^+ + P^{out}_n} \right\rangle , \tag{5.30}$$

where we used the small letter g for the power from generators and storage and the notation $[\cdot]^{+,-}$ for the positive and negative part, as in Section 4.3, particularly Eq. 4.5.

The average storage inflow for the 64-bus network model in Fig. 5.9 confirms that hydrogen storage is mainly utilised for wind power generation, whereas battery storage mostly receives inflow from solar power generation. The capacity distribution of pumped hydro has not been optimised but as a short-term storage option is primarily used for power generated by PV panels, too.

5.5 CONCLUSION AND OUTLOOK

Flow tracing is a well-known method to dissect the power flows on a network according to shares attributed to the network-injecting source nodes [169, 170]. Such an attribution of power flows and thus network usage has been proposed as an essential component for a fair allocation of both operational (for instance losses or stability measures) and grid infrastructure costs [161, 175, 176, 187]. In the present contribution we show how a reformulation of the flow tracing method serves as powerful tool set to analyze the complex spatio-temporal

2 The work behind this section has been prepared as "Flow-based analysis of storage usage in a low-carbon European electricity scenario" by Tranberg et al. [189].

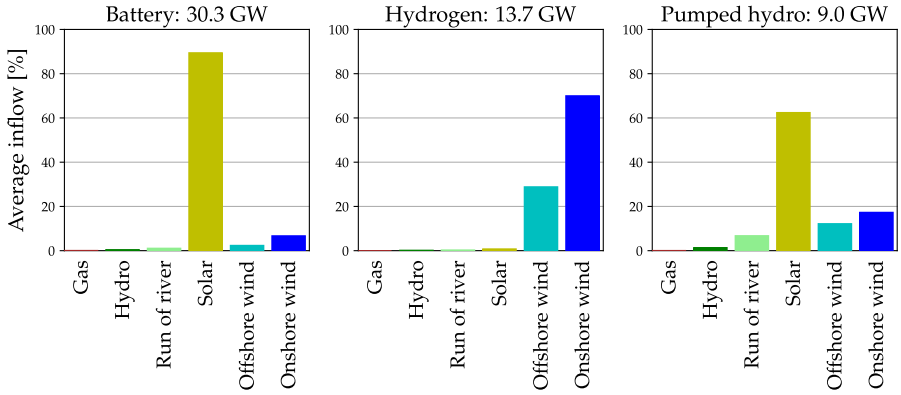


Figure 5.9: Average hourly inflow per storage technology in the 64 bus network with a moderate transmission expansion of $x = 1.5$ from Chapter 4. Reproduced from Tranberg et al. [189].

patterns of generation, consumption and power flows in interconnected large-scale electricity systems, in particular those with a high share of renewable generation. At the point of injection into the network, the power flow can straightforwardly be assigned to a specific geographical location, mix of different generation technologies or any other attribution of interest. Following the composition of ingoing flow from the net generators through the network, the algorithm yields in an intuitive way the respective shares of the total power flow and of the outflow to the net consumers.

The potential of this method is illustrated in the context of the Scenario 2023B of the 118-bus transmission benchmark case with renewable capacities [168]. We dissect the power flows into components associated with the geographical origin and generation technology (wind offshore/onshore, solar, conventional), yielding a selection of measures about the respective transmission system usage and the corresponding import-export relations between the system nodes. The need of incorporating relevant correlations in the aggregation procedure from full high-dimensional results to a lower-dimensional expression is discussed by comparing different transmission capacity usage measures assigned to the geographical regions of the benchmark network.

The discussion in the present chapter suggests future work on different levels. From a technical point of view, it will be interesting to transfer the idea of more general in-partitions to alternative methods of flow allocation [164, 179, 190]. First steps in this direction have been taken by [166] for allocation methods based on power transfer distribution factors, but a rigorous discussion is still lacking in the literature. In the present paper we focus on illustrating the versatility of the reformulated flow tracing method by considering a well-defined and fully open benchmark system.

6 CONCLUSIONS AND OUTLOOK

In this thesis we have considered the design of the energy system for very high shares of renewable energy. In particular we have focused on the roles the different spatial scales play in the energy system.

The conclusions fall into two categories: (1) Physics-inspired improvement of the methodology for energy system planning, and (2) concrete insights on the family of cost-optimal solutions of the European electricity system.

6.1 ENERGY SYSTEM PLANNING

The dynamics of the planning and operation of an electricity system have been smeared over several scales by the technological advancements of wind and solar generation: Their coupling to spatio-temporal weather correlations leads to a dependence on external sources of flexibility, which have to be evaluated from a systemic point of view. The hourly to yearly to decadal temporal scales, as well as the smaller-than-country up to the continental spatial scales, are all found to be inter-dependent for assessing the system's evolution.

The form-invariance of the linearized power flow equations to the Kirchhoff's laws for DC circuits was exploited to identify several power flow formulations by transforming the complementary equations individually to the dual graph. In the Kirchhoff formulation, the voltage law (KVL) as well as the current law (KCL) are directly expressed in terms of the branch flow variables by using a cycle basis, instead of the conventionally-used voltage angles. This formulation preserves the sparseness of the system of equality constraints in the LOPF and clarifies the coupling between the power flow constraints and the central nodal energy balance constraint. An extensive benchmark shows that this formulation reduces the solution time of the LOPF problem in respect to the other formulations by up to two orders of magnitude and scales better for larger networks and longer covered time periods. The application to the co-optimization of generation, storage and transmission expansion as discussed in Chapter 4 reveals that models of up to 512 buses with every third hour of a year or alternatively in hourly resolution with 196 buses, which consistently fail to converge with the conventional Angles formulation, become tractable for the commercial solver Gurobi on strong compute nodes. The solution occupies around 100 GB of memory and four processors for 20 to 24 hours (for each of the 5-7 iteration steps).

Long-term power system analysis has been strongly impeded by the scattered availability of the required data: Topology and electrical parameters of the transmission network, the current capacities of generation and storage units, the spatio-temporal time-series of both the electricity demand and the

potential generation of solar and wind installations. Several projects associated with the Open Energy Modelling Initiative like OPSD, the OEDB and SciGRID have recently started to address this need, and the European electricity system model PyPSA-Eur described in Chapter 3 pulls from several of them and combines them in a semi-automated process to a complete model for the self-developed PyPSA framework. It is freely available, easily maintainable and will continue to be extended from within the follow-up projects.

A k-means based clustering method on the spatial configuration of the nodes has been used to reduce the spatial complexity of the European renewable electricity network and makes it amenable to commercial high-performance solvers. The interpolation between the common "aggregated" one-node-per-country electricity system models towards the highest-voltage network models[1] indicates: The adverse effects of the extra transmission constraints represented in the model are approximately cancelling with the additional freedom to distribute the renewable generation capacities, when focusing on the total system cost. The optimal generation mix of low resolution models, though, is skewed in favour of offshore wind and underestimates the spatially varying potential of onshore wind generation. The solutions of the optimization family stabilize approximately for model details corresponding to more than 200 buses for the European continent with the current transmission volume. If the grid bottlenecks can be adressed by a small amount of expansion, solutions for networks with 120-150 buses corresponding to specific lengths between 100 to 200 km (25%-75% quantiles) are already close to the detailed networks. This might be an indication that the energy system models need to resolve the correlation lengths of varying renewable generation, in particular of wind with 300 to 600 km. The co-optimization of generation, storage and transmission capacities in a spatially resolved electricity system model achieves a significant cost reduction by a selective expansion of transmission corridors.

The flowtracing method, which originally segments the instantaneous power flow in an electricity network to contributions of individual generators, has been straightforwardly extended to the analysis of sections of the network or flexible groupings of generators. In a realistic examplary network, the principal flow patterns have been determined and split into wind, solar and conventionally generated energy. The varying generation of renewable generation, here too, increases the importance of considering the distribution of the flow over time: The energy mix in a link was demonstrated to be strongly dependent on the line loading and led to a more adequate measure for transmission usage. In the example, an above-proportional amount of the transmission volume is attributed to wind turbines.

1 The full models have not been solved for the identified European spatial and temporal scope, yet.

6.2 THE COST-OPTIMAL DECARBONIZED EUROPEAN ELECTRICITY SYSTEM

The optimal geographical distribution of the renewable generation capacities is strongly driven by resource quality: In the South PV panels generate most of the energy. Their production is strongly correlated over large areas and has to be shifted on diurnal time-scales to match demand using short-term storage, batteries and the available pumped hydro storage. In these areas, the current transmission grid is already strong enough to transport the solar energy to the demand centers, except for the capacities at the distribution level which can not be assessed by the current model. The energy demand in Northern Europe is optimally fed from wind turbines in high-capacity factor regions typically situated in coastal areas. The most cost-efficient grid is to build a high-volume transmission band along the wind installations for averaging their feed-in across the continent, since the synoptic scale variability needs to be buffered by expensive hydrogen storage, otherwise.

Today's transmission capacities with an ideal coordination of dispatchable hydro and (CO_2 emission capped) gas generation can support an energy mix based on 35% of solar and 50% wind energy[2] by the addition of 40 GWh of batteries and 80 GW of electrolysers and fuel cells for hydrogen. Though, a third of the hydrogen storage and two thirds of the batteries can be avoided with a moderate 50% expansion of the transmission volume, which allows to replace 15 percentage points of solar by offshore wind. The average system cost, thus, reduces from 82 to 68 Eur/MWh. A further, but minor, cost decrease to 66 Eur/MWh is possible by expanding transmission to three times its current volume and leveraging very remote high-capacity factor onshore wind potentials to displace solar, as well as offshore wind energy. The additional reach for spatial smoothing reduces the need for hydrogen converters to 30 GW, then. Repeated delays of previously planned transmission corridors due to public opposition may be helped by a compromise solution, though [82].

This non-linear benefit of transmission has already been observed in [31, 38]. In a simple model of the European electricity system, in which each country installed wind and solar capacities to minimize their average local residual load, the overall amount of energy from dispatchable thermal capacities fell sharply with increasing capacities of the tie-lines between countries: Expanding them to two times of today's net transfer capacities (NTC) avoids already 70% of the emissions that could be saved with unconstrained power flows at 12×NTC. In [191], the distribution of the renewable capacities was instead chosen so as to minimize the system levelized cost of electricity (LCOE) and, importantly, the benefit of transmission was shown to translate to a non-linear system cost reduction by about 10%. Schlachtberger et al. [40] demonstrated in a linear techno-economic one-node-per-country model that the cost benefit persists even when short- and long-term storage is available and carbon-free feed-in from hydro generation can be dispatched flexibly. The model found an

2 In respect to total demand after losses from curtailment and storage efficiency.

expansion to about nine times the volume of today's NTC[3] to be cost-optimal, but most of the cost reduction to an average system cost of 68 Eur/MWh would already be available from a 4× expansion of the NTC to 125 TW km as a compromise. These findings are consistent with this work, mainly because NTCs have been deliberately constrained to a very low level to reflect country-internal bottlenecks and the inefficient usage before installing flow-based market coupling (FBMC) [192]. The analoguous volume for the cross-country transmission lines of the detailed model in Chapter 3 adds already up to $\mathcal{V}_{today} = 83\,\text{TW km}$. Using the detailed model as a reference, instead, the compromise solution is an expansion to $1.5 \times \mathcal{V}_{today}$ and the cost-optimal grid is $3.4 \times \mathcal{V}_{today}$.

In summary, by using a more detailed model that can represent regions within each country and important intra-country bottlenecks, we have shown that less drastic grid expansion is required than the one-node-per-country results indicate. Modelling at this level of detail was only made possible by the reformulation of the optimal power flow problem presented in Chapter 2 and the clustering algorithm introduced in Chapter 4. We expose low-carbon configurations of the European electricity system integrating high shares of varying wind and solar sources with long- and short-term storage and flexible hydro and gas capacities at today's system costs, by a high CO_2 price ($\sim 200\,\text{Eur}/\text{tCO}_2$) in an integrated European electricity market.

6.3 OUTLOOK

Model extensions

Coupling electricity to other energy demand sectors, such as heating and transport, and to other energy carriers, such as gas and heat, is a necessary part of reducing GHG emissions from the use of fossil fuels in the energy system. Since chemical and heat carriers are significantly easier to store than electrical energy, they are able to replace much of the flexibility locking in up to a fifth of the costs in the presented solutions essentially for free. Brown et al. [193] demonstrate the cost benefit from sector coupling to be higher than the one from transmission in a one-node-per-country setting. At the spatial scales identified in this thesis, the large flexible demand of synthetic fuel production, coupled with cheap thermal energy storage in the heating sector might eliminate the economic benefit of transmission expansion beyond today's capacities altogether.

The model in this work is only able to assess the grid expansion on the UHV and EHV levels. Meanwhile at least in studies with a German focus, the expansion of the distribution grid added between 10% and 15% to system costs, which could affect the optimal mix between the lower-voltage connected PV against the medium-voltage or high-voltage connected onshore and offshore wind. Ancillary services, like reserves for voltage and frequency, traditionally

3 Like in the models presented in this work, the assumed length of the tie-lines stretches between country centres.

should be controlled for with simple estimations in a post analysis to guard against unexpected interactions [16, 41, 194–196].

Model simplifications

Especially the increase in technological scope brought by sector coupling will strongly impact model complexity and add to the computational burden. Several venues of simplification should be explored:

Schäfer et al. [197] found analytical scaling laws for the transmission capacities and transmission volume of a quantile transmission layout for the unconstrained flows arising from a simple dispatch scheme in coarse-grained renewable electricity networks. If these held also for an economic or transmission-constrained dispatch, they could inform scale-dependent cost penalties lessening the effect of a low-resolution representation. Similarly, technology dependent transmission line loading or underestimated resource variance could be captured by factors derived from analysing the micro-scale statistics – like the LCOE shifts from Becker et al. [198].

The scale thresholds for modelling do not necessarily have to be the same for different technologies of the energy system, like we have assumed in this work. The observed effect of better resource availability winning over constraints by grid bottlenecks might suggest that the spatial scale available for the distribution of renewable capacities could be chosen at a finer resolution than the grid detail. This technique has already been successfully employed in the one-node-per-country studies of the FRESNA group, for which up to four onshore wind turbine distributions were available within larger countries [40, 193, 199].

Final remarks

As models integrate more components, detail and interlinkages, so grows the need for systemic analysis to understand the results. In this thesis we have presented work that goes in both directions: we have demonstrated a methodology for modelling in more spatial detail, as well as presenting techniques, such as flow tracing, to help interpret the complex results. While we have demonstrated the necessity of spatial detail to avoid over-simplification, this has to go hand in hand with new techniques to make the results transparent and comprehensible. Combining complexity with comprehensibility is one of the great challenges of the energy transition.

A TRANSMISSION LINE MODEL

A single transmission line is commonly modeled as a two-port network: At each end of the transmission line current enters the cables and leaves by the ground with a specific voltage difference. An infinitesimal length $\mathrm{d}l$ of the line is characterised by four parameters as shown in Fig. A.1.

SERIES REACTANCE $x \approx 0.49\,\Omega/\mathrm{km}$ (marginal energy "loss" by self induction)

SERIES RESISTANCE $r \approx 0.05\,\Omega/\mathrm{km}$ (energy loss by heat dissipation)

SHUNT SUSCEPTANCE $b \approx 3.4\,\mu/\Omega\mathrm{km}$ (provides/balances reactive power)

SHUNT CONDUCTANCE $g \ll 1\,\mu/\Omega\mathrm{km}$ (energy loss by charge dissipation into ground)

The specified typical values for a 50 Hz-230 kV-line have been taken from [200]. The parameters add up to a complex *series impedance* $z = r + ix$ and a *shunt admittance* $y^{sh} = g + ib$. Note that the series impedance is several magnitudes larger than the shunt admittance.

The response of any two-port network can also be reproduced by a simple substitute circuit consisting of two admittances Y_1^Π, Y_2^Π connecting the upper and the lower branches and one impedance Z^Π on the upper branch, shown in Fig. A.2. Due to the typical arrangement of the circuit diagram this substitute is known as a Π-equivalent circuit. To prove the equivalence one first uses the general loop and node rules once on the equivalent circuit and then on the $\mathrm{d}l$-increment and integrates along the full length of the line [59], resulting in two matrices relating (V_1, I_1) and (V_2, I_2). They are identical for the choice

$$Z^\Pi = Z_L = Z\frac{\sinh \gamma l}{\gamma l} \approx Z\,, \qquad Y_{1,2}^\Pi = Y_L = Y\frac{\tanh \gamma l/2}{\gamma l/2} \approx Y \tag{A.1}$$

$$Z = zl\,, \qquad Y = yl\,, \qquad \gamma = \sqrt{zy}\,. \tag{A.2}$$

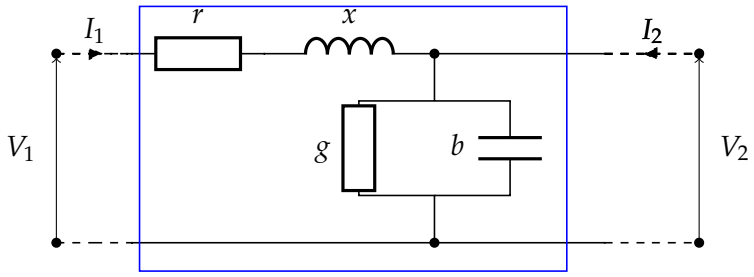


Figure A.1: Transmission line model. The blue box describes an infinitesimal part of the transmission line. To find the relation between the input and output voltages V_i and currents I_i, the line has to be integrated along its length.

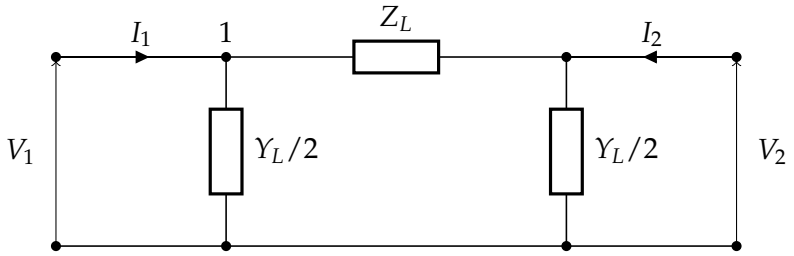


Figure A.2: Π-equivalent circuit of a transmission line. The shunt admittance can be distributed at the nodes while the series impedance clumps naturally.

The approximation holds for short-/medium-length lines ($\lesssim 300\,e\,km$), for which it, thus, suffices to distribute the admittance Y at the nodes and the impedance Z in between, both linear in the length l.

B SWING EQUATION

The AC power flow equation 2.3 described in Section 2.2 is valid at each point in time, but does not describe the evolution of the energy network with time, although the parameters S_n and P_n do vary considerably in time and space, as we will discuss in the next section. What is even more, in the case of large-scale disturbances like transmission line or generator contingencies, feedback mechanisms of the generators and loads lead to complex transient network dynamics. Considerable attention in the literature has been spent on variants of the swing equation with the form

$$\frac{2H_n}{\omega_R}\ddot{\delta}_n + \frac{D_n}{\omega_R}\dot{\delta}_n = A_n - \sum_{m=1, m\neq n} K_{n,m} \sin\left(\delta_n - \delta_m - \gamma_{n,m}\right) \ . \tag{B.1}$$

They essentially describe the phase (δ) dynamics of a damped (D_n) rotational mass ($\sim H_n$) driven by a synchronous voltage source ($\sim e^{i\delta_n}$) coupled to the phases of other electrical components δ_m. The coupling terms $K_{n,m}$ depend on the model for the dynamic response of the load; refer to [66] for a comparison. Depending on the specific research question, this classical 1-dimensional generator model might also have to be replaced by a two-axis generator model or incorporate the nonlinear response of the governors for frequency and voltage magnitude [69].

ZUSAMMENFASSUNG

Die Dekarbonisierung des Energiesystems ist angesichts der erwarteten Risiken der globalen Erwärmung eine der drängendsten Aufgaben unserer Zeit. Die Erzeugung elektrischer Energie aus der Verbrennung fossiler Brennstoffe wird von erneuerbaren Energiequellen wie Wasserkraft, Wind und Sonne abgelöst, unterstützt durch staatliche Steuerungsmechanismen und fallende Kosten aus technologischem Fortschritt und Skaleneffekten in der Herstellung. Der unvermeidliche Wechsel von flexibel einsetzbarer Erzeugung zu wetterabhängigen, räumlich und zeitlich fluktuierenden Kraftwerken verknüpft die Erzeugung und Verteilung von Elektrizität zu einem eng verflochtenen komplexen System in mehreren Dimensionen und Disziplinen:

- In der Zeit, hängen die langen Skalen in Jahren und Dekaden für die Planung und den Bau neuer Kapazität von den vielen Skalen des Wettersystems ab. Die Strahlung der Sonne verändert sich hauptsächlich täglich und saisonal, während Windgeschwindigkeiten von einem breiten Frequenzband aus Fluktuationen mit Perioden mehrerer Wochen zu einem Jahr gekennzeichnet sind.

- Im Raum werden die Energiesystemlösungen von den langreichweitigen Korrelationen sowie von den lokalen Variationen des Wetters, als auch durch lokale Engpässe in den Übertragungsnetzen, beeinflußt. Wie in Deutschland bereits beobachtet werden kann, schränken diese Engpässe den Energiefluß aus Offshorewindenergieanlagen in den Süden ein.

- Die Entscheidungen über Technologiemix und räumlicher Verteilung der gebauten Kapazitäten werden in der Regel von Investoren auf Basis ökonomischer Prinzipien getroffen,

- innerhalb von Rahmenbedingungen die in sozialen und politischen Rückkopplungsschleifen aus öffentlicher Meinung und Lobbyeinflüssen gebildet werden.

Die vorliegende Arbeit beschreibt die Entwicklung einer selbst-konsistenten Familie von Modellen des europäischen Elektrizitätssystems, die die Steuerung des physikalischen Gleichgewichtszustands im Stromnetz, Investitionen in Infrastruktur und vereinfachte politische Rahmenbedingungen abbilden. Gegenüber vergleichbaren Analysen des europäischen Stromsystems konnte die räumliche Auflösung bei gleichem zeitlichem Detail deutlich über vorherige Rechengrenzen erhöht werden. Das höhere räumliche Detail der erneuerbaren Potentiale und die simultane Optimierung von Erzeugungs-, Speicher und Übertragungskapazitäten findet Lösungen den Europäischen Elektrizitätsbedarf zu heutigen Kosten mit nur 5% der Kohlenstoffdioxid-Emissionen zu decken; und bedeutenderweise unterscheidet sich der optimale Technologiemix systematisch von den Ergebnissen mit geringer Auflösung.

Kapitel 1 beginnt mit einer knappen Einführung in die Geschichte und die Erkenntnisse der Forschung zum Klimawandel und den Vorhersagen und wahrscheinlichen Folgen der globalen Erwärmung, um die Notwendigkeit zu verdeutlichen, den Ausstoß von Klimagasen in den nächsten Jahrzenten weitgehend zu beenden. Die wichtigsten erneuerbaren Energiequellen Wasserkraft, Wind und Sonne werden eingeführt und die hauptsächlichen Frequenzen ihrer Fluktuationen mit den Betriebs- und Planungszeitskalen heutiger Elektrizitätssysteme verglichen. Ergebnisse aus der Literatur zum Zusammenspiel erneuerbarer Energien mit Kurz- und Langzeitspeichern, Übertragungsnetzkapazitäten und dem Energiebedarf aus anderen Energiesektoren werden knapp umrissen, um eine tagesaufgelöste Darstellung mindestens eines Jahres des gesamten europäischen Elektrizitätssystems in einem techno-ökonomischen Optimierungsansatz zu begründen. Ein solcher Ansatz erlaubt die wechselseitigen Abhängigkeiten zwischen den unterschiedlichen Skalen der räumlichen und zeitlichen Dimensionen in einem systemischen Rahmen abzubilden.

In Kapitel 2 werden die Bestimmungs-Gleichungen der elektrischen Leistungsflüße durch das Stromnetz linearisiert und auf die Kirchhoffschen Knoten und Maschenregeln in Gleichstromschaltkreisen zurückgeführt. Mit Hilfe graphen-theoretischer Zerlegungen zeigen wir, dass diese Gleichungen in sieben zueinander äquivalenten linearen Gleichungssystemen formuliert werden können. Diese Gleichungssyteme bilden die zentralen Zwangsbedingungen des linearen optimalen Leistungsflußes, der den ökonomischen Kraftwerkseinsatz und damit auch Rentabilität und die Verteilung von Investitionen bestimmt. Die Lösungszeit der daraus resultierenden neuen Formulierungen ist für viele Probleme um eine oder mehr Größenordnungen kleiner und skaliert wesentlich besser in der Anzahl der Netzwerkknoten und -kanten als die bisherigen Formulierungen, die auf Phasenwinkeln oder Leistungsverteilungsfaktoren beruhen. Für die Bestimmung der optimalen Erzeugungs- und Übertragungskapazitäten in einem europäischen Netzwerk mit mehr als 200 Knoten, um die stündlichen Lasten eines Referenzjahres mit erneuerbaren Energien, Kurz- und Langzeitspeichern zu liefern, und Gaskraftwerken, um die letzten Leistungsspitzen zu bedienen, wachsen die linearen Optimierungsprobleme zu ungefähr 15 Millionen Ungleichungen für etwa 10 Millionen Variablen und konvergieren erst in der neuen Kirchhoff-Formulierung.

Kapitel 3 beschreibt die Sammlung, Ergänzung und Zusammenfassung aller Daten zu einem Modell des europäischen Elektrizitätssystems. Die Netzwerktopologie der 220 kV und 380 kV Spannungsebenen wird aus einer Onlinekarte des Zusammenschlusses der Netzbetreiber (ENTSO-E) extrahiert und um typische elektrische Parameter und Transformatoren zwischen den Spannungsebenen aus der Literatur ergänzt. Konventionelle Kraftwerkskapazitäten werden algorithmisch aus sechs unvollständigen veröffentlichten Kraftwerkslisten entnommen, standardisiert und mit statistischen Methoden zu einem weitgehend konsistenten Datensatz zusammengefasst. Die zeitliche Verfügbarkeit der Leistung erneuerbarer Energieerzeugung aus Wasserkraft, Windenergieanlagen und Solaranlagen wird aus einem Reanalyse-Wetterdatensatz für das Gebiet um jeden Netzwerkknoten abgeschätzt. Ihre Ausbaupotentiale werden aus Landnutzungsdaten berechnet. Der Elektrizitätsbedarf wird aus dem

Landesverbrauch auf die einzelnen Netzwerkknoten disaggregiert. Zuletzt werden die einzelnen Modellbestandteile mit veröffentlichten aggregierten Daten verglichen und vertretbare Abweichungen gefunden.

Das Kapitel 4 ist der zentrale Teil der Arbeit. Zunächst gibt es eine kurze Übersicht über Methoden zur Netzwerkreduktion in der Literatur. Viele Methoden beziehen sich auf die Vereinfachung des Einflußes eines externen Netzwerkteils auf ein Kerngebiet eines Elektrizitätsnetzwerkes. Dagegen ist das Ziel dieser Arbeit grobkörnigere Näherungen des detaillierten Netzwerkmodells aus dem vorherigen Kapitel zu erzeugen. Dies geschieht in zwei Schritten:

1. Die Netzwerkknoten jedes Landes werden mit dem sogenannten k-means Algorithmus aus dem Feld des Maschinellen Lernens auf der Basis ihrer räumlichen Verteilung und Gewicht im Netzwerk in eine gegebene Anzahl an Zonen eingeteilt. Das Gewicht eines Knotens wird hierbei als die ihm relativ zugewiesene mittlere Last und Kapazitäten thermischer Kraftwerke angenommen.

2. Die Zonen werden zu Netzknoten eines gröberen Netzwerks aggregiert und einzelne Kraftwerke nach Technologie zusammengefasst. Erneuerbare Erzeugungszeitreihen und Ausbaupotentiale müssen gewichtet addiert werden. Alle Übertragungsleitungen zwischen zwei Zonen werden mit einer einzelnen äquivalenten Leitung ersetzt, die einem Übertragungskorridor entspricht.

Die lineare Optimierung der Kapazitäts- und der Energiekosten des europäischen Elektrizitätssystems mit Wind-, Solar- und Wasserkrafterzeugung, Lang- und Kurzzeitspeichern und Gaskraftwerken wird für sieben unterschiedliche Näherungsstufen gelöst, vom 37-Knoten Modell mit einem Knoten pro synchroner Zone und Land, zu einem hochaufgelösten 362-Knoten Modell.

In den Ergebnissen kann man zwei gegenläufige Effekte beobachten: (1) Zum einen erlaubt es die Abbildung von feineren räumlichen Skalen im hochaufgelösten Modell erneuerbare Kraftwerke gezielt an den ertragreichsten Orten zu platzieren. Die gewichtete Mittelung mit der die Resourcen in den Modellen mit weniger Knoten zusammengefasst werden, reduziert die vorhandenen Kapazitätsfaktoren. (2) Auf der anderen Seite, sind in den Modellen mit mehr Knoten auch mehr der Übertragungsengpässe innerhalb eines Landes vertreten, die den Transport von Windstrom aus Regionen mit hohem Kapazitätsfaktor entlang der Küsten und im Meer zu den Lastzentren im Landesinneren verhindern. Wenn keine Übertragungskapazitäten ausgebaut werden können, kompensieren sich die Auswirkungen der beiden Effekte auf die Systemkosten ungefähr. Auf den optimalen Energiemix wirken sie sich aber dennoch deutlich aus: Da Strom aus Offshore-Windanlagen schwerer ins Landesinnere zu transportieren ist, und mehr und bessere lastnahe Wind und Solarerzeugung mit höherer Auflösung zur Verfügung steht, halbiert sich der Anteil an Offshore Windkapazitäten vom gröbsten zum feinsten Modell zugunsten der Onshore Erzeugung.

Diese Trends werden vor allem deutlich, betrachtet man die Veränderung der Kapazitäten im deutschen Netz: Im Modell mit nur einem Knoten werden in Deutschland noch 40 GW Offshore-Wind und 46 GW Solarmodule gebaut. Bei

höheren Auflösungen stellt sich der Offshore-Wind allerdings als eine Fehlinvestition heraus, da die Netzengpässe innerhalb Deutschlands verhindern, dass der Windstrom in zu den Lastzentren im Westen und Süden Deutschlands transportiert wird. Nur 12 GW Offshore-Wind lassen sich ohne Netzausbau in Deutschland integrieren und müssen mit zusätzlichen 50 GW Solarkapazität ausgeglichen werden.

Ab einer Auflösung von in etwa 200 Netzknoten in Europa werden die Verteilungen der Erzeugungskapazitäten stabiler und verändern sich nur noch graduell.

Sobald eine kleine Erweiterung der vorhandenen Transmissionkapazitäten erlaubt ist, können die Netzengpässe mit geringem Kostenaufwand gelockert werden. Tatsächlich gilt, dass im feineren Netzwerk ein Netzengpass mit geringerem Aufwand aufgehoben werden kann. Im Resultat senken sich die Systemkosten mit einer hohen Zahl an Netzwerkknoten bereits bei geringem Netzausbau. Netzausbau wurde allerdings im letzten Jahrzehnt immer wieder durch den Protest von lokalen Organisationen und Initiativen zum Schutz von Umwelt und Heimat verhindert, und ist trotz seiner vergleichsweise geringen Investitionskosten nur eingeschränkt realisierbar. Der Netzausbau zum ökonomischen Optimum auf das 3-fache Netzvolumen senkt die Systemkosten um ca. 20% auf 66 €/MWh, ist aber wahrscheinlich nicht erreichbar. Um die Bildung eines Kompromisses zwischen sozialer Akzeptanz und den Kostenvorteilen zu analysieren, wird der Ausbau des elektrischen Netzvolumens schrittweise eingeschränkt. Es zeigt sich ein deutlich nicht-lineares Verhalten: Mehr als die Hälfte des Kostenvorteils kann bereits mit einem Ausbau des Netzvolumens um 25% erreicht werden und zwischen dem 2-fachen und dem 3-fachen Netzvolumen verändern sich die Kosten nur noch im Promillebereich. Die Verteilung der erneuerbaren Erzeugungskapazitäten verschiebt sich bei höherem Übertragungsvolumen von PV-Panelen und Offshore-Wind zu Onshore-Wind. Bei sehr begrenzter Übertragungskapazität wird Sonnenkraft in der Nähe von Lastzentren gebaut und ihre täglichen Schwankungen mit Batteriespeichern ausgeglichen, während der Windstrom aus windreichen Gegend in Nordeuropa sich nicht in großen Mengen zum Strombedarf transportieren lässt. Winderzeugung profitiert besonders von zusätzlichen Netzkapazitäten entlang den Küsten, die dazu genutzt werden die zeitlichen synoptischen Fluktuationen der Windgeschwindigkeiten mit Hilfe der räumlichen Variabilität auszugleichen. Diese Fluktuationen haben Perioden von mehreren Wochen und könnten sonst nur mit teuren Langzeitspeichern ausgeglichen werden, die Wasserstoff oder Gas synthetisieren, speichern und anschließend wieder unter hohen Verlusten verstromen.

Bei ausreichend hoher räumlicher Auflösung haben die kosteneffizienten Lösungen für das europäische Elektrizitätssystem eine klare an den Kapazitätsfaktoren ausgerichtete räumliche Struktur: Der Energiebedarf im Süden wird aus Solarstrom gedeckt, die Residuallast wird durch Batterien und Pumpspeicherkraftwerke und steuerbarer Wasserkraft ausgeglichen. Bereits heute bestehende Übertragungskapazitäten sind dafür ausreichend. Im Norden wird Windenergie an den Küsten und Offshore erzeugt, und mit einem neuen Leitungsband entlang dieser Windinstallationen ausgeglichen, um den Bedarf an

Langzeitspeichern möglichst gering zu halten. Wird mehr Netzausbau zugelassen, ist es günstiger den Windstrom weiter in den Süden zu transportieren und damit Solarmodule und Batteriespeicher zu ersetzen.

Die Analyse des komplexen Wechselspiels der einzelnen Komponenten erfordert einen systemischen quantitativen Ansatz, aber auch die zusätzlichen Schritte auf Zusammenhänge zwischen den Annahmen und Parametern, die teilweise, wie im zentralen Fall der Kosten, nur auf Schätzungen und unsicheren Extrapolationen beruhen, und den Ergebnissen zu schließen, um robuste Erkenntnisse über das systemische Verhalten zu gewinnen. Eine zentrale Rolle kommt hierbei den Flüssen im Netzwerk zu, die systemisch zunächst nur durch eine künstliche Einschränkung der Kapazität einzelner oder einer Gruppe von Übertragungsleitungen untersucht werden kann. Im Kapitel 5 wird aus dieser Motivation eine Methode mit dem Namen Flowtracing weiterentwickelt, die auf einem vektoriellen Diffusionsprozess auf dem gerichteten Flussgraph im Netzwerk beruht. Sie zerlegt den Leistungsfluss durch das Stromnetz in einzelne Energieflüsse von Quellen zu Senken und kann in der Erweiterung auch flexibel dafür eingesetzt werden, die Energieflüsse zwischen Regionen und von bestimmten Technologiegruppen zu untersuchen. Es wird demonstriert, dass die synoptischen Schwankungen der Windenergieerzeugung zu langreichweitigen Flüssen durch das Netzwerk bei hoher Leitungsbeladung führen. Bezieht man diese Korrelation in ein Leitungsauslastungsmaß für verschiedene Technologie ein, nehmen in einem realistischen Versuchsnetz sowohl Wind- als auch Solarenergie mehr als doppelt soviel Leitungsvolumen ein, als sie zum Energiemix beitragen. Die Bestimmung des Strommixes bei dem einzelne Speichertypen geladen werden, bestätigt die wechselseitige Beziehung, dass Solarstrom durch Kurzzeitspeicher wie Batterien und Pumpspeicherkraftwerke über den Tag verteilt werden, während die volle Ausnutzung von Windstrom in einem defossilierten Energiesystem auf Langzeitspeichern beruht.

Die vorliegende Arbeit kombiniert Methoden der statistischen Physik, der Physik komplexer Systeme und der komplexen Netzwerktheorie, um die Planung erneuerbarer Elektrizitätsnetzwerke an die Realität variierender wetterabhängiger Einspeisung und des damit verbundenen Flexibilitätsbedarfs an Speicher, Stromnetz und flexibler Erzeugung anzupassen. Dadurch kann der systemische Standpunkt einer techno-ökonomischen Optimierung mit hohem CO_2 Preis über die heutigen Marktgrenzen hinweg auf Skalen zwischen 100 und 200 km eingenommen werden, die die Korrelationslängen der Variation erneuerbarer Energien auflösen, und Lösungen mit deutlich reduziertem Netzausbau finden, die nur geringe Mengen an Offshore-Windenergie verwenden.

Für die Interpretation muss allerdings bedacht werden, dass ungefähr zwei Drittel der anthropogenen Kohlenstoffdioxidemissionen nicht aus der Elektrizitätserzeugung sondern aus der Energieverwendung in anderen Sektoren resultiert, wie in der Einleitung aufgeschlüsselt wird. Die heutige Sicht ist, dass Sektoren, wie der Wärme- oder der Transportsektor, in der Zukunft mehr und mehr elektrifiziert werden sollen. Zum einen wird das zu zusätzlichen Lasten führen die hier noch nicht berücksichtigt wurden und zusätzliche Erzeugungskapazitäten erfordern, zum anderen sind die Energieformen dieser zusätzlichen Lasten flexibel speicherbar und können damit laut aktueller Studien zur Flexibilität im Elektrizitätssektor benutzt werden, solange diese systemisch koordiniert werden!

CURRICULUM VITAE

Name: Jonas Hörsch

Birth: 26th September 1985, in Bamberg

EDUCATION

Dec 2014 - Dec 2018	Physics PhD student at the Renewable Energy System and Network Analysis group, Frankfurt Institute for Advanced Studies (FIAS), in close collaboration with Aarhus University, Denmark, under supervision of Prof. Dr. Stefan Schramm and Prof. Dr. Martin Greiner
Sep 2006 - Nov 2014	Studies in Physics at the University of Potsdam
	Diploma in Physics at the University of Potsdam, grade "Sehr gut" (very good)
	Diploma exam in theoretical physics, experimental physics, mathematics and quantum optics
	Diploma thesis: "Quantum Synchronization" at the Freie Universität Berlin under supervision of Prof. Dr. Jens Eisert and Prof. Dr. Markus Aspelmeyer, University of Vienna
Apr 2011 - Mar 2012	Research stay at the University of Vienna, collaboration with Prof. Dr. Aspelmeyer
Sep 1996 - Aug 2005	Attendance at Franz-Ludwig-Gymnasium, Bamberg
	Graduation: Allgemeine Hochschulreife (higher education entrance qualification), grade: 1.7 (good).
	Focus subjects: Physics, French.

TEACHING

2016 - 2017	Master thesis co-supervision of Fabian Hofmann and Clara Steinebach
Oct 2015 - Jun 2017	Tutor for three Complex renewable energy networks courses
Apr 2014 - Aug 2014	Tutor for Quantum mechanics
Oct 2007 - Jul 2009	Tutor for several Theoretical Physics and Mathematics courses

ACKNOWLEDGEMENTS

First of all, I would like to thank my external supervisor, Martin Greiner, for the opportunity to work with him. He gave me hospitality and support, whenever I came to Aarhus or strayed afar. He never stops pointing at the fundamental questions when others are caught in modelling frenzy. I am equally grateful to Stefan Schramm for accepting me in the group, for stoically meeting under-prepared talks by me with encouraging advice. The space and flexibility he gave me to explore this wide topic are greatly appreciated.

I am indebted to Sarah Becker who watched over my first steps in the topic and at FIAS. She never failed to answer boring questions for even a third time and enthusiastically embraced my half-baked refactoring attempts. She and Bjørn introduced and welcomed me to the finer Hessian traditions.

Mirko Schäfer with his clear analytical understanding and reasoning provided me with the perfect counter-balance to my numerical endeavors. Thanks for helping me understand the gist of complex network theory and sitting with me through my first and close to last publishing attempts.

I am happy to have shared an office and some of the more unusual working hours together with my fellow PhD student David Schlachtberger, whom I greatly admire for his structured and steady way of working. It was a joy to work with and co-supervise – patchily, I admit – our bright master students Fabian Hofmann, Clara Steinebach and Markus Schlott. Go do great stuff!

I was lucky that Tom Brown joined our group when I started coming to grips with the fundamental and simplistic reasoning about power systems. His idealism, motivation and goals were catching and shaped large parts of this work. He relentlessly provided support and counsel. I can only deceive myself into thinking we are equally profiting from the encounter, hopefully for some time. He also introduced me to the open energy modelling initiative, where I was engagingly received by like-minded enthusiasts.

Besides those I can explicitly mention, there are several more friendly people at the FIAS in Frankfurt and the AUESG in Aarhus who deserve credit for providing a supportive and friendly atmosphere and infrastructure.

Thanks to my friends and family who supported my nomadic months of writing with encouragement, acceptance and an open ear. Finally, I would like to thank Lena for her bursts of support and continuous caring nudges.

BIBLIOGRAPHY

[1] C. D. Rodgers and C. D. Walshaw. "The computation of infra-red cooling rate in planetary atmospheres". In: *Quarterly Journal of the Royal Meteorological Society* 92.391 (Jan. 1966), pp. 67–92. ISSN: 1477-870X. DOI: 10.1002/qj.49709239107.

[2] Syukuro Manabe and Richard T. Wetherald. "Thermal Equilibrium of the Atmosphere with a Given Distribution of Relative Humidity". In: *Journal of the Atmospheric Sciences* 24.3 (May 1967), pp. 241–259. ISSN: 1520-0469. DOI: 10.1175/1520-0469(1967)024<0241:teotaw>2.0.co;2.

[3] R. K. Pachauri et al. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by R.K. Pachauri and L. Meyer. Geneva, Switzerland: IPCC, 2014, p. 151. DOI: 10013/epic.45156.d001.

[4] Syukuro Manabe and Richard T. Wetherald. "The Effects of Doubling the CO2-Concentration on the climate of a General Circulation Model". In: *Journal of the Atmospheric Sciences* 32.1 (Jan. 1975), pp. 3–15. ISSN: 1520-0469. DOI: 10.1175/1520-0469(1975)032<0003:teodtc>2.0.co;2.

[5] Michael E. Schlesinger and John F. B. Mitchell. "Climate model simulations of the equilibrium climatic response to increased carbon dioxide". In: *Reviews of Geophysics* 25.4 (1987), p. 760. ISSN: 8755-1209. DOI: 10.1029/rg025i004p00760.

[6] A. A. Lacis, G. A. Schmidt, D. Rind, and R. A. Ruedy. "Atmospheric CO2: Principal Control Knob Governing Earth's Temperature". In: *Science* 330.6002 (Oct. 2010), pp. 356–359. ISSN: 1095-9203. DOI: 10.1126/science.1190653.

[7] Colin P. Kelley, Shahrzad Mohtadi, Mark A. Cane, Richard Seager, and Yochanan Kushnir. "Climate change in the Fertile Crescent and implications of the recent Syrian drought". In: *Proceedings of the National Academy of Sciences* 112.11 (Mar. 2015), pp. 3241–3246. ISSN: 1091-6490. DOI: 10.1073/pnas.1421533112.

[8] UNFCCC. *Adoption of the Paris Agreement. Report No. FCCC/CP/2015/L.9*. 2015. URL: http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf.

[9] Detlef P. van Vuuren, Andries F. Hof, Mariësse A. E. van Sluisveld, and Keywan Riahi. "Open discussion of negative emissions is urgently needed". In: *Nature Energy* 2.12 (Dec. 2017), pp. 902–904. ISSN: 2058-7546. DOI: 10.1038/s41560-017-0055-2.

[10] Greet Janssens-Maenhout et al. "EDGAR v4.3.2 Global Atlas of the three major Greenhouse Gas Emissions for the period 1970-2012". In: *Earth System Science Data Discussions* (Aug. 2017), pp. 1–55. ISSN: 1866-3591. DOI: 10.5194/essd-2017-79.

[11] J.P. Deane, B.P. Ó Gallachóir, and E.J. McKeogh. "Techno-economic review of existing and new pumped hydro energy storage plant". In: *Renewable and Sustainable Energy Reviews* 14.4 (May 2010), pp. 1293–1302. ISSN: 1364-0321. DOI: 10.1016/j.rser.2009.11.015.

[12] Dave Jones, Alice Sakhel, Matthias Buck, and Patrick Graichen. *The European Power Sector in 2017. State of Affairs and Review of Current Developments*. Tech. rep. Agora Energiewende and Sandbag, 2018. URL: https://www.agora-energiewende.de/fileadmin/Projekte/2018/EU_Jahresrueckblick_2017/Agora_EU-report-2017_WEB.pdf.

[13] ENTSO-E. *Electricity in Europe 2016*. Online, retrieved May 2018. 2016. URL: https://www.entsoe.eu/Documents/Publications/Statistics/electricity_in_europe/entsoe_electricity_in_europe_2016_web.pdf.

[14] Adrian Whiteman, Tobias Rinke, Javier Esparrago, Iana Arkhipova, and Samah Elsayed. "Renewable capacity statistics 2018". In: *IRENA* (Mar. 2018), p. 47. URL: http://www.irena.org/publications/2018/Mar/Renewable-Capacity-Statistics-2018.

[15] Fraunhofer ISE. *Photovoltaics Report*. Tech. rep. Updated on 26 February 2018. 2018. URL: https://www.ise.fraunhofer.de/content/dam/ise/de/documents/publications/studies/Photovoltaics-Report.pdf.

[16] T. W. Brown, T. Bischof-Niemz, K. Blok, C. Breyer, H. Lund, and B. V. Mathiesen. "Response to 'Burden of proof: A comprehensive review of the feasibility of 100% renewable-electricity systems'". In: *Renewable and Sustainable Energy Reviews* (2018). DOI: 10.1016/j.rser.2018.04.113. arXiv: 1709.05716v1.

[17] Open Power System Data. *Data Package Time series. Version 2017-07-09*. July 2017. URL: https://data.open-power-system-data.org/time_series/2017-07-09/.

[18] James R Holton and Gregory J Hakim. *An Introduction to Dynamic Meteorology (5th ed.)* Elsevier, 2012. ISBN: 9780123848666. DOI: 10.1016/c2009-0-63394-8.

[19] Clara M St. Martin, Julie K Lundquist, and Mark A Handschy. "Variability of interconnected wind plants: correlation length and its dependence on variability time scale". In: *Environmental Research Letters* 10.4 (Apr. 2015), p. 044004. ISSN: 1748-9326. DOI: 10.1088/1748-9326/10/4/044004.

[20] D. Connolly, H. Lund, B.V. Mathiesen, and M. Leahy. "A review of computer tools for analysing the integration of renewable energy into various energy systems". In: *Applied Energy* 87.4 (2010), pp. 1059–1082. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2009.09.026.

[21] Andrea Herbst, Felipe Toro, Felix Reitze, and Eberhard Jochem. "Introduction to Energy Systems Modelling". In: *Swiss Journal of Economics and Statistics* 148.2 (Jan. 2012), pp. 111–135. ISSN: 2235-6282. DOI: 10.1007/bf03399363.

[22] S. Jebaraj and S. Iniyan. "A review of energy models". In: *Renewable and Sustainable Energy Reviews* 10.4 (2006), pp. 281–311. ISSN: 1364-0321. DOI: 10.1016/j.rser.2004.09.004.

[23] Stefan Pfenninger, Adam Hawkes, and James Keirstead. "Energy systems modeling for twenty-first century energy challenges". In: *Renewable and Sustainable Energy Reviews* 33 (May 2014), pp. 74–86. ISSN: 1364-0321. DOI: 10.1016/j.rser.2014.02.003.

[24] Darryl R. Biggar and Mohammad Reza Hesamzadeh. *The Economics of Electricity Markets*. Wiley, 2014.

[25] Alexander Zerrahn and Wolf-Peter Schill. "Long-run power storage requirements for high shares of renewables: review and a new model". In: *Renewable and Sustainable Energy Reviews* 79 (2017), pp. 1518–1534. DOI: 10.1016/j.rser.2016.11.098.

[26] Wolf-Peter Schill and Alexander Zerrahn. "Long-run power storage requirements for high shares of renewables: Results and sensitivities". In: *Renewable and Sustainable Energy Reviews* 83 (Mar. 2018), pp. 156–171. ISSN: 1364-0321. DOI: 10.1016/j.rser.2017.05.205.

[27] D. Connolly, H. Lund, B.V. Mathiesen, E. Pican, and M. Leahy. "The technical and economic implications of integrating fluctuating renewable energy using energy storage". In: *Renewable Energy* 43 (2012), pp. 47–60. ISSN: 0960-1481. DOI: 10.1016/j.renene.2011.11.003.

[28] Yvonne Scholz. "Renewable energy based electricity supply at low costs. Development of the REMix model and application for Europe". PhD thesis. Universität Stuttgart, 2012. DOI: 10.18419/opus-2015.

[29] Venkat Krishnan, Jonathan Ho, Benjamin F. Hobbs, Andrew L. Liu, James D. McCalley, Mohammad Shahidehpour, and Qipeng P. Zheng. "Co-optimization of electricity transmission and generation resources for planning and policy analysis: review of concepts and modeling approaches". In: *Energy Systems* 7.2 (2016), pp. 297–332. ISSN: 1868-3975. DOI: 10.1007/s12667-015-0158-4.

[30] Schaber, K., Steinke, F., Mühlich, P., and Hamacher, T. "Parametric study of variable renewable energy integration in Europe: Advantages and costs of transmission grid extensions". In: *Energy Policy* 42 (Mar. 2012), pp. 498–508. DOI: 10.1016/j.enpol.2011.12.016.

[31] Sarah Becker, Rolando Rodríguez, Gorm Andresen, Stefan Schramm, and Martin Greiner. "Transmission grid extensions during the build-up of a fully renewable pan-European electricity supply". In: *Energy* 64 (Jan. 2014), pp. 404–418. DOI: 10.1016/j.energy.2013.10.010. arXiv: 1307.1723.

[32] D. Connolly, H. Lund, and B.V. Mathiesen. "Smart Energy Europe: The technical and economic impact of one potential 100% renewable energy scenario for the European Union". In: *Renewable and Sustainable Energy Reviews* 60 (July 2016), pp. 1634–1653. ISSN: 1364-0321. DOI: 10.1016/j.rser.2016.02.025.

[33] Katrin Schaber. "Integration of variable renewable energies in the European power system: a model-based analysis of transmission grid extensions and energy sector coupling". PhD thesis. Technische Universität München, 2013.

[34] Paul Nahmmacher, Eva Schmid, Lion Hirth, and Brigitte Knopf. "Carpe diem: A novel approach to select representative days for long-term power system modeling". In: *Energy* 112 (Oct. 2016), pp. 430–442. ISSN: 0360-5442. DOI: 10.1016/j.energy.2016.06.081.

[35] Kris Poncelet, Erik Delarue, Daan Six, Jan Duerinck, and William D'haeseleer. "Impact of the level of temporal and operational detail in energy-system planning models". In: *Applied Energy* 162 (Jan. 2016), pp. 631–643. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2015.10.100.

[36] Gregor Czisch. "Szenarien zur zukünftigen Stromversorgung". PhD thesis. Universität Kassel, 2005.

[37] Heide, D., von Bremen, L., Greiner, M., Hoffmann, C., Speckmann, M., and Bofinger, S. "Seasonal optimal mix of wind and solar power in a future, highly renewable Europe". In: *Renewable Energy* 35.11 (2010), pp. 2483–2489.

[38] Rolando A. Rodriguez, Sarah Becker, Gorm Andresen, Dominik Heide, and Martin Greiner. "Transmission needs across a fully renewable European power system". In: *Renewable Energy* 63 (Mar. 2014), pp. 467–476. DOI: 10.1016/j.renene.2013.10.005. arXiv: 1306.1079.

[39] Christian Bussar, Melchior Moos, Ricardo Alvarez, Philipp Wolf, Tjark Thien, Hengsi Chen, Zhuang Cai, Matthias Leuthold, Dirk Uwe Sauer, and Albert Moser. "Optimal allocation and capacity of energy storage systems in a future European power system with 100% renewable energy generation". In: *Energy Procedia* 46 (2014), pp. 40–47. DOI: 10.1016/j.egypro.2014.01.156.

[40] D.P. Schlachtberger, T. Brown, S. Schramm, and M. Greiner. "The benefits of cooperation in a highly renewable European electricity network". In: *Energy* 134.Supplement C (2017), pp. 469–481. ISSN: 0360-5442. DOI: 10.1016/j.energy.2017.06.004.

[41] *Kombikraftwerk 2: Abschlussbericht*. Tech. rep. Fraunhofer IWES et al., Aug. 2014. URL: http://www.kombikraftwerk.de/mediathek/abschlussbericht.html.

[42] Jonas Egerer, Clemens Gerbaulet, and Casimir Lorenz. "European Electricity Grid Infrastructure Expansion in a 2050 Context". In: *The Energy Journal* 37.01 (Sept. 2016). ISSN: 0195-6574. DOI: 10.5547/01956574.37.si3.jege.

[43] Schaber, K., Steinke, F., and Hamacher, T. "Transmission grid extensions for the integration of variable renewable energies in Europe: Who benefits where?" In: *Energy Policy* 43 (2012), pp. 123–135. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2011.12.040.

[44] Tom Brown, Thomas Ackermann, Eckehard Tröster, and Peter-Philipp Schierhorn. "Optimising the European transmission system for 77% renewable electricity by 2030". In: *IET Renewable Power Generation* 10.1 (Jan. 2016), pp. 3–9. ISSN: 1752-1424. DOI: 10.1049/iet-rpg.2015.0135.

[45] S. Hagspiel, C. Jägemann, D. Lindenberger, T. Brown, S. Cherevatskiy, and E. Tröster. "Cost-optimal power system extension under flow-based market coupling". In: *Energy* 66 (2014), pp. 654–666. ISSN: 0360-5442. DOI: 10.1016/j.energy.2014.01.025.

[46] Joshua Adam Taylor. *Convex Optimization of Power Systems*. Cambridge University Press, 2015.

[47] Jonas Hörsch, Henrik Ronellenfitsch, Dirk Witthaut, and Tom Brown. "Linear optimal power flow using cycle flows". In: *Electric Power Systems Research* 158 (2018), pp. 126–135. ISSN: 0378-7796. DOI: 10.1016/j.epsr.2017.12.034. arXiv: 1704.01881.

[48] K. Purchala, L. Meeus, D. Van Dommelen, and R. Belmans. "Usefulness of DC power flow for active power flow analysis". In: *IEEE Power Engineering Society General Meeting*. June 2005, 454–459 Vol. 1. DOI: 10.1109/PES.2005.1489581.

[49] B. Stott, J. Jardim, and O. Alsac. "DC Power Flow Revisited". In: *IEEE Trans. Power Syst.* 24.3 (2009), p. 1290. DOI: 10.1109/TPWRS.2009.2021235.

[50] Fred C. Schweppe, Michael C. Caramanis, Richard D. Tabors, and Roger E. Bohn. "Spot Pricing of Electricity". In: (1988). DOI: 10.1007/978-1-4613-1683-1.

[51] Barbara Burstedde. "Essays on the Economic of Congestion Management - Theory and Model-based Analysis for Central Western Europe". PhD thesis. Universität zu Köln, 2012.

[52] Allen J. Wood, Bruce F. Wollenberg, and Gerald B. Sheblé. *Power Generation, Operation and Control*. New York: John Wiley & Sons, Nov. 2013, p. 656.

[53] G. Latorre, R.D. Cruz, J.M. Areiza, and A. Villegas. "Classification of publications and models on transmission expansion planning". In: *IEEE Transactions on Power Systems* 18.2 (May 2003), pp. 938–946. ISSN: 0885-8950. DOI: 10.1109/tpwrs.2003.811168.

[54] T. Pesch, H.-J. Allelein, and J.-F. Hake. "Impacts of the transformation of the German energy system on the transmission grid". In: *The European Physical Journal Special Topics* 223.12 (June 2014), pp. 2561–2575. ISSN: 1951-6401. DOI: 10.1140/epjst/e2014-02214-y.

[55] A.J. Conejo and J.A. Aguado. "Multi-area coordinated decentralized DC optimal power flow". In: *IEEE Transactions on Power Systems* 13.4 (1998), pp. 1272–1278. ISSN: 0885-8950. DOI: 10.1109/59.736264.

[56] A.G. Bakirtzis and P.N. Biskas. "A decentralized solution to the DC-OPF of interconnected power systems". In: *IEEE Transactions on Power Systems* 18.3 (Aug. 2003), pp. 1007–1013. ISSN: 0885-8950. DOI: 10.1109/tpwrs.2003.814853.

[57] Olivier Megel, Goran Andersson, and Johanna L. Mathieu. "Reducing the computational effort of stochastic multi-period DC optimal power flow with storage". In: *2016 Power Systems Computation Conference (PSCC)* (June 2016). DOI: 10.1109/pscc.2016.7541033.

[58] Ariana Minot, Yue M. Lu, and Na Li. "A parallel primal-dual interior-point method for DC optimal power flow". In: *2016 Power Systems Computation Conference (PSCC)* (June 2016). DOI: 10.1109/pscc.2016.7540826.

[59] J.J. Grainger and W.D. Stevenson. *Power system analysis*. McGraw-Hill series in electrical and computer engineering: Power and energy. McGraw-Hill, 1994. ISBN: 9780070612938.

[60] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas. "MATPOWER: Steady-State Operations, Planning and Analysis Tools for Power Systems Research and Education". In: *IEEE Trans. Power Syst.* 26 (2011), p. 12. DOI: 10.1109/TPWRS.2010.2051168.

[61] Richard Lincoln. *PYPOWER Version 5.0*. 2015. URL: https://github.com/rwl/PYPOWER.

[62] DIgSILENT GmbH. *PowerFactory*. 2016. URL: http://digsilent.de/.

[63] Henrik Ronellenfitsch, Marc Timme, and Dirk Witthaut. "A Dual Method for Computing Power Transfer Distribution Factors". In: *IEEE Transactions on Power Systems* (2016). ISSN: 1558-0679. DOI: 10.1109/tpwrs.2016.2589464. arXiv: 1510.04645.

[64] Henrik Ronellenfitsch, Debsankha Manik, Jonas Hörsch, Tom Brown, and Dirk Witthaut. "Dual theory of transmission line outages". In: *IEEE Transactions on Power Systems* PP.99 (2017). ISSN: 0885-8950. DOI: 10.1109/TPWRS.2017.2658022. arXiv: 1606.07276.

[65] Burak Kocuk, Hyemin Jeon, Santanu S. Dey, Jeff Linderoth, James Luedtke, and Xu Andy Sun. "A Cycle-Based Formulation and Valid Inequalities for DC Power Transmission Problems with Switching". In: *Operations Research* 64.4 (2016), pp. 922–938. DOI: 10.1287/opre.2015.1471.

[66] Takashi Nishikawa and Adilson E Motter. "Comparative analysis of existing models for power-grid synchronization". In: *New Journal of Physics* 17.1 (Jan. 2015), p. 015012. ISSN: 1367-2630. DOI: 10.1088/1367-2630/17/1/015012.

[67] Debsankha Manik, Marc Timme, and Dirk Witthaut. "Cycle flows and multistability in oscillatory networks". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27.8 (Aug. 2017), p. 083123. ISSN: 1089-7682. DOI: 10.1063/1.4994177. arXiv: 1611.09825v2.

[68] G. Filatrella, A. H. Nielsen, and N. F. Pedersen. "Analysis of a power grid using a Kuramoto-like model". In: *The European Physical Journal B* 61.4 (Feb. 2008), pp. 485–491. ISSN: 1434-6036. DOI: 10.1140/epjb/e2008-00098-8. arXiv: 0705.1305v2.

[69] Paul M. Anderson and A. A. Fouad. *Power System Control and Stability*. IEEE, 2002. ISBN: 0470545577. DOI: 10.1109/9780470545577.

[70] D. Van Hertem, J. Verboomen, K. Purchala, R. Belmans, and W.L. Kling. "Usefulness of DC power flow for active power flow analysis with flow controlling devices". In: *The 8th IEEE International Conference on AC and DC Power Transmission*. 2006, pp. 58–62.

[71] Reinhard Diestel. *Graph Theory*. Springer Berlin Heidelberg, 2010. ISBN: 9783642142796. DOI: 10.1007/978-3-642-14279-6.

[72] M. E. J. Newman. *Networks – An Introduction*. Oxford: Oxford University Press, 2010.

[73] P. Kundur, N.J. Balu, and M.G. Lauby. *Power system stability and control*. EPRI power system engineering series. McGraw-Hill, 1994. ISBN: 9780070359581.

[74] Juan M. Morales, Antonio J. Conejo, Henrik Madsen, Pierre Pinson, and Marco Zugno. "Integrating Renewables in Electricity Markets: Operational Problems". In: *International Series in Operations Research & Management Science* (2014). ISSN: 2214-7934. DOI: 10.1007/978-1-4614-9411-9.

[75] T. Brown, J. Hörsch, and D. Schlachtberger. *PyPSA: Python for Power System Analysis Version 0.8*. 2017. URL: http://pypsa.org.

[76] Gurobi Optimization Inc. *Gurobi Optimizer Reference Manual*. 2016. URL: http://www.gurobi.com.

[77] Carsten Matke, Wided Medjroubi, and David Kleinhans. *SciGRID - An Open Source Reference Model for the European Transmission Network (v0.2)*. July 2016. URL: http://www.scigrid.de.

[78] Gorm B. Andresen, Anders A. Søndergaard, and Martin Greiner. "Validation of Danish wind time series from a new global renewable energy atlas for energy system analysis". In: *Energy* 93, Part 1 (2015), pp. 1074–1088. DOI: 10.1016/j.energy.2015.09.071. arXiv: 1409.3353.

[79] Johannes Köster and Sven Rahmann. "Snakemake: a scalable bioinformatics workflow engine". In: *Bioinformatics* 28.19 (2012), p. 2520. DOI: 10.1093/bioinformatics/bts480.

[80] *A Framework Strategy for a Resilient Energy Union with a Forward-Looking Climate Change Policy*. Tech. rep. European Commission, 2015. URL: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2015:80:FIN.

[81] BNetzA and BKartA. *Monitoringbericht 2016*. 2016. URL: http://tinyurl.com/bnetza-monitoringbericht-2016.

[82] A. Battaglini, N. Komendantova, P. Brtnik, and A. Patt. "Perception of barriers for expansion of electricity grids in the European Union". In: *Energy Policy* 47 (2012), pp. 254–259. DOI: 10.1016/j.enpol.2012.04.065.

[83] *ENTSO-E Pan-European Power Transmission Grid Datasets*. Dec. 2016. URL: https://www.entsoe.eu/stum/.

[84] Qiong Zhou and J.W. Bialek. "Approximate model of European interconnected system as a benchmark system to study effects of cross-border trades". In: *Power Systems, IEEE Transactions on* 20.2 (May 2005), pp. 782–788. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2005.846178.

[85] Rajan Gupta, Harihar Shankar et. al. *Global Energy Observatory*. Online, retrieved March 2017. 2017. URL: http://globalenergyobservatory.org/.

[86] N. Hutcheon and J.W. Bialek. "Updated and validated power flow model of the main continental European transmission network". In: *IEEE PowerTech (POWERTECH), 2013 Grenoble*. June 2013. DOI: 10.1109/PTC.2013.6652178.

[87] Tue V. Jensen and Pierre Pinson. "RE-Europe - a large-scale dataset for modeling a highly renewable European electricity system". In: *Scientific Data* 4 (Nov. 2017), p. 170175. ISSN: 2052-4463. DOI: 10.1038/sdata.2017.175. URL: http://pierrepinson.com/docs/JensenPinson2017.pdf.

[88] Tue V. Jensen, Hugo de Sevin, Martin Greiner, and Pierre Pinson. *The RE-Europe data set*. Dec. 2015. DOI: 10.5281/zenodo.620228.

[89] *OpenStreet Map*. 2017. URL: http://www.openstreetmap.org/.

[90] Malte Scharf and Arjuna Nebel. *osmTGmod: Open source German transmission grid model based on OpenStreetMap v0.1.3*. 2017. URL: https://github.com/wupperinst/osmTGmod.

[91] Jonas Egerer. "Open Source Electricity Model for Germany (ELMOD-DE)". In: *DIW Data Documentation* (2016). URL: https://www.diw.de/sixcms/detail.php?id=diw_01.c.528929.de.

[92] Jonas Hörsch. *PyPSA-Eur: An Open Optimization Model of the European Transmission System (dataset)*. 2017. URL: https://github.com/FRESNA/pypsa-eur.

[93] T. Brown, J. Hörsch, and D. Schlachtberger. *PyPSA: Python for Power System Analysis version 0.9.0*. Apr. 2017. DOI: 10.5281/zenodo.582307. URL: http://pypsa.org.

[94] *ENTSO-E Interactive Transmission System Map*. Jan. 2017. URL: https://www.entsoe.eu/map/.

[95] Bart Wiegmans. *GridKit: GridKit 1.0 'for Scientists'*. Mar. 2016. DOI: 10.5281/zenodo.47263.

[96] Bart Wiegmans. *GridKit extract of ENTSO-E interactive map*. June 2016. DOI: 10.5281/zenodo.55853.

[97] Wikipedia. *List of HVDC projects — Wikipedia, The Free Encyclopedia*. Online, accessed Sep 2016. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_HVDC_projects&oldid=782297059.

[98] D. Oeding and B.R. Oswald. *Elektrische Kraftwerke und Netze*. 7th ed. Springer, 2011. ISBN: 9783642192463. DOI: 10.1007/978-3-662-06960-8.

[99] Eurostat. *Infrastructure - electricity - annual data*. Dataset nrg_113a. Online, retrieved May 2017. 2017. URL: http://ec.europa.eu/eurostat/.

[100] ENTSO-E. *Net generating capacity*. 2015. URL: https://www.entsoe.eu/db-query/miscellaneous/net-generating-capacity.

[101] ENTSO-E. *Scenario Outlook & Adequacy Forecast*. 2015. URL: https://www.entsoe.eu/outlooks/maf/.

[102] Open Power System Data. *Data Package National generation capacity. Version 2017-07-07*. July 2017. URL: https://data.open-power-system-data.org/national_generation_capacity/2017-07-07/.

[103] Open Power System Data. *Data Package Conventional power plants. Version 2017-07-07*. July 2017. URL: https://data.open-power-system-data.org/conventional_power_plants/2017-07-07/.

[104] ENTSO-E Transparency Platform. *Installed Capacity Per Production Unit*. Online, retrieved Jul 2017. 2017. URL: https://transparency.entsoe.eu/generation/r2/installedCapacityPerProductionUnit/show.

[105] *Kraftwerksliste der Bundesnetzagentur*. 2017. URL: http://tinyurl.com/bnetza-kraftwerksliste.

[106] Jonas Hörsch, Fabian Hofmann, and Fabian Gotzens. *Powerplantmatching: Set of tools to combine multiple power plant databases*. 2017. URL: https://github.com/FRESNA/powerplantmatching.

[107] Sandia National Laboratories. *DOE Global Energy Storage Database*. Online, retrieved March 2017. 2017. URL: http://www.energystorageexchange.org/projects.

[108] David Wheeler and Kevin Ummel. "Calculating CARMA: global estimation of CO2 emissions from the power sector". In: *Available at SSRN 1138690* (2008).

[109] Kevin Ummel. "CARMA revisited: an updated database of carbon dioxide emissions from power plants worldwide". In: *Center for Global Development Working Paper* 304 (2012).

[110] World Resources Institute. *PowerWatch: An open-source global power plant database project*. Online, retrieved March 2017. 2017. URL: https://github.com/Arjay7891/WRI-Powerplant.

[111] Lars Marius Garshol. *Duke: a fast and flexible deduplication engine, Version 1.2*. Feb. 2014. URL: https://github.com/larsga/Duke/.

[112] A. K. Elmagarmid, P. G. Ipeirotis, and V. S. Verykios. "Duplicate Record Detection: A Survey". In: *IEEE Transactions on Knowledge and Data Engineering* 19.1 (Jan. 2007), pp. 1–16. ISSN: 1041-4347. DOI: 10.1109/TKDE.2007.250581.

[113] Pedro Domingos and Michael Pazzani. "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss". In: *Machine Learning* 29.2 (Nov. 1997), pp. 103–130. ISSN: 1573-0565. DOI: 10.1023/A:1007413511361.

[114] I. Androutsopoulos, J. Koutsias, K. V. Chandrinos, and C. D. Spyropoulos. "An Experimental Comparison of Naive Bayesian and Keyword-Based Anti-Spam Filtering with Personal E-mail Messages". In: *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Aug. 2000). ., pp. 160–167. arXiv: cs/0008019v1.

[115] Fabian Gotzens, Jonas Hörsch, and Heidi Heinrichs. "Setting up a European TIMES Electricity System Model in a transparent way – the issue of data quality in power plant portfolios". In: *EMP-E Special Issue, Energy Strategy Reviews* (2017). Submitted.

[116] Kies, A., Chattopadhyay, K., von Bremen, L., Lorenz, E., and Heinemann, D. *RESTORE 2050 Work Package Report D12: Simulation of renewable feed-in for power system studies.* Tech. rep. in preparation. RESTORE 2050, 2016.

[117] Benjamin Pfluger, Frank Sensfuß, Gerda Schubert, and Johannes Leisentritt. "Tangible ways towards climate protection in the European Union (EU Long-term scenarios 2050)". In: *Fraunhofer ISI* (2011).

[118] Saha, S. et al. "The NCEP Climate Forecast System Reanalysis". In: *Bull. Amer. Meteor. Soc.* 91.8 (2010), pp. 1015–1057.

[119] U.S. Energy Information Administration. *Hydroelectricity Net Generation Europe 2000-2014*. 2017. URL: http://tinyurl.com/EIA-hydro-gen-EU-2000-2014 (visited on 03/01/2017).

[120] EEA. *Corine Land Cover (CLC) 2012, Version 18.5.1*. Apr. 2012.

[121] EEA. *Natura 2000 data - the European network of protected sites*. 2016. URL: http://www.eea.europa.eu/data-and-maps/data/natura-7.

[122] Iain Staffell and Stefan Pfenninger. "Using bias-corrected reanalysis to simulate current and future wind power output". In: *Energy* 114 (2016), pp. 1224–1239. ISSN: 0360-5442. DOI: 10.1016/j.energy.2016.08.068.

[123] Douglas T Reindl, William A Beckman, and John A Duffie. "Diffuse fraction correlations". In: *Solar energy* 45.1 (1990), pp. 1–7.

[124] Soteris Kalogirou. *Solar energy engineering: processes and systems*. 1st ed. Burlington San Diego London: Elsevier Inc., 2009.

[125] DT Reindl, WA Beckman, and JA Duffie. "Evaluation of hourly tilted surface radiation models". In: *Solar energy* 45.1 (1990), pp. 9–17.

[126] Hans Georg Beyer, Gerd Heilscher, and Stefan Bofinger. "A robust model for the MPP performance of different types of PV-modules applied for the performance check of grid connected systems". In: *Eurosun. Freiburg* (2004).

[127] Gorm Andresen, Jonas Hörsch, and Tom Brown. *Atlite: Light-weight version of Aarhus RE Atlas for converting weather data to power systems data*. 2017. URL: https://github.com/FRESNA/atlite.

[128] Robert Carl Pietzcker, Daniel Stetter, Susanne Manger, and Gunnar Luderer. "Using the sun to decarbonize the power sector: The economic potential of photovoltaics and concentrating solar power". In: *Applied Energy* 135 (2014), pp. 704–720. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2014.08.011.

[129] Richard Müller, Uwe Pfeifroth, Christine Träger-Chatterjee, Jörg Trentmann, and Roswitha Cremer. "Digging the METEOSAT treasure—3 decades of solar surface radiation". In: *Remote Sensing* 7.6 (2015), pp. 8067–8101. URL: http://www.mdpi.com/2072-4292/7/6/8067.

[130] Stefan Pfenninger and Iain Staffell. "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data". In: *Energy* 114 (Nov. 2016), pp. 1251–1265. ISSN: 0360-5442. DOI: 10.1016/j.energy.2016.08.060.

[131] *Ludwig Hülk*, Lukas Wienholt, Ilka Cußmann, Ulf Philipp Müller, Carsten Matke, and *Editha Kötter*. "Allocation of annual electricity consumption and power generation capacities across multiple voltage levels in a high spatial resolution". In: *International Journal of Sustainable Energy Planning and Management* 13 (2017), pp. 79–92. DOI: 10.5278/ijsepm.2017.13.6. URL: https://journals.aau.dk/index.php/sepm/article/view/1833.

[132] Viktor Slednev, Manuel Ruppert, Valentin Bertsch, Wolf Fichtner, Nico Meyer-Hübner, Michael Suriyah, Thomas Leibfried, Philipp Gerstner, Michael Schick, and Vincent Heuveline. "Regionalizing Input Data for Generation and Transmission Expansion Planning Models". In: *Advances in Energy System Optimization: Proceedings of the first International Symposium on Energy System Optimization*. Ed. by Valentin Bertsch, Wolf Fichtner, Vincent Heuveline, and Thomas Leibfried. Springer International Publishing, 2017, pp. 205–217. ISBN: 978-3-319-51795-7. DOI: 10.1007/978-3-319-51795-7_13.

[133] ENTSO-E. *Country-specific hourly load data*. 2011. URL: https://www.entsoe.eu/data/data-portal/consumption/.

[134] ENTSO-E. *Lengths of circuits*. 2015. URL: https://www.entsoe.eu/db-query/miscellaneous/lengths-of-circuits.

[135] Inderjit S Dhillon, Yuqiang Guan, and Brian Kulis. "Kernel k-means: spectral clustering and normalized cuts". In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2004, pp. 551–556.

[136] E. Cotilla-Sanchez, P. D. H. Hines, C. Barrows, S. Blumsack, and M. Patel. "Multi-Attribute Partitioning of Power Networks Based on Electrical Distance". In: *IEEE Transactions on Power Systems* 28.4 (Nov. 2013), pp. 4979–4987. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2013.2263886.

[137] *Potenzial der Windenergienutzung an Land*. Tech. rep. Bundesverband WindEnergie, 2012. URL: https://www.wind-energie.de/shop-potenzial-der-windenergienutzung-land.

[138] *Potenzial der Windenergie an Land*. Tech. rep. Umweltbundesamt, 2013. URL: https://www.umweltbundesamt.de/sites/default/files/medien/378/publikationen/potenzial_der_windenergie.pdf.

[139] Norman Gerhardt et al. *Interaktion EE-Strom, Wärme und Verkehr*. Tech. rep. Fraunhofer IWES, 2015. URL: http://www.energiesystemtechnik.iwes.fraunhofer.de/de/projekte/suche/laufende/interaktion_strom_waerme_verkehr.html.

[140] Hans-Martin Henning and Andreas Palzer. *100% erneuerbare Energien für Strom und Wärme in Deutschland*. Tech. rep. Fraunhofer ISE, 2012. URL: https://www.ise.fraunhofer.de/de/veroeffentlichungen/veroeffentlichungen-pdf-dateien/studien-und-konzeptpapiere/studie-100-erneuerbare-energien-in-deutschland.pdf.

[141] Tom Brown, Jonas Hörsch, and David Schlachtberger. "PyPSA: Python for Power System Analysis". In: *Journal of Open Research Software* 6 (2018). ISSN: 2049-9647. DOI: 10.5334/jors.188. arXiv: 1707.09913v2.

[142] Jonas Hörsch and Tom Brown. "The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios". In: *14th International Conference on the European Energy Market*. Accepted after peer-review. June 2017, pp. 1–7. DOI: 10.1109/EEM.2017.7982024. arXiv: 1705.07617.

[143] Maximilian Schumacher and Lion Hirth. "How Much Electricity Do We Consume? A Guide to German and European Electricity Consumption and Generation Data". In: *FEEM Working Paper No. 88.215* (2015). . DOI: 10.2139/ssrn.2715986. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2715986#.

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

[145] S. Deckmann, A. Pizzolante, A. Monticelli, B. Stott, and O. Alsac. "Studies on Power System Load Flow Equivalencing". In: *IEEE Transactions on Power Apparatus and Systems* PAS-99.6 (Nov. 1980), pp. 2301–2310. ISSN: 0018-9510. DOI: 10.1109/TPAS.1980.319798.

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

[147] J. A. Hartigan and M. A. Wong. "Algorithm AS 136: A K-means clustering algorithm". In: *Applied Statistics* 28.1 (1979), pp. 100–108.

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

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

[150] Camille Hamon, Ebrahim Shayesteh, Mikael Amelin, and Lennart Söder. "Two partitioning methods for multi-area studies in large power systems". In: *International Transactions on Electrical Energy Systems* 25.4 (Jan. 2014), pp. 648–660. ISSN: 2050-7038. DOI: 10.1002/etep.1864.

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

[152] E. Shayesteh, C. Hamon, M. Amelin, and L. Söder. "REI method for multi-area modeling of power systems". In: *International Journal of Electrical Power & Energy Systems* 60 (2014), pp. 283–292. ISSN: 0142-0615. DOI: 10.1016/j.ijepes.2014.03.002.

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

[154] *eHighways 2050 Final Reports*. Tech. rep. ENTSO-E et al., 2015. URL: http://www.e-highway2050.eu.

[155] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

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

[157] Andreas Schröder, Friedrich Kunz, Jan Meiss, Roman Mendelevitch, and Christian von Hirschhausen. *Current and prospective costs of electricity generation until 2050*. eng. Data Documentation, DIW 68. Berlin: Deutsches Institut für Wirtschaftsforschung (DIW), 2013. URL: http://hdl.handle.net/10419/80348.

[158] Jonas Hörsch, Mirko Schäfer, Sarah Becker, Stefan Schramm, and Martin Greiner. "Flow tracing as a tool set for the analysis of networked large-scale renewable electricity systems". In: *International Journal of Electrical Power & Energy Systems* 96 (2018), pp. 390–397. ISSN: 0142-0615. DOI: 10.1016/j.ijepes.2017.10.024. arXiv: 1609.02977.

[159] M Aguado, R Bourgeois, J Bourmaud, J Van Casteren, M Ceratto, M Jakel, B Malfiet, C Mestdag, P Noury, M Pool, et al. "Flow-based market coupling in the central western european region-on the eve of implementation". In: *CIGRE, C5-204* (2012). URL: http://www.cigre.org/content/download/17044/680596/version/1/file/C5_204_2012.pdf.

[160] European Network of Transmission System Operators for Electricity. *Ten-Year Network Development Plan (TYNDP) 2016*. Tech. rep. ENTSO-E, 2016. URL: http://tyndp.entsoe.eu/.

[161] CONSENTEC and Frontier Economics. *Study on the further issues relating to the inter-TSO compensation mechanism*. 2006.

[162] Mario Neukirch. "Protests against German electricity grid extension as a new social movement? A journey into the areas of conflict". In: *Energy, Sustainability and Society* 6.1 (Feb. 2016), p. 4. ISSN: 2192-0567. DOI: 10.1186/s13705-016-0069-9.

[163] A. J. Conejo, J. Contreras, D. A. Lima, and A. Padilha-Feltrin. "Zbus Transmission Network Cost Allocation". In: *IEEE Transactions on Power Systems* 22.1 (Feb. 2007), pp. 342–349. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2006.889138.

[164] Y. C. Chen and S. V. Dhople. "Power Divider". In: *IEEE Transactions on Power Systems* 31.6 (Nov. 2016), pp. 5135–5143. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2016.2519900.

[165] H. Rudnick, R. Palma, and J. E. Fernandez. "Marginal pricing and supplement cost allocation in transmission open access". In: *IEEE Transactions on Power Systems* 10.2 (May 1995), pp. 1125–1132. ISSN: 0885-8950. DOI: 10.1109/59.387960.

[166] Tom Brown. "Transmission network loading in Europe with high shares of renewables". In: *IET Renewable Power Generation* 9.1 (Jan. 2015), pp. 57–65. ISSN: 1752-1424. DOI: 10.1049/iet-rpg.2014.0114.

[167] Jonas Hörsch, Tom Brown, and Stefan Schramm. "Spatial Scale Dependence in Joint Optimisation of Generation and Transmission for Highly Renewable Scenarios". In: *15th International Workshop on Large-Scale Integration of Wind Power into Power Systems*. Paper: WIW16-124. Nov. 2016.

[168] H. Barrios, A. Roehder, H. Natemeyer, and A. Schnettler. "A benchmark case for network expansion methods". In: *PowerTech, 2015 IEEE Eindhoven*. June 2015, pp. 1–6. DOI: 10.1109/PTC.2015.7232601.

[169] J. Bialek. "Tracing the flow of electricity". In: *Generation, Transmission and Distribution, IEE Proceedings-* 143.4 (July 1996), pp. 313–320. ISSN: 1350-2360. DOI: 10.1049/ip-gtd:19960461.

[170] D. Kirschen, R. Allan, and G. Strbac. "Contributions of individual generators to loads and flows". In: *Power Systems, IEEE Transactions on* 12.1 (Feb. 1997), pp. 52–60. ISSN: 0885-8950. DOI: 10.1109/59.574923.

[171] J. Bialek. "Topological generation and load distribution factors for supplement charge allocation in transmission open access". In: *Power Systems, IEEE Transactions on* 12.3 (Aug. 1997), pp. 1185–1193. ISSN: 0885-8950. DOI: 10.1109/59.630460.

[172] J. Bialek. "Allocation of transmission supplementary charge to real and reactive loads". In: *Power Systems, IEEE Transactions on* 13.3 (Aug. 1998), pp. 749–754. ISSN: 0885-8950. DOI: 10.1109/59.708576.

[173] G. Strbac, D. Kirschen, and S. Ahmed. "Allocating transmission system usage on the basis of traceable contributions of generators and loads to flows". In: *Power Systems, IEEE Transactions on* 13.2 (May 1998), pp. 527–534. ISSN: 0885-8950. DOI: 10.1109/59.667378.

[174] D. Kirschen and G. Strbac. "Tracing active and reactive power between generators and loads using real and imaginary currents". In: *Power Systems, IEEE Transactions on* 14.4 (Nov. 1999), pp. 1312–1319. ISSN: 0885-8950. DOI: 10.1109/59.801890.

[175] J.W. Bialek and S. Ziemianek. "Tracing based transmission pricing of cross-border trades: fundamentals and circular flows". In: *Power Tech Conference Proceedings, 2003 IEEE Bologna*. Vol. 3. June 2003, 8 pp. Vol.3-. DOI: 10.1109/PTC.2003.1304456.

[176] J.W. Bialek, S. Ziemianek, and R. Wallace. "A methodology for allocating transmission losses due to cross-border trades". In: *Power Systems, IEEE Transactions on* 19.3 (Aug. 2004), pp. 1255–1262. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2004.831296.

[177] Ignacio Pérez-Arriaga, Luis Olmos Camacho, and Francisco Javier Rubio Odériz. *Report on cost components of cross border exchanges of electricity*. 2002.

[178] Luis Olmos Camacho and Ignacio J. Pérez-Arriaga. "An assessment of inter-TSO compensation algorithms in the Internal Electricity Market of the European Union". In: *International Journal of Electrical Power & Energy Systems* 29.10 (Dec. 2007), pp. 699–712. ISSN: 0142-0615. DOI: 10.1016/j.ijepes.2007.05.004.

[179] Delberis A. Lima, Antonio Padilha-Feltrin, and Javier Contreras. "An overview on network cost allocation methods". In: *Electric Power Systems Research* 79.5 (2009), pp. 750–758. ISSN: 0378-7796. DOI: 10.1016/j.epsr.2008.10.005.

[180] A.A. Abou El Ela and R.A. El-Sehiemy. "Transmission usage cost allocation schemes". In: *Electric Power Systems Research* 79.6 (2009), pp. 926–936. ISSN: 0378-7796. DOI: 10.1016/j.epsr.2008.12.005.

[181] F. J. Rubio-Oderiz and I. J. Perez-Arriaga. "Marginal pricing of transmission services: a comparative analysis of network cost allocation methods". In: *IEEE Transactions on Power Systems* 15.1 (Feb. 2000), pp. 448–454. ISSN: 0885-8950. DOI: 10.1109/59.852158.

[182] Mirko Schäfer, Bo Tranberg, Sabrina Hempel, Stefan Schramm, and Martin Greiner. "Decompositions of injection patterns for nodal flow allocation in renewable electricity networks". In: *The European Physical Journal B* 90.8 (Aug. 2017). ISSN: 1434-6036. DOI: 10.1140/epjb/e2017-80200-y.

[183] Carlos Vazquez, Luis Olmos, and Ignacio J Pérez-Arriaga. "On the selection of the slack bus in mechanisms for transmission network cost allocation that are based on network utilization". In: *Proceedings of the 15th Power Systems Computing Conference, PSCC, Seville*. 2002.

[184] P.A Kattuman, R.J Green, and J.W Bialek. "Allocating electricity transmission costs through tracing: a game-theoretic rationale". In: *Operations Research Letters* 32.2 (2004), pp. 114–120. ISSN: 0167-6377. DOI: 10.1016/S0167-6377(03)00095-6.

[185] Mirko Schäfer, Sabrina Hempel, Jonas Hörsch, Tranberg Bo, Stefan Schramm, and Martin Greiner. "Power flow tracing in complex networks". In: *New Horizons in Fundamental Physics*. Ed. by Stefan Schramm and Mirko Schäfer. FIAS Interdisciplinary Science Series. Nov. 2016, pp. 357–373. DOI: 10.1007/978-3-319-44165-8_26.

[186] C. Achayuthakan, C.J. Dent, J.W. Bialek, and W. Ongsakul. "Electricity Tracing in Systems With and Without Circulating Flows: Physical Insights and Mathematical Proofs". In: *Power Systems, IEEE Transactions on* 25.2 (May 2010), pp. 1078–1087. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2009.2037506.

[187] Bo Tranberg, Anders B Thomsen, Rolando A Rodriguez, Gorm B Andresen, Mirko Schäfer, and Martin Greiner. "Power flow tracing in a simplified highly renewable European electricity network". In: *New Journal of Physics* 17.10 (Oct. 2015), p. 105002. ISSN: 1367-2630. DOI: 10.1088/1367-2630/17/10/105002.

[188] Tiago Soares, Fábio Pereira, Hugo Morais, and Zita Vale. "Cost allocation model for distribution networks considering high penetration of distributed energy resources". In: *Electric Power Systems Research* 124 (2015), pp. 120–132. ISSN: 0378-7796. DOI: 10.1016/j.epsr.2015.03.008.

[189] Bo Tranberg, Mirko Schäfer, Tom Brown, Jonas Hörsch, and Martin Greiner. "Flow-based analysis of storage usage in a low-carbon European electricity scenario". In: *15th International Conference on the European Energy Market*. 2018. DOI: 10.1109/EEM.2018.8469951. arXiv: 1806.02549.

[190] Delberis A. Lima and Antonio Padilha-Feltrin. "Allocation of the costs of transmission losses". In: *Electric Power Systems Research* 72.1 (2004), pp. 13–20. ISSN: 0378-7796. DOI: 10.1016/j.epsr.2004.03.002.

[191] Emil H. Eriksen, Leon J. Schwenk-Nebbe, Bo Tranberg, Tom Brown, and Martin Greiner. "Optimal heterogeneity in a simplified highly renewable European electricity system". In: *Energy* 133 (Aug. 2017), pp. 913–928. ISSN: 0360-5442. DOI: 10.1016/j.energy.2017.05.170.

[192] CEER ACER. *Annual Report on the Results of Monitoring the Internal Electricity and Gas Markets in 2016*. Tech. rep. Agency for the Cooperation of Energy Regulators (ACER), 2016. URL: http://www.acer.europa.eu/Official_documents/Acts_of_the_Agency/Publication/ACER%20Market%20Monitoring%20Report%202016%20-%20ELECTRICITY.pdf.

[193] T. Brown, D. Schlachtberger, A. Kies, S. Schramm, and M. Greiner. "Synergies of sector coupling and transmission extension in a cost-optimised, highly renewable European energy system". In: (2018). arXiv: 1801.05290v1 [physics.soc-ph].

[194] Deutsche Energie-Agentur. *DENA-Netzstudie II*. Tech. rep. 2010. URL: http://www.dena.de/publikationen/energiesysteme.

[195] Deutsche Energie-Agentur. *dena-Studie Systemdienstleistungen 2030*. Tech. rep. 2014. URL: https://www.dena.de/themen-projekte/projekte/energiesysteme/dena-studie-systemdienstleistungen-2030/.

[196] T. Ackermann, S. Untsch, M. Koch, and H. Rothfuchs. *Verteilnetzstudie Rheinland-Pfalz*. Tech. rep. 2014. URL: https://www.oeko.de/oekodoc/1885/2014-008-de.pdf.

[197] M. Schäfer, S. Bugge Siggaard, Kun Zhu, C. Risager Poulsen, and M. Greiner. "Scaling of transmission capacities in coarse-grained renewable electricity networks". In: *EPL (Europhysics Letters)* 119.3 (Aug. 2017), p. 38004. ISSN: 1286-4854. DOI: 10.1209/0295-5075/119/38004.

[198] Sarah Becker, Bethany A. Frew, Gorm B. Andresen, Timo Zeyer, Stefan Schramm, Martin Greiner, and Mark Z. Jacobson. "Features of a fully renewable US electricity system: Optimized mixes of wind and solar PV and transmission grid extensions". In: *Energy* 72 (Aug. 2014), pp. 443–458. ISSN: 0360-5442. DOI: 10.1016/j.energy.2014.05.067. arXiv: 1402.2833.

[199] David P. Schlachtberger, Tom Brown, Mirko Schäfer, Stefan Schramm, and Martin Greiner. "Cost optimal scenarios of a future highly renewable European electricity system: Exploring the influence of weather data, cost parameters and policy constraints". In: *Energy* (2018). submitted in Mar. 2018. arXiv: 1803.09711v1 [physics.soc-ph].

[200] J. Machowski, J. Bialek, and J. Bumby. *Power System Dynamics: Stability and Control*. Wiley, 2008. ISBN: 9780470714140.