Statistical Physics of Power Flows on Networks with a High Share of Fluctuating Renewable Generation

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich Physik

der Johann Wolfgang Goethe-Universität

in Frankfurt am Main

von

Dominik Heide

aus Frankfurt

Frankfurt, 2010

(D30)

vom Fachbereich Physik der

Johann Wolfgang Goethe - Universität als Dissertation angenommen.

Dekan: Prof. Dr. Michael Huth

Gutachter: Prof. Dr. Martin Greiner, Prof. Dr. Joachim Maruhn

Datum der Disputation:

Zusammenfassung

Regenerative Energien haben in der jüngsten Zeit eine große Bedeutung bekommen. Diese können die zwei größten Probleme der fossilen Energiequellen lösen: Die Begrenztheit der Ressourcen und das Anfallen von Abfällen bzw. Abgasen. Allerdings birgt ein Energieerzeugungssystem mit einem hohem Anteil erneuerbarer Energien neue Herausforderungen: Da wichtige erneuerbare Energien in Abhängigkeit von externen Einflüssen fluktuieren, stellen sie ein Problem für die Versorgungssicherheit dar. Dies ist vor allem bei Wind- und Solargeneratoren der Fall. Der Einfluß der Fluktuationen auf Leistungsflüsse im Transportnetz ist bislang nicht ausreichend verstanden. Um Übergänge von heutigen zu zukünftigen Stromversorgungssystemen planen zu können, ist jedoch ein gutes Verständnis von möglichen Zukunftsszenarien wichtig.

In dieser Arbeit wird daher zunächst die Machbarkeit von Stromversorgungssystemen mit einem hohen Anteil erneuerbarer Energien untersucht. Dabei wird sich herausstellen, daß Energietransport innerhalb Europas notwendig ist. Daher werden im Anschluss Transportflüsse mit Hilfe eines vereinfachten Modells analysiert und analytische Näherungen der Leistungsflüsse hergeleitet. Schließlich werden die aus den Last- und Erzeugungszeitreihen resultierenden Leistungsflüsse analysiert und mit den Modell-Ergebnissen verglichen.

Da Stromversorgungssysteme einen wichtigen Teil der Infrastruktur moderner Gesellschaften ausmachen, sind diese zum Gegenstand der Forschung komplexer Netze geworden. Dabei werden Methoden der statistischen Physik zur Analyse der Zusammenhänge von Struktur und Funktion von Netzwerken benutzt. Bisher wurden Transportnetzwerke hauptsächlich mit Bezug auf ihre Toleranz bei Ausfällen oder gezielten Attacken untersucht. Zur Beschreibung des Transportes werden in der Literatur hauptsächlich zwei Paradigmen genutzt: Zum einen der Transport entlang der kürzesten Pfade zwischen Quelle und Senke und zum anderen der sogenannte "DC-Fluß", der eine Näherung der vollen physikalischen Leistungsflußgleichungen darstellt.

Ein erster Ansatz, Flüsse in einem Netzwerk mit fluktuierenden Quellen zu verstehen, wird vorgeschlagen, die Wahrscheinlichkeitsverteilungen der Flüsse und die Stabilität des gesamten Transportnetzes untersucht. Diese erste Modellierung geschieht im Rahmen der in der Literatur benutzten Methoden, insbesondere des Transportparadigmas, das auf kürzesten Pfaden beruht. Transportkapazitäten werden aus den Analysen hergeleitet und den Verbindungen zugeordnet, sowie die Toleranz des Netzwerks bezüglich der fluktuierenden Flüsse getestet. Auftretende Fehlerkaskaden, die durch Umverteilung der Last einer überlasteten Verbindung auftreten, werden analysiert. Das vorgeschlagene Kapazitätslayout ermöglicht es, für gegebene Überlastungswahrscheinlichkeiten die

notwendigen Kapazitäten festzulegen. Dieser Ansatz ist jedoch nicht geeignet, um Leistungsflüsse in Transportnetzen zu beschreiben.

Um ein gutes Verständnis von Stromversorgungssystemen mit einem hohen Anteil erneuerbarer Energien zu erlangen, werden Verbrauchs- und Erzeugungszeitreihen von elektrischer Energie analysiert. Die Erzeugung wird dabei vor allem für Wind- und Solarenergie betrachtet. Deren Zeitreihen werden aus Wetterdaten generiert und mit den Lastdaten verglichen. Die Erzeugungskapazitäten werden nach den politischen Planungen für das Jahr 2020 räumlich zugeordnet. Um die Erzeugungs- und Lastreihen für ganz Europa zu erhalten werden die räumlich hoch aufgelösten Zeitreihen aufsummiert. Eine erste Betrachtung der Zeitreihen ergibt für Wind- und Soloarenergie, wie auch für den Verbrauch, eine starke saisonale Abhängigkeit. Die aggregierten europäischen Erzeugungsdaten zeigen, daß Winderzeugung und Last mit gleicher Phase saisonal variieren: Beide sind im Winter höher als im Sommer. Die Erzeugung der Windkraft ist allerdings im Winter höher als die Last und im Sommer signifikant niedriger, so daß bei einer Versorgung mit elektrischer Energie ausschließlich aus Windkraft Speicher vonnöten sind, die überschüssige Energie im Winter für den Sommer speichern. Solarenergie zeigt ein umgekehrtes Verhalten: Im Winter ist die europaweite Energieausbeute niedriger als im Sommer. Die 180°-Phasenverschiebung der Soloarenergie zur Last hat zur Konsequenz, daß in einem Szenario, in dem ausschließlich Sonnenenergie genutzt wird, größere Speicherkapazitäten benötigt wird, verglichen mit dem Fall, in dem elektrische Energie ausschließlich durch Windkraft erzeugt wird.

Da es häufig Zeiträume gibt, in denen die Erzeugung von Wind- und Sonnenenergien nicht ausreicht, um den Bedarf zu decken, werden zusätzliche Anlagen benötigt, die Energie speichern oder bei Bedarf kurzfristig Ausgleichsenergie erzeugen können. Die saisonalen Einflüsse, die in den Zeitreihen gefunden werden, legen nahe, daß eine Kombination aus Wind- und Solarenergie es ermöglicht, daß Last und Erzeugung einander angeglichen werden können. Dabei sollen die benötigten Speicherkapazitäten und die benötigte Ausgleichsenergie minimiert werden. Unter der Annahme, daß nur erneuerbare Energie aus Sonne und Wind genutzt wird, werden hierzu die Zeitreihen von Last und Erzeugung auf eins normiert. Dabei werden die Anteile von Wind-, bzw. Solarenergie mit den Faktoren a, bzw. b bezeichnet, mit der Bedingung $a + b = 1$. Dies impliziert, daß die Kapazitätszuweisungen skaliert werden, unter Beibehaltung der Verhältnisse der räumlich zugeordneten Erzeugungskapazitäten.

Die optimalen Verhältnisse von Wind- zu Solarenergie, ausgedrückt durch den Anteil der Windenergie $a = b - 1$, werden in Bezug auf den benötigten Speicher und die benötigte Ausgleichsenergie minimiert. Für den Fall einer ausschließlich regenerativen Energieerzeugung mit Wind- und Solarenergie finden wir eindeutige Minima der benötigten Speicher und Ausgleichsenergie. Diese liegen im Falle der Minimierung des Speichers bei $a \approx 0.6$ und für die Minimierung der Ausgleichsenergie bei $a \approx 0.8$. Die unterschiedlichen optimalen Mixe können durch Dynamiken in den Datenzeitreihen auf verschiedenen Zeitskalen erklärt werden. Tag-Nacht-Wechsel, und die dadurch benötigte Ausgleichsenergie um die Nachts fehlende Solarenergie zu kompensieren, führen zu einem optimalen Mix mit einem hohen Anteil von Windenergie, bei einer Minimierung der Aus-

gleichsenergie. Die zu speichernde Energie wird von den saisonalen Zeitskalen dominiert und führt daher zu einem anderen optimalen Mix. Analysiert man die 24-Stunden-Mittelwerte der Zeitreihen, so bleibt der optimale Mix bezüglich der Speicherenergie unverändert, der optimale Mix bezüglich der Ausgleichsenergie konvergiert ebenfalls zu $a \approx 0.6$.

Die Speicherenergie im optimalen Mix beträgt 10% des jährlichen Verbrauchs. Dies bedeutet, daß ungefähr 300TWh gespeichert werden müssen. Es ist nicht abzusehen, daß ein derartiger Speicherbedarf realisierbar wäre: Die Energie, die mit etablierten Technologien gespeichert werden kann, ist um Größenordnungen kleiner. Die Umwandlung von elektrischer Energie in Wasserstoff, dessen Speicherung in Salzkavernen und eine Rückverstromung des gespeicherten Wasserstoffs ist eine hypothetische Technologie, die absehbar die höchste Speicherkapazität bietet. Deren Größenordnung liegt im optimalen Fall bei 10 bis 20TWh. Die Effizienz dieser Speichertechnologie ist jedoch gering. Eine Möglichkeit, den Speicherbedarf in diesen voraussichtlich machbaren Bereich zu bringen, besteht in einer Überinstallation der Wind- und Solarenergiekapazitäten. Durch diese Überinstallation verändert sich der optimale Mix nur minimal. Die benötigten Speicherkapazitäten sinken jedoch bei 50% Überinstallation auf ungefähr 1% des jährlichen Verbrauchs.

Das europäische Szenario impliziert Transport von überschüssiger Leistung unabhängig von der räumlichen Lage von Erzeugung und Last. Ebenfalls interessant ist es, den Fall zu betrachten, daß einzelne Länder ausschließlich ihren eigenen Bedarf decken und daher kein Transport über Landesgrenzen hinweg stattfindet. Mit diesem und dem europäischen Szenario sind zwei Extremfälle abgedeckt. Für jedes Land wurden die benötigten Ausgleichs- und Speicherenergien ermittelt und ebenfalls eindeutige Minima gefunden, die als optimale Verhältnisse von Wind- und Sonnenenergie für das jeweilige Land interpretiert werden. Auffällig ist, daß in Abhängigkeit vom Breitengrad des analysiertem Landes, bei Betrachtung der benötigten Speicherenergie, ein höherer Anteil von Solarenergie im optimalen Mix gefunden wird. Griechenland, zum Beispiel, hat einen optimalen Mix bei $a \approx 0$, wohingegen Irland, mit $a \approx 0.8$, die benötigte Speicherenergie mit hauptsächlich Windenergie minimiert. Bei Betrachtung der Ausgleichsenergie findet sich der gleiche Effekt, allerdings in einem weit geringeren Maße mit Werten von $0.6 \leq a \leq 0.9$. Um die benötigten Speicher- und Ausgleichskapazitäten der beiden Szenarien zu vergleichen, werden die jeweiligen Kapazitäten der einzelnen Länder aufsummiert, mit dem Ergebnis, daß man für beide Szenarien einen über Europa aggregierten Wert erhält. Bei dem Vergleich ergibt sich, daß europaweiter Transport die benötigten Speicher und Ausgleichskapazitäten um einen Faktor von ungefähr 1.6 erhöhen, wenn es keinen Austausch zwischen den Ländern gibt. Europaweiter Energietransfer ermöglicht daher, fluktuierende Energieerzeugung in einem höheren Maß auszugleichen. Diese Ergebnisse illustrieren die Wichtigkeit des Transports von Energie, dem sich die vorliegende Arbeit daher im Folgenden zuwendet.

Zu der Frage, welche Auswirkungen fluktuierende Energiequellen auf die Leistungsflüsse in Transportnetzen haben, gab es bisher noch keine wissenschaftlichen Erkenntnisse.

Wir formulieren daher zunächst ein Modell mit dessen Hilfe diese Leistungsflüsse unter kontrollierten Bedingungen analysiert werden können. Dieses Modell basiert auf generalisierten und vereinfachten Beobachtungen in den Daten und berücksichtigt die probabilistischen Eigenschaften der Erzeugung. Die räumliche Verteilung von Erzeugung und Last sowie deren Verhältnis κ sind parametrisiert und die resultierenden Leistungsflüsse werden für verschiedene Transportnetzstrukturen mit Hilfe der Monte Carlo-Methode simuliert. In Anlehnung an andere Arbeiten werden nur quasi-stationäre Zustände berücksichtigt, da die Komplexität des Modells anderenfalls zu groß würde. Für die Berechnung der Transportflüsse wird die sogenannte "DC-Näherung" aus der Elektrotechnik benutzt, die auf verschiedenen, empirisch hergeleiteten Näherungen beruht, und nur die Wirkleistung berücksichtigt sowie Transportverluste vernachlässigt. In stationären Zuständen muß die Leistungsbilanz Null sein, d.h. die komplette eingespeiste Energie von den Verbrauchern entnommen werden. Dies ist bei fluktuierenden Quellen im Allgemeinen nicht der Fall, so daß zwei Ausgleichsverfahren definiert werden, die im Falle von Überproduktion oder Unterproduktion Verbraucher und Erzeuger so anpassen, daß diese übereinstimmen: Im Falle von überschüssiger Energie werden Erzeuger herunter geregelt, bei Energiemangel werden ausreichende Hilfsgeneratoren angenommen oder es wird Last abgeschaltet. Das sogenannte "Minimale Dissipation"-Ausgleichsverfahren minimiert die Summe der quadratischen Leistungsflüsse im Netz. Bei Nutzung des "Globaler Faktor"-Ausgleichsverfahren werden Überproduktion oder Unterdeckung gleichmäßig auf alle Erzeuger oder Verbraucher relativ zu ihrer Stärke verteilt. Die Simulationen zeigen, daß das "Minimale Dissipation"-Ausgleichsverfahren starke kurzreichweitige Korrelationen von Erzeugung und Last einführt. Bei dem "Globaler Faktor"-Ausgleichsverfahren sind die Korrelationen jedoch vernachlässigbar.

Die ungerichteten Leistungsflüsse in den Leitungen haben aufgrund der fluktuierenden Erzeugung einen probabilistischen Charakter. Ein Ergebnis der Simulation ist, daß diese Verteilungen mit wenigen Ausnahmen sehr gut durch generalisierte Gammaverteilungen beschrieben werden können. Im Folgenden leiten wir analytisch Näherungen der Momente der ungerichteten Leistungsflüsse her und können mit deren Hilfe die Verteilungen der Leistungsflüsse in sehr guter Näherung vorhersagen. Diese analytischen Näherungen basieren auf der zuvor diskutierten Beobachtung, daß die Erzeugung und Last für den Fall des "Globaler Faktor"-Ausgleichsverfahren unkorreliert sind. Dadurch lassen sich Theoreme der Wahrscheinlichkeitstheorie anwenden, um sowohl den Einfluß des Ausgleichsverfahrens als auch die Verteilung der Leitungsflüsse in den Leitungen zu beschreiben. Die zweiten Momente der Verteilungen der Leitungsflüsse lassen sich ohne weitere Näherungen bestimmen. Für die ersten und dritten Momente läßt sich der analytisch berechnete Ausdruck für relevante Netzgrößen nicht numerisch auswerten. Daher werden zwei Näherungen für diesen Ausdruck hergeleitet. Die erste Näherung beruht auf einer Beschreibung von Last und Generationen mittels Gauß-Verteilungen. Mit dieser Näherung können die ungerichteten Leistungsflussverteilungen in den Leitungen gut beschrieben werden. Eine Kombination der gaußschen Näherung mit der exakten Beschreibung der Momente erlaubt eine numerische Berechnung und verbessert die Qualität der Näherungen signifikant. Dieses Verfahren ist auf Grund der kurzreichweitigen Korrelationen im Allgemeinen nicht anwendbar, wenn das "Minimale

Dissipation"-Ausgleichsverfahren benutzt wird. Für den Fall, daß Last und Erzeugung im Mittel übereinstimmen, wird gezeigt, daß sich das qualitative Verhalten der mit den beiden vorgestellten Ausgleichsverfahren ermittelten Flüsse ähnelt. Die analytischen Näherungen beschreiben die Leistungsflussverteilungen, die in den Simulationen gefunden wurden, in diesem Fall in sehr guter Näherung.

Basierend auf den analytischen Näherungen der Leistungsflüsse lassen sich Leitungskapazitäten zuweisen, so daß diese mit gegebener Wahrscheinlichkeit nicht überlastet werden. Eine Überprüfung dieser Kapazitäten mit Hilfe der Simulation zeigt, daß die zugewiesenen Kapazitäten eine Überlastung durch fluktuierende Erzeugung mit der gewünschten Wahrscheinlichkeit verhindern. Dies erlaubt im Rahmen des Modells die benötigten Kapazitäten, nur basierend auf der Netzwerkstruktur und den Mittelwerten sowie Varianzen der Erzeugung und Last an den Netzwerkknoten zu berechnen, ohne auf aufwendige Simulationen zurückgreifen zu müssen.

Am Ende dieser Arbeit werden, basierend auf den Datenzeitreihen, die Leistungsflüsse berechnet und mit den Modellergebnissen verglichen. Wir finden, daß einige ausgewählte Leitungen, beispielsweise die Verbindung von Spanien nach Frankreich, in der Spitze Leistungen bis zu 100 GW transportieren müßten. Auch gibt es eine generelle Tendenz, daß tagsüber Leitungsflüsse von Süden nach Norden führen und dies nachts umgekehrt ist. Überraschenderweise sinkt im Falle einer Überproduktion der benötigte Transport nicht oder nur minimal. Zur Verringerung der benötigten Transportkapazitäten müssen die Erzeugungskapazitäten daher homogener verteilt werden, oder die Leistungsflüsse durch lokale Speicher verstetigt werden, so daß die Leistungsfluss-Spitzen weniger ausgeprägt sind. Der Vergleich mit den Ergebnissen der Analyse des Modells zeigt einige Übereinstimmungen. Die Leistungsflussverteilungen lassen sich größtenteils sehr gut durch generalisierte Gammaverteilungen beschreiben. Auch sind die analytischen Näherungen im Mittel gut, allerdings sind die Abweichungen für einige Leitungen groß.

Zukünftige Arbeiten können sich beispielsweise der Quantifizierung der Effekte von gemischtem Betrieb von Speichern und Ausgleichsgeneratoren oder den Auswirkungen einer homogeneren Verteilung der Erzeugungskapazitäten auf den Transport widmen. Auch die analytisch hergeleiteten Näherungen können noch an die spezifischen Eigenschaften der Erzeugung der Daten angepaßt werden, um eine bessere Vorhersage der resultierenden Leistungsflüsse zu ermöglichen.

In der vorliegenden Arbeit konnten Rahmenbedingungen für den zuverlässigen Betrieb von Energieversorgungssystemen mit hohem Anteil erneuerbarer Energien hergeleitet werden. Zudem konnten Transportflüsse in solchen Energieversorgungssystemen in einem Modell in sehr guter Näherung beschrieben werden. Diese Näherungen beschreiben auch Transportflüsse, basierend auf echten Daten, im Mittel gut. Damit bietet diese Arbeit einen Beitrag an, für ein besseres Verständnis von Transportflüssen in Übertragungsnetzwerken mit einem hohen Anteil von fluktuierender erneuerbarer Energieerzeugung.

Contents

1. Introduction

Renewable energy resources have become increasingly important in the last decades [29, 30, 50, 74, 98, 99, 118, 120]. Especially for wind and solar power generation the potential and growth is expected to be high [98]. Renewable energy generation can solve the main problems associated with conventional fossil and nuclear power generation: Limited resources and waste. When employing a high share of renewable energy sources, new challenges arise and the general conditions for reliable and stable power systems as well as the resulting power transmission requirements need to be studied.

In this work, I focus on electrical power systems. Historically, the ability to transport electricity over long distances was a breakthrough. Transmission of electrical power was the dawn of a new era as it freed power from geographycal constraints. Before, only cities and towns that were close to energy sources such as coal, wood, and hydropower could become industrial centers. With the ability to transport electrical energy, all regions potentially gained access to energy sources [83]. The feasibilty of generating power in one location and transmitting it to consumers over long distances was shown at the International Electro-Technical Exhibition of 1891 in Frankfurt/Main. The enthusiasm at that time can be seen in the official poster in Figure 1.1. The power to illuminate a display of a thousand incandescent lamps was generated at Lauffen am Neckar and transported over 175 km to Frankfurt using three-phase alternating currents at $15kV$. This successfull demonstration ended the discussion on the most economical means of transmitting electrical energy [83].

Figure 1.1.: *The official poster for the International Electro-Technical Exhibition of 1891.*

Power grids evolved to highly interdependent complex systems [5], consisting of various turbines, generators, transformers, high voltage cables, circuit breakers et cetera [107]. Recent advances of power electronics allow for more versatile components that

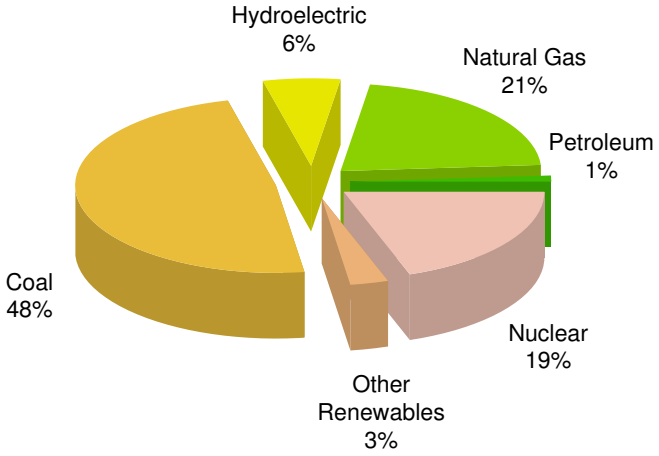


Figure 1.2.: *US electrical power generation by source in the year 2008 [119].*

facilitate control and increase the power transfer capabilities of the power transmission network [25]. However, the system has to be monitored in order to detect faults and to ensure reliable, save, and economic operation with sufficient contingency reserves [126].

Today, electrical power generation is dominated by fossil and nuclear fuels, see e.g. the shares of different energy source in the U.S. for 2008 in Figure 1.2. Existing infrastructure and control algorithms are therefore adjusted to the characteristics of these generators. However, as all resources are limited, the usage of fossil and nuclear energy sources is not sustainable in the future. The amount of remaining resources and reserves is under debate, the important fossil fuels can be expected to be depleted within the next hundred or maximally two hundred years [see e.g. 29, 69, 76, 129, 130]. Negative consequences, such as higher prices for scarce resources or environmental damages when exploiting sites that are difficult of access, can be expected to set in much earlier. Coal, for example, was the source of 48% of the U.S. electrical power in 2008 (see Figure 1.2) but is estimated to reach peak production around 2025 [130]. In Figure 1.3, the uranium resources according to Zittel and Schindler [129] are shown. Although significant stocks from different sources, e.g. from the conversion of nuclear weapons, exist, the proven resources together with the existing stock are projected to be exhausted within the next 30 years. Possible resources including all estimated discovered resources with extraction costs of up to 130 $/kg will be exhausted within around 70 years while predictions on further undiscovered resources are highly speculative [129].

Fossile and nuclear fuels also cause severe waste problems. Despite reprocessing technologies, the current nuclear technology produces waste that may remain highly radioactive for a million years [101]. More than a quarter-million tons of commercial high-level nuclear waste is in the need of disposal worldwide and no acceptable solution for long time disposal has been implemented yet [60, 101]. There is accumulating evidence for the driving role of green house gases, emitted when burning fossil fuels, in the increase of the average temperature of Earth's near-surface air since the mid-20th

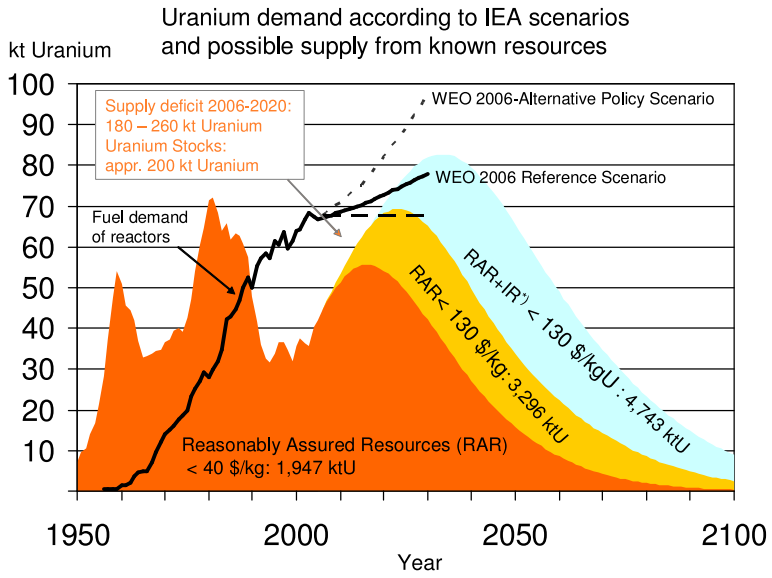


Figure 1.3.: *Uranium supply and demand taken from Zittel and Schindler [129] based on the World Energy Outlook 2006 report of the International Energy Agency (IEA). RAR denotes reasonably assured resources and IR the inferred resources. The fuel demand of reactors currently operating is shown by the black line. Future demands are estimated based on the scenarios in the World Energy Outlook (WEO 2006) of the International Energy Agency. The red area indicates the uranium already mined and the forcast of reasonably assured resources at a price below 40$ per kg Uranium. Between 40 $ per kilogram and 130$ per kilogram, there are higher reasonably assured resources indicated by the orange area. Estimates of undiscovered uranium sites are shown in blue.*

century [see e.g. 84, 110]. The possible consequences of this effect, well-known as "global warming", are very serious, so that there is a common agreement that greenhouse gas emissions are to be avoided and reduced in the future [29, 30, 50, 74].

Limited resources or the waste problems do not occur when employing renewable energy sources. The most important renewable sources in terms of the total available power are wind, solar, geothermal, biomass and hydroelectric generation [29, 99]. While wind and solar power are capable to meet future energy demands [29, 30, 50, 74], the use of other renewable energy sources can supplement wind and solar power generation or take advantage of local conditions, e.g. geothermal power in Iceland. However, as the energy delivered by wind and solar irradiation is driven by weather conditions, it is highly intermittent [72, 74]. Due to specific spatial features, e.g. good conditions for solar energy in the south and high wind speeds close to oceans, the resulting transport requirements for energy will be high [30, 50].

Consequently, understanding transmission in power systems is an important topic. Not only work with respect to detailed problems of technical implemention, but also a good theoretical knowledge of the whole system is important. As power transmission

systems are crucial infrastructure networks, they are also interesting from the point of view of statistical physics of complex networks [5]. They allow to refine models describing transport on networks and are one of the many real world networks, where the relation between topology and function is studied [5, 9, 43, 75]. In many fields of science, networks and their properties play an important role. Approaches from complex network theory are used, in e.g. neuroscience [see e.g. 24, 116], gene expression or coexpression analysis [see e.g. 106, 114], for chemical applications [see e.g. 127], or the study of various transport networks as well as synchronization properties [see e.g. 6, 10] and social networks [see e.g. 3, 123]. Because of their importance, many studies were devoted to general transport infrastructure networks [see e.g. 3, 58, 75, 85]. Especially the vulnerability to failures and attacks was assessed. With respect to power grids, understanding blackouts dynamics is an important aim [9, 26, 27, 43].

1.1. Objective and Methods of this Work

The question, to what extent a fully renewable power system is generally feasible, has to be addressed. Fluctuating wind and solar energy generation requires storage and transport that allow for a spatio-temporal compensation of excess power with deficits in generation. Additional backup balancing generators, based e.g. on hydroelectric power generation, might also be needed. The questions that arise are: How much storage and balancing is needed? How much wind, solar, hydro and geothermal power is good for Europe? Is there an optimal mix between them? Furthermore, the transport flows in such systems need to be quantified and understood. These questions have to be analyzed with respect to technological constraints and capacity limits.

In this work, the above questions are studied based on generation data of solar and wind power derived from metrological data and on empirical load data. For this purpose, we use a macroscopic approach to quantify the required storage and balancing energies, while for specific problems more details are included. For example, to calculate transport flows, the spatial arrangement of generation and consumption as well as the transmission network are considered.

To quantify power flows, we employ methods from complex network theory, engineering, and probability theory. To gain a deeper understanding of fluctuating flows in transport networks, a generic model is introduced that incorporates stylized facts found in the data. The model is evaluated using Monte-Carlo simulations and the statistics of the resulting transport flows are analyzed. Analytical approximations are derived using methods from probability theory and compared to the simulation results and the data. Due to the high complexity of power systems, all work is done in a quasi static framework, so that transient dynamics are neglected.

1.2. Overview

This thesis is organized as follows. A brief theoretical background on complex networks and simple transport models is given in Chapter 2 with a focus on the applicability to power systems. Based on a simple transport paradigm, a first approach to transport flows elicited by fluctuating sources is proposed.

For a better understanding of the characteristics of renewable power generation, we examine time series of generation and consumption data in Chapter 3. In particular, we analyze the Europe-wide required storage and balancing capacities for different scenarios. The results are compared to storage and balancing capacities that are potentially available in Europe and the feasibility of a power system with a high share of renewable energy generation is evaluated. This is followed by an examination of the properties of individual countries in Chapter 4 and an assessment of the advantages of European cooperation compared to a scenario without power exchange among countries.

We introduce a generic model in Chapter 5 that allows to gain a deeper understanding of transport flows in power systems within a simplified framework. The model incorporates stylized facts of the data from the preceding chapters. The results of Monte-Carlo type simulations of the model are discused and approximations to the power flow distributions analytically derived. The transport resulting from the weather-based generation data is analyzed in Chapter 6 and compared to the results obtained from the model.

Finally, a discussion of the results of this thesis together with an outlook of future work is presented in Chapter 7. Background information on power systems and power transmission are given in the Appendix.

2. Transport Flows on Networks

Statistical physics of complex networks offers a rich set of methods to analyze and describe various systems. Power grids can be understood as large, highly interdependent complex networks [5]. The main purpose of this section is to evaluate existing models on how well these can be used to understand power systems with a high share of fluctuating renewable power generation. As the impact of source fluctuations on the flow of the individual link was not been considered yet, an extension to existing models is proposed and analyzed.

In Section 2.1 the basic notions and methods of graph theory and complex networks are introduced. Structual properties of power grids are discussed in Section 2.1.3.

Motivated by modern infrastructures and their cascading failures, much work has been devoted to transport in complex networks. Models using transport along the shortest path from the sending to the receiving vertex are introduced in Section 2.2.1. Blackouts in power systems, where small disturbances may lead to a failure of large parts of the system, inspired researchers to consider cascading failures. Studies of cascading failures based on shortest path transport are also introduced in Section 2.2.1. Models based on approximations to the physical power flow equations of power systems can be found in Section 2.2.2. Fluctuating sources are a central point in this study and existing work on that topic is presented in Section 2.2.3. The results from literature are discussed in Section 2.2.4.

Previous work did not consider fluctuations in the flows of the transported quantities. Cascading failures can be caused by external fluctuations and may not only be important for power systems but also for other applications, like internet traffic for example. We introduce a new approach based on shortest path transport in Section 2.3 and derive and analyze a new capacity layout that is, with a given probability, stable with respect to overloads caused by fluctuations. A summary of this chapter can be found in Section 2.4.

2.1. Statistical Physics of Complex Networks

"Statistical physics [...] consists in the study of the special laws which govern the behavior and properties of macroscopic bodies (that is, bodies formed of a very large number of individual particles, such as atoms and molecules). To a considerable extent the general character of these laws does not depend on the mechanics (classical or quantum) which describes the motion of the individual particles in a body" [Landau

and Lifshitz, 79]. Typically systems with a large number of elements are analyzed using methods of probability theory and statistics. Many real world systems can be described by or reduced to networks. The aim is to describe systems as diverse as physical or engineered networks, information networks, biological networks, cognitive, semantic, and social networks within a unified framework. Statistical physics of complex networks aims to find universal behavior and general laws which are independent of the microscopic properties of the system. It allows for a simplified approach to study complex systems, whose complexity makes an analysis considering all details infeasible.

Starting with work from Watts and Strogatz [123] and Barabási and Albert [14] methods and tools from statistical physics were applied to complex networks. Interesting connections can be found. For example, consider the probability that a random walker, starting at point i, will reach point s before point t on a finite grid or network. It is equal to the voltage of the vertex i in a network of uniform resistors[1], if the voltage at s is set to one and set to zero at t [47]. If the observed network is not random, organizational principles of the real world system should be reflected and encoded in its topology [3]. This approach provides a new way to gain insights on complex systems and their dynamics. In various fields of science where networks and their properties play an important role, complex networks theory was applied, e.g. neuroscience [see e.g. 24, 116], gene expression or coexpression analysis [see e.g. 106, 114], for chemical applications [see e.g. 127] or the study of various transport networks as well as social networks [see e.g. 3, 123] and synchronization properties [see e.g. 6, 10].

In this section an introduction to graph theory and complex networks is given, to the basic notions in Section 2.1.1 and to fundamental findings in Section 2.1.2. The properties of power grids in terms of network theory found in the literature are sketched in Section 2.1.3.

2.1.1. Basic Notions

A network or graph $G = (\mathcal{V}, \mathcal{E})$ consists of a set of vertices $\mathcal{V}$ that are connected by a set of edges $\mathcal{E}$. The set of vertices or nodes is defined by $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ with $N = N(G)$ elements. The number of elements of a set is also denoted as $|\cdot|$, e.g. $N = |\mathcal{V}|$. The set of edges or links $\mathcal{E} = \{e_{ij}, \ldots\}$ with $M = |\mathcal{E}|$ elements is a subset of the set $\mathcal{V}^2$ of unordered vertex pairs. An edge from vertex i to j exists if the pair i, j is an element of $\mathcal{E}$ and is written as e_{ij}. Edges are also denoted by an unique index k, $e_{ij} = e_k$.

Two vertices i, j are said to be neighbors or adjacent, if they are connected by an edge e_{ij} of G, which is equivalent to the condition $\{v_i, v_j\} \in \mathcal{E}$. The vertices v_i and v_j

[1] The probabilities for the next step of the walker have to be uniform over all possible next steps.

are said to be incident with edge e_{ij}. This allows to define the adjacency matrix

$$\mathbf{A}_{ij} = \begin{cases} 1 & v_i \text{ has a link to } v_j \\ 0 & \text{otherwise} \end{cases} , \tag{2.1}$$

and the incidence matrix, that maps from the space of edges $\mathcal{E}$ to the space of vertices $\mathcal{V}$,

$$\mathbf{K}_{ij} = \begin{cases} 1 & v_i \text{ is the initial vertex of edge } e_j \\ -1 & v_i \text{ is the terminal vertex of edge } e_j \\ 0 & \text{otherwise} \end{cases} , \tag{2.2}$$

where $\mathbf{A}$ is an $N \times N$ and $\mathbf{K}$ an $N \times M$ matrix. Both matrices play an important role in various fields and for various applications, see e.g. Section 5.1.1.1 and A.2.4. In this work, only undirected networks are considered, in the sense that from $\{v_i, v_j\} \in \mathcal{E}$ follows $\{v_j, v_i\} \in \mathcal{E}$. However, the direction of the link e_{ij} from v_i to v_j serves as a reference for the sign of the flow f_{ij} between these vertices. Positive f_{ij} indicate a flow from i to j, negative f_{ij} the opposite. This notion of direction is used also for the incidence matrix. The initial and terminal vertices of the links are assigned randomly.

For a given vertex v_i, the set of neighbors $\mathcal{N}_i$ is given by

$$\mathcal{N}_i = \{v_j \in \mathcal{V} : \{v_i, v_j\} \in \mathcal{E}\} = \{v_j \in \mathcal{V} : e_{ij} \in \mathcal{E}\} . \tag{2.3}$$

For undirected graphs, that are considered in this work, the number of neighbors of a vertex i is called its degree, $k_i = |\mathcal{N}_i|$. For directed graphs, the in and out degree typically differs. An important class of graphs are "regular graphs", where the degree $k_i = k$ is identical for all vertices. Random graphs, whose vertices are connected somehow randomly, are classified by the type of the degree distribution $p(k)$, that is the probability for a randomly chosen vertex to have degree k. This is discussed in more detail in Section 2.1.2.

Based on the incidence or the adjacency matrix, the Laplace matrix, also referred to as combinatorial Laplacian or Kirchhoff matrix in the literature [19], is defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{A} = \mathbf{K}\,\mathbf{K}^T . \tag{2.4}$$

where $\mathbf{D}_{ii} = k_i$ is an $N \times N$ diagonal matrix with the vertex degrees on the diagonal. $\mathbf{X}^T$ denotes the transposed matrix. The Laplacian can be interpreted as a discrete version of the Laplace operator[2] that allows to consider problems like diffusion and random walks on networks [89, 115]. In this work, the Laplace matrix will be used in the context of the Kirchhoff flow [89], see Section A.2.4.

[2] On a grid, the Laplace matrix can be obtained as a finite difference approximation of the Laplace operator using a stencil of size one [103] as the adjacency matrix only considers next neighbors, see Equation (2.4). This approximation yields an discretization error of the order $\mathcal{O}(h^2)$, where h is the grid spacing of the system [92]. For a random network, the vertices are interpreted as grid points and local connectivity of the vertex determines the finite difference approximation stencil at this grid point.

The clustering coefficient C_i of a vertex i denotes the relative number of edges between its neighbors to the maximally possible number. It is defined for each vertex i as [123]

$$C_i = \frac{1}{k_i(k_i-1)} \sum_{j \in \mathcal{N}_i} |(\mathcal{N}_i \cap \mathcal{N}_j)| = \frac{2T_i}{k_i(k_i-1)} \ , \tag{2.5}$$

where $\mathcal{N}_i \cap \mathcal{N}_j$ denotes the subgraph of the neighborhood shared by both vertices i and j. Each vertex, that is in the neighborhood of two adjacent vertices, forms a triangle. By symmetry, $\sum_{j \in \mathcal{N}_i} |(\mathcal{N}_i \cap \mathcal{N}_j)|$ is twice the number T_i of triangles passing through vertex i and this is normalized to the number of maximally possible $k(k-1)/2$ triangles to obtain the clustering coefficient. It is a local measure with values between zero and one by construction. The average clustering coefficient cc of the network is given by averaging over all C_i

$$cc = \langle C \rangle = \frac{1}{N} \sum_{i \in \mathcal{V}} C_i. \tag{2.6}$$

Graphs with a high average clustering coefficient consist of clusters of vertices with many edges among each other. A low average clustering coefficient can be found for more uniform networks.

A path $[i \to j]$ from vertex i to j is a subgraph consisting of a non recurring set of vertices $\mathcal{V}([i \to j]) = \{x_i, x_{p_1}, \ldots, x_{p_l} (= x_j)\}$ and the respective edges $\mathcal{E}([i \to j]) = \{e_{ip_1}, e_{p_1p_2}, \ldots, e_{p_{l-1}p_l}\}$ of G [19]. The number of edges in $\mathcal{E}([i \to j])$ defines the length $l = |\mathcal{E}([i \to j])|$ of the path. In general, there is more than one path from i to j. The shortest path refers to the path with the minimal length l and is denoted by $[i \to j]_s$. The shortest path may be degenerate in the sense that there are more paths with the same minimal length. In the literature different metrics to define the lengths in the network are used that lead to different definitions of the shortest paths and path lengths [see e.g. 104]. In the following, only the so-called hop-metric, where each edge has length one, is considered as described above. A different way to describe the shortest path is the path function,

$$\text{path}([j \to k]; i) = \begin{cases} \xi & i \in \mathcal{V}([j \to k]_s) \\ 0 & \text{else.} \end{cases} \ . \tag{2.7}$$

The value of ξ is one for the case of a non-degenerate path. Otherwise, the initial value of one is divided equally at each bifurcation point over all branches of the remaining path[3].

A graph is called connected, if for every pair of vertices i and j there exists a path $[i \to j]$. A subgraph that includes all vertices that are connected by a path is called a component[4]. For a connected graph the number of components is one. The distance d_{ij} between two vertices i and j is given by the length of the shortest path $[i \to j]_s$

[3] To the knowledge of the author, there are no approaches to further simplify the path function.
[4] In other words, it is the maximal connected subgraph [19].

and defined to be infinite for a pair of vertices which is not connected by a path. The average distance within connected graphs or components is calculated as the average over all pairs of vertices

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i,j \in \mathcal{V}_c, i \neq j} d_{ij} \ , \tag{2.8}$$

where $\mathcal{V}_c$ denotes the set of vertices of the graph or component.

In many publications [see e.g. 85, 122], the size of the largest component or giant component N_{gc} is used to quantify the effect of changes in the network, e.g. due to link or vertex removal. Starting from a connected graph, the quantity N_{gc}/N gives a normalized measure of the decomposition of the graph after the removal of vertices or links. This measure has the disadvantage, that the average distance of a graph may increase when deleting an edge, while N_{gc}/N might stay constant. A measure that is also sensitive to the distances is the efficiency of the network, defined as

$$E = \frac{1}{N(N-1)} \sum_{i,j \in \mathcal{V}, i \neq j} \frac{1}{d_{ij}} \ , \tag{2.9}$$

which is effectively the harmonic mean over all distances. A fully connected network, has efficiency one. Increasing the distances decreases the efficiency. A pair of vertices that is not connected is defined to have an infinite distance, so that the contribution of this pair to the efficiency is zero. To quantify changes, the relative efficiency, E/E_0 is considered, where E_0 is the efficiency of the initial network. It is one, if there are no changes in the network and zero if there are no paths between the vertices.

2.1.2. Complex Networks

Erdõs and Rény [49] presented the first very influential paper on random graphs [3]. Previous work focused on regular or fully connected graphs [19]. The network model analyzed by Erdõs and Rény (ER) starts with N vertices and each pair of vertices is connected by a link with probability p. The probability of a given degree at a given vertex is then distributed binomially [18] and the number of nodes with degree k follow as Poisson distribution in the large N limit [3]. Also for $N \to \infty$, the probability to find connected graphs, the probability distribution of the size of the giant component, and other measures were calculated analytically [3].

Starting with Watts and Strogatz [123], Faloutsos et al. [51], and Barabási and Albert [14], characteristic degree distributions were found in many real world systems. These degree distributions deviate significantly from the Poisson distribution. This is a strong hint that the structure is not random. Most real world networks exhibit a degree distribution with a power-law or exponential tail [see e.g. 88]. A review of studied networks was published by Albert and Barabási [3], that cover a wide range of different

disciplines and topics from movie actor collaboration networks and protein folding networks to Internet topologies.

Powerlaw, also called scalefree, degree distributions are given by

$$p(k) \propto k^{-\gamma} \ . \tag{2.10}$$

This distribution is discrete for vertex degrees and has usually a minimum and maximum degree or an exponential cutoff [7]. Empirical values of γ have been found to be typically between two and three, although occasionally γ may lie outside these bounds [3].

Besides characteristic degree distributions, two other properties of networks play a prominent role, the average distance and clustering [3]. Despite the large size of many real world networks, the average distance between all pairs of vertices is relatively small. Random graphs as proposed by Erdõs and Rény also exhibit this property. The clustering coefficient influences the finer structure of the network. A high clustering coefficient means that many cliques are found, subsets of vertices with relatively many links within them. The average distance for a given degree distribution can be expected to increase for increasing clustering coefficient, as the fact that there are many links within a clique means that there are few links connecting the different cliques. Surprisingly, many real world networks have a small average distance and a high clustering coefficient that is orders of magnitudes larger than the clustering coefficient of ER-random networks with the same size N and average degree $\langle k \rangle$ [3]. Networks with these features were called small world networks by Watts and Strogatz [123].

A prominent example of a network structure model is the "preferential attachment" (PA) model by Barabási and Albert [14]. It argues that the scalefree nature of many real networks is caused by two generic mechanisms, growth and preferential attachment. An initially small network is assumed to grow over time and new vertices are assumed to attach to vertices with a high degree with high probability. For the network of web pages, for example, these are plausible assumptions, as constantly new pages appear and they will link to highly known pages with already have a high degree. The probability Π that a new node will attach to a node i with degree k_i is modelled with $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$[14]. The networks generated applying these rules are scale-free with an exponent of $\gamma = 3$ and have a clustering that is significantly higher than found for ER random networks. The average distance in the network scales logarithmically with the size N of the network and is a bit higher than for ER random networks but still exhibits the "small world" property [3]. These findings are in good agreement with many scale-free networks and thus support the assumptions that growth together with preferential attachment plays an important role in the evolution of some networks. The concept of preferential attachment was refined in following studies, e.g. with respect to aging [7, 45]. Further models exist that aim to explain specific network topologies, e.g. the Internet topologies at different levels, as discussed in Krioukov et al. [78].

A different approach to network generation, is to generate networks with specific properties without implementing rules to explain specific observed structures. This

helps, for example, to study the dependence of network properties on specific parameter variations. The so called "configuration model" [see e.g. 8] is a simple method designed to generate networks with special properties, like arbitrary degree distributions or clustering coefficients. The basic idea is that, in a first step, each vertex gets a number of edges assigned that are not yet connected to other vertices, so-called stubs. The number of stubs is drawn from the desired degree distribution, that can be arbitrarily chosen. In a next step, pairs of vertices i, j with at least one stub are selected and a stub of each vertex replaced by a connecting edge. This is repeated until there are no stubs left. The pairs of vertices to be connected can be selected randomly or such that additional constraints are met. For example, the method proposed by Ángeles Serrano and Boguñá [8] allows to tune the clustering coefficient and the degree-degree distribution.

Scholz [105] proposes a network generation method, referred to as "geometric p-model" based on the configuration model that interpolates between random and geometric networks and thus allows to tune the clustering coefficient within a certain range. A geometric network is embedded in a two dimensional Euclidean space spatially close-by vertices are connected. For networks embedded in a two dimensional box a clustering coefficient around 0.6 is found independent of the size N of the network [39]. The network generation according to the "geometric p-model" is done with a simple method: Vertices not only get a number of stubs assigned, drawn from the desired degree distribution, but also a position on a two-dimensional plane. The position is drawn from a uniform distribution between zero and one, independly for both coordinates. When connecting a stub of vertex i, the vertex with the smallest Euclidean distance that has a free stub is chosen as the target vertex j with probability p_{cc}. With probability $1 - p_{cc}$ a random vertex with a free stub is chosen. Thus, a value of $p_{cc} = 0$ creates a random network with a vanishing clustering coefficient and when using $p_{cc} = 1$ a geometric network with a large clustering coefficient. The variation of p_{cc} allows for a smooth interpolation between the two types of networks and their respective clustering coefficients in the large N limit [105].

2.1.3. Power Grids and Graphs

We consider only the high voltage transmission level of power grids in this work. For a more detailed discussion, see Section A.1. When working on power transmission systems using graph theory, the first question is: What are the properties of power grid networks in terms of graph theory? The most important is the degree distribution besides other parameters like the number of nodes and links as well as clustering coefficients, degree-degree correlations and so on.

Since there are no data of power grids publicly available[5], we revert to parameters that can be found in the literature. Rosas-Casals et al. [102] reproduced the network data of the European power grid from a map that is available as image from the

[5]To the knowledge of the author.

European Network of Transmission System Operators for Electricity. The found degree distribution is close to an exponential distribution

$$p(k) = \gamma^{-1} \exp\left(-\frac{k}{\gamma}\right) \tag{2.11}$$

for $k \gtrsim 2$. The probability for degree one does not follow this distribution. The number of nodes is $N \approx 3000$, the average number of degrees over Europe $\langle k \rangle = 2.8$, and the parameter of the distribution $\gamma_{EU} = 1.8$ for all Europe. Some subnetworks are also given, e.g. for the UK $\gamma_{UK} = 0.91$ and for Portugal $\gamma_{PT} = 2.71$ is found.

These values are consistent with the findings by Albert et al. [5], who also find an exponential degree distribution with $\gamma = 2$ for the North American power grid, with $N = 14099$ and $M = 19657$. According to Holmgren [68], the Western U.S. power grid and the Nordic power grid also have a degree distribution close to an exponential distribution with an average degree $\langle k_{WUS} \rangle = 2.33$ and $\langle k_N \rangle = 2.67$, respectively [68].

The clustering coefficient of all mentioned power grids is significantly higher than the clustering coefficient of random networks [68, 102]. The same is true for the average path length. The degree-degree correlations of the European power grid are essentially zero [102].

The networks used when modelling power systems in Chapter 5 have an exponential degree distribution with $1 \lesssim \gamma \lesssim 3$, the bounds found for the subnetworks of the European power grid. The clustering coefficient is varied, since it was found to be large compared to random networks. As mentioned above, the number of vertices with degree one is significantly lower than the expected number from the degree distribution. To account for that, the probability for degree one vertices is set to a given fraction of the probability of the degree two vertices. The resulting degree distribution is renormalized accordingly. To generate an ensemble of graphs, the method introduced by Scholz [105] as described in Section 2.1.2 is used, which allows for an arbitrary degree distribution and clustering coefficient.

2.2. Review of Transport Network Models

As much of the transport infrastructure in modern societies is organized in complex networks, transport networks have been studied widely. As failures may have dramatic consequences, it is important to understand the properties and vulnerabilities of these systems. In this section some approaches are reviewed. In this context the load and capcities can be defined for links or vertices depending on which is the limiting factor in the system. For example, the bandwidth in the internet infrastructure is limited by the throughput of the routers [46].

A first approach was proposed by Albert et al. [4]. The resilience of a network to failures and attacks was quantified by analyzing the diameter of the network and, in

case the network fragments into more than one cluster, in terms of the size of the largest cluster and the average size of the isolated clusters. The tolerance of the network to failures was examined by removing vertices in a random fashion. To inflict the most damage, an intentional attack would damage the most connected nodes [4], thus an attack is simulated by removing specifically the vertices with the highest degree. A fraction f of links was removed in networks with scalefree and with exponential degree distributions. For the exponential network the average distance increases slowly with f independent whether the removal was due to failure or attack. For scale-free networks a completely different behavior is found: On failures the distance does not change while for attacks the average distance increases with f. This increase is stronger than the increase for exponential networks. From the largest cluster and the average size of the isolated clusters, one can infer, that with increasing f the exponential network disintegrates into many components. The same is true when considering scale-free networks for the case of attacks, while random node removal disconnects only few small components, leaving the main network intact. These findings were confirmed analytically using percolation theory [32, 33]. Since the Internet has a scale-free topology, the authors conclude that it has a high robustness against failures whereas it is vulnerable to attacks against the most connected vertices.

For a better understanding, transport in networks was modelled in more detail. Many studies consider transport based on shortest path, as will be defined in Section 2.2.1, while in Section 2.2.2 models are introduced for the special case of power grids that consider the DC-flow, a transport paradigm approximating the full power flow equations. As our focus is on fluctuating flux due to intermittent renewable power generation, studies considering this topic are presented in Section 2.2.3. A discussion of the reviewed literature is done in Section 2.2.4.

All of these models consider a quasi-static case, where transient behavior is neglected and the assumption is made that load redistribution takes place on a much faster timescale than changes in the structure of the network.

2.2.1. Models Based on Shortest-Path Transport

Inspired by the betweenness centrality which was introduced to identify influential people in a social network [87], Goh et al. [58] propose a simple approach for the load of a vertex: Every vertex sends a packet of size one to every other vertex along the shortest paths. The load L_i of vertex i then follows as the number of paths that pass through it

$$L_i = \frac{1}{N(N-1)} \sum_{j,k \neq j} \mathrm{path}([j \to k]; i) \; . \tag{2.12}$$

The normalization by the number of pairs $N(N-1)$ yields a load of one for a vertex where all paths pass through. If there is no path between any two vertices, the path

function is zero for all vertices. Using this approach, Goh et al. [58] found a scaling law of the load distribution of all vertices $P_L(l) \propto l^{-\delta}$ with $\delta \approx 2.2$ when considering scale-free networks with γ between two and three. They propose that this scaling is universal. For $\gamma > 3$, the exponent δ increases and the distribution eventually becomes an exponential distribution for $\gamma \to \infty$.

Albert et al. [5] used the load as introduced by Goh et al. [58], to study power grids. The work is based on the network of the North American high voltage power grid. The properties of the real power grid were reflected in the analysis by differentiating between generators, transmission substations and distribution substations that connect to the low voltage distribution grid. The stability is measured in terms of the connectivity loss C_L defined as the fraction of generators not connected to a distribution substation. The first finding is that the grid is stable to a removal of generators due to redundancy at the power generation level. This is true for both, a random as well as degree based removal. However, degree and load based removal of transmission substations has a large effect, for example the connectivity loss C_L goes up to 60% if only 6% of transmission vertices are removed. Load based removal is defined so that the load is calculated for the intact network and the transmission substations with the highest load are taken out of the network, according to the initial calculation.

Inspired by the fact that in the three decades before 2000, blackouts have been happening in the U.S. electric transmission grid on average every 13 days [28], so called cascading failures came into the focus of research. A typical phenomenon of these blackouts is that small disturbances, leading to failures in one component, can trigger a cascade of further failures due to the subsequent redistribution of the load [85]. In the approach of Albert et al. [5], cascading failures were modelled by recalculating the load after every removal of a transport substation. The load based removal can be understood as a simultaneous failure of a fraction of the transport substations. Removing each vertex individually, recalculating the load and selecting the next vertex with the highest new load for removal, until the desired fraction of nodes is removed, mimics the effect of a cascading failure. Using this approach, a high instability of the network was found: Removing 6% of the transport substations cause a connectivity loss C_L of over 90% [5].

An important effect of cascading failures is that a small failure can lead to a breakdown of arbitrary large parts of the network. This not be captured by the cascading failure approach by Albert et al. [5], as the number of failing vertices is given a priori in the model. Motter and Lai [85] introduced capacities for each vertex. Using the same definition for the load L_j as in Equation (2.12), the capacities

$$C_j = (1 + \alpha)\, L_j \,, \qquad j = 1, 2, \ldots, N \tag{2.13}$$

are assigned to each vertex, with $\alpha \geq 0$. A vertex is assumed to fail if the load is bigger than its capacity, $L_j > C_j$, and thus removed from the network. Based on the new shortest paths, the recalculated load might cause further failures. This scheme is repeated until no further vertices fail, thus implementing a cascading failure with a size that depends on the assigned capacities and the properties of the network. Interesting

to note is that a loss of parts of the network reduces the total load in the network, as disconnected vertices can not cause any load on other vertices, see Equation (2.12).

Different network structures were tested by Motter and Lai [85] with respect to the effect of a cascading failure in response to the removal of a node, either randomly, with the highest degree, or with the highest load. The stability was measured in terms of the size of the largest component after a removal N' relative to its initial size N by $G = \frac{N'}{N}$. For a regular graph with $k_i = \langle k \rangle = 3$, the network disintegrates for $\alpha = 0$ but no cascading failures occur for slightly larger α. Scale-free networks are stable only for random vertex removal with $\alpha > 0$. When removing high degree vertices, G goes down to around 0.2 for $\alpha \to 0$ and only slightly increases for higher α values. The same is found when removing high load vertices [85].

These results were analyzed further by Zhao et al. [128]. The connectivity loss C_L due to cascading failures was characterized as a phase transition between the disintegrated network and an intact network as a function of the tolerance parameter α. They present an analytical approach to approximate the critical α_c which quantifies the onset of the phase transition.

To allow for a stable operation of transmission systems, Wang and Kim [122], for example, modified the capacity layout in Equation (2.13), such that mainly vertices with a high load get an additional capacity. They show that this approach makes the network more robust while reducing the cost. A different approach is proposed by Schäfer et al. [104]. They prove that an intelligent routing, which distributes the load more homogeneous over the network, can reduce the required capacities and allows for a much higher stability to cascading failures for a given investment compared to the routing along the shortest hop paths.

2.2.2. DC-Flow Models

Modern power grids use alternating currents (AC) to transport energy from generators to consumers, for more details consult Sections A.1 and A.2. This is mainly due to the fact that in AC-systems the voltage can be transformed, allowing high voltage transmission to reduce losses, up to 765 kV are in operation [107], and lower voltages for distribution to increase security, see Section A.2.5. While for direct current (DC) systems capacitances and inductances play only a role in transient responses to changes, these have to be considered for AC-systems. As discussed in detail in Sections A.2.1 and A.2.2, for an exact treatment of the power flows not only the active power has to be taken into account but also reactive power.

The power flow equations are second order in the voltage. To solve these nonlinear equations, well established methods exist in the literature from electro-technical [see e.g. 90, 107, 126] or mathematical theory [see e.g. 23, 103]. However two important approximations to the network flows exist [see e.g. 90, 126], the decoupled flow and

the DC flow approximation. Based on empirical observations the power flow equations can be linearized such that the active and reactive power are described by independent equations, thus referred to as decoupled power flow. Assuming a homogeneous voltage across the network allows to neglect the reactive power flow, so that the power flow is approximated by a linear equation consisting only of the active power, referred to as "DC flow"[6]. It is important to notice that, in this power flow approximation only reactances influence the flow as empirically the resistances are found to be small compared to the reactances of the lines and thus neglected. This implies that no losses are considered.

The structure of the network, together with the reactances of the links, are described by the matrix $\mathbf{B}'$. This matrix resembles the Laplace matrix in Equation (2.4), as discussed in Section A.2.4. For a formulation of the full power flow equations, currents and voltages in AC-systems are described by their absolute values and phases, see Section A.2.1. Due to the approximations, only the phases of the voltages at the vertices $\vec{\delta}$ are relevant and the DC power flow is given by

$$\vec{P} = \mathbf{B}'\vec{\delta}\,. \tag{2.14}$$

Positive P_i denotes active power injected at vertex i and negative values the power that is withdrawn. The in and out flow of the network has to be equal for the energy in the system to be conserved. Thus, the condition $\sum_i P_i = 0$ has to be ensured, referred to as balancing or matching generation and consumption. The DC flow equations are modelled on the exact power flow equations, so they can expected to be a better description of flows in power grids that shortest path flows. The exact derivation together with the details on the approximations are discussed in Section A.2.3.

In a series of studies [26, 27, 43], the dynamics of blackouts in power grids was studied, both by analyzing blackout time series from 1984 to 1998 and by reproducing the findings with a simple model. After analyzing the probability distributions and the long time correlations, the data of the blackout size was found to be consistent with avalanche sizes in a running sandpile and the authors conclude that this is an evidence of self organized criticality in the blackout dynamics [27]. The model, they propose, is specifically tuned to reproduce the avalanche behavior. Initial capacities are assigned to generators and links and these grow on a slow, daily time scale. Links also get capacities assigned, but these are only increased after a failure. Balancing is done using a linear program with respect to the assigned capacities and to a cost function that discourages load shedding. Using this approach, avalanche behavior found in power grids has been reproduced. Anghel et al. [9] extended this study with a more detailed modelling of line failures as well as a utility response model and obtained first results on evaluating repair strategies.

An approach that is capable of reproducing the dynamics of power systems is proposed by Filatrella et al. [53]. Large generators and loads are typically synchronous machines

[6]Since this approximation is computationally cheap compared to the full flow equations, it is also referred as fast power flow [126].

in which the constantly magnetized rotor is synchronous to the magnetic field of the stator and thus synchronous to the transmission network. The power balance equation, considering the power stored in the rotating mass P_i^{acc}, the losses due to dissipation P_i^{diss}, the power transmitted P_i^{trans}, and the power injected or withdrawn P_i at vertex i, can be written as

$$P_i = P_i^{diss} + P_i^{acc} + P_i^{trans} \ . \tag{2.15}$$

A second order diffenerential equation for the phase at each generator can be derived,

$$\ddot{\theta}_i = P - \alpha_i \, \dot{\theta}_i + P_{max} \sum_{j \in \mathcal{N}(i)} \sin(\theta_j - \theta_i) \tag{2.16}$$

with $\alpha = \frac{2K_D\Omega}{I}$ the ratio of the dissipation rate K_D and the synchronous frequency Ω to the inertia I of the generator. P_{max} denotes the maximal power that can be transferred over a link. The dynamic stability with respect to synchronization in response to faults was analyzed and a qualitative agreement to empirical observations was found. Equation (2.16) also can be interpreted as a network of coupled Kuramoto oscillators with an additional inertia term. The Kuramoto model is one of the most important models of coupled phase oscillators that is used to study synchronization phenomena in physical, biological, chemical, and social systems [see e.g. 1]. In this approach transmission losses are neglected, so that the DC-flow approximation is the steady state solution of Equation (2.16) when neglecting the power consumed by dissipation at the synchronous frequency Ω. This can be shown using linear stability analysis [113].

2.2.3. Fluctuations

Few studies considered fluctuations in network transport. An importantant work was done by Barabási et al. [15]. Analyzing fluctuating fluxes in different networks, they found a characteristic scaling of the standard deviation σ of the flow in dependence of the mean $\langle f \rangle$,

$$\sigma \propto \langle f \rangle^{\alpha} \ , \tag{2.17}$$

with exponents either in the vicinity of $\alpha \simeq 1$ or $\alpha \simeq 0.5$. The flow analyzed for the daily traffic measurements of 374 geographically distinct Internet routers and currents in a microprocessor, yielded $\alpha = 0.5$. The value of $\alpha = 1$ was found for the number of visitors an Internet site receives, highway traffic measurements at 127 locations and the flow in rivers probed on 3495 locations on several U.S. rivers. All of these systems clearly have an underlying network on which the transport takes place. Barabási et al. [15] claimed that these scalings are universal with a scaling of $\alpha = 0.5$ for systems that are dominated by external fluctuations and $\alpha = 1$, if the behaviour of the system is dominated by its internal collective fluctuations.

Kim and Motter [75] studied the relation between capacities and load. Based on real data, they found for various networks[7], that only vertices with a high load L_i operate close to their capacity C_i. For vertices with a small load the capacity is significantly higher than the load. The strength of this effect varies for the different cases. The air transportation network has a load close to the capacity, while infrastructure networks, like highways or power grids, have a capacity much higher than the load, especially for vertices with a small load. Their empirical and theoretical analysis provides evidence that in all the systems analyzed, the determination of capacities results from a decentralized trade-off between cost and robustness. Robustness was tested with respect to fluctuations in the quantity of the transported goods. The optimal capacities were then derived with respect to an objective function consisting of the failure probability for a given capacity and the associated costs, that are assumed to be linear. A parameter allows to weight the importance of the two objectives. The optimization reveals that system-specific characteristics of the observed nonlinear behavior of the capacity - load relation are mainly determined by the weight given to the cost. An interesting remark is that the capacity allocation is rather independent of the network structure [75].

2.2.4. Discussion

Various studies of transport properties of complex networks have been presented in the previous sections, most of them analyze stability against failures or attacks. In the context of statistical physics of complex systems, a rich set of methods and characteristic features have been developed to analyze and describe transport networks.

However, there is also some criticism. Doyle et al. [46] argue that some of the claims do not hold when tested with real data or when examined by domain experts. For the case of the Internet, the theoretical results e.g. by Barabási and Albert [14] or Albert et al. [4] suggest a "robust yet fragile" nature with respect to vertex removal due to failures or attacks. This finding is neither supported by engineering data nor theory [46]. It is argued, that the topology of the Internet at router level is severly restraint by physical and economical constraints. For these networks no central high degree hubs are found whose failure leads to breakdown of the network into small clusters, as argued by Albert et al. [4]. Further, the Internet has a layered architecture that uses feedback control to enable a robust performance in an heterogeneous environment that experiences many disruptions. While the lowest layers have strong physical and economic constraints, higher layers define their own, often unique network topologies. They make the Internet robust to perturbations that they are designed for, but may be vulnerable to other perturbations. According to Doyle et al. [46], this is the real "robust yet fragile" property of the Internet. In this work, we try to avoid oversimplification when analyzing the power transmission system by a careful analysis of generation data that is presented in Chapters 3 and 4.

[7] The data was derived from the U.S. and international air transportation network, airplane movements on the network of routes between airports, highways, power grids, and Internet routers.

To calculate power flows in transmission networks, two transport paradigms are used in the presented literature, shortest path transport and the DC-flow approximation. While the latter is directly derived from the physical flow equations, the quality of the approximation using a betweenness centrality based load that uses shortest path transport needs to be addressed. Kim and Motter [75] show that the correlation between the empirical load and the betweenness centrality that is based on shortest paths, see Equation (2.12), is not meaningfully strong when considering power grids. This suggest that shortest path transport is not a good transport paradigm to analyze power grids.

Most presented studies only considered quasi-static states, neglecting all transient effects. Using a conceptually simple dynamical phenomenological flow model that captures stylized facts of electrical networks, Simonsen et al. [109] find that the dynamics can play an important role for the network robustness. The capacity layout as introduced by Motter and Lai [85] was used and a link defined to fail if the load is bigger than the capacity for at least the period τ. The robustness of the network exhibits a strong dependence on τ. The static case shows a much higher robustness compared to the case in which links immediately fail, $\tau = 0$. The two cases consititute the upper and lower bound on the robustness. This illustrates the importance the network dynamics may have.

2.3. Fluctuations using Shortest Path Flow

The impact of fluctuating fluxes to cascading failures in transport networks was not studied yet. As will be shown in Chapter 3, fluctuations are important when considering power systems with a high share of renewable energy so there is a strong need to study the impact of fluctuating fluxes on the robustness of networks. Vertices may fail either directly due to an accumulation of extreme flux or due to a subsequent overload cascade. Immediate questions that arise are: How does an efficient capacity layout look like, which is able to cope with the fluctuating fluxes? Given various classes of fluctuating fluxes, how do they determine the resulting fluctuations in the accumulated vertex loads? Are there correlations between the accumulated loads of different vertices and how do they look like?

This study is based on the approach of Motter and Lai [85] which uses the shortest path transport paradigm. It is not directly applicable to power systems but more to get first insights into the effects of fluctuating fluxes within a well studied framework [see e.g. 58, 67, 85, 87, 104], see also Section 2.2.1. The results presented in the following are published in Heide et al. [63].

As in Section 2.2, every vertex i of the network $G{=}(\mathcal{V}, \mathcal{E})$, sends a unit flux $s_{ij} = 1$ to every other vertex $f{\neq}i$ along the shortest-hop paths $[i{\to}f]_s$. This results in an

accumulated vertex load

$$L_v = \frac{1}{N(N-1)} \sum_{i,j \in \mathcal{V}} \mathrm{path}([i \to j]; v) s_{ij} \ . \tag{2.18}$$

The value of the path function is either 1 or 0, depending on whether the vertex v is part of the shortest-hop path or not. In case of shortest-path degeneracy, the value of the path function is reduced by a factor depending on the degrees and depths of the respective branching points.

Based on the load in Equation (2.18), the capacities

$$C_v(\alpha) = (1+\alpha) \langle L_v \rangle \ , \tag{2.19}$$

where $\langle X \rangle$ denotes the average over X, are assigned to the vertices. If for some reason one or more vertices fail, a network-wide redistribution of the loads in Equation (2.18) occurs due to a modification of the shortest paths. The new load L_v of vertex v may become larger than its capacity C_v and subsequent failures can occur that trigger a cascading failure.

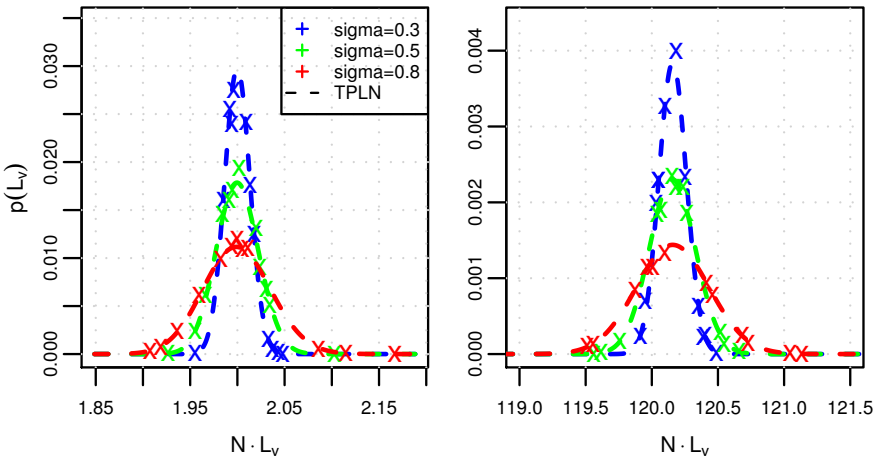
2.3.1. Load Fluctuations Resulting from Flux Fluctuations

In previous studies s_{ij} was always assumed to be constant and uniform. Here, flux fluctuations are introduced into the modeling of the load in Equation (2.18) by varying the strengths s_{ij} according to some distribution. For demonstration, we pick a lognormal distribution with mean $\langle s \rangle = 1$, on average the results of [85] should be found. The fluctuation strength is defined as its standard deviation $\sigma = \sqrt{\langle (s-1)^2 \rangle}$. We distinguish two fluctuation scenarios. The first is denoted as path-like, where all s_{ij} are drawn independently from each other. For the second, which we denote as source-like, all $s_{ij} = s_i$ belonging to the same source vertex i are given the same value sampled from the lognormal distribution.

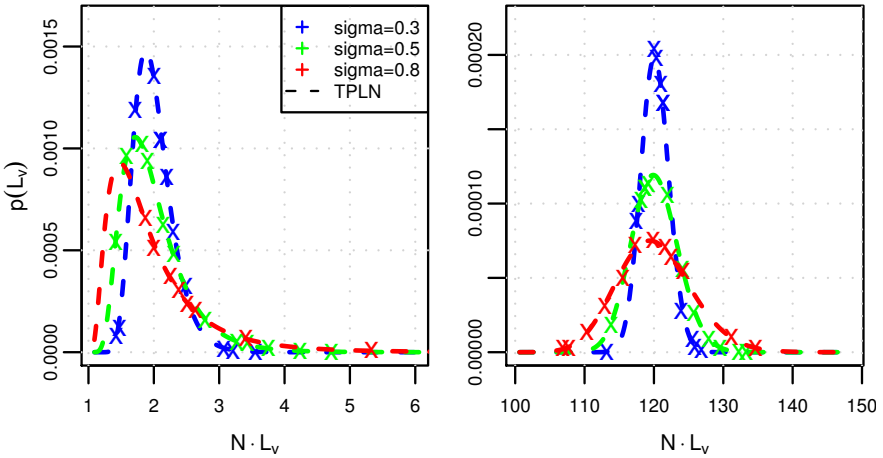
In order to develop a new capacity layout beyond the mean-flux case in Equation (2.19), a good understanding is needed on how the flux fluctuations determine the load distributions across the network. We begin by looking at the one-point distribution $p_v(L_v)$. Equation (2.18) can be read as a weighted sum of independently and identically distributed random fluxes s_{ij}. In case of log-normal fluxes with fluctuation strengths $\sigma \lesssim 1$, this sum can be approximated again by a lognormal distribution [100]. To allow for some more flexibility we choose a three-parameter generalization of the log-normal distribution

$$p_v(L_v) = \frac{1}{\sqrt{2\pi}\eta_v (L_v - \kappa_v)} \exp\left[-\frac{(\log[L_v - \kappa_v] - \mu_v)^2}{2\eta_v^2} \right] \tag{2.20}$$

for the description of the load distribution. The parameters expressed in terms of the first three central moments of the load at v, the mean m_v, variance σ_v and the third



(a) *Path-like fluctuations*



(b) *Source-like fluctuations*

Figure 2.1.: *Simulated load distributions $p(L_v)$ due to path-linke (a) and source-like (b) flux fluctuations for three fluctuation strengths. Two vertices with minimum (left) and maximum (right) load are depicted. One typical realization of a random scale-free network, for which the number of vertices, scale-free exponent and minimum degree have been set to N=1000, γ=2.5 and $k_{\min}$=2, has been used with 10^4 fluctuation realizations. The dashed curves correspond to the three-parameter log-normal distribution [TPLN, see Equation (2.20)] with predicted parameters as defined in Equation (2.24) and (2.25).*

central moment γ_v, follow as

$$\eta_v^2 = \log\left(b_v^{\frac{1}{3}} + b_v^{-\frac{1}{3}} - 1\right) = \log\left(d_v\right) \tag{2.21}$$

$$\mu_v = \frac{1}{2}\ln\left(\frac{\sigma_v^2}{d_v^2 - d_v}\right) \tag{2.22}$$

$$\kappa_v = m_v - \sigma_v\sqrt{\frac{1}{d_v - 1}}\,. \tag{2.23}$$

with $b_v = \frac{1}{2}\left(\gamma_v\sqrt{\gamma_v^2+4}+\gamma_v^2+2\right)$ and $d = b_v^{\frac{1}{3}} + b_v^{-\frac{1}{3}} - 1$. These equations are directly derived from the central moments of the three-parameter generalization of the log-normal distribution, as given by e.g. Cohen and Whitten [31].

To fit the load distribution at vertex v, the three parameters μ_v, η_v, κ_v are calculated from the first three cumulants of Equation (2.20) which have to be equal to the first three cumulants of the load in Equation (2.18). This is called method of moments in the literature [see e.g. 112]. These are for path-like fluxes

$$\langle L_v^n\rangle_c = \langle s^n\rangle_c \sum_{i\neq j=1}^{N}\left(\frac{\mathrm{path}([i\to j];v)}{N(N-1)}\right)^n , \tag{2.24}$$

and for source-like fluxes

$$\langle L_v^n\rangle_c = \langle s^n\rangle_c \sum_{i=1}^{N}\left(\sum_{j=1(j\neq i)}^{N}\frac{\mathrm{path}([i\to j];v)}{N(N-1)}\right)^n . \tag{2.25}$$

The parameters μ_v, η_v, κ_v in Equations (2.21) to (2.23) can be directly calculated by inserting the moments from Equation (2.24) or (2.25) for m_v, σ_v, and γ_v.

Figure 2.1 compares the predicted three-parameter log-normal distribution with simulated one-point load distributions, which have been sampled from a large number of independent flux fluctuation realizations on a typical random scale-free network. For path- as well as source-like flux fluctuations and for all vertices ranging from minimum to maximum average load, the analytical distributions fit the numerical data very well. – Note, that the one-point load distribution in Equation (2.20) should not be confused with the distribution $p(\langle L_v\rangle)$ of average loads across all vertices of the network. For the latter we reproduce the result $p(\langle L\rangle) \sim \langle L\rangle^{-\delta}$ with $\delta\approx 2.2$, which has been shown [58] to be universal for all scale-free networks with exponent $2<\gamma\leq 3$.

The $n = 2$ cumulants of Equations (2.24) and (2.25) are depicted in Figure 2.2 as a function of the averaged vertex loads. For path-like flux fluctuation, shown in the inset of Figure 2.2, a scaling relation of the type $\sqrt{\langle L_v^2\rangle_c} \sim \langle L_v\rangle^\beta$ is found with exponent $\beta = 0.5$. This dispersion relation has already been observed in [40] and related to internal collective dynamics on the network. It has not been clear whether the found value of the scaling exponent is universal or not. For source-like flux fluctuations no good overall scaling is observed. This is due to the fact that only the load distribution for high average loads are Gaussian shaped, whereas load distributions of vertices with small loads have a long tail that increases the variance; consult again Figure 2.1. The asymptotic high/low load regimes are in accordance with $\beta = 0.5$ and 0.1, respectively, indicating that the scaling exponent $\beta \approx 0.5$ is not universal.

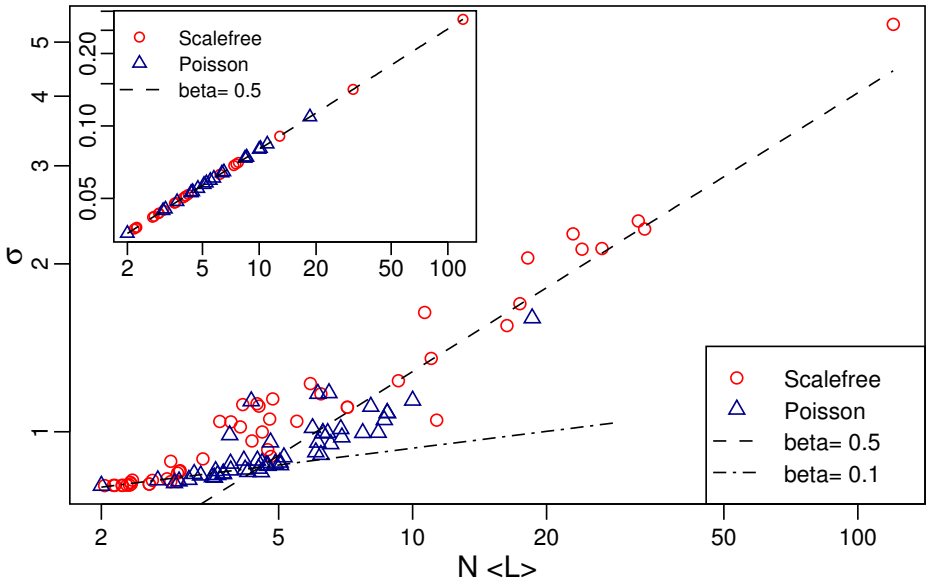


Figure 2.2.: *The moment* $\sigma(L_v) = \sqrt{\langle L_v^2 \rangle_c}$ *of Equations (2.24) and (2.25) as a function of the average vertex load* $\langle L_v \rangle$ *for one realization of path- (inset) and source-like fluctuations on random scale-free (circles) and Poisson (triangles) networks. Parameters are* $N = 1000$, $\gamma = 2.5$, $k_{\min} = 2$ *for the scale-free networks, and* $N = 1000$, $\langle k \rangle = 5$ *for the Poisson networks. 50 out of* 10^3 *vertices are shown. The dashed and dash-dotted straight lines represent the scaling exponents* $\beta = 0.5$ *and* 0.1.

2.3.2. Capacity Layout and Failure Probabilities

The good agreement of the predicted three-parameter log-normal distributions with the vertex loads allows for a direct construction of a new capacity layout, which is robust against flux fluctuations up to some confidence level. For a single vertex the quantile

$$q = \int_0^{C_v} p_v(L_v) dL_v = F_v(C_v) \tag{2.26}$$

describes the confidence level that its load L_v remains smaller than its capacity C_v. Since $1 - q$ describes the probability that the vertex will fail due to direct overloading, a confidence level very close to one is desirable. A typical value in engineering is $q = 0.9999$. By presetting the confidence level to such a targeted value, the capacity $C_v(q) = F_v^{-1}(q)$ needed at the vertex is obtained from the inverse of the cumulative distribution function F_v. Since $p_v(L_v)$ is three-parameter log-normal, F_v can be expressed in terms of the inverse of the cumulative distribution function $\Phi(q)$ of a centered normal distribution with unit variance. This leads to the capacity assignment

$$C_v(q) = e^{\eta_v \Phi^{-1}(q) + \mu_v} + \kappa_v \ , \tag{2.27}$$

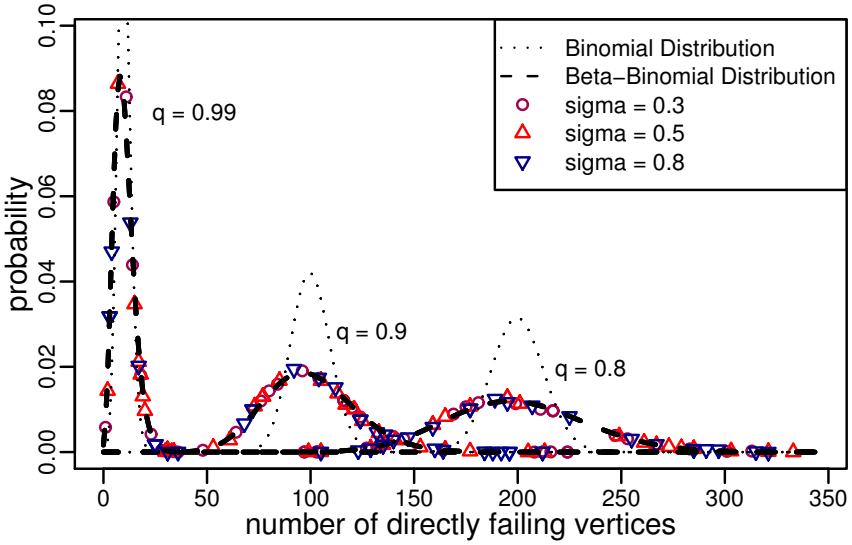


Figure 2.3.: *Probability distributions for the number of directly failing vertices due to source-like flux fluctuations in a scale-free network. Network parameters are as in the previous figures.* 10^4 *flux realizations have been used for various fluctuation strengths. The distributions only depend on the quantile, see Equation (2.26):* $q = 0.8$ *(right),* 0.9 *(centre),* 0.99 *(left). The dotted and dashed curves correspond to binomial and beta-binomial distributions with the same mean* $(1-q)N$*. The correlation parameter* ρ *of the beta-binomial distribution has been calculated with Equation (2.29) and (2.30).*

with μ_v, η_v, and κ_v obtained from Equations (2.20) to (2.23) by applying the method of moments. In principle, different q values could be assigned to different vertices, but for simplicity we chose the same q for all vertices.

By this construction, the distribution of the number of directly failing vertices M due to the fluctuating fluxes will have a mean of $\langle M \rangle = (1-q) \cdot N$. As can be seen in Figure 2.3, the actual number may deviate much from this mean. Note that the shown distributions only depend on the quantile q and not on the strength of the flux fluctuations. The distributions would be binomial if the direct failure of a vertex were independent of the other vertices, however, this is not the case. The probabilities of directly failing vertices are correlated since all vertices on a shortest path receive the same flux strength from the transmitting vertex.

A good approximation to the observed distributions is provided by the beta-binomial distribution

$$p(M; a, b) = \binom{N}{M} \frac{B(M+a, N-M+b)}{B(a,b)} , \tag{2.28}$$

where $B(\cdot,\cdot)$ is the Beta function. It is known to describe correlated Bernoulli random variables [66]. The two parameters a and b can be rewritten as the mean $\langle M \rangle / N = (1-q) = \frac{a}{a+b}$ and the correlation measure $\rho = \frac{1}{a+b+1}$. From best fits of Equation (2.28)

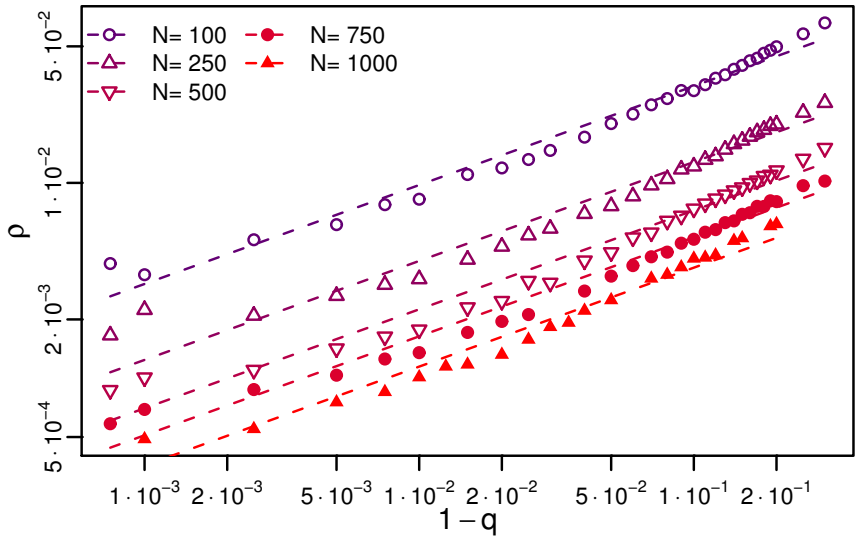


Figure 2.4.: *Relation between the parameters ρ and $(1-q)$ of the beta-binomial distributions, see Equation (2.28), which have been directly fitted to the sampled distributions of Figure 2.3.*

to the distributions of Figure 2.3, we find the empirical relation

$$\rho = \omega_N \, (1-q)^{\xi} \ , \tag{2.29}$$

see Figure 2.4. Within acceptable precision the exponent ξ turns out to be independent of the network size. For ω_N the N-dependence

$$\omega_N = \omega \, N^{-\nu} \ . \tag{2.30}$$

is found. The following table lists the fitted parameter values for path- and source-like flux fluctuations on scale-free ($\gamma = 2.5$, $k_{\min} = 2$) as well as Poisson networks ($\langle k \rangle$=5).

path	ω	ξ	ν
Poisson	1.50	0.53	0.80
scale-free	1.47	0.41	0.80

source	ω	ξ	ν
Poisson	5.40	0.70	0.80
scale-free	3.34	0.51	0.80

The good description by the beta-binomial distribution in Equation (2.28) allows to make an analytical prediction of the probability that with the capacity layout in Equation (2.27) no vertex of the network will fail due to flux fluctuations. This probability $p(M{=}0; 1{-}q, \rho)$ with ρ from Equations (2.29) and (2.30) is equal to the probability that the network efficiency [37], as introduced in Section 2.1.1,

$$E = \frac{1}{N(N-1)} \sum_{i \neq f \in \mathcal{V}} \frac{1}{d_{if}} \tag{2.31}$$

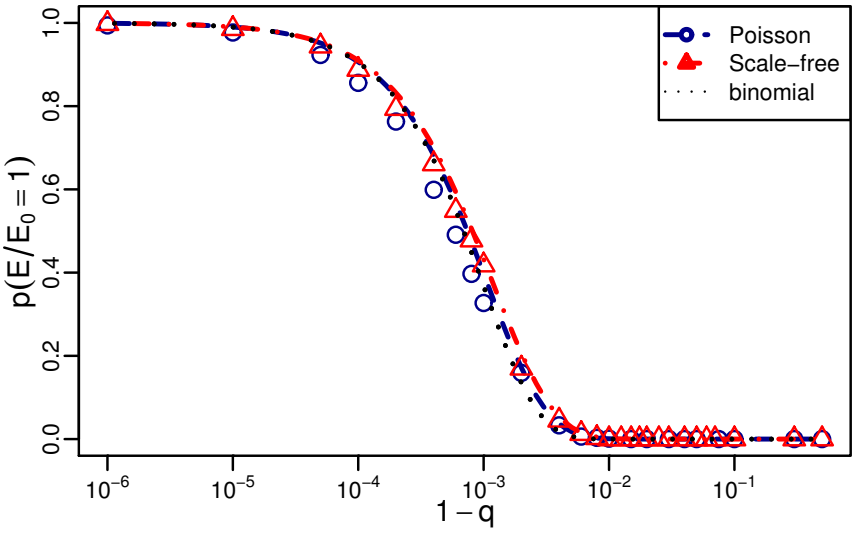


Figure 2.5.: *Probability for the relative efficiency to remain* $^{E}/_{E_0} = 1$ *given the capacity layout in Equation (2.27) against source-like flux fluctuations. The symbols represent the data from a simulation of* 10^3 *fluctuation realizations with strength* $\sigma = 0.5$ *on a representative random scale-free (triangles) and Poisson (circles) network of size* $N = 1000$ *(other parameters as stated before). The dashed and dotted curves represent the analytical prediction based on the beta-binomial distribution in Equation (2.28) and the binomial simplification, respectively.*

remains equal to its initial value E_0 of the intact network, thus $p\left(\frac{E}{E_0} = 1\right) = p(M = 0; 1 - q, \rho)$. The network efficiency represents a measure to evaluate the quality of a capacity layout. It includes direct as well as cascading failure of vertices. d_{if} is the shortest-hop distance between vertices i and f. Figure 2.5 compares the predicted $p(0; 1-q, \rho)$ with numerical data.

Since the correlation ρ of the Beta-binomial distribution goes to zero as q goes to one, the probability $p\left({}^{E}/_{E_0} = 1\right)$ can also be approximated using the binomial distribution $p_{bin}\left(M{=}0; 1{-}q\right)$; see again Figure 2.5. This gives a parameter free approximation to the probability that no vertex fails. The same relation also holds for path-like flux fluctuations, shown in Figure 2.6.

Finally, we compare the investment costs $I = \sum_{v \in \mathcal{V}} C_v = I(\alpha) = I(q)$ relative to $I_0 = \sum_{v \in \mathcal{V}} \langle L_v \rangle$ of the two capacity layouts in Equation (2.19) and Equation (2.27). These are functions of the tolerance parameter α and the quantile q, respectively. Figure 2.7 shows the efficiency as defined in Equation (2.31) of a scale-free network as a function of I. For source-like flux fluctuations the efficiency of Equation (2.27) remains close to zero up to a critical investment cost, only then to jump up and to overtake the efficiency of Equation (2.19). In the limit $E/E_0 \to 1$, the investment costs into the newly proposed capacity layout defined in Equation (2.27) are significantly smaller than for the standard layout in Equation (2.19). For path-like flux fluctuations both capacity

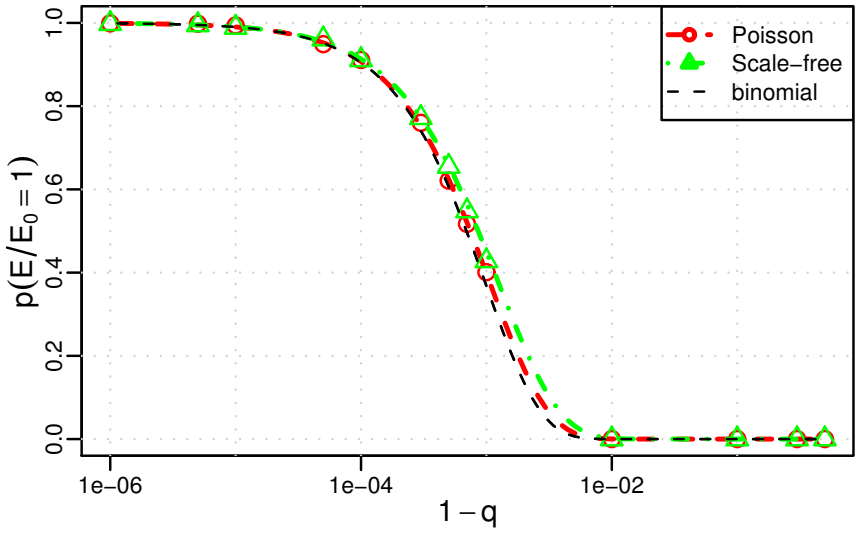


Figure 2.6.: *Same as Figure 2.5 but for path-like fluctuations.*

layouts reveal an abrupt transition from low to high efficiency at very low investment costs.

2.4. Discussion

Statistical physics of complex networks has developed a wide variety of methods and measures to quantify and analyze various systems. Transport systems were analyzed, mainly with respect to stability to link or vertex removal.

Based on the model proposed by Motter and Lai [85], we analyzed the impact of fluctuating fluxes on the robustness of networks. The cumulants of the flows over each link in the network are derived based on topological information and the cumulants of the source fluctuations. An approximation of the probability distributions of the flows is found using the method of moments. This allows to develop a new robust capacity layout. It is able to cope with the load fluctuations induced by flux fluctuations transported on the network. Within given confidence level it supports the network to operate at full efficiency and guarantees robustness against a cascading failure. We conclude that within simple framework, fluctuations have a strong impact on the efficiency of the network. A good understanding allows the construction of a capacity layout efficient with good robustness at low cost.

In the next sections, renewable energy generation based on metrological data are studied as a base for further modelling of power grids and renewable energy generation.

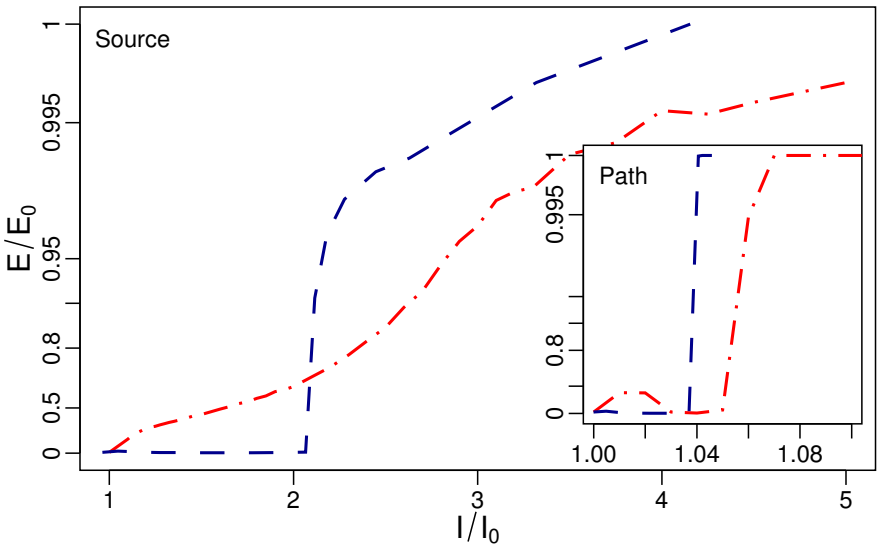


Figure 2.7.: *Relative efficiency of a typical N=1000 scale-free network as a function of investment costs against direct vertex failures as well as subsequent cascading failures induced by path- (inset) and source-like fluctuations of strength $\sigma = 0.8$. Dash-dotted and dashed curves represent the capacity layouts for Equations (2.19) and (2.27), respectively. All curves come with confidence level of* 0.99*, meaning that* 10 *out of the simulated* 10^3 *fluctuation realizations result in relative efficiencies below the curve.*

3. Weather Data Driven Analysis of Renewable Energy Generation

We investigate the design of a future European power supply system with a high share of renewables energy generation. These days wind power has emerged as the most dominant contributor to renewable power generation, with potential for more in the future. A straightforward and simple answer to the first question has been pointed out [72, 81]: there are enough wind resources around the globe to supply all continents with only wind power. Also, the potential for solar power is high [30]. The different renewable power sources and their potential are discussed in Section 3.1. Due to its weather-driven fluctuations wind and solar power generation is not a simple stand-alone solution. An enormous amount of balancing and storage will be needed on top of this wind-only scenario.

As the weather is the driving force of fluctuations in wind and solar power generation, an analysis of a future power system with a high share of wind and solar generation has to be based on data that accounts for that influence. A eight year time series of wind and solar generation based on weather data with hourly time and a high spatial resolution is thus considered. The consumption, that has to be met by generation, is considered with the same temporal resolution. The details of the time series for consumption and generation are discussed in Section 3.2.

In this chapter, the whole European generation and consumption is analyzed on a macroscopic scale. This allows for general considerations on important characterisctics of the system as discussed in Section 3.3. Storage and balancing requirements strongly depend on the mixture between wind and solar power generation. An optimal mix, that minimizes storage or balancing capacities, is discussed for a hundred percent renewable generation scenario in Section 3.4 and for a transitional scenario in Section 3.5. Over-installation of generation capacities has advantages and is discussed in Section 3.6.

In the generation and the consumption data, distinct time scales are found. Their influence and possibilities to exploit these are discussed in Section 3.7. In Section 3.8, the results of this chapter are summarized. Large parts of the results shown here are published in Heide et al. [64].

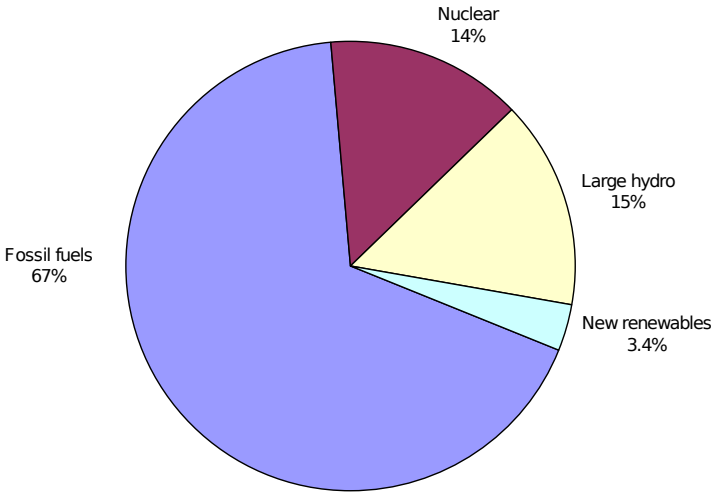
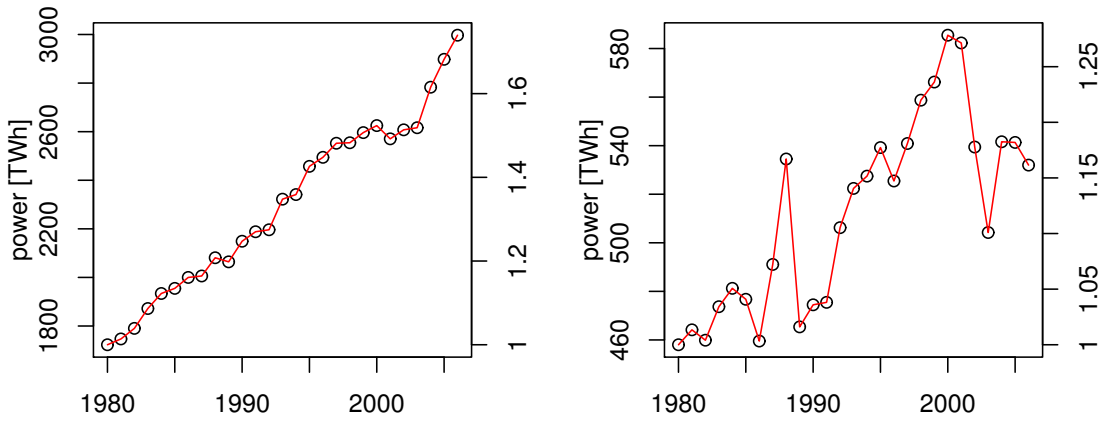


Figure 3.1.: *Share of global electricity from renewable energy in 2006 [98].*



(a) *World* **(b)** *Europe*

Figure 3.2.: *Accumulated hydroelectric power for (a) Europe and (b) Europe between 1980 to 2006. On the right axis the generation is normalized to the values of 1980. Data source : U.S. Energy Information Administration [119].*

3.1. Renewable Energy Generation

There are various technologies to generate energy. In Figure 3.1, the global shares of different electricity sources in 2006 are shown. Electricity generation is dominated worldwide by fossil fuels and nuclear power. Around 15% is contributed by large hydro[1] electricity, a much bigger contribution than the remaining renewables that produce

[1]Electricity generated from water flowing downhill, typically from behind a dam. Large hydroelectric power usually entails a substancial reservoir and refers to generation capacities larger than 10 MW. The definition can vary from country to country [99].

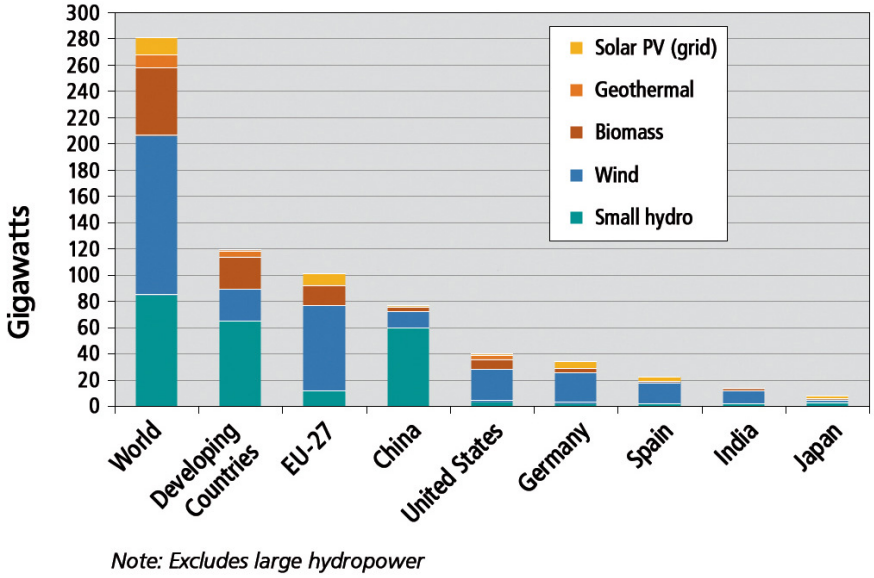


Figure 3.3.: *Renewable power capacities in 2008 for electricity generation worldwide, the developing countries, EU, and top six countries, taken from REN21 [99]. Large hydropower generates around a factor of three more than all other renewables together and is thus excluded, see also Figure 3.1. For the definition of large hydropower see text. The group of developing countries include China.*

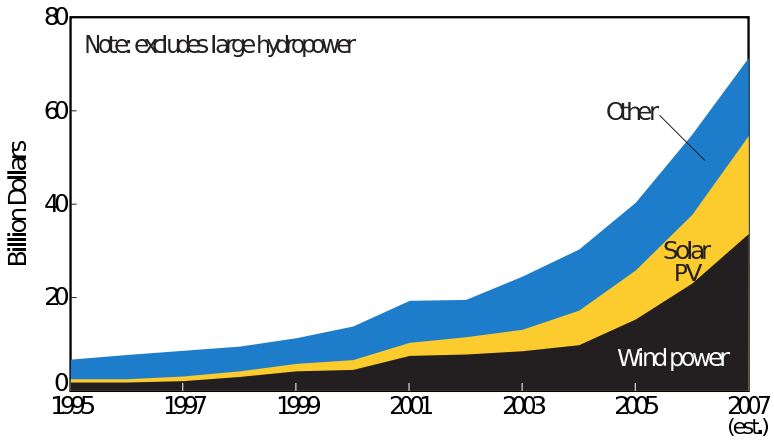
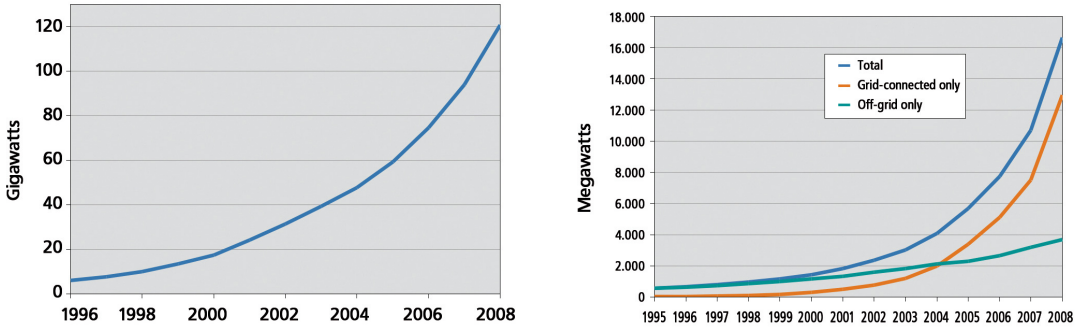


Figure 3.4.: *Annual investments in renewable energy capacities, taken from REN21 [98].*

around 3.4% of the electrical power worldwide. From 1980 to 2006, the electricity generation from hydro increased by almost 80% worldwide, see Figure 3.2a. In the same time span, the European hydro electricity increased by only 30% as shown in Figure 3.2b. This is due to the fact that most of the economically feasible hydroelectric resources in Europe had to a large extend already been developed. Further renewable energy generation capacities within Europe have to be build up using other technologies [118].



(a) *Wind power capacities* **(b)** *Solar power capacities*

Figure 3.5.: *Worldwide installed capacities for wind and solar power [99].*

The most important renewable energy sources, excluding large hydro, are shown in Figure 3.3 for selected countries and regions. Wind power generation makes already a large contribution to the renewable generation, while solar power plays a smaller role. The highest potential for new renewable capacities is expected for wind and solar power [50]. This forecast is supported by the fact that the investments from 1996–2006, shown in Figure 3.4, have increased almost exponentially and is also reflected in installed wind and solar capacities, see Figure 3.5. This strongly demonstrates the importance these energy sources are going to have in the next decades [see e.g. 50, 118, 120].

In the following, we focus on wind and solar power generation. However, hydro power can be very important to balance shortages as large capacities exist that can be activated on short time scales, see Sections 3.6 and 3.7.

3.2. Data Time Series

Key to the modeling of wind- and solar-power generation is a large weather data set with good spatial and temporal resolution all over Europe. Its convolution with future-projected wind and solar power capacities reveals how much wind and solar power is generated across Europe together with the spatial and temporal information. The following subsections will explain the details. The load modeling is described in the last subsection of this section.

3.2.1. Weather Data

Weather data for all of Europe is available from various sources with different spatial and temporal resolutions. For our purposes three selection criteria have been important:

- The correct modeling of intra-day solar- and wind-power ramps require a good

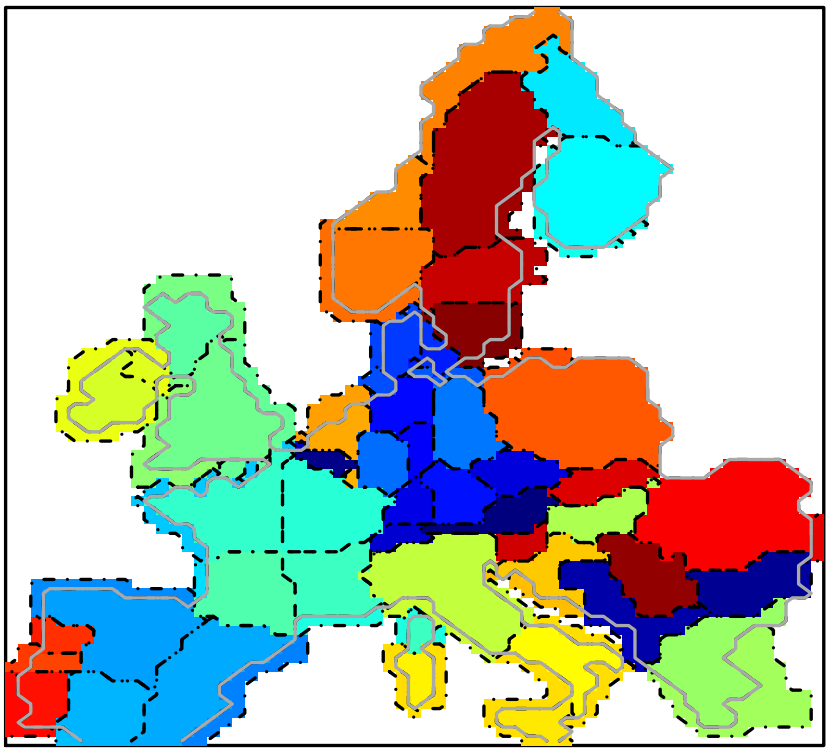


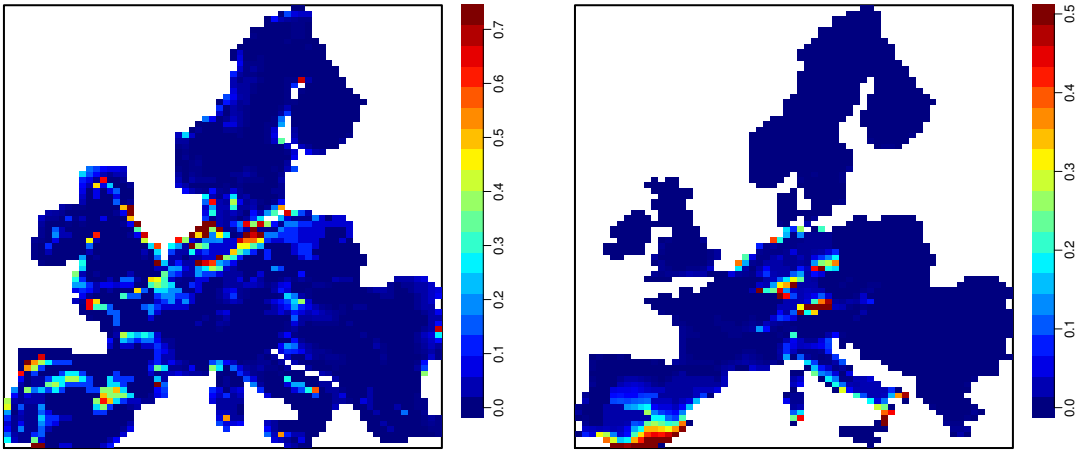
Figure 3.6.: *The region partintioning. Due to the fact that load data is only available for regions larger than the grid size of the weather data, the generation data is aggregated accordingly. The gray line indicates the onshore area, regions outside are considered as offshore. The borders between the regions are indicated by the black lines.*

time resolution of at least 1h.

- In order to resolve the passing of synoptic systems related to high winds and opaque clouds a spatial resolution of at least $50\times50\text{km}^2$ is required.
- In order to gain representative and significant statistics covering all possible seasonal and extreme weather situations a rather long time window is required, ranging over a couple of years.

These criteria have been met by the private weather service provider WEPROG (Weather & Wind Energy Prognosis) [124]. With regional models it downscales medium-resolved analysis data from the US Weather Service NCEP (National Center for Environmental Prediction) [86] down to $47\times48\text{km}^2$ spatial and 1h time resolution over an eight-years period (2000-2007).

This high-resolution data provides direct information on the wind speed and direction 100m above ground. The solar global radiation is not a standard output, but can be computed directly from the data on the net short wave radiation at the surface, the



(a) *Wind power capacities* **(b)** *Solar photovoltaics power capacities*

Figure 3.7.: *Expected wind power and solar photovoltaics power capacities [GW] per grid-cell across Europe in 2020. The spatial grid-cell resolution of* $47 \times 48\text{km}^2$ *has been adapted to the weather data. For a better visualization capacities larger than 0.73 GW for wind and 0.50 GW for PV are indicated in dark red.*

total cloud cover, and a standard cloud and surface albedo.

3.2.2. Assumed Wind- and Solar-Power Capacities and Generation

The national 2020 targets serve as guidance for a rough distribution of wind and photovoltaic capacities in Europe. Figure 3.7 illustrates the expected installed wind-power and solar photovoltaics power capacities across Europe. They total to 227 and 68 GW, respectively. 66 GW of wind power is assumed to be installed offshore. The subsequent finer distribution within each country onto the grid cells of the weather data is done empirically, giving more capacity to those grid cells with large average wind speed and large average global radiation, respectively.

The conversion of hourly WEPROG wind speeds into wind power at each grid cell was done using typical wind power curves at 100m hub height. Different power curves have been assigned for on- and offshore grid points. Losses due to wake effects have been modeled explicitly for offshore grid points by assuming a park layout of 7x7 turbines in offshore wind farms. Additional 7% losses have been introduced due to electrical losses and turbine non-availability. The same 7% of losses have also been applied to onshore grid points. The turbine cut-off due to extreme winds is empirically parameterized by an additional modification of the power curve, which mimics the gradual power-lowering-behavior of wind turbines with storm-control.

The solar photovoltaic power generation within the grid cells has been calculated based

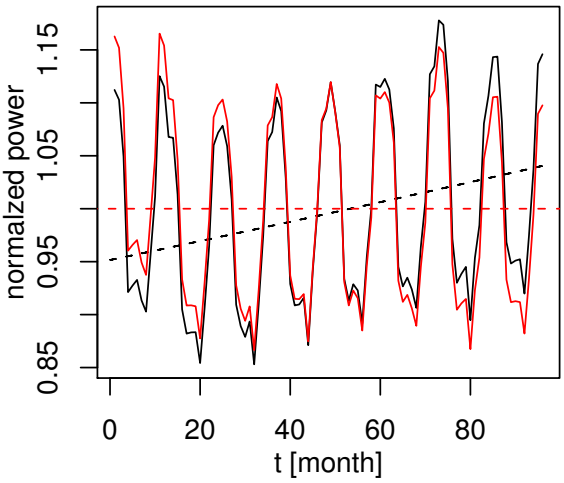


Figure 3.8.: *The monthly averaged load accumulated over Europe. The black curve indicates the data. An exponential trend is removed resulting in the load shown by the red curve. The black dashed line shows the fit of an exponential trend based on a running one year average.*

on the available meteorological data (global radiation, air temperature), assumptions on the characteristics of the photovoltaic plants (tilt angle, orientation, fixed or with solar tracker) and the geographical coordinate of the grid cell considered. A mix of different photovoltaic plant technologies was considered for each grid cell [16].

This convolution of the weather data with the wind and solar power capacities produces spatio-temporal power-generation patterns across Europe. These patterns are important for the calculation of power flows, see Chapter 6.

3.2.3. Load modeling

There are no sources to obtain the load data on the same spatial $47\times48\text{km}^2$ grid-cell resolution as for generation. But for the work presented here, a coarser resolution is sufficicent. For almost all European countries the load profiles have been downloaded either from the UCTE-homepage [117] or from the national transmission providers. At least for the two recent years those have an hourly resolution. For the remaining years they have been replicated with the known relative annual electric power consumption; special care was given to a proper handling of the weekend effect.

The load increases over time, see Figure 3.8. Since for generation the fixed capacities expected for 2020 are used, the load is detrended. An exponential trend[2] is removed according to $y = a\exp(b \cdot t)$. Every country is detrended individually, since the growth factor strongly deviates from country to country. For all of Europe the load increases

[2]The exponential detrending was not done in Heide et al. [64], so the results shown here differ slightly.

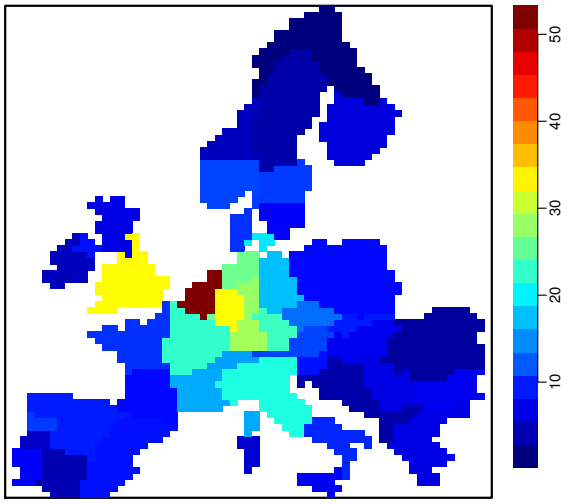


Figure 3.9.: *Average annual load [TWh] per grid cell in the 50 coarse-grained onshore regions.*

by 1.138% per year over the considered eight years.

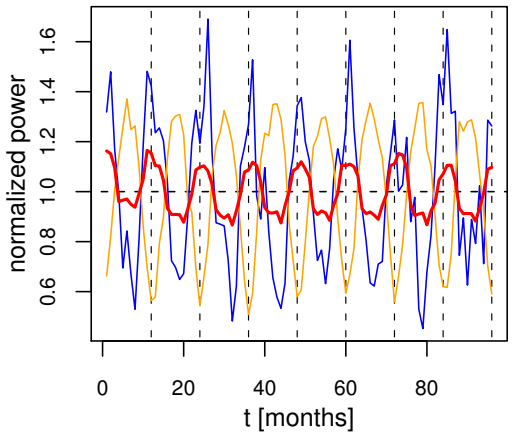
Some countries, especially the larger ones, come with a large average load. Those have been further subdivided into regions, with some spatial correlation to the territories of the respective network transmission providers. The regional load profiles have been obtained from the country profiles with a multiplicative factor obtained from a linear regression between the annual electric power consumption on the one hand and population and gross domestic product on the other hand.

Figure 3.9 shows the average annual load of the 50 onshore regions during the years 2000-2007. Offshore regions come with no load and are not shown. The sum over all regions totals to 2995 TWh average annual consumption. Its seasonal dependence is shown in Figure 3.10.

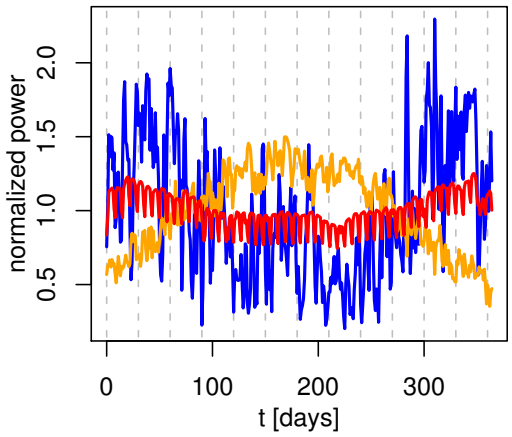
3.3. The European Perspective

In a renewable future well beyond 2020 the share of wind-power generation may well increase beyond fifty percent [50, 72]. For such a large amount, the spatial and temporal scales that have to be looked at increase substantially. Regions with a momentary excess of wind power will try to export it, whereas deficit regions are depending on import. This spatial horizon beyond regions and countries helps to smoothen short-term wind-power fluctuations and to reduce the short-term balancing and storage needs.

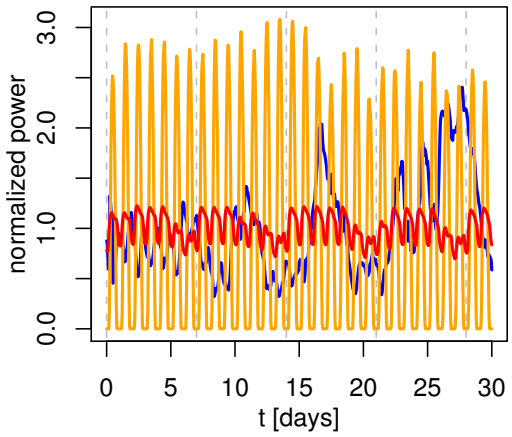
Besides the short-term time scales ranging from minute to a few days, the weather follows a distinct seasonal pattern. Winds across Europe are stronger during winter than in summer. As a consequence, the wind power aggregated over all of Europe is



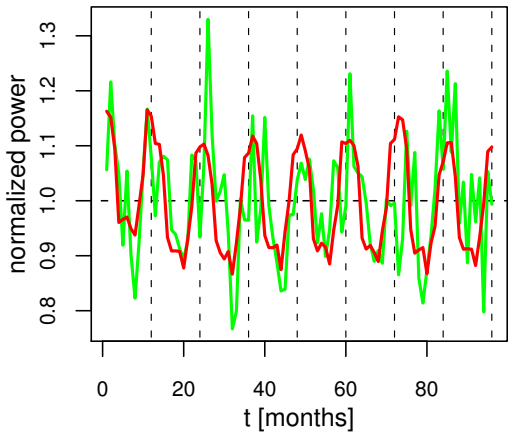
(a) *Wind-power-generation (blue), solar-power-generation (orange) and load (red). One month averages shown over all eight years are shown.*



(b) *Same as (a) for one day averages over a year.*



(c) *Same as (a) for one hour averages over a month (may).*



(d) *The wind- and solar-power generation time series are combined with a 60%/40% weighting (green) together with the load (red).*

Figure 3.10.: *Normalized wind-power-generation, solar-power-generation and load time series aggregated over Europe. Each series is shown in one-month resolution and is normalized to its 8-years average. The vertical lines indicate (a,d) months and (b,c) weeks.*

larger in winter than in summer. This is shown in Figures 3.10a to 3.10c. In fact, the winter maximum is about double the summer minimum.

If wind were the only power source in a fictitious future Europe, then the seasonal wind power curve has to be directly compared to the European load curve. This is

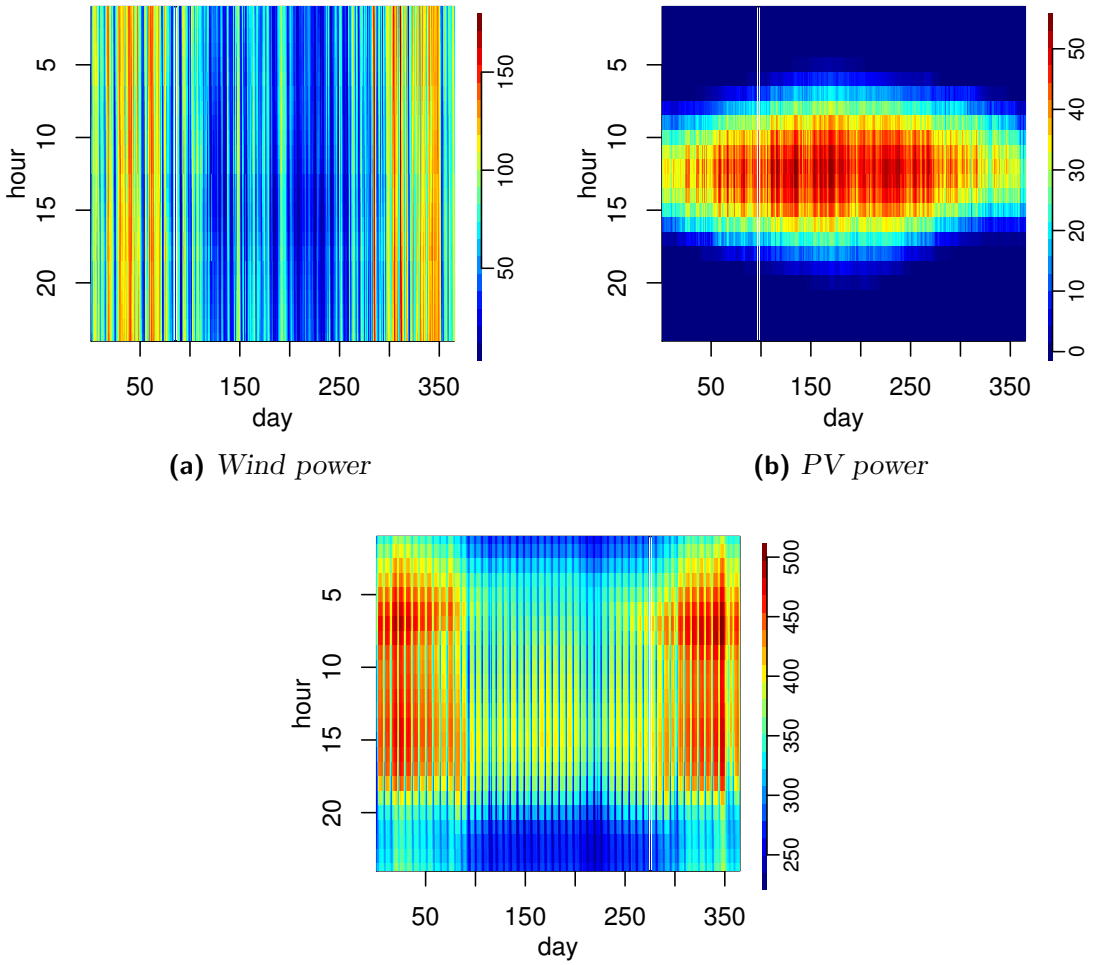
also illustrated in Figure 3.10a for the monthly average values. In this 100% wind-only scenario the yearly average of wind-power generation and load is the same. However, the seasonal behavior is different. The seasonal load curve also comes with a maximum in winter and a minimum in summer, but the seasonal variation strength is much smaller than for the wind-power generation. As a consequence an enormous amount of storage and balancing energy is required. Over summer the storage and balancing plants have to feed the deficit. During winter the large wind-power excess is put into the storage.

Like wind, also the solar community has its own solar-only answer to the question how much wind, solar etc. is needed [72]. If solar were the only power source in another fictitious European future, then the seasonal generation curve would look like the orange one in Figure 3.10a. The solar power generation is much larger during summer than in winter. Since it anticorrelates with the seasonal load curve, a 100% solar-only scenario will lead to even larger seasonal storage and balancing needs than for the wind-only case.

Let us summarize Figure 3.10a in another way. For Europe the seasonal wind-power generation nicely correlates with the seasonal load behavior. On the other hand, the seasonal solar-power generation anticorrelates with the seasonal load behavior. However, the seasonal wind- and solar-power variation strengths are roughly the same. Both are significantly larger than for the seasonal load.

When considering these facts set by the European weather conditions, an idea is created immediately: Future Europe is able to counterbalance seasonal wind with solar power generation. Their share should be almost the same, with a small extra contribution from wind power due to its seasonal correlation with the load. Figure 3.10d takes 60% from the wind curve and 40% from the solar curve of Figure 3.10a. The resulting curve is able to nicely follow the seasonal load curve. It is expected that this optimal mix brings seasonal storage and balancing needs to a minimum.

A different approach to look at the data is shown in Figure 3.11, where the generated wind and solar power and the load of the year 2000 aggregated over Europe are shown, colorcoded for every hour (y-axis) for each day (x-axis). This allows to look at all data for a given year at once. The 24-hour structuring takes the day/night changes with it's effects into account, thus showing temporal dependencies. The wind power generation has small fluctuations from hour to hour, but the seasonal time scale dominates. In both winter and summer, there are periods in the order of a few days, where there is no wind in winter and strong generation in summer. The solar power generation dominated by the day-night changes and also varying on a seasonal timescale. The load is influenced by many factors, seasonal and day/night effects as well as weekends and vacations, e.g. around Christmas.



(a) *Wind power*

(b) *PV power*

(c) *Load*

Figure 3.11.: *Generated power and load for a given hour (y-axis) on a given day (x-axis) of the year 2000. The colors code the power in GW.*

3.4. The 100% wind-plus-solar-only scenario

Based on seasonal time series such as shown in Figures 3.10a to 3.10c, it is straightforward to quantify a seasonal optimal mix between wind and solar power generation in a 100% wind-plus-solar-only scenario for a future Europe, in which the generation from wind and solar power on average match the consumption. Key to such quantifications is the mismatch energy

$$\Delta(t) = \gamma \left(a \frac{W(t)}{\langle W \rangle} + b \frac{S(t)}{\langle S \rangle} \right) - \frac{L(t)}{\langle L \rangle} \, . \tag{3.1}$$

$W(t)$ represents the total European wind power generation at hour t, and $\langle W \rangle$ its average over all data contained in the eight-years-long time series. $S(t)$ and $L(t)$ are

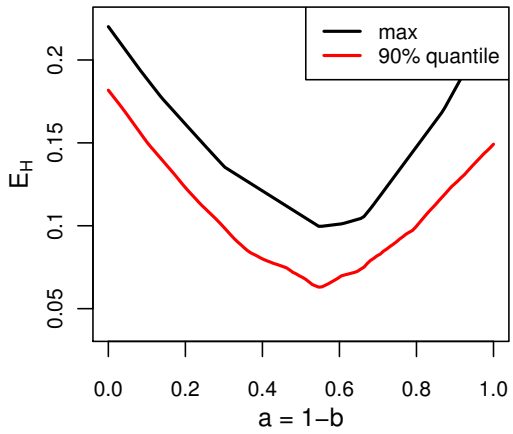


Figure 3.12.: *Storage energy, see Equation (3.3), as a function of the wind fraction $a = 1 - b$ in a 100% wind-plus-solar-only scenario for a future Europe based on hourly time resolution. For the storage energy its maximum (black) and 90% quantile (red) is shown. The unit of the storage energy is in average anual load.*

the respective solar power and load time series. The coefficients a and b tell how much of the load is on average covered by wind and solar power generation. For the scenario considered here, these coefficients are constrained to $a + b = 1$. The overproduction factor γ is needed to compensate for losses and is set to one in this section unless noted otherwise.

To evaluate the seasonal optimal mix, a simple storage model is constructed out of the mismatch energy in Equation (3.1)

$$H(t) = H(t-1) + \begin{cases} \eta_{\text{in}}\Delta(t) & \text{if } \Delta(t) \geq 0 \\ \eta_{\text{out}}^{-1}\Delta(t) & \text{if } \Delta(t) < 0 \end{cases} . \tag{3.2}$$

Whenever the mismatch is positive, the surplus generation is storage with efficiency η_{in}. In case of a negative mismatch the generation deficit is taken out of the storage with efficiency η_{out}. The time series $H(t)$ describes the filling level of the storage. The approach in Equation (3.2) is based on the property that on average the generation is equal to the load, so that the filling level at the end of the data set is equal to the initial level, $H(t = 0) = H(t = 8\text{y})$. For a first simple approach, its maximum and minimum determines the required storage energy

$$E_H = max_t H(t) - min_t H(t) . \tag{3.3}$$

This quantity is shown in Figure 3.12 as a function of the wind fraction $a = 1 - b$ normalized to the annual load. It comes with a rather flat minimum at $a = 0.47$. The storage efficiencies have been set to $\eta_{\text{in}} = \eta_{\text{out}} = 1$. Also shown in Figure 3.12 is the 90% quantile $Q(95\%) - Q(5\%)$ of the storage energy, which is determined from the distribution $p(H)$ and $\int_0^Q p(H)dH = 0.95$ (0.05). This second variant of the required storage energy reveals a pronounced minimum at $a = 0.57$.

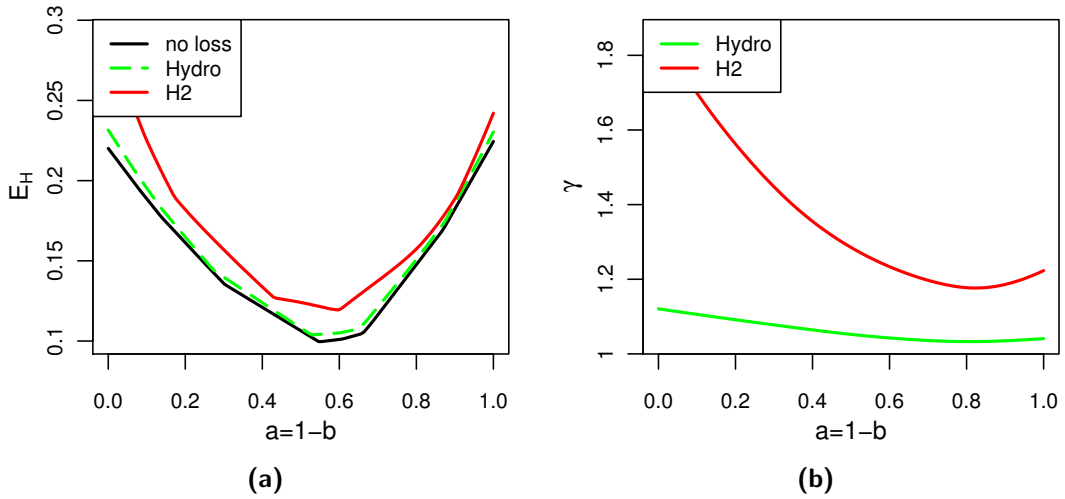


Figure 3.13.: *(a) Comparison of the maximum storage energies for an idealized storage (black), pumped hydro (green) and hydrogen storage (red), derived from hourly wind, solar, and load time series. The storage energy is normalized to the average annual load. (b) Overproduction factor γ of wind and solar power needed to compensate the storage losses.*

The used idealized storage efficiencies $\eta_{\text{in}} = \eta_{\text{out}} = 1$ are not realistic. Pumped hydro has $\eta_{\text{in}} = \eta_{\text{out}} = 0.9$ and hydrogen storage has $\eta_{\text{in}} = \eta_{\text{out}} = 0.6$. Since efficiencies smaller than one lead to storage losses, the wind and solar power generation has to be increased in order to compensate for the losses. The overproduction factor $\gamma = a + b > 1$ is determined from the requirement that the storage level $H(t = 8\text{y}) = H(t = 0)$ reached after 8 years is equal to the initial storage level. The formula yielding the overproduction factor is derived in Section A.4. Figure 3.13a shows that the smaller the storage efficiencies turn out to be the larger the maximum required storage energy becomes. However, the location of the seasonal optimal mix does not change significantly. The overproduction factor is illustrated in Figure 3.13b. For $a = b$ it amounts to $\gamma = 1.05$ (pumped hydro) and 1.28 (hydrogen).

As can be seen in Figures 3.12 and 3.13a, the required maximum storage energy has to be 10% (without storage losses) and 12% (for hydrogen storage) of the anual load and amounts to 300 and 360 TWh, respectively. These are very large numbers. They will even double once future Europe decides to switch to a wind-only or a solar-only scenario.

In the absence of storage, the needed balancing power can be also defined based on Equation (3.1). In case of a negative mismatch, additional power is needed, e.g. from gas turbines or hydro power. It is defined as

$$B(t) = \begin{cases} -\Delta(t) & \text{if } \Delta(t) < 0 \\ 0 & \text{if } \Delta(t) \geq 0 \end{cases} . \tag{3.4}$$

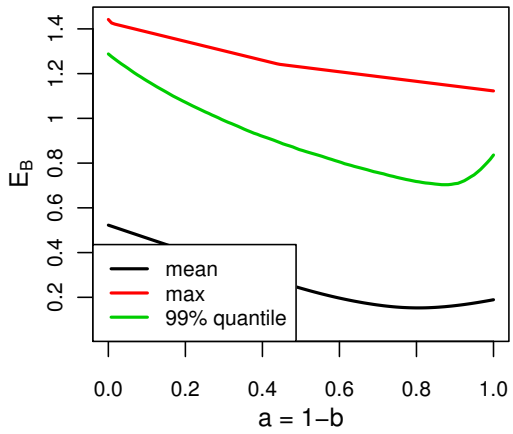


Figure 3.14.: *Needed balancing power, see Equation (3.4), in dependence of the wind fraction $a = 1 - b$. The balancing power is normalized by the average load.*

The mean, maximal and the 99% quantile of the required balancing power are shown in Figure 3.14. The minimum of the mean comes at a fraction $a = 0.80$ of wind power with 0.15 times the average load. For the 99% quantile that corresponds to around 12 hours per year, 0.70 times the average load is needed as balancing power and the maximal balancing amounts to 1.12 times the average load, that is needed for one hour over the period of eight years. The optimal mix is shifted towards $a = 1$ for higher quantiles. The finding, that storage and balancing energy yield different optimal mixes, is discussed in Section 3.7. The needed balancing capacities based on the average hourly load amount to 51 GW on average, which have to supply 447 TWh per year. The capacities that ensure balancing for the 99% quantile have to be 239 GW.

The storage energy lead to a seasonal optimal mix of 50-60% wind and 50-40% solar power generation. These results agree nicely with the intuition obtained from Figure 3.10. From these findings on the seasonal optimal mix a few more conclusions can be given for a 100% wind-plus-solar-only scenario in a future Europe. Just for demonstration we consider the average load from 2000 to 2008, which is 2995 TWh. 55% of this makes 1647TWh and requires 627GW of installed wind power capacity; here the wind load factor 0.30 has been used, which has been directly determined from the used weather data. The remaining 45% make 1348 TWh and require 733 GW of installed solar photovoltaic power; here the PV load factor 0.21 has been used, which again has been directly determined from the used weather data.

627GW of installed wind power capacity across Europe translate into around 314000 build 2 MW turbines, or 126000 5MW turbines, or around 4000 wind farms of the size of the first offshore wind farms Horns Rev I and Nysted. As a rule of thumb [41], one MW of installed capacity requires 0.07km^2 onshore and 0.11km^2 offshore, respectively. This translates 627GW into around 50000km^2 onshore or 75000km^2 offshore. For comparison, Denmark has an area of 43000km^2.

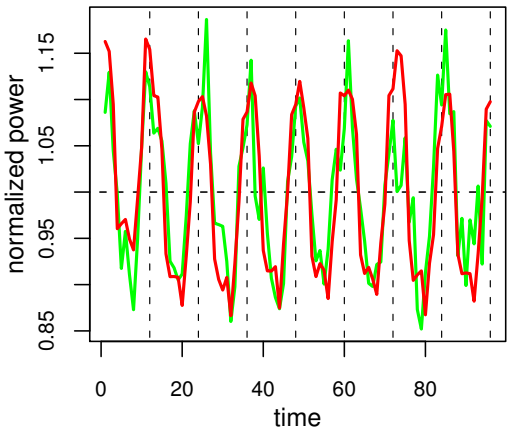


Figure 3.15.: *A mix of 27% wind, 0% solar and 73% fossil-nuclear power generation (green) is also able to follow the (red) seasonal load curve. See Figure 3.10 for comparison.*

The spatial and temporal mean global radiation 169W/m^2 is computed from the weather data and translates 1460TWh-producing 810GW-installed solar photovoltaic power into a PV-panel area of 5000km^2. For comparison, Germany has the potential to cover 1330km^2 of roofs with ideal slopes and direction [96].

Currently, Germany has about 190 GWh of pumped hydro facilities in operation, with only little room for more. The exact amount of pumped hydro across all of Europe is not known to the author. Even if it is a factor of ten more, still two orders of magnitude are missing to reach the required 300-480 TWh. This is discussed further in Section 3.6.

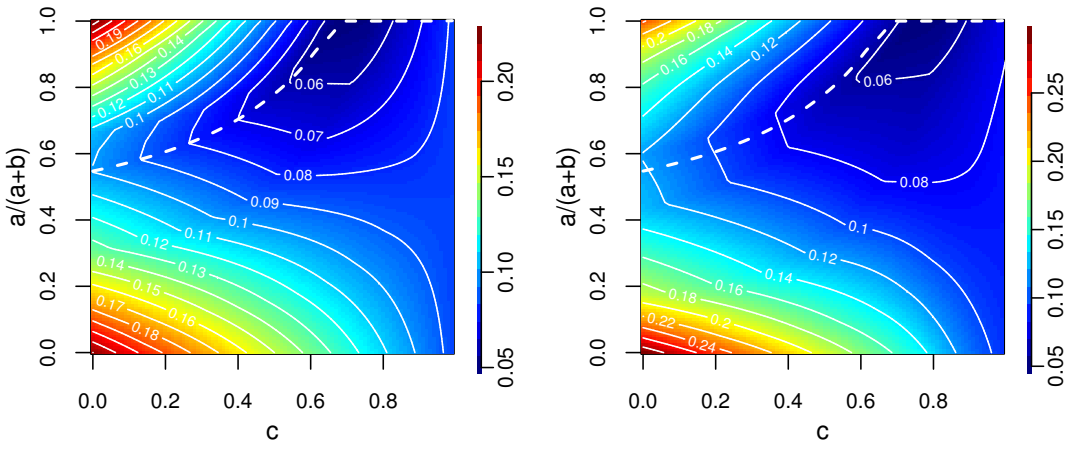
3.5. Transitional scenarios with wind, solar, and fossile-nuclear power

The investigations on the 100% wind-plus-solar-only scenario can be extended to transitional scenarios by modifying the mismatch energy defined in Equation (3.1) to

$$\Delta(t) = \gamma \left(a\frac{W(t)}{\langle W \rangle} + b\frac{S(t)}{\langle S \rangle} \right) + c\frac{F(t)}{\langle F \rangle} - \frac{L(t)}{\langle L \rangle} \,. \tag{3.5}$$

$F(t) = \langle F \rangle$ represents fossil-nuclear power generation and is assumed to be time-independent as these are optimized for continuous operation. It may even include a contribution from geothermal power. The transitional strategy discussed in this section, results in the 100% scenario as introduced in Section 3.4.

The three coefficients $a + b + c = 1$ add up to one and match the average load. The choice $a = 0.27$, $b = 0.00$, $c = 0.73$ leads to Figure 3.15. The seasonal power generation curve follows the seasonal load curve more closely than for the $a = 0.60$, $b = 0.40$,



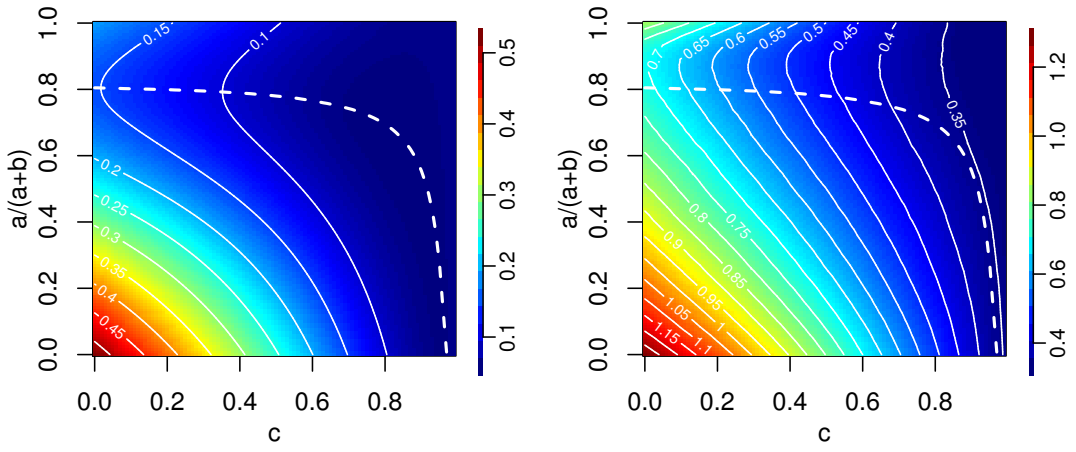
(a) $\eta = 1$ **(b)** $\eta = 0.6$

Figure 3.16.: *Maximum storage energy, see Equations (3.3) and (3.5), required for all of Europe as a function of the two independent coefficients c and $a/(a+b)$. The dashed curve represents the seasonal optimal mix between wind and solar power generation as a function of the remaining fossil-nuclear power generation for $\eta = 1$. The unit of the contours is given in average annual load over Europe.*

$c = 0.00$ example shown in Figure 3.10d. This indicates already, that as long as a fraction of fossil-nuclear power remains in the generation mix the need for storage energy will be smaller than for the 100% wind-plus-solar-only scenario.

Figure 3.16a shows the required maximum storage energy, see Equation (3.3), which has been deduced from Equation (3.5) with $\eta_{\text{in}} = \eta_{\text{out}} = 1$ and Equation (3.2). It is a function of the two independent coefficients c and $a/(a+b)$. The smallest storage energy is obtained for $a = 0.27$, $b = 0.00$, $c = 0.73$, which are the values used for Figure 3.15. For large fossil-nuclear fractions $1 \geq c \geq 0.73$ the storage energy reaches a minimum when only wind power generation covers the remaining fraction $a = 1 - c$. This is because the seasonal wind power generation curve correlates with the seasonal load curve; confer again Figure 3.10a. Due to its anticorrelation it is not favorable to include solar power into the fossil-nuclear-dominated generation mix. The same is true for $\eta = 0.6$, as shown Figure 3.16b. The qualitative behaviour is similar to the case of $\eta = 1$.

Solar power is needed once the fossil-nuclear power generation is reduced below $c<0.73$. The seasonal solar power generation then has to counterbalance the seasonal wind power generation. Otherwise the absolute seasonal wind-power variation would become larger than the absolute seasonal load variation. The dashed curve in Figure 3.16a follows the bottom of the storage valley and represents the seasonal optimal mix between wind and solar power generation as a function of the remaining fossil-nuclear power generation. In the limit $c \to 0$ the wind and solar coefficients $a \approx b \approx 0.5$ become about the same. Obviously, this result agrees with the earlier result obtained in Figure 3.12.



(a) *average* **(b)** 99% *quantile*

Figure 3.17.: *Balancing power needed, see Equations (3.4) and (3.5), required for all of Europe as a function of the two independent coefficients c and $a/(a+b)$. The dashed curve represents the seasonal optimal mix between wind and solar power generation as a function of the fossil-nuclear power generation. The unit of the contours is given in average load over Europe.*

Let us follow the dashed optimal-mix-curve once more, from right to left. At $c = 1$ the required maximum storage energy amounts to 8% of the annual load. From $c = 1$ down to $c = 0.73$ the required maximum storage energy decreases down to 5% of the annual load. From $c = 0.73$ to $c = 0$ the required maximum storage energy increases again and reaches 10% of the annual load at $c = 0$.

The balancing, shown in Figure 3.17, exhibits a different qualitative behaviour for both the average and the quantile values. The optimal mix for $c \ll 1$ is around $a = 0.8$. For $c \to 1$, the wind share goes down to zero. The reasons behind this deviation of the optimal mix minimizing the storage energy from the case minimizing the balancing power is discussed in Section 3.7. However, for the transitional scenario and $c \to 1$, the variations in the needed balancing power are small for different a. Therefore, the optimal mix obtained for storage could be used, with only minor increase in the needed balancing power.

We close this Section with an additional remark. If the seasonal load curve had come with a maximum in summer and a minimum in winter, then the optimal-mix curve would have been different. This is discussed in more detail in Section 4.1.

3.6. Over Installation

Even when spatially aggregated over the whole continent, the fluctuation strength of wind and solar power remains large. In order to absorb these fluctuations, enormous amounts of storage and balancing are required. For a simplified Europe based on 100% wind and solar power generation the required storage energy has been estimated in Sections 3.4 and 3.5. It amounts to 10% of the annual European consumption. Given the consumption rate of 2007, this corresponds to 320 TWh. This large number is already an optimal minimum, where 60% wind and 40% solar power generation are mixed together so that their opposite strong seasonal dependences almost cancel each other and follow the weaker seasonal load behaviour (see Figure 3.10d).

For pumped hydro an energy capacity of several hundred TWh is fully out of European reach. As discussed in Section 3.1, hydro power is almost fully developed in Europe. Pumped-storage hydroelectricity requires two reservoirs at different heights and water is either pumped up for storage or released back through a turbine to generate energy. Because of its dependency on suitable topological sites and the needed altitude difference for the reservoirs, the potential of future pumped hydro storage capcities is limited [61]. A hypothetical hydrogen storage in mostly North German salt caverns would be able to store energy in the order of $10 - 20$TWh. These numbers are derived from the potential volume of working gas[3] in all European salt caverns of $9.09 \cdot 10^9$ m^3 with planned extensions of another $3.5 \cdot 10^9$ m^3 [56]. The energy density is $170\,{}^{kWh}/_{m^3}$ at an operating pressure difference of 120 bar [44], yielding the above numbers. The energy density of adiabatic compressed air storage is two orders of magnitude smaller [44] amounting to 0.1 TWh, if all available salt cavern would be used. Even the hypothetical hydrogen storage would still be one order of magnitude away from the estimated several hundred TWh required that need to be stored. At present no other large-scale round-trip storage technologies are in sight. Storage lakes represent a different form of storage. They do not store excess electricity, they only balance electricity deficits. Scandinavia has most of the storage lakes in Europe. They have a seasonal energy capacity of about 100 TWh. This is also by far not sufficient for a 100% wind-plus-solar Europe, defined as a scenario.

A way has to be found how to reduce the enormous amount of storage needs for a fully renewable Europe. A straightforward solution is to go into the over-installation regime, where wind and solar power generate more on average than the load. Negative mismatches in the fluctuating balancing between wind-plus-solar-power generation and load will occur less frequently. The question then is, how much excess generation is needed? We want to learn in a quantitative manner how the storage capacities decrease as a function of excess generation. At the same time, we want to look not only at storage energy, but also on balancing power, and ask again the same questions.

[3] The working volume is not the volume of the storage site, but the amount of gas which can be stored and extracted again.

The calculations in this section are based on the mismatch as defined in Equation (3.1) only here with an over installation factor $\gamma \geq 1$. The amplitudes γa and $\gamma b = \gamma(1-a)$ tell how much of the load is on average covered by wind and solar power generation. $(\gamma - 1) \geq 0$ represents the average excess generation. a and $b = (1-a)$ are equal to the share of average wind- and solar-power generation, respectively.

When considering the simple storage model defined in Equation (3.2), we find that, due to the overproduction, the time series of the storage $H(t)$ has a positive linear drift, see Figure 3.18 (middle). In this case, the simple subtraction in Equation (3.2) of the overall minimum from the overall maximum of the storage-level as used in Section 3.4 time series does not make sense. The new definition of the storage energy

$$E_{\mathrm{H}} = max_t \left(H(t) - min_{t' \geq t} H(t') \right) \ . \tag{3.6}$$

takes care of the positive drift. At time t the non-constrained storage level is $H(t)$. For all larger times $t' \geq t$ the non-constrained storage level does not drop below $min_{t' \geq t} H(t')$. Their difference represents the required storage energy at time t. Its maximum over all times yields the required overall storage energy E_{H}. Note, that the alternative definition $E_{\mathrm{H}} = max_t \left(max_{t' \leq t} H(t') - H(t) \right)$ is fully equivalent to Equation (3.6). With the definition in Equation (3.6), the non-constrained time series in Equation (3.2) can be transformed into a constrained storage-level time series:

$$H_{\mathrm{c}}(t) = \begin{cases} E_{\mathrm{H}} & \text{if } E_{\mathrm{H}} - H_{\mathrm{c}}(t-1) < \eta_{\mathrm{in}} \Delta(t) \ , \\ H_{\mathrm{c}}(t-1) + \eta_{\mathrm{in}} \Delta(t) & \text{if } E_{\mathrm{H}} - H_{\mathrm{c}}(t-1) > \eta_{\mathrm{in}} \Delta(t) > 0 \ , \\ H_{\mathrm{c}}(t-1) + \eta_{\mathrm{out}}^{-1} \Delta(t) & \text{if } H_{\mathrm{c}}(t-1) \geq -\eta_{\mathrm{out}}^{-1} \Delta(t) \geq 0 \ , \\ 0 & \text{if } H_{\mathrm{c}}(t-1) + \eta_{\mathrm{out}}^{-1} \Delta(t) < 0 \ . \end{cases} \tag{3.7}$$

By construction, the constrained storage level never exceeds the storage energy given by Equation (3.6) and never drops below zero. An example of a constrained storage level time series is shown in Figure 3.18 (bottom).

3.6.1. Storage Energy

The storage energy as defined in Equation (3.6) depends on the excess generation γ and the relative share a of wind-power generation via Equations (3.1) and (3.2). This dependence is illustrated in Figure 3.19a. The storage efficiencies have been set to $\eta_{\mathrm{in}} = \eta_{\mathrm{out}} = 1$. The contour plot reveals a strong dependence on a and γ. At $\gamma = 1$ the minimum storage energy is at $a \approx 0.6$. This represents the optimal mix between wind and solar power generation, which has been found in Sections 3.4 and 3.5. As can be seen from the dashed line, this optimal mix does not change much for $\gamma > 1$. For $1.1 \leq \gamma \leq 1.4$ it is around $a \approx 0.75$ and for $1.6 \leq \gamma \leq 1.9$ it is around $a \approx 0.45$. Note, that the respective absolute values $E_{\mathrm{H}}(\gamma, a)$ do not differ much from $E_{\mathrm{H}}(\gamma, a = 0.6)$. Figure 3.19b shows several one-dimensional cuts through the two-dimensional landscape of Figure 3.19a. The cuts along the optimal mix and along $a = 0.6$ are almost identical.

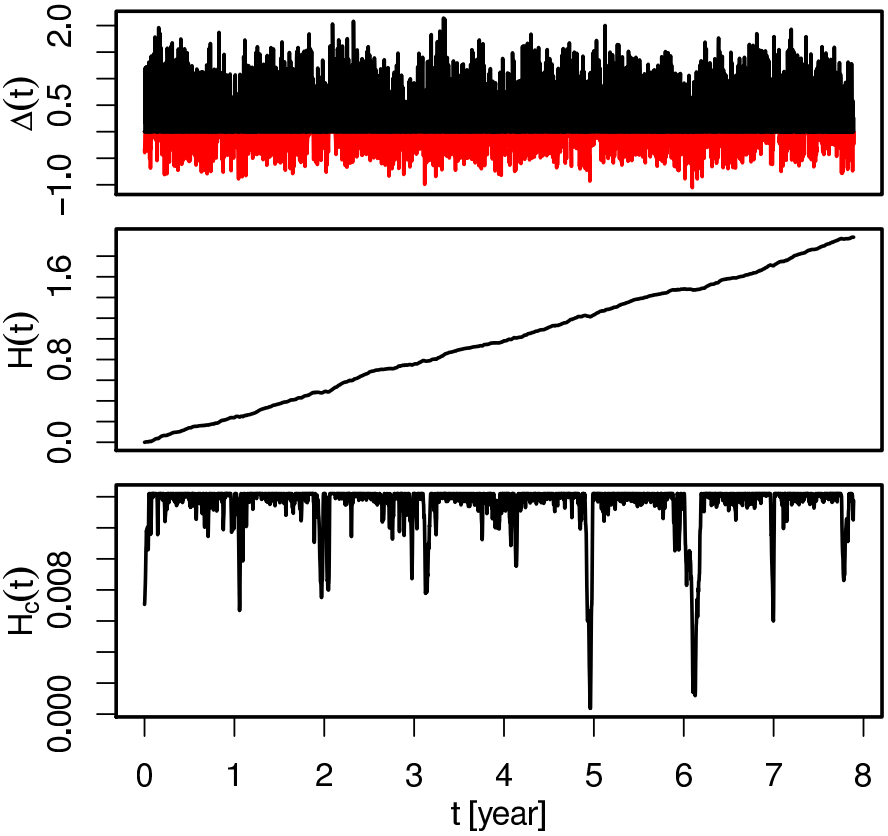


Figure 3.18.: *Time series of (top) mismatch power from Equation (3.1) (middle) non-constrained storage level following from Equation (3.2), and (bottom) storage level as defined in Equation (3.7) constrained with Equation (3.6). The unit of the mismatch power is given in average hourly load. The unit of the storage levels is given in annual load. Parameters have been set* $\gamma = 1.25$, $a = 0.60$, $\eta_{\text{in}} = \eta_{\text{out}} = 1$.

For these two cases the storage energy decreases much faster with increasing γ than along the wind-only $a = 1$ and solar-only $a = 0$. At $\gamma = 1.05$ the storage energy $E_{\text{H}}(\gamma = 1.05, a = 0.6)$ is half of $E_{\text{H}}(\gamma = 1, a = 0.6)$, at $\gamma = 1.15$ it is a quarter, and at $\gamma = 1.35$ it is a tenth.

Figure 3.20 is similar to Figure 3.19, except that the storage efficiencies have been set to $\eta_{\text{in}} = \eta_{\text{out}} = 0.6$. These values are valid for hydrogen storage [36]. Due to the conversion losses, some extra generation is needed to make up for the losses. The left border in Figure 3.20a represents the extra generation in order to fully match the average load. For $a = 1.0, 0.7, 0.0$ this extra generation amounts to $\gamma = 1.23, 1.20, 1.87$, respectively. Again, the storage energy reveals a strong dependence on a and γ. The optimal mix between wind and solar power generation, which minimizes the storage energy, is at $a \approx 0.7$. Compared to $E_{\text{H}}(\gamma = 1.20, a = 0.7)$, the storage energy $E_{\text{H}}(\gamma, a = 0.7)$ reduces to a half at $\gamma = 1.23$, to a quarter at 1.30, and to a tenth at 1.57; see Figure 3.20b. The wind-only storage energy $E_{\text{H}}(\gamma, a = 1.0)$ is found to be significantly larger than for the

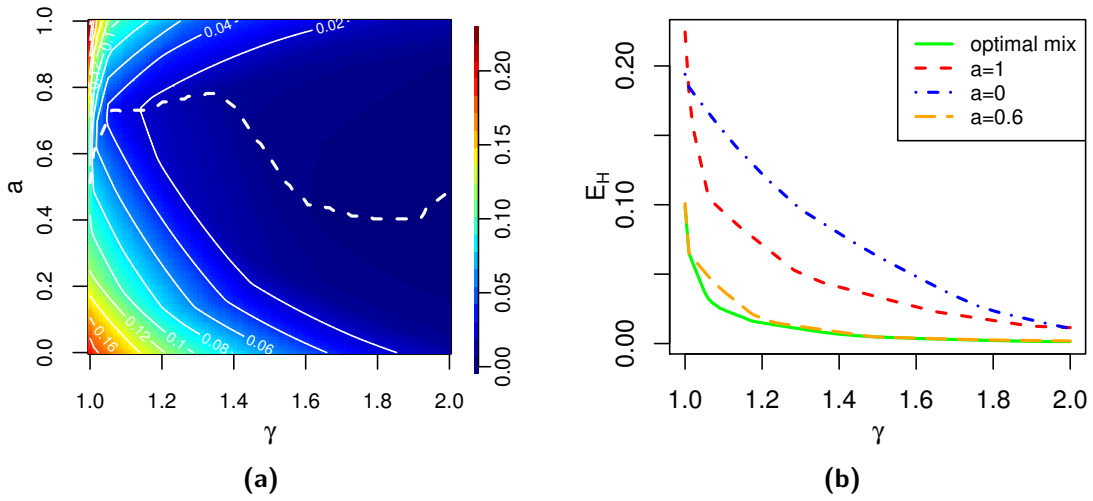


Figure 3.19.: *(a) Storage energy as defined in Equation (3.6) as a function of the excess generation γ and the share a of wind-power generation. The storage efficiencies have been set equal to $\eta_{\text{in}} = \eta_{\text{out}} = 1$. The contour lines represent constant storage energy and their attached numbers are measured in average annual load. (b) Cuts through (a) at $a = 1.0$ (red), 0.6 (orange), 0.0 (blue), and along the dashed optimal-mix line (green).*

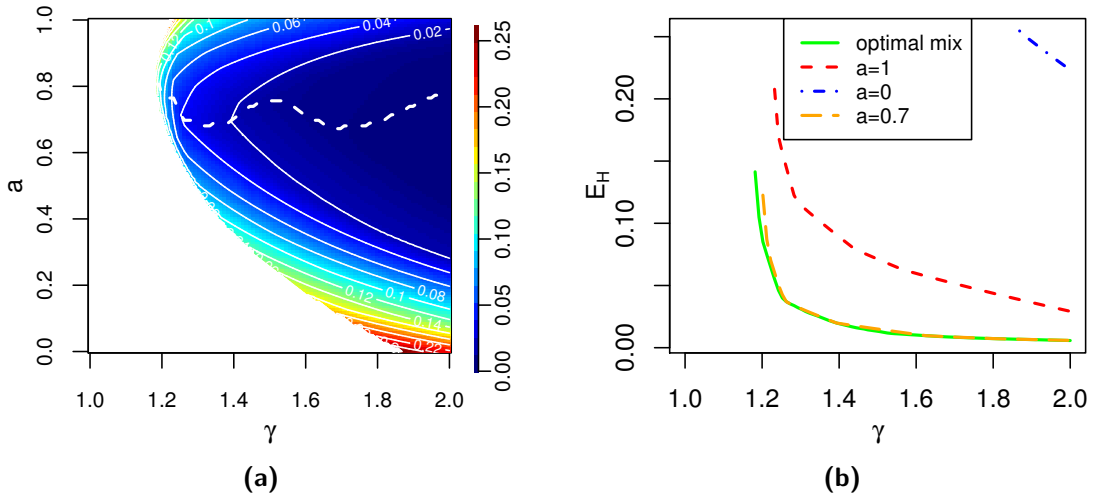


Figure 3.20.: *Same as Figure 3.19, but for storage efficiencies $\eta_{\text{in}} = \eta_{\text{out}} = 0.6$.*

optimal-mix storage energy $E_{\text{H}}(\gamma, a = 0.7)$. The solar-only storage energy $E_{\text{H}}(\gamma, a = 0.0)$ is beyond reach.

The storage energy in Figures 3.19 and 3.20 is measured in annual load. The average annual load for all of Europe amounted to 3240TWh in 2007. In case of ideal storage with $\eta = 1$ the value $E_{\text{H}}(\gamma = 1.00, a = 0.6) = 0.10$ then corresponds to 330TWh

of storage energy. With 35% excess generation this value is reduced to a tenth, i.e. $E_{\mathrm{H}}(\gamma = 1.35, a = 0.6) = 33\mathrm{TWh}$. Because of the conversion losses, hydrogen storage requires 20% excess generation per se. Its storage energy $E_{\mathrm{H}}(\gamma = 1.20, a = 0.7) = 0.12$ corresponds to 400TWh. A 57% excess generation is needed to reduce this value down to $E_{\mathrm{H}}(\gamma = 1.57, a = 0.7) = 40\mathrm{TWh}$. This reduced value is in the order the estimated $10 - 20\mathrm{TWh}$ potential of hydrogen storage in North German salt caverns, see the introduction of Section 3.6.

The required storage energies for pumped hydro and compressed air storage lie inbetween those for ideal and hydrogen storage. An excess generation somewhere inbetween 35-55% will lead to a storage energy around 40TWh.

3.6.2. Balancing Energy

A different form of storage is provided by storage lakes. It does not make use of the positive mismatch power and is only used to balance the negative mismatch. Storage lakes and gas turbines represent examples for balancing storage.

As defined in Equation (3.4), the hourly balancing power is defined by the negative mismatch. When multiplied with the number of hours $T = 8760$ contained in a year, its average leads to a measure

$$E_{\mathrm{B}} = T \langle B \rangle \ , \tag{3.8}$$

which we denote as the annual balancing energy.

Figure 3.21 illustrates E_{B} as a function of the excess generation γ and the relative share a of wind-power generation. The contour plot reveals that E_{B} has a minimum at $a \approx 0.8$. Compared to its storage counterpart at $a \approx 0.6$, this optimal mix is different and shifted towards a larger fraction of wind power generation.

In order to understand this, we introduce the average daily profiles $l(t)$, $w(t)$, $s(t)$ of load, wind and solar power generation. For each $0 < t \leq 24$ hour they are averaged over all available eight years. For example, $l(t) = (1/N_i) \sum_i L(t + 24i)$ with i running through all $N_i = 8{\times}365$ days within the 8 years. Together with selected combinations of the form $aw(t) + (1 - a)s(t)$ these profiles are illustrated in Figure 3.22a. The load and solar-power generation profile show the expected behaviour. Although small, but visible, the profile $w(t)$ indicates more wind-power generation during night than during day-time hours; see also the respective time series of Figure 3.10c, where this effect is also observed. The difference

$$E_{\mathrm{profile}} = \frac{1}{2 \cdot 24} \int_0^{24} |l(t) - [aw(t) + (1 - a)s(t)]| \, dt \tag{3.9}$$

of the profiles represents a measure for the balancing needs. Figure 3.22b shows it as a function of the wind-power fraction a and reveals a minimum at $a = 0.92$. By looking

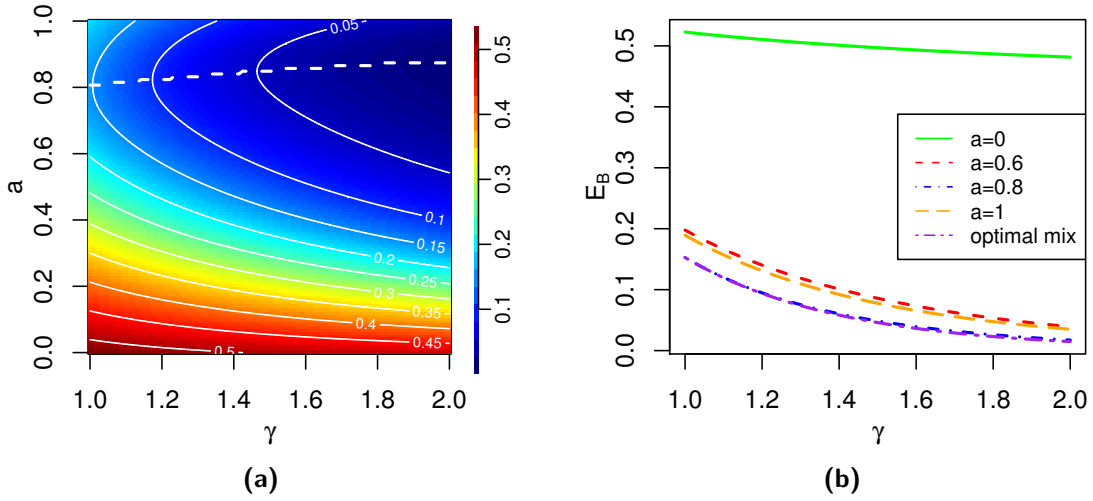


Figure 3.21.: *(a) Balancing energy, defined in Equation (3.8), as a function of the excess generation γ and the share a of wind-power generation. The contour lines represent constant balancing energy and their attached numbers are measured in annual load. (b) Cuts through (a) at $a = 1.0$ (yellow), 0.8 (blue), 0.6 (red), 0.0 (green), and along the dashed optimal-mix line (purple).*

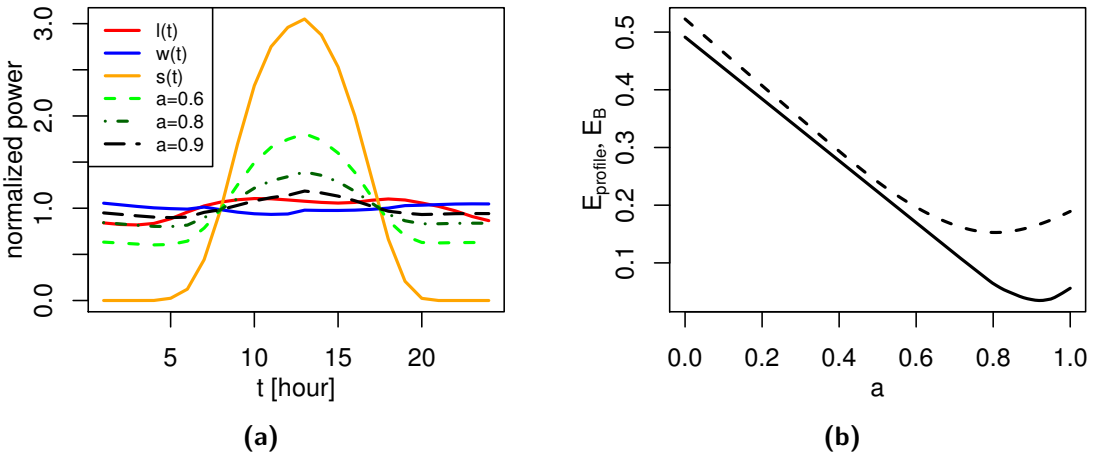


Figure 3.22.: *(a) Average daily profiles of (red) load, (blue) wind power generation, (yellow) solar power generation, (dashed green) 60%/40%, (dash-dotteded dark-green) 80%/20% and (long-dashed black) 90%/10% mix of wind/solar power generation. The profiles have been averaged over 8 years and are normalized to one. (b) Balancing integral, see Equation (3.9), as a function of the share a of wind-power generation. The excess generation has been fixed to $\gamma = 1$. For comparison, the balancing energy from Equation (3.8) with $\gamma = 1$ is shown as the dashed curve.*

again at the profiles of Figure 3.22a this outcome is obvious. The profile with $a = 0.9$ is closest to the load profile.

Figure 3.22b also compares $E_{\text{profile}}(a)$ with the balancing energy $E_{\text{B}}(a, \gamma = 1)$. Up to $a \approx 0.6$ both quantities are almost identical, but beyond that they differ. At $a = 0$ the solar-only profile exceeds the load profile during day times and is zero during night times. With other words, no balancing is needed for half of the day and full balancing is needed for the other half of the day. This explains $E_{\text{profile}}(a = 0) \approx E_{\text{B}}(a = 0, \gamma = 1) \approx 0.5$, which is half of the average load. The small difference between the two quantities is due to fluctuations of the solar power generation, which are present on hourly, daily and seasonal time-scales. In the other limit, $a = 1$, only fluctuations caused by wind power generation are present. They have a big impact on the balancing energy. $E_{\text{B}}(a = 1, \gamma = 1) = 0.19$ is about three times as large as $E_{\text{profile}}(a = 1) = 0.06$. Furthermore, the minimum of $E_{\text{profile}}(a)$ at $a = 0.92$ is shifted to $a = 0.81$ for $E_{\text{B}}(a, \gamma = 1)$.

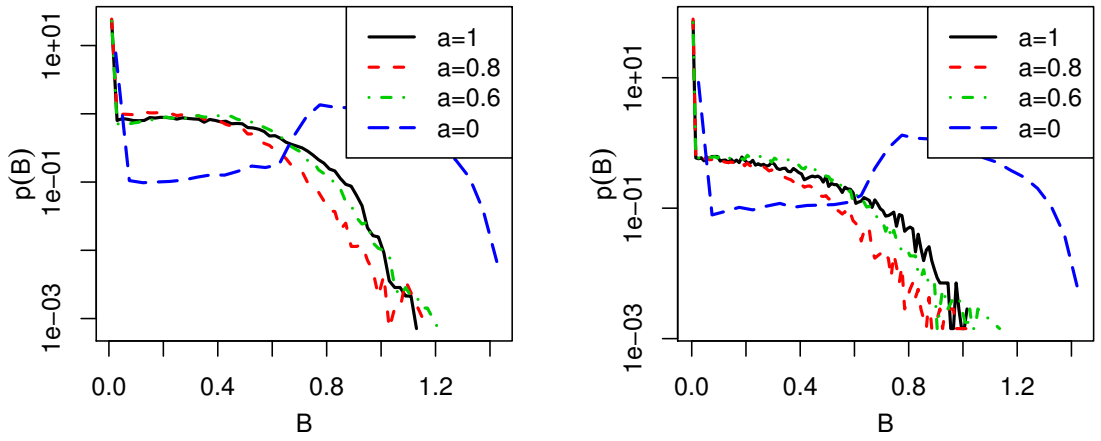
Figure 3.21b shows the γ dependence of the balancing energy for various fixed a. $a = 1$ and $a = 0.6$ lead to an almost identical behaviour. The solar-only $a = 0$ leads to balancing energies, which remain close to the previously discussed 50% of the average load at $\gamma = 1$. This is much larger than for the $0.6 \leq a \leq 1.0$ cases. Along $a = 0.8$ the balancing energy decreases from 15% of the average load at $\gamma = 1$, via 10% at $\gamma = 1.18$, to 5% at $\gamma = 1.49$.

Considering again the annual European load of 3240TWh of the year 2007, 15%, 10% and 5% translate into required balancing energies of 480, 320 and 160 TWh, respectively. A balancing energy of 100 TWh, which roughly corresponds to the capacity of the Scandinavian storage lakes, would require an excess generation $\gamma = 1.72$ at $a = 0.8$.

3.6.3. Balancing Power

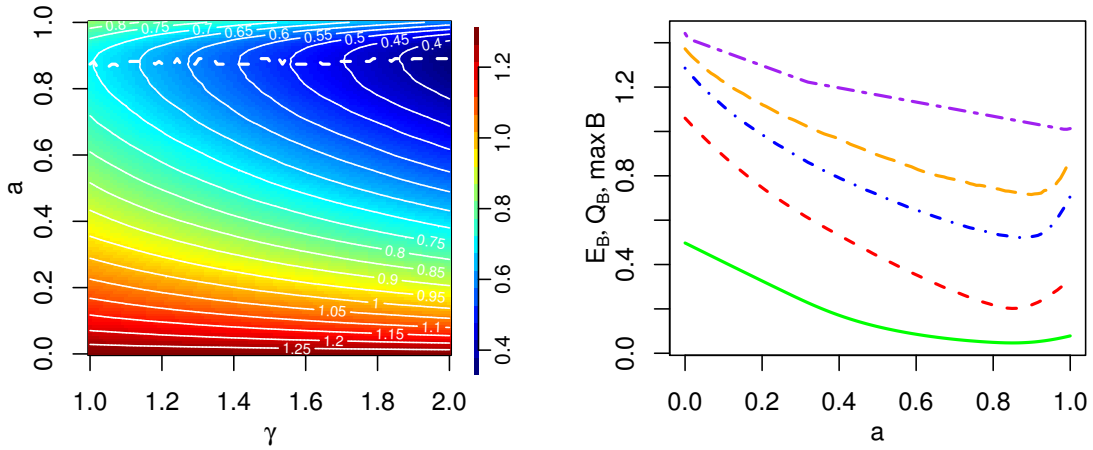
The estimates of Sections 3.6.1 and 3.6.2 have been concerned with the required amount of storage or balancing energy. An equally important characteristic to look at is balancing power. The driving question is: how large can the hourly mismatch get, and how often does this occur? In more general terms, how does the statistics of the balancing power from Equation (3.4) look like, and how does it depend on a and γ? The answers will help to complete the quantification of the respective storage technology needs.

Figure 3.23 illustrates the probability distributions $p(B)$ of the hourly balancing power in Equation (3.4) for selected combinations of a and γ. They have been sampled over the whole eight years of available data. All distributions have a pronounced peak at $B = 0$. For $\gamma = 1$ these peaks have a probability mass $0.47, 0.49, 0.46, 0.39$ at $a = 1, 0.8, 0.6, 0$, indicating that for a little less than half of the time balancing is not needed. For $\gamma = 1.5$ we find $0.73, 0.79, 0.70$ for $a = 1, 0.8, 0.6$, indicating that balancing is needed for only about 25% of the time. In case of the solar-only $a = 0$, the distributions almost do not depend on γ. During day hours the solar power generation exceeds the load, which



(a) $\gamma = 1.0$ **(b)** $\gamma = 1.5$

Figure 3.23.: *Probability distributions $p(B)$ of hourly balancing power from Equation (3.4) for $a = 1.0$, 0.8, 0.6, 0.0. The probability distributions have been sampled over eight years.*



(a) **(b)**

Figure 3.24.: *(a) 99% hourly-balancing quantile Q_{B} as a function of a and γ. (b) From top to bottom maximum balancing $\max B(t)$, 99.9%, 99% and 90% quantiles Q_{B}, as well as the average balancing energy E_{B} as a function of a for fixed $\gamma = 1.5$.*

again explains the peak at $B = 0$ with probability mass ≈ 0.4. During night hours no sun is shining, so that independent of the value for γ full balancing prevails. The maximum balancing results to be $\max B = 1.4$ times the average hourly load. The maximum balancing for the wind-only $a = 1$ is smaller. It turns out to be $\max B = 1.12$ for $\gamma = 1$ and decreases a little to 1.01 for $\gamma = 1.5$. The probability distributions for $a = 1$ result to be very similar to those with $a = 0.6$. For a values inbetween these two, like $a = 0.8$, the tail of the distribution is shifted slightly to the left.

γ	a	$E_{\mathrm{H}}^{\eta=1}$	$E_{\mathrm{H}}^{\eta=0.6}$	E_{B}	$Q_{\mathrm{B}}^{q=0.9}$	$Q_{\mathrm{B}}^{q=0.99}$	$Q_{\mathrm{B}}^{q=0.999}$
1	0.6	0.101	-.-	0.198	0.554	0.807	0.973
	0.7	0.116	-.-	0.166	0.501	0.759	0.931
	0.8	0.147	-.-	0.153	0.465	0.718	0.899
	0.9	0.182	-.-	0.163	0.479	0.708	0.877
1.25	0.6	0.014	0.069	0.129	0.452	0.724	0.897
	0.7	0.013	0.043	0.100	0.386	0.668	0.849
	0.8	0.017	0.050	0.084	0.337	0.624	0.817
	0.9	0.034	0.078	0.089	0.343	0.610	0.788
1.5	0.6	0.005	0.019	0.087	0.356	0.649	0.834
	0.7	0.005	0.015	0.063	0.280	0.588	0.784
	0.8	0.007	0.018	0.049	0.218	0.541	0.745
	0.9	0.015	0.037	0.050	0.218	0.526	0.716

Table 3.1.: *Storage energy E_{H}, balancing energy E_{B}, and hourly 90%, 99%, 99.9% balancing quantiles Q_{B} for $a = 0.6$, 0.7, 0.8, 0.9 and $\gamma = 1$, 1.25, 1.5. The storage and balancing energies are normalized to the average annual load. The balancing quantiles are normalized to the average hourly load.*

The average of these distributions has already been shown in Figure 3.21. Figure 3.24a now illustrates the 99% ($q = 0.99$) quantile Q_{B} as a function of a and γ. It is defined as $\int_0^{Q_{\mathrm{B}}} p(B)dB = q$. For every γ its minimum comes at $a = 0.88$. Along this minimum line, the 99% quantile $Q_{\mathrm{B}}(a = 0.88, \gamma)$ decreases linearly with γ, from 70% of the average hourly load at $\gamma = 1$ down to 35% at $\gamma = 2$. These rather large values guarantee that in 99% of the time the hourly balancing power remains below these quantiles. With other words, in 1% of the time, which is equivalent to 700 hours within 8 years or 1.7 hours per week, the hourly balancing power will be larger. Figure 3.24b depicts other quantiles for $\gamma = 1.5$. The quantile with $q = 0.999$ means that only nine hours within one year result in a larger balancing power. The optimal a, which minimizes the quantile, increases with q. For $q = 0.9$, 0.99 and 0.999, it turns out to be $a = 0.85$, 0.89 and 0.90, respectively.

Table 3.1 summarizes the hourly 90%, 99%, 99.9% balancing quantiles for all combinations of $a = 0.6$, 0.7, 0.8, 0.9 and $\gamma = 1$, 1.25, 1.5. The unit is in average hourly load. Given again the annual European consumption of 3240 TWh in the year 2007, the average hourly load amounts to 370 GW. This translates for example the 99% quantiles with $a = 0.6$, $\gamma = 1$ and 1.5 into 300 and 240 GW, respectively.

3.7. Separation of Time Scales

When looking at Figures 3.19, 3.20, 3.21a, and 3.24a, we observe that storage energy, balancing energy and balancing quantile come with differing optimal combinations

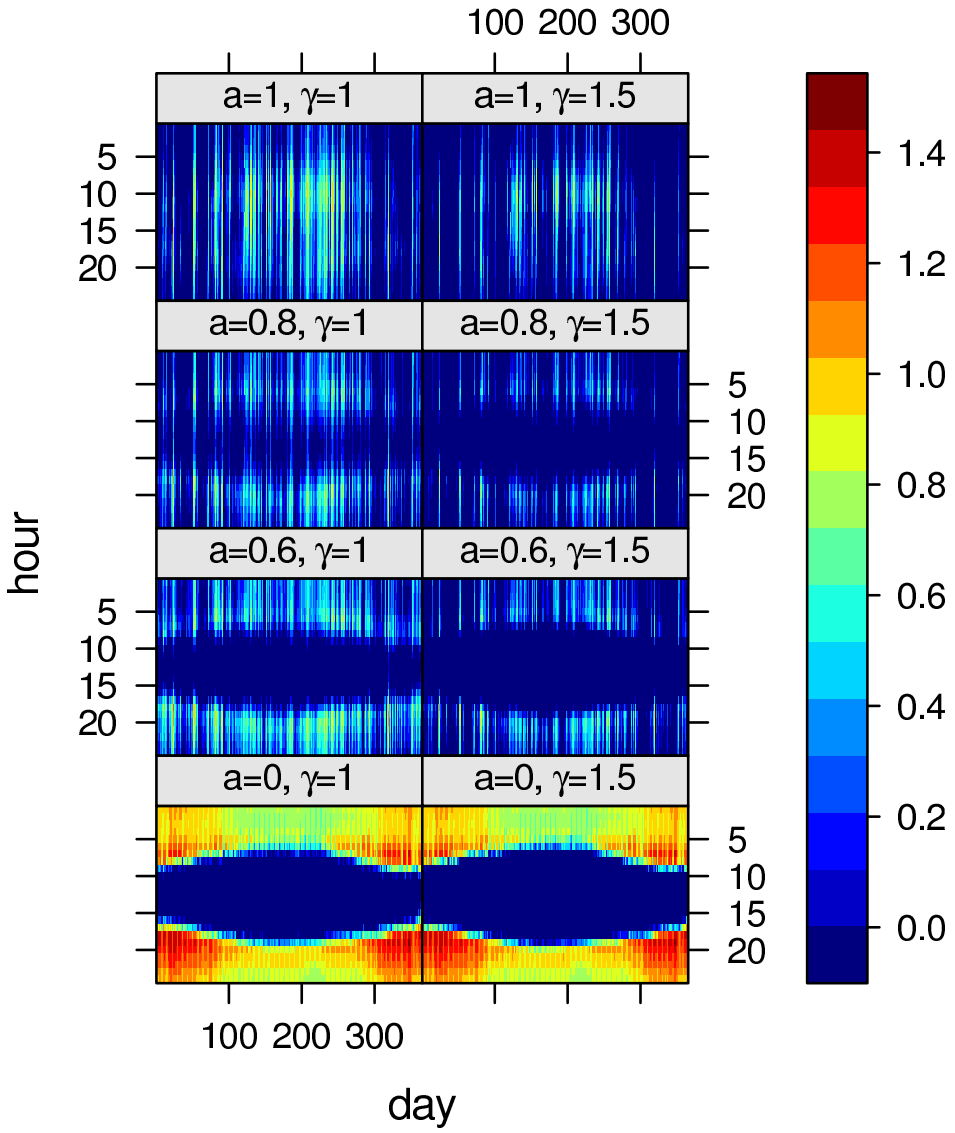


Figure 3.25.: *Fluctuation pattern of balancing power over all hours (y-axis) and days (x-axis) within one arbitrary year. Eight different combinations for a and γ are shown. The unit of the balancing power is given in average hourly load.*

between wind and solar power generation; consult also Table 3.1. The minimum of E_{H} is at $a \approx 0.5 - 0.7$, the minimum of E_{B} is at $a \approx 0.8$, and the minimum of the 99% quantile Q_{B} is at $a \approx 0.9$. Of course this can be seen as the result of three different optimization objectives, but from an infrastructure investor's point of view it would have been nicer if the three outcomes had agreed more.

The distributions of the hourly balancing power, shown in Figure 3.23, do not provide information on when exactly which amount of balancing power is needed. Although neglecting fluctuations, the average daily profiles of Figure 3.22a contain at least some information: for $a \leq 0.92$ it is likely that more balancing is needed during night hours

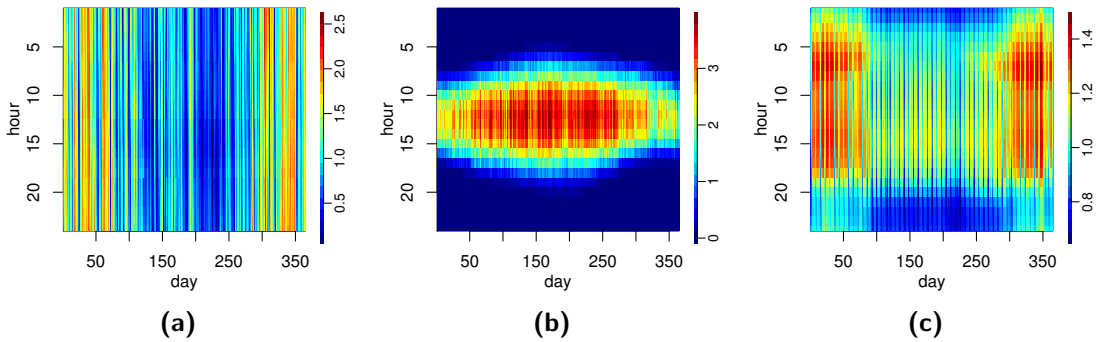


Figure 3.26.: *Fluctuation pattern of (a) wind power generation, (b) solar power generation and (c) load over all hours and days within one arbitrary year.*

than during daytime hours; for $a \geq 0.92$ it is the opposite. More information is given by Figure 3.25, which depicts the fluctuation pattern of balancing power over all hours and days of one full year for various combinations of a and γ.

These patterns result from combinations of respective patterns for wind power generation, solar power generation and load. The latter three are illustrated in Figure 3.26. Since they are very different from each other, different combinations of the form as given in Equation (3.1) lead to different fluctuation patterns for the balancing power, defined in Equation (3.4). As can be seen in the first row of Figure 3.25, the limit $a \approx 1$ is dominated by the wind pattern of Figure 3.26a. Mostly during groups of low-wind summer days balancing is needed, for almost all 24 hours per day. The limit $a \approx 0$ is shown in the last row of Figure 3.25. It is dominated by the solar pattern of Figure 3.26b. During daylight times absolutely no balancing is required. During the entire night the balancing equals the full load. The fluctuation patterns in the intermediate regime $a \approx 0.6$ are inbetween the two extremes. The third row of Figure 3.25 reveals that again absolutely no balancing is needed during daylight times. During nighttime hours balancing is needed to some degree, not all the time up to full load, but with reduced amplitude over many small groups of consecutive days, which are more or less continuously spread over the whole year. The fluctuation patterns for $a = 0.8$ are similar to those for $a = 0.6$, with two small exceptions: for $\gamma \approx 1$ some balancing is occasionally also needed during daylight times, and for larger $\gamma \approx 1.5$ the reduced balancing during nighttime hours is occurring with fewer and over the year a little bit more intermittently distributed day groupings.

The fluctuation patterns of Figure 3.25 reveal, that for $0 \leq a \leq 0.9$ the power mismatch shows a regular intra-day behavior. During daytime only excess power is generated. Negative mismatches occur only during nighttime. Because of this, it makes sense to introduce a separation of times scales into the storage and balancing dynamics, i.e. a dynamics above one day and a dynamics below one day. In the following Subsection we will look at the dynamics above one day.

γ	a	$E_{\mathrm{H}}^{\eta=1}$	$E_{\mathrm{H}}^{\eta=0.6}$	E_{B}	$Q_{\mathrm{B}}^{q=0.9}$	$Q_{\mathrm{B}}^{q=0.99}$	$Q_{\mathrm{B}}^{q=0.999}$
1	0.6	0.100	-.-	0.094	0.292	0.549	0.648
	0.7	0.115	-.-	0.110	0.331	0.557	0.666
	0.8	0.147	-.-	0.129	0.379	0.583	0.695
	0.9	0.182	-.-	0.151	0.436	0.633	0.745
1.25	0.6	0.014	0.034	0.028	0.103	0.388	0.503
	0.7	0.013	0.025	0.037	0.156	0.407	0.526
	0.8	0.016	0.035	0.055	0.221	0.447	0.570
	0.9	0.034	0.071	0.077	0.297	0.528	0.635
1.5	0.6	0.004	0.010	0.009	0.000	0.246	0.364
	0.7	0.005	0.009	0.012	0.000	0.275	0.395
	0.8	0.006	0.011	0.021	0.074	0.316	0.444
	0.9	0.015	0.035	0.038	0.164	0.438	0.540

Table 3.2.: *Same as Table 3.1, but based on daily instead of hourly mismatches.*

So far the wind power generation $W(t)$, the solar power generation $S(t)$, the load $L(t)$, the mismatch $\Delta(t)$ and the balancing power $B(t)$ have been modeled with a one-hour time resolution. Now the time resolution is changed to one day. The daily wind power generation then is

$$\overline{W}(\tau) = \frac{1}{24} \sum_{t=24(\tau-1)+1}^{24\tau} W(t) \; ; \tag{3.10}$$

it is divided by 24 in order to normalize its average to one. The time τ proceeds in steps of one day. The daily solar power generation $\overline{S}(\tau)$ and the daily load $\overline{L}(\tau)$ are obtained in complete analogy. Contrary to the daily mismatch

$$\begin{aligned} \overline{\Delta}(\tau) &= \gamma \left[a\overline{W}(\tau) + (1-a)\overline{S}(\tau) \right] - \overline{L}(\tau) \\ &= \frac{1}{24} \sum_{t=24(\tau-1)+1}^{24\tau} \Delta(t) \; , \end{aligned} \tag{3.11}$$

the daily balancing power can not be calculated in two ways. Its proper definition is analogous to Equation (3.10):

$$\overline{B}(\tau) = \begin{cases} -\overline{\Delta}(\tau) & \text{if } \overline{\Delta}(\tau) < 0 \\ 0 & \text{otherwise} \end{cases} \; . \tag{3.12}$$

This expression is not identical to the daily average of the hourly balancing power. Actually Equation (3.12) is smaller than $(1/24) \sum_{t=24(\tau-1)+1}^{24\tau} B(t)$, because the latter does not take into account the compensating positive mismatches occurring over the day.

Based on the daily mismatch $\overline{\Delta}(\tau)$ and the daily balancing power $\overline{B}(\tau)$, the storage energy, the balancing energy and the balancing quantiles are determined completely

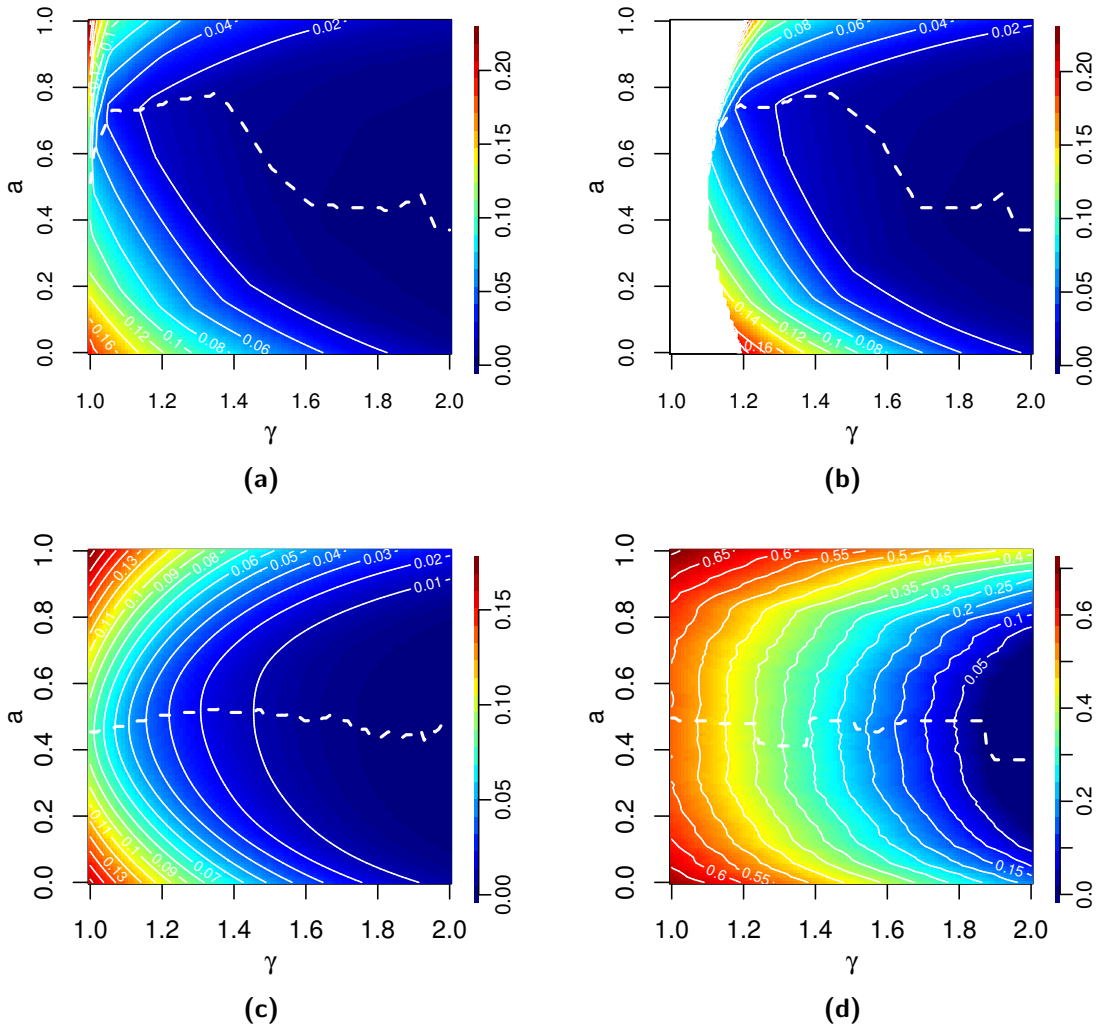


Figure 3.27.: *Storage energy E_H with (a) $\eta_{in} = \eta_{out} = 1$ and (b) $\eta_{in} = \eta_{out} = 0.6$, (c) balancing energy E_B and (d) 99% balancing quantile Q_B based on the daily mismatch,defined in Equation (3.11) and balancing power from Equation (3.12). The unit of the contour lines for the storage and balancing energies is the average annual load. The unit of the contour lines for the balancing quantile is the average hourly load.*

analogous to the previous Sections. The results are shown in Figure 3.27 as a function of γ and a. Figure 3.27a and Figure 3.19a are almost indistinguishable, which means that the ideal ($\eta_{in} = \eta_{out} = 1$) storage energies based on hourly and daily mismatches are almost identical; compare also the third columns of Tables 3.1 and 3.2. As can be seen from the storage time series of Figure 3.18, the storage energy is only determined by fluctuations on the synoptic and seasonal time scales, which are larger than one day. For hydrogen storage with reduced efficiencies $\eta_{in} = \eta_{out} = 0.6$ the storage energies based on daily and hourly mismatches are not identical; compare Figure 3.27b and

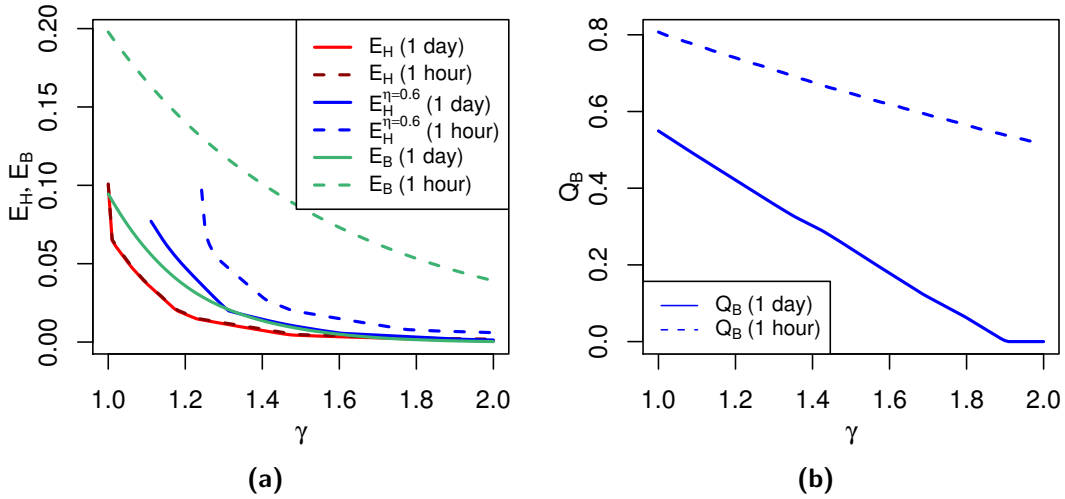


Figure 3.28.: *Comparison of the differing impacts between (dashed) hourly and (solid) daily mismatches in Equations (3.1) and (3.11) on (a) storage energy E_{H} with (red) $\eta_{\mathrm{in}} = \eta_{\mathrm{out}} = 1$ and (blue) $\eta_{\mathrm{in}} = \eta_{\mathrm{out}} = 0.6$, (green) balancing energy E_{B}, and (b) (blue) 99% balancing quantile Q_{B}. The unit of the storage and balancing energies is given in average annual load. The unit of the balancing quantiles is given in average hourly load. The parameter a has been fixed to $a = 0.6$.*

Figure 3.20a. They are similar for $0.85 \leq a \leq 1$, but differ for $0 \leq a \leq 0.85$. This difference is specified in the fourth columns of Tables 3.1 and 3.2. For $a = 0.6$ the storage energy based on daily mismatches is only about half of the storage energy based on hourly mismatches.

Figures 3.27c and 3.27d illustrate the balancing energy and the 99% balancing quantile based on the daily mismatches. They are completely different to their counterparts based on the hourly mismatches; consult again Figures 3.21a and 3.24a. The new minima are now found at $a \approx 0.5$. This result is independent of γ and coincides nicely with the $a \approx 0.6$ obtained for the storage energy. From an infrastructure perspective this is good news: based on the daily mismatches the optimal mixes between wind and solar power generation for storage energy, balancing energy and high quantiles of balancing power are almost the same at around $a = 0.5 - 0.6$.

Compared to the hourly mismatches, the balancing energy and the quantiles of balancing power based on daily mismatches are greatly reduced. Compare Figures 3.27c and 3.27d with Figures 3.21a and 3.24a. Table 3.2 lists the obtained values for daily-based E_{B} and Q_{B} for various combinations of γ and a. The E_{B} values 0.094, 0.028, 0.009% for $a = 0.6$ and $\gamma = 1$, 1.25, 1.5 are significantly smaller than the respective 0.153, 0.084, 0.049 for the hourly-based E_{B} at its minimizing $a = 0.8$. A similar finding is obtained for the balancing quantiles, where for example the 52.6% and 71.6% of the hourly-based 99% and 99.9% quantiles Q_{B} at $\gamma = 1.5$, $a = 0.9$ are reduced by a factor of two down to 24.6% and 36.4% for the respective daily-based quantiles at $a = 0.6$.

Figure 3.28 summarizes the results obtained in Figure 3.27 and Table 3.2 in a different way. For fixed $a = 0.6$ it compares the γ dependence of the smaller daily-based with the larger hourly-based storage energies, balancing energies and balancing quantiles. With this separation of time scales we anticipate a separation into a short-term and long-term storage.

The long-term storage takes care of the daily mismatch. Over the day its hourly contribution would be more or less constant and sum up to the daily mismatch. Its storage/balancing energy and power quantiles would correspond to those based on the daily mismatches. With another look into Table 3.2 at for example $\gamma = 1.5$ and $a = 0.6$ the required numbers would be $E_{\mathrm{H}}(\eta = 1) = 0.004$ and $Q_{\mathrm{B}}(q = 0.99) = 0.246$ for ideal round-trip storage, which, given the annual European load of 2007, translate into $E_{\mathrm{H}}(\eta = 1) = 15\mathrm{TWh}$ and $Q_{\mathrm{B}}(q = 0.99) = 90\mathrm{GW}$. For hydrogen storage the respective numbers are $E_{\mathrm{H}}(\eta = 0.6) = 0.010 = 35\mathrm{TWh}$ and $Q_{\mathrm{B}}(q = 0.99) = 0.246 = 90\mathrm{GW}$. For one-way storage reservoirs like storage lakes the respective numbers are $E_{\mathrm{B}} = 0.009 = 30\mathrm{TWh}$ and $Q_{\mathrm{B}}(q = 0.99) = 0.246 = 90\mathrm{GW}$.

The short-term storage takes care of the hourly mismatch around the daily mismatch. Roughly, its required balancing power corresponds to the difference between the quantile based on the hourly mismatch and the quantile based on the daily mismatch. Again for $\gamma = 1.5$ and $a = 0.6$, this results in $Q_{\mathrm{B}}^{\mathrm{hourly}}(q = 0.99) - Q_{\mathrm{B}}^{\mathrm{daily}}(q = 0.99) = 0.40 = 140\mathrm{GW}$. A safe upper bound for the required storage energy for a roundtrip storage would then be $E_{\mathrm{H}}^{\mathrm{short-term}} = 140\mathrm{GW} \times 12\mathrm{h} = 1.68\mathrm{TWh} < 2\mathrm{TWh}$. The multiplication with 12h is due to the night hours, where balancing is needed; see again Figure 3.25. Candidates for such a short-term storage would be pumped hydro, electric cars, compressed air, and any combination thereof.

3.8. Summary

Besides short term fluctuations, wind and solar power generation across Europe follow the seasonal cycle of the weather. Wind power generation in winter is much stronger than during summer. For solar power generation the summer season produces much larger yields than during winter. In this way mother nature determines how to design a future European power supply system based on a very high share of renewables. When mixed together in a specific ratio, the opposite strong seasonal behaviors of wind and solar power generation almost cancel each other and follow the weaker seasonal load behavior.

Depending on the objectives to minimize storage energy, balancing energy or balancing power, different optimal mixes between wind and solar power generation have been found. In case of storage energy, the optimal mix is 60% wind and 40% solar power generation for ideal roundtrip storage, and 70% wind and 30% solar power generation for hydrogen storage. In case of balancing energy, the optimal mix is 80% wind and

20% solar power generation, for balancing power it is 90% wind and 10% solar power generation. Furthermore, all optimal mixes turn out to be more or less independent from the amount of excess power generation. For transitional scenarios with a high fraction of fossil-nuclear power generation left in the system, the share of wind power generation increases when minimizing the required storage size and decreases when minmizing balancing energy compared to the only renewable scenario. Note, that the different optimal mixes are mainly caused by the intra-day mismatch dynamics between wind plus solar power generation and load. Once the intra-day time scales are neglected, the optimal mixes for storage energy, balancing energy and balancing power would become identical at around 60% wind and 40% solar power generation. The storage energy for all of Europe amounts to around 10% of the average annual load. Compared to other scenarios like wind-only or solar-only, the optimal mix reduces the need for storage energy by a factor of two.

With no excess wind plus solar power generation, the required storage needs turn out to be very large. However, they decrease very fast with the introduction of excess power generation. In the following we list the concrete numbers for the required storage needs, given the European consumption load of 3240 TWh for 2007; consult again with Table 3.1. With the objective to minimize the storage energy, the required needs for roundtrip storage with ideal efficiencies $\eta_{\text{in}} = \eta_{\text{out}} = 1$ amount to $E_{\text{H}} = 320$ TWh storage energy and $Q_{\text{B}}(q = 0.99) = 300$ GW balancing power for $\gamma = 1$ and $a = 0.6$. For excess generation with $\gamma = 1.5$ and $a = 0.6$ the numbers reduce to $E_{\text{H}} = 16$ TWh storage energy and $Q_{\text{B}}(q = 0.99) = 240$ GW. However, the installed wind and solar-photovoltaic power capacities across Europe would each increase from 750 to 1100 GW. For comparison, hydrogen storage with non-ideal efficiencies $\eta_{\text{in}} = \eta_{\text{out}} = 0.6$ would require around 40 TWh and 220 GW for minimum storage energy and non-minimum balancing power at $\gamma = 1.5$ and $a = 0.7$ with installed 1300 GW wind and 830 GW solar power capacities. If we were to choose the other objective to minimize the balancing power, then the optimal $a = 0.9$ at $\gamma = 1.5$ leads to required 50 TWh (ideal) and 120 TWh (hydrogen) storage energies, and 195 GW balancing power, with installed 1650 GW wind and 275 GW solar power capacities.

For one-way balancing storage the two objectives to minimize either the balancing energy or to minimize the balancing power lead to quite similar results. The optimal share of generated wind power amounts to $a \approx 0.8$-0.9. At zero excess generation $\gamma = 1$ the required European balancing energy and balancing power result to be 500 TWh and 265 GW. They are reduced down to 160 TWh and 200 GW once the excess generation is increased to $\gamma = 1.5$.

The presented results demonstrate that excess wind and solar power generation will be one key to reduce the required storage needs for a fully renewable European power system. In fact, a fully renewable power system is only fully defined when including of a reasonable amount of excess generation. However, with a reasonable amount of excess wind and solar power generation alone the resulting storage energies, balancing energies and balancing powers are still very large. With 50% excess generation, roundtrip storage

still comes with a required storage energy of the order 20-50 TWh and balancing power of the order 200 GW. For one-way balancing storage it will be 160 TWh storage energy and 200 GW balancing power.

With the considered storage technologies, central storage facilities seem to be unavoidable due to the high energy that needs to be storage in a storage only scenario. These facilities are required to provide high capacities for balancing power and can be expected to cause large power flows. A paradigm change so that the load follows the intra-day generation could help to overcome these problems. A shift of consumption to day time and less load during night time would reduce both the required balancing power, see Figure 3.25, and the extreme transport flows, as will be shown in Section 6.1. However, this approach highlights the need for accurate prediction of wind and solar power for at least 24 hours. Local, distributed storage facilities with limited storage energy also have a great potential to match the generation and consumption over a day.

A good understanding of possible future scenarios has to be the basis for deciding on efficient transitions from the existing power system towards a fully renewable scenario. It has to be noted, that the two extreme scenarios, only storage or only balancing power, are discussed here. Both scenarios, were shown to be feasible. A mix of both approaches potentially further reduces the storage and the balancing needs.

4. Advantages of Cooperation

In Chapter 3, weather data driven time series of wind and solar photovoltaic power generation and load time series have been analyzed with respect to European storage and balancing needs. Implicitly a European cooperation was assumed, with the advantage that the pooling over this large area can be expected to help counterbalancing spatially localized fluctuations. Another scenario is, that the European countries do not cooperate and implement strategies that are optimal for their specific needs. In Section 4.1, optimal mixes for individual countries are presented and in Section 4.2 compared to the cooperative scenario. For simplicity only the one hundred percent scenario, as introduced in Section 3.4, is considered. In that case, $c = 0$, $\gamma = 1$, and $a = 1 - b$ is used, see Equation (3.1). The data with the one-hour temporal resolution is analyzed and only ideal storage with $\eta_{in} = \eta_{out} = 1$ is considered, see also Section 3.4.

4.1. Optimal Mixes for Countries

We analyze the optimal mixes for the individual countries. The respective wind and solar generation and load data are aggregated over the sets of regions of each country. The respective region sets, along with the abbreviations for each country, are shown in Figure 4.1. The optimal mixes are discussed in detail for Germany, Ireland, Spain, and Greece. For comparison, the all-European scenario, based on an uniform optimal mix for all countries and power exchange, results for storage and balancing energies and power are again shown in Figure 4.2.

For storage, the results for each country are shown in Figure 4.3. The values are normalized to the average annual consumption of the respective country. The optimal mix varies, between almost zero and $a \approx 0.8$. For Greece, the storage energy for $a = 1$ is around a factor five higher than in the optimal mix at $a \approx 0$. For Ireland on the other hand, the optimal mix is almost $a = 0.8$. The other countries considered here are in between these cases.

When considering the balancing energy and quantiles, as shown in Figure 4.4, only a small variation of a is found for the optimal mix, with $a \approx 0.6$ and up to $a \approx 1$ for the higher quantiles. All countries exhibit almost the same behavior when looking at the average values. With increasing quantiles, the minimum is less pronounced, indicating that extreme events cannot be counterbalanced, independent of the mix. This is a significant difference to the European scenario in Figure 4.2b.

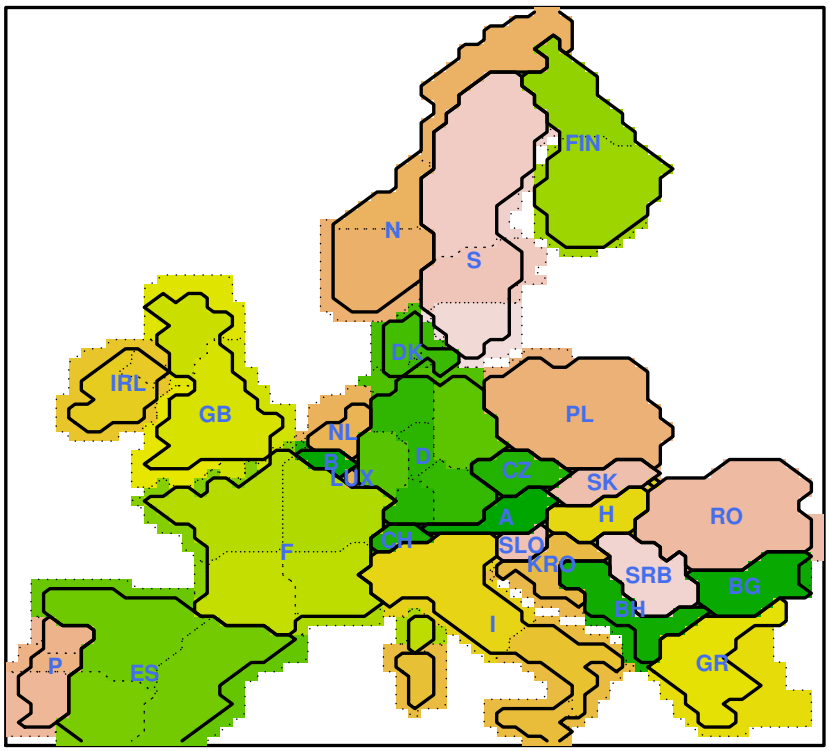
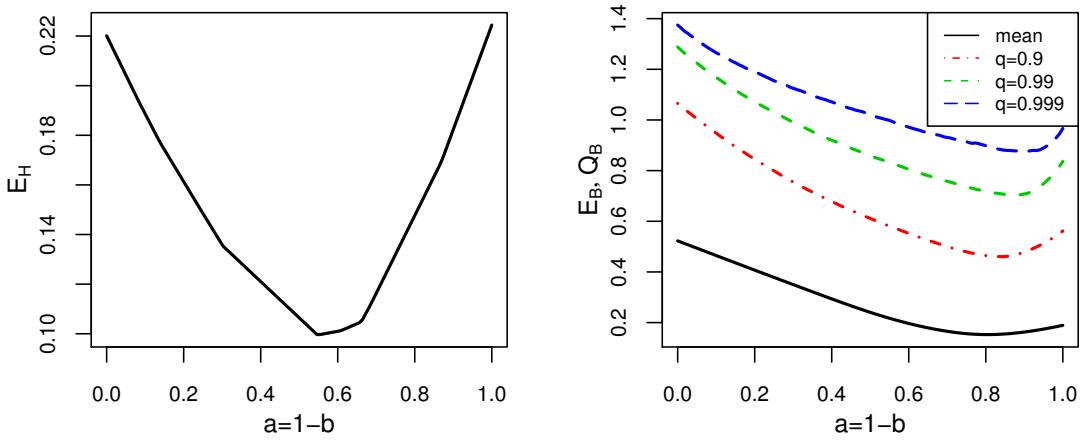


Figure 4.1.: *Map of the countries with the respective abbreviations. The colors, together with the dotted lines, indicate the regions. The offshore regions are not drawn within the country borders, for easier recognition of the map. Offshore generation is added to the respective countries.*

To explain the different behavior of the optimal mixes obtained when minimizing the storage and balancing energies, the timeseries of the generation and the load has to be examined. The seasonal characteristics of the load have a big influence on the optimal mix. For each country, the timeseries of the monthly averages are shown in Figure 4.5. Both Spain and Greece show a much smaller fluctuation strength in the photovoltaic generation compared to Germany and Ireland, consult the orange curves in Figure 4.5. For Germany and Ireland in Figure 4.5a and 4.5b, the load time series correlates strongly with the wind series. This behavior is weak for Spain, see Figure 4.5c. For Greece in Figure 4.5d, the load is rather anti-correlated to the wind generation and thus more correlated with the photovoltaic time series, see Figure 4.5d. The main reason for high electricity usage in winter is illumination and heating. In summer cooling is an important seasonal influence on the load. The winter peak can be seen, to a different extent, for all countries shown in Figure 4.5. For Spain and especially for Greece, the peak in Summer is strong, while for Germany it is small. As the optimal mix that minimizes the storage energy is dominated by seasonal timescales, $a \approx 0$ for Greece and $a \approx 0.75$ for Ireland are in agreement with the timeseries of the load and generation for the respective countries.



(a) *Storage capacity* **(b)** *Balancing capacity and power*

Figure 4.2.: *Storage energy ($\eta_{in} = \eta_{out} = 1$), balancing energy and power shown for $c = 0$ and varying a for whole Europe as discussed in Chapter 3. The values in (a) are normalized to annual average load and in (b) to hourly average load.*

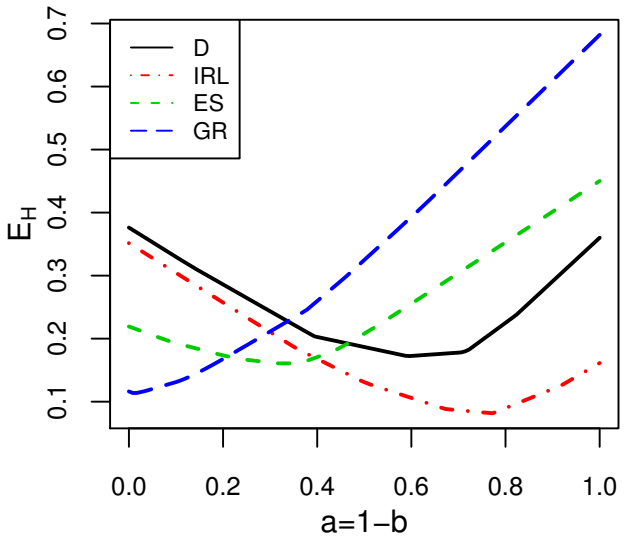
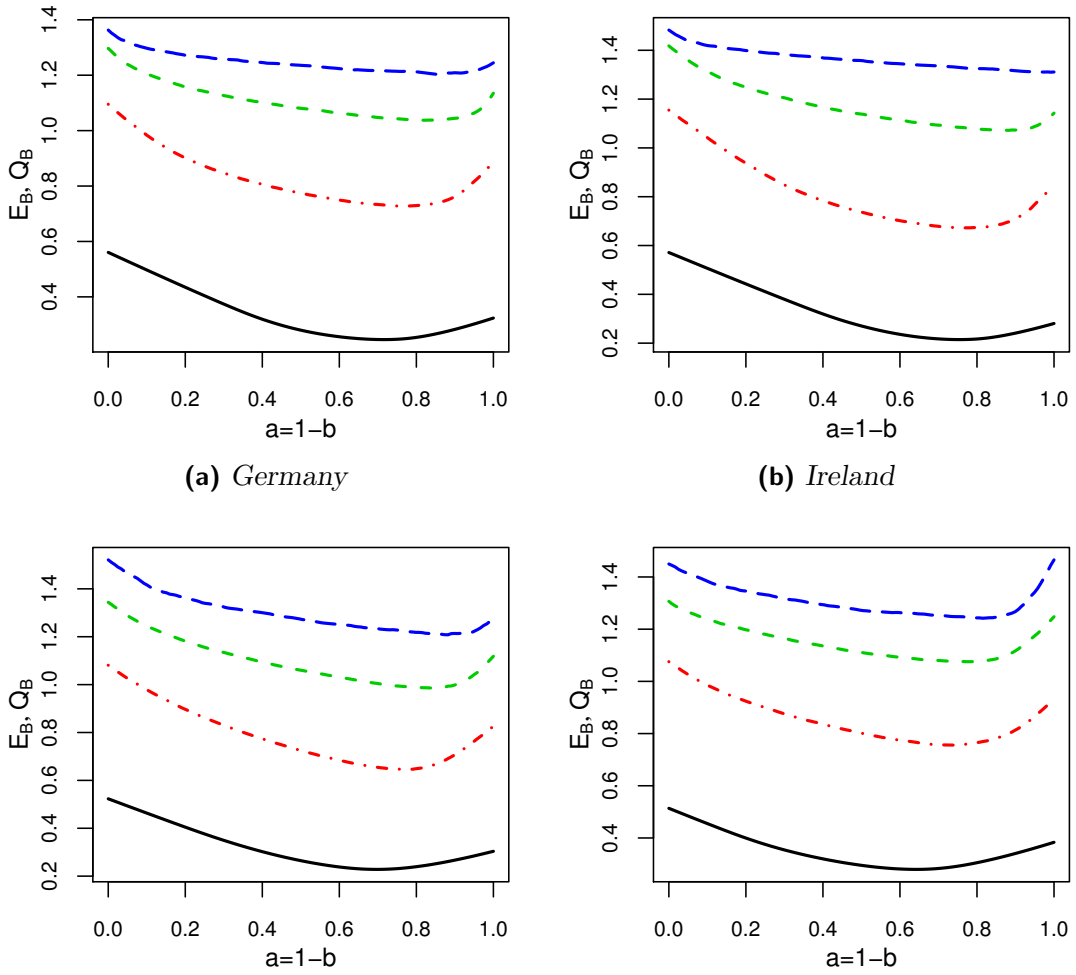


Figure 4.3.: *Required storage energies ($\eta_{in} = \eta_{out} = 1$) for selected countries. The values are normalized to the average annual consumption of the respective country.*

The balancing energy, on the other hand, is dominated by short timescales. When employing a high share of photovoltaic power generation, there is a high probability of a shortage of power during night. This is independent of the seasonal characteristics. To minimize the required balancing energy and power, the day night fluctuations have to be minimized and so the ratio between solar and wind is similar for all countries.

The observation, that the optimal mix for storage varies strongly between countries

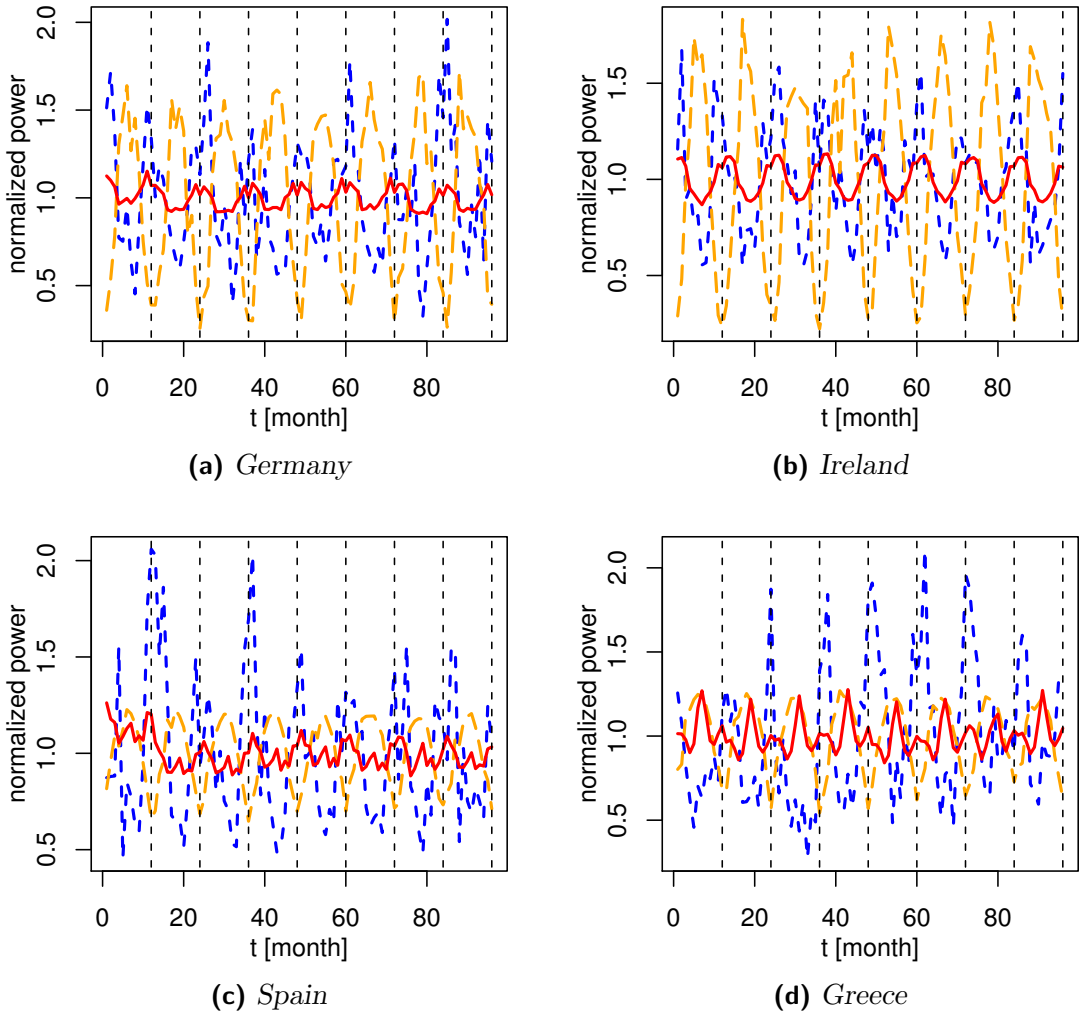


(a) *Germany* **(b)** *Ireland*

(c) *Spain* **(d)** *Greece*

Figure 4.4.: *Balancing needs for selected countries. As in Figure 4.2b the balancing energy is indicated with a solid black curve, the quantiles $q = 0.9, 0.99, 0.999$ in red, green, and blue, respectively. The quantile values are normalized to the average hourly load.*

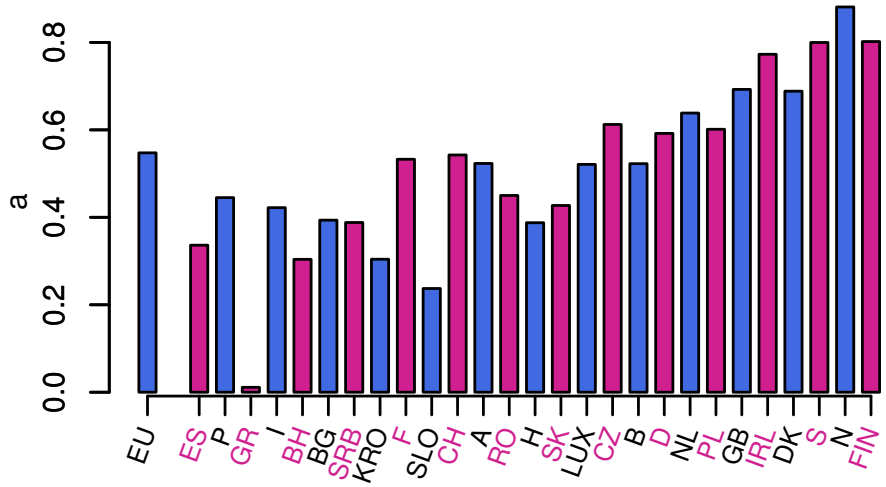
as well as to a smaller extent the optimal mix for balancing, is also found when looking at all European countries within the data set. This can be seen in Figure 4.6. An ordering from left to right roughly corresponds to an ordering from south to north. In good approximation, we find that the more south the country is located, the lower the fraction of wind power generation, when minimizing the storage energy. The same is found for balancing energy but to a much smaller extent, as shown in Figure 4.6b.



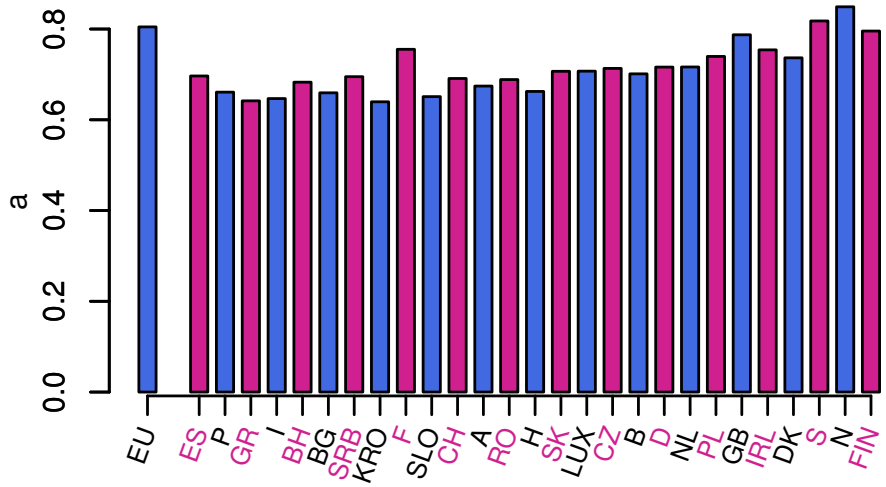
(a) *Germany* **(b)** *Ireland*

(c) *Spain* **(d)** *Greece*

Figure 4.5.: *Load (red), wind (blue), and photovoltaics (orange) power generation. Normalized time series of the monthly averages for selected countries for the years 2000 – 2008.*



(a) E_H



(b) E_B

Figure 4.6.: *Overview of the optimal mixes for each country, optimizing either (a) the required storage energy E_H or (b) balancing energy E_B. The abbreviations for the countries are shown in Figure 4.1. The countries are roughly sorted according to their latitude. Countries in the south are on the left side and northern countries on the right.*

4.2. Cooperation over Europe versus Selfish Countries

A scenario contrary to European cooperation would be the implementation of autarkic national policies. The optimal mix can be calculated for every country, as shown in the previous section. In this scenario, the load of each country has to be satisfied only with the generation within the country as well as by its balancing and storage facilities.

To asses the question, if a cooperative strategy, with power flows across borders, is to be favored over a selfish approach for each country, without power flows between the countries, the required storage and balancing energies and power are compared. These are shown in Figure 4.7. The energies and quantiles for the European case is straightforward as done in Chapter 3. The values for the storage and balancing energies and power quantiles within countries are calculated by examining each country individually using the respective optimal mixes and adding up the optimal mix values of each country.

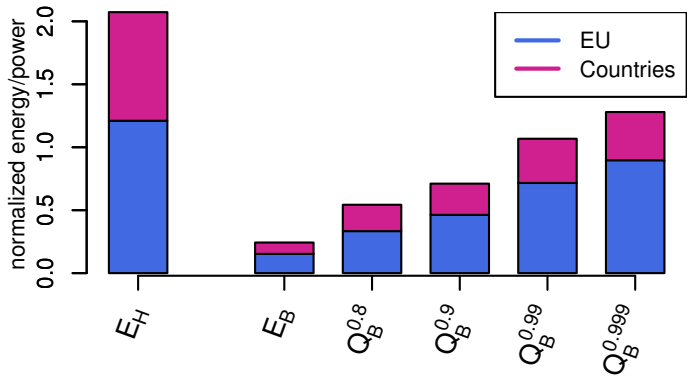


Figure 4.7.: *Comparison of a cooperative scenario (blue) versus a selfish scenario (red). In the latter countries are optimized with respect to their own needs but no power is exchanged. The cooperative scenario assumes European cooperation so that there are no limits for power flows across the borders. The storage energy is normalized to the average monthly load, whereas the reserve energies and powers to the average hourly load.*

The comparison of the different scenarios is depicted in Figure 4.7. Note, that the storage energies are normalized to the average monthly load, while the balancing energy and power quantiles are plotted relative to the average hourly load. It is obvious that the cooperation results in significantly less storage and balancing energies. This is also true for the quantiles of the balancing power. The upscaling factors for the selfish scenario relative to the cooperative case are shown in Figure 4.8. The storage energies aggregated over all countries would have to be scaled up by a factor around 1.7. The same is true for the balancing. On average 1.6 times more balancing is needed. For the quantiles above 0.9, the factor is slightly lower.

The benefits of the cooperative approach, have to be considered in the context that

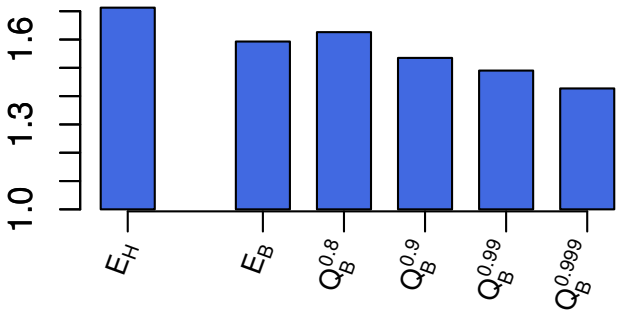


Figure 4.8.: *The ratios of the energies and power quantiles for the selfish scenario with respect to the cooperative scenario. The factor gives the required upscaling of the energies and power quantiles for the case without power flows over country borders.*

the needed capacities to transport the power have to be available, but also that storage as well as balancing energies are limited. The required transport capacities will be the topic of the next chapter.

5. Flows in Power Grids: Model

As shown in Chapter 4, power exchange between countries reduces the required storage and balancing energies. Also, the scenarios analyzed in Chapter 3 imply transport of energy. This illustrates the importance of power transmission. Sufficient transport capacities have to be installed to allow for the required power flows.

Power transport systems are large and highly complex [5]. Consequently, the study of power flows is highly nontrivial as the whole system has to be taken into account [90]. The effect of fluctuating power generation on the power flows has not been studied yet. Therefore, we propose a generic model to understand transport flows within a framework with known properties. In the next chapter, the gained insights are then tested against the data presented in Chapter 3.

The models described in Section 2.2 are not appropriate to analyze power flows resulting from fluctuating generation as these either use the shortest path transport paradigm[1] or are designed to describe specific problems with no relation to fluctuating generation. The power flow in real power systems is determined by physical laws and the specific properties of the power grid and its components. A usefull approximation to the exact power flow equations is the so-called "DC power flow" approximation. It is introduced in Section 5.1 and derived in detail in Section A.2. We introduce a model that is based on stylized facts from the data in Chapter 3. The model is defined in Section 5.2 and an overview of the results is presented in Section 5.3.

Fluctuating sources cause flows in transport networks that have a probabilistic nature. Similar to the work in Section 2.3, we estimate these probability distributions by using analytically derived approximations and propse a capacity layout. The case when generation and load between different vertices are uncorrelated is presented in Section 5.4, while the correlated case is discussed in Section 5.5.

To obtain a model that is numerically and analytically treatable, only quasi-static states are analyzed, as done in other studies [5, 26, 43]. This approach neglects all transient effects. However, when analyzing cascading failures, as done in Section 2.3, the transient dynamics can play an important role [109]. Therefore, the reliability of calculated cascading failures in our static model is unclear and we only estimate the robustness against failures of single links.

[1] The load based on shortest path transport has no significant correlation to the flow in power grids [75].

5.1. DC Power Flow Approximation

In Section 2.2, we discussed two transport paradigms that are used in the literature to quantify transport in power grids. The shortest path transport assumes that transported commodities are routed along the shortest hop path. However, it was shown [75] that this paradigm does not adequately describe power flows. A transport paradigm approximating the physical power flow equations is the so called DC-flow. Since the DC-flow is well established in the literature [see e.g. 90, 126] and widely used [9, 26, 27, 43], we chose this approach to examine transport flows in transmission systems (see also Section 2.2.2).

Here, a short outline of the implications of the DC-flow approximation is given. For a more detailed derivation consult Section A.2. AC power systems can be well described by using complex admittances Y, an extension to the laws governing DC circuits, e.g. Ohm's law $I = Y \cdot U$. It allows to apply DC-current concepts to systems with sinusoidal voltages and currents with constant frequency. The complex admittance $Y = G + jB$ is the sum of the conductance, the inverse of the resistance, and the susceptance that arises from capacitances and inductances. Because of the changing voltages and currents, reactive power has to be considered, see Section A.2.1. The full steady state power flow equations[2] can be written as

$$S_i = P_i + jQ_i = U_i \left(\sum_k \mathbf{Y}_{ik} U_k \right)^* , \tag{5.1}$$

where j denotes the imaginary unit and $U_i = |U_i| e^{j\delta_i}$ is the voltage at vertex i with phase angle δ_i. P_i and Q_i describe the active and reactive powers and S_i denotes the complex power. The entries of the complex admittance matrix $\mathbf{Y}$ consist of the admittances of the links. For known voltages, the active and reactive powers at each vertex can be calculated using Equation (5.1).

Typically, the injected or withdrawn complex power is known and the voltages phase angles have to be calculated, so that the power flows over the links can be derived. Since Equation (5.1) is second order in the voltage, it cannot be solved directly. The specific properties of power grids allow for approximations, so that the power flow equations can be greatly simplified. Based on the observation by engineers [90, 126], that the resistance of the links is small compared to the capacitance, the resistances are neglected. Furthermore, the phase angles δ_i are assumed to be small and voltage amplitudes to be uniform. Using these approximations, Equation (5.1) can be simplified to a linear system of equations

$$\vec{P} = \mathbf{B}' \vec{\delta} , \tag{5.2}$$

[2] The description in Equation (5.1) is only valid for steady states. For transient behavior, the differential equations describing the system have to be solved.

with

$$\mathbf{B}'_{ik} = B_{ik} \qquad \text{for } i \neq k \tag{5.3}$$

$$\mathbf{B}'_{ii} = -\sum_{k=1,k\neq i}^{N} B_{ik} \tag{5.4}$$

based on the susceptance B_{ik} of the link $i \to k$. This is an equation only for the active power, since changes in the reactive power vanish due to the simplifications. Because of this similarity to DC-systems, the approximation is referred to as DC-flow approximation [90, 126].

The linear system in Equation (5.2) is much easier to solve than the full power flow equations (5.1). It is used by engineers to obtain a fast approximation of the flows in the network or as initial solution when solving Equation (5.1). If all susceptances B_{ik} are set to one, the matrix $\mathbf{B}'$ is equal to the Laplace matrix $\mathbf{L}$ (see also Section A.2.4). There is also a formal analogy with the flow of currents in a resistor network that is governed by Kirchhoff's equations as discussed in Section A.2.4. This allows to apply results found for the latter to be applied to Equation (5.2), but with the conceptual difference, that the entries of the matrix are not conductances but only susceptances and that the variable on the right hand side of Equation (5.2) is the phase angle and not the voltage.

Solving Equation (5.2) for given $\vec{P}$ cannot be done directly, since $\mathbf{B}'$ has a zero eigenvalue and is thus not invertible [see e.g. 89]. A simple method is introduced in Section A.2.4.1, using the Moore-Penrose pseudoinverse, so that

$$\vec{\delta} = \mathbf{B}'^{+}\vec{P}\,, \tag{5.5}$$

with the constraint that the in- and out- flows are balanced $\sum_i P_i = 0$. The balancing condition is necessary so that Equation (5.5) is consistant with Equation (5.2). To avoid confusion with the balancing power from the previous chapters, we refer to the balancing condition as "matching the in- and out-flows". The power on the link $j \to k$ is given by [126]

$$f_{jk} = B_{jk}\left(\delta_j - \delta_k\right)\ . \tag{5.6}$$

Using the transpose of the incidence matrix $\mathbf{K}^T$, see Section 2.1.1, the power flows can be written as

$$\vec{f} = \mathbf{B}\,\mathbf{K}^T\vec{\delta}\,, \tag{5.7}$$

where $\mathbf{B}$ is a $M \times M$ diagonal matrix with the suceptances $B_{ik} = B_q$ on the diagonal and $\vec{f}$ is the vector of flows along the links. The ordering of $\vec{f}$ and B_q has to be defined with respect the same unique index for the edges as used for $\mathbf{K}$ in Equation (2.2).

5.1.1. Matching Policies

From energy conservation follows, that the power flowing into the power system has to either flow out or be consumed in the system. The generation and load at vertex i are denoted by g_i and l_i, respectively. The mismatch, defined as sum over all generation and load $m = \sum (g_i + l_i)$, is typically nonzero. To ensure the conservation of energy, the load and the generation have to be modified by some method so that they match, i.e. $\sum g_i^b = \sum l_i^b$. There are unlimited degrees of freedom to modify the load and generation, accordingly. We introduce two methods, that we refer to as "matching policies".

The minimum dissipation policy, described in Section 5.1.1.1, minimizes the sum over the quadratic flows. It is a benchmark as this minimizing property is known a priory. In Section 5.1.1.2, we introduce the global factor policy, which distributes the mismatch over all vertices relative to their load or generation and can be treated analytically.

Various other policies are possible. In real systems, the employment of capacities is usually assigned based on the costs of the generators and transport [107]. The implementation of such a policy would incorporate a modelling of different classes of generators with their cost functions, their spatial distribution, and rules on economic decision making. This is beyond the scope of this work. However, as the cost of transmission are also considered, the economic dispatch can be expected not to deviate strongly from the minimum dissipation policy.

In the following, the in- and out-flows before the policy is applied are referred to as "unmatched". The matched in- and out-flows, that fullfill the conservation of energy condition, are denoted by $s_j^b = l_j^b + g_j^b$.

5.1.1.1. Minimum Dissipation

A policy that minimizes the aggregated flows in the network is of great advantage. It defines a lower bound that can be used for comparison. Trivially, to minimize the flows, the load at vertex i should be satisfied by the generation at the same vertex, as this elicits no flow in the network. Thus, we define the unmatched in- and out-flow at vertex i as $s_i = g_i + l_i$ and the matched in- and out-flows for all vertices by

$$\vec{s}^{\,b} = \vec{s} + \vec{s}^{\,c} \,, \tag{5.8}$$

with s_i^c the correction at each vertex so that $\sum_i s_i^b = 0$. The minimal flow has to meet some constraints. The flow needs to satisfy Equation (5.2). For positive s_j, referred to as sources, the constraint $0 \leq s_j^b \leq s_j$ has to be met. For sinks, with $s_j \leq 0$, it is the constraint $s_j \leq s_j^b \leq 0$, in order not to introduce artificial generation or load. Further, $\vec{s}^{\,c}$ has to satisfy $\sum_i s_i = m = -\sum_i s_i^c$.

Another difficulty is that the flow f_{ij} along the link $i \to j$ can be positive or negative, depending on the direction. Using the absolute value in the objective function would

introduce a rather inconvenient nonlinearity. Minimizing the quadratic flow f_{ij}^2 solves this problem, leading to a constrained nonlinear optimization problem, also referred to as nonlinear programming, which is not easy to solve. Gertz and Wright [55] propose a method to solve problems of the form

$$\min_{\vec{x}} \vec{x}\,\mathbf{Q}\,\vec{x} \tag{5.9}$$

with the constraints $\mathbf{A}\vec{x} = \vec{b}$ and $x_i^l \leq x_i \leq x_i^u$, where x_i^l and x_i^u are the respective upper and lower bounds. To convert our problem into this form, we define the vector $\vec{x}$ as

$$\vec{x} = \begin{pmatrix} \vec{f} \\ \vec{s^c} \end{pmatrix} , \tag{5.10}$$

where $\vec{f}$ is a vector of length M[3] denoting the flows on the links. The in- and out-flow correction $\vec{s^c}$ for each vertex is of length N. Since only the flows should be minimized, the matrix $\mathbf{Q}$ has the form

$$\mathbf{Q} = \begin{pmatrix} 1_M & 0 \\ 0 & 0_N \end{pmatrix} , \tag{5.11}$$

where 1_M denotes a diagonal $M \times M$ unit matrix and 0_N an $N \times N$ zero matrix.

The constraints have to be of the form $\mathbf{A}\vec{x} = \vec{b}$. As shown in Section A.2.4, the in/out flows $\vec{s^b}$ are related to the flow on the links $\vec{f_{ij}}$ by the incidence matrix $\mathbf{K}$, $\vec{s^b} = \mathbf{K}\vec{f_{ij}}$. Using this relation, Equation (5.8) can be written as

$$\vec{s} = \vec{s^b} - \vec{s^c} = \mathbf{K}\vec{f_{ij}} - \vec{s^c} = \mathbf{A}'\vec{x'} \tag{5.12}$$

with

$$\mathbf{A}' = \begin{pmatrix} \mathbf{K} & I_N \end{pmatrix} . \tag{5.13}$$

The constraint $\sum_j s_j^c = -m$ can be incorporated into the linear constrain equation $\mathbf{A}\vec{x} = \vec{b}$ via

$$\mathbf{A} = \begin{pmatrix} \mathbf{K} & I_N \\ 0 & 1 \ldots 1 \end{pmatrix} . \tag{5.14}$$

and

$$\vec{b} = \begin{pmatrix} \vec{s} \\ -m \end{pmatrix} , \tag{5.15}$$

Additional conditions for all sources are $0 \leq s_j^b \leq s_j$ and for sinks $s_j \leq s_j^b \leq 0$.

So far, Equation (5.2), $\vec{P} = \mathbf{B}'\vec{\delta}$, is not implemented explicitly. As argued in Section A.2.4, the matrix $\mathbf{B}'$ is formally equal to the admittance matrix $\mathbf{Y}_{res}$ describing

[3] M denotes the number of edges.

a resistor network. The flow obeying Kirchhoff's laws is known to minimize the dissipation in a resistor network [19]. The dissipation p_{ij}^d at the link $i \to j$ is the product of the current f_{ij} and the voltage difference $(u_i - u_j)$ at the resistor R_{ij}. With Ohm's law, we find

$$p_{ij}^d = f_{ij}\,(u_i - u_j) = R_{ij} f_{ij}^2 \ . \tag{5.16}$$

The dissipation of the whole network is defined as the sum of the energy dissipated at all links

$$p_d = \sum_{e_k \in \mathcal{E}} R_{e_k} f_{e_k}^2 \tag{5.17}$$

with $\mathcal{E}$ the set of edges and e_k denoting an edge $i \to j$, so that e.g. $R_{e_k} = R_{ij}$. For $R_{ij} = 1$, Equation (5.17) is equal to the objective function in Equation (5.9). As shown in Section A.2.4, Kirchhoff's current law can be written as $\vec{s}^{\,b} = \mathbf{K}\vec{f}_{ij}$, where $\mathbf{K}$ is the incidence matrix. By Thomson's Principle[4] [73], we know that if the current law is met and the flow is minimal, Kirchhoff's potential law is also fullfilled. Because of the formal equality of $\mathbf{B}'$ and $\mathbf{Y}_{res}$ and of the conservation of energy, $\vec{P} = \mathbf{B}'\vec{\delta}$ (see Equation (5.2)) has to be satisfied for $B_{ij} = 1$. Arbitrary B_{ij} can easily be implemented by extending the vector $\vec{x}$ in Equation (5.10) with a vector containing the voltage phases $\vec{\delta}$. Further, the constraint $\mathbf{X}\,\mathbf{K}^T\vec{\delta} - \vec{f} = 0$, where $\mathbf{X}$ is a diagonal matrix with the susceptances on the diagonal, is added to the matrix $\mathbf{A}$ in Equation (5.14) together with further zeros in $\mathbf{Q}$ in Equation (5.9). This is not done here, since we use only uniform susceptances $\mathbf{B}_{ij} = 1$ in this work. Equation (5.5) is consequently fullfilled because the flow is minimal and the additional constraints would only cause a higher complexity for numerical evaluation that is thus avoided.

5.1.1.2. Global Factor Policy

A very simple approach is to divide mismatches with equal shares over all vertices. For shortages, all vertices have to shed a fraction of the load and the same is true for surpluses concerning generation. Sinks and sources at each vertex are then multiplied with a common factor in order to match generation and load,

$$g_i^b = (1 - \alpha')g_i \qquad l_i^b = (1 - \beta')l_i \ . \tag{5.18}$$

To ensure that for $m = \sum_i s_i = \sum_i g_i + l_i > 0$ only the generation is reduced, we require that $0 \leq \alpha' \leq 1$ with $\beta' = 0$. $\alpha' = 0$ and $0 \leq \beta' \leq 1$ causes loads to be shedded only for negative m.

[4]The currents in a resistor network that fullfill Kirchhoff's circuit laws, are the unique flows that minimize the dissipation [73].

With the requirement that $\sum_i \left(g_i^b + l_i^b\right) = 0$, the values of α' and β' follow for positive mismatch $m > 0$ as

$$\alpha' = \frac{m}{\sum_i g_i} = 1 + \frac{\sum_i l_i}{\sum_i g_i} \qquad \beta' = 0 \; . \tag{5.19}$$

For $m < 0$, we find

$$\beta' = \frac{m}{\sum_i l_i} = 1 + \frac{\sum_i g_i}{\sum_i l_i} \qquad \alpha' = 0 \; . \tag{5.20}$$

This policy can be implemented very efficiently, as the factors α' and β' are calculated simply by adding up all load and generation. These factors can then be applied to all vertices, yielding a low numerical complexity.

5.2. Definition of the Model

The model presented here, serves as a base for the following studies. It is focused on fluctuating generation and various distributions of generators and consumers in the network. The generation g_i is defined to be a random variable to account for fluctuations in the renewable power generation. The load is modeled by a constant value, since the timescale of variations in the load is much larger than the timescale of fluctuations in the renewable power generation. This is plausible on the level of the transport network, since only aggregated loads have to be considered. Besides that, the strength of the variations in the load is much smaller than the variation in the generation, see Chapter 3, and the daily load curves are well known by utility companies [107]. Consequently, mainly fluctuations in the generation due to e.g. weather effects constitute a problem, eventually leading to shortages or excess of power. We assume that for cases of shortages, load is either shedded or met with additional local balancing generation, and that excess generation is discarded.

To test different scenarios for the distribution of the load and the generation over the vertices, the mean values of the generation and consumption are varied in the network. We interpolate between a homogeneous distribution and a separation between center and periphery in terms of the degrees of the vertices, so that load is located in the center and generation in the periphery. A vertex with high degree is considered to be more central than low degree vertices. This is plausible since e.g. a large share of wind power can be expected to be generated offshore or close to the sea and each generator to be connected only by few links. Consumption centers should be connected to the grid by many links to ensure a high reliability of the system. Various other spatial separations could be defined, but to restrict degrees of freedom of the model, only these scenarios are considered.

The parameters β_g and β_l, determine the strength of the spatial separation of the

generation and load. The mean generation at vertex i is defined to be

$$\langle g_i \rangle = c_g \left(k_i - k_{min} + 1\right)^{-\beta_g} \tag{5.21}$$

and the load

$$\langle l_i \rangle = l_i = -c_l \left(k_{max} + 1 - k_i\right)^{-\beta_l} \ , \tag{5.22}$$

with the normalization factors c_l and c_g. The higher β_L and β_g the more power is consumed in the center and produced in the periphery. k_{max} and k_{min} denote the maximal and minimal degree of the network.

To compare different networks, the sum of the generation and of the load are normalized so that $\sum_i^N \langle l_i \rangle = -1$ and $\sum_i^N \langle g_i \rangle = \kappa$. The parameter κ is the ratio of the power produced to the power consumed. The prefactors c_l and c_g then follow as

$$c_l = \left(\sum_i \left(k_{max} + 1 - k_i\right)^{-\beta_l} \right)^{-1} \qquad c_g = \kappa \left(\sum_i \left(k_i - k_{min} + 1\right)^{-\beta_g} \right)^{-1} \ . \tag{5.23}$$

The choice of the probability distribution of the generation was made such, that it is "easy to work with" but on the other hand plausible with respect to real data. In the model, the generation g_i is drawn from a uniform distribution, since it is the simplest continuous distribution supported on a bounded interval and has e.g. a well defined characteristic function. The uniform distribution is denoted by $U(x; a, b)$, where a and b denote the upper and lower bound. The lower bound for the generation is defined to be zero and the upper bound follows as twice the mean generation $\langle g_i \rangle$. The generation at each vertex is assumed to be independent and successive fluctuation realizations to be uncorrelated.

The transport networks are created with the geometric p network approach described in Section 2.1.2. Power grids were modelled by networks with exponential degree distribution $p(k) = \gamma^{-1} \exp\left(-\frac{k}{\gamma}\right)$ with $\lambda = \{1, 1.8\}$ as these mimic the European grid, see Section 2.1.3. For comparison, also scale-free degree distributions, with $\alpha = \{2.3, 2.8\}$, and Poisson degree distributions, with $\lambda = \{1.5, 1.8\}$, are analyzed. The latter also resemble the simplified grid in Chapter 6, as this is close to a geometric network with Poisson degree distribution [39]. Networks for the above configurations were created with $p_{cc} = 0$ and $p_{cc} = 1$, to obtain a low and a high clustering coefficient, respectively, as described in Section 2.1.2. The probability that a randomly selected vertex has degree one is set to 10% of the probability of a vertex with degree two to account for the features found for the network of power grids, see Section 2.1.3. One typical realization of each network is considered and their properties are summarized in Table 5.1. For simplicity all susceptances of the links are set to one, $B_{ij} = 1$.

The model is evaluated using the Monte Carlo method. In the following 10^4 fluctuation realizations for the generation g_i are drawn for each vertex from the respective uniform

	Exponential				Scale-free			
	$\lambda = 1$		$\lambda = 1.8$		$\alpha = 2.3$		$\alpha = 2.8$	
p_{cc}	0.	1.	0.	1.	0.	1.	0.	1.
$\langle k \rangle$	2.498	2.498	2.108	2.108	3.772	3.772	2.898	2.898
$\langle$distance$\rangle$	9.701	21.469	22.343	69.217	4.809	7.068	6.667	11.582
cc	0.002	0.115	0.000	0.054	0.009	0.249	0.002	0.166
	Poisson							
	$\lambda = 1.5$		$\lambda = 2.5$					
p_{cc}	0.	1.	0.	1.				
$\langle k \rangle$	2.554	2.554	3.172	3.172				
$\langle$distance$\rangle$	9.401	23.226	6.529	14.284				
cc	0.000	0.129	0.000	0.134				

Table 5.1.: *Measures of the graphs that are used in the simulations. p_{cc} is the parameter of the network generator to tune the clustering coefficient (cc), see Section 2.1.2.*

distribution. For each fluctuation realization the resulting transport flows are calculated. This is done as follows: One of the matching policies introduced in Section 5.1.1 is applied to the load l_i and generation g_i for each vertex, yielding the balanced load l_i^b and generation g_i^b that satisfy $\sum_i \left(l_i^b + g_i^b\right) = 0$. The balanced load allows to define the flow into or out of vertex i, given by

$$s_i^b = l_i^b + g_i^b \; . \tag{5.24}$$

The DC flow approximation is used to calculate the flows in the transport network; see Section 5.1 and Section A.2.3. From the equations of the DC-flow follows that

$$\vec{s^b} = \mathbf{B}'\vec{\delta} \, , \tag{5.25}$$

with the voltage phase angle δ_i at vertex i. To calculate the phases the Moore–Penrose pseudoinverse $\mathbf{B}'^{+}$ is used (see Section A.2.4.1) since the matrix $\mathbf{B}'$ is not directly invertible:

$$\vec{\delta} = \mathbf{B}'^{+}\vec{s^b} \; . \tag{5.26}$$

The flow f_{ij} along a the link $i \to j$ is given by

$$f_{ij} = B_{ij}\left[\delta_i - \delta_j\right] = B_{ij}\left[\sum_{k=1}^{N}\left(\mathbf{B}'^{+}_{ik} - \mathbf{B}'^{+}_{jk}\right) s_k^b\right] \; . \tag{5.27}$$

When considering all fluctuation realizations, an ensemble of flows for each link is found. This allows to derive probability distributions for the flows of each link.

The proposed model greatly simplifies the findings in Chapter 3. This includes the systematic behavior on the seasonal and daily timescales for the load and generation. On the other hand, this approach reduces the parameters of the model and allows for a systematic analysis and analytical estimations of the flows.

5.3. Simulation Results Overview

To get an overview of the behavior of the model and the dependencies on the parameters, we sampled a part of the parameter space. All combinations of the parameters $\beta_S = \{0, 1\}$, $\beta_L = \{0, 1\}$, $\kappa = \{0.5, 1, 2\}$, and $p_{cc} = \{0, 1\}$ were used togeher with the matching policies introduced in Section 5.1.1. Networks with exponential degree distribution, with $\lambda = \{1, 1.8\}$, scale-free degree distribution, with $\alpha = \{2.3, 2.8\}$, and Poisson degree distribution, with $\lambda = \{1.5, 1.8\}$ are studied. Their main properties are shown in Table 5.1.

The aggregated flows over all links

$$F^A = \sum_{k \in \mathcal{E}} |f_k| \tag{5.28}$$

averaged over 10^4 fluctuation realizations are shown in Figure 5.1. As defined in Section 5.2, the sum of all unbalanced loads are normalized to one and the unit of the flow is arbitrary. The difference between the data for different clustering coefficients, $p_{cc} = 0$ and $p_{cc} = 1$, is mainly an up-scaled sum of the flow when using the global factor policy. For the minimum dissipation policy this is only true for $\kappa = 1$. For β_l or β_g equal to one, the sources and sinks are spatially separated, so that the aggregated flows increase compared to $\beta_l = \beta_g = 0$. Both cases of β_l with $\beta_g = 1$ show a higher flow than the corresponding cases with $\beta_g = 0$. This is not surprising as the load of all vertices is constant and so for $\beta_g = 0$ there is a high probability for a generator to be close to a load. For $\beta_g = 1$, the generation is located in the periphery and needs to be transported over longer distances, causing a higher aggregated flow. Varying the ratio between sources and loads κ has a different effect on the global factor policy compared to the minimum dissipation policy. The latter show the maximal flow for $\kappa = 1$ and the flow for both the $\kappa = 0.5$ and $\kappa = 2$ cases is significantly lower. This is due to the fact that the minimum dissipation policy first serves the local demand. For $\kappa = 0.5$, most of the time the source can not even satisfy the local load and so little power is injected into the network causing a small aggregated flow. The opposite takes place for $\kappa = 2$, almost all loads can be satisfied locally, so that little transport is needed. For $\kappa = 1$, the probability that a load can not be satisfied locally or within it's near neighborhood is high, so that the average accumulated flow is relatively large. For the global factor policy there is no preference to serve local loads first. Any pair i, j of vertices might be chosen for matching so that there is a high probability for transport for all κ.

When looking at the average distances of the networks in Table 5.1 together with the

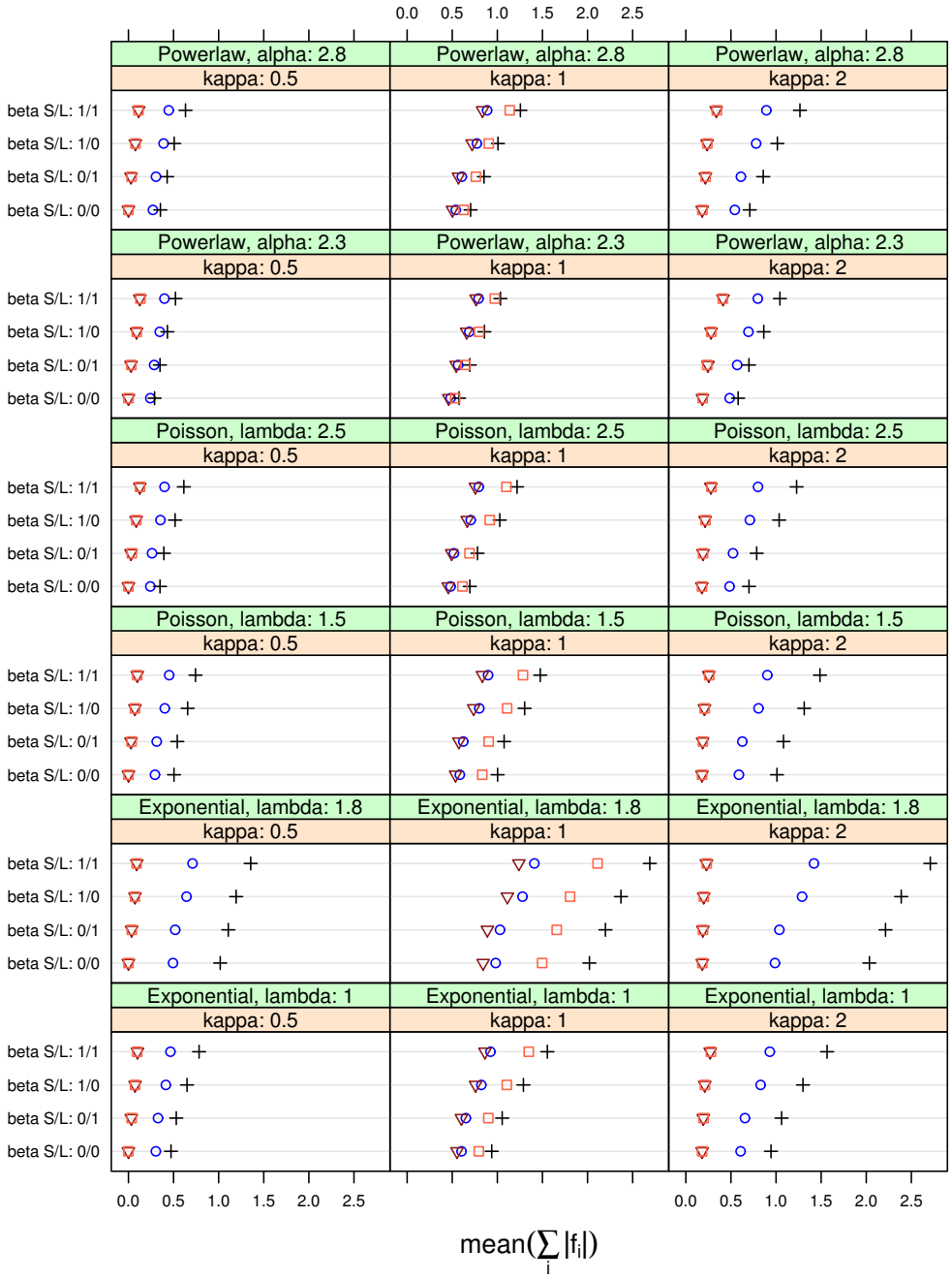
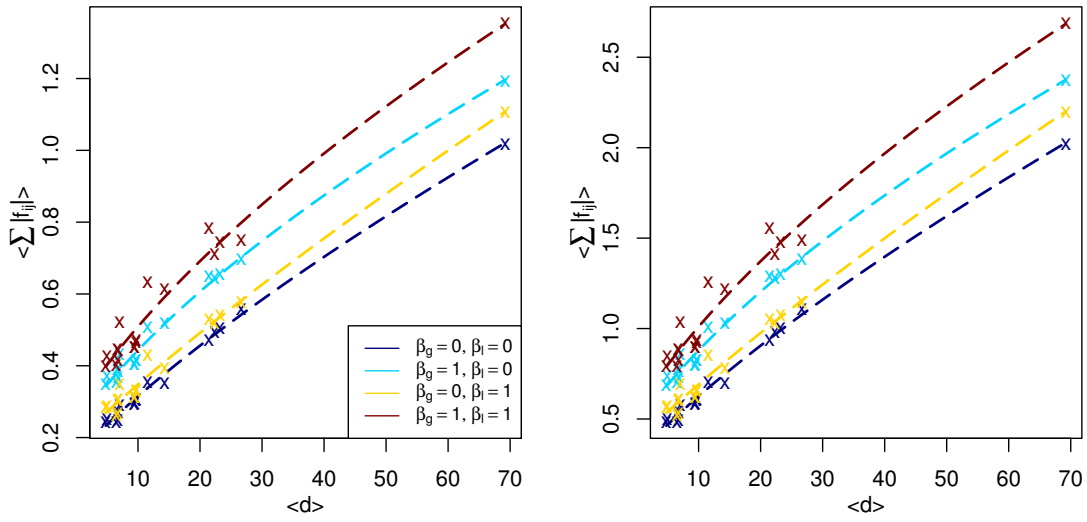


Figure 5.1.: *The average sum over all flows* $\langle \sum_{k \in \mathcal{E}} |f_k| \rangle$ *for different configurations for one network realization of each degree distribution as denoted in the plot. The flows have arbitrary units. Colors / symbols:* ▽ *Minimum dissipation with* $p_{cc} = 0$, □ *minimum dissipation with* $p_{cc} = 1$, ○ *global factor with* $p_{cc} = 0$, *and* + *global factor with* $p_{cc} = 1$.



(a) $\kappa = 0.5$, *global factor policy*

(b) $\kappa = 1$, *global factor policy*

(c) $\kappa = 2$, *global factor policy*

(d) $\kappa = 1$, *minimum dissipation policies*

Figure 5.2.: *Scaling of the averages of the sums of all flows* $\left\langle \sum_j f_{e_j} \right\rangle$ *with the mean distances of all networks listed in Table 5.1 with different* κ*. The crosses indicate the average aggregated flow and the mean distances for a each network. For the fitted dashed line see text.*

aggregated flows in Figure 5.1, a scaling between the average accumulated flow and the mean distances of the networks seems likely. In Figure 5.2, this scaling is shown for all networks generated with $p_{cc} = 0$ and $p_{cc} = 1$. The different average distances arise from different degree distributions and clustering coefficients, see Table 5.1. For the global factor policy and the minimum dissipation policy for $\kappa = 1$, there is a dependence of the mean aggregated flow to the average distance that can be fitted in good approximation with

$$\left\langle \sum_j f_{e_j} \right\rangle \propto (\langle d \rangle - c)^a \ . \tag{5.29}$$

For the configurations shown in Figure 5.2, $0.6 \leq a \leq 0.825$ was found indicated by the dashed lines. No such scaling could be found for the minimum dissipation policy with $\kappa = 0.5$ and 2. This finding is only discussed qualitatively, here. The global factor policy implements a matching that ensures that all vertices receive a fraction of their load independent of their positions. Larger distances between generator and load cause a higher aggregated flow, so the scaling with the average network distance is not surprising. With respect to the aggregated flow, the network properties, like degree distribution, clustering coefficient etc., are dominated by the resulting average network distance. The interesting observation is that both policies show a similar qualitative behavior for $\kappa = 1$, indicating that in this regime their influence on the flows is small. Further, the comparison with the shortest path transport reveals important properties of the DC flow. With a simple reasoning, we find $a \approx 1$ for shortest path transport. The mean distance is the sum over all distances between all vertex pairs normalized by the number of pairs of vertices $N(N-1)$. This can also be calculated by counting all edges that are part of a shortest path between two vertices normalized to the number of all vertex pairs, since both sums have to be the same. This yields

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{k \in \mathcal{E}} \left(\sum_{j \in \mathcal{V}} \sum_{i \in \mathcal{V} \neq j} \mathrm{path}([i \to j]; e_k) \right) \tag{5.30}$$

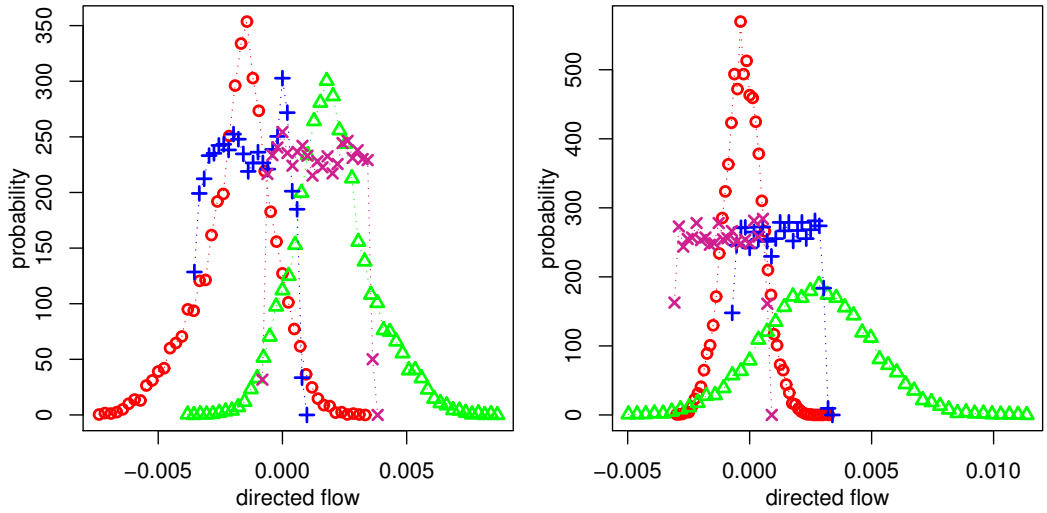
(see also Equation (2.18)). The flow over link f_k using shortest paths is given by

$$f_k = \sum_{j \in \mathcal{V}} \sum_{i \in \mathcal{V} \neq j} \alpha_{ij} \mathrm{path}([i \to j]; e_k) \ , \tag{5.31}$$

where α_{ij} is the flow from i to j. If we assume that the flows between two vertices are similarly distributed for all pairs of vertices[5], the average of the factors $\langle \alpha_{ij} \rangle$ can be drawn in front of the summations, so that the average accumulated flow $F^A = \sum_{k \in \mathcal{E}} f_k$ in the network follows as

$$F^A = \sum_{k \in \mathcal{E}} f_k = \langle \alpha_{ij} \rangle \sum_{j \in \mathcal{V}} \sum_{i \in \mathcal{V} \neq j} \mathrm{path}([i \to j]; e_k) = N(N-1) \langle \alpha_{ij} \rangle \langle d \rangle \propto \langle d \rangle \tag{5.32}$$

[5]E.g. for the case of $\beta_L = \beta_S = 0$, the average flows between all pairs of vertices are equal.



(a) *Minimum dissipation policy* **(b)** *Global factor policy*

Figure 5.3.: *Histograms of typical flow distributions on links using the global factor and the minimum dissipation policy, here: Exponential degree distribution with* $\lambda = 1.8$, $\beta_S = 1$, $\beta_L = 1$, $N = 1000$ *and* $\kappa = 1$. *The symbols indicate different links.*

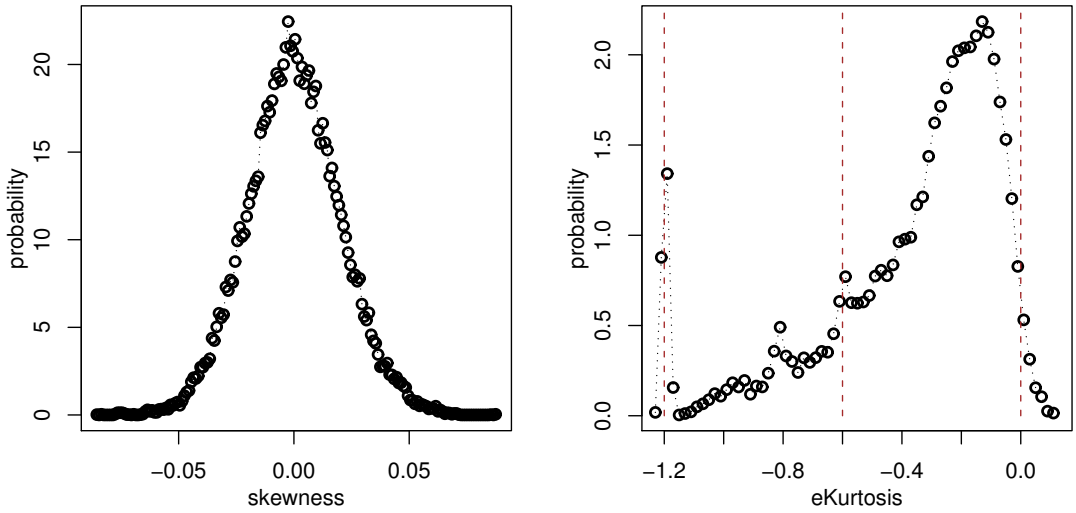
and show a scaling exponent of one. The fact that the DC flow exhibits $a < 1$ is thus remarkable. It can be explained by the fact that the DC flow is indistinguishable, in the sense that a load can not discriminate between the generating vertices it receives power from. Because of the potentials that follow from Equation (5.2), a large fraction of the generation flows to the nearest load, so that there is an inherent tendency for the effective flows between the vertices to be minimal. Therefore, the increase of the average aggregated flow increases slower than the average distance.

After considering the aggregated flow, we now analyze the probability distributions of the flow over the individual links. In general, it can be expected that the flows over many links are distributed close to a normal distribution as will be argued in the following. However, some links may be dominated by the flow of only few vertices, so that the probability distribution of the flows may be close to the distribution of the dominating vertex, which is a uniform distribution by definition.

Some typical flow distributions are shown in Figure 5.3. As expected, some distributions are close to uniform and normal distributions. To describe the flow distributions, the third and fourth standardized moments, called the skewness γ_1 and the kurtosis γ_2',

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \qquad \gamma_2' = \frac{\mu_4}{\sigma^4}\,, \tag{5.33}$$

are examined. Hereby, μ_3 and μ_4 denote the third and fourth central moments. The skewness is a measure of the asymmetry of a distribution and the kurtosis measures



(a) *Distribution of the skewness*

(b) *Distribution of the excess kurtosis, the brown lines indicate the kurtosis of the uniform, the triangular, and the normal distribution from left to right*

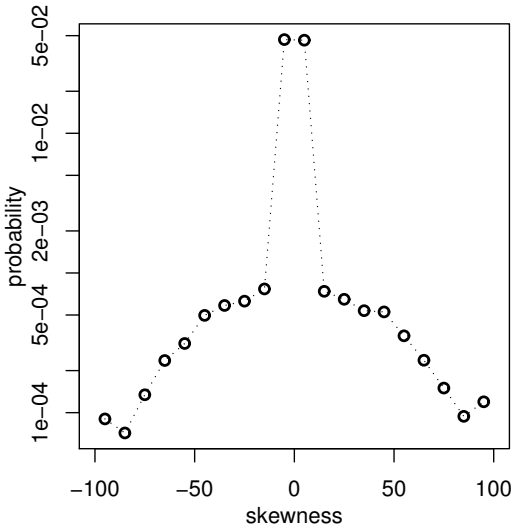
Figure 5.4.: *Distributions of the skewness and the excess kurtosis of all links for all configurations shown in Figure 5.1 that employed the global factor policy.*

how strongly a distribution is peaked. Commonly the excess kurtosis is used, defined as

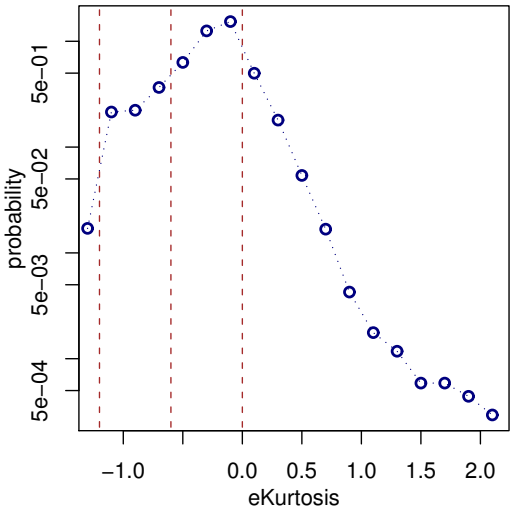
$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{\kappa_4}{\kappa_2^2} \,, \tag{5.34}$$

where κ_n is the n-th cumulant and $\gamma_2 \geq -2$ by construction [71]. The excess kurtosis has the advantage that the value for the normal distribution is zero. Thus, the kurtosis can be used as a test for the Gaussianity of a distribution [71]. The uniform distribution has an excess kurtosis of $\gamma_2 = -\frac{6}{5}$ and the triangular distribution, which is the convolution of two identical uniform distributions, has $\gamma_2 = -\frac{3}{5}$.

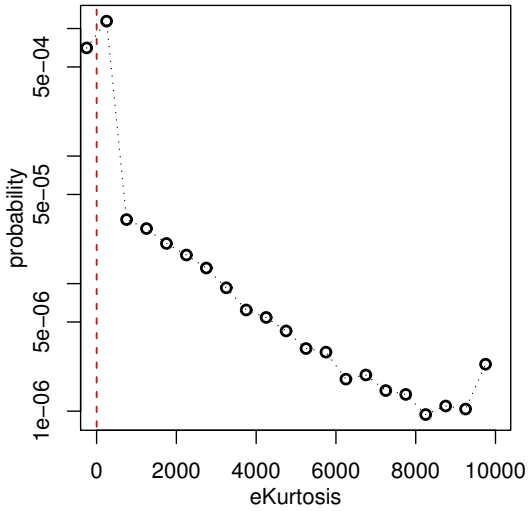
The distributions of the skewness and kurtosis are shown in Figure 5.4 using the global factor policy for all parameters that were considered in Figure 5.1. The skewness is scattered around zero with a small variance, so that the flow distributions are in good approximation symmetric about the mean. The kurtosis therefore carries the most information. The kurtosis lies in almost all cases between the excess kurtosis of the uniform distribution, $-\frac{6}{5}$, and zero, i.e. the kurtosis of the normal distribution. The highest probability is an excess kurtosis close to zero, a strong evidence that the probability distributions of the flows on most of the links are close to a normal distribution. Values above zero indicate that the distribution is stronger peaked than a normal distribution. The first two standardized moments of the link flows using the minimum dissipation policy, shown in Figure 5.5, are also close to a normal distribution with the highest probability but the deviations are much stronger. The excess kurtosis



(a) *Distribution of the skewness*



(b) *Distribution of the excess kurtosis for* $\kappa = 1$



(c) *Distribution of the excess kurtosis for* $kappa = \{0.5, 1, 2\}$

Figure 5.5.: *Distributions of the skewness and the excess kurtosis of all links for all configurations shown in Figure 5.1 that employed the minimum dissipation policy. The y-axis is on a logarithmic scale.*

with $\kappa = 1$ of all links is shown in Figure 5.5b. Compared to the global factor policy, the number of links with a flow distribution that is more peaked than the normal distribution, indicated by an excess kurtosis large than zero, is significantly higher. When including the values for $\kappa = 0.5$ and $\kappa = 2$ in Figure 5.5c, we see very high values for the excess kurtosis. This is explained by the fact that for $\kappa = 0.5$ and $\kappa = 2$, the

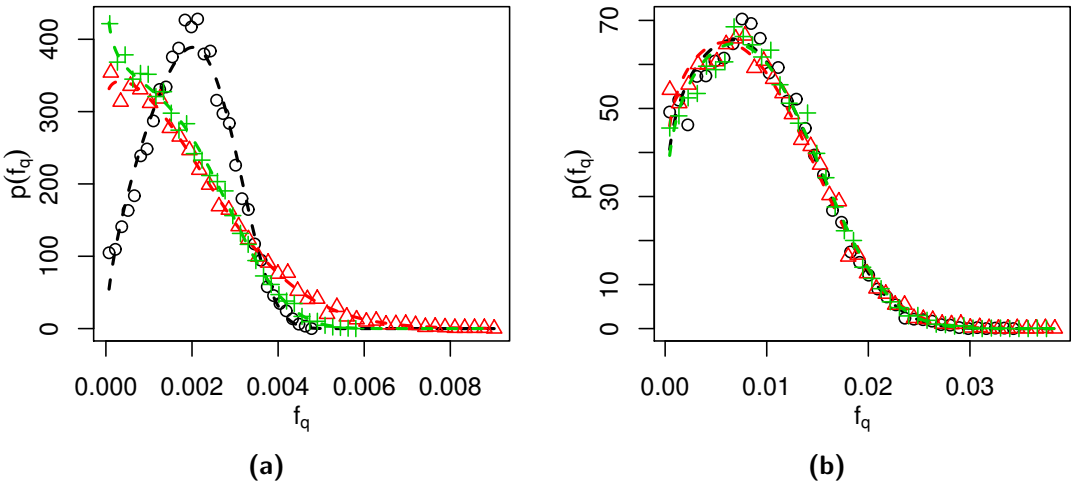


Figure 5.6.: *Typical probability distributions of the flows on links using the global factor policy on a exponential network with $\lambda = 1$ and $N = 1000$. In (a) links with a low average flow are shown, in (b) links with a high average. The network parameters are chosen as in Figure 5.3. The dashed lines indicate the fitted generalized gamma distribution.*

power flows in the network are small, so that the distributions are strongly peaked at zero, consult also Figure 5.1.

The fact that there is a high probability to find a distribution close to a normal distribution is not surprising. The DC flow from vertex i to j flows over all possible paths from i to j. For a given fluctuation realization, the flow over link q can be written as

$$f_q = \sum_{k=1}^{N} \varrho_{qk} s_k \ , \tag{5.35}$$

where ϱ_{qk} denotes the fraction of the in- and outflow of vertex k that flows over link q. If N is large, enough prefactors ϱ_{qk} are similar and assuming that the correlations are small, the central limit theorem applies and the distribution is in good approximation a normal distribution [54].

For some problems, e.g. to determine the required capacity of a link, the direction of the flow is not important but only its magnitude. Therefore, the absolute flow f^a over the link q, defined as

$$f_q^a = |f_q| \ , \tag{5.36}$$

is examined. In Figure 5.6, typical distributions for the absolute flow are shown using the global factor and in Figure 5.7 for the minimum dissipation policy. In most cases, these distributions are very good approximated by a generalized gamma distribution.

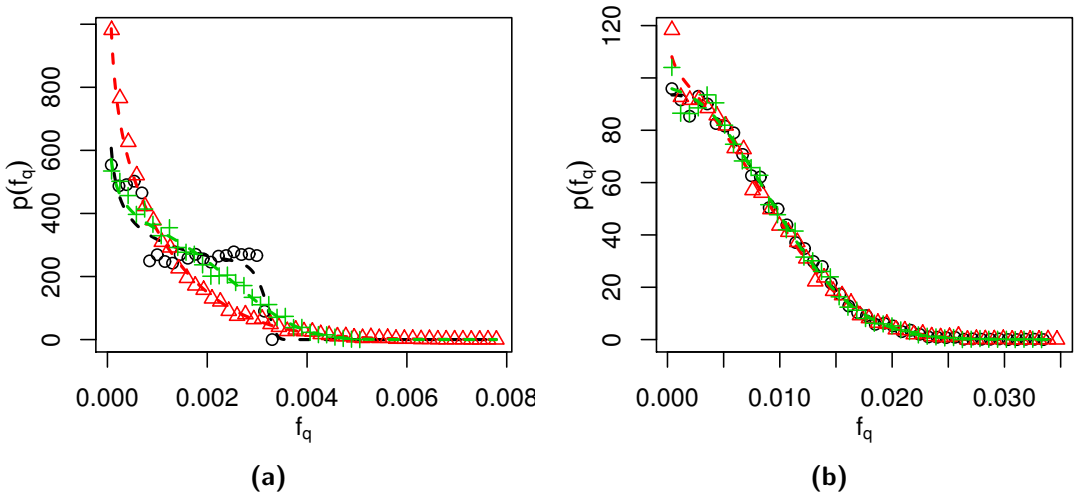


Figure 5.7.: *Typical probability distributions of the flows on links using the minimum dissipation policy. The parameters are the same as in Figure 5.6.*

The generalized gamma distribution is defined as

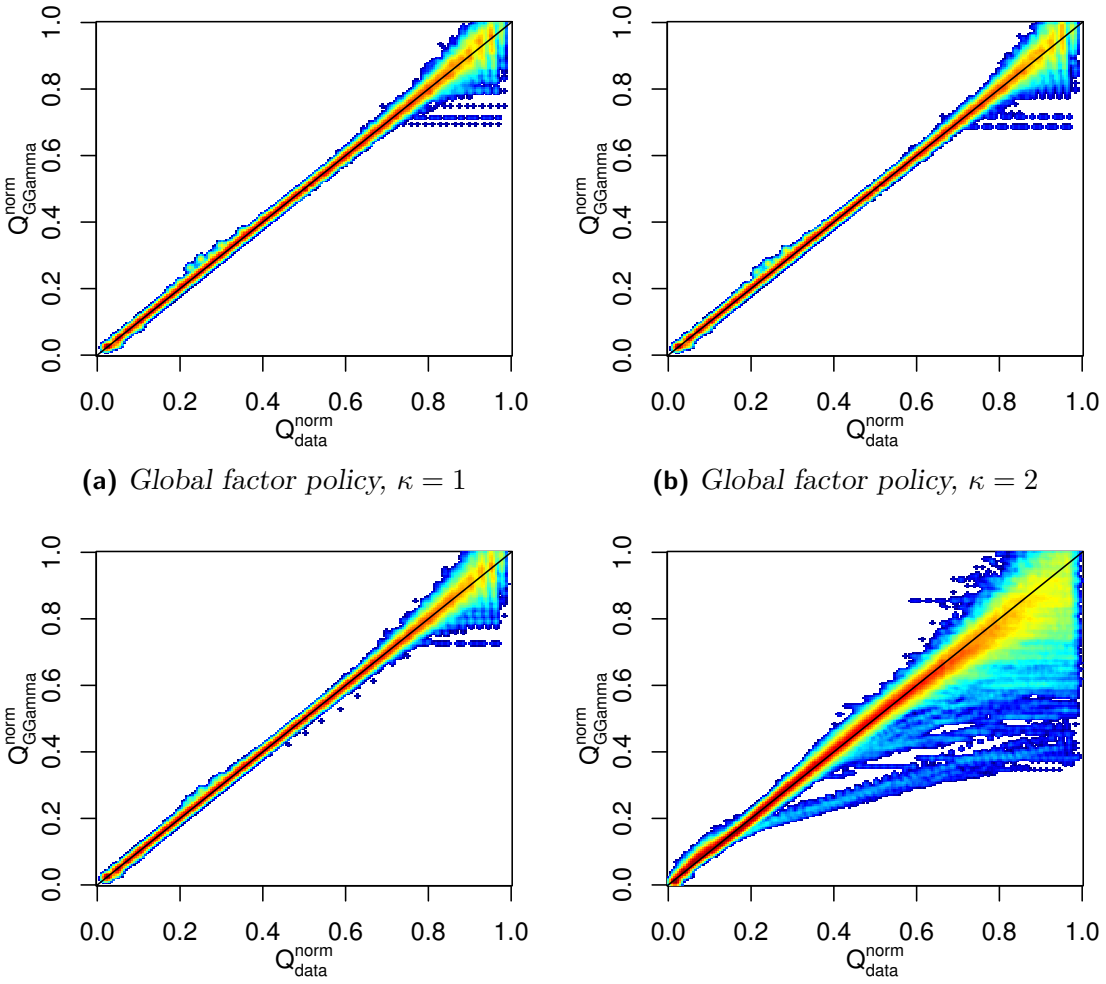
$$f(x;b,d,k) = d\,\frac{x^{dk-1}}{b^{dk}\Gamma(k)}\exp\left(-\left(\frac{x}{b}\right)^{d}\right)\,, \tag{5.37}$$

with positive b, d and k, and the gamma function $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,dt$ [112]. The dashed lines in Figure 5.6 and 5.7 indicate the fitted generalized gamma distributions. The parameters were fitted to the first three raw moments, given by

$$\mu_n' = E(X^n) = b^n\frac{\Gamma\left(k-\frac{n}{d}\right)}{\Gamma(k)}\,, \tag{5.38}$$

as described in Section A.3.2.

A good method to compare two distributions is the so-called Q-Q plot [125], where the quantiles of the data are plotted against the quantiles of the fitted distribution, in our case the generalized gamma distribution. The quantile Q denotes the value with the property, that with probability q the values of the probability distribution $p(x)$ are smaller than Q, $\int_{-\infty}^{Q} p(x)dx = q$. The Q-Q plots of all links for $\beta_l = \beta_g = 0$ and $\beta_l = \beta_g = 1$ are shown in Figure 5.8 as a density. All quantiles were normalized to the maximal quantile in the data of the respective link, so that $0 < Q_{data}^{norm} < 1$ for all quantiles of the data. The dark red colors correspond to pairs of quantiles found in almost all links and the dark blue regions to quantiles only found for single links. The two policies are shown separately for $\kappa = 1$ and $\kappa = 2$. For $\kappa = 1$, both policies show the highest densities along the $x = y$ line, indicating, that the quantiles of the data and the fitted generalized gamma distribution are equal. Only for high quantiles, small deviations are found that can be expected to be due to the sampling. This is also true for the global factor policy for $\kappa = 2$. However, there are large deviations between the



(a) *Global factor policy,* $\kappa = 1$

(b) *Global factor policy,* $\kappa = 2$

(c) *Minimum dissipation policy,* $\kappa = 1$

(d) *Minimum dissipation policy,* $\kappa = 2$

Figure 5.8.: *Q-Q plot for comparing the quantiles of the data* Q^{norm}_{data} *to the quantiles of the fitted generalized gamma distribution* Q^{norm}_{GGamma} *for all links. The quantiles are normalized to the values of the maximal quantile of the data for each link to map all quantiles on the range between zero and one. The quantile plots of all links are shown as density, red corresponds to a pair of quantiles found in all links and the dark blue regions are only found for single links. The isolated point are due to the binning of the data. An exponential network with* $\lambda = 1$, $\beta_S = \beta_L = \{0, 1\}$ *is used and* $N = 1000$.

distributions of the absolute flows when using the minimum dissipation policy with $\kappa = 2$ and the fitted generalized gamma distribution on some links. This is discussed in more detail in Section 5.5.

5.4. Analytical Approximations for the Uncorrelated Case

In this section, approximations to the flow distributions for each link are derived analytically. This allows to predict the flows in the network and assign appropriate capacities to the links.

In probability theory, most theorems assume uncorrelated random variables. The unmatched in- and out-flows are uncorrelated by definition in the model. The matching introduces negative correlations between generation and loads due to the condition that the sum over the generation has to be equal to the sum over the loads, $\sum_i g_i^b = -\sum_i l_i^b$. By systematicly matching generation and load, the correlations between the matched in- and out-flows s_j^b, for different vertices are determined by the policy.

Considering a pair of vertices i and j, the correlation between the random variables of the matched in- and out-flow S_i^b and S_j^b is defined by [77]

$$\rho_{i,j} = \mathrm{corr}(S_i^b, S_j^b) = \frac{\mathrm{cov}(S_i^b, S_j^b)}{\sigma_{S_i^b}\sigma_{S_i^b}} = \frac{E[(S_i^b - \mu_{S_i^b})(S_j^b - \mu_{S_j^b})]}{\sigma_{S_i^b}\sigma_{S_j^b}} \,, \tag{5.39}$$

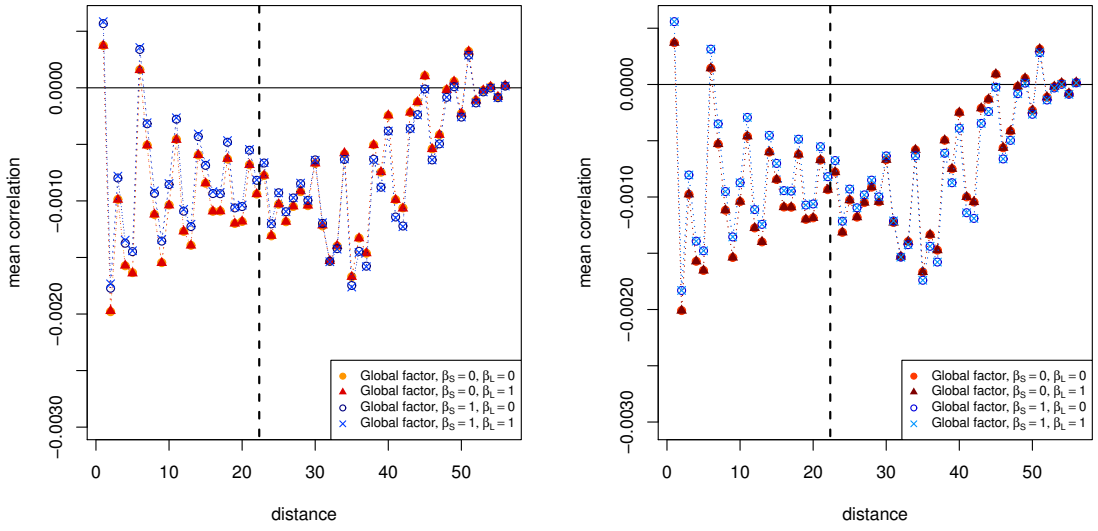
with $E[X]$ being the expectation value of the random variable X with mean μ_X and standart deviation σ_X. The correlation is defined so that $-1 \leq \rho_{i,j} \leq 1$. Figure 5.9 shows the average correlation of the respective s_i^b for all pairs of vertices at a given distance, when using the global factor policy. The correlations found in the data are in the order of $\mathcal{O}(N^{-1})$. We conclude, that when using the global factor policy, the correlations are small and can be neglected for $N \gtrsim 100$. The derivation done in this section takes advantage of this fact. As the distributions of the absolute flow are in good agreement with a generalized gamma distribution, we aim to derive the respective parameters for each link.

The in- and out-flow at each vertex is given by a probability distribution $p_i(s_i)$. The effects of the matching will be discussed in detail in Section 5.4.2 and the distributions $p(s_j^b)$ derived. For now, we neglect the influence of the matching policy and assume $s_j^b = s_j$. The cumulants and the cumulant generating functions are used for the analysis, as these allows for an approach similar to the method used in Section 2.3. The properties of the cumulants are described in Section A.3.1.

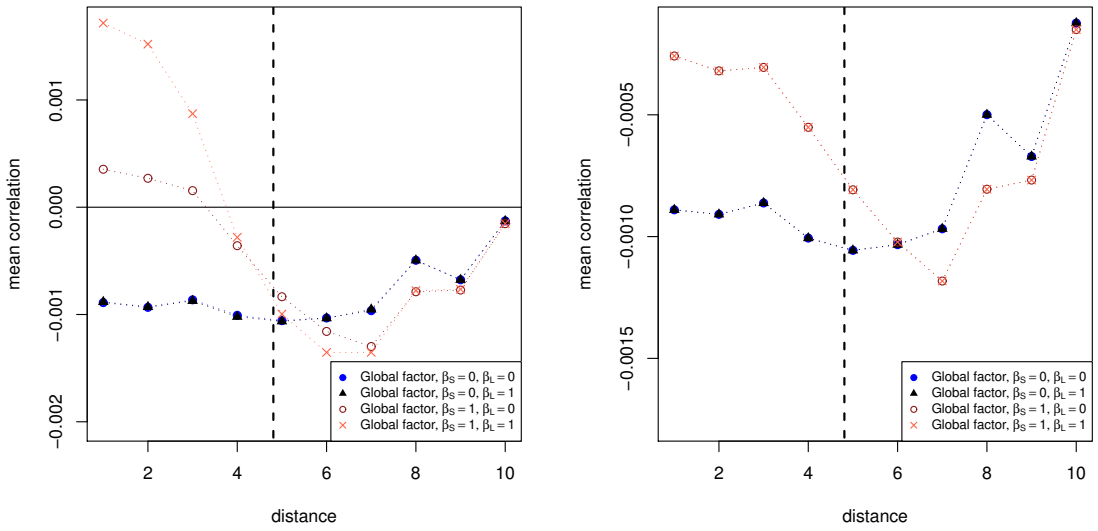
For a given realization $\vec{S^b}$ of the in- and out-flow at each vertex, the flow over the link $q \to r$ is given by

$$F_{qr} = B_{qr} \sum_{k=1}^{N} \left(Y_{qk}^+ - Y_{rk}^+\right) S_k^b \,, \tag{5.40}$$

consult also Equation (5.27). For notational simplicity, $\mathbf{Y}$ is used to describe the DC-flow instead of $\mathbf{B}'$, see Section 5.1. F_{qr} is thus a weighted sum over N random



(a) $\kappa = 1$, *exponential degree distribution* **(b)** $\kappa = 2$, *exponential degree distribution*



(c) $\kappa = 1$, *scalefree degree distribution* **(d)** $\kappa = 1$, *scalefree degree distribution*

Figure 5.9.: *Average correlations of the matched in- and out-flows s_j of all pairs of vertices at a given distance introduced by the global factor matching policy. The data for a graph with exponential degree distribution with $\lambda = 1.8$ and for a scalefree network with $\alpha = 2.3$ are shown, both with $N = 1000$ and $p_{cc} = 0$. The dashed line indicates the mean distance of the graph.*

values. This means that the probability distribution of the flow at each edge is given by the convolutions of the probability distributions of the in- and out-flows $p\left(s_j^b\right)$

$$p_{f_{qr}}(f_{qr}) = p_{s_1^b}\left(s_1^b; \gamma_{qr}^1\right) * p_{s_2^b}\left(s_2^b; \gamma_{qr}^2\right) * \cdots * p_{s_N^b}\left(s_N^b; \gamma_{qr}^N\right) \ , \tag{5.41}$$

with $\gamma_{qr}^j = B_{qr}\left(Y_{r,j}^+ - Y_{q,j}^+\right)$. $p_{s_j^b}\left(s_j^b; a_j\right)$ describes random values, S_j^b, drawn from the

probability distributions $p_{s_j^b}(s_j^b)$ scaled with the factor a_j. The convolution is denoted by $*$.

By definition, the in- and out-flows S_j are drawn from a uniform distribution $U(x; a', b')$, so that all values on the closed interval from a' to b' have an equal probability and since the balancing is neglected here, we have

$$p_{f_{qr}}(f_{qr}) = U\left(s_1^b; \gamma_{qr}^1 a_1', \gamma_{qr}^1 b_1'\right) * U\left(s_2^b; \gamma_{qr}^2 a_2', \gamma_{qr}^2 b_2'\right) * \cdots * U\left(s_N^b; \gamma_{qr}^N a_N', \gamma_{qr}^N b_N'\right) \ . \quad (5.42)$$

The scaling factors γ_{qr}^j from Equation (5.41) are implemented by scaling the bounds of the distribution, see Section A.3.1. Still calculating Equation (5.41) and (5.42) is rather nontrivial. Bradley and Gupta [21] derived a series representation of the sum of N non-identically distributed uniform variables and a closed form solution could not be found in literature. The distribution of the sum of N nonidentical independent random variables uniformly distributed on the closed interval $[c_i - a_i, c_i + a_i]$ is then given by [21]

$$f_N(x) = \frac{1}{(N-1)!\, 2^{N+1}} \left(\prod_{j=1}^{N} a_j\right)^{-1} \left[\sum_{\vec{\varepsilon} \in \{-1,1\}^N} \left(z_{\vec{\varepsilon}}(x)\right)^{N-1} \operatorname{sign}(z_{\vec{\varepsilon}}(x)) \prod_{j=1}^{N} \varepsilon_j\right] \quad (5.43)$$

with

$$z_{\vec{\varepsilon}}(x) = x + \sum_{j=1}^{N} (\varepsilon_j a_j - c_j) \ . \quad (5.44)$$

The derivation of Equation (5.43) is outlined in Section A.3.3. The sum over $\vec{\varepsilon}$ indicates that the summation has to be done over all 2^N vectors of signs $\vec{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)$ with $\varepsilon_j = \pm 1$. Using Equation (5.43) the probability distribution of the directed flows can be calculated when inserting $\gamma_{qr}^1 a_1'$ and $\gamma_{qr}^1 b_j'$ for $c_i - a_i$ and $c_i + a_i$, respectively. This equation cannot be numerically evaluated[6] as will be discussed in Section 5.4.1.1 on Page 98.

5.4.1. The Moments of the Absolute Flows

The absolute flow $f_{ij}^a = |f_{ij}|$, without the information on the direction, is of great interest. It was discussed in Section 5.3 and is crucial to define quantities like the capacities for the links. The probability distribution of the absolute flow f_{ij}^a can be calculated from the distribution of the directed flows by

$$p_{ij}^a\left(f_{ij}^a\right) = H(f_{ij}) \left(p_{ij}(f_{ij}) + p_{ij}(-f_{ij})\right) \quad (5.45)$$

[6] Using high precision numerics, Equation (5.43) was evaluated for small N but the needed resources were high, so that a Monte-Carlo simulation as done in Section 5.3 is the better approach.

where $H(x)$ is the Heaviside function that is one for $x > 0$, zero for $x < 0$, and 0.5 at $x = 0$. The probability on the negative semiaxis is mirrored and added to the positive semiaxis.

To calculate the moments, a method similar to the derivation in Bradley and Gupta [21] is used. They take advantage of the properties of the characteristic function as the sum over N nonidentical uniformly distributed random variables, like in Equation (5.41), becomes a product over the respective characteristic functions (see Section A.3.1).

The characteristic function[7] of Equation (5.45) is found to be

$$\tilde{p}^a_{ij}(k) = \frac{1}{2\pi}\tilde{H}(k) * \Big(\tilde{p}_{ij}(k) + \tilde{p}_{ij}(-k)\Big) \tag{5.46}$$

and the characteristic function, or Fourier transform, of the Heaviside function is given by

$$\tilde{H}(k) = \pi\delta(k) + \frac{1}{ik} \tag{5.47}$$

so that

$$\tilde{p}^a_{qr}(k) = \underbrace{\frac{1}{2}\cdot\Big(\tilde{p}_{qr}(k) + \tilde{p}_{qr}(-k)\Big)}_{\tilde{g}^{qr}_1} + \underbrace{\frac{1}{2\pi}\frac{1}{ik} * \Big(\tilde{p}_{qr}(k) + \tilde{p}_{qr}(-k)\Big)}_{\tilde{g}^{qr}_2} \ . \tag{5.48}$$

For $\tilde{g}^{qr}_2$ to be finite, the characteristic functions $\tilde{p}_{qr}(k)$ and $\tilde{p}_{qr}(-k)$ are assumed to vanish sufficiently fast for high k.

The characteristic function of the uniform distribution $U(x; c_j - a_j, c_j + a_j)$ is given by the modulated sinc function[8] $\tilde{U}(k; c_j - a_j, c_j + a_j) = e^{ic_j k}\,\mathrm{sinc}(a_j k)$ [21]. From Equation (5.41), we get

$$\tilde{p}^a_{qr}(k) = \prod_{j=1}^{N} \tilde{U}\big(k; \gamma^j_{qr}a'_j, \gamma^1_{qr}b'_j\big) = \prod_{j=1}^{N} e^{ic_j k}\,\mathrm{sinc}(a_j k) \ , \tag{5.49}$$

with $c^{qr}_j = \gamma^j_{qr}\frac{a'_j+b'_j}{2}$ and $a^{qr}_j = \left|\gamma^j_{qr}\frac{b'_j-a'_j}{2}\right|$[9].

For the general case, further properties of the moments can be calculated independent of the underlying distributions $p_{s^b_j}\big(s^b_j; \gamma^j_{qr}\big)$. The n-sth raw moment[10] can be calculated

[7] The characteristic function of f is marked by a tilde, $\tilde{f}$.

[8] The modulation, given by $e^{ic_j k}$, accounts for translation and is zero for uniform distributions centered at zero.

[9] For simplicity, we define $a^{qr}_j \geq 0$ without loss of generality

[10] The raw moments denote the moments about zero, $\mu'_k = E\left[X^k\right]$, in contrast to the central moments that are centered about the mean, $\mu_k = E\left[\left(X - \langle X\rangle\right)^k\right]$.

from the characteristic function by evaluating the n-sth derivation of the characteristic function at zero [77]:

$$\langle x^q \rangle = i^{(-q)} \left[\frac{d^q}{dk^q} \tilde{p}(k) \right]_{k=0} . \tag{5.50}$$

The raw moments of the absolute flow follow from

$$\left\langle \left(f_{ij}^a \right)^n \right\rangle = i^{(-n)} \left[\frac{d^n}{dk^n} \tilde{p}_{qr}^a(k) \right]_{k=0} = i^{(-n)} \left[\frac{d^n}{dk^n} \tilde{g}_1^{qr}(k) \right]_{k=0} + i^{(-n)} \left[\frac{d^n}{dk^n} \tilde{g}_2^{qr}(k) \right]_{k=0} . \tag{5.51}$$

As mentioned above, one of the aims is to estimate the parameters of the generalized gamma distribution that approximates the distribution of the absolute flows, so we focus on the first three moments since these are sufficient to calculate the parameters of the generalized gamma distribution.

The moments of terms $\tilde{g}_1^{qr}$ and $\tilde{g}_2^{qr}$ in Equation (5.48) are analyzed separately as follows from Equation (5.51). By using symmetry considerations, the moments of the term $\tilde{g}_1^{qr}$ can be simplified, whereas $\tilde{g}_1^{qr}$ is always an even function by definition[11]. The derivative of an odd function is always an even function and vice versa [22]. By construction, odd functions are zero at the origin, so that

$$\langle (\tilde{g}_1^{qr})^n \rangle = i^{-n} \left[\frac{d^n}{dk^n} \tilde{g}_1^{qr} \right]_{k=0} = 0 \qquad \text{for odd } n \tag{5.52}$$

follows[12]. From the property of additivity of the cumulants (see Section A.3.1) follows for $n = 2$

$$\left\langle (\tilde{g}_1^{qr})^2 \right\rangle = \left(\sum_{j=1}^{N} \left(\mathbf{Y}_{qj}^+ - \mathbf{Y}_{rj}^+ \right)^2 \sigma_j^2 \right) + \left(\sum_{j=1}^{N} \left(\mathbf{Y}_{qj}^+ - \mathbf{Y}_{rj}^+ \right) \mu_j \right)^2 , \tag{5.53}$$

where μ_j^2 and σ_j^2 are the mean and the variance of the in- and out-flow s_j^b at vertex j.

To analyze the convolution part of Equation (5.48), $\tilde{g}_2^{qr}$ is plugged into Equation (5.50) and $\langle (\tilde{g}_2^{qr})^n \rangle$ is expanded, so that

$$\langle (g_2^{qr})^n \rangle = \frac{i^{-(n+1)}}{2\pi} \left[\frac{d^n}{dk^n} \left(\int_{-\infty}^{\infty} \frac{1}{k-s} \left(\tilde{p}_{qr}(s) + \tilde{p}_{qr}(-s) \right) ds \right) \right]_{k=0} . \tag{5.54}$$

The derivation can not be drawn under the integral since $\frac{1}{k-s}$ is not continuous for k equal to s. To calculate the integral, $\frac{1}{k-s}$ is expanded into a Taylor series around $k = 0$,

[11] The function $f(k) = \tilde{p}_{qr}(k) + \tilde{p}_{qr}(-k)$ always satisfies $f(k) = f(-k)$

[12] This can also be shown using the cumulants.

$\frac{1}{k-s} = -\left(\frac{1}{s} + \frac{k}{s^2} + \frac{k^2}{s^3} + \frac{k^3}{s^4} + \ldots\right)$, yielding

$$\begin{aligned}\langle (g_2^{qr})^n \rangle &= \frac{1}{2\pi i^{n+1}} \left[\frac{d^n}{dk^n} \left(\int_{-\infty}^{\infty} \left(\frac{1}{s} + \frac{k}{s^2} + \frac{k^2}{s^3} + \frac{k^3}{s^4} + \ldots \right) \right.\right. \\ &\quad \left.\left. (\tilde{p}_{qr}(s) + \tilde{p}_{qr}(-s))\, ds \right) \right]_{k=0} \qquad (5.55) \\ &= -\frac{n!}{2\pi i^{n+1}} \int_{-\infty}^{\infty} \frac{1}{s^{n+1}} \left(\tilde{p}_{qr}(s) + \tilde{p}_{qr}(-s)\right) ds \qquad (5.56)\end{aligned}$$

which follows from the fact that k can be drawn in front of the integral and that the n-th derivative gives a prefactor of the form $\left(\frac{1}{s^{n+1}} + \frac{k^1}{s^{n+2}} + \ldots\right)$ to the integral, so that only the term constant in k is left at $k = 0$. With the same argument as for Equation (5.52), we find

$$\langle (g_2^{qr})^n \rangle = 0 \qquad \text{for odd } n\,, \qquad (5.57)$$

since the sum of $\tilde{p}_{qr}(s)$ and $\tilde{p}_{qr}(-s)$ is an even function that is multipied with an odd function $\frac{1}{s^{n+1}}$, yielding an odd function. The integral from minus infinity to plus infinity over an odd function is zero.

To summarize, for the first three moments of the absolute flow we find

$$\begin{aligned}\left\langle \left(f_{ij}^a\right) \right\rangle &= \langle g_2^{qr} \rangle \qquad (5.58) \\ \left\langle \left(f_{ij}^a\right)^2 \right\rangle &= \sum_{j=1}^{N} \sigma_j^2 + \mu_j^2 \qquad (5.59) \\ \left\langle \left(f_{ij}^a\right)^3 \right\rangle &= \left\langle (g_2^{qr})^3 \right\rangle\,, \qquad (5.60)\end{aligned}$$

so $\langle (g_2^{qr})^n \rangle$ for odd n is left to calculate. For the following, Equation (5.56) is separated into $\left\langle \left(g_{2+}^{qr}\right)^n \right\rangle$ and $\left\langle \left(g_{2-}^{qr}\right)^n \right\rangle$, with

$$\begin{aligned}\langle (g_2^{qr})^n \rangle &= -\frac{n!}{2\pi i^{n+1}} \int_{-\infty}^{\infty} \frac{1}{s^{n+1}} \left(\tilde{p}_{qr}(s) + \tilde{p}_{qr}(-s)\, ds\right) \qquad (5.61) \\ &= -\Bigg[\underbrace{\frac{n!}{2\pi i^{n+1}} \int_{-\infty}^{\infty} \frac{1}{s^{n+1}} \left(\tilde{p}_{qr}(s)\, ds\right)}_{\left\langle \left(g_{2+}^{qr}\right)^n \right\rangle} + \underbrace{\frac{n!}{2\pi i^{n+1}} \int_{-\infty}^{\infty} \frac{1}{s^{n+1}} \left(\tilde{p}_{qr}(-s)\, ds\right)}_{\left\langle \left(g_{2-}^{qr}\right)^n \right\rangle} \Bigg]\,. \qquad (5.62)\end{aligned}$$

In the next sections, $\langle (g_2^{qr})^n \rangle$ for odd n will be calculated exactly as well as using different approximations.

5.4.1.1. Exact Moments of the Absolute Flow

First, an exact solution of $\langle (g_2^{qr})^n \rangle$ is derived. Inserting the characteristic function from Equation (5.49) into $\left\langle \left(g_{2+}^{qr}\right)^n \right\rangle$, we have

$$\begin{aligned} \left\langle \left(g_{2+}^{qr}\right)^n \right\rangle &= \frac{n!}{2\pi i^{n+1}} \int_{-\infty}^{\infty} \frac{1}{s^{n+1}} \prod_{j=1}^{N} e^{ic_j k} \operatorname{sinc}(a_j k)\, ds \\ &= \frac{n!}{2\pi i^{n+1}} \int_{-\infty}^{\infty} \frac{1}{s^{n+1}} e^{iyk} \prod_{j=1}^{N} \operatorname{sinc}(a_j k)\, ds \end{aligned} \tag{5.63}$$

with $y = \sum_{j=1}^{N} c_j$. This last equation is the same as solved by Bradley and Gupta [21] except for an additional factor of $\frac{1}{s^{n+1}}$ under the integral (see Equation (A.85) in Section A.3.3). Therefore, Equation (5.63) can be solved with the same method as described in Section A.3.3. The final result is given by

$$\begin{aligned} \left\langle \left(g_{2+}^{qr}\right)^n \right\rangle &= \left(\frac{2\pi n!}{2^N (N+n)!}\right) \left(\prod_{j=1}^{N} a_j^{qr}\right)^{-1} \left[\sum_{\varepsilon \in \{-1,1\}^N} \prod_{j=1}^{N} \varepsilon_i \right. \\ &\quad \left. \cdot \left(\sum_{j=1}^{N} c_j^{rq} + \vec{\varepsilon} \cdot \vec{a}^{qr}\right)^{N+n} \cdot \operatorname{sign}\left(\sum_{j=1}^{N} c_j^{rq} + \vec{\varepsilon} \cdot \vec{a}^{qr}\right) \right] \end{aligned} \tag{5.64}$$

where the summation over $\vec{\varepsilon}$ is done over all vectors of length N with entries of ± 1 as in Equation (5.43). The same can be done for $\left\langle \left(g_{2-}^{qr}\right)^n \right\rangle$, yielding the exact result of $\left\langle \left(f_{ij}^{a}\right)^n \right\rangle$ for odd n.

Although we found the exact result in Equation (5.64) for the odd moments, it has some flaws, as it cannot be computed numerically for large N. First, there are 2^N possible vectors for $\vec{\varepsilon}$ and thus an equal number of summands that have to be evaluated in a loop. For $N = 100$, the number of summands is around $1.268 \cdot 10^{30}$, which is infeasible. Furthermore, the expression $\sum_{j=1}^{N} c_j^{rq} + \vec{\varepsilon} \cdot \vec{a}^{qr}$ is raised to the power of $N + n$ yielding a large number that poses problems on numerical evaluation. Although the latter problem can be circumvented by using high precision numerics, evaluating the logarithm of some expressions and using the exponential function on aggregated values, both problems make Equation (5.64) impractical since there is no advantage over a Monte Carlo simulation to find the moments.

5.4.1.2. Gaussian Approximation

A much simpler approach to approximate Equation (5.62),

$$\langle (g_2^{qr})^n \rangle = -\Bigg[\underbrace{\frac{n!}{2\pi i^{n+1}} \int_{-\infty}^{\infty} \frac{1}{s^{n+1}} \tilde{p}_{qr}(s)\, ds}_{\langle (g_{2+}^{qr})^n \rangle} + \underbrace{\frac{n!}{2\pi i^{n+1}} \int_{-\infty}^{\infty} \frac{1}{s^{n+1}} \tilde{p}_{qr}(-s)\, ds}_{\langle (g_{2-}^{qr})^n \rangle} \Bigg] , \tag{5.65}$$

for odd n is to use a Gaussian approximation to the probability distributions of the in- and out-flows s_j at the vertices. The moments of the directed flow are given by $\mu_{f_{ij}} = \mu_{ij} = \sum_{k=1}^{N} \left(Y_{ik}^{+} - Y_{jk}^{+}\right) s_k$ and $\sigma_{f_{ij}} = \sigma_{ij} = \sum_{k=1}^{N} \left(Y_{ik}^{+} - Y_{jk}^{+}\right)^2 s_k^2$. The characteristic function of the Gaussian distribution $N(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ is given by $e^{i\mu t - \frac{\sigma^2 t^2}{2}}$.

To solve Equation (5.65), the following integrals are important and thus solved first. These integrals and the derivations shown in the following are also needed in the next section. The principal value (PV) of the integral $\int_{-\infty}^{\infty} t^{-1} e^{ist} e^{-\frac{\sigma^2 t^2}{2}} dt$ can be evaluated to [21]

$$\begin{aligned} \mathrm{PV} \int_{-\infty}^{\infty} t^{-1} e^{i\mu t} e^{-\frac{\sigma^2 t^2}{2}} dt &= \lim_{\epsilon \to 0} \Bigg[\int_{-\infty}^{-\epsilon} t^{-1} e^{i\mu t} e^{-\frac{\sigma^2 t^2}{2}} dt + \int_{\epsilon}^{\infty} t^{-1} e^{i\mu t} e^{-\frac{\sigma^2 t^2}{2}} dt \Bigg] \qquad (5.66) \\ &= \lim_{\epsilon \to 0} \left[2i \int_{\epsilon}^{\infty} t^{-1} \sin(\mu t)\, e^{-\frac{\sigma^2 t^2}{2}} dt \right] \\ &= 2i \int_{0}^{\infty} t^{-1} \sin(\mu t)\, e^{-\frac{\sigma^2 t^2}{2}} dt \, . \qquad (5.67) \end{aligned}$$

To calculate the integral, it is easier to use the fact that the integral in 5.66 is a Fourier

transformation:

$$\begin{aligned}
\mathrm{PV}\int_{-\infty}^{\infty}\frac{1}{t}e^{-\frac{\sigma^2t^2}{2}}e^{ist}dt &= \mathcal{F}^{-1}\left(\frac{1}{t}e^{-\frac{\sigma^2t^2}{2}}\right)=\mathcal{F}^{-1}\left(\frac{1}{t}\right)*\mathcal{F}^{-1}\left(e^{-\frac{\sigma^2t^2}{2}}\right)\\
&= \frac{i}{2}\operatorname{sign}(s)*\frac{\sqrt{2\pi}}{\sigma}e^{-\frac{s^2}{2\sigma^2}}\\
&= -\frac{i\pi}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\operatorname{sign}(t-s)\,e^{-\frac{t^2}{2\sigma^2}}dt\\
&= -\frac{i\pi}{\sqrt{2\pi}\sigma}\left(\int_{s}^{\infty}e^{-\frac{t^2}{2\sigma^2}}dt-\int_{-\infty}^{s}e^{-\frac{t^2}{2\sigma^2}}dt\right)\\
&= -i\pi\left(1-\frac{2}{\sqrt{2\pi}\sigma}\int_{-\infty}^{s}e^{-\frac{t^2}{2\sigma^2}}dt\right)\\
&= -i\pi\left[1-\left(1+\operatorname{erf}\left(\frac{s}{\sqrt{2}\sigma}\right)\right)\right]\\
&= i\pi\operatorname{erf}\left(\frac{s}{\sqrt{2}\sigma}\right)\,, \qquad (5.68)
\end{aligned}$$

where $\operatorname{erf}(\cdot)$ is the error function. Again, the convolution is assumed to be finite. We also used the properties of the Gaussian distribution with respect to the Fourier transformation and furthermore $\mathcal{F}^{-1}\left(\frac{1}{t}\right)=i\pi\operatorname{sign}(s)$. A second important integral is [22]

$$\int_{-\infty}^{\infty}e^{i\mu t}e^{-\frac{\sigma^2t^2}{2}}dt = \frac{\sqrt{2\pi}}{\sigma}e^{-\frac{\mu^2}{2\sigma^2}}\,. \qquad (5.69)$$

We first calculate $\left\langle\left(g_{2+}^{qr}\right)^n\right\rangle$ from Equation (5.65). With the product rule we find for $n\geq 1$

$$\left\langle\left(g_{2+}^{qr}\right)^n\right\rangle=\frac{n!}{2\pi i^{n+1}}\left(i\mu_{qr}\int_{-\infty}^{\infty}\frac{1}{nt^n}e^{i\mu_{qr}t-\frac{\sigma_{qr}^2t^2}{2}}dt-\sigma^2\int_{-\infty}^{\infty}\frac{1}{nt^{n-1}}e^{i\mu_{qr}t-\frac{\sigma_{qr}^2t^2}{2}}dt\right)\,. \qquad (5.70)$$

For higher moments, we can use the fact that repeated integration by parts of the integral in Equation (5.70) and resubstituting the result back always yields terms of the

form

$$\begin{aligned}
& \phi_\gamma^m \int_{-\infty}^{\infty} t^{-(m-\gamma)} e^{i\mu t - \frac{\sigma^2 t^2}{\mu}} dt + \psi_\gamma^m \int_{-\infty}^{\infty} t^{-(m-\gamma-1)} e^{i\mu t - \frac{\sigma^2 t^2}{\mu}} dt \\
= \; & \phi_\gamma^m \frac{i\mu}{m-\gamma-1} \int_{-\infty}^{\infty} t^{-(m-\gamma-1)} e^{i\mu t - \frac{\sigma^2 t^2}{\mu}} dt + \\
& -\phi_\gamma^m \frac{\sigma^2}{m-\gamma-1} \int_{-\infty}^{\infty} t^{-(m-\gamma-2)} e^{i\mu t - \frac{\sigma^2 t^2}{\mu}} dt + \\
& \psi_\gamma^m \int_{-\infty}^{\infty} t^{-(m-\gamma-1)} e^{i\mu t - \frac{\sigma^2 t^2}{\mu}} dt \\
= \; & \left(\phi_\gamma^m \frac{i\mu}{m-(\gamma+1)} + \psi_\gamma^m \right) \int_{-\infty}^{\infty} t^{-(m-(\gamma+1))} e^{i\mu t - \frac{\sigma^2 t^2}{\mu}} dt \\
& -\phi_\gamma^m \frac{\sigma^2}{m-(\gamma+1)} \int_{-\infty}^{\infty} t^{-(m-(\gamma+1))} e^{i\mu t - \frac{\sigma^2 t^2}{\mu}} dt + \\
= \; & \phi_{\gamma+1}^m \int_{-\infty}^{\infty} t^{-(m-(\gamma+1))} e^{ist - \frac{\sigma^2 t^2}{s}} dt + \psi_{\gamma+1}^m \int_{-\infty}^{\infty} t^{-(m-(\gamma+1)-1)} e^{ist - \frac{\sigma^2 t^2}{s}} dt
\end{aligned} \tag{5.71}$$

for $\gamma \leq m-1$. From the above equations follows

$$\phi_\gamma^m = \phi_{\gamma-1}^m \frac{i\mu}{m-\gamma} + \psi_{\gamma-1}^m \qquad \psi_\gamma^m = -\phi_{\gamma-1}^m \frac{\sigma^2}{m-\gamma} \,, \tag{5.72}$$

with $\phi_0^m = 1$ and $\psi_0^m = 0$. It can be shown by induction that ϕ_γ^m and ψ_γ^m are the prefactors of the respective integrals after γ times integrating by parts. For the case of $m = n+1$ and $\gamma = m-1 = n$, the two integrals in Equation (5.70) are given by

$$\left\langle \left(g_{2+}^{qr}\right)^n \right\rangle = \frac{n!}{2\pi i^{n+1}} \left(\phi_n^{n+1} \int_{-\infty}^{\infty} t^{-1} e^{i\mu_{qr} t - \frac{\sigma_{qr}^2 t^2}{\mu_{qr}}} dt + \psi_n^{n+1} \int_{-\infty}^{\infty} e^{i\mu_{qr} t - \frac{\sigma_{qr}^2 t^2}{\mu_{qr}}} dt \right) \tag{5.73}$$

$$= \frac{n!}{2\pi i^{n+1}} \left(\phi_n^{n+1} \, i\pi \, \mathrm{erf}\!\left(\frac{\mu_{qr}}{\sqrt{2}\sigma_{qr}} \right) + \psi_n^{n+1} \frac{\sqrt{2\pi}}{\sigma_{qr}} e^{-\frac{\mu_{qr}^2}{2\sigma_{qr}^2}} \right) \,. \tag{5.74}$$

The integrals in Equation (5.73) were calculated in the beginning of this section.

For $n = 1$, we find

$$\begin{aligned}
\left\langle g_{2+}^{qr} \right\rangle &= -\frac{1}{2\pi} \left(i\mu_{qr} \int_{-\infty}^{\infty} \frac{1}{t} e^{i\mu_{qr} t - \frac{\sigma_{qr}^2 t^2}{2}} dt - \sigma^2 \int_{-\infty}^{\infty} e^{i\mu_{qr} t - \frac{\sigma_{qr}^2 t^2}{2}} dt \right) \\
&= \frac{\mu}{2} \, \mathrm{erf}\!\left(\frac{\mu_{qr}}{\sqrt{2}\sigma_{qr}} \right) + \frac{\sigma_{qr}}{\sqrt{2\pi}} e^{-\frac{\mu_{qr}^2}{2\sigma_{qr}^2}} \,,
\end{aligned} \tag{5.75}$$

using Equation (5.72). Since for the characteristic function of the Gaussian distribution $\tilde{N}(-s; \mu_{qr}, \sigma_{qr}) = \tilde{N}(s; -\mu_{qr}, \sigma_{qr})$ holds, we have $\left\langle \left(g_{2+}^{qr}\right)^n \right\rangle = \left\langle \left(g_{2-}^{qr}\right)^n \right\rangle$, and the first

moment of the absolute value of the Gaussian distribution is

$$\left\langle f_{qr}^{a} \right\rangle = \mu_{qr} \operatorname{erf}\left(\frac{\mu_{qr}}{\sqrt{2}\sigma_{qr}} \right) + \frac{\sqrt{2}\sigma_{qr}}{\sqrt{\pi}} e^{-\frac{\mu_{qr}^2}{2\sigma_{qr}^2}} \ . \tag{5.76}$$

The third moment is given by

$$\left\langle \left(f_{qr}^{a}\right)^3 \right\rangle = \left\langle \left(g_{2+}^{qr}\right)^3 \right\rangle = \left(\mu_{qr}^3 + 3\mu_{qr}\sigma_{qr}^2\right) \operatorname{erf}\left(\frac{\mu_{qr}}{\sqrt{2}\sigma_{qr}} \right) + 2\frac{\mu_{qr}^2\sigma_{qr} + 2\sigma_{qr}^3}{\sqrt{2\pi}} e^{-\frac{\mu_{qr}^2}{2\sigma_{qr}^2}} \tag{5.77}$$

using $\phi_3^4 = -\frac{i}{6}\left(\mu_{qr}^3 + 3\sigma_{qr}^2\mu_{qr}\right)$ and $\psi_3^4 = \frac{1}{6}\left(\mu_{qr}^2\sigma_{qr}^2 + 2\sigma_{qr}^4\right)$, see Equation (5.72).

This distribution of the absolute values of Gaussian distributed random variables can be found as the "folded normal" distribution in the literature [80]. The derivation was shown here, because ϕ_{m-1}^m and ψ_{m-1}^m are needed in the next section. We define the recursive expressions

$$\phi_{n+1}(\mu,\sigma) = \phi_n^{n+1}(\mu,\sigma) \qquad \psi_{n+1}(\mu,\sigma) = \psi_n^{n+1}(\mu,\sigma) \tag{5.78}$$

with $\phi_n^{n+1}(\mu,\sigma)$ and $\psi_n^{n+1}(\mu,\sigma)$, see Equation (5.72).

5.4.1.3. Exact-Gaussian Mix Approximation

The sum of independent identical uniformly distributed random variables converges in good approximation for only a few summands to a Gaussian distribution. But if there are only few summands that differ substantially from the other summands, e.g. because of a big prefactor, the distribution of the sum can change drastically. Thus, we propose the following approach: Groups of vertices with in/out flows that have a similar distribution and thus similar influence on the link are approximated by a Gaussian distributions. The in/out flows with strongly deviating prefactors are treated exact using the solution given in Equation (5.64). Since the convolution of Gaussian distributions is again a Gaussian, all of the above defined groups can be collected into one distribution with mean m_a and variance σ_a^2. With Q being the number of vertices whose influence on the link $q \to r$ are treated exactly and w_j^{qr} and u_j^{qr} denoting the respective c_k^{qr} and a_k^{qr} with adopted indices, we find

$$\tilde{p}_{qr}(k) = \tilde{G}\left(k;\mu_a,\sigma_a^2\right) \cdot \prod_{j=1}^{Q} \tilde{U}\left(k; w_j^{qr} - u_j^{qr}, w_j^{qr} + u_j^{qr}\right) \ . \tag{5.79}$$

Using this approach, we calculate $\langle (g_2^{qr})^n \rangle = \langle \left(g_{2+}^{qr}\right)^n \rangle + \langle \left(g_{2-}^{qr}\right)^n \rangle$ as defined in Equation (5.62). The integral to be solved is

$$\begin{aligned}
\langle \left(g_{2+}^{qr}\right)^n \rangle &= \frac{n!}{2\pi i^{n+1}} \Bigg(\int_{-\infty}^{\infty} \frac{1}{s^{n+1}} \tilde{G}\left(s; \mu_a, \sigma_a^2\right) \cdot \\
&\quad \prod_{j=1}^{Q} \tilde{U}\left(s; w_j^{qr} - u_j^{qr}, w_j^{qr} + u_j^{qr}\right) ds \Bigg) \\
&= \frac{n!}{2\pi i^{n+1}} \left(\prod_{j=i}^{Q} u_j^{rq} \right)^{-1} \Bigg[\int_{-\infty}^{\infty} \frac{1}{s^{Q+n+1}} \tilde{G}\left(s; \mu_a, \sigma_a^2\right) \\
&\quad \cdot \exp\left(is \sum_{j=1}^{Q} w_j^{qr} \right) \left(\prod_{s=1}^{Q} \sin(u_s^{qr} s) \right) ds \Bigg] \\
&= \frac{n!}{2\pi i^{n+1}} \left(\prod_{j=i}^{Q} u_j^{rq} \right)^{-1} \Bigg[\int_{-\infty}^{\infty} \frac{1}{s^{Q+n+1}} \left(\prod_{s=1}^{Q} \sin(u_s^{qr} s) \right) \\
&\quad \cdot \exp\left(is \left(\mu_a + \sum_{j=1}^{Q} w_j^{qr} \right) - \frac{\sigma_a^2 s^2}{2} \right) ds \Bigg] . \qquad (5.80)
\end{aligned}$$

With the same reasoning as in Section 5.4.1, we find

$$\begin{aligned}
\langle \left(g_{2+}^{qr}\right)^n \rangle &= \frac{n!}{2\pi 2^Q i^{Q+n+1}} \left(\prod_{j=i}^{Q} u_j^{rq} \right)^{-1} \Bigg[\int_{-\infty}^{\infty} \frac{1}{s^{Q+n+1}} \\
&\quad \exp\left\{ is \left(\mu_a + \sum_{j=1}^{Q} \varepsilon_j u_j^{rq} + \sum_{j=1}^{Q} w_j^{qr} \right) - \frac{\sigma_a^2 s^2}{2} + \right\} ds \Bigg] . \qquad (5.81)
\end{aligned}$$

Using that μ_a is the sum over all c_j^{qr} treated as Gaussian distributions and $\sum_{j=1}^{2Q} w_j^{qr}$ is the sum over the remaining c_j^{qr}, the variable $\theta_{\vec{\varepsilon}}$ is defined to be

$$\begin{aligned}
\theta_{\vec{\varepsilon}} &= \mu_a + \left(\sum_{j=1}^{Q} w_j^{qr} \right) + \left(\sum_{j=1}^{Q} \varepsilon_j u_j^{rq} \right) \\
&= \left(\sum_{j=1}^{N} c_j^{qr} \right) + \left(\sum_{j=1}^{Q} \varepsilon_j u_j^{rq} \right) . \qquad (5.82)
\end{aligned}$$

This can be used to transform Equation (5.81) into

$$\begin{aligned}\left\langle \left(g_{2+}^{qr}\right)^n \right\rangle &= \frac{n!}{2\pi 2^Q i^{Q+n+1}} \left(\prod_{j=i}^{Q} u_j^{rq}\right)^{-1} \left[\sum_{\varepsilon\in\{-1,1\}^{2N}} \rho_{\vec{\varepsilon}} \right. \\ &\left. \cdot \int_{-\infty}^{\infty} \frac{1}{s^{Q+n+1}} \exp\left\{ is\theta_{\vec{\varepsilon}} - \frac{\sigma_a^2 s^2}{2} \right\} ds \right] \end{aligned} \tag{5.83}$$

with the 2^Q vectors of signs $\vec{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_Q) \in \{-1, 1\}^Q$, $\rho_\varepsilon = \prod_{j=1}^{Q} \varepsilon_j$. It can be seen that the integral in Equation (5.83) is the same as Equation (5.74). The solution is thus

$$\begin{aligned}\left\langle \left(g_{2+}^{qr}\right)^n \right\rangle &= \frac{n!}{2^Q i^{Q+n+1}} \left(\prod_{j=i}^{Q} u_j^{rq}\right)^{-1} \left[\sum_{\varepsilon\in\{-1,1\}^Q} \left(\prod_{j=1}^{Q} \varepsilon_j\right) \right. \\ &\left. \left(\psi_{(Q+n)}(\theta_{\vec{\varepsilon}}, \sigma_a) \frac{1}{\sqrt{2\pi}\sigma_a} e^{-\frac{\theta_{\vec{\varepsilon}}^2}{2\sigma_a^2}} + \phi_{(Q+n)}(\theta_{\vec{\varepsilon}}, \sigma_a) \frac{i}{2} \operatorname{erf}\left(\frac{\theta_{\vec{\varepsilon}}}{\sqrt{2}\sigma_a} \right) \right) \right] \end{aligned} \tag{5.84}$$

with $\phi_n(\cdot,\cdot)$ and $\psi_n(\cdot,\cdot)$ as defined in Equation (5.72) and Equation (5.78),

$$\phi_{n+1}(\mu,\sigma) = \phi_n^{n+1}(\mu,\sigma) = \phi_\gamma^m = \phi_{\gamma-1}^m \frac{i\mu}{m-\gamma} + \psi_{\gamma-1}^m \tag{5.85}$$

$$\psi_{n+1}(\mu,\sigma) = \psi_n^{n+1}(\mu,\sigma) = \psi_\gamma^m = -\phi_{\gamma-1}^m \frac{\sigma^2}{m-\gamma} \,. \tag{5.86}$$

with $\phi_0^m = 1$ and $\psi_0^m = 0$.

In the limit $Q \to N$, the Equation (5.86) has to be equal to the exact solution in Equation (5.64). In that case the Gaussian distribution can be understood to have zero mean and $\sigma_a \to 0$. For that case, the error function converges to the $\operatorname{sign}(\theta_{\vec{\varepsilon}})$ function and $\phi_{(Q+n)}(\theta_{\vec{\varepsilon}}, \sigma_a) = \frac{i^{(Q+n)}}{(Q+n)!}\, \theta_{\vec{\varepsilon}}^{(Q+n)}$. The behavior of the first part for σ_a can be derived from Equation (5.73): The integral corresponding to the term $e^{-\frac{\theta_{\vec{\varepsilon}}^2}{2\sigma_a^2}}$ goes to zero for $\sigma_a \to 0$. Thus, in the limit $Q \to N$, Equation (5.86) is equal to the result found in Section 5.4.1.1.

For the same reasons as in Section 5.4.1.2, we have $\left\langle \left(g_{2+}^{qr}\right)^n \right\rangle = \left\langle \left(g_{2-}^{qr}\right)^n \right\rangle$, so that

$$\begin{aligned}
\left\langle f_{qr}^{a} \right\rangle &= \frac{2}{2^Q i^{Q+2}} \left(\prod_{j=i}^{Q} u_j^{rq} \right)^{-1} \left[\sum_{\varepsilon \in \{-1,1\}^Q} \left(\prod_{j=1}^{Q} \varepsilon_j \right) \right. \\
&\left. \left(\psi_{(Q+1)}(\theta_{\vec{\varepsilon}}, \sigma_a) \frac{1}{\sqrt{2\pi}\sigma_a} e^{-\frac{\theta_{\vec{\varepsilon}}^2}{2\sigma_a^2}} + \phi_{(Q+1)}(\theta_{\vec{\varepsilon}}, \sigma_a) \frac{i}{2} \operatorname{erf}\!\left(\frac{\theta_{\vec{\varepsilon}}}{\sqrt{2}\sigma_a} \right) \right) \right] \quad (5.87) \\
\left\langle \left(f_{qr}^{a}\right)^3 \right\rangle &= \frac{12}{2^Q i^{Q+4}} \left(\prod_{j=i}^{Q} u_j^{rq} \right)^{-1} \left[\sum_{\varepsilon \in \{-1,1\}^Q} \left(\prod_{j=1}^{Q} \varepsilon_j \right) \right. \\
&\left. \left(\psi_{(Q+3)}(\theta_{\vec{\varepsilon}}, \sigma_a) \frac{1}{\sqrt{2\pi}\sigma_a} e^{-\frac{\theta_{\vec{\varepsilon}}^2}{2\sigma_a^2}} + \phi_{(Q+3)}(\theta_{\vec{\varepsilon}}, \sigma_a) \frac{i}{2} \operatorname{erf}\!\left(\frac{\theta_{\vec{\varepsilon}}}{\sqrt{2}\sigma_a} \right) \right) \right] \quad (5.88)
\end{aligned}$$

For the result in Equation (5.88) to have computational advantages over Equation (5.64), Q has to be maximally in the order of ten, so that Equation (5.88) is numerically feasible. The approximations can be expected to be better than the results in Section 5.4.1.2, as at least some of the source and sink distributions are treated exactly. The question that arises is which sources and sinks should be treated exactly and which can be collected into the Gaussian approximation.

As can be seen in Equation (5.64), only the contribution of the $\vec{a}^{qr}$ to the exact solution varies for changing $\vec{\varepsilon}$, while the contribution of the $\vec{c}^{qr}$ stays constant. Thus, the $\vec{a}^{qr}$ values of Equation (5.64) have to be considered for deciding which sources or sinks to treat exactly for a given link. Two methods were tested. First, the $\vec{a}^{qr}$ values are clustered, so that the prefactor a_j^{qr} is assigned to the cluster q_k when the distance $\left| q_k - a_j^{qr} \right| < r_{max}$ is smaller than a given maximal distance r_{max}. As sums of n i.i.d. random values drawn from a uniform distribution converge to a Gaussian distribution very fast, e.g. $n = 4$ yields a distribution that has only little deviations, all prefactors a_j^{qr} that belong to clusters with less than 4 elements are now treated exactly[13]. The second method is to consider the Q largest a_j^{qr} values, as these can be expected to dominate the moments.

Both methods were found to perform equally good for a given number of Q. However, the first method described is harder to control, as the number of Q, which follows from a given r_{max} is not fixed. The selected a_j^{qr} are almost always the largest in both methods, so consequently in the following, we are using the second method and treat only the Q largest values of the a_j^{qr} exactly in Equation (5.82). The following results are shown for $Q = 6$.

[13] As all a_j^{qr} belonging to one cluster deviate by no more than $2r_{max}$, the convergence of the sum of the in- and out-flows weighted by these a_j^{qr} are assumed to have a similar convergence as the case of i.i.d. uniform distributed random variables.

5.4.2. Matching of Load and Generation

As said before, there are unlimited possibilities how to balance a mismatch $\sum_j g_j + l_j \neq 0$. Only for the global factor policy (see Section 5.1.1.2), we found an analytical description of the distribution of the matched in- and out-flow at each vertex. This is presented in this section.

We assume that the generation and load at each node are given by the random variables g_i and l_i, whereas the generation is drawn from a uniform distribution of values between a and b. The balanced generation g_i^b and load l_i^b are defined by

$$g_i^b = \alpha g_i \qquad \text{and} \qquad l_i^b = \beta l_i \ , \tag{5.89}$$

where α and β are chosen such that

$$\sum_i \left(g_i^b + l_i^b\right) = 0 \ , \tag{5.90}$$

with $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$ to ensure that no artificial generation and load is introduced. This constraint is equal to $0 \leq g_i^b \leq g_i$ and $0 \leq l_i^b \leq l_i$. The mismatch in the unbalanced in- and out-flow is given by

$$M = \sum_i g_i + \sum_i l_i \ . \tag{5.91}$$

For $M > 0$, there is an excess of power, so that $\beta = 1$ and

$$\alpha = -\frac{\sum_i l_i}{\sum_i g_i} \ . \tag{5.92}$$

ensures that Equation (5.90) holds. With the same reasoning, we find $\alpha = 1$ and

$$\beta = -\frac{\sum_i g_i}{\sum_i l_i} \tag{5.93}$$

for $M < 0$.

The distribution of α for $M > 0$ is a ratio distribution of the sum of all generation and all load. Let q_G be the probability that $M > 0$, so that α is one and thus $g_j^b = g_j$. With probability $1 - q_G$, we have $M < 0$ and the matched in- and out-flow is given by αg_j so that

$$p_{g_i^b}(x) = q_G \cdot p_{g_i^b}(x|M > 0) + (1 - q_G)\, p_{g_j}(x) \ . \tag{5.94}$$

The distribution of g_i^b consists of two distributions weighted by q_G.

To calculate $p_{g_i^b}(x|M > 0)$, the ratio distribution of $\alpha = -\frac{\sum_i l_i}{\sum_i g_i}$ has to be calculated and subsequently the product distribution of $\alpha \cdot g_i$. The sum of all loads and the sum

of all generations are independent by definition. For the product distribution of $\alpha \cdot g_i$, we assume that the contribution of the single g_i to the sum $\sum_i g_i$ is small enough so that the correlation of each generator with the sum is negligible and further α and g_i are assumed to be independent. The same has to be done for β and the distribution $p_{l_i^b}(x|M<0)$.

In general, the distribution function of the ratio $Z = \frac{Y}{X}$ with independent $Y \sim p_Y(y)$ and $X \sim p_X(x)$ with $X, Y > 0$ is given by [70]

$$p_Z(z) = \int_0^\infty y\, p_Y(y)\, p_X(zy) dy\ . \tag{5.95}$$

For large enough N, the sums $\sum_i l_i$ and $\sum_i g_i$ may converge to a Gaussian distribution and we assume that this is the case. Furthermore, the sum of all loads and the sum of all generation are uncorrelated by construction. The ratio distribution of two independent Gaussian distributions was shown to be [65]

$$p_Z(z) = \frac{b(z)\cdot c(z)}{a^3(z)} \frac{1}{\sqrt{2\pi}\sigma_x\sigma_y}\left[2\Phi\left(\frac{b(z)}{a(z)}\right) - 1\right] + \frac{1}{a^2(z)\cdot \pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{\mu_x^2}{\sigma_x^2}+\frac{\mu_y^2}{\sigma_y^2}\right)} \tag{5.96}$$

with

$$a(z) = \sqrt{\frac{1}{\sigma_x^2}z^2 + \frac{1}{\sigma_y^2}}\ , \qquad b(z) = \frac{\mu_x}{\sigma_x^2}z + \frac{\mu_y}{\sigma_y^2}\ , \tag{5.97}$$

$$c(z) = e^{\frac{1}{2}\frac{b^2(z)}{a^2(z)} - \frac{1}{2}\left(\frac{\mu_x^2}{\sigma_x^2}+\frac{\mu_y^2}{\sigma_y^2}\right)}\ , \tag{5.98}$$

and $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du$. If $\mu_x = \mu_y = 0$, Equation (5.96) becomes a Cauchy distribution [38, 82].

Assuming that the absolute values of the coefficients of variation $|c_v| = \left|\frac{\sigma}{\mu}\right|$ are small enough, Equation (5.96) can be approximated by [65]

$$p_\alpha(z) = \frac{\sigma_Y^2\mu_X + \sigma_X^2\mu_Y z}{\sqrt{2\pi\left(\sigma_Y^2 + \sigma_X^2 z^2\right)^3}} \exp\left(-\frac{\left(\mu_Y - \mu_X\alpha\right)^2}{2\left(\sigma_Y^2 + \sigma_X^2 z^2\right)}\right)\ . \tag{5.99}$$

Hayya et al. [62] approximate the mean ω_i^α and variance $\left(\zeta_i^\alpha\right)^2$ of p_α for the case of a normal ratio distribution with a second order Taylor series expansion, finding

$$\begin{aligned} \omega_i^\alpha &\simeq -\left(\frac{\mu_l}{\mu_g} + \frac{\sigma_g^2\mu_l}{\mu_g^3}\right) \\ \left(\zeta_i^\alpha\right)^2 &\simeq \frac{\sigma_g^2\mu_l^2}{\mu_g^4} + \frac{\sigma_l^2}{\mu_g^2}\ , \end{aligned} \tag{5.100}$$

where μ_g, σ_g^2 and μ_l, σ_l^2 are the mean and the variance of the aggregated generation and load, respectively. The minus sign for ω_i^α is due to the sign in Equation (5.92).

For the following calculations, Equation (5.99) is not practical. The lognormal, Weibull and gamma distributions are found to be very similar to the ratio distribution in Equation (5.99), at least for parameter ranges considered here. Using a gamma distribution, the calculation of the product distribution of $\alpha \cdot g_i$ has been found to be feasible.

The gamma distribution is defined as

$$\text{Gamma}(x; k, \theta) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \, \Gamma(k)} \, , \tag{5.101}$$

for $x > 0$ and $k, \theta > 0$. With the mean $k\theta$ and the variance $k\theta^2$, the distribution p_α can be approximated by a gamma distribution with the parameters

$$k_i^\alpha = \frac{(\omega_i^\alpha)^2}{(\zeta_i^\alpha)^2} \tag{5.102}$$

$$\theta_i^\alpha = \frac{(\zeta_i^\alpha)^2}{\omega_i^\alpha} \tag{5.103}$$

by the use of Equation (5.100).

To find the distribution of the product αg_i, we then used

$$p_{g_i^b}(z|M>0) = \int_{x=0}^{1} \int_{y=0}^{\infty} \text{Gamma}(x; k_i^\alpha, \theta_i^\alpha) \; U(y; 0, b) \, \delta(z - xy) \, dxdy \, , \tag{5.104}$$

for $z > 0$ and $p_{g_i^b}(z < 0|M > 0) = 0$. The integration for x from zero to one reflects that $0 \leq \alpha \leq 1$ by construction. The generation is drawn from a uniform distribution $U(y; 0, \gamma_i^g)$ by definition of the model, see Section 5.2, where γ_i^g denotes the upper bound of the generation at vertex i. The product distribution $X \cdot Y$ for two independent random variables is given by [70]

$$p(x) = \int_0^\infty \frac{1}{y} f\left(\frac{x}{y}\right) g(y) \, dy \, , \tag{5.105}$$

where $X > 0$ is drawn from the distribution $f(x)$ and $Y > 0$ from the distribution $g(x)$. Thus using $U(x; 0, \gamma_i^g) = \frac{1}{\gamma_i^g} \left(H(x) - H(x - \gamma_i^g)\right)$, the product distribution becomes

$$p_{g_i^b}(z|M>0) = \frac{1}{\gamma_i^g} \int_0^1 \frac{1}{x} \text{Gamma}(x; k_i^\alpha, \theta_i^\alpha) \left(H(z/x) - H(z/x - \gamma_i^g)\right) dx \tag{5.106}$$

for $z > 0$, where $H(\cdot)$ is the Heaviside function. Since z/x under the integration is always positive or zero, as z is describing the generation, the first Heaviside is always one

and the second $H(z/x - \gamma_i^g)$ is zero for $x > z/\gamma_i^g$. This defines the range where the two Heaviside functions do not cancel each other, so that

$$p_{g_i^b}(z|M>0) = \frac{1}{\gamma_i^g} \int_{z/\gamma_i^g}^{1} \frac{1}{x} \text{Gamma}(x; k_i^\alpha, \theta_i^\alpha)\, dx \tag{5.107}$$

for $z \leq \gamma_i^g$ and $p_{g_i^b}(z > \gamma_i^g) = 0$, because of the $\alpha \leq 1$ condition. This integral evaluates to

$$p_{g_i^b}(z|M>0) = \frac{1}{\gamma_i^g\, \theta_i^\alpha} \frac{\Gamma\left(k_i^\alpha - 1, \frac{z}{\gamma_i^g\, \theta_i^\alpha}\right) - \Gamma\left(k_i^\alpha - 1, \frac{1}{\theta_i^\alpha}\right)}{\Gamma(k_i^\alpha)}\,, \tag{5.108}$$

where $\Gamma(\cdot)$ and $\Gamma(\cdot,\cdot)$ are the Gamma and the incomplete Gamma function, respectively and we find

$$p_{g_i^b}(z) = \frac{1}{\gamma_i^g\, \theta_i^\alpha} \frac{\Gamma\left(k_i^\alpha - 1, \frac{z}{\gamma_i^g\, \theta_i^\alpha}\right) - \Gamma\left(k_i^\alpha - 1, \frac{1}{\theta_i^\alpha}\right)}{\Gamma(k_i^\alpha)} + (1 - q_G)\, p_{g_j}(z) \tag{5.109}$$

for $0 \leq z \leq \gamma_i^g$.

The raw moments are calculated by evaluating the integral

$$\mu_i^n = \int_0^{\gamma_i^g} x^n p_{g_i^b}(x) dx\,, \tag{5.110}$$

so that we find

$$\mu_{g_i^b} = \frac{\gamma_i^g}{2}\, \theta_i^\alpha \left(k_i^\alpha - \frac{\Gamma(k_i^\alpha + 1, 1/\theta_i^\alpha)}{\Gamma(k_i^\alpha)}\right) + (1 - q_G)\, \frac{\gamma_i^g}{2} \tag{5.111}$$

$$\sigma^2_{g_i^b} = \frac{\left(\gamma_i^g\right)^2}{3}\, \left(\theta_i^\alpha\right)^2 \left(k_i^\alpha\, (k_i^\alpha + 1) - \frac{\Gamma(k_i^\alpha + 2, 1/\theta_i^\alpha)}{\Gamma(k_i^\alpha)}\right) + (1 - q_G)\, \frac{\left(\gamma_i^g\right)^2}{3} - \left(\mu_i^G\right)^2\,. \tag{5.112}$$

The contribution from the first term, representing the influence of the balancing does not have a factor of q_G since this is already considered in the integration limits that follow from $0 \leq \alpha \leq 1$. To ensure the normalization of the distribution, the q_G used for evaluating Equation (5.111) and (5.112) is calculated from

$$q_G = \int_0^{\gamma_i^g} p_{g_i^b}(z) dz = 1 - \frac{\Gamma\left(k_i^\alpha, \frac{1}{\theta_i^\alpha}\right)}{\Gamma(k_i^\alpha)}\,, \tag{5.113}$$

as due to the approximations $\int_0^\infty p_{g_i^b}(z) dz = 1$ is not guaranteed.

To calculate the distribution of the matched load, the same method can be employed. The distribution of the factor β is approximated by a gamma distribution with the

parameters

$$k_i^\beta = \frac{\left(\omega_i^\beta\right)^2}{\left(\zeta_i^\beta\right)^2} \qquad \text{and} \qquad \theta_i^\beta = \frac{\left(\zeta_i^\beta\right)^2}{\omega_i^\beta} \tag{5.114}$$

which are derived from the first central moments

$$\omega_i^\beta \simeq -\left(\frac{\mu_g}{\mu_l} + \frac{\sigma_l^2 \mu_g}{\mu_l^3}\right) \tag{5.115}$$

$$\left(\zeta_i^\beta\right)^2 \simeq \frac{\sigma_l^2 \mu_g^2}{\mu_l^4} + \frac{\sigma_g^2}{\mu_l^2} \quad . \tag{5.116}$$

As argued in Section 5.2, the load fluctuates on a much smaller timescale than the generation and thus is assumed to be constant with value γ_i^l. The constant load is multiplied by β, so that we find for the distribution of the balanced load

$$p_{l_i^b}(z|M<0) = \int_0^1 p_\beta(x)\,\delta\big(x\gamma_i^l - z\big)\,dx = \frac{1}{|\gamma_i^l|} p_\beta\left(\frac{z}{\gamma_i^l}\right) . \tag{5.117}$$

This yields

$$p_{l_i^b}(z) = \frac{1}{|\gamma_i^l|} p_\beta\left(\frac{z}{\gamma_i^l}\right) + (1-q_L)\,p_{l_j}(z) \tag{5.118}$$

for $\gamma_i^l \le z \le 0$, with $p_\beta(x) \simeq \text{Gamma}\Big(x; k_i^\beta, \theta_i^\beta\Big)$. Analogous to Equation (5.94), we find the mean

$$\begin{aligned} \mu_{l_i^b} &= \gamma_i^l\,\theta_i^\beta \frac{\Gamma\Big(k_i^\beta+1\Big) - \Gamma\Big(k_i^\beta+1, {}^1\!/_{\theta_i^\beta}\Big)}{\Gamma\Big(k_i^\beta\Big)} + (1-q_L)\,\gamma_i^l \\ &= \gamma_i^l\,\theta_i^\beta \left(k_i^\beta - \frac{\Gamma\Big(k_i^\beta+1, {}^1\!/_{\theta_i^\beta}\Big)}{\Gamma\Big(k_i^\beta\Big)}\right) + (1-q_L)\,\gamma_i^l \end{aligned} \tag{5.119}$$

and the variance

$$\begin{aligned} \sigma_{l_i^b}^2 &= \left(\gamma_i^l\right)^2 \left(\theta_i^\beta\right)^2 \frac{\Gamma\Big(k_i^\beta+2\Big) - \Gamma\Big(k_i^\beta+2, {}^1\!/_{\theta_i^\beta}\Big)}{\Gamma\Big(k_i^\beta\Big)} - \left(\mu_{l_i^b}\right)^2 \\ &= \left(\gamma_i^l\right)^2 \left(\theta_i^\beta\right)^2 \left(k_i^\beta\left(k_i^\beta+1\right) - \frac{\Gamma\Big(k_i^\beta+2, {}^1\!/_{\theta_i^\beta}\Big)}{\Gamma\Big(k_i^\beta\Big)}\right) - \left(\mu_{l_i^b}\right)^2 . \end{aligned} \tag{5.120}$$

The probability that $M < 0$ is given by

$$q_L = \int_0^{\gamma_i^l} p_{g_i^l}(z)dz = 1 - \frac{\Gamma\left(k_i^\beta, \frac{1}{\theta_i^\beta}\right)}{\Gamma\left(k_i^\beta\right)} \,. \tag{5.121}$$

Both the central moments of the generation and of the load contain a term of the form $\frac{\Gamma(k+n,z)}{\Gamma(k)}$ with $n \geq 1$. This is numerically not feasible for large k since the gamma function diverges so that the numerator and the denominator are beyond numerical limits even if the ratio is numerically feasible. Using $\Gamma(k+1) = k\Gamma(k)$ and the regularized gamma function $Q(s,x) = \frac{\Gamma(s,x)}{\Gamma(s)}$, this term can rewritten to

$$\frac{\Gamma(k+1,z)}{\Gamma(k)} = kQ(k+n,z) \quad \text{and} \quad \frac{\Gamma(k+2,z)}{\Gamma(k)} = k(k+1)Q(k+n,z) \,. \tag{5.122}$$

The moments thus can be written as

$$\mu_{g_i^b} = \frac{\gamma_i^g}{2}\,\theta_i^\alpha\, k_i^\alpha \left[1 - Q(k_i^\alpha + 1, 1/\theta_i^\alpha)\right] + (1 - q_G)\,\frac{\gamma_i^g}{2} \tag{5.123}$$

$$\sigma^2_{g_i^b} = \frac{(\gamma_i^g)^2}{3}\,(\theta_i^\alpha)^2\; k_i^\alpha\,(k_i^\alpha + 1)\left[1 - Q(k_i^\alpha + 2, 1/\theta_i^\alpha)\right] + (1 - q_G)\,\frac{(\gamma_i^g)^2}{3} - \left(\mu_i^G\right)^2 \tag{5.124}$$

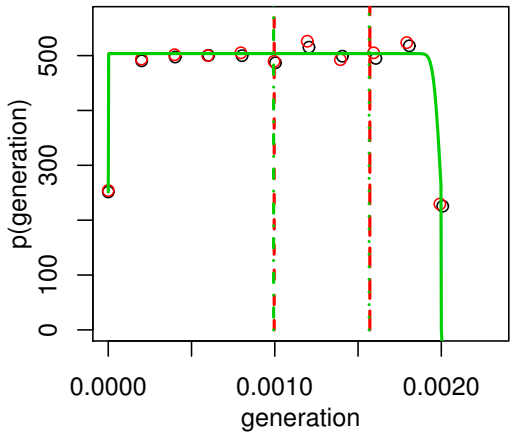
and

$$\mu_{l_i^b} = \gamma_i^l\,\theta_i^\beta\, k_i^\beta \left[1 - Q\left(k_i^\beta + 1, 1/\theta_i^\beta\right)\right] + (1 - q_L)\,\gamma_i^l \tag{5.125}$$

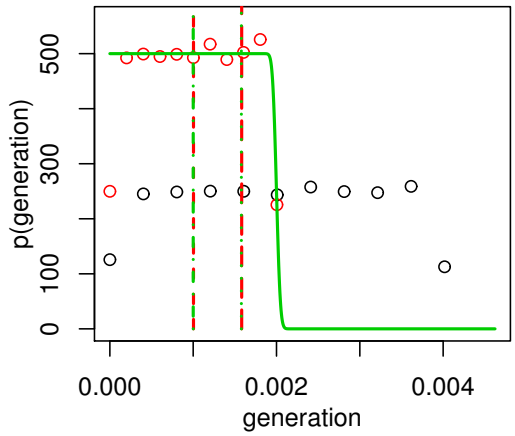
$$\sigma^2_{l_i^b} = \left(\gamma_i^l\right)^2 \left(\theta_i^\beta\right)^2 k_i^\beta \left(k_i^\beta + 1\right)\left[1 - Q\left(k_i^\beta + 2, 1/\theta_i^\beta\right)\right] - \left(\mu_{l_i^b}\right)^2 \,. \tag{5.126}$$

The regularized gamma function $Q(s,x)$ is implemented in numerical libraries, like the Gnu Scientific Library [57] that is used in the simulations. Using Equations (5.123) to (5.126), no numerical problems occur for large k.

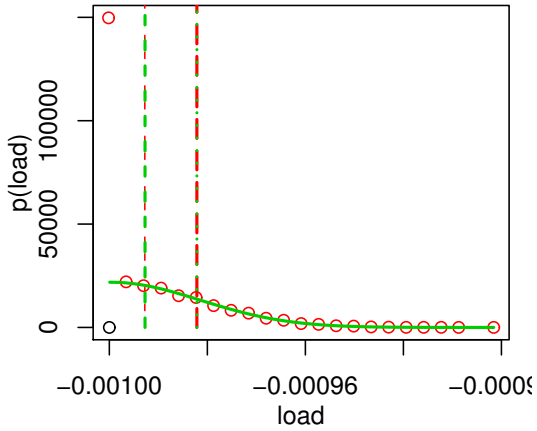
The derived equations and moments are compared to the data in Figure 5.10. As can be seen, the mean and the variance for the generation in Equation (5.123) and Equation (5.124), as well as the load in Equation (5.125) and Equation (5.126) match the moments from the data with very good precision. The distribution functions of the generation Equation (5.109) and the load Equation (5.117) are in good agreement with the data. Only for higher values of the generation there are larger deviations. This can be expected to be due to small correlations of the generation at a vertex to the sum over all generation, which were neglected in the calculations.



(a) $\kappa = 1$, *generation*



(b) $\kappa = 2$, *generation*



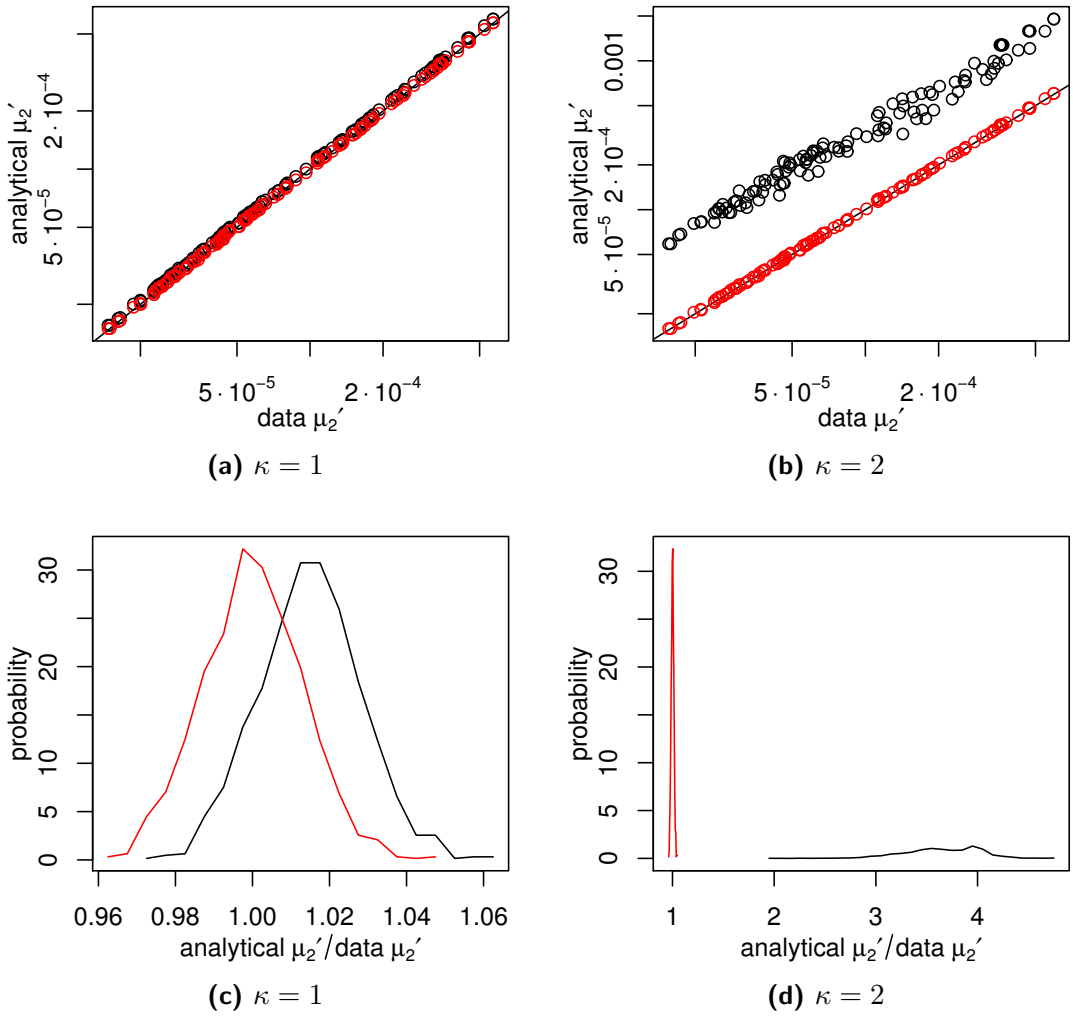
(c) $\kappa = 1$, *consumption*

Figure 5.10.: *The data for the matched generation and load and the analytical approximations. The black circles indicate the unmatched load and generation, the red circles the respective matched values. The green lines indicate the analytical derived distributions. The red vertical lines show the mean and the second raw moments for the data and the green the moments as derived analytical. For the consumption, the delta function is not shown and the unmatched load is indicated by one black circle. The figures are done for* $N = 1000$, $\beta_g = \beta_l = 0$ *and an exponential network with* $\lambda = 1$.

5.4.3. The Flow Distributions of the Links

In the previous sections, approximations to the moments of the flow over the links and the mean and variance of the matched generation and load were calculated. In this section, these results are put together and compared to simulation data.

First, the variance of the absolute flow can be calculated without any approximations



(a) $\kappa = 1$ **(b)** $\kappa = 2$

(c) $\kappa = 1$ **(d)** $\kappa = 2$

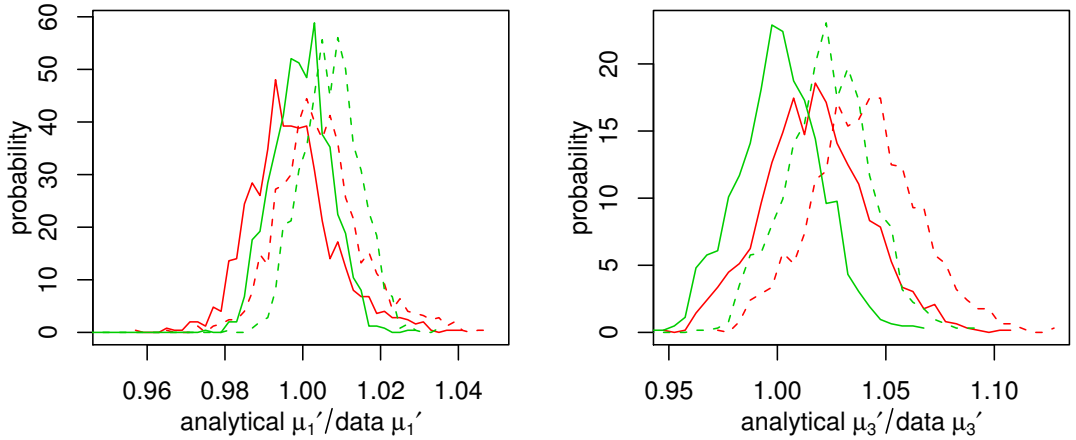
Figure 5.11.: *The second raw moment from the data (black) compared to the analytical derived results (red) shown in (a) and (b) for a network with exponential degree distribution with* $\lambda = 1$ *and* $\beta_L = \beta_S = 1$ *for* 10^4 *fluctuation realizations. In (c) and (d) histograms of the analytical results relative to the second moments of the data are plotted. For details see text.*

when neglecting the matching, see Equation (5.53). To incorporate the effects of the matching, the mean and variance of the matched in/out flow (see Equations (5.123) to (5.126)) are used for each vertex instead of the unmatched moments. This approach is a correction of the first two moments of the in/out flow, neglecting all higher order corrections.

In Figure 5.11, the second raw moments calculated analytically are compared to the second raw moments from the data. The black points represent the unmatched analytical calculated second moments against the second moments from the data, the red points are the analytical calculated second moments using the matching correction.

As can be seen in Figure 5.11a and Figure 5.11b, the red dots lie on the line $x = y$, indicating that the analytical second moments are close to being equal to the second moments from the data. In both cases, the matchinging correction improves the result, especially for $\kappa = 2$ in Figure 5.11b. The histograms of the analytical second moments normalized by the second moments calculated from the data are shown in Figure 5.11c and Figure 5.11d. Figure 5.11c shows the distributions of the relative deviations for $\kappa = 1$. The deviations are smaller than four percent when the matching corrections are used.

The first and third moments can only be calculated using approximations as derived in Section 5.4.1. Therefore, the deviation for the second moments is the best result that can be achieved with our approach, as no approximations were used. The case of $\kappa = 2$ in Figure 5.11d shows the large deviations for the unmatched flow and that the matching correction achieves the goal of correcting the moments. The plots of the moments are shown for $N = 100$ for clarity, the histograms are done with $N = 1000$, in both cases 10000 fluctuation realizations are used to obtain the simulation data.

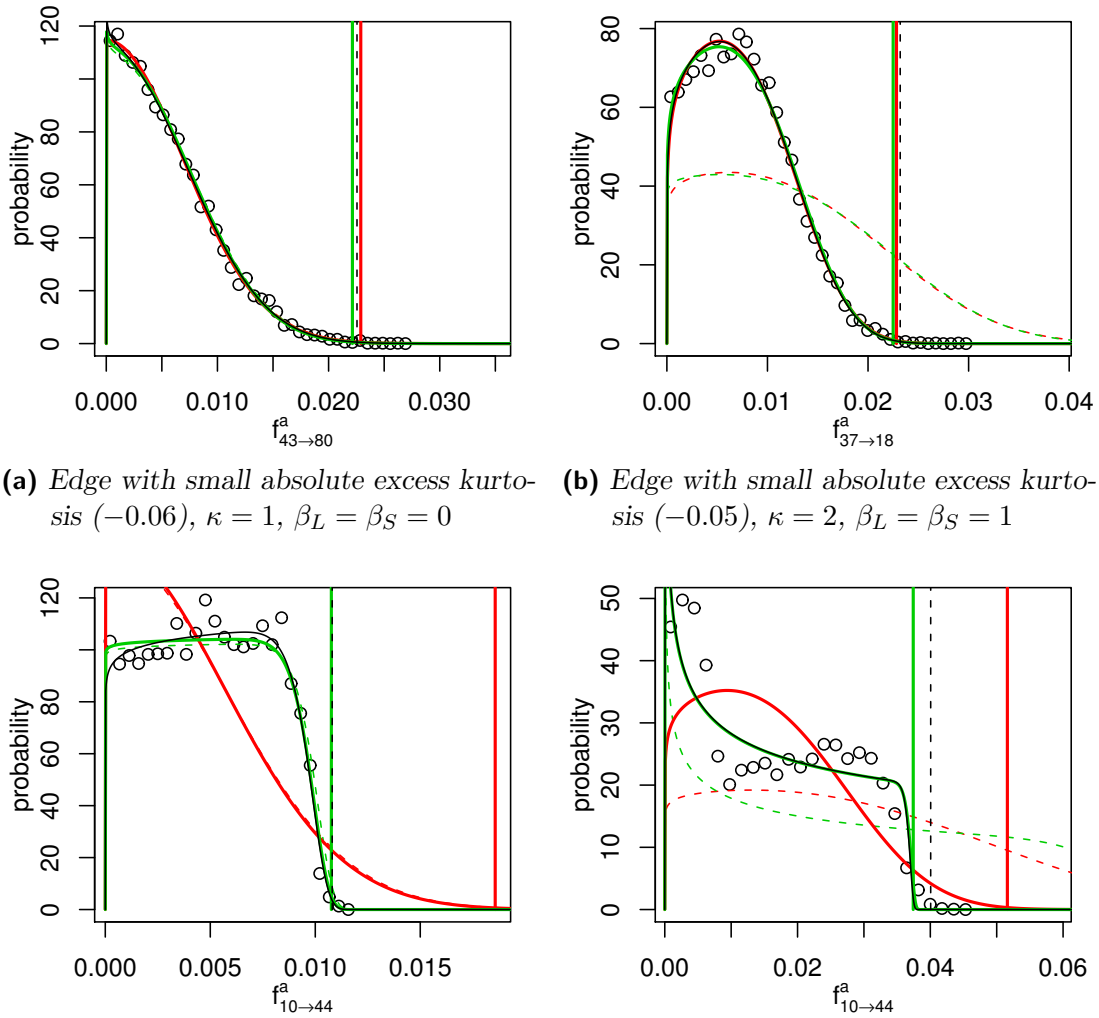


(a) *first raw moment* **(b)** *third raw moment*

Figure 5.12.: *The relative deviations of the first and third raw moments to the moments of the data using the Gaussian approximation (red) and the mixed approximation (green) for a network with exponential degree distribution with $N = 1000$, $\lambda = 1$, and $\kappa = 1$ as well as $\beta_L = \beta_S = 1$. The dashed lines show the respective approximations without the matching correction, for comparison.*

The approximations for the first and third raw moments, derived in Section 5.4.1.2 and 5.4.1.3, are tested against the moments of the data in Figure 5.12. Both approximations show only small deviations from the data, comparable to the deviations of the second moments in Figure 5.11c. However, the mixed approximation exhibits smaller deviations.

This finding is supported by Figure 5.13, where histograms of the absolute flows for selected links are shown. Similar to the work in Section 5.3, for each link the generalized gamma distribution is fitted to the moments (see Section A.3.2) using the moments



(a) *Edge with small absolute excess kurtosis* (-0.06), $\kappa = 1$, $\beta_L = \beta_S = 0$

(b) *Edge with small absolute excess kurtosis* (-0.05), $\kappa = 2$, $\beta_L = \beta_S = 1$

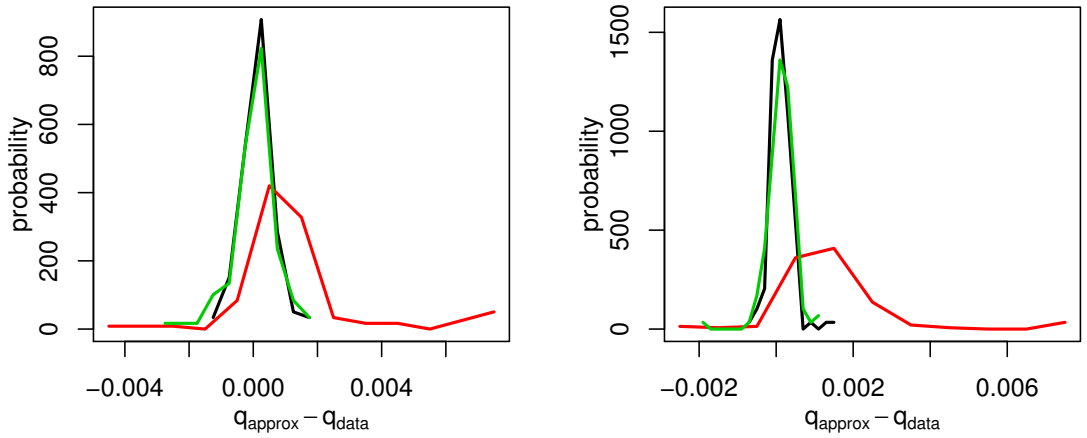
(c) *Edge with high absolute excess kurtosis* (-1.2), $\kappa = 1$, $\beta_L = \beta_S = 0$

(d) *Edge with high absolute excess kurtosis* (-1.18), $\kappa = 2$, $\beta_L = \beta_S = 1$

Figure 5.13.: *Typical distributions of the absolute flow and the analytical approximations. The black dots indicate the histogram of the data, the black line the fit with the generalized gamma distribution based on the moments of the data, and the vertical dashed line the* 99.9% *quantile calculated directly from the data. The red lines indicate the generalized gamma distribution based on the Gaussian approximation and the* 99.9% *quantile from this distribution, the green lines the corresponding using the mixed approximation. The dashed colored lines indicate the distributions based on the respective approximations without the matching correction. The plots are done for a network with exponential degree distribution with* $\lambda = 1$ *and* $N = 100$ *with* 10000 *fluctuation realizations.*

approximated analytically. For some edges the fitted distribution describes the data very good and collapses to almost one line together with the generalized gamma distribution that is calculated from the moments of the data, see Figure 5.13a and Figure 5.13b. For

edges with a high excess kurtosis of the directed flow distribution, the approximation based on the Gaussian approximation shows larger deviations from the data. The mixed approximation is in good agreement with the data for the distribution and the quantile, as can be seen in Figure 5.13c and Figure 5.13d. The dashed lines indicate the respective approximations without the matching correction, showing the approach of correcting the mean and variance of the in/out flow of each vertex significantly improves the results.

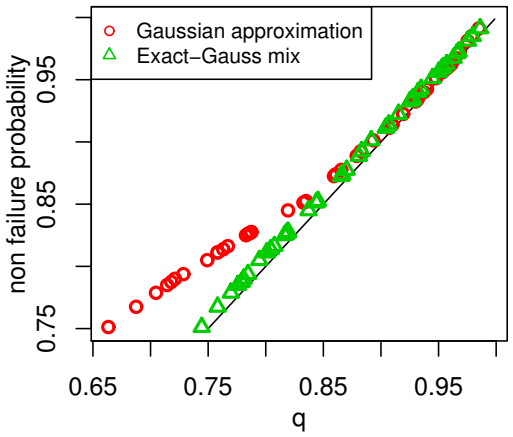


(a) *Histogram of the deviations, $\kappa = 1$, $\beta_L = \beta_S = 0$ for a network with exponential degree distributions with $\lambda = 1$*

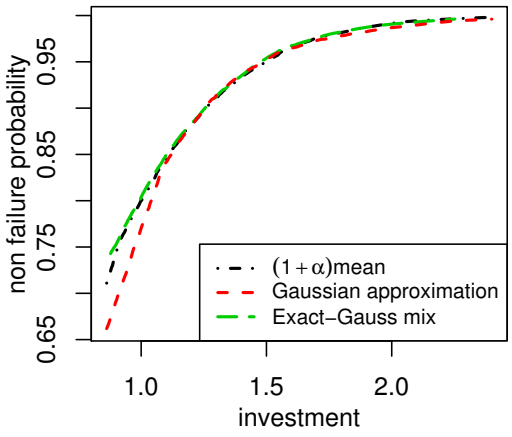
(b) *Histogram of the deviations, $\kappa = 1$, $\beta_L = \beta_S = 0$ for a network with scale degree distributions with $\alpha = 2.3$*

Figure 5.14.: *Deviations of the 99.9% quantile calculated from the data and from the generalized gamma distribution with parameters fitted to the moments obtained from the data (black), the Gaussian approximation (red) and the mixed approximation (green). The plots are done for a networks with $N = 1000$.*

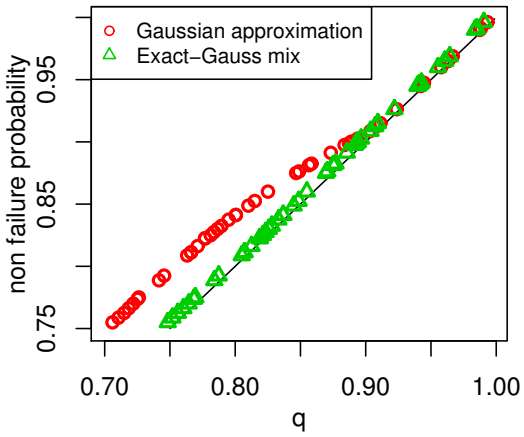
By inspecting all edges, we find that the generalized gamma distribution based on the moments of the Exact-Gaussian mix approximation is in very good agreement with the generalized gamma distribution based on the moments of the data. This is also supported by Figures 5.14a and 5.14b. The black curve illustrates the deviations between the simulation data quantiles and the quantiles calculated from the generalized gamma distribution directly fitted to the data. The deviations using the Exact-Gaussian mix approximation almost collapses with the deviations when the fitting is done directly to the similation results. To show that this does not only hold for networks with exponential degree distribution, results for a network with scalefree degree distribution are also shown in Figure 5.14.



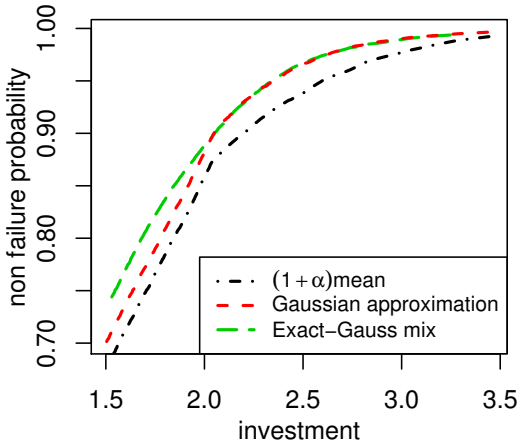
(a) *Quantiles versus minimal non-failure probabilities,* $\kappa = 1$, $\beta_L = \beta_S = 0$



(b) *Investment versus minimal non-failure probabilities,* $\kappa = 1$, $\beta_L = \beta_S = 0$



(c) *Quantiles versus minimal non-failure probabilities,* $\kappa = 2$, $\beta_L = \beta_S = 1$



(d) *Investment versus minimal non-failure probabilities,* $\kappa = 2$, $\beta_L = \beta_S = 1$

Figure 5.15.: *Non-failure probabilities for capacity layouts based on the Gaussian and the Exact-Gaussian mix approximation and the* $(1-\alpha)\cdot$*mean layout using the global factor matching policy on a network with exponential degree distributions with* $\lambda = 1$.

5.4.4. A Capacity Layout Based on the Flow Distributions

Based on the approximated flow distributions, we are able to define capacity layouts for each link. These layouts are compared to a very simple layout [85]

$$c_i = (1+\alpha)\, m_i \,, \tag{5.127}$$

where m_i is the mean of the absolute flow of link i, referred as the "$(1-\alpha)\cdot$mean" layout in the following. For the capacity layouts based on the Gaussian approximation

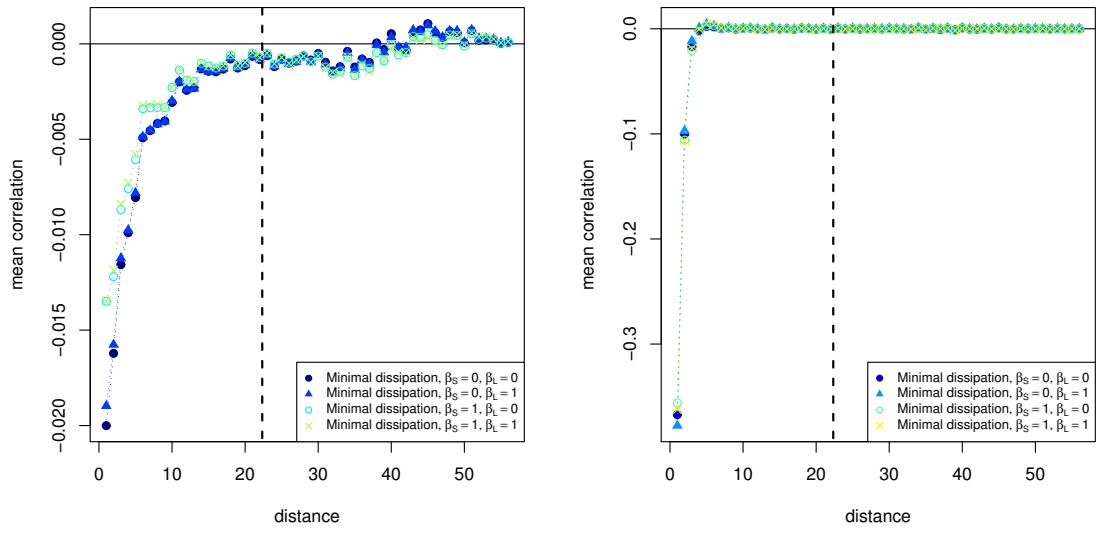
and the mixed approximation, the capacity is chosen such, that the probability that an edge is not failing is q. In Figure 5.15a and Figure 5.15c, the quantiles are plotted against the minimal non-failure probabilities of all links. We see that the Exact-Gaussian mix approximation is almost on the $y = x$ line, while the Gaussian approximation overestimates the non-failure probability for q smaller than ≈ 0.85. If we consider the investment costs, defined as the sum over all capacities, the difference between the layouts in terms of the investment cost for high q is small. But it has to be kept in mind that for the $(1 - \alpha) \cdot$ mean layout, the mean values have to be calculated and the non-failure probability is unknown.

5.5. Analytical Approach for the Correlated Case

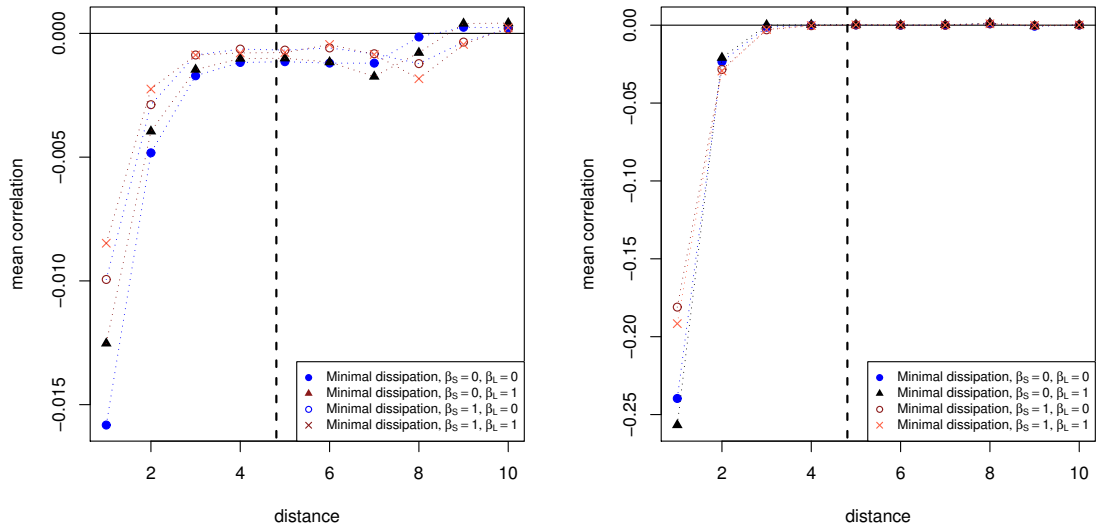
As discussed in Section 5.4, the correlations in the matched in- and out-flows s_j^b are important for the applicability of theorems from probability theory. Figure 5.16 shows the correlations of s_j^b for pairs of vertices with a given distance, generated using the minimum dissipation policy. In all cases the correlations between vertices that are only few hops apart, is significantly higher than for pairs that have a distance larger than the average distance of the network. This indicates, that generation is assigned to the closest loads in terms of the hop distance, when using the minimum dissipation policy. The order of magnitude of the maximal correlations for $\kappa = 1$ is comparable to the correlations found in Section 5.4 for the global factor policy. This is in agreement with the similarities between the flows when applying the two policies for $\kappa = 1$, that are described in Section 5.3. For $\kappa = 2$, the short range correlations amount to almost -0.4 and have to be considered.

We have not found an analytical description of the distribution of the s_j^b for the case of the minimum dissipation policy. Furthermore, the approach from Section 5.4 can not be applied if the correlations s_j^b are significant. On the other hand, we have seen in Section 5.3 that for $\kappa = 1$ the qualitative behavior of the two discussed policies is similar. Therefore, the results of Section 5.4 are applied to data, which is calculated using the minimum dissipation policy. As done in Section 5.4.3, the minimal non-failure probability of the absolute flow are compared with the quantiles derived from the approximations and the resulting investments illustrated. This is shown in Figure 5.17. For $\kappa = 1$ in Figure 5.17a, the non-failure probability is slightly underestimated for $q \gtrsim 0.85$, but in good agreement with the data for both approximations. The case $\kappa = 2$ shown in Figure 5.17c, reveals that the high correlations introduce qualitative different flows in the network and the predicted non-failure probabilities of the capacity layout from Section 5.4 are not adequate.

In terms of the needed investment, the case of $\kappa = 1$ behaves like the global policy in Section 5.4, see Figure 5.17b, whereas for $\kappa = 2$ in Figure 5.17d the approximations perform much worse than the $(1 - \alpha) \cdot$ mean capacity layout. Thus, for $\kappa = 1$, the Gaussian and the Exact-Gaussian mix approximation can be used to obtain an estimate



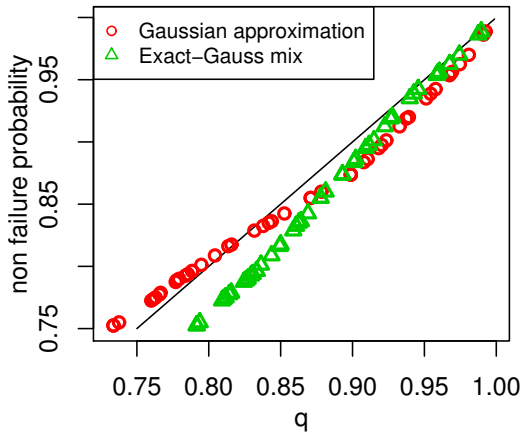
(a) $\kappa = 1$, *exponential degree distribution* **(b)** $\kappa = 2$, *exponential degree distribution*



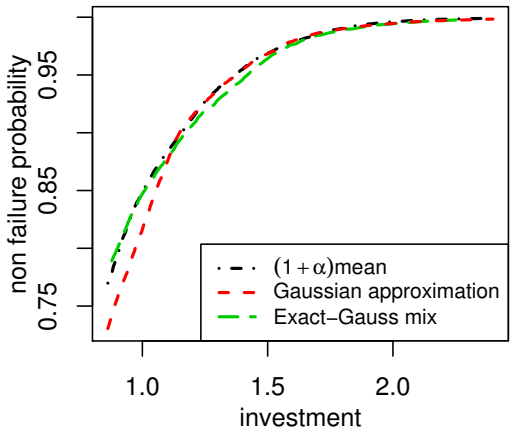
(c) $\kappa = 1$, *scalefree degree distribution* **(d)** $\kappa = 1$, *scalefree degree distribution*

Figure 5.16.: *Average correlations of the matched in- and out-flows s_j of all pairs of vertices at a given distance introduced by the minimum dissipation matching policy. The data for a graph with exponential degree distribution with $\lambda = 1.8$ and for a scalefree network with $\alpha = 2.3$ are shown, both with $N = 1000$ and $p_{cc} = 0$. The dashed line indicates the mean distance of the graph.*

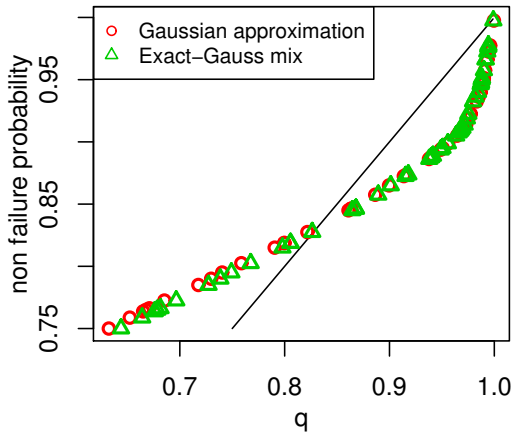
of the capacities needed at each link when using the minimum dissipation policy.



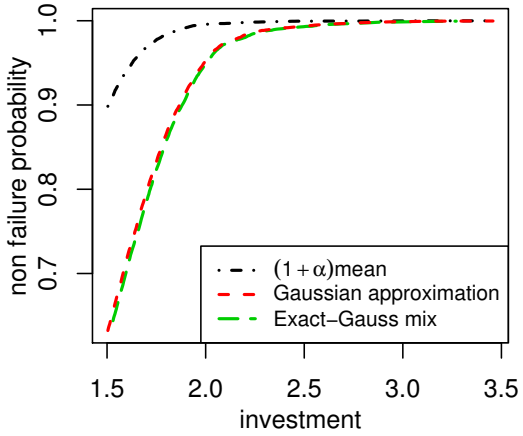
(a) *Quantiles versus minimal non-failure probabilities,* $\kappa = 1$, $\beta_L = \beta_S = 0$



(b) *Investment versus minimal non-failure probabilities,* $\kappa = 1$, $\beta_L = \beta_S = 0$



(c) *Quantiles versus minimal non-failure probabilities,* $\kappa = 1$, $\beta_L = \beta_S = 1$



(d) *Investment versus minimal non-failure probabilities,* $\kappa = 1$, $\beta_L = \beta_S = 0$

Figure 5.17.: *Non-failure probabilities for capacity layouts based on the Gaussian and the Exact-Gaussian mix approximation and the* $(1-\alpha) \cdot$ *mean layout using the minimum dissipation matching policy for a network with exponential degree distributions with* $\lambda = 1$.

5.6. Summary

We have introduce a model that allows to analyze fluctuating generation and various distributions of generators and consumers in the transmission network. It is based on stylized facts that were found in the data. Two policies to match generation and load were proposed. Using the Monte Carlo method, the effect of fluctuations on the flow in the network was studied. The distributions of the absolute flows of the links was found to be close to a generalized gamma distribution.

The policies influence significantly the flows in the network. The minimum dissipation policy was found to assign load and generation such that they are close together thus introducing short range correlations. The global factor policy introduces only small correlations and analytical approximations to the flows of each link could be derived. These derivations include a description of the global factor policy and the flows in the network. Based on the approximated flow distributions, a capacity layout was defined. The non-failure probability can be predicted in very good agreement with the simulation results. Although the approximations do not improve the security for a given investment, compared to a very simple capacity layout, they allow to tune the investment such that a given non-failure probability can be achieved.

For the case that the generation is on average equal to the load, $\kappa = 1$, the effect of the two policies are very similar. This allows to apply the results derived for the global factor policy to the minimal policy.

From the analytical considerations, we conclude that for uncorrelated in- and out-flows only the influence of the in- and out-flows of a few vertices have to be considered exactly. The in- and out-flows of the remaining vertices can be summarized by a Gaussian distribution. Which vertices are important can be quantified in terms of the pseudo inverse of the admittance matrix and the variances of the vertices.

6. Transport Flows Based on the Weather Data

In Chapter 4, it was found that European cooperation, in the sense that power is exchanged, significantly reduces the required balancing and storage energies. Further, we analyzed a simplified model of power systems with fluctuating generation in Chapter 5 and analytically derived approximations to the power flow distributions of the links. In this chapter, the power flows based on the data introduced in Chapter 3 are discussed. The two objectives are to quantify the transport flows resulting from weather data driven fluctuations of the power generation and to compare the model based results to real world data.

In Section 6.1, the transmission network and the scenario that is analyzed is introduced. The resulting flows are discussed in Section 6.2. Finally, the approximations derived in Chapter 5 are tested and discussed in Section 6.3.

6.1. Transmission Network and Methods

The transport flows are evaluated on a simplified transmission network. This is due to the fact that the network data of the European power grids are not publicly available. Further, the consumption data is only available on the scale of regions, so that a fine grained transport analysis would involve an estimation of the consumption on a finer grid, see also Section 3.2.3. Thus, the approach chosen here is to estimate the transport flows between the regions. The corresponding network is derived based on the partitioning of the regions and their spatial arrangement. We basically connect neighboring regions as shown in Figure 6.1, yielding a network, that is reminiscent to a geometric network, see Section 2.1.2.

Various scenarios were analyzed in Chapter 3 with respect to the required balancing and storage energies and power. The case using only balancing generators resembles the assumptions that were used in Chapter 5. Analyzing transport flows incorporating storage would give rise to the additional questions how hypothetical future storage capacities will be set up and what their spatial distribution will be. When using the transitional scenario, as discussed in Section 3.5, a similar problem occurs: A fraction of conventional generation has to be assigned to every region. The spatial distributions of storage and conventional generation has a big influence on the resulting transport flows and requires a detailed evaluation that is is beyond the scope of this work. For these

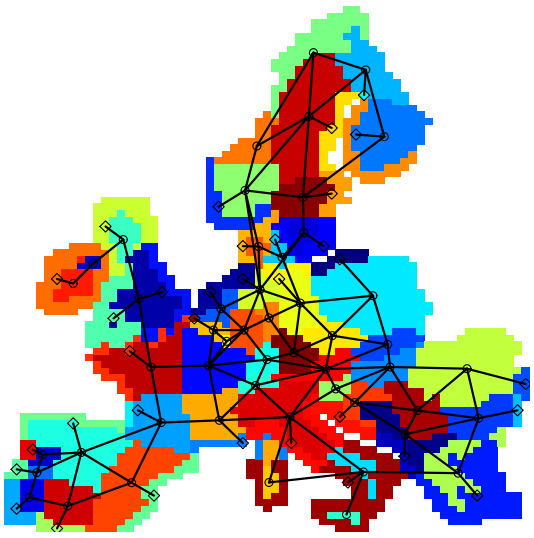


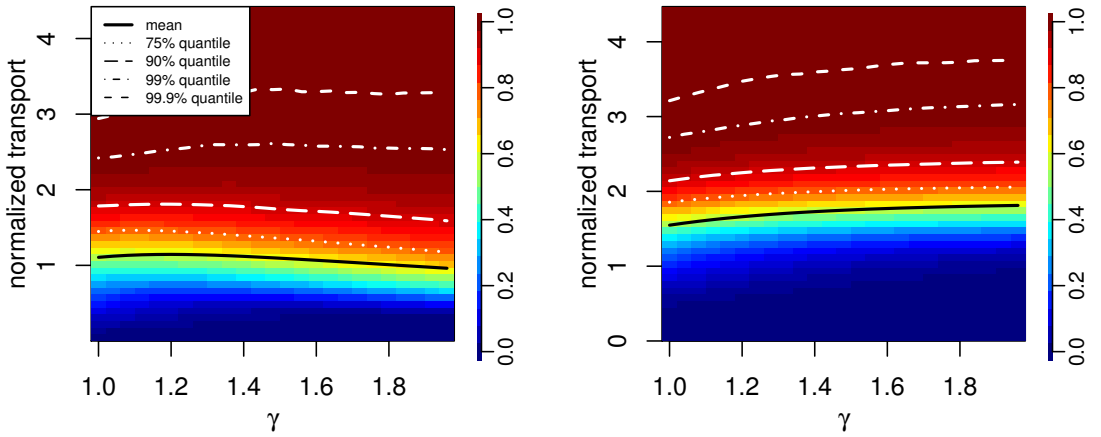
Figure 6.1.: *Transport network for the inter-region flows. Different regions are indicated by different colors, the links are drawn between the center of masses of the regions.*

reasons, we focus on the hundred percent and the over installation scenario using only local balancing generation to compensate shortages, see Sections 3.4 and 3.6. Sufficient balancing generation capacities are assumed for every region. The optimal mix is chosen according to the minimal balancing energy, $a \simeq 0.8$, as shown in Figure 3.21.

The flow along the links is calculated using the DC power flow approximation as described in Section 5.1. For simplicity uniform suceptances of one are assigned to all links. The matching is done using the minimum dissipation and the global factor policies as described in Section 5.1.1.

6.2. Aggregated Flows, Link Flows and Matching Policies

A first quantity studied, is the aggregated absolute flow in the whole network. It is calculated by summing up the absolute flows $|f_{ij}|$ of all links, for each hour separately. The case of over installation with one hour averaging is shown in Figure 6.2a and 6.2b, for $c = 0$ and $a = 0.8$. The values for $\gamma = 1$ are equal to the hundred percent scenario discussed in Section 3.4. For that scenario, the average transport using the minimum dissipation policy is in the order of the average European consumption. This is equal to the case, that on average the whole power consumed has to be transported from one region to a neighboring one. The maximal aggregated inter-region flow is around three times the average load, meaning that for these cases the whole generated power has to be transported over a large part of Europe, see Figure 6.1. The average transport using the minimum dissipation is about one third of the average hourly load smaller than the



(a) *minimum dissipation,* $a = 0.8$ *and* $b = 0.2$*, one hour averages*

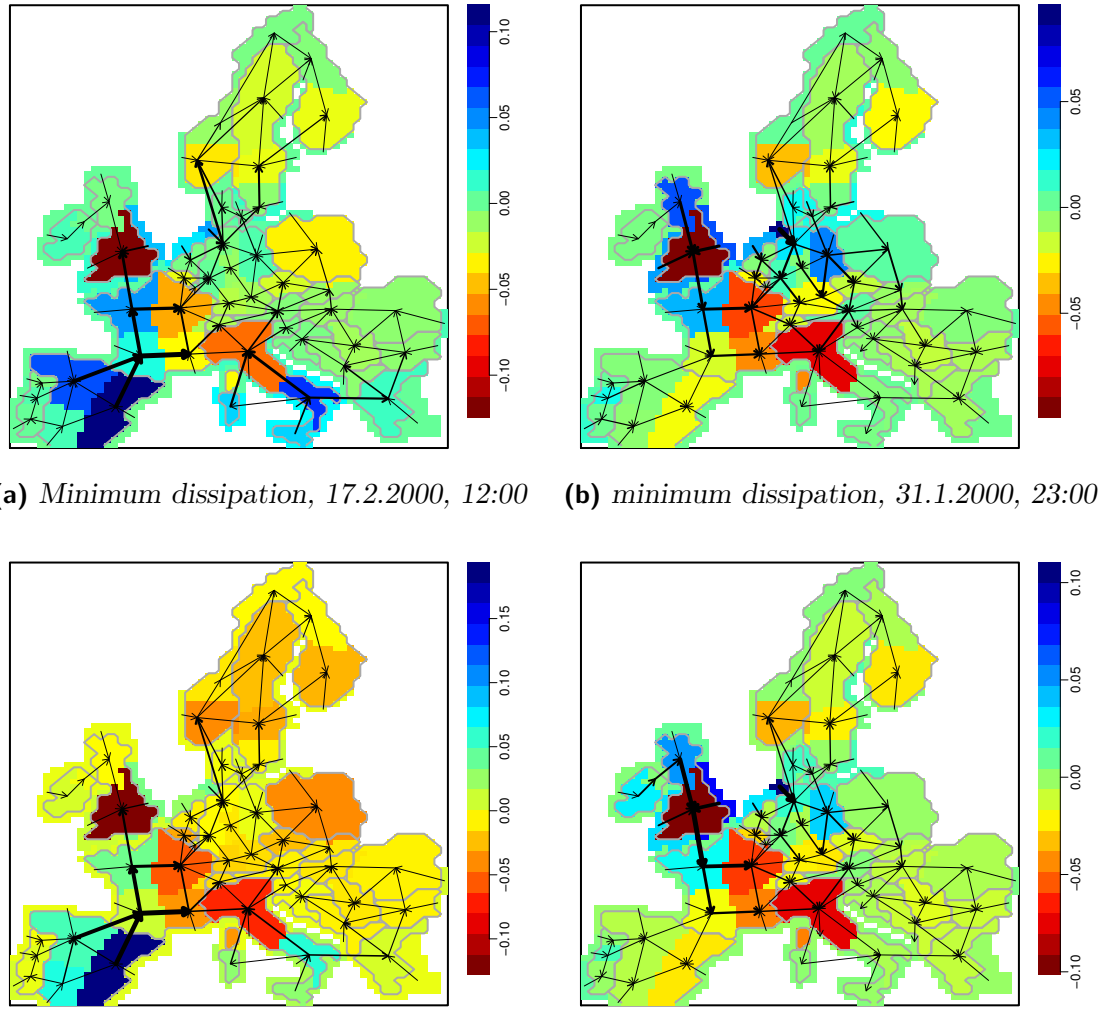
(b) *Global factor,* $a = 0.8$ *and* $b = 0.2$*, one hour averages*

Figure 6.2.: *The quantiles of the aggregated inter-region transport for overproduction* $\gamma \geq 1$ *considering one hour averages normalized to the mean load per hour over Europe. The minimum dissipation and the global factor policy are shown. The inter-region transport is the aggregated flow of all links, with the summation done for each hour separately. The colors indicate the quantiles, meaning that e.g. for the value of* 0.9*, all but* 10% *of the aggregated flows are lower than the corresponding aggregated flow. The parameters used are* $a = 0.8$ *and* $c = 0$*.*

inter-region transport when using the global factor policy.

For increasing overproduction, the aggregated inter-region transport exhibits a different behaviour for the two policies. For the minimum dissipation policy, the average aggregated flow increases slightly first, before it decreases to a value a bit lower than for $\gamma = 1$. This is a disagreement with the findings for the model in Figure 5.1, where the flows are significantly lower for $\gamma = 2$ compared to the case of $\gamma = 1$ when using the minimum dissipation policy. When employing the global factor policy the aggregated inter-region flows increase for higher overproduction, a result that is close to the findings for the model (see Figure 5.1). For both matching policies and $\gamma = 1$, a large fraction of the probability mass is close to the mean. The difference of the mean value to the 90% quantile is around two thirds of the average European load, whereas the difference from the mean to the 99.9% quantile amounts to almost twice the average load. For an overproduction factor of $\gamma > 1$, the spread increases for the minimum dissipation policy while it stays in good approximation constant for the global factor policy indicating that for rare events large transport flows can not be avoided.

In Figure 6.3, the flows are shown for two points in time. For both policies the maximal inter-region flow is illustrated that occurs at noon and at eleven pm. Furthermore, the matched in and out flows normalized to the average European hourly load for each region are indicated by the color of the respective region. These flows capture characteristic



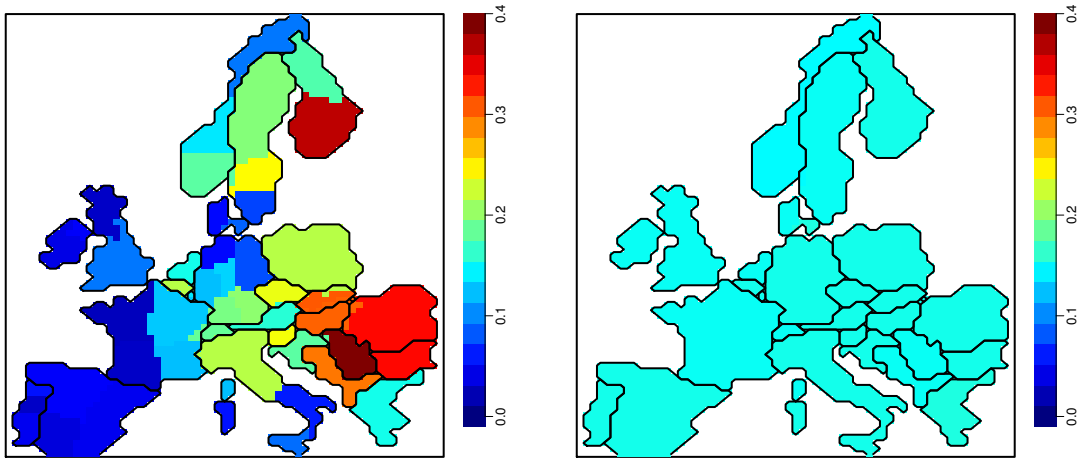
(a) *Minimum dissipation, 17.2.2000, 12:00* **(b)** *minimum dissipation, 31.1.2000, 23:00*

(c) *Global factor, 17.2.2000, 12:00* **(d)** *Global factor, 31.1.2000, 23:00*

Figure 6.3.: *Matched in and out flow for each region and the flows on the links for the maximal aggregated inter-region flow at noon and midnight. The colors indicate the matched in and out flow s_j normalized to the average hourly load for each region. The thickness of the arrows indicate the magnitude of the flow. The data for the minimum dissipation and global factor matching policies are shown. The parameters are $a = 0.8$, $c = 0$ and $\gamma = 1$, see Equation (3.1).*

features found for all time steps. The links from Spain to France transport relatively large amounts of power but almost always from Spain to the North and mostly during daytime. Another link that has to carry large amounts of energy is the connection from France to Great Britain but flows in both directions were found, see Figure 6.3. Besides that, there is a general tendency that, during daytime, the transport takes place from South to North while at night the power flows from North to South. Also, from offshore regions more power flows onshore during night compared to the situation at

noon, see e.g. the German, Portuguese or Swedish offshore regions in Figure 6.3. This is presumably due to the solar generation during daytime. The largest consumption can be found in England around London, Eastern France and the North of Italy. North Germany on the other hand exports power almost all of the time.



(a) *minimum dissipation* **(b)** *Global factor*

Figure 6.4.: *Average reserve power needed relative to the average load for every region. Results for the minimum dissipation and global factor matching policy are shown. For the latter the values are almost the same for all regions with deviations around two percent. The parameters used are* $a = 0.8$, $\gamma = 1$, *and* $c = 0$.

The installed capacities planed according to the national targets for 2020 were discussed in Section 3.2.2 and Figure 3.7. We see that the wind and photovoltaic generation capacities are planed to be installed mainly in the North, South, and West. For Eastern Europe only small capacities for renewable generation are planed. This spatial segregation of the generation capacities can be expected to strongly influence the spatial distribution of the mismatch and the required matching capacities. The difference between the minimum dissipation and global factor matching policy seems to be minor, when considering Figure 6.3. But the two policies cause a different spatial allocation of the employment of balancing power. In Figure 6.4 the average balancing power for each region normalized to its average load is shown. The minimum dissipation policy causes a strong spatial separation, almost no balancing is used in Western Europe, Northern Germany and Denmark, while Eastern Europe and the North have on average a high need for balancing power (see Figure 6.4a). In contrast, the global factor policy causes a uniform average reserve power employment and thus a higher flow.

The spatial distribution of the flows in Figure 6.3 for different hours of the day can be an explanation of the discrepancy between the aggregated flows in the model and in Figure 6.2a when using the minimum dissipation policy. The characteristic behaviour of solar and wind power generation over a day, together with the spatial separation of these power sources, poses additional constraints on the matching policy. During night

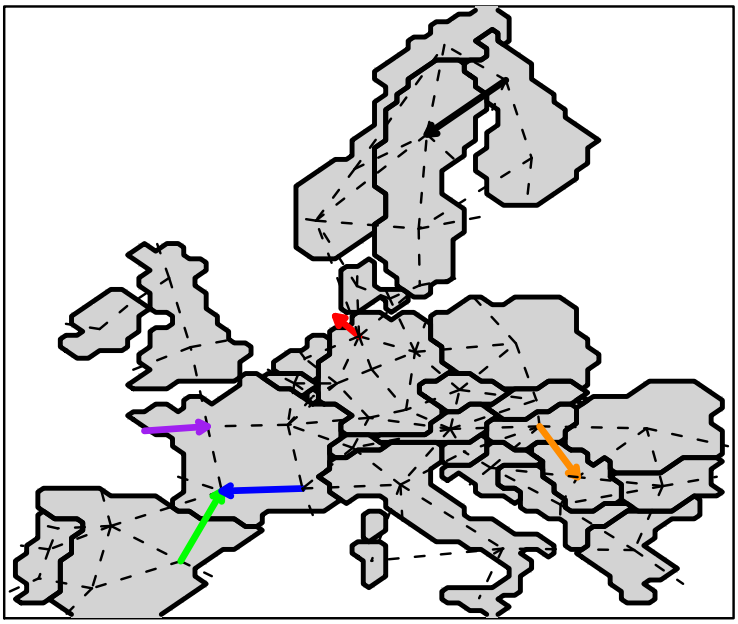
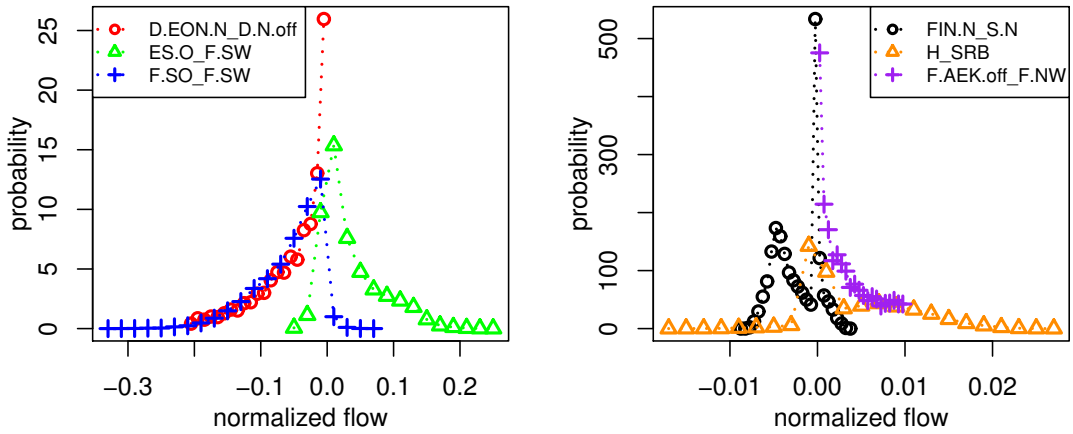


Figure 6.5.: *Selected edges, the colors correspond to the colors in the histograms in Figure 6.6 and 6.7.*

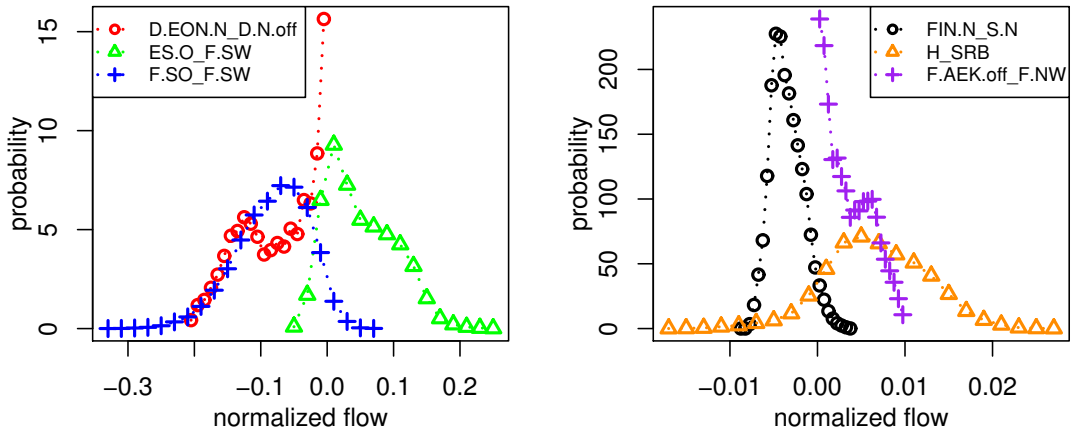
no solar power is generated, so that only the power from wind energy, that is mainly located in the North and North East, can be allocated to consumers. In summer, when the wind generation is lower, see Figure 3.10, during daylight mainly the solar generation is available. Therefore even for high overproduction, the aggregated transport decreases only little when using the minimum dissipation policy for the data.

To further study the flow, the probability distributions of the flows along the links are discussed. The network of Europe considered here consists of of 129 links but only the histograms of some typical links, indicated in Figure 6.5, can be shown. Links with the highest average flow are shown in Figure 6.6a and 6.6c for the minimum dissipation and the global factor policy, respectively. We see, that the magnitudes of the flows in the network can be very high for some links. The links from offshore regions have to transport all power generated there and thus the maximal flow depends on the installed offshore capacities. The links from Northern Spain to South-West France and from there to South-East France carry in the extreme cases a large fraction of the power generated on the Iberian Peninsular and South West France to Northern Europe. This amounts to a peak load of around a third of the European hourly consumption that corresponds to around 110GW as the average aggregated hourly consumption of Europe from our data is 342 GW. It is unlikely that these capacities can be build up. In the opposite direction the flow is comparably low, although the flow towards the Iberian Peninsular is high compared to the other links, as shown in Figure 6.6b and 6.6d.



(a) *minimum dissipation policy, high average flows*

(b) *minimum dissipation policy, low average flows*



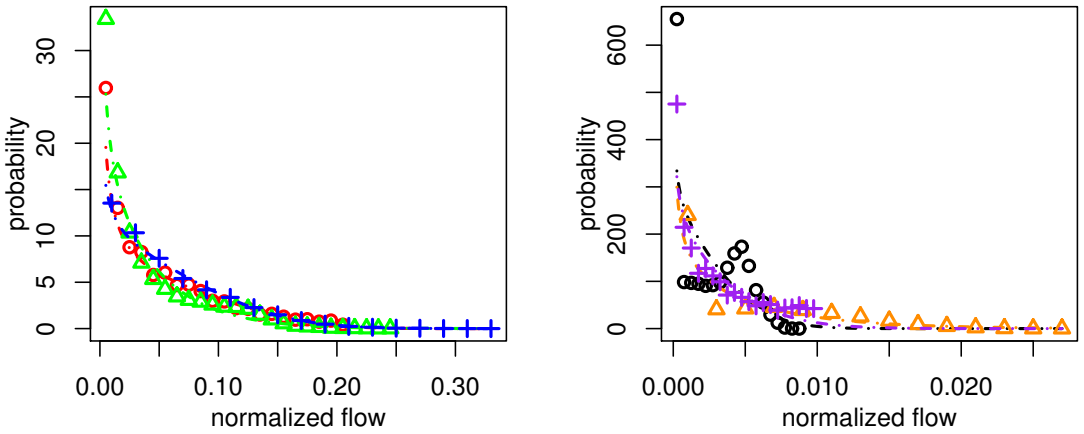
(c) *Global factor policy, high average flows*

(d) *Global factor policy, low average flows*

Figure 6.6.: *Histograms of the flow over selected edges. The inter-region power flows are calculated using the global factor and the minimum dissipation policies and normalized to the average European consumption. The spatial positions of the links are shown in Figure 6.5. The parameters used are* $a = 0.8$, $\gamma = 1$, *and* $c = 0$.

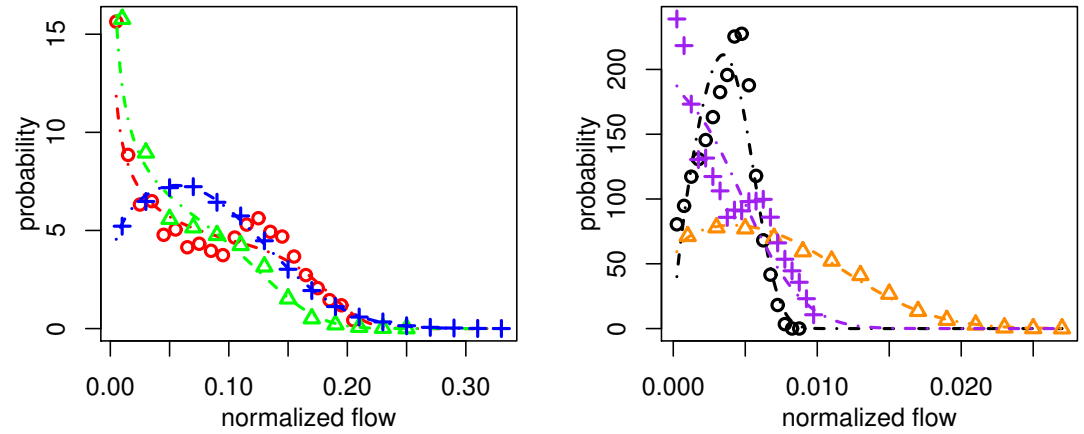
The distributions are quite diverse, some have only one distinct peak but e.g. the link from North of Finland to North of Sweden is almost bimodal when using the minimum dissipation policy. Only few vertices have a distribution close to a normal distribution, which is surprising as in real systems many random variables converge to a normal distribution [54]. The shapes of the distributions belonging to the different matching policies are not quantitatively different, but considering the individual link the distributions are significantly different when comparing the flows based on the two matching policies.

An important quantity is the undirected flow, given by the absolute value of the flow.



(a) *minimum dissipation policy, high average flows*

(b) *minimum dissipation policy, low average flows*

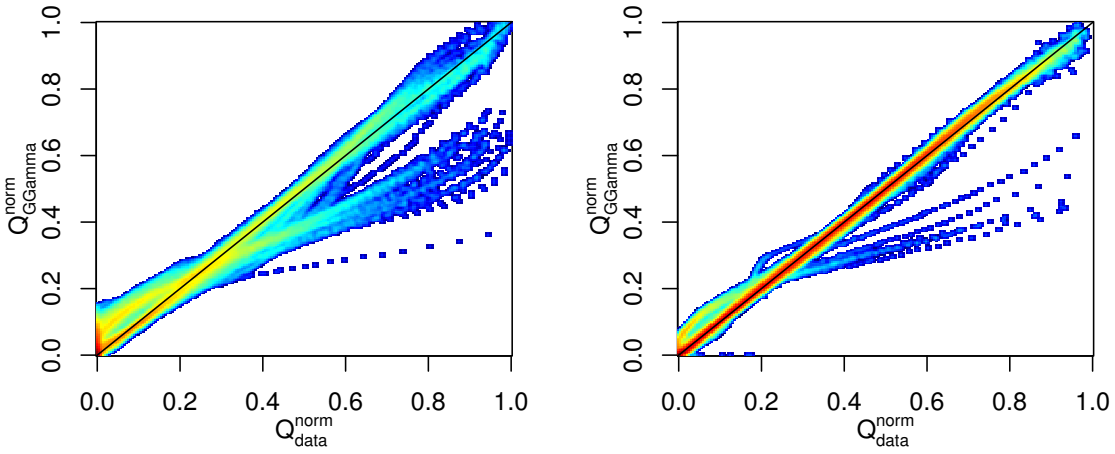


(c) *Global factor policy, high average flows*

(d) *Global factor policy, low average flows*

Figure 6.7.: *Histograms of the absolute flows over selected edges. The inter-region power flows are calculated using the global factor and the minimum dissipation policies and normalized to the average European consumption. The spatial positions of the links are shown in Figure 6.5. The dot-dashed lines indicate the fitted generalized gamma distributio, see text. The parameters used are* $a = 0.8$, $\gamma = 1$, *and* $c = 0$.

These are shown in Figure 6.7 for the same links as in Figure 6.6. The absolute flows illustrate even more the high power flows on some links. The qualitative form of the distributions resemble the distributions found in Chapter 5, shown in Figures 5.6 and 5.7. Therefore, the generalized gamma distribution is fitted to the moments calculated from the data, indicated by the dot-dashed lines in Figure 6.7. The resemblance when using the minimum dissipation policy is worse than for the case of the global factor policy. As in Section 5.3, the agreement of the generalized gamma distribution is tested using a Q-Q plot, where the quantiles of the data are compared to the quantiles of



(a) *Minimum dissipation policy* **(b)** *Global factor policy*

Figure 6.8.: *QQ-plots of the quantiles of the fitted generalized gamma distribution to the qunatiles of the data. As in Figure 5.8, the quantiles are normalized to the maximal values of each link and shown as a density for all links in the network. The parameters used are* $a = 0.8$, $\gamma = 1$, *and* $c = 0$.

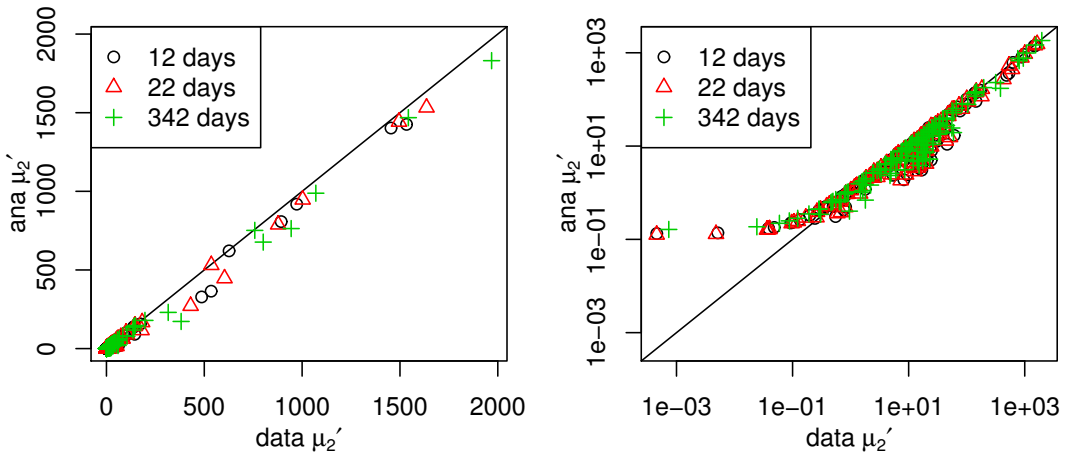
the fitted generalized gamma distribution. The normalized quantiles for all links are plotted as a density for both matching policies in Figure 6.8. Besides few links that show large deviations, the quality of the approximation for the global factor matching policy in Figure 6.8a is surprisingly good. When using the minimum dissipation policy in Figure 6.8a, the deviations are significantly larger, which can be attributed to the correlations introduced by the minimum dissipation policy, as discussed in Section 5.5.

6.3. Analytical Results and Data

Looking at the results from Section 5.4 and 5.5, the questions arises how to describe the flow distributions along the links of a power system considering real data, therefore the derived approximations are tested for how good the inter-region transport can be approximated.

As the matching in Section 5.4.2 is derived for uniformly distributed fluctuations only, the matched in and out flows from the data are inserted directly into Equations (5.59), (5.76), and (5.77). To avoid systematic effects due to day/night changes and due to temporal correlations, only the data at noon and the hour before and after are considered.

The second raw moments from the data in Chapter 6 are plotted against the analytical second raw moments in Figure 6.9 for tranport using the global factor policy, $a = 0.8$, and $\gamma = 1$. The analytical moments are calculated using Equation (5.59). Almost all



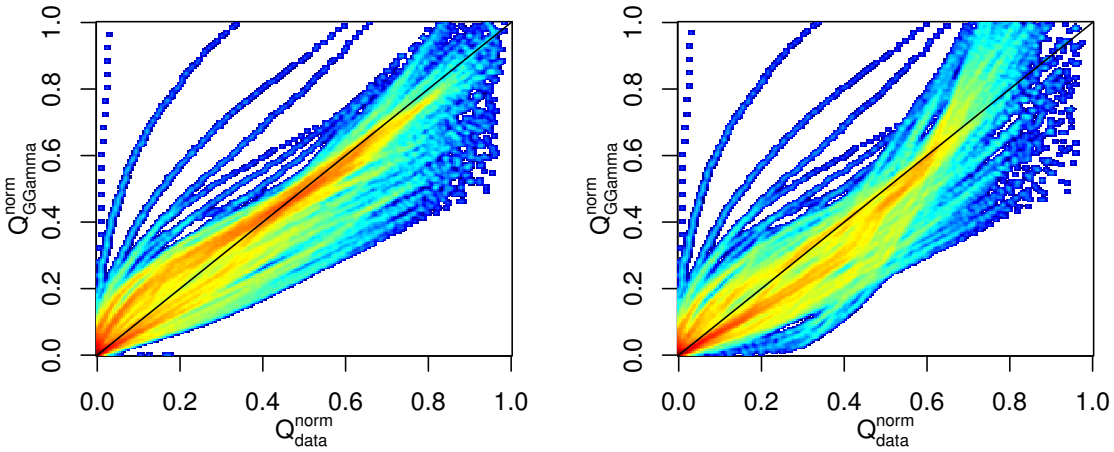
(a) *minimum dissipation policy* **(b)** *minimum dissipation policy*

Figure 6.9.: *Second raw moments from the data from Chapter 6 compared to the analytical second raw moments in a (a) semi-logarithmic and (b) double-logarithmic plot. The transport is calculated using the global factor policy and $a = 0.8$, $\gamma = 1$, and $c = 0$. Details see text. The different colors and and symbols indicate a pooling of data over different time spans centered on the first of Juli.*

points scatter around the $y = x$ line, indicating that the approach yields results that are close to the results calculated from the data. There are some systematic deviations for the small second raw moments and for values around 10. As no approximation enters Equation (5.59) besides the assumption that the data is uncorrelated, the deviations can be expected due to spatial correlations in the in- and out-flows of nearby regions. For most links, the data is in good agreement with the analytic predictions, so that the correlations are assumed to be small enough. To test for a seasonal dependence in the approximations, the calculations are done for different periods of time around the first of Juli. The results for 12, 22, and 342 days are very similar, so that we conclude that the seasonal influence does not affect the approximations.

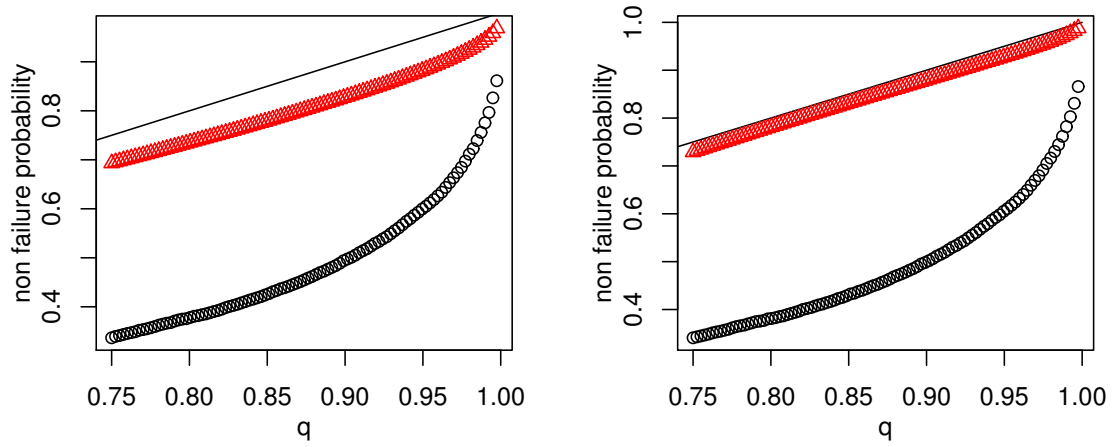
The analytical approximations to the first and third moments, derived in Sections 5.4.1.2 and 5.4.1.3, were used to estimate the parameters of the generalized gamma distribution. As before the agreement between the transport flow distributions of all links and the approximated generalized gamma distribution are compared in a Q-Q plot shown in Figure 6.10. Large deviations are found between analytical approximations and the data for some links. The gaussian approximation performs better than the Exact-Gaussian mix approximation, which is not surprising, as the latter is specifically tuned for in- and out-flows that are uniformly distributed.

Based on the approximated distributions a capacity layout can be defined. For a given non failure probability q, the capacities of the links are chosen such that the flow in the network is smaller than the capacity with probability q, see also Section 2.3.2. The capacity layouts are tested in Figure 6.11. The black circles show that the approximation



(a) *Gaussian approximation* **(b)** *Exact-Gaussian mix approximation*

Figure 6.10.: *Q-Q plots of the quantiles of the generalized gamma distribution derived using the Gaussian approximation and the Exact-Gaussian mix approximation, see Section 5.4.1. As in Figure 5.8, the quantiles are normalized to the maximal values of each link and shown as a density for all links in the network. The parameters used are $a = 0.8$, $\gamma = 1$, and $c = 0$.*



(a) *Gaussian approximation* **(b)** *Exact-Gaussian mix approximation*

Figure 6.11.: *Non failure probabilities for capacity layouts based on the Gaussian and the Exact-Gaussian mix approximation using the global factor matching policy with $a = 0$, $\gamma = 1$, and $c = 0$. The black circles indicate the failure probability of the worst link, the red triangles the average failure probability.*

for the link, whose non failure probability is overstimated the most, is far from the desired q. On average, the approximation performs pretty well, indicated by the red triangles. The Exact-Gaussian mix approximation overestimates the required capacities, so that it seems to perform better on average but as can be seen when looking at the

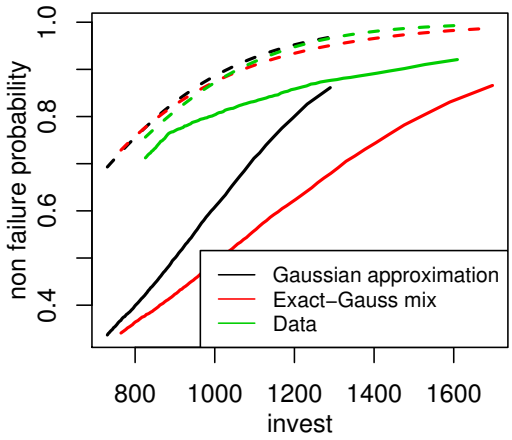


Figure 6.12.: *Investment costs for capacity layouts based on the generalized gamma distribution, whose parameters are derived using the Gaussian approximation, the Exact-Gauss mix approximation and directly from the data, against the resulting non failure probabilities q. The solid lines indicate the worst case link the dashed lines the average non failure probability. The investment costs, given in arbitrary units, are assumed to be equal for all links and calculated as the sum over all capacities.*

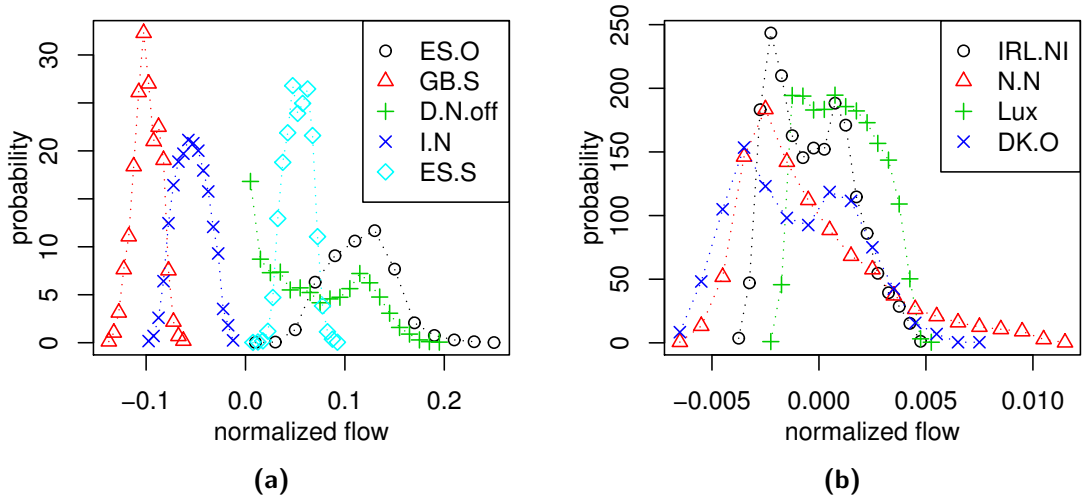


Figure 6.13.: *Selected in and out flows for selected regions with (a) high and (b) low mean values of the in and ouflows. The parameters used are $a = 0.8$, $\gamma = 1$, and $c = 0$.*

investment costs, this is due to the fact that most of the capacities are chosen too high. The investment costs are assumed to be equal for all links and evaluated by summing up all capacities. These are shown in Figure 6.12 compared to the non failure probability q. When comparing the investment costs to the non-failure probabilities, we see that only the capacity layout using the generalized gamma distribution derived directly from the data performs acceptably for the worst case link. However, on average the three tested layouts perform similar.

The fact that some links are pretty badly approximated with the derived approximations from the model in Chapter 5 is not surprising. In the model, the generation is approximated by an uniform distribution and the load assumed to be constant but the data time series is much more complex. In Figure 6.13, in- and out-flow distributions for select regions are shown. While the uniform distribution is a good approximation for e.g. Luxembourg, the distribution for Eastern Spain or Northern Norway deviate strongly. However, we have seen in Section 5.4.4, that a specific treatment of the most influencial vertices of a link can improve the estimation of the flow distributions. The good agreement of the second moment with the data suggest that the approach generally works. Also, the global factor matching policy can be most probably adjusted to the more complex requirements of the data.

6.4. Discussion

Highly intermittend renewable energy generation causes large power flows over Europe. Here, only a simplified transport network was analyzed to estimate the power flows between the regions. The findings, using this approach, indicate that over installation of generation capacities, in the sense that the national targets for 2020 are scaled up, does not decrease the needed transport capacities. From the point of view of a minimization of the transport capacities, a more homogeneous distribution of the generation capacities could be desireable as generators located close to the consumers trivially elicit less transport and the day/night effect in the transport could be weakened.

The distribution of the flows show that there are links, that exhibit very high power flows. For links connecting offshore regions, this is not surprising, because flows up to the installed offshore generation capacities can arise. The link connecting Spain to France or other links in Southern France need to transport up to 100 GW in extreme cases, a value that is extremely high and unlikely to be installed. Regional storages are expected to cut extreme flows and allow for a higher utilization of the links.

When testing the analytically derived approximations of the flows in the model with the data, we find that the proposed capacity layouts perform on average acceptably good. However, the approximation of the flow distributions of some links are far away from the flows in the data. A more detailed analysis of the in and out flows of the regions is expected to improve these results.

7. Summary and Discussion

Due to limited resources and waste problems, generation from renewable energy sources will play a crucial role in the future supply of energy. However, fluctuations in the generation pose new challenges when implementing a reliable power system. This work consists of an evaluation of the required energy and power to balance shortages in power systems with a high share of renewable energy generation. For that purpose, the storage of excess generation or the usage of balancing generators is assumed. An analysis of power flows in networks is also presented, for both a generic model and for a data time series derived from weather data. Furthermore, the effects of fluctuating sources on cascading failures, when using a transport paradigm that is widely used in the literature of complex networks, has been studied analytically and numerically.

For planning an efficient transition of todays to future power systems, it is important to understand the properties of these future systems. This is also crucial when defining a model to evaluate power flows within a controlable framework. Therefore, we analyzed a data time series of wind and solar generation with a high temporal and spatial resolution, together with a load time series based on publicly available consumption data from utility companies.

Based on the data time series, we found that optimal mixes between the installed capacities of wind and solar generators exist, so that the storage or balancing energies can be minimized. This is due to distinct seasonal characteristics of the said energy sources and the consumption. Using an optimal linear combination of the different energy sources, the mismatch between generation and load can be minimized. This finding is analyzed for different scenarios. We show that an over installation of wind and solar generation capacities of around 50% of the average consumption brings the required storage or balancing energies down to values that will presumably be feasible. This is analyzed for the cases where either only storage or only balancing generation is employed. Two distinct optimal mixes were found, a ratio of 60% wind and 40% solar generation when minimizing the required storage energy, and 80% wind and 20% solar generation to minimize the balancing energy. These different ratios are caused by the fact that the storage and the balancing energy are sensitive to the dynamics on different time scales. For the balancing energy, fluctuations on the daily time scale are crucial, especially the day/night changes. The storage energy is determined by time scales in the order of weeks or months.

These optimal mixes can also be found for individual countries. Depending on the geographical locations, the seasonal characteristics of the wind and solar generation and the load change. In the South, the fluctuation strength of the solar generation is

much lower than for Northern countries. Also, the seasonal consumption differs. This leads to different optimal mixes for each country, with a significant correlation to the latitude of the respective center of mass. The optimal mixes for the storage energy show a strong dependence on the latitude, while the deviations in the optimal mixes for balancing energy are small. This again can be explained by the dependence of the storage and balancing energy on different timescales. Comparing the aggregated storage and balancing energies over the individual countries to these calculated for whole Europe, which are calculated with a uniform optimal mix of wind and solar generation for all regions and with the assumption of unlimited power exchange between the regions, we find that the storage and balancing energies for the latter case are significantly smaller. This illustrates the importance of power transmission.

Fluctuating generation results in power flows on the links with a probabilistic nature. No previous studies have systematically assessed their properties. Therefore, we introduce a model based on stylized facts derived from the data. This approach allows to quantify the resulting flows in a framework with a priori known properties. We account for spatial separation of generation and consumption by assigning the former to the periphery and the latter to the center. A homogeneous spatial distribution is also considered. The generation at each node is described by a random variable, drawn independently from a uniform distribution for each node. The load is assumed to be constant, and the ratio of the average generation to the load fixed by the parameter κ. The "DC power flow" approximation used by engineers, is an appropriate paradigm to describe power flows. To ensure the conservation of energy, two policies are introduced that match the in- and out-flows of the network, the "minimum dissipation" policy, that has the a priori known property to minimize the sum of the quadratic flows in the network, and the "global factor" policy, that assigns shortages and excess generation uniformly over all vertices relative to their load and generation. The policies were found to have a strong influence on the resulting flows, only for the case of $\kappa = 1$, the qualitative behavior of the flows based on both policies are equal. For all cases, the generalized gamma distribution was found to be in very good agreement with the distributions of the undirected flows on the links.

The correlations in the matched generation and load of different vertices, introduced by the "global factor" policy, can be neglected. This allows for an analytical approximation of the undirected flow distributions on each link in the network. First, the raw moments of the flows are derived. For the second raw moment, no approximations are needed. The exact equations for the odd raw moments cannot be numerically evaluated for a system with a realistic size, and so we derive two approximations. A simple approach, in which the undirected flows are approximated by Gaussian distributions, is presented. For a given link, these results can be significantly improved by combing the exact equations, describing the most influential vertices, with a Gaussian distribution that approximates the flows from the remaining vertices. The distributions of the matched generation and load for each vertex could be derived for the "global factor" policy. The factors, that scale the generation and load for each vertex, so that they match, can be expressed as

a ratio distribution. This distribution is approximated by a gamma distribution that allows to calculate the probability distributions of the matched load and generation for each generator and consumer in very good approximation. The estimated first moments only exhibit deviations of maximally two percent compared with the first moments of the data, for the third moments, maximally five percent are observed. Using these results, the parameters of the respective generalized gamma distribution are calculated. For almost all links, these distributions are in very good agreement with the flow distributions found in the simulations. We show, that the quality of the estimated flow distributions, based on the Exact-Gaussian mix approximation, is as good as possible, when the undirected flow distributions are assumed to be distributed according to a generalized gamma distribution.

The derived distributions of the undirected flows for each link allow to define capacities for each link, so that the flow on the link is, with a given probability, smaller than the respective capacity. Using the predicted capcities, the non-failure probability of each link was tested and we found that the proposed capacities almost exactly ensure the desired security of operation.

The "minimum dissipation" policy is shown to introduce short range correlations that are significant for $\kappa \neq 1$ but can be neglected for $\kappa = 1$. We could not derive an analytical approach similar to the one presented for the "global factor" policy, so we tested the existing analytical approximations with that policy. For the case of $\kappa = 1$, the results for the high quantiles are in very good agreement with the data and large deviations are found for $\kappa = 2$.

Transport flows based on the data time series were also calculated. We found that some links exhbit very large flows, e.g. the link connecting Spain to France has a peak load of around 100 GW. There is a general tendency that power flows during daytime go from South to the North while the opposite is found during night. This is attributed to the spatial separation of the installed wind and solar generation capacities. The sum over undirected transport flows is significantly higher when using the "global factor" matching policy compared to the case using the "minimal dissipation policy". In the over installation regime, the flows decrease only marginally. This is a difference to the model and is expected to be also connected with the spatial distribution of the generation capacities.

The power flow distributions of the undirected flows can be described in good agreement for most of the links with a generalized gamma distribution when using the "global factor" policy. The deviations are significantly larger for the case of the "minimal dissipation" policy. For some links, the analytical approximations only poorly estimate the undirected flow distributions. This is not surprising as the model only assumes uniform distributed generation which is not appropriate for the case of the data. Nevertheless, for many links the analytical approximations from the model yield a acceptable description of the data.

7.1. Discussion and Future Work

In this work, we show that using an over installation of around 50% of the average consumption, the required storage and balancing energies are in an order of magnitude that are presumably feasible in the future. However, only the cases using storage only or balancing generation only are analyzed in this work. These cases can be understood as the extreme cases, that define the upper bounds of the required storage and balancing energies. An interesting extension is to combine both cases. Limiting the storage energy to a fraction of the upper bound allows to evaluate the required balancing power from the power that cannot be satisfied using the storage. From the discussed time series of the constrained storage level, we conjecture that the storage energy can be drastically reduced with only little average annual balancing generation as the storage energy is dominated by only few large peaks. On the other hand the required balancing generation capacity can be expected to be high while being only seldomly in operation.

The spatial assignment of the generation capacities is based on the political defined national targets for the year 2020. For our analysis these capacities are scaled up to obtain the desired generation capacities. While this is a valid assumption for a first assessment as the degrees of freedom are thus reduced, a different capacity layout might allow for further optimization. A rough estimate of the optimization potential, can be derived from the analysis of the optimal mixes for the individual countries. The effect for the balancing power can be expected to be small, as the optimal mix for all countries is close to the European optimum of 80% wind and 20% solar power generation. Regarding the storage energy, we conjecture on the base of the heterogeneous optimal mixes for the individual countries that a more optimized assignment of the generation capacities could reduce the required storage energy. Also, the optimization of the spatial distribution of the generation capacities should be done with respect to the transport flows, as some of the found link flows are so high that it is unrealistic that these transport capacities can be set up. The usage of local short time storage can also help to decrease the transport flows, so that an assessment of the transport flows using local storages for each or for selected regions is an important scenario for future work. It is also interesting to calculate the flows on a finer grid that is closer to the topology of the European transmission grid but the problem is the availability of the respective network data. Applying the matching policies and calculating the transport flows is time consuming, so that for transport related optimizations the analytical approximations might be helpful as a rough estimate.

The distributions of the transport flows within the presented modelling approach were predicted with high accuracy by the analytical approximations when using the "global factor" policy. For the "minimum dissipation" policy with $\kappa = 2$ the results are not accurate enough. An extension to the case of the "minimum dissipation" policy seems to be possible. We showed that the correlation length is short and concluded that the matching takes place in a small neighborhood. A possible approach to describe the "minimum dissipation" policy is thus to calculate the matching within the neighborhood

as done for the "global factor" policy and calculated the resulting transport flows. The remaining generation can be matched using the same approach with the network outside the considered neighborhood. Due to the linearity of the model, the resulting flows of the two flow distributions for each link can be superimposed. Furthermore, the prediction of the model based approximations when analyzing the real data could be improved by systematically analyzing the distributions of the generated energy for each region and deriving approximations to replace the model assumption of uniformly distributed energy generation.

Finally, the effect of transient behavior is important to be assessed. As shown in the literature these can have a significant influence but are neglected in this work. Filatrella et al. [53], for example, proposed a model that is reduced to the essential characteristics of power generators and able to account for transient effects. Only little modifications have to be made to incorporate renewable fluctuating generation.

A. Appendix

A.1. Power and Transmission Systems Overview

Modern societies are to a high degree depended on a safe, reliable, and economic supply of electric power [59]. Power systems evolved over the last hundred years to highly complex and interdependent systems [5]. They consist of a large number of components, that can be roughly grouped into generators, loads, and transmission components [107]. Furthermore, for all these components various devices for protection and control are employed. The challenging task for engineers to implement and control such a system becomes even more complex with increased power generation from distributed or fluctuating renewable energy sources.

At the end of the 19th century, during the initial years of electricity distribution DC-transmission was the system of choice. AC transmission was not understood well enough, e.g. there was no practical AC motor available [83]. Thomas Edison promoted strongly DC-transmission, despite it's obvious drawbacks. As shown in Section A.2.5, the losses of transmission decrease quadratically with the increase of the voltage, so that a high voltage is desireable for transmission. But since the voltage in a DC-power system can not be transformed, the power has to be transmitted at the voltage level it is used at the consumers side. Thus, for security reasons the voltage can not be too high, so that huge losses are inevitable and generators had to be close to the consumers. The work of Nikola Tesla formed the basis of modern AC-Power systems and, together with George Westinghouse, he commercialized the system in the USA [83]. In Germany the question how to transport electric power was settled after the International Electro-Technical Exhibition of 1891 with a successfull demonstration of AC-power transmission [83].

The biggest advantage of AC-Systems is that the voltage can be easily transformed, so that high voltages can be used for transportation to reduce losses (see Section A.2.5) and for distribution a lower voltage ensures security and usability. Therefore, AC-Systems with different voltage levels are used for almost all transmission and distribution systems today [91].

AC-Systems have an oscillating voltage and current. This allows the use of transformers but also introduces some effects that complicate generation and transmission. All components have to be able to work with oscillating voltage and current. Almost all generated power is produced with synchronous machines that are used as generators. All of these generators have to be in synchrony as otherwise large voltage and current peaks can occur. If synchronism is lost, the generator has to be detached from the

power grid or otherwise damage to the generator is possible or even a breakdown of the whole grid can occur [90]. Engineers take also advantage of the special properties of AC-Systems. Power to meet additional loads is first taken from the energy stored in the rotating masses of the generators. That leads to a decreasing frequency that serves the system as signal to increase the energy input [107]. Besides the stability of the system, efficient economic dispatch is an important aim, so that operating and control of power grids is a highly nontrivial task.

Transmission systems are organized hierarchical. The transport network usually operates at voltages above 220 kV, for large distances up to 765 kV. It is usually meshed, so that most nodes have more than two connections, ensuring that a breakdown of one link can be compensated by other links, the so-called $N-1$ stability. The direction of the power flows usually changes with the demand or locations of the currently cheapest available generators [107]. Power flows are governed by Kirchhoff's circuit laws and Ohm's law and disperse over all parallel paths from the sources to the loads. Subtransmission systems work with voltages around 110 kV. They can work as transport networks with a meshed topology and changing power flow directions or as distribution networks with a tree-like structure and a dominating powerflow direction towards the consumers [107]. The distribution systems usually have a voltage between 10kV and 30kV and transfer the power over a tree-like network structure to the consumer [107].

AC-Transmission also comes with some disadvantages. The Skin effect, for example, reduces the effective diameter of wires thus increasing resistances. Furthermore, the impedances of the links, like mutual impedance or capacities to the ground, gain importance and have to be taken into consideration [90]. This is discussed in more detail in Section A.2. All these problems do not occur for DC-Systems. For special circumstances DC-transmission is used for selected connections in an AC environment to take advantages of the lower losses and of the insensitivity to reactances. At both ends of the DC-line an inverter and rectifier couples the line to the AC system. For offshore windparks with underwater cables DC-transmission is used as well as for transmission lines longer than 1000km [107]. The lower transmission losses face losses in the inverters and large investment costs for the additional components. Another application for this setup is a connection for power exchange between two unsynchronous AC systems.

A.2. Power flow

In this section, methods to describe and calculate the power flow in AC power systems are presented. In Section A.2.1, the concepts of active and reactive power are introduced. The exact power flow equations are analyzed in Section A.2.2. Approximations to these equations are derived in Section A.2.3. The currents flowing in a resistor network according to Kirchhoff's rules and the formal analogy to the approximations of power flows is discussed in Section A.2.4 along with methods to solve the flow equations. The advantages of high voltages when transmitting power are shown in Section A.2.5.

A.2.1. Power Factor, Active and Reactive Power

Alternating current (AC) exhibits a more complex behavior of the power flows compared to DC systems. Not only resistances have to be considered but also capacitances and inductances. A common way to work with this kind of systems is to assume sinusoidal voltages and currents with constant frequency, so that the concept of resistances in DC systems can be extended to capacitances and inductances. The complex-valued functions for voltage and current are defined as

$$U(t) = Ue^{j(\omega t+\phi_U)} = \underline{\mathrm{U}}e^{j\omega t} \qquad \text{and} \qquad I(t) = Ie^{j(\omega t+\phi_I)} = \underline{\mathrm{I}}e^{j\omega t} \ , \tag{A.1}$$

with the phases ϕ, frequency ω, and the complex amplitudes $\underline{\mathrm{U}} = Ue^{j\phi_U}$ and $\underline{\mathrm{I}} = Ie^{j\phi_I}$.

Ohm's law can be applied, yielding $I(t) = Y \cdot U(t)$, where Y is the complex admittance. For a resistor, the admittance follows as $Y = {}^1\!/_R$ and for a capacity, that is described by $i(t) = C\frac{dv(t)}{dt}$, we find $Y = j\omega C$. The complex admittance for an AC circuit with constant frequency is thus defined as

$$Y = G + jB \tag{A.2}$$

with the conductance G and the susceptance B.

When using alternating currents, the concept of power has to be extended. For purely resistive loads the current and the voltage change signs at the same time, so that the dissipated power is always positive. Energy is dissipated at the resistor, referred to as active power. For an ideal capacitance or inductance, the current and the voltage are 90° out of phase, so that the product of the current and the voltage is positive for half a cycle and negative for the other. Thus, for reactive loads the consumed net power is zero. The power that flows is referred to as reactive power [52, 107]. A good example is the an LC-circuit, consisting of an ideal inductance and a capacitance. It can be mathematically described as a harmonic oscillator and the energy flows back and forth between both components. For ideal components no energy is dissipated and only reactive power is flowing. In real systems, all components have a resistance, so that the oscillations will eventually die out.

In the following, a description of active and reactive power is derived. Using the AC voltage $u(t) = U\cos(\omega t)$ and the current $i(t) = I\cos(\omega t - \varphi)$, so that φ denotes the phase difference, the instantaneous power at time t follows as

$$p(t) = u(t)i(t) = U \cdot I\cos(\omega t)\cos(\omega t - \varphi) \ . \tag{A.3}$$

Applying the trigonometric identity $\cos(\omega t - \varphi) = \cos(\omega t)\cos(\varphi) - \sin(\omega t)\sin(\varphi)$, the instantaneous power follows as [35]

$$p(t) = \underbrace{U \cdot I\cos(\varphi)}_{P}\left(1 + \cos^2(2\omega t)\right) + \underbrace{U \cdot I\sin(\varphi)}_{Q}\sin(2\omega t) \ . \tag{A.4}$$

When looking at Equation (A.4), we find that $\langle p(t) \rangle = P$. The second part, containing the reactive power Q averages to zero. The average power P is referred to as active power. The power oscillates around P with amplitude P. The second term in Equation (A.4) denotes the power flowing in and out of the component with the amplitude Q. The reactive power is not consumed since it is the part of the power flow where U and I have a 90 degree phase difference. In power systems this can be problematic, e.g. some of the transported power over long transmission lines is stored in the capacitance of the line, thus reducing the transported active power. When using the complex voltage and current from Equation (A.1), the complex power is defined by

$$\underline{\mathrm{S}} = \underline{\mathrm{U}}\underline{\mathrm{I}}^* = UIe^{j\varphi} = Se^{j\varphi} \ , \tag{A.5}$$

with φ again the phase difference between the current and the voltage. Either U or I have to be complex conjugated, so that the complex power is equal to the power for DC systems. The complex conjugated current is used by convention [90]. The absolute value $S = |\underline{\mathrm{S}}| = S\,S^* = \sqrt{P^2 + Q^2}$ is referred to as apparent power. Using $e^{j\varphi} = \cos(\varphi) + j\,\sin(\varphi)$, the complex power can be written as [2]

$$\underline{\mathrm{S}} = UI\cos(\varphi) + jUI\sin(\varphi) = P + jQ \tag{A.6}$$

Thus, the complex power $\underline{\mathrm{S}}$ yields all information if the instantaneous power in Equation (A.4).

Another important parameter to describe power flow is the ratio of the real power to the apparent power

$$p_f = \frac{P}{\underline{\mathrm{S}}} = \frac{|P|}{|\underline{\mathrm{S}}|} = \cos(\varphi) \ . \tag{A.7}$$

Both the active and reactive power have to balanced for AC power systems to be stable. In good approximation, unbalanced active power leads to deviations of the frequency and unbalanced reactive power to deviations from the desired voltage [90, 107]. Both can be understood by looking at the mode of operation of the synchronous generators. In the former case the power taken out is larger than the external propulsive power, so the power stored in the rotating mass is consumed leading to a decaying rotation speed and thus decaying frequency. In the latter case, the voltage decreases when the reactive power in the system increases because of an inductive reaction of the armature [35, 90, 107].

A.2.2. Calculation of the Exact Power Flow

When calculating power flows in a power grid, the admittance matrix $\mathbf{Y}$ is used to describe the system. The entries $\mathbf{Y}_{ik} = G_{ik} + jB_{ik}$ consist of the admittance of the

link $i \to k$ and the diagonal is given by $\mathbf{Y}_{ii} = \sum_{k \neq i} \mathbf{Y}_{ik}$. In the stationary state, the complex power at each node is given by

$$S_i = P_i + jQ_i = U_i \left(\sum_k \mathbf{Y}_{ik} U_k \right)^* . \tag{A.8}$$

To describe the transient behavior, the full differential equations of the system have to be solved. The in/outflow of active and reactive power at vertex j has to equal to the power that is transfered to or from neighboring vertices [90]. For given complex power S_i at each vertex, the voltage $\vec{U}^s$ that solves Equation (A.8) can be found by solving $f_j\left(\vec{U}^s\right) = 0$, with

$$f_i\left(\vec{U}\right) = P_i + jQ_i - \left(U_i \left(\sum_k \mathbf{Y}_{ik} U_k \right)^* \right) \tag{A.9}$$

for all vertices. It is an equation second order in the voltage, so that it has a well defined solution but it can not be derived in a closed form. To find the solution of Equation (A.9), Newton's method can be used. It is a well established method for root finding of nonlinear equations [see e.g. 103, 107]. Newton's method iteratively refines an initial guess by repeatedly linearizing the nonlinear function around the current guess and subsequently solving the linear problem [see e.g. 23, 103, 108]. Newton's method is only guaranteed to converge in the vicinity of the solution of the nonlinear system. Various extensions are proposed to ensure global convergence and to improve the numerical performance [23, 48]. To find the solution $f(x) = b$, the method is defined as

$$f(x_j) - b = f'(x_j)\, \Delta x_j \tag{A.10}$$

with x_0 the initial guess and $\Delta x_j = x_{j+1} - x_j$.

The entries of $\mathbf{Y}_{ik}$ and the voltages are complex valued quantities and can be expressed in cartesic or polar coordinates,

$$U_k = |U_k| e^{j\delta_k} \qquad \text{and} \qquad Y_{ik} = |Y_{ik}| e^{j\alpha_{ik}} . \tag{A.11}$$

Thus, the potentials $U_k(|U_k|, \theta_k)$ are to be found that solve Equation (A.8) for given complex power.

In the following, we will show the derivation of Newton's method for the power flow in Equation (A.9) as presented by Oeding and Oswald [90]. Equation (A.8) can be rearranged to

$$\mathbf{U}\, \mathbf{Y}\, \vec{U} - \left(\vec{P} + j\vec{Q} \right) = \vec{0} \,, \tag{A.12}$$

with the diagonal matrix $\mathbf{U} = \mathrm{diag}(U_1, \ldots, U_N)$. Since Equation (A.12) has to be fullfilled for both real and the imaginary part, the equations

$$\Re\left[\mathbf{U}\,\mathbf{Y}\,\vec{U}\right] - \vec{P} = \Delta\vec{P} \tag{A.13}$$

$$\Im\left[\mathbf{U}\,\mathbf{Y}\,\vec{U}\right] - \vec{Q} = \Delta\vec{Q} \tag{A.14}$$

have to be solved, so that $\Delta\vec{P}$ and $\Delta\vec{Q}$ are zero. The voltages at iteration step $\nu+1$ are normalized to the voltages to the previous iteration step

$$|u_i^{(\nu+1)}| = \frac{|U_i^{(\nu+1)}|}{|U_i^{(\nu)}|}\,. \tag{A.15}$$

Defining the state vector

$$\vec{x} = (\delta_1, \ldots, \delta_N, u_1, \ldots, u_N)^T\,, \tag{A.16}$$

Equations (A.13) and (A.14) yield for the next iteration step $\nu+1$ using Newton's method

$$\left(\frac{\partial\Delta\vec{P}}{\partial\vec{x}}\right)_{(\nu)}\Delta\vec{x}_{(\nu+1)} + \Delta\vec{P}_{(\nu)} = \vec{0} \qquad \text{and} \qquad \left(\frac{\partial\Delta\vec{Q}}{\partial\vec{x}}\right)_{(\nu)}\Delta\vec{x}_{(\nu+1)} + \Delta\vec{Q}_{(\nu)} = \vec{0}\,, \tag{A.17}$$

with $\Delta\vec{x}_{(\nu+1)} = \vec{x}_{(\nu+1)} - \vec{x}_{(\nu)}$ and the partial derivations calculated at the point of the last solution $\vec{x}_{(\nu)}$. These can be merged to

$$\begin{pmatrix} \frac{\partial\Delta\vec{P}}{\partial\vec{x}} \\ \frac{\partial\Delta\vec{Q}}{\partial\vec{x}} \end{pmatrix}_{(\nu)}\Delta\vec{x}_{(\nu+1)} = \underbrace{\begin{pmatrix} \mathbf{H} & \mathbf{K} \\ \mathbf{M} & \mathbf{L} \end{pmatrix}_{(\nu)}}_{\mathbf{J}_{(\nu)}}\Delta\vec{x}_{(\nu+1)} = -\begin{pmatrix} \Delta P \\ \Delta Q \end{pmatrix}_{(\nu)}\,, \tag{A.18}$$

where the Jacobian matrix $\mathbf{J}_{(\nu)}$ can be written in terms of four submatrices. For every iteration step, the Jacobian matrix has to be calculated and Equation (A.18) to be solved to find the refined solution $\vec{x}_{(\nu+1)} = \vec{x}_{(\nu)} + \Delta\vec{x}_{(\nu+1)}$. There are various algorithms to solve linear systems of equations of the form $\mathbf{A}\vec{x} = \vec{b}$ [see e.g. 103].

Oeding and Oswald [90] show that using the matrix

$$\mathbf{R} = \begin{pmatrix} |U_1||Y_{11}||U_1|e^{j(\delta_{11}-\alpha_{11})} & \cdots & |U_1||Y_{1N}||U_N|e^{j(\delta_{1N}-\alpha_{11})} \\ \vdots & \ddots & \vdots \\ |U_N||Y_{N1}||U_1|e^{j(\delta_{N1}-\alpha_{N1})} & \cdots & |U_N||Y_{NN}||U_N|e^{j(\delta_{NN}-\alpha_{NN})} \end{pmatrix} \tag{A.19}$$

with $\delta_{ik} = \delta_i - \delta_k$[1], the submatrices of the Jacobian matrix $\mathbf{J}_{(\nu)}$ follow as

$$\begin{aligned}
\mathbf{H}_{(\nu)} &= \Im\left[\mathbf{R}\right] - \mathrm{diag}\left(\vec{Q}_{(\nu)}\right) && \text{(A.20)}\\
\mathbf{M}_{(\nu)} &= -\Re\left[\mathbf{R}\right] - \mathrm{diag}\left(\vec{P}_{(\nu)}\right) && \text{(A.21)}\\
\mathbf{K}_{(\nu)} &= \Re\left[\mathbf{R}\right] - \mathrm{diag}\left(\vec{P}_{(\nu)}\right) && \text{(A.22)}\\
\mathbf{L}_{(\nu)} &= \Im\left[\mathbf{R}\right] - \mathrm{diag}\left(\vec{Q}_{(\nu)}\right) \ , && \text{(A.23)}
\end{aligned}$$

with $\mathrm{diag}(\vec{y})$ indicating a matrix with the vector $\vec{y}$ on the diagonal. On the diagonals of the matrices $\mathbf{H}$, $\mathbf{M}$, $\mathbf{K}$, and $\mathbf{L}$, the active and the reactive power of the current solution $\vec{U}_{(\nu)}$, given by

$$\begin{aligned}
\vec{P}_{(\nu)} &= \Re\left[\mathbf{U}_{(\nu)}\,\mathbf{Y}\,\vec{U}_{(\nu)}\right]\\
\vec{Q}_{(\nu)} &= \Im\left[\mathbf{U}_{(\nu)}\,\mathbf{Y}\,\vec{U}_{(\nu)}\right] \ , && \text{(A.24)}
\end{aligned}$$

are subtracted.

A.2.3. Decoupling of the Power Flow and DC Approximation

The Jacobian matrix $\mathbf{J}_{(\nu)}$ has to be calculated for every iteration step and might be expensive to evaluate. Using some approximations, the problem can be simplified, so that the Jacobian has to be calculated only once. In power grids the differences of the voltage angles δ_{ik} are typically small and the impedances dominated by the reactances. Then the approximations

$$\cos(\delta_{ik}) = 0 \qquad \text{and} \qquad \sin(\delta_{ik} - \alpha_{ik}) = 0 \tag{A.25}$$

hold for the case of power grids [90].

The matrix $\mathbf{R}$ in Equation (A.19) can then be simplified and the entries of the matrices $\mathbf{M}$ and $\mathbf{K}$ in Equations (A.23) go to zero. The remaining system of equations with the simplified matrices $\mathbf{H}$ and $\mathbf{L}$, denoted by $\mathbf{H}'$ and $\mathbf{L}'$, is given by

$$\mathbf{H}'\Delta\delta = -\Delta p \qquad \text{and} \qquad \mathbf{L}'\Delta u = -\Delta q \ . \tag{A.26}$$

In power grids, the voltages are tried to be kept as uniform as possible. The influence of the voltages amplitudes $|U_k|$ in Equation (A.8) on vertex i is thus approximated by the voltage amplitude $|U_i| \approx |U_k|$. Together with the above approximations, we then find

$$S_i = P_i + jQ_i = |U_i|^2 \sum_k \mathbf{Y}^*_{ik} e^{j(\delta_{ik} - \alpha_{ik})} \ . \tag{A.27}$$

[1]Since $\delta_{ii} = 0$, the diagonal elements simplify to $|U_i|^2|Y_{ii}|e^{-j\alpha_{ii}}$

Equations (A.26) and (A.27) can be rewritten to

$$\mathbf{B}'\Delta\vec{\delta} = \mathbf{U}^s\Delta P \tag{A.28}$$
$$\mathbf{B}''\Delta\vec{u} = \mathbf{U}^s\Delta Q \tag{A.29}$$

with the matrix elements $\mathbf{U}^s_{ii} = |U_i|^2$ on the diagonals and zero for the nondiagonal elements. The entries of the matrices $\mathbf{B}'$ and $\mathbf{B}''$ are given by [107]

$$\mathbf{B}'_{ik} = \mathbf{B}''_{ik} = B_{ik} \qquad \text{for } i \neq k \tag{A.30}$$
$$\mathbf{B}'_{ii} = -\sum_{k=1,k\neq i}^{N} B_{ik} \tag{A.31}$$
$$\mathbf{B}''_{ii} = 2B_{ii} + \sum_{k=1,k\neq i}^{N} B_{ik}\ . \tag{A.32}$$

Only the reactances B_{ik} enter in this approximation, as the resistances in the addmittance matrix $\mathbf{Y}$ are assumed to be small. B_{ii} describe the reactances that connects all vertices i to a common reference node to account for losses at the nodes. The matrices $\mathbf{B}'$ and $\mathbf{B}''$ are constant in $\vec{u}$, so that the system to be solved is linear. This approach is referred to as "decoupled power flow".

Further approximations can be made to derive the "DC power flow". A uniform voltage $|U_i| = U$ is assumed, so that Equation (A.29) is zero on the left side and the reactive power does not need to be considered [90, 107, 126]. Further, the $\mathbf{B}_{ii}$ are assumed to be small. The active power in Equation (A.13) can then be rewritten as

$$\Delta\vec{P}_i = \left[|U_i|^2 \sum_k |\mathbf{Y}_{ik}| \cos\left(\delta_{ik} - \alpha_{ik}\right)\right] - P_i\ . \tag{A.33}$$

Again, the resistances for $\Delta\vec{P}_i$ are neglected and assumed that the inductive part of the lines dominates the admittances, so that $\alpha_{ik} \to -\pi/2$. The term $\cos\left(\delta_{ik} - \alpha_{ik}\right)$ is thus replaced by $-\sin\left(\delta_{ik}\right)$. For small voltage angle differences δ_{ik}, the sine function can be approximated, $\sin\left(\delta_{ik}\right) \approx \delta_{ik}$, so that [90]

$$\Delta\vec{P}_i \approx -\left[U^2 \sum_k |\mathbf{Y}_{ik}| \left(\delta_i - \delta_k\right)\right] - P_i = U^2 \left(\mathbf{B}'\vec{\delta}\right)_i - P_i\ . \tag{A.34}$$

In matrix notation, this can be expressed as

$$\vec{P} = U^2\mathbf{B}'\vec{\delta}\ . \tag{A.35}$$

For simplicity, usually a voltage $U = 1$ is assumed. The power on the line $i \to k$ is given by [126]

$$P_{jk} = B_{jk}\left(\delta_j - \delta_k\right)\ . \tag{A.36}$$

A.2.4. Kirchhoff Flow

Kirchhoffs circuit laws consist of the current law (KCL) and the potential law (KPL) and state that:

KCL the sum over all currents entering a node in an electrical network has to be equal to the sum of all currents leaving that node,

KPL the sum of all potential differences around any closed loop is zero.

To transport electrical charges, the currents I_j are injected or drawn at each vertex j. The current from vertex k to k is denoted by I_{jk}. With that definition the KCL is given by

$$I_j + \sum_{k \in \mathcal{N}_j} I_{jk} = 0 \; . \tag{A.37}$$

The currents I_j are defined to be positive if current is injected at vertex j and negative for current drawn from the vertex. A direction is associated to the links[2] and a positive current indicates a flow in the direction of the link, a negative the opposite. To every vertex a potential U_j is assigned. For a closed loop s, where $s = \{s_1, \ldots, s_l, s_1\}$ gives the indices of the vertices, the KPL can be written as

$$\sum_{i=1}^{l} \left(U_{s_i} - U_{s_{i+1}} \right) = 0 \; . \tag{A.38}$$

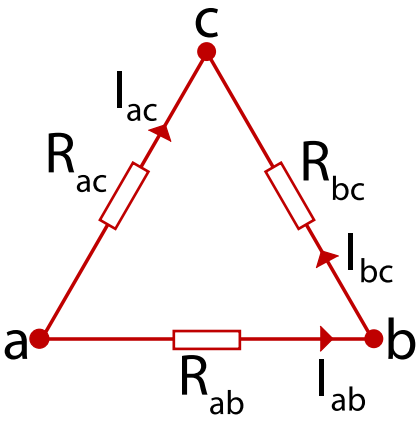


Figure A.1.: *A circuit that consists of a network of three resistors*

Using the KCL in Equation (A.37) and Ohm's law, $U = R \cdot I$, the currents in the network can be calculated for given potentials that satisfy the KPL. For a simple circuit, as shown in Figure A.1 the currents follow as

$$\begin{aligned}
I_a &= I_{ab} + I_{ac} = (U_a - U_b)\, R_{ab}^{-1} + (U_a - U_c)\, R_{ac}^{-1} \\
&= \left(R_{ab}^{-1} + R_{ac}^{-1} \right) U_a - R_{ab}^{-1} U_b - R_{ac}^{-1} U_c \\
I_b &= I_{bc} - I_{ab} = -R_{ab}^{-1} U_a + \left(R_{ab}^{-1} + R_{bc}^{-1} \right) U_b - R_{bc}^{-1} U_c \\
I_c &= -I_{ac} - I_{bc} = -R_{ac}^{-1} U_a - R_{bc}^{-1} U_b + \left(R_{ac}^{-1} + R_{bc}^{-1} \right) U_a \; .
\end{aligned} \tag{A.39}$$

The signs are chosen such that a current into the vertex is positive. Equation (A.39) can be generalized to

$$\vec{I} = \mathbf{Y} \vec{U} \; , \tag{A.40}$$

[2] As $I_{jk} = -I_{kj}$ the direction of the link is not of importance as long as it is known from which node to which the current flows.

where $\mathbf{Y}$ is the admittance matrix with the properties

$$\mathbf{Y}_{ij} = -R_{ij}^{-1} \tag{A.41}$$

for $i \neq j$ and

$$\mathbf{Y}_{ii} = \sum_j R_{ij}^{-1} \, . \tag{A.42}$$

Using Ohm's law, the current along the line from i to j is given by $I_{ij} = R_{ij}^{-1} \, (U_j - U_i)$.

When comparing Equation (A.40) with Equation (A.35), we see that formally the equations are equal. The matrix $\mathbf{Y}$ has the same structure as $\mathbf{B}'$, the diagonal elements are the sums of the respective columns, with the difference that the entries of $\mathbf{B}'$ consist of the reactances only. The total power sums up to zero as do the currents, to fullfill the conservation of energy. Thus, current flows governed by the Kirchhoff equations are equal to power flows resulting from the approximation in Equation (A.35).

The properties of the Kirchhoff flow will be elaborated in the following in more details, mostly based on Bollobás [19]. These definitions were derived in the context of graph theory which was introduced in Section 2.1. The incidence matrix as introduced in Equation (2.2) is defined by

$$\mathbf{K}_{ij} = \begin{cases} 1 & v_i \text{ is the initial vertex of edge } e_j \\ -1 & v_i \text{ is the terminal vertex of edge } e_j \\ 0 & \text{otherwise} \end{cases} \tag{A.43}$$

and has a distinct meaning in the context of flows. Since each row of the transposed incidence matrix $\mathbf{K}^t$ has a one at the initial and a minus one at the terminal vertex, $\mathbf{K}^t \, u$ gives the potential differences $(u_i - u_j)$ at every edge $e_{i \to j}$. Kirchhoffs potential law (KPL) postulates that the sum of all potential differences around any cycle is zero. The sum over the potential differences of a cycle $\{v_1, \ldots, v_k, v_1\}$ is given by

$$\left(\sum_j^{j-1} u_j - u_{j+1} \right) - u_1 = (u_1 - u_2) + (u_2 - u_3) + \cdots + (u_{k-1} - u_k) - u_1 = 0 \, . \tag{A.44}$$

Thus, the KPL is always fullfilled for voltage differences defined by $\mathbf{K}^t \, \vec{u}$. The flow along a link is according to Ohm's law $f_{ij} = \frac{u_i - u_j}{r_{ij}}$, which can be put into the form

$$\vec{f} = \mathbf{R}^{-1} \mathbf{K}^t \vec{u} \, , \tag{A.45}$$

with $\mathbf{R}^{-1}$ a $M \times M$ matrix with the admittances of the respective links on the diagonal,

$$\mathbf{R}_{ij}^{-1} = \begin{cases} R_{e_i}^{-1} & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \, , \tag{A.46}$$

and $\vec{f}$ a vector of length M that gives the flows along the links. Kirchhoff's current law (KCL) as defined in Equation (A.40) can be written as $\mathbf{K}\vec{f} = \vec{s}$ for all flows that satisfy Equation (A.45). Thus, Kirchhoff's law can be put into the form

$$\vec{s} = \mathbf{K}\mathbf{R}^{-1}\mathbf{K}^t\vec{u}\,. \tag{A.47}$$

The matrix

$$\mathbf{Y} = \mathbf{K}\mathbf{R}^{-1}\mathbf{K}^t \tag{A.48}$$

maps directly from the potentials to the in and out flows at each vertex. For $R_{ij} = 1$, this matrix is equal to the Laplace matrix, see Section 2.1.1. Equation (A.47) can then be written as

$$\vec{s} = \mathbf{Y}\,\vec{u}\,. \tag{A.49}$$

or as

$$\vec{s} = \mathbf{L}\,\vec{u}\,. \tag{A.50}$$

for $R_{ij} = 1$.

A.2.4.1. Solving the Flow Equations

The Laplace matrix $\mathbf{L}$ of a graph consisting of one connected component is positive-semidefinite with one eigenvalue that is zero[3]. Therefore, Equation (A.49) is an under-determined system of equations, as $\vec{s}$ is invariant to a translation of $\vec{u}$ with the eigenvector $\vec{v}_0$, that fullfills $\mathbf{L}\vec{v}_0 = 0$, so that $\vec{s} = \mathbf{L}\,(\vec{s} + \vec{v}_0) = \mathbf{L}\vec{s}$. $\mathbf{L}$ is thus not invertible and the solution to $\vec{s} = \mathbf{L}\vec{u}$ for given $\vec{s}$ can not be calculated directly.

In the literature, this problem is often avoided by fixing the potential of a reference vertex to zero and thus removing the corresponding column and row from the Laplace matrix. The resulting matrix consists of independent columns and has a unique solution [see e.g. 89].

A different approach to solve Equation (A.49) for $\vec{u}$ is the use of the Moore-Penrose pseudoinverse $\mathbf{L}^+$ [109, 121], in the following only referred to as pseudo inverse [93]. It provides a least squares solution to a system of linear equations [94]. Given a system of linear equations

$$\mathbf{A}\vec{x} = \vec{b} \tag{A.51}$$

it may not be possible to find a vector $\vec{x}$ which solves the equation for an over-determined system. For an under-determined system, the solution is not unique. For our problem in Equation (A.49), we have an under-determined system and the pseudoinverse approach

[3]The multiplicity of the zero eigenvalue is equal to the number of components [17].

yields the solution that has smallest Euclidean norm $|x|$ [94]. The Laplace matrix is a symmetric matrix, so a eigendecomposition of the form

$$\mathbf{L} = \mathbf{M}\mathbf{\Lambda}\mathbf{M}^t \tag{A.52}$$

is possible, where $\mathbf{M}$ is an orthogonal matrix and $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues λ_i. Since $\mathbf{L}$ is positive-semidefinite all eigenvalues satisfy $\lambda_i \geq 0$. Using the eigendecomposition, the pseudo inverse is defined as [94]

$$\mathbf{L}^+ = \mathbf{M}\mathbf{\Lambda}^+\mathbf{M}^t \tag{A.53}$$

with

$$\mathbf{\Lambda}^+_{jj} = \begin{cases} \lambda_j^{-1} & \text{for } \lambda_j \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{A.54}$$

The solution $\vec{u} = \mathbf{Y}^+\vec{s}$ and $\vec{u} = \mathbf{L}^+\vec{s}$ for Equations (A.49) and Equation (A.50) thus gives the potential vector $\vec{u}$ with the smallest Euclidean norm and is used in this work to calculate the potential vector $\vec{u}$ for a given in/out flow vector $\vec{s}$.

A.2.5. High Voltage Transmission

Transmitting power always implies losses due to the resistances of the wires. The system should be designed so that these losses are minimal. The power dissipation by a resistance of the wire is given by $P_{wire} = \Delta U \cdot I$, where ΔU is the voltage drop along the line. Using Ohm's law and the power to be transfered $P = U \cdot I$, we find

$$P_{wire} = R_{wire} \cdot I^2 = R_{wire}\frac{P^2}{U^2} \ . \tag{A.55}$$

For given line resistance R_{wire} and power to transport P, the losses P_{wire} are minimal if the voltage is maximal. On the other hand, increasing the voltage will result in a higher current by Ohm's law or it has to be assumed that the resistance of the consumer is changed so that the current decreases.

To understand why high voltages are used for transmission, the whole transmission system and the load has to be taken into account. A schematic circuit diagram is shown in Figure A.2. The system consists of a transmission line with resistance R_{link} and a transformer with a impedance Z_t in the primary circuit. On the secondary side the load has an impedance Z_{load}. For an ideal transformer the incoming power on the primary side has to equal the outgoing power on the secondary side, $P_{in} = U_p \cdot I_p = P_{out} = U_s \cdot I_s$, so that the ideal transformer equation follows as:

$$\frac{V_s}{V_p} = \frac{N_s}{N_p} = \frac{I_p}{I_s} \ , \tag{A.56}$$

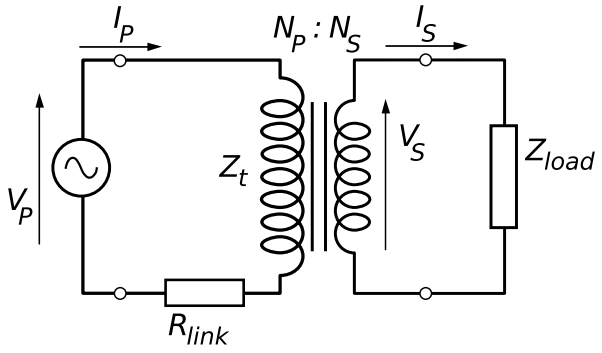


Figure A.2.: *Simplified circuit diagram of the transmission from a voltage source over over a transmissionline with resistance R_{link} using a transformer to convert the voltage for the load with impedance Z_{load}.*

where N_P and N_S are the number of turns of the transformer on the primary and secondary side, respectively. By Ohm's law, we have the relations between the impedances, currents, and voltages

$$Z_{load} = \frac{V_s}{I_s} , \qquad Z_t = \frac{V_p}{I_p} . \tag{A.57}$$

Using Equation (A.56) and Equations (A.57), the impedance of the transformer can be expressend in dependence of the load impedance and the ratio of the voltages on the primary and secondary side,

$$Z_t = Z_{load} \left(\frac{N_p}{N_s}\right)^2 = Z_{load} \left(\frac{V_p}{V_s}\right)^2 . \tag{A.58}$$

We see that the total impedance in the primary circuit $Z_t + R_{link}$ increases when increasing the primary voltage relative to the secondary voltage thus decreasing the current I_p. This explains why the current can be assumed to decrease in Equation (A.55) for increasing voltage.

Given the power to be consumed at the load $P^* = Z_{load} \cdot I_s^2$, we find in terms of the primary side

$$P^* = V_t \cdot I_p = Z_{load} \left(\frac{V_p}{V_s}\right)^2 \cdot I_p^2 \tag{A.59}$$

and using the losses at the line

$$P_{loss} = R \cdot I_p^2 , \tag{A.60}$$

the relative loss follows as

$$\frac{P_{loss}}{P^*} = \frac{R}{Z_{load}} \left(\frac{V_s}{V_p}\right)^2 \propto \left(\frac{V_s}{V_p}\right)^2 . \tag{A.61}$$

Thus, we obtain the result that increasing the voltage of the primary side, the voltage used for transmission, by a factor of ten means that the ratio of the transmission loss to the consumed power decreases by a factor of hundred.

A.3. Probability Distributions

A.3.1. Properties of Probability Distributions

Some basic properties of probability distribution used in this work are summarized here. The raw moments of a random variable X drawn from the probability distribution $p(x)$ are defined as $\mu'_k = E\left[X^k\right]$, with the expectation operator

$$E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)\, p(x) dx \ .$$

The central moments are defined as deviations from the mean μ, $\mu_k = E\left[(X-\mu)^k\right]$. The cumulants generating function of the random variable X is defined as [34]

$$g(t) = \log\big(E\left[e^{tX}\right]\big) \ , \tag{A.62}$$

The cumulants κ_n follow as the derivatives of the cumulant generating function at $k = 0$ [97]

$$\kappa_1 = \mu = g'(0), \tag{A.63}$$
$$\kappa_2 = \sigma^2 = g''(0), \tag{A.64}$$
$$\vdots$$
$$\kappa_n = g^{(n)}(0) \ . \tag{A.65}$$

The first cumulant equals the mean, while the second and third are equal to the second and third central moments.

First, the additivity of cumulants is shown. The cumulants of the sum of two independent random variables X and Y are calculated using

$$\begin{aligned} g_{X+Y}(t) &= \log\big(E\left[e^{t\cdot(X+Y)}\right]\big) = \log\big(E\left[e^{tX}\right] \cdot E\left[e^{tY}\right]\big) \\ &= \log\big(E\left[e^{tX}\right]\big) + \log\big(E\left[e^{tY}\right]\big) = g_X(t) + g_Y(t) \ , \end{aligned} \tag{A.66}$$

so that the cumulant κ_j of the sum distribution p_{X+Y} is given by the sum of the respective cumulants of the two distributions. With the same method, it is easy to show that for the cumulants and the central moments

$$\kappa_n(cX) = c^n \kappa_n(X) \ , \qquad \mu_n(cX) = c^n \mu_n(X) \tag{A.67}$$

holds. Thus, for random variable $Y = \alpha X$, with X drawn from a normal distribution $N(x;\mu,\sigma^2)$, follows that Y is distributed as $N(y;\alpha\mu,\alpha^2\sigma^2)$. If X is uniformly distributed $U(x;a,b)$, with a and b the upper and lower bound, the distribution of Y is given by $U(y;\alpha a,\alpha b)$.

Another important relation is that the distribution of the sum S_{X+Y} of two random variables X and Y drawn independently from the distributions $p_X(x)$ and $p_Y(y)$ is given by the convolution of the two distributions

$$S_{X+Y}(z) = (p_X * p_Y)(z) = \int_{-\infty}^{\infty} p_X(z-y) \cdot p_Y(y) dy \tag{A.68}$$

This is proved by

$$\begin{aligned} S_{X+Y}(z) &= \iint_{-\infty}^{\infty} p_X(x) \cdot p_Y(y) \delta(z-(x+y))\, dxdy & \text{(A.69)} \\ &= \int_{-\infty}^{\infty} p_X(x) \cdot p_Y(z-x) dx = \int_{-\infty}^{\infty} p_X(z-y) \cdot p_Y(y) dy & \text{(A.70)} \end{aligned}$$

where the integration is done either over x or y. Further, we see that $(p_X * p_Y)(z) = (p_Y * p_X)(z)$.

A.3.2. Fitting the Generalized Gamma Distribution

The generalized gamma distribution was introduced by Stacy [111]. It is given by

$$f(x; b, d, k) = d\, \frac{x^{dk-1}}{b^{dk}\Gamma(k)} \exp\left(-\left(\frac{x}{b}\right)^d\right) , \tag{A.71}$$

with positive b, d and k and the gamma function $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt$ [112]. The raw moments are given by

$$\mu'_n = b^n \frac{\Gamma\left(k + \frac{n}{d}\right)}{\Gamma(k)} . \tag{A.72}$$

Prentice [95] derived a maximum likelihood solution to determine the parameters b, d, and k from a set of data, but the method often fails to converge and gave in our simulations poor results. Stacy and Mihram [112] derived a way to calculate the parameters from the data, but used the mean, variance and skewness of the logarithm of the data. The latter is not available from the derivations in Chapter 5 so this method can not be used. The method of moments [see e.g. 112], where the analytical moments depending on the parameters are compared to the moments calculated from the data, was far more successfull. The first raw moments μ'_n in dependence of the parameters b, d, and k are given by

$$\begin{aligned} \mu'_1 &= b \frac{\Gamma\left(k + \frac{1}{d}\right)}{\Gamma(k)} & \text{(A.73)} \\ \mu'_2 &= b^2 \frac{\Gamma\left(k + \frac{2}{d}\right)}{\Gamma(k)} & \text{(A.74)} \\ \mu'_3 &= b^3 \frac{\Gamma\left(k + \frac{3}{d}\right)}{\Gamma(k)} . & \text{(A.75)} \end{aligned}$$

We therefore solve Equations (A.73) to (A.75) directly using an iterative method similar to the work by Ashkar et al. [11]. The system of two equations

$$b(\mu_1', b_i, d_i, k_i)^2 \cdot \frac{\Gamma\left(k_i + \frac{2}{d_i}\right)}{\Gamma(k_i)} - \mu_2' = 0 \tag{A.76}$$

$$b(\mu_1', b_i, d_i, k_i)^3 \cdot \frac{\Gamma\left(k_i + \frac{3}{d_i}\right)}{\Gamma(k_i)} - \mu_3' = 0\,, \tag{A.77}$$

with $b(\mu_1', b, d, k) = \mu_1' \frac{\Gamma(k)}{\Gamma\left(k+\frac{1}{d}\right)}$, is solved iteratively starting from an initial guess b_0, d_0, and k_0. Equations (A.76) and (A.77) can be solved using a Newton-Krylov method. In an outer loop the system is linearized using Newton's method and in the inner loop the linear system is approximately solved using a method based on Krylov subspace approximations. Here, the GMRES solver is used [103] to solve the linearized system. For a nonlinear system Newton's method is known to converge q-quadratically in the vicinity of the solution [42], but far away from the solution it is not guaranteed to converge and might even diverge. To ensure global convergence some methods have been proposed as summarized by Brown and Saad [23]. Further techniques to speed up solving the nonlinear root finding problem, is to decrease the accuracy of the inner linear solver far away from the solution [48]. The above described methods are implemented in the package PETSc [12, 13], that we used to solve Equations (A.76) and (A.77) numerically.

Newton's method needs the Jacobian matrix of the Equations (A.73) to (A.75), which is given in Ashkar et al. [11]. Since the gamma functions are highly nonlinear for negative arguments and b, d, and k larger have to be larger than zero by definition, we substituted the term $\frac{\Gamma\left(k+\frac{n}{d}\right)}{\Gamma(k)}$ in Equation (A.73) to Equation (A.75) by

$$\gamma_n = \begin{cases} \frac{\Gamma\left(k+\frac{n}{d}\right)}{\Gamma(k)} & \text{for } k > 0 \text{ and } d > 0 \\ (-d+1)^\alpha + (-k+1)^\alpha & \text{else} \end{cases}\,, \tag{A.78}$$

with $\alpha \ll 1$. This definition forces negative values of d and k back to the positive semiaxis. Numerically the values of the gamma functions might give extremely large values while the ratio may stay small. This is implemented by taking the logarithm of the fraction using functions that return the logarithm of the gamma function $\Gamma_L(\cdot)$ that are implemented for example in the C++ library boost [20], and taking the exponent of the differences

$$\frac{\Gamma\left(k+\frac{n}{d}\right)}{\Gamma(k)} = \exp\left(\Gamma_L\left(k+\frac{n}{d}\right) - \Gamma_L(k)\right) \tag{A.79}$$

allows to calculate the fraction of the gamma functions numerically. Further, the raw moments are sometimes not easy to handle numerically, as they might be very large or

small, due to the exponents. This can be circumvented by using

$$\begin{aligned} \beta\mu_1' &= b'\gamma_1 && \text{(A.80)} \\ \beta^2\mu_2' &= b'^2\gamma_2 && \text{(A.81)} \\ \beta^3\mu_3' &= b'^3\gamma_3 \, . && \text{(A.82)} \end{aligned}$$

so that b is given by $b = \frac{b'}{\beta}$. The factor β is chosen larger than one if the higher moments are small, which is the case for our problems. To calculate the parameters of the generalized gamma distribution in Chapter 5, $\beta = 10$ is used.

With this approach the parameters of the generalized gamma distribution could be estimated robustly from the first three raw moments.

A.3.3. Sum of N Nonidentically Distributed Uniform Random Variables

Here, the derivation of the results from Bradley and Gupta [21] are briefly outlined. The distribution of the sum of N nonidentically independent distributed uniform random variables is calculated. The probability density of $\sum_{j=1}^{N} X_j$, where the X_j are drawn from the mutually independent probability density functions (PDFs) $p_j(x)$, is given by $f_N(x) = (p_1 * p_2 * \cdots * p_N)(x)$. The PDFs $p_j(x)$ are uniform distributions on the interval $[c_j - a_j, c_j + a_j]$.

The characteristic function of $f_N(x)$ is given by

$$\tilde{f}_N(x) = \prod_{j=1}^{N} \tilde{p}_j = \prod_{j=1}^{N} e^{itc_j} \operatorname{sinc}(a_j t) \ . \tag{A.83}$$

Bradley and Gupta [21] start with a Fourier inversion that gives

$$f_N(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \tilde{f}_N(x) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \prod_{j=1}^{N} \operatorname{sinc}(a_j t) \ , \tag{A.84}$$

with $y = x - \sum_{j=1}^{N} c_j$. A change of the variable $t \to -t$ yields

$$f_N(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ity} \prod_{j=1}^{N} \operatorname{sinc}(a_j t) \, dt \, . \tag{A.85}$$

The sinc function is replaced by $\operatorname{sinc}(a_j t) = \frac{\sin(a_j t)}{t} = \frac{e^{ia_j t} - e^{-ia_j t}}{2it}$, so that we get

$$f_N(x) = \frac{1}{2\pi} \left(\frac{1}{2i} \right)^N \left(\prod_{j=1}^{N} a_j^{-1} \right) \int_{-\infty}^{\infty} t^{-N} e^{ity} \prod_{j=1}^{N} \left(e^{ia_j t} - e^{-ia_j t} \right) dt \tag{A.86}$$

and the product of exponentials is expanded so that the result is

$$f_N(x) = \frac{1}{2\pi}\left(\frac{1}{2i}\right)^N \left(\prod_{j=1}^{N} a_j^{-1}\right)\left[\sum_{\varepsilon\in\{-1,1\}^N}\left(\prod_{j=1}^{N}\varepsilon_j\right)\int_{-\infty}^{\infty} t^{-N} e^{it(y+\vec{\varepsilon}\cdot\vec{a})}dt\right] . \tag{A.87}$$

The sum has to be carried out over each of the 2^N vectors of signs $\vec{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)$, with $\varepsilon_j = \pm 1$. Since Equation (A.86) is known to converge, we know that the singularities in Equation (A.87) must cancel, therefore the required finite integral is given by its principal value. Equation (A.87) can be integrated $N-1$ times by parts, yielding

$$\begin{aligned} f_N(x) \;=\; & \frac{1}{2\pi}\left(\frac{1}{2i}\right)^N \left(\prod_{j=1}^{N} a_j^{-1}\right)\frac{i^{N-1}}{(N-1)!}\left[\sum_{\varepsilon\in\{-1,1\}^N}\left(\prod_{j=1}^{N}\varepsilon_j\right)\right. \\ & \left.(y+\vec{\varepsilon}\cdot\vec{a})^{N-1}\cdot \mathrm{P.V.}\int_{-\infty}^{\infty} t^{-1}e^{it(y+\vec{\varepsilon}\cdot\vec{a})}dt\right] . \end{aligned} \tag{A.88}$$

The last integral evaluates to [21]

$$\mathrm{P.V.}\int_{-\infty}^{\infty} t^{-1}e^{it(y+\vec{\varepsilon}\cdot\vec{a})}dt = i\pi\,\mathrm{sign}(y+\vec{\varepsilon}\cdot\vec{a}) \ , \tag{A.89}$$

so that the final result is given by

$$f_N(x) = \frac{1}{(N-1)!\,2^{N+1}}\left(\prod_{j=1}^{N} a_j\right)^{-1}\left[\sum_{\vec{\varepsilon}\in\{-1,1\}^N}\Big(z_{\vec{\varepsilon}}(x)\Big)^{N-1}\mathrm{sign}(z_{\vec{\varepsilon}}(x))\prod_{j=1}^{N}\varepsilon_j\right] \tag{A.90}$$

with

$$z_{\vec{\varepsilon}}(x) = x + \sum_{j=1}^{N}(\varepsilon_j a_j - c_j) \ . \tag{A.91}$$

A.4. Overproduction Factor

The 100% renewable generation scenario in Section 3.4, assumes that generation and consumtion are equal over the eight years of data. Losses in the storage for $\eta_{in} < 1$ or $\eta_{out} < 1$ therefore have to be compensated by overproduction. Here, the derivation of the appropriate overproduction factor is derived.

The surplus generation factor γ is determined from the requirement that the storage level reached after 8 years is equal to the initial storage level,

$$H(t = 8\mathrm{y}) = H(t = 0) \ , \tag{A.92}$$

and thus $H(t = 8\text{y}) - H(t = 0) = 0$. Starting from the mismatch

$$\Delta(t) = \tau \left[\gamma \left(a \frac{W_\tau(t)}{\langle W \rangle} + b \frac{S_\tau(t)}{\langle S \rangle} \right) + c \frac{F_\tau(t)}{\langle F \rangle} - \frac{L_\tau(t)}{\langle L \rangle} \right] \tag{A.93}$$

and the definition of the storage

$$H(t) = H(t-1) + \begin{cases} \eta_{\text{in}} \Delta(t) & \text{if } \Delta(t) \geq 0 \ , \\ \eta_{\text{out}}^{-1} \Delta(t) & \text{if } \Delta(t) < 0 \ , \end{cases} \tag{A.94}$$

we find, using the Heaviside function $\Theta(x)$,

$$\begin{aligned} 0 &= \eta_{in} \sum_t \Delta(t)\Theta(\Delta(t)) + \eta_{out}^{-1} \sum_t \Delta(t)\Theta(-\Delta(t)) \\ &= \eta_{in} \sum_t \Delta(t) + \left(\eta_{out}^{-1} - \eta_{in}\right) \sum_t \Delta(t)\Theta(-\Delta(t)) \ . \end{aligned} \tag{A.95}$$

Using the balancing

$$B(t) = \begin{cases} -\Delta(t) & \text{if } \Delta(t) < 0 \\ 0 & \text{if } \Delta(t) \geq 0 \end{cases} \ . \tag{A.96}$$

and inserting Equation (A.93), we find

$$0 = \eta_{in} \left(\gamma a + \gamma b + \gamma c - 1\right) - \frac{\eta_{out}^{-1} - \eta_{in}}{N_t} \sum_t B(t; a, b, c, \gamma) \ ,$$

where N_t is the number of timesteps. Since $a + b + c = 1$ by definition, γ is given by

$$\gamma = \frac{\left(\eta_{out}^{-1} - \eta_{in}\right)}{N_t \eta_{in}} \left[\sum_t B(t; a, b, c, \gamma) \right] + 1 \ . \tag{A.97}$$

We use an iterative approach, to approximate γ

$$\gamma_{j+1} = \frac{\left(\eta_{out}^{-1} - \eta_{in}\right)}{N_t \eta_{in}} \left[\sum_t B(t; a, b, c, \gamma_j) \right] + 1 \ . \tag{A.98}$$

For a simple function describing the load and generation, we can show that this expression converges. We consider the continuous functions $L(t) = \frac{2}{t_{max}} t$ and $G(t) = 1$ for the load and generation, that have both a mean value of one. The solution of for the continuous version of Equation (A.98) then has the solution

$$\gamma^* = \frac{2\left(1 - \sqrt{\frac{\eta_{in}}{\eta_{out}}}\right)}{1 - \frac{\eta_{in}}{\eta_{out}}} \tag{A.99}$$

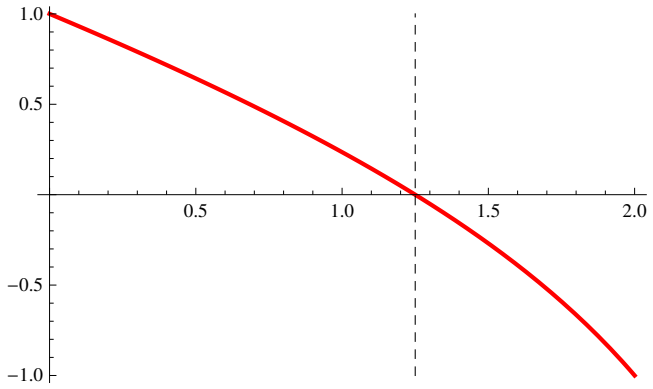


Figure A.3.: *Plot of γ_j against $\Delta\gamma_j$, the dashed line is γ^*.*

Plugging in the continuous load and generation into Equation (A.98), we find

$$\gamma_{j+1} = \frac{4 + \gamma_j^2 \left(1 - \frac{\eta_{in}}{\eta_{out}}\right)}{4 + 2\gamma_j \left(1 - \frac{\eta_{in}}{\eta_{out}}\right)} , \tag{A.100}$$

and the difference between γ_j and γ_{j+1} is given by

$$\Delta\gamma_j = \gamma_{j+1} - \gamma_j = \frac{4 + \gamma_j^2 \left(1 - \frac{\eta_{in}}{\eta_{out}}\right)}{4 + 2\gamma_j \left(1 - \frac{\eta_{in}}{\eta_{out}}\right)} - \gamma_j . \tag{A.101}$$

In Figure A.3, we plot $\Delta\gamma_j$ against γ_j in . As the plot shows $\Delta\gamma_j$ has a negative slope for all possible values for γ_j so that it always converges to the solution γ^*, where $\Delta\gamma_j$ is zero. We numerically verified that Equation (A.98) converges for the time series of the data in Chapter 3.

Bibliography

[1] J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler. The kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, 77:137–185, Apr. 2005. doi: 10.1103/RevModPhys.77.137.

[2] H. Akagi, E. H. Watanabe, and M. Aredes. *Instantaneous Power Theory and Applications to Power Conditioning*. Wiley InterScience, April 2007.

[3] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47–97, Jan 2002. doi: 10.1103/RevModPhys.74.47.

[4] R. Albert, H. Jeong, and A.-L. Barabaási. Error and attack tolerance of complex networks. *Nature*, 2000.

[5] R. Albert, I. Albert, and G. L. Nakarado. Structural vulnerability of the north american power grid. *Phys. Rev. E*, 69(2):025103, Feb 2004. doi: 10.1103/PhysRevE.69.025103.

[6] J. A. Almendral and A. Díaz-Guilera. Dynamical and spectral properties of complex networks. *New Journal of Physics*, 9(6):187, 2007.

[7] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America*, 97(21):11149–11152, 2000.

[8] M. Ángeles Serrano and M. Boguñá. Tuning clustering in random networks with arbitrary degree distributions. *Phys. Rev. E*, 72(3):036133, Sep 2005. doi: 10.1103/PhysRevE.72.036133.

[9] M. Anghel, K. Werley, and A. E. Motter. Stochastic model for power grid dynamics. 2007.

[10] A. Arenas, A. Diaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou. Synchronization in complex networks. *Physics Reports*, 469(3):93 – 153, 2008. ISSN 0370-1573. doi: DOI:10.1016/j.physrep.2008.09.002.

[11] F. Ashkar, B. Bobée, D. Leroux, and D. Morisette. The generalized method of moments as applied to the generalized gamma distribution. *Stochastic Hydrology and Hydraulics*, 2(3):161–174, September 1988.

[12] S. Balay, V. Eijkhout, W. D. Gropp, L. C. McInnes, and B. F. Smith. Efficient management of parallelism in object oriented numerical software libraries. In E. Arge, A. M. Bruaset, and H. P. Langtangen, editors, *Modern Software Tools in Scientific Computing*, pages 163–202. Birkhäuser Press, 1997.

[13] S. Balay, K. Buschelman, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang. Petsc users manual. Technical Report ANL-95/11 - Revision 2.1.5, Argonne National Laboratory, 2004.

[14] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509 – 512, 1999.

[15] A.-L. Barabási, M.A. de Menezes, S. Balensiefer, and J. Brockman. Hot spots and universality in network dynamics. *Eur. Phys. J. B*, 38(2):169–175, 2004.

[16] H. Beyer, G. Heilscher, and S. Bofinger. A robust model for the mpp performance of different types of pv-modules applied for the performance check of grid connected systems. *EUROSUN (ISES Europe Solar Congress) Freiburg*, 2004.

[17] T. Biyikoglu, J. Leydold, and P. F. Stadler. *Laplacian Eigenvectors of Graphs*. Springer, Berlin, Oktober 2007.

[18] B. Bollobás. The diameter of random graphs. *Transactions of the American Mathematical Society*, 267(1):41–52, 1981.

[19] B. Bollobás. *Modern Graph Theory*. Springer, 1998.

[20] Boost. http://www.boost.org/.

[21] D. M. Bradley and R. C. Gupta. On the distribution of the sum of n non-identically distributed uniform random variables. *Annals of the Institute of Statistical Mathematics*, 54(3):689–700, 2002.

[22] Bronstein, Semendjajew, Musiol, and Mühlig. *Taschenbuch der Mathematik*. Verlag Harri Deutsch, 2001.

[23] P. N. Brown and Y. Saad. Hybrid krylov methods for nonlinear systems of equations. *SIAM SIAM J. SCI. STAT. COMPUT.*, 11(3):450, 1990.

[24] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. 10:186–198, March 2009.

[25] L. Cai, I. Erlich, and G. Stamtsis. Optimal choice and allocation of facts devices in deregulated electricity market using genetic algorithms. pages 201 – 207 vol.1, oct. 2004. doi: 10.1109/PSCE.2004.1397562.

[26] B. Carreras, V. Lynch, M. L. Sachtjen, I. Dobson, and D. E. Newman. Modelling blackout dynamics in power trasmission networks with simple structure. *Hawaii International Conference on System Sciences, Maui, Hawaii.*, 2001.

[27] B. Carreras, D. E. Newman, I. Dobson, and A. B. Poole. Evidence for self-organized criticallity in electric power system blackouts. *Hawaii International Conference on System Sciences, Maui, Hawaii.*, 2001.

[28] B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman. Critical points and transitions in an electric power transmission model for cascading failure blackouts. *Chaos*, 12(4):985–994, Dec 2002.

[29] A. Cho. Energy's Tricky Tradeoffs. *Science*, 329(5993):786–787, 2010. doi: 10.1126/science.329.5993.786.

[30] D. Clery. Sending African Sunlight to Europe, Special Delivery. *Science*, 329 (5993):782–783, 2010.

[31] A. C. Cohen and B. J. Whitten. Estimation in the three - parameter lognormal distribution. *Journal of the American Statistical Association,*, 75:399–404, 2007.

[32] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.*, 85(21):4626–4628, Nov 2000. doi: 10.1103/PhysRevLett.85.4626.

[33] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Breakdown of the internet under intentional attack. *Phys. Rev. Lett.*, 86, 2001.

[34] L. Comtet. *Advanced Combinatorics*. D. Reidel Publishing Company, 1974.

[35] V. Crastan. *Elektrische Energieversorgung 1*. Springer Berlin Heidelberg, 2. edition, 2007.

[36] F. Crotogino and R. Hamelmann. Wasserstoff-Speicherung in Salzkavernen. In *Tagungsband "14. Symposium zur Nutzung regenerativer Energiequellen und Wasserstofftechnik*, Nov 2007.

[37] P. Crucitti, V. Latora, M. Marchiori, and A. Rapisarda. Efficiency of scale-free networks: error and attack tolerance. *Physica A: Statistical Mechanics and its Applications*, 320:622–642, 2003. ISSN 0378-4371.

[38] J. H. Curtiss. On the distribution of the quotient of two chance variables. *The Annals of Mathematical Statistics*, 12(4):409–421, 1941. ISSN 00034851.

[39] J. Dall and M. Christensen. Random geometric graphs. *PHYSICAL REVIEW E*, 66, 2002.

[40] M. A. de Menezes and A.-L. Barabási. Fluctuations in network dynamics. *Phys. Rev. Lett.*, 92(2):028701, Jan 2004. doi: 10.1103/PhysRevLett.92.028701.

[41] DENA. Energiewirtschaftliche Planung für die Netzintegration von Windenergie in Deutschland an Land und Offshore bis zum Jahr 2020. *Köln*, 2 2005.

[42] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. 1983.

[43] I. Dobson, B. Carreras, V. Lynch, and D. E. Newman. An initial model for complex dynamics in electric power system blackouts. *Hawaii International Conference on System Sciences, Maui, Hawaii.*, 2001.

[44] S. Donadei and F. Crotogino. Energy storage in salt caverns – today and tomorrow. In *9TH WORLD SALT SYMPOSIUM, Beijing, China, September 4 to 9th,*, 2009.

[45] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. pages 1079–1187, 2002.

[46] J. C. Doyle, D. L. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger. The "robust yet fragile" nature of the Internet. *Proceedings of the National Academy of Sciences of the United States of America*, 102(41): 14497–14502, 2005.

[47] P. G. Doyle and J. L. Snell. Random walks and electric networks. 2007. arXiv:math/0001057v1.

[48] S. C. Eisentat and H. F. Walker. Choosing the forcing terms in an inexact newton method. *SIAM J. Sci. Stat. Comput.*, 17:16, 1996.

[49] A. Erdõs and Rény. On random graphs. *Publ. Math. Debrecen*, 6:290–297, 1959.

[50] European Climate Foundation. Roadmap 2050. Technical report, 2010. URL `www.roadmap2050.eu`.

[51] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. pages 251–262, 1999.

[52] R. Fetea and A. Petroianu. Reactive power: A strange concept? *Second European Conference on Physics Teaching In Engineering Education*, 2000.

[53] G. Filatrella, A. Hejde Nielsen, and N. Falsig Pedersen. Analysis of a power grid using a kuramoto-like model. *EPJ B*, 61(4), 2008. doi: 10.1140/epjb/e2008-00098-8.

[54] T. Fließbach. *Statistische Physik*. Elsevier, 4. edition, 2007.

[55] E. M. Gertz and S. J. Wright. Object-oriented software for quadratic programming. *ACM Transactions on Mathematical Software*, 29:58–81, 2003.

[56] A. Gillhaus. Natural gas storage in salt caverns - present status, developments and future trends in europe. In *SMRI Spring Meeting 2007, Basel*, 2007.

[57] GNU Scientific Library. http://www.gnu.org/software/gsl/.

[58] K.-I. Goh, B. Kahng, and D. Kim. Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.*, 87, 2001.

[59] L. L. Grigsby, editor. *Power System Stability and Control*. CRC Press Inc, 2006.

[60] R. W. Grimes and W. J. Nuttall. Generating the Option of a Two-Stage Nuclear Renaissance. *Science*, 329(5993):799–803, 2010. doi: 10.1126/science.1188928.

[61] F. Hannig, T. Smolinka, P. Bretschneider, S. Nicolai, S. Krüger, F. Meißner, and M. Voigt. Stand und Entwicklungspotenzial der Speichertechniken für Elektroenergie – Ableitung von Anforderungen an und Auswirkungen auf die Investitionsgüterindustrie. Bmwi-auftragsstudie 08/28, Bundesministeriums für Wirtschaft und Technologie, 2009. URL `http://bmwi.de/BMWi/Redaktion/PDF/Publikationen/Studien/speichertechniken-elektroenergie.pdf`.

[62] J. Hayya, D. Armstrong, and N. Gressis. A note on the ratio of two normally distributed variables. *Management Science*, 21(11):1338–1341, 1975. ISSN 00251909.

[63] D. Heide, M. Schäfer, and M. Greiner. Robustness of networks against fluctuation-induced cascading failures. *Phys Rev E*, 2008.

[64] D. Heide, L. von Bremen, M. Greiner, C. Hoffmann, M. Speckmann, and S. Bofinger. Seasonal optimal mix of wind and solar power in a future, highly renewable europe. *Renewable Energy*, 35(11):2483 – 2489, 2010. ISSN 0960-1481. doi: DOI:10.1016/j.renene.2010.03.012.

[65] D. V. Hinkley. On the ratio of two correlated normal random variables. *Biometrika*, 56(3):pp. 635–639, 1969. ISSN 00063444.

[66] M. Hisakado and S. Kitsukawa, Kenji an Mori. Correlated binomial models and correlation structures. *arXiv*, 2006.

[67] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han. Attack vulnerability of complex networks. *Phys. Rev. E*, 65(5):056109, May 2002. doi: 10.1103/PhysRevE.65.056109.

[68] A. J. Holmgren. Using graph models to analyze the vulnerability of electric power networks. *Risk Anal*, 26(4):955–969, Aug 2006. doi: 10.1111/j.1539-6924.2006.00791.x.

[69] M. K. Hubbert. Nuclear Energy and the Fossil Fuels 'Drilling and Production Practice'. *Spring Meeting of the Southern District. Division of Production. American Petroleum Institute. San Antonio, Texas*, pages 22–27, 1956.

[70] E. V. Huntington. Frequency distribution of product and quotient. *The Annals of Mathematical Statistics*, 10(2):195–198, 1939. ISSN 00034851.

[71] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13:411–430, 2000.

[72] M. Jacobson and M. Delucchi. A path to sustainable energy by 2030. *Scientific American*, pages 58–65, Nov. 2009.

[73] F. P. Kelly. Network routing. *Philosophical Transactions: Physical Sciences and Engineering*, 337(1647):pp. 343–367, 1991. ISSN 09628428.

[74] R. A. Kerr. Do We Have the Energy for the Next Transition? *Science*, 329(5993): 780–781, 2010. doi: 10.1126/science.329.5993.780.

[75] D.-H. Kim and A. E. Motter. Fluctuation-driven capacity distribution in complex networks. *New Journal of Physics*, 10(5):053022, 2008.

[76] G. Knies. Global energy and climate security through solar power from deserts. Technical report, Trans-Mediterranean Renewable Energy Cooperation in cooperation with The Club of Rome, 2006. URL `http://www.desertec.org/downloads/deserts_en.pdf`.

[77] U. Krengel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. 5. edition, 2000.

[78] D. Krioukov, F. Chung, kc claffy, M. Fomenkov, A. Vespignani, and W. Willinger. The workshop on internet topology (wit) report. *ACM SIGCOMM Computer Communication Review*, 37(1), January 2007.

[79] L. D. Landau and E. M. Lifshitz. *Statistical Physics Part 1*, volume 5 of *Course of theoretical phyaics*. PERGAMON PRESS, 3rd edition, 1980.

[80] F. C. Leone, L. S. Nelson, and R. B. Nottingham. The folded normal distribution. *Technometrics*, 3(4):pp. 543–550, 1961. ISSN 00401706.

[81] X. Lu, M. McElroy, and J. Kiviluoma. Global potential for wind-generated electricity. *Proc. Natl. Acad. Sci.*, 106:10933–10938, 2009.

[82] G. Marsaglia. Ratios of normal variables and ratios of sums of uniform variables. *Journal of the American Statistical Association*, 60(309):193–204, 1965. ISSN 01621459.

[83] T. McNichol. *AC/DC: The savage tale of the first standards war*. Jossey-Bass, San Francisco, CA, 2006.

[84] G. A. Meehl, W. M. Washington, C. M. Ammann, J. M. Arblaster, T. M. L. Wigley, and C. Tebaldi. Combinations of natural and anthropogenic forcings in twentieth-century climate. *Journal of Climate*, 17:3721–3727, 2004.

[85] A. E. Motter and Y.-C. Lai. Cascade-based attacks on complex networks. *Phys. Rev. E*, 66(6):065102, Dec 2002. doi: 10.1103/PhysRevE.66.065102.

[86] NCEP. http://www.ncep.noaa.gov. National Center for Environmental Prediction.

[87] M. E. J. Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Phys. Rev. E*, 64(1):016132, Jun 2001.

[88] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003. doi: 10.1137/S003614450342480.

[89] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2):026113, Feb 2004. doi: 10.1103/PhysRevE.69.026113.

[90] D. Oeding and B. Oswald. *Elektrische Kraftwerke und Netze*. Springer, 6. edition, 2004.

[91] B. Oswald. *Berechnung von Drehstromnetzen*. Vieweg+Teubner, 2009.

[92] M. Patra and M. Karttunen. Stencils with isotropic discretisation error for different operators. *Numerical Methods for Partial Differential Equations*, 22, 2005. doi: 10.1002/num.20129.

[93] R. Penrose. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, pages 406–413, 1955.

[94] R. Penrose and J. A. Todd. On best approximate solutions of linear matrix equations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 52: 17–19, 1956.

[95] R. L. Prentice. A log gamma model and its maximum likelihood estimation. *Biometrika*, 61(3):539–544, 1974.

[96] V. Quaschning. Systemtechnik einer klimaverträglichen Elektrizitätsversorgung in Deutschland für das 21. Jahrhundert. *Fortschritt Berichte VDI, Reihe 6, Nr. 437, VDI Verlag, Düsseldorf*, 2000.

[97] L. E. Reichl. *A Modem Course in Statistical Physics*. John Wiley & Sons, Inc., 2nd edition, 1998.

[98] REN21. Renewables global status report 2007. `http://www.ren21.net/globalstatusreport/g2007.asp`, 2006.

[99] REN21. Renewables global status report 2009. `http://www.ren21.net/globalstatusreport/g2009.asp`, 2009.

[100] M. Romeo, V. Da Costa, and F. Bardou. Broad distribution effects in sums of lognormal random variables. *The European Physical Journal B*, 32:513–525, 2003.

[101] E. A. Rosa, S. P. Tuler, B. Fischhoff, T. Webler, S. M. Friedman, R. E. Sclove, K. Shrader-Frechette, M. R. English, R. E. Kasperson, R. L. Goble, T. M. Leschine, W. Freudenburg, C. Chess, C. Perrow, K. Erikson, and J. F. Short. Nuclear Waste: Knowledge Waste? *Science*, 329(5993):762–763, 2010. doi: 10.1126/science.1193205.

[102] M. Rosas-Casals, S. Valverde, and R. Solé. Topological vulnerability of the european power grid under errors and attacks. *International Journal of Bifurcation and Chaos*, 17(7):2465–2475, 2007.

[103] Y. Saad. *Iterative Methods for Sparse Linear Systems*. 2000.

[104] M. Schäfer, J. Scholz, and M. Greiner. Proactive robustness control of heterogeneously loaded networks. *PRL*, 96, 2006.

[105] J. Scholz. *Unknown*. PhD thesis, Frankfurt Institute for Advanced Studies, 2010.

[106] J. Scholz, M. Dejori, M. Stetter, and M. Greiner. Noisy scale-free networks. *Physica A*, 350:622–642, 2005.

[107] A. J. Schwab. *Elektroenergiesysteme*. Springer, 2. edition, 2009.

[108] J. N. Shadid, R. S. Tuminaro, and H. F. Walker. An inexact newton method for fully coupled solution of the navier stokes equations with heat and mass transport. *Journal of Computational Physics*, 137:155, 1997.

[109] I. Simonsen, L. Buzna, K. Peters, S. Bornholdt, and D. Helbing. Transient dynamics increasing network vulnerability to cascading failures. *Physical Review Letters*, 100(21):218701, 2008. doi: 10.1103/PhysRevLett.100.218701.

[110] Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, editors. *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, 2007.

[111] E. W. Stacy. A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, pages 1187–1192, 1962.

[112] E. W. Stacy and G. A. Mihram. Parameter estimation for a generalized gamma distribution. *Technometrics*, 7(3):349–358, 1965.

[113] S. H. Strogatz. *Nonlinear Dynamics and Choas*. Perseus Books Publishin, LLC, Dec. 2000.

[114] J. M. Stuart, E. Segal, D. Koller, and S. K. Kim. A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules. *Science*, 302(5643):249–255, 2003. doi: 10.1126/science.1087447.

[115] P. Tetali. Random walks and the effective resistance of networks. *Journal of Theoretical Probability*, 4:101–109, 1991. ISSN 0894-9840.

[116] M. Timme, F. Wolf, and T. Geisel. Topological speed limits to network synchronization. *Phys Rev Lett*, 92(7):074101, Feb 2004.

[117] UCTE. http://www.ucte.org/resources/dataportal/consumption.

[118] U.S. Energy Information Administration. International energy outlook 2009. http://www.eia.doe.gov/oiaf/ieo/pdf/electricity.pdf, 2009.

[119] U.S. Energy Information Administration. International electricity generation, all countries, 1980-2006. `http://www.eia.doe.gov/emeu/international/electricitygeneration.html`, 2010. downloaded May 2010.

[120] VDMA. Strommix in der EU27, March 2010. URL `http://www.vdma.org/`.

[121] U. von Luxburg. A tutorial on spectral clustering. *Stat Comput*, 17:395–416, 2007.

[122] B. Wang and B. J. Kim. A high-robustness and low-cost model for cascading failures. *EPL (Europhysics Letters)*, 78(4):48001, 2007.

[123] D. J. Watts and S. H. Strogatz. Collective dynamics of "small-world" networks. *Nature*, 393:440–442, 1998.

[124] WEPROG. http://www.weprog.com.

[125] M. B. Wilk and R. Gnanadesikan. Probability plotting methods for the analysis of data. *Biometrika*, 55(1):pp. 1–17, 1968. ISSN 00063444.

[126] A. J. Wood and B. F. Wollenberg. *Power Generation, Operation, and Control*. John Wiley & Sons, New York, 1996.

[127] W. Xiao and I. Gutman. Resistance distance and laplacian spectrum. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling*, 110(4), 2003.

[128] L. Zhao, K. Park, and Y.-C. Lai. Attack vulnerability of scale-free networks due to cascading breakdown. *Phys. Rev. E*, 70(3):035101, Sep 2004.

[129] W. Zittel and J. Schindler. Uranium resources and nuclear energy. Technical report, Energy Watch Group, 2006.

[130] W. Zittel and J. Schindler. Coal: Resources and future production. Technical report, Energy Watch Group, 2007.