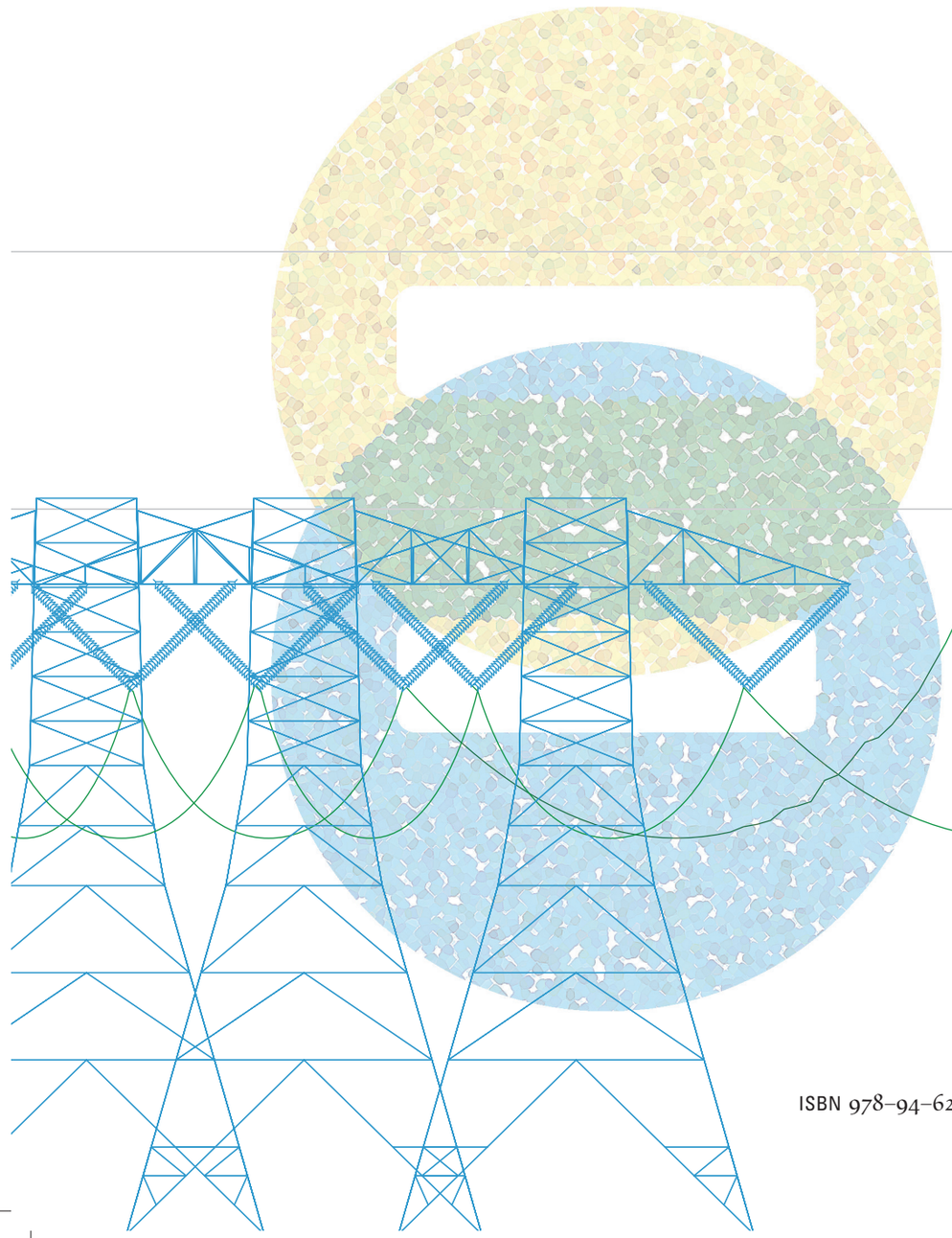
Assessing power grid reliability using rare event simulation



Assessing power grid reliability using rare event simulation

ISBN 978-94-6299-059-3

Wadman

Wander Wadman

WANDER WADMAN

ASSESSING POWER GRID RELIABILITY USING RARE EVENT SIMULATION

The research described in this thesis was financially supported by the Netherlands Organisation for Scientific Research (NWO) and Centrum Wiskunde & Informatica (CWI), the Dutch national research institute for mathematics and computer science. The research was performed in collaboration with DNV GL - Energy.

Cover by Joran A. Kuijper

Printing: Ridderprint BV

ASSESSING POWER GRID RELIABILITY USING RARE EVENT SIMULATION

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. D.C. van den Boom

ten overstaan van een door het College voor Promoties ingestelde

commissie, in het openbaar te verdedigen in de Agnietenkapel

op donderdag 18 juni 2015, te 10 uur

door Wander Sybe Wadman

geboren te Utrecht

Promotiecommissie:

Promotor:	prof. dr. D.T. Crommelin	Universiteit van Amsterdam
Copromotor:	prof. dr. ir. J.E. Frank	Universiteit Utrecht
Overige leden:	prof. dr. M.R.H. Mandjes	Universiteit van Amsterdam
	prof. dr. J.H. van Zanten	Universiteit van Amsterdam
	prof. dr. R. Núñez Queija	Universiteit van Amsterdam
	dr. A.A.N. Ridder	Vrije Universiteit
	prof. dr. A.P. Zwart	Technische Universiteit Eindhoven
	prof. dr. ir. J.A. La Poutré	Technische Universiteit Delft
Faculteit:	Faculteit der Natuurwetenschappen, Wiskunde en Informatica	

Voor Nora

ACKNOWLEDGMENTS

This dissertation is the result of four years of research at CWI Amsterdam. During those years I enjoyed many intensive discussions, exchanged thoughts on international platforms and came into close contact with the power industry. Many people were involved in making this experience so memorable.

First, many thanks go to my promotor Daan Crommelin for the way he introduced me to the academic world and how he coached me over the past years. Despite his many other duties, he was always willing to discuss the next hurdle I encountered. In fact, it was often Daan walking into my room to suggest a paper or a promising next step in my research. Even a small question often ended up in an hour long discussion next to his whiteboard.

Also many thanks go to my copromotor Jason Frank for his willingness to give me advice at any time. I learned a lot from him on numerical methods, academic writing and the scientific community in general.

This research was performed in collaboration with DNV GL Energy. Gabriël Bloemhof is gratefully acknowledged for sharing his knowledge on power grids, for linking me to experts at DNV GL and for his help choosing research topics that will make a difference. In every discussion we had I learned something new about the fascinating world of electricity.

Special thanks go to Bert Zwart for introducing me to the research area of rare event simulation. I gratefully benefited from his recommendations on relevant conferences and on viable next steps in my research. I also like to express my gratitude to Kees Oosterlee, who mentioned this PhD position to me. Sometimes small hints can make a world of difference for one's career.

Keith, thank you for your everlasting curiosity for all those technical problems I encountered and for tolerating me as a grumpy roommate from time to time. Bart, Debarati, Halldora, Janis, Jesse, Nick, Zsolt, thanks for all the passionate discussions we had during lunch and dinner.

I thank my PhD colleagues Anton, Benjamin, Bram, Christoph, Daan, Eleni, Emmanuelle, Folkert, Gebre, Joost, Lech, Linda, Lisanne, Marjon, Marwin, Nicolas, Qian, Shashi, Simon, Sonja, Thibault, Tom, Willem and Yagiz for creating such a great atmosphere at CWI.

The kind help of the CWI staff — especially Bikkie, Hans, Léon, Minnie and Nada — is very much appreciated. Joran, thank you for putting the finishing touch to my dissertation, and in particular for including the AMS Euler font.

Aan mijn lieve ouders, zus en oma: dank voor jullie rotsvaste vertrouwen in de goede afloop. En voor al jullie wijze woorden en de geruststellende toon waarmee jullie deze altijd over weten te brengen.

Als laatste Nora: van alle persoonlijke steun kwam die van jou uiteraard van het dichtstbij. Dank je wel lieve Noor, voor die ongelooflijke hoeveelheden optimisme, hulp en liefde. Laten we er een mooi feest van maken het komende jaar!

SUMMARY

ASSESSING POWER GRID RELIABILITY USING RARE EVENT SIMULATION

Renewable energy generators such as wind turbines and solar panels supply more and more power in modern electrical grids. Although the transition to a sustainable power supply is desirable, considerable implementation of distributed and intermittent generators may strain the power grid.

First, most renewable sources have an uncertain generation pattern: predicting the amount of generation of a wind turbine or solar panel can be very challenging. Therefore, implementation of many renewable energy generators in a power grid increases the risks of connection overloads and voltage deviations. Grid operators may have to curtail power to prevent violation of grid stability constraints. These events are obviously undesirable, especially as modern societies have grown accustomed to a very reliable power grid.

Second, since many renewable generators are distributed over the power grid, the generation of power is decentralized instead of operating in a 'top-down' fashion. For this reason it is in general not clear what typical combinations of generation and consumption patterns — or stated more general, power injection patterns — will lead to a violation of stability constraints.

Since grid operators are responsible for a highly reliable power grid, they want to estimate to what extent these constraint violations occur. To assess grid reliability over a period of interest, various reliability indices exist. The main challenge of this research is to develop reliability assessment methods for a power grid given the uncertainty of power injections.

Deterministic methods for power flow analyses are well-established in the power system community. Many power flow analyses employ the steady state AC power flow equations. These equations form a

nonlinear algebraic system relating the power injections at all nodes to the nodal voltages and other variables that determine if the grid is stable. We model the power injections as stochastic processes to account for their uncertainty. At each time step the mapping from the state of these processes to the occurrence of a constraint violation requires solving the nonlinear system. Since this solution is only defined implicitly, we can not derive the reliability indices directly and we estimate them by a Monte Carlo simulation.

We illustrate the feasibility of Monte Carlo reliability estimation of a stochastic power flow model with wind power modeled as an autoregressive-moving-average (ARMA) model. Instead of enforcing current constraints on connections, short-term current overloading is allowed by enforcing the more realistic temperature constraints. We show that — especially when intermittent power constitutes a significant part of the power supply — a model allowing for temporary current overloading may save costs by avoiding over-investments.

Power curtailments are typically rare in modern power grids. Using conventional Monte Carlo or Crude Monte Carlo (CMC) simulation for grid reliability estimation may therefore require a prohibitively large number of samples to achieve a sufficient level of accuracy. Over the last decades, rare event simulation techniques have been developed to reduce the variance of estimators for very small probabilities. This research area is an established methodology in molecular biology, telecommunications and finance, but in power systems applications have emerged only recently and in limited number. Splitting and importance sampling are the two main categories of rare event simulation. In this thesis we extend a Crude Monte Carlo (CMC) method with a splitting technique to efficiently compute unbiased estimators for several indices.

Splitting techniques replicate sample paths whenever the rare event is presumed considerably more likely given the current chain state. Crucial for a significant variance reduction of the estimator is choosing a suitable importance function. This function ideally maps each system state to the probability that a sample path starting from that state will hit the rare event set. In an example of a small power grid, splitting estimation with

a heuristically chosen importance function already requires orders of magnitude less workload than CMC estimation to obtain a fixed accuracy.

However, a heuristically chosen importance function may replicate sample paths that are relatively unlikely to hit the rare event set, especially in case of many stochastic processes. In this case the variance of the estimator is hardly reduced — it may in fact have become larger than the variance of a CMC estimator. In power grid reliability analyses with many distributed uncertain power injections this is a substantial problem as it is not clear a priori what multidimensional sample path will typically lead to rare event occurrences. Specifically, we address this problem of heuristically choosing an importance function for a disconnected rare event set.

Asymptotic results from large deviations theory give an insight into the typical path to the rare event set. These results also yield an approximation of the rare event probability starting from a specified system state. We construct an importance function based on these results to compute splitting estimates for connection overload probabilities in a power grid. Experiments show that for a fixed accuracy the splitting technique based on large deviations computationally outperforms CMC estimators by orders of magnitude. The computational gain remains of similar size even for a larger number of stochastic processes.

In an example we compare the performance of this importance function to that of a heuristically chosen importance function based on the Euclidean distance to the rare event set. This example showed that heuristic splitting — unlike large deviations based splitting — requires even more workload than CMC simulation to obtain a fixed accuracy of probability estimates. This clearly illustrates that accelerating Monte Carlo simulation by using a splitting technique requires a carefully chosen importance function.

SAMENVATTING

BETROUWBAARHEIDSSCHATTINGEN VAN ELEKTRICITEITSNETWERKEN DOOR MIDDEL VAN 'RARE EVENT SIMULATION'

Duurzame energiebronnen zoals windturbines en zonnepanelen voeden in steeds grotere mate hedendaagse elektriciteitsnetwerken. Hoewel de transitie naar een duurzame energievoorziening wenselijk is, kunnen aanzienlijke hoeveelheden decentrale, onzekere opwekking de betrouwbaarheid van het elektriciteitsnetwerk aantasten.

Ten eerste vormt de onzekere hoeveelheid opwekking een risico: windturbines en zonnepanelen wekken een hoeveelheid energie op die vaak moeilijk te voorspellen is. Een aanzienlijke bijdrage van weersafhankelijke generatoren aan de totale energievoorziening zal daarom het risico verhogen dat leidingen — ondergrondse kabels of bovengrondse lijnen — overbelast raken of dat voltages op aansluitingen teveel afwijken van het nominale spanningsniveau. Netbeheerders moeten soms de energietoevoer onderbreken zodat het netwerk aan stabiliteitsvoorwaardes blijft voldoen. Deze onderbrekingen zijn uiteraard onwenselijk, te meer omdat moderne samenlevingen hoge eisen stellen aan netwerkbetrouwbaarheid.

Ten tweede ondergaat een elektriciteitsnetwerk met een groeiend aantal duurzame generatoren een transitie van een centrale energievoorziening naar een decentrale voorziening. Hierdoor is het op voorhand onduidelijk welke combinaties van opwekkings- en consumptiepatronen — of in het algemeen vermogensinjectiepatronen — de grootste risico's vormen voor een overschrijding van stabiliteitsvoorwaardes.

Aangezien netbeheerders verantwoordelijk zijn voor een hoge netwerkbetrouwbaarheid, is het in hun belang te weten in welke mate genoemde stabiliteitsvoorwaardes overschreden zullen worden. Er bestaan verschillende indicatoren voor netwerkbetrouwbaarheid. Het doel van dit promotieonderzoek is het ontwikkelen van rekenmethodes ter

beoordeling van de netwerkbetrouwbaarheid gegeven de onzekerheid van alle vermogensinjecties.

Deterministische methodes voor netberekeningen zijn al langere tijd gemeengoed in de elektriciteitswereld. Veel analyses van netwerken nemen de *AC power flow* evenwichtsvergelijkingen aan. Dit niet-lineaire stelsel van vergelijkingen beschrijft de relatie tussen enerzijds de vermogensinjecties en anderzijds de voltages op knooppunten en andere variabelen die bepalen of het netwerk stabiel is. Om de onzekerheid van vermogensinjecties mee te nemen, modelleren we deze als (tijdsdiscretisaties van) stochastische processen. Om te bepalen of een toestand van dit proces op een zeker tijdstip een overschrijding van een stabiliteitsvoorwaarde impliceert, dient het niet-lineaire stelsel opgelost te worden. Aangezien deze oplossing enkel impliciet bekend is, zijn expliciete oplossingen voor de betrouwbaarheidsindicatoren niet beschikbaar. We schatten deze indicatoren daarom door middel van een Monte Carlo simulatie.

We gebruiken een Monte Carlo simulatie om de netwerkbetrouwbaarheid te schatten waarbij een *autoregressive-moving-average* (ARMA) model de onzekere windopwekkingen voorstelt. Als stabiliteitsvoorwaarde bij iedere leiding hanteren we een maximum temperatuur in plaats van een maximum stroomsterkte. We laten zien dat een model dat tijdelijke overbelastingen toelaat overinvesteringen in leidingen en dus onnodige kosten kan voorkomen. Deze besparing is met name aanzienlijk als onzekere vermogensinjecties een groot deel van het totale vermogen leveren.

Vermogensonderbrekingen zijn zeldzame gebeurtenissen in moderne elektriciteitsnetwerken. Conventionele Monte Carlo of Crude Monte Carlo (CMC) simulatie van netwerkbetrouwbaarheid vereist daarom geregeld een onpraktisch groot aantal trekkeningen om een geschikt nauwkeurigheidsniveau te bereiken. Het vakgebied *rare event simulation* is de laatste decennia ontwikkeld om de variantie te verkleinen van Monte Carlo schatters voor zeer kleine kansen. Het vakgebied heeft al veel ontwikkeling ondergaan in toepassingsgebieden als moleculaire biologie, telecommunicatie en financiële wiskunde. Toepassingen in elektriciteitsnetwerken zijn echter nog weinig voorkomend en van recente datum. Splitting en importance sampling zijn de twee voornaamste

categorieën van *rare event simulation*. In dit promotieonderzoek breiden we een CMC simulatie uit met een splitting methode om zuivere schatters voor verschillende betrouwbaarheidsindicatoren efficient te berekenen.

Een splitting methode splitst gerealiseerde paden op in meerdere kopieën zodra vanuit de huidige toestand de *rare event* een beduidend hogere kans tot voltrekken heeft. Een geschikte keuze voor de zogenaamde importance functie is cruciaal om de variantie van de schatter aanzienlijk te verminderen. Idealiter beeldt deze functie elke toestand van het stochastische process af op de kans dat de *rare event* zich zal voordoen gegeven deze toestand. In een experiment op een klein elektriciteitsnetwerk vereist splitting met een heuristische keuze voor de importance functie ordegroottes minder rekenkracht dan CMC om een gewenste nauwkeurigheid te bereiken.

Echter, een heuristisch gekozen importance functie zou paden kunnen opsplitsen van waaruit een voltrekking van de *rare event* in werkelijkheid relatief onwaarschijnlijk is, met name in modellen met veel stochastische processen. In dit geval nemen *rare event* gebeurtenissen nauwelijks toe en is de variantiereductie miniem — de variantie van de splitting schatter zou zelfs groter kunnen zijn dan die van een CMC schatter. Dit vormt een probleem in betrouwbaarheidsanalyses van netwerken met veel decentrale onzekere opwekking aangezien het op voorhand niet duidelijk is welke combinaties van injectiepaden de *rare event* voornamelijk veroorzaken. We behandelen dit probleem van een heuristisch gekozen importance functie in het specifieke geval dat de *rare event set* uit meerdere onsamenhangende deelverzamelingen bestaat.

Asymptotische resultaten van *large deviations theory* geven inzicht in de meest kansrijke paden die tot de *rare event* leiden. Deze resultaten leveren bovendien een benadering op van de kans op de *rare event* gegeven een zekere begintoestand. We baseren een keuze voor de importance functie op deze resultaten en berekenen hiermee splitting schattingen voor de kansen op overbelastingen van leidingen in een elektriciteitsnetwerk. Uit experimenten blijkt dat splitting gebaseerd op *large deviations theory* ordegroottes minder rekenkracht vereist voor een gewenste nauwkeurigheid dan CMC simulatie. Zelfs voor een groter aantal stochastische processen is de rekenkundige winst van vergelijkbare grootte.

In een voorbeeld vergelijken we deze importance functie met een heuristisch gekozen importance functie die gebaseerd is op de Euclidische afstand tot de *rare event set*. Uit dit experiment blijkt dat de heuristische splitting methode — in tegenstelling tot de op *large deviations theory* gebaseerde splitting methode — nog meer rekenkracht vergt voor een gewenste nauwkeurigheid dan CMC simulatie. Dit weerspiegelt de noodzaak van een zorgvuldig gekozen importance functie voor het versnellen van Monte Carlo simulatie met een splitting methode.

CONTENTS

Part I

MOTIVATION OF RESEARCH

1

RENEWABLE ENERGY GENERATION: A CHALLENGE FOR GRID RELIABILITY

On 20 March 2015 a solar eclipse [. . .] will be visible across Europe. The reduction in solar radiation will directly affect the output of the photovoltaics (PV) and for the first time this is expected to have a relevant impact on the secure operation of the European power system.

— European network of transmission system operators for electricity [28]

The development of a sustainable and reliable electricity supply that fulfills the ever growing global demand is one of the biggest challenges of the 21st century. Fortunately, the installed capacity of many forms of renewable energy like wind, (concentrating) solar photovoltaics and geothermal power has increased exponentially since the start of the century [29, 69, 70][1]. In 2013 wind power accounted for 33% of the electricity demand in Denmark, and for 21% in Spain; in the same year photovoltaic systems delivered 7.8% of the total electricity demand in Italy [70].

As a result, modern electrical power grids are evolving: the model of a top-down energy supply is steadily transforming into a decentralized model in which the grid is substantially fed by distributed power generators.

Although this development is desirable, the intermittent nature of many renewable energy generators will form a challenge for the reliability of the power grid. For example, the amount of generated solar and wind power depends on meteorological conditions that are difficult to predict and

1 We especially recommend reading [70], a clear report on the global development of renewable energy using various statistics.

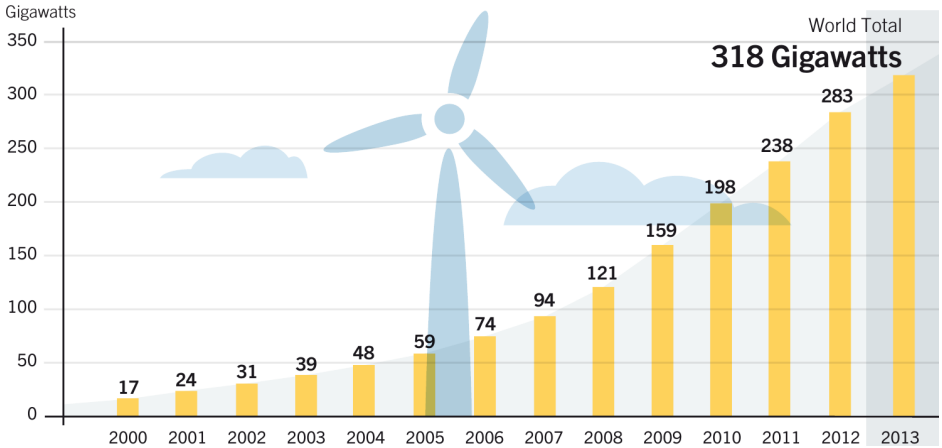


Figure 1.1: Historical data of global wind power capacity (GW). Source: [70]

impossible to control [63]. Different sources exhibit different time patterns and their peaks occur at different moments in time. Power imbalances induced by generation intermittency pose a threat to the stability of the grid [57, Chapter 26].

Grid operators are responsible for preserving grid stability. To prevent violation of stability constraints, the grid operator has to reschedule power generation or even to curtail power [7, 96]. The latter means that power demanded by consumers is not delivered or — in privatized energy markets — that power generation that was originally scheduled is repealed. Both forms of curtailments are obviously undesirable as modern societies have grown accustomed to a very reliable power grid [13].

It may be tempting to reason that any emergent risk of a grid instability induced by generation intermittency can easily be resolved by curtailing the power generated by the renewable source. However, this is only true when it is an *excess* of renewable power that causes a grid instability risk. Second, in privatized energy markets both generators and consumers are regarded as customers of the power grid [59, Chapter 2]. Besides curtailing power *consumption*, curtailing power *generation* is therefore undesirable too.

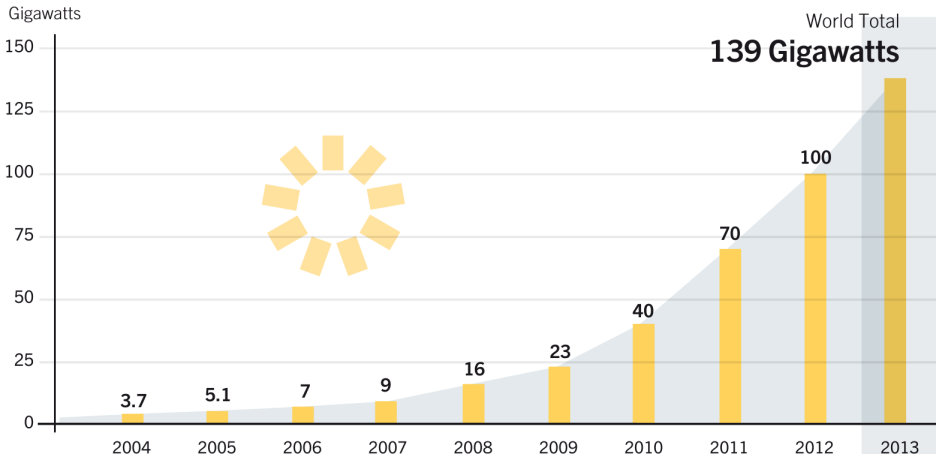


Figure 1.2: Historical data of global photovoltaics capacity (GW). Source: [70]

Power consumption may also be uncertain — although on higher level grids one can predict the consumption of many consumers aggregated at one node to some extent. Electric vehicles may aggravate the uncertainty of consumption in the near future given their large charging loads at uncertain times and locations [16]. In this thesis we therefore assume power generation as well as power consumption to be uncertain in general. Stated more generally, we assume uncertain *power injections*, where positive power injections correspond to generation and negative power injections correspond to consumption.

Several means to mitigate the risk caused by power injection uncertainty have been proposed. These include demand response — i.e., changes in usage by consumers in response to dynamical electricity pricing —, storage devices and strategic dispatch of controllable generators [1, 4, 5, 67]. Above all, a quantitative reliability assessment will help grid operators to decide how to act on the risks. This anticipation can refer to both long-term investment questions and short-term operational strategies. Besides avoiding instability risks, a reliability assessment also helps to avoid unnecessary investments if the power grid reliability proves to be adequate.

Grid operators have already developed reliability assessment methods for what has colloquially been called 'the world's biggest machines' [39, 48]. However, many conventional assessments assume the power injections to be deterministic: typically a few 'worst case' injection patterns are defined and the resulting system state of the grid is examined [14]. In case of many distributed uncertain power injections, this practice may result in a misleading risk analysis as it is not clear a priori what combination of power injections can be described as the worst case [15]. As many national European grids aim for one large pan-European grid [27], this problem of complexity will become more severe. Modeling the power injections as stochastic processes, as will be the case in this thesis, is more general in this sense.

1.1 GRID RELIABILITY INDICES

A reliability assessment of a power grid involves estimating to what extent electrical constraints are violated during a specified time interval — say several minutes, an hour or even a year. *Grid reliability indices* can broadly be divided in probability, duration, frequency and severity of constraint violations. Different consumers are interested in different indices, depending on the behavior of their connected devices during and shortly after a power outage. Examples include:

- A farm factory manager desires to maintain the farm temperature, and thus is mainly concerned with curtailments that are longer than say one hour.
- A factory with an industrial machine that needs hours to restart after a curtailment is mainly concerned with the curtailment frequency. The manufacturer is relatively indifferent towards the duration of the curtailment.

The SAIDI, SAIFI, CAIDI, CAIFI indices [17] account for these different preferences. The System Average Interruption Duration Index is the

average interruption (outage) duration for each customer served, and is calculated as:

$$\text{SAIDI} = \frac{\sum_{\text{year}} \text{customer interruption durations}}{\#[\text{customers served}]}$$

Similarly, the System Average Interruption Frequency Index (SAIFI), the Customer Average Interruption Duration Index (CAIDI) and the Customer Average Interruption Frequency Index (CAIFI) are defined as

$$\begin{aligned}
\text{SAIFI} &= \frac{\#[\text{customer interruptions}]}{\#[\text{customers served}]} \\
\text{CAIDI} &= \frac{\sum_{\text{year}} \text{customer interruption durations}}{\#[\text{customer interruptions}]} = \frac{\text{SAIDI}}{\text{SAIFI}} \\
\text{CAIFI} &= \frac{\#[\text{customer interruptions}]}{\#[\text{distinct customers interrupted}]}
\end{aligned}$$

Figure 1.3 displays SAIFI indices for different European countries. Consumers in all displayed countries suffer from only a few interruptions per year. These indices are relative to the number of (distinct) customers or interruptions. Other indices quantify the occurrences of all load curtailments in the grid, where these occurrences are in fact stochastic variables [7, 48]. Here, load curtailments refer to curtailments of power *consumption*. However, in contemporary privatized energy markets grid operators must ensure power *generation* by electricity producers as well. Therefore, load-based definitions are easily generalized to power curtailments — where power refers to both consumption and generation. After defining

$$C = \{\text{A power curtailment occurs during } [0, T]\}, \tag{1.1}$$

the event of a power curtailment during the time interval $[0, T]$, we describe four of these indices below:

1. The Probability of Power Curtailments during $[0, T]$

$$\text{PPC}(T) := \mathbb{P}(C), \tag{1.2}$$

with C as defined in (1.1).

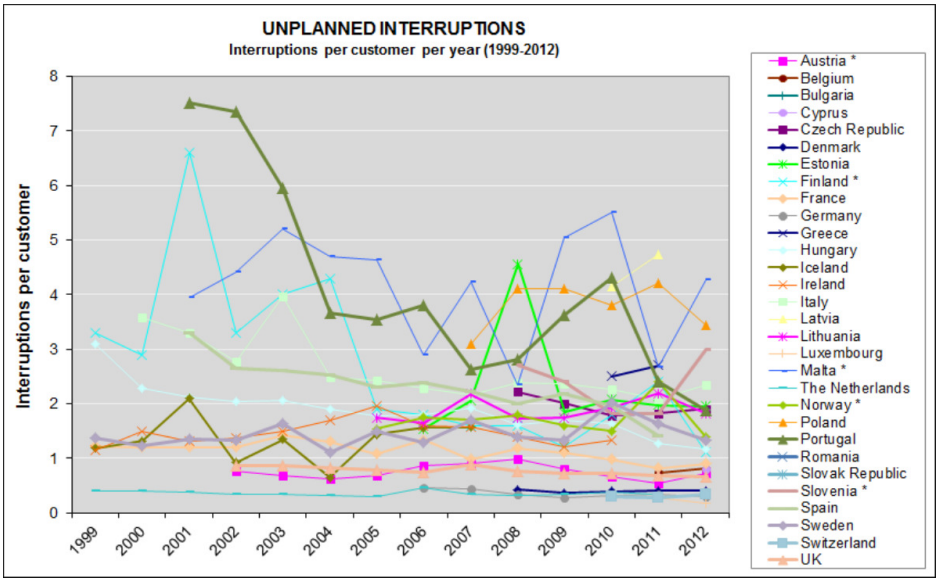


Figure 1.3: Reliability survey within Europe, by comparing SAIFI indices for different countries over different years. Source: [13]

2. The Expected Duration of Power Curtailments

$$\mathrm{EDPC}(T) := \mathbb{E}\left[D(T)\right], \tag{1.3}$$

with $D(T)$ the total duration of curtailments during $[0, T]$.

3. The Expected Number of Power Curtailments

$$\mathrm{ENPC}(T) := \mathbb{E}\left[N_C(T)\right], \tag{1.4}$$

with $N_C(T)$ the number of distinct power curtailments during $[0, T]$.

4. The Expected Energy Not Supplied accounts for the severity of the curtailment

$$\mathrm{EENS}(T) := \mathbb{E}\left[\int_0^T S_C(t)dt\right], \tag{1.5}$$

with $S_C(t)$ the size of power curtailment (in MW) at time t.

We should note that grid component failures are in general uncertain too. In [7, 48], two main categories of power system reliability are distinguished. *Power system adequacy* refers to the ability of the system to supply power to customers at all times, whereas *power system security* refers to the ability to withstand sudden disturbances like short circuits and component failures. We restrict the scope of this thesis to grid instabilities due to uncertain power injections only, i.e., we focus on power system adequacy.

1.2 RESEARCH AIM AND OUTLINE OF THE LITERATURE INVESTIGATION

In the previous section we discussed that uncertain power injections will form a challenge for grid reliability. Furthermore, we illustrated that grid operators want an insight into how reliable a power grid is. Given the uncertain nature of power injections, such an assessment is not straightforward.

The aim of this research is therefore to develop (efficient) computational methods to assess power grid reliability given the uncertain nature of the power injections. A specific example is a methodology to compute reliability indices (1.2) – (1.5) given the distribution of power injections. While conducting this research we found that our proposed Crude Monte Carlo estimation of grid reliability may require a prohibitively large number of samples. Therefore, we focused on *efficient* computation of grid reliability indices for the remainder of the research, and in particular on rare event simulation of power grids.

Estimation of reliability indices enables the grid operator to act on potential risks. This anticipation will be driven by economic factors. One example is the Value of Lost Load, denoting the amount consumers are willing to pay to avoid an outage [68]. Economic factors however are outside the scope of this thesis, which focuses on the *physical state* of the power grid. Nevertheless, the complexity of typical economic optimization problems for grid operators supports the aim of this research: estimation methods of reliability indices developed in

this thesis can be employed to evaluate each state in an optimization procedure for operational strategies or investment questions. Developing computationally efficient estimation of indices will enable optimizations that are otherwise too complex.

In Chapter 2 we introduce existing power flow models that assume deterministic parameters. Since the AC power flow equations will form the computational bottleneck of several Monte Carlo estimation methods in this work, we will describe efficient numerical solvers for this nonlinear system of equations in detail. In parts of Chapter 7 we will assume the alternative DC power flow equations, so we introduce those equations too. Furthermore, we discuss conventional deterministic approaches by grid operators to dimension power grid connections.

In Chapter 3 we first discuss models for stochastic power injections. Note that we have not yet specified the stochastic processes describing the power injections. This is because models will vary over the subsequent chapters. For some models high veracity will be an attractive property while others are analytically more tractable. Second, we show that power curtailments are typically rare events. We illustrate the high computational cost of Crude Monte Carlo Simulation for estimating very small probabilities, and introduce two existing categories of rare event simulation techniques. At the end of Chapter 3 we address the coherence of the subsequent chapters which are all based on published or submitted material.

Part II

LITERATURE INVESTIGATION

2 DETERMINISTIC POWER FLOW ANALYSIS

Any power flow model starts with the definition of its topology. The topology can be modeled by a connected graph with N nodes (also called buses) and M edges (also called branches or connections). The connection from node i to j is referred to as (i, j). In this chapter the power injections (both consumption and generation) are constant and deterministic. We will distinguish two models for the power flow equations: Alternating Current (AC) and Direct Current (DC). Both AC and DC equations are steady state equations: they relate the power injections at all nodes to the power flows through all grid connections assuming the latter immediately reach an equilibrium state. As opposed to steady state power flow models, time-dependent power flow models exist too [10, 78], but we will not consider them in this work. Such models employ time frames ranging from milliseconds to seconds and are often used to decide on sudden contingencies like a short circuit or lightning strike. Since a step size on the order of minutes will capture the typical variability of power injections, the steady state power flow equations are sufficiently accurate for our purposes. As we will see, the DC power flow equations are less accurate than AC partly because nodal voltages are assumed to be equal over all nodes, but their linear form makes it an appealing model for fast computation and analytic tractability. The AC power flow equations are nonlinear and require numerical methods to compute the systems state. Many books on power flow analysis describe both AC and DC equations in detail [18, 37, 39].

2.1 THE ALTERNATING CURRENT POWER FLOW MODEL

To avoid confusion with vector indices, $\imath$ denotes the imaginary unit. We first introduce the main variables.

- The complex nodal voltage $V_i \in \mathbb{C}$ at grid node i. We introduce the polar notation

$$V_i := |V_i| e^{\imath \delta_i}, \tag{2.1}$$

  where $|V_i| \in [0, \infty)$ is the voltage magnitude and $\delta_i \in (-\pi, \pi]$ is the voltage angle (or phase angle) at node i. The voltage drop over connection (i, j) is $V_j - V_i$, and the voltage on the ground is assumed to be zero.

- The complex current injection $I_i \in \mathbb{C}$ at node i.

- The complex current $I_{ij} \in \mathbb{C}^{N \times N}$ flowing through connection (i, j) from node i to node j.

- The complex power $S_i \in \mathbb{C}$ injected at node i. We can write

$$S_i := P_i + \imath Q_i, \tag{2.2}$$

  where the real part $P_i \in \mathbb{R}$ is called the active power and the imaginary part $Q_i \in \mathbb{R}$ is called the reactive power. For both P_i and Q_i positive values correspond to net generation at node i, whereas negative values correspond to net consumption at node i.

- The admittance $y_{ij} \in \mathbb{C}$ of connection (i, j), given by

$$y_{ij} := \frac{1}{R_{ij} + \imath X_{ij}},$$

  with $R_{ij} \in \mathbb{R}$ the resistance and $X_{ij} \in \mathbb{R}$ the reactance of connection (i, j). There exists no connection between node i and $j \neq i$ if and only if $y_{ij} = 0$.

We will derive the AC power flow equations from the complex, vector valued generalizations of Kirchoff's current law, Ohm's law and the definition of power. Kirchoff's current law states that at any node the sum of currents flowing into that node is equal to the sum of currents flowing out of that node:

$$-I_i + \sum_{k=1}^{N} I_{ik} = 0, \tag{2.3}$$

at each node i. Figure 2.1 shows an example.

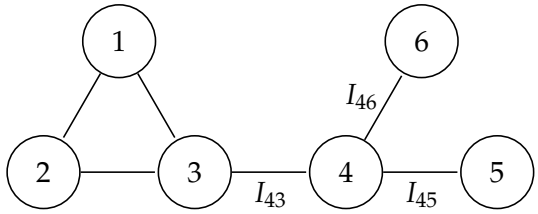


Figure 2.1: The current entering any junction is equal to the current leaving that junction: $I_4 = I_{43} + I_{46} + I_{45}$, with I_4 the current injection at node 4.

Note that Kirchoff's voltage law is also present but redundant in the model: the directed sum of the voltages around any closed circuit will by definition of the nodal voltages always be equal to zero. The second law of importance is Ohm's law: the current through a connection between two nodes is directly proportional to the voltage drop over these nodes. Ohm's law can be stated by

$$y_{ij}(V_i - V_j) = I_{ij}. \tag{2.4}$$

Combining Kirchoff's current law and Ohm's law, we can write

$$I_i = \sum_{j=1}^{N} Y_{ij} V_j, \tag{2.5}$$

or

$$I = YV \tag{2.6}$$

in matrix-vector notation for an appropriately chosen matrix $Y \in \mathbb{C}^{N \times N}$. This matrix is called the admittance matrix. It is easy to check that

$$Y_{ij} := \begin{cases} -y_{ij} & \text{if } i \neq j, \\ \sum_{k=1}^{N} y_{ik} & \text{if } i = j, \end{cases} \tag{2.7}$$

are the elements of the admittance matrix. Note that for $i \neq j$, $Y_{ij} = 0$ if and only if there is no connection between node i and j. In this sense,

Y encodes the topology of the power grid. Furthermore, y_{ii} denotes the shunt or admittance-to-ground at node i, which can be used to ensure that the nodal reactive power remains in a specified interval. Capacitor shunts and inductor shunts respectively inject or consume reactive power, resulting in a higher or lower nodal voltage, respectively. The shunt y_{ii} will cause an extra amount of current I_{is} to be injected at node i. Since the voltage on the ground is zero by definition, we have $I_{is} = y_{ii}(V_i - 0) = y_{ii}V_i$. This means that in the admittance matrix, an extra term y_{ii} has to be added to the diagonal term Y_{ii}.

We conclude the discussion on the admittance matrix by introducing two different notations for Y_{ij}:

$$Y_{ij} = G_{ij} + \imath B_{ij} = |Y_{ij}|e^{\imath\theta_{ij}}. \tag{2.8}$$

Here $G_{ij}, B_{ij} \in \mathbb{R}$ are the conductance and susceptance, respectively, of connection (i, j) if $i \neq j$. The most right-hand side expression is simply the polar notation using admittance angle $\theta_{ij} \in (-\pi, \pi]$. Now we introduce the definition of complex power

$$S_i = V_i I_i^*,$$

at node i, where x^* denotes the complex conjugate of x. Then, using (2.2) and (2.5), the complex conjugate of S_i is

$$S_i^* = P_i - \imath Q_i = V_i^* \sum_{j=1}^{N} Y_{ij} V_j. \tag{2.9}$$

Using (2.1) and (2.8) we rewrite the right-hand side in polar notation

$$P_i - \imath Q_i = \sum_{j=1}^{N} |V_i Y_{ij} V_j| e^{\imath(\theta_{ij} + \delta_j - \delta_i)}.$$

Expanding this equation and equating real and reactive (i.e. imaginary) parts results in the *AC Power Flow Equations* (AC PFEs). That is,

$$P_i = \sum_{j=1}^{N} |V_i Y_{ij} V_j| \cos(\theta_{ij} + \delta_j - \delta_i) \quad \text{for all nodes } i, \tag{2.10}$$

$$Q_i = -\sum_{j=1}^{N} |V_i Y_{ij} V_j| \sin(\theta_{ij} + \delta_j - \delta_i) \quad \text{for all nodes } i. \tag{2.11}$$

The power flow equations (2.10) and (2.11) form the central model of many power flow analyses. The admittance matrix elements $|Y_{ij}|$ and θ_{ij} are given, as well as power injections P_i, Q_i. The system of equations should be solved for the voltage angles δ_i at all nodes and the voltage magnitudes $|V_i|$ at most nodes, as will be explained in the next section.

2.1.1 *A modified Newton-Raphson solver for the AC power flow equations*

In this section we will describe the details of solving the AC PFEs (2.10) – (2.11) using a modified Newton-Raphson method. First, we distinguish three types of grid nodes:

1. *The slack node.*
   A power flow model will contain exactly one slack node (also called the swing bus), where the residual power of the network is either generated or consumed. Hence, no power flow equations have to be solved at the slack node. In this thesis, node 1 is always the slack node. Its voltage magnitude $|V_1|$ is given and without loss of generality, we set $\delta_1 = 0$.

2. *PQ nodes.*
   If node i is a PQ node, a specified amount of real power P_i and reactive power Q_i is injected at that node. Voltage magnitude $|V_i|$ and voltage angle δ_i are unknown in the AC PFEs for each PQ node i. Typical examples of PQ nodes are nodes where power is consumed only, and thus they are also known as load nodes. However, small-scale generators often control the real and reactive power, and we assume in this thesis that all nodes with an uncertain power injection are PQ nodes.

3. *PV nodes.*
   These nodes are also known as voltage-controlled nodes, since apart from P_i voltage magnitude $|V_i|$ is kept at a specified value at each PV node i. At node i voltage angle δ_i is therefore the only unknown to be solved in AC PFEs (2.10) – (2.11). The amount of reactive power Q_i is not given at this node but follows immediately from

the solution of the AC PFEs by substituting this solution in (2.11). One should think of these nodes as locations where large, controllable power plants like fossil-fuel power stations are connected to the grid. By tuning the turbine real power P_i is controlled, and the voltage magnitude is controlled by adjusting the generator excitation. For this reason, PV nodes were also referred to as generator nodes before the rise of small-scale generators. However, small-scale generators like wind turbines or solar panels are often insufficiently powerful to control the nodal voltage in a power grid, so corresponding nodes are often PQ nodes.
Note that the AC PFEs (2.10) – (2.11) are expressed in polar form, and not for example as in (2.9). The reason is that it is the voltage magnitude $|V_i|$ that is given at PV nodes, and not the real and imaginary parts of V_i.

Suppose that the network consists of N_{PQ} PQ nodes, N_{PV} PV nodes and of course one slack node. Then we can list the numbers of specified quantities, available equations and state variables as given in Table 2.1.

node type	*# nodes*	*Quantities specified*	*# available equations*	*# state variables* $\delta_i, \lvert V_i\rvert$
Slack	1	$\delta_1, \lvert V_1\rvert$	0	0
PV	N_{PV}	$P_i, \lvert V_i\rvert$	N_{PV}	N_{PV}
PQ	N_{PQ}	P_i, Q_i	$2N_{PQ}$	$2N_{PQ}$
Total	$N_{PV} + N_{PQ} + 1$	$2(N_{PV} + N_{PQ} + 1)$	$N_{PV} + 2N_{PQ}$	$N_{PV} + 2N_{PQ}$

Table 2.1: Summary of the AC Power Flow Equations.

There is no closed-form solution available for the nonlinear system of AC PFEs. In fact, the solution may not exist for a given set of parameters. This case can be interpreted as the generators and the slack node being incapable of delivering the specified demand in the power grid. In most practical situations a voltage collapse occurs instead. Typically, voltage

constraints (as will be described in Section 2.3) are violated long before a voltage collapse occurs, and we ignore this possibility.

Two suitable iterative methods used to solve the AC PFEs are the Gauss-Seidel method and the Newton-Raphson method. The latter is known to outperform the former in speed-accuracy ratio for all except very small systems [39]. We will give an overview of the Newton-Raphson method.

1. We choose initial values $\delta_i^{(0)}, |V|_i^{(0)}$ for all state variables. A typical choice is $\delta_i^{(0)} = 0$ for all nonslack nodes i and $|V|_i^{(0)} = 1$ for all PQ nodes i. Set the index of the Newton-Raphson iteration k to zero.

2. We substitute approximations $\delta_i^{(k)}, |V|_i^{(k)}$ into the PFEs (2.10) – (2.11) to calculate the power injection approximations $P_i^{(k)}, Q_i^{(k)}$ for all nodes i. Compute mismatches

   $$\Delta P_i^{(k)} := P_i^{(k)} - P_i,$$

   for all nonslack nodes i. Similarly, compute mismatches

   $$\Delta Q_i^{(k)} := Q_i^{(k)} - Q_i,$$

   for all PQ nodes i.

3. We compute a modified Jacobian of the system. For notational convenience, we introduce the Jacobian matrix equation assuming initially that all N nodes are PQ nodes:

   $$\begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} \begin{pmatrix} \Delta\delta^{(k)} \\ \Delta|V|^{(k)} \end{pmatrix} = \begin{pmatrix} \Delta P^{(k)} \\ \Delta Q^{(k)} \end{pmatrix}. \tag{2.12}$$

   Here $\Delta\delta^{(k)}, \Delta|V|^{(k)}, \Delta P^{(k)}, \Delta Q^{(k)} \in \mathbb{R}^N$ denote the vector differences of the voltage angle, voltage magnitude, active powers and reactive

powers, respectively, at iteration k. The Jacobian block-matrices J_{11}, J_{12}, J_{21}, $J_{22} \in \mathbb{R}^{N\times N}$ are given by

$$(J_{11})_{ij} := \frac{\partial P_i}{\partial \delta_j}, \quad (J_{12})_{ij} := \frac{\partial P_i}{\partial |V_j|},$$
$$(J_{21})_{ij} := \frac{\partial Q_i}{\partial \delta_j}, \quad (J_{22})_{ij} := \frac{\partial Q_i}{\partial |V_j|}.$$

The modification of the Jacobian in (2.12) is based on the following equivalent equation:

$$\begin{pmatrix} J_{11} & J_{12}D \\ J_{21} & J_{22}D \end{pmatrix} \begin{pmatrix} \Delta\delta^{(k)} \\ D^{-1}\Delta|V|^{(k)} \end{pmatrix} = \begin{pmatrix} \Delta P^{(k)} \\ \Delta Q^{(k)} \end{pmatrix}, \tag{2.13}$$

where matrix $D \in \mathbb{R}^{N\times N}$ is diagonal with nonzero elements $D_{ii} = |V_i|^{(k)}$. The matrix in (2.13) is the modified Jacobian. This modification saves the computation of half of the matrix elements as they become related to each other. That is, it is readily checked from (2.10) – (2.11) that

$$\begin{aligned}
|V_j|\frac{\partial P_i}{\partial |V_j|} &= -\frac{\partial Q_i}{\partial \delta_j} = |V_i V_j Y_{ij}| cos(\theta_{ij} + \delta_j - \delta_i), \qquad \text{for } i \neq j, \\
|V_j|\frac{\partial Q_i}{\partial |V_j|} &= \frac{\partial P_i}{\partial \delta_j} = -|V_i V_j Y_{ij}| sin(\theta_{ij} + \delta_j - \delta_i), \qquad \text{for } i \neq j, \\
|V_i|\frac{\partial Q_i}{\partial |V_i|} &= -\frac{\partial P_i}{\partial \delta_i} - 2|V_i|^2 B_{ii} = Q_i - |V_i|^2 B_{ii}, \\
|V_i|\frac{\partial P_i}{\partial |V_i|} &= \frac{\partial Q_i}{\partial \delta_i} + 2|V_i|^2 G_{ii} = P_i + |V_i|^2 G_{ii}.
\end{aligned} \tag{2.14}$$

We started this Newton-Raphson step by assuming all nodes are PQ nodes. However, since not all nodes are PQ nodes certain equations and terms should be removed from (2.13) (see the node type descriptions at the start of Section 2.1.1). That is, since node 1 is a slack node:

a) Angle $\delta_1 = 0$ and $|V_1| = 1$ are given so $\Delta\delta_1^{(k)} = \Delta|V_1|^{(k)} = 0$. Elements $\Delta\delta_1^{(k)} = \Delta|V_1|^{(k)}$ as well as the first column of

$J_{11}, J_{12}D, J_{21}$ and $J_{22}D$ can therefore be removed from linear system (2.13).

b) Mismatches $\Delta P_1^{(k)}$ and $\Delta Q_1^{(k)}$ are not defined so the first elements of $\Delta P^{(k)}$ and $\Delta Q^{(k)}$ and the first row of $J_{11}, J_{12}D, J_{21}$ and $J_{22}D$ should be removed from (2.13).

Additionally, for each PV node i:

a) The voltage magnitude $|V_i|$ is given so $\Delta|V_i|^{(k)} = 0$. Elements $\Delta|V_i|$ as well as the i-th column of $J_{12}D$ and $J_{22}D$ can therefore be removed from (2.13).

b) The mismatch $\Delta Q_i^{(k)}$ is not defined so the i-th element of $\Delta Q^{(k)}$ and the i-th row of J_{21} and $J_{22}D$ should be removed from (2.13).

The resulting system reads

$$\begin{pmatrix} \tilde{J}_{11} & \tilde{J}_{12}\tilde{D} \\ \tilde{J}_{21} & \tilde{J}_{22}\tilde{D} \end{pmatrix} \begin{pmatrix} \Delta(\delta)_P^{(k)} \\ \Delta|V|_Q^{(k)}/|V|_Q^{(k)} \end{pmatrix} = \begin{pmatrix} \Delta P^{(k)} \\ \Delta Q^{(k)} \end{pmatrix}, \tag{2.15}$$

Here $\tilde{J}_{11}, \tilde{J}_{12}\tilde{D}, \tilde{J}_{21}$ and $\tilde{J}_{22}\tilde{D}$ are the submatrices obtained by removing rows and columns as described from the modified Jacobian in (2.13). $(\delta)_P$ is the subvector of δ with all elements corresponding to PV and PQ nodes. $|V|_Q$ is the subvector of $|V|$ with all elements corresponding to PQ nodes. The division $\Delta|V|_Q^{(k)}/|V|_Q^{(k)}$ is performed elementwise.

4. We solve equation (2.15) for $\Delta(\delta)_P^{(k)}$ and $\Delta|V|_Q^{(k)}/|V|_Q^{(k)}$. We compute the next step approximations

$$\delta_i^{(k+1)} = \delta_i^{(k)} + \Delta\delta_i^{(k)},$$

for all PV and PQ nodes i, and

$$|V_i|^{(k+1)} = |V_i|^{(k)} \left(1 + \frac{\Delta|V_i|^{(k)}}{|V_i|^{(k)}}\right),$$

for all PQ nodes i.

5. We use these new approximations for the state variables for step 2 and iterate steps 2 to 5 until $\Delta P_i^{(k)}, \Delta Q_i^{(k)}$ are within a desired tolerance.

Once the nodal voltages are found, all connection currents I_{ij} (and thus the power flowing through all connections) can be computed using Ohm's law (2.4).

2.1.2 *Fast Decoupled Power Flow*

To improve the computational efficiency of the described Newton-Raphson method, the Fast Decoupled Power Flow (FDPF) has been developed [81]. In the last decades, the Fast Decoupled Power Flow method has become prevalent in industry to solve power flow equations [20, 52, 83]. The acceleration is based on six relatively weak assumptions under which the Jacobian is constant over all iterations. The resulting approximate version of the Newton-Raphson method typically requires more iterations, but each iteration will be computationally less intensive, and the FDPF often requires less workload than the original Newton-Raphson method in Section 2.1.1. The first two assumptions are:

1. A change in the voltage magnitude leaves the flow of real power unchanged:

$$\frac{\partial P_i}{\partial |V_j|} = 0,$$

for $i, j = 1, \ldots, N$.

2. A change in the voltage angle δ leaves the flow of reactive power Q unchanged:

$$\frac{\partial Q_i}{\partial \delta_j} = 0,$$

for $i, j = 1, \ldots, N$.

Then $J_{12} = J_{21} = 0$, so linear system (2.15) can be split into two systems:

$$J_{11}\Delta(\delta)_P^{(k)} = \Delta P^{(k)}, \tag{2.16}$$

and

$$J_{22}\Delta|V|_Q^{(k)} = \Delta Q^{(k)}. \tag{2.17}$$

This is the *decoupling* part of the algorithm. The *fast* part of the algorithm involves four assumptions based on the following rules of thumb:

3. The angular differences $\delta_i - \delta_j$ are usually so small that

$$\cos(\delta_j - \delta_i) \approx 1,$$
$$\sin(\delta_j - \delta_i) \approx \delta_j - \delta_i.$$

4. The connection susceptances B_{ij} are usually much larger than the connection conductances G_{ij} so that

$$G_{ij}\sin(\delta_j - \delta_i) \ll B_{ij}\cos(\delta_j - \delta_i).$$

5. Q_i at node i satisfies

$$Q_i \ll |V_i|^2 B_{ii},$$

6. The voltage magnitude at node i is usually close to the nominal value:

$$|V_i| \approx 1.$$

We will use these approximations 3-6 to simplify J_{11} and J_{22}, whose off-diagonal elements are given by

$$\frac{\partial P_i}{\partial \delta_j} = |V_j|\frac{\partial Q_i}{\partial |V_j|} = -|V_i V_j Y_{ij}|\sin\left(\theta_{ij} + \delta_j - \delta_i\right).$$
$$= -|V_i V_j|\Big[B_{ij}\cos\left(\delta_j - \delta_i\right) + G_{ij}\sin\left(\delta_j - \delta_i\right)\Big].$$

First, approximations 3 and 4 yield

$$\frac{\partial P_i}{\partial \delta_j} = |V_j| \frac{\partial Q_i}{\partial |V_j|} \approx -|V_i V_j| B_{ij}. \tag{2.18}$$

Second, approximation 5 reduces the diagonal elements (2.14) of J_{11} and J_{22} to

$$\frac{\partial P_i}{\partial \delta_i} \approx |V_i| \frac{\partial Q_i}{\partial |V_i|} \approx -|V_i|^2 B_{ii}. \tag{2.19}$$

The resulting approximation of the first decoupled equation (2.16) reads

$$-\begin{pmatrix} |V_2||V_2|B_{22} & \dots & |V_2||V_N|B_{2N} \\ \vdots & \ddots & \vdots \\ |V_N||V_2|B_{N2} & \dots & |V_N||V_N|B_{NN} \end{pmatrix} \Delta(\delta)_P^{(k)} = \Delta P^{(k)}, \tag{2.20}$$

Finally, applying assumption 6 to the first voltage magnitude of each matrix element simplifies (2.20) to

$$-B_P \;\; \Delta(\delta)_P^{(k)} = \Delta P^{(k)} / |V|_P^{(k)}. \tag{2.21}$$

Here $|V|_P^{(k)}$ is the subvector of all elements of $|V|^{(k)}$ that correspond to PV or PQ nodes and the division is again performed elementwise. The approximate Jacobian $-B_P \in \mathbb{R}^{(N-1)\times(N-1)}$ simply consists of the elements $-\mathrm{Im}\{Y_{ij}\}$ for all PV and PQ nodes i and j. Similarly, the approximation of the second decoupled equation (2.16) becomes

$$-B_Q \;\; \Delta|V|_Q^{(k)} = \Delta Q^{(k)} / |V|_Q^{(k)}. \tag{2.22}$$

Here the approximate Jacobian $-B_Q \in \mathbb{R}^{N_{PQ}\times N_{PQ}}$ similarly consists of the elements $-\mathrm{Im}\{Y_{ij}\}$ for all PQ nodes i and j. Note that the approximate Jacobians $-B_P$ and $-B_Q$ remain constant over all Newton-Raphson iterations. They can be computed before iterations are commenced and each iteration can therefore be evaluated relatively fast. One disadvantage is the fact that more iterations may be necessary due to the error of approximations. However, the idea is that the approximations are accurate

enough for the convergence to be faster than the convergence of the conventional Newton-Raphson method of Section 2.1.1. In the original article of FDPF examples are shown where convergence required a factor 5 less workload than when the exact Jacobian is used as in Section 2.1.1 [81].

2.1.3 *Sparse computations*

The typical number N of power grid nodes depends on what is defined as one grid. Most definitions consider either a transmission grid (transporting electricity at higher voltage levels) or a distribution grid (delivering electricity to individual consumers), since grid operators are typically responsible for only one of the two. Assuming this distinction, N will range from tens to hundreds or thousands grid nodes [39]. One notable exception is the Eastern Interconnection Eastern US power grid with as many as $N = 49\,000$ nodes [41]. This number may increase even further as more power grids become connected [44, 45].

Although a power grid in theory contains M connections with $N - 1 \leq M \leq N(N-1)/2$, M is typically on the order of N. The admittance matrix Y is therefore sparse in most power grids. In this section we will explain how to benefit computationally from the sparsity of Y. To compute the mismatches $\Delta P_i/|V_i|$ as proposed in the previous subsection, one needs to evaluate the current Newton-Raphson iteration approximation for

$$\begin{aligned} P_i/|V_i| &= \sum_{j=1}^{N} |Y_{ij}V_n| \cos(\theta_{ij} + \delta_j - \delta_i), \\ &= \sum_{j=1}^{N} |V_j| \Big(G_{ij} \cos(\delta_j - \delta_i) + B_{ij} \sin(\delta_j - \delta_i) \Big), \end{aligned}$$

for all nodes i. The second equality follows from a trigonometric identity and the definition (2.8) of Y. One can write this in the vector form

$$P/|V| = A(G, B, \boldsymbol{\delta})|\boldsymbol{V}|, \tag{2.23}$$

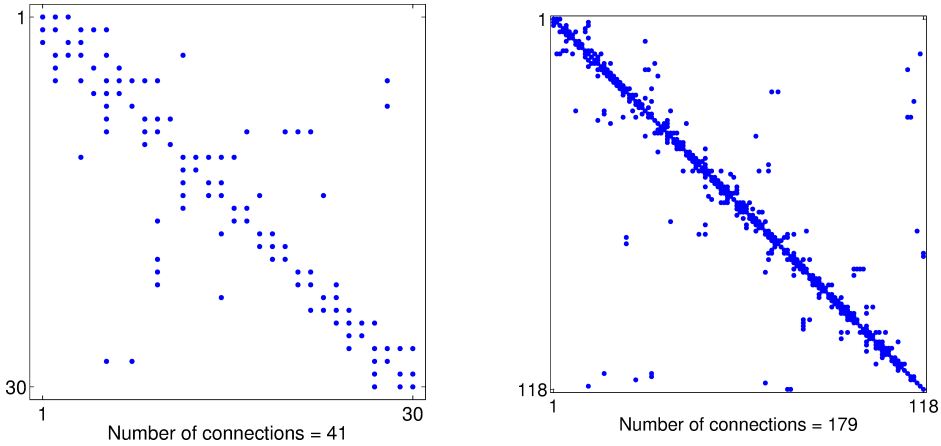


Figure 2.2: Sparsity of Y for the IEEE-30 and IEEE-118 test cases, respectively. Test cases can be found in [87]

with vectors $\boldsymbol{P}, |\boldsymbol{V}|, \boldsymbol{\delta} \in \mathbb{R}^N$, where the division on the left-hand side is performed elementwise, and where matrix $A(G, B, \boldsymbol{\delta}) \in \mathbb{R}^{N \times N}$ depends on $G = (G_{ij}), B = (B_{ij})$ and $\boldsymbol{\delta}$:

$$A = (A_{ij}), \quad \text{with} \quad A_{ij} = G_{ij} \cos(\delta_j - \delta_i) + B_{ij} \sin(\delta_j - \delta_i).$$

Now note that A will be at least as sparse as $Y = G + \imath B$. Therefore, to evaluate (2.23), workload will be reduced by only computing the necessary terms in the summand by precaching the indices of nonzero elements of Y. Neglecting the cost of inversion of matrices B_P and B_Q — which is reasonable in Monte Carlo simulations of the following chapters since we can reuse the inverse every time step and sample — it is readily checked that the computational complexity of one sparse FDPF iteration grows as $\mathcal{O}(M)$, with M the number of power grid connections. This compares to $\mathcal{O}(N^2)$ for nonsparse FDPF as described in Section 2.1.2. If used in a Monte Carlo simulation, conventional Newton-Raphson as explained in Section 2.1.1 will certainly be computationally inferior to FDPF: at every time step in every sample a linear system must be solved instead of a matrix-vector multiplication. An experiment comparing sparse FDPF and the conventional Newton-Raphson method showed

a decrease in CPU time for all but the smallest IEEE test cases (results not shown here). In the IEEE-300 test case with $N = 300, M = 411$ the method converged around twice as fast, confirming that a sparse FDPF method accelerates the conventional Newton-Raphson method for the AC PFEs. We will use both sparse computations and the FDPF method to solve the AC PFEs in this work.

Using Table 2.2 we will give an insight in the computational costs of different parts of the sparse FDPF solver for different IEEE test cases [87]. All average CPU times are based on 100 measurements. The initialization

IEEE-N test case, $N =$	14	30	57	118	300
Initialization	0.112	0.15	0.26	0.71	6.62
Inversion of B	0.084	0.13	0.23	1.09	9.66
Inversion of $\bar{B}$	0.031	0.061	0.13	0.25	4.81
Sparse indexing	0.031	0.037	0.060	0.17	1.33
Newton-Raphson	0.52	0.46	1.16	1.16	6.88
Post-processing	0.07	0.11	0.26	0.61	6.53
Total	0.85	0.95	2.09	4.00	35.8
# Newton-Raphson iterations	7	9	16	7	16

Table 2.2: Average CPU times (ms) of parts of the sparse FDPF solver.

refers to the extraction and processing of input data. The first inversion is that of the Jacobian in (2.21). The second inversion is that of the Jacobian in (2.22), which is a smaller matrix since the PV node equations are omitted here. The part of sparse indexing refers to the collection of indices of nonzero elements of Y, as well as the corresponding indices of other matrices. In the Newton-Raphson loop, the solution for the state variables $|\boldsymbol{V}|$ and $\boldsymbol{\delta}$ is derived. From this solution, all power injections, connection currents and connection power flows are derived in the post-processing part. One can see from this table that for larger networks, the CPU time of the inversions becomes significant. For $N = 118$ or $N = 300$ this part is computationally more intensive than the part of the Newton-Raphson iterations. Nevertheless, Monte Carlo simulation will

require both inversions only once after which the resulting inverse can be used each time step and sample path. Therefore, the workload of the two inversions will in our case most probably be insignificant, and we will not attempt to improve it.

2.2 THE DIRECT CURRENT POWER FLOW MODEL

The alternative Direct Current (DC) PFEs can easily be derived from the FDPF method described in Section 2.1.2. The two main assumptions are the following:

1. The voltage magnitudes are assumed to be equal to the nominal value: $|V_i| = 1$ at each PQ node i. Note that the FDPF method assumed this to be true for some values $|V_i|$ in the *Jacobian* only, whereas the DC model assumes nominal voltages in the system itself.

2. Shunts y_{ii} are ignored in the admittance matrix Y (see (2.7)). The resulting susceptance matrix is denoted by B'_P.

The first assumption implies $\Delta|V|_Q^{(k)} = 0$ in (2.22) and thus this equation becomes redundant. It remains to solve (2.21) for the voltage angles δ_i at all nonslack nodes only. Since $|V_i| = 1$, this linear system becomes

$$-B'_P \begin{pmatrix} \Delta\delta_2 \\ \vdots \\ \Delta\delta_N \end{pmatrix} = \begin{pmatrix} \Delta P_2 \\ \vdots \\ \Delta P_N \end{pmatrix}.$$

The computational complexity of the algorithm is $O(M)$ when using sparse computations. This is the same as that of one sparse FDPF iteration of the AC PFEs and DC solvers are therefore faster than AC solvers. Furthermore, the linear form of the DC model enables a closed-form solution for the state variables. For this reason an analytic approach is more often viable when assuming the DC power flow model than when assuming the nonlinear AC power flow model. However, the DC power flow model assumes voltages to be equal to the nominal voltage value.

This is a strong assumption, so the DC model is considered less accurate than the AC model for power grids with alternating current.

2.3 GRID STABILITY

We call a power grid stable (also called *in normal operation*) if the following constraints (also called *operating limits*) are satisfied [96, 7]:

1. *Connection constraints*.
   For each connection (i, j), the temperature $T_{ij}(t)$ should be bounded at all time t:

   $$T_{ij}(t) < T_{ij}^{\max}. \tag{2.24}$$

   Violation of this stability constraint will cause the corresponding line to loose its tensile strength or sag. In turn, this will influence the admittance of the line, although we neglect this phenomenon in this thesis. Grid operators have to take this constraint into account when dimensioning a new cable or line. One sufficient condition for constraint (2.24) to hold is that the connection current is bounded

   $$|I_{ij}(t)| \leq I_{ij}^{\max}, \tag{2.25}$$

   or equivalently that the power flowing through the connection is bounded:

   $$|P_{ij}(t)| \leq P_{ij}^{\max}. \tag{2.26}$$

   This condition is in general too strong as the temperature incurs some lag time, as we will illustrate in Chapter 4.

2. *Voltage (magnitude) constraints*. The voltage magnitudes should lie between acceptable bounds at all PQ nodes at all time t.

   $$V^{\min} \leq V(t) \leq V^{\max}, \tag{2.27}$$

   If a voltage constraint is violated, equipment connected at the corresponding node will get damaged.

3. *Reactive power constraints*. The reactive power should lie between acceptable bounds at all PV nodes.

$$Q^{\min} \leq Q(t) \leq Q^{\max}. \tag{2.28}$$

Grid operators are responsible for monitoring the stability of the power grid and act on predicted violations of stability constraints (2.24), (2.27) and (2.28). Conventionally, corrective actions like rescheduling generation has been used as a first attempt to avoid predicted violations. If not all violations can be prevented in this way, the grid operator will curtail loads. That is, at specific nodes demanded power is not delivered. However, as explained in Section 1.1, grid operators can not easily reschedule generation in privatized electricity markets since power suppliers are market players just as power consumers. Therefore, rescheduling generation can be viewed as a curtailment that is similar to a load curtailment, and we regard both as a *power curtailment* in general. In fact, to resolve grid instability, an Optimal Power Flow problem has to be solved, where the most economic dispatch of both generation and consumption is chosen such that all constraints are satisfied.

Optimal Power Flow is a research area in itself (see e.g. Soliman and Mantawy [80, Chapter 5]), and is outside the scope of this thesis. Instead, we assume that a violation of a stability constraint like (2.24), (2.27) or (2.28) immediately induces a power curtailment and is as such undesirable. However, the complexity of many Optimal Power Flow problems justifies the aim of this research to develop accelerated simulation techniques of grid violations occurrences: a natural extension of this research would be an Optimal Power Flow solver where such a simulation method evaluates each state in the optimization procedure. In this way relatively complex Optimal Power Flow models incorporating power injection uncertainty can be solved within a reasonable amount of time.

2.4 DETERMINISTIC HEURISTICS ASSESSING POWER SYSTEM RELIABILITY

In this chapter we introduced deterministic models for power flow analysis. Many deterministic power flow analyses were developed in the twentieth century when centralized, controllable (fossil-fuel) power stations supplied the electricity in a 'top-down' fashion. Deterministic AC power flow equations were used to compute the nodal voltages from predicted constant power injections (or piecewise constant functions of time). Using these values the grid stability could then be evaluated.

Instead of using a probabilistic approach, the grid state can in principle be evaluated using different *scenarios*, including the scenario under normal operation and specified worst case scenarios. A widely used example of a worst case scenario is the $n-k$ criterion [48]: given n grid components (connections, generators, transformers, etc.), is the grid stable if k components — with $k = 0, 1, 2$ being typical values — fail? Iterating over all possible combinations of component failures gives an insight what types of contingent events the power grid can withstand. However, the probability that a combination of components fail simultaneously is not taken into account. Therefore, some scenarios may be very unlikely, or even worse, a likely and catastrophic scenario of more than k component failures is neglected in the analysis. Furthermore, only the state of components are assumed uncertain, and not the power injections.

Other deterministic approaches have been used to account for power flow uncertainty. One example is the Strand-Axelson model [82] that heuristically relates the maximum load $P^{\max}$ to the annual energy consumption E_y by a consumer:

$$P^{\max} \approx \alpha E_y + \beta \sqrt{E_y}. \tag{2.29}$$

The coefficients α and β have to be determined empirically, and will typically depend on the considered area and the connection type [98]. Note that the maximum loads of different consumers $P_i^{\max}$ will in general occur at different times. This implies that the maximum load $P_{\text{cable}}^{\max}$ of a

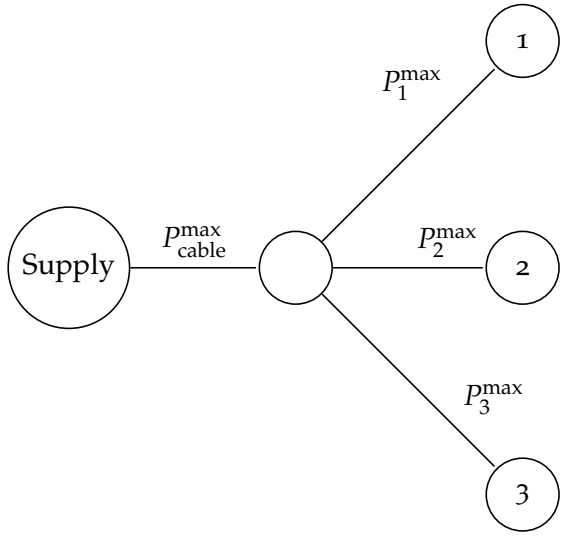


Figure 2.3: As peak loads of consumers do not occur simultaneously in general, $P_{\text{cable}}^{\max} = \sum_{i=1}^{3} P_i^{\max}$ does not necessarily hold.

cable will be less than the sum $\sum_{i=1}^{n} P_i^{\max}$ over all lower level lines fed by this cable (see Figure 2.3).

The Rusck model [74] heuristically relates $P_{\text{cable}}^{\max}$ to $P_i^{\max}$ assuming homogeneous patterns of all consumers i:

$$P_{\text{cable}}^{\max} \approx n P^{\max,1} \left(s + (1-s)/\sqrt{n} \right). \tag{2.30}$$

Here n is the number of consumers and simultaneity factor s is to be found empirically.

3

STOCHASTIC SIMULATION OF POWER FLOW MODELS

When power grids are significantly fed by uncertain and distributed generation, deterministic power flow analyses using heuristics as explained in the Section 2.4 become inaccurate. For example, the Strand-Axelson model in (2.29) and the Rusck model in (2.30) assume homogeneous consumption over all consumers, which is unrealistic if consumers also generate power. Power will flow 'two-way' rather than 'top-down', so the sum of lower level power flows will not be bounded by power flows on aggregated levels. Furthermore, it will become challenging to determine worst case scenarios for the power grid. That is, it is not clear a priori what typical combination of power injections cause a violation of a stability constraint.

A stochastic model for power injections avoids these problems by incorporating the typical variability of nodal generation (and consumption). In this thesis, we assume the power injections to be the only stochastic sources. By inertia of meteorological systems, temporal correlation of weather-dependent generation can be considered significant. Similarly, we can realistically assume that other power injections such as consumption and fossil-fuel power generation are continuous processes in time. In this thesis we choose a model that can incorporate temporal correlations by modeling the power injections as stochastic processes, discretized in time. A time step size on the order of several minutes will incorporate the typical variability of renewable generation.

For renewable generation, modeling the meteorological source — e.g., the wind speed or solar radiation – directly instead of the corresponding amount of generation is often preferred, since properties of the former are often well-known and data is more often available. As an example, Chapter 4 discusses desirable properties of a wind speed model, including

a realistic stationary distribution, time-continuity, temporal periodicities and spatial dependence between wind speeds at different locations.

3.1 SUMMARY OF THE STOCHASTIC MODEL

Let us summarize the model choices and assumptions made so far in this thesis.

1. A stochastic processes for the power injections is specified and discretized in time. The choice for the stochastic processes will vary over the coming chapters.

2. The AC PFEs

$$P_i = \sum_{j=1}^{N} |V_i Y_{ij} V_j| \cos(\theta_{ij} + \delta_j - \delta_i) \quad \text{for all nodes } i, \tag{3.1}$$

$$Q_i = -\sum_{j=1}^{N} |V_i Y_{ij} V_j| \sin(\theta_{ij} + \delta_j - \delta_i) \quad \text{for all nodes } i, \tag{3.2}$$

   as derived in Section 2.1 relate the power injections to all nodal voltages $|V_i|e^{-\imath\delta_i}$ and thus to all other electrophysical quantities of interest.

3. As explained in Section 2.3, we assume that a violation of a grid stability constraint like

$$|I_{ij}(t)| < I_{ij}^{\max}$$

   or

$$V^{\min} \leq V(t) \leq V^{\max}$$

   immediately induces a power curtailment.

4. As described in Section 1.2, we want to compute the value of reliability indices $\mathrm{PPC}(T)$, $\mathrm{EDPC}(T)$, $\mathrm{ENPC}(T)$ and $\mathrm{EENS}(T)$ (see (1.2) – (1.5)) over a specified period $[0, T]$, given the distribution

of the power injections. Note that we can write each of these reliability indices as an expectation $\mathbb{E}[I]$. Then using the law of total expectation, we can write the index as a product of a probability and a conditional expectation:

$$\mathbb{E}[I] = \mathbb{P}(C)\mathbb{E}[I|C] + \mathbb{P}(C^c)\mathbb{E}[I|C^c] = \mathbb{P}(C)\mathbb{E}[I|C],$$

with

$$C = \{\text{A power curtailment occurs during } [0, T]\}. \tag{3.3}$$

At each time step the mapping of the power injections to the occurrence of a curtailment C involves solving the AC PFEs (3.1) – (3.2). The solution of the nonlinear algebraic system (3.1) – (3.2) is only implicitly defined, so we can not derive grid reliability indices like $\mathbb{P}(C)$ directly.

3.2 MONTE CARLO SIMULATION

Monte Carlo simulation can be employed to estimate expectations (and thus probabilities as $\mathbb{P}(A) = \mathbb{E}[\mathbf{1}_A]$) that are too complex to derive directly [72]. Suppose we want to compute the expectation $\mathbb{E}[H(X)]$ of a deterministic function H of a random vector X containing all random variables in the model as elements. Then a Crude Monte Carlo (CMC) estimator is defined by

$$\hat{H}_n := \frac{1}{n}\sum_{i=1}^{n} H(X_i). \tag{3.4}$$

Here the random samples $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables with the same distribution as X. Then estimator $\hat{H}_n$ is unbiased as $\mathbb{E}[\hat{H}_n] = \mathbb{E}[H(X_1)] = \mathbb{E}[H(X)]$, and its variance

$$\operatorname{Var} \hat{H}_n = \frac{1}{n} \operatorname{Var} H(X_1) \tag{3.5}$$

is inversely proportional to the number of samples n. In fact, the Central Limit Theorem asserts that as $n \to \infty$,

$$\frac{\hat{H}_n - \mathbb{E}[\hat{H}_n]}{\sqrt{\operatorname{Var} \hat{H}_n}} = \sqrt{n} \frac{\hat{H}_n - \mathbb{E}[H(X_1)]}{\sqrt{\operatorname{Var} H(X_1)}} \xrightarrow{d} N(0,1) \tag{3.6}$$

converges in distribution to a standard normal random variable. Therefore, by choosing n sufficiently large, the value of $\hat{H}_n$ will be sufficiently close to the correct value $\mathbb{E}[H(X)]$. Note that this holds for a general number of dimensions of X and for any function H, illustrating the robustness of Monte Carlo simulation in the number of dimensions and in the complexity of the model.

We propose estimating $\mathbb{P}(C)$ using Monte Carlo simulation. First, we sample the discretized stochastic processes and solve the AC PFEs (3.1) – (3.2) at each time step. Second, we evaluate whether all grid stability constraints are satisfied and if not, we say that a power curtailment occurs at that time step. Finally, repeating this for all time steps in $[0, T]$ yields one realization of $\mathbf{1}_C$, and the average over many such realizations constitutes a Monte Carlo estimate for $\mathbb{E}[\mathbf{1}_C] = \mathbb{P}(C)$.

3.2.1 *Computational inefficiency of Crude Monte Carlo*

Power curtailments are undesirable. As explained Chapter 1 we may expect their occurrence to be rare in modern societies. When T is equal to one week, values of $\mathbb{P}(C) \approx 10^{-4}$ or much smaller are not uncommon [12, 11]. For smaller time intervals this probability becomes even smaller. In fact, having *zero* contingent disruptions per year is known to be a recent target of Transmission System Operators [84, Chapter 2.5].

In case the value of interest is a very small probability $\mathbb{P}(C)$, the CMC estimator in (3.4) becomes

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{C \text{ in sample } i\}},$$

which is the average of n Bernoulli random variables with probability $\mathbb{P}(C)$. Each Bernoulli variable has variance $\mathbb{P}(C)(1-\mathbb{P}(C))$, so (3.5) becomes

$$\operatorname{Var} \hat{p}_n = \frac{\mathbb{P}(C)(1-\mathbb{P}(C))}{n}.$$

The Central Limit Theorem in (3.6) showed that $\hat{p}_n$ is approximately normal, so the half-width of an approximate 95% confidence interval for $\mathbb{P}(C)$ will be

$$2\sqrt{\frac{\mathbb{P}(C)(1-\mathbb{P}(C))}{n}} \approx 2\sqrt{\mathbb{P}(C)/n},$$

where the approximation is justified as $\mathbb{P}(C)$ is very small. A confidence interval $[0, 2\hat{p}_n]$ will not be very helpful since it contains any value arbitrarily close to zero. It seems therefore reasonable to require $2\sqrt{\mathbb{P}(C)/n} < \mathbb{P}(C)/2$, or equivalently

$$n > 16/\mathbb{P}(C). \tag{3.7}$$

So for very small $\mathbb{P}(C)$, we require a very large number n of Monte Carlo samples to obtain a fixed accuracy of the estimate. We can rephrase this reasoning in terms of the relative variance

$$\frac{\operatorname{Var} \hat{p}_n}{\mathbb{P}(C)^2} = \frac{1-\mathbb{P}(C)}{\mathbb{P}(C)n} \tag{3.8}$$

of the CMC estimator: the relative variance diverges to infinity as $\mathcal{O}(1/\mathbb{P}(C))$ when $\mathbb{P}(C) \to 0$. This implies that to fix the relative variance, approximately m times as many samples are required for an m times as small probability $\mathbb{P}(C)$. We conclude that the fast rate of divergence of the relative variance induces a computational inefficiency of CMC simulation for very small probabilities [58, 71, 72].

Note that in our power grid model, each Monte Carlo sample is a function of a sample path defined on many time steps. Therefore, one CMC sample already involves solving the nonlinear system of AC PFEs (3.1) – (3.2) a large number of times. For example, assuming

$\mathbb{P}(C) = 10^{-5}$, T equal to one week, and a time step of 5 minutes, (3.7) suggests using 1.6×10^6 samples, so $1.6 \times 10^6 \times 12 \times 24 \times 7 = 3.2 \times 10^9$ nonlinear systems should be solved. Assuming that one Newton-Raphson requires several milliseconds (see Table 2.2 in Section 2.1.3), this CMC estimation is expected to require more than a thousand hours. Even worse, note that many investment questions for power grids involve choosing between different scenarios or even optimizing potential investments, requiring iteration of the described Monte Carlo simulation for different potential configurations of the grid. We conclude that CMC estimation is computationally too intensive for grid reliability analyses.

3.3 RARE EVENT SIMULATION

The computational burden of CMC simulation for grid reliability can be reduced by extending CMC with a rare event simulation technique. The rest of this chapter will be an introduction to rare event simulation, as we will apply one such technique called splitting to our problem in Chapters 5, 6 and 7.

Rare event simulation techniques have been developed to accurately estimate very small probabilities with moderate computational effort. The research area has found various applications, for example in telecommunications, molecular biology and insurance risk analysis [19, 24, 35]. Applications of rare event simulation to power systems have emerged only recently and in limited number [6, 8, 38, 43, 47, 51, 66, 75, 76, 77]. Rare event simulation techniques extend CMC simulation in an attempt to reduce the variance of the CMC estimate using a similar workload [72]. By including extra information on the event of interest, one may achieve a very large variance reduction of orders of magnitude — millions is not uncommon in case the probability is very small. The next two subsections will treat two main variants of rare event simulation: importance sampling and (multilevel) splitting.

In importance sampling, samples are drawn from a chosen alternative distribution instead of from the original distribution [71]. To correct for the induced bias, each realization is multiplied by a weighting factor

before the average is computed. To achieve significant variance reduction, one should find an alternative distribution that substantially increases rare event occurrences.

In contrast, splitting techniques replicate sample paths whenever the rare event is presumed substantially more likely given the current state [31, 50]. Crucial for variance reduction is a suitable choice for a function that maps system states to this probability. Although conceptually distinct, both variants similarly advance variance reduction by including information on typical paths to the rare event set.

3.3.1 *Importance sampling*

Importance sampling is a Monte Carlo variance reduction technique [58, 71, 72]. We assume that $f : \mathcal{S} \to [0, \infty)$ is the continuous probability density function (pdf) of random variable X with values in $\mathcal{S}$. Importance sampling is based on the following identity:

$$\mathbb{E}_f[H(X)] = \int_{\mathcal{S}} H(x) f(x) dx = \int_{\mathcal{S}} \frac{H(x) f(x)}{g(x)} g(x) dx = \mathbb{E}_g \left[H(Y) \frac{f(Y)}{g(Y)} \right].$$

Here, $g : \mathcal{S} \to [0, \infty)$ a pdf satisfying $g(x) > 0$ for all $x \in \mathcal{S}$ and random variable $Y \sim g$. This identity suggests to sample random variables Y_i for $i = 1, \ldots, n$ from the alternative density g and to correct each realization $H(Y_i)$ by the likelihood ratio $f(Y_i)/g(Y_i)$ before averaging. The resulting Importance Sampling estimator is

$$\hat{J}[g]_n = \frac{1}{n} \sum_{i=1}^{n} H(Y_i) \frac{f(Y_i)}{g(Y_i)}, \quad \text{with i.i.d. } Y_1, \ldots, Y_n \sim g. \tag{3.9}$$

As $\mathbb{E}[\hat{f}[g]_n] = \mathbb{E}[H(Y_1)f(Y_1)/g(Y_1)] = \mathbb{E}[H(Y)f(Y)/g(Y)] = \mathbb{E}[H(X)]$, (3.9) is an unbiased estimator for $\mathbb{E}[H(X)]$. We want to choose g such that

$$\begin{aligned}\operatorname{Var}\left(\hat{f}[g]_n\right) &= \frac{1}{n}\operatorname{Var}\left(\frac{H(Y)f(Y)}{g(Y)}\right)\\ &= \frac{1}{n}\left(\mathbb{E}\left[\frac{H^2(Y)f^2(Y)}{g^2(Y)}\right] - \mathbb{E}^2[H(X)]\right)\\ &= \frac{1}{n}\left(\int_{\mathcal{S}} \frac{H^2(x)f^2(x)}{g(x)}dx - \mathbb{E}^2[H(X)]\right) \end{aligned} \tag{3.10}$$

is small. It is possible to define an alternative density $g^* : \mathcal{S} \to [0,\infty)$ for which the corresponding importance sampling estimator has zero variance:

$$g^*(x) := \frac{H(x)f(x)}{\int_{\mathcal{S}} H(x)f(x)dx} = \frac{H(x)f(x)}{\mathbb{E}[H(X)]}. \tag{3.11}$$

Indeed, g^* is a pdf and for $g(x) = g^*(x)$ the variance in (3.10) becomes

$$\operatorname{Var}\left(\hat{f}[g^*]_n\right) = \frac{1}{n}\left(\left[\int_{\mathcal{S}} H(x)f(x)dx\right]^2 - \mathbb{E}^2[H(X)]\right) = 0. \tag{3.12}$$

For this reason, $\hat{f}[g^*]_n$ is called the zero-variance estimator. One sample will already yield the exact value. In fact, if $H(X) = \mathbf{1}_A$ for event $A = A(X)$, the value of interest is probability $\mathbb{E}[H(X)] = \mathbb{P}(A)$ and g^* denotes the conditional pdf of random variable X given A. However, constructing g^* requires knowing the value of interest $\mathbb{E}[H(X)]$, which would defeat the point of simulation. Nevertheless, the lesson is to find a pdf g close to g^*.

Methods such as the Cross Entropy Method or the Variance Minimization Method search a suitable alternative pdf for importance sampling [73]. Assuming a certain type of parametric distribution, these methods iteratively attempt to improve distribution parameters such that rare events occur more frequently.

3.3.2 *Multilevel splitting*

The scope of a splitting technique is — unlike that of importance sampling — limited to a Markov process $\{X_t, t \geq 0\}$ with values in $\mathcal{X}$. Often the rare event of interest is defined as X_t entering a specified rare event set $B \subset \mathcal{X}$ before a stopping time ζ. The probability $\mathbb{P}(C)$ in our context of power grid reliability fits this framework (see (3.3)). We define

$$\tau_B = \inf\{t > 0 : X_t \in B\}$$

as the first entrance time in B and $x_0 \notin B$ as the initial state. We want to estimate the rare event probability

$$\gamma := \mathbb{P}(\tau_B < \zeta | X_0 = x_0).$$

Two typical choices for ζ are a fixed finite time $\zeta = T$ and stopping time $\zeta = \tau_A := \inf\{t > 0 : \exists\ 0 \leq s < t \text{ s.t. } X_s \notin A \text{ and } X_t \in A\}$, the first time the process reenters A, where $A \subset \mathcal{X}$ is a set that contains x_0.

Details of the (multilevel) splitting, or importance splitting, can be found in Rubino and Tuffin [71], L'Ecuyer et al. [50], Garvels [31], Botev and Kroese [9]. The key ingredient of any splitting technique is the importance function

$$h : \mathcal{X} \mapsto \mathbb{R}$$

that assigns a real value to each state X_t. It is constructed such that $h(x) \geq 1$ precisely when x corresponds to a rare event occurrence and $h(x_0) = 0$. Note that a time-dependent rare event set fits this framework simply by viewing the time as a dimension of the state space. The interval $[0, 1]$ is divided into m subintervals with intermediate thresholds $0 = l_0 < l_1 < \cdots < l_m = 1$. Let $T_k = \inf\{t > 0 : h(X_t) \geq l_k\}$ be the first time of entering the k-th level and $D_k = \{T_k < T\}$ the event that the k-th level

is hit during $[0, T]$. Obviously, $\mathbb{P}(D_m)$ is the value of interest as it is equal to γ. Also, $\mathbb{P}(D_0) = 1$. Since $D_m \subset D_{m-1} \subset \cdots \subset D_0$, we can write

$$\begin{aligned}\gamma &= \mathbb{P}(D_m) = \mathbb{P}(D_m \cap D_{m-1}) = \mathbb{P}(D_m|D_{m-1})\mathbb{P}(D_{m-1}) \\ &= \dots \\ &= \prod_{k=1}^{m} \mathbb{P}(D_k|D_{k-1}).\end{aligned}$$

That is, γ is a product of m conditional probabilities $p_k := \mathbb{P}(D_k|D_{k-1})$, which we will estimate separately. Independent sample paths from the conditional distribution of the entrance state $(T_{k-1}, X_{T_{k-1}})$ given D_{k-1} would give us an estimate for p_k. However, we do not know this distribution for levels $k > 1$, and we use its empirical distribution instead, obtained from samples of the previous level. We proceed recursively in this way, and at each level k we estimate p_k by the proportion $\hat{p}_k$ of sample paths for which D_k occurs (see Figure 3.1). Then the product

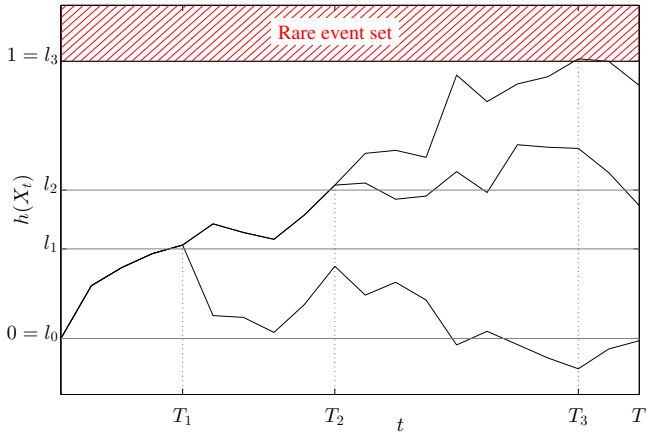


Figure 3.1: A minimal example of a splitting simulation. Source: [92]

$$\hat{\gamma} := \prod_{k=1}^{m} \hat{p}_k$$

is an unbiased estimator for γ for several variants of the splitting technique, for example Fixed Splitting (FS), Fixed Effort (FE) and Fixed Number of Successes (FNS) [3, 22, 2]. These three variants differ in which of the following variables are deterministic constants, and which are random:

- The number c_k of offspring of each sample path that entered level k.
- The total number n_k of sample paths at level k (the number of trials).
- The number r_k of sample paths that entered level k (the number of successes).

Here we give a brief description of the three splitting variants mentioned above.

1. *Fixed Splitting*
   Fixed Splitting determines the number of offspring $c_k \geq 1$ per level hit in advance. That is, we start a specified number n_0 of sample paths, resulting in r_1 samples that have exceeded threshold l_1, with $r_1 \geq 0$ a random integer. Then c_1 sample paths are continued from each of these r_1 entrance states. After repeating this at all levels k we compute the conditional probability estimates

   $$\hat{p}_k = \frac{r_k}{c_{k-1} r_{k-1}} \quad \text{for } k = 1, \ldots, m,$$

   where $c_0 := 1$ and $r_0 := n_0$. Obviously, $n_k = c_k r_k$ is also random. A proof that the estimator $\hat{\gamma} = \prod \hat{p}_k$ is unbiased can be found in Asmussen and Glynn [3]. An advantage of this variant is that we can simulate each sample path up untill stopping time ζ independent of the other sample paths. Furthermore, as all n_0 sample path trees are independent, one can compute one Fixed Splitting estimator from each tree and compute the sample variance of the mean of these n_0 estimates. In this way, the accuracy of the Fixed Splitting estimate can easily be assessed. A disadvantage of Fixed Splitting is that paths may hit the next level many times or never, especially

when the number of levels and the thresholds are far from optimally chosen. The former situation results in a high computational burden, the latter situation in a useless estimate $\hat{\gamma} = 0$.

2. *Fixed Effort*
   Fixed Effort mitigates the risks of path explosion and extinction by fixing the total number n_k of samples at each level k. Again we start a specified number n_0 of sample paths, resulting in r_1 samples that have exceeded threshold l_1, with $r_1 \geq 0$ a random integer. At the next level we start n_1 sample paths, each continuing from an entrance state chosen uniformly random from the r_1 entrance states. We iterate this for all levels and compute the conditional probability estimates

   $$\hat{p}_k = \frac{r_k}{n_{k-1}} \quad \text{for } k = 1, \ldots, m.$$

   Note that c_k is not defined since the number of offspring at level k may vary per level entrance. A proof that the estimator $\hat{\gamma} = \prod \hat{p}_k$ is unbiased can be found in Del Moral and Garnier [22]. Fixed Effort suffers from the risks of path explosion and extinction to a lesser extent than Fixed Splitting. Nevertheless, as the realized number of hits may differ from the expected number of hits, the risks persist.

3. *Fixed Number of Successes*
   The FNS variant fixes the number of hits r_k for all levels k. That is, we start sample paths until a specified number r_1 of samples have exceeded threshold l_1. As a result now the number of required samples $n_0 \geq r_1$ is now a random integer. At the next level, we choose one of the r_1 entrance states at random and continue the sample path from this state. We repeat this until we observed r_2 paths that exceeded l_2, and so on. We iterate this for all levels and compute the conditional probability estimates

   $$\hat{p}_k = \frac{r_k - 1}{n_{k-1} - 1} \quad \text{for } k = 1, \ldots, m.$$

A proof that the estimator $\hat{\gamma} = \prod \hat{p}_k$ is unbiased can be found in Amrein and Künsch [2]. The advantage of this variant is that the accuracy of the estimator is controlled by adjusting the computational effort.

Crucial for variance reduction of a splitting estimator is choosing a suitable importance function [31, 50]. The importance function should 'reward good behavior' by splitting trajectories that are more likely to hit the rare event set. Glasserman et al. [36] show that ideally, the most likely path to the rare event set should coincide with the most likely path to any intermediate level. In the setting of a stationary Markov process and $\zeta = \tau_A$ a stopping time as defined at the start of this section, Garvels et al. [32] suggest using the importance function

$$h^*(x) := \mathbb{P}(\tau_B < \tau_A | X_0 = x). \tag{3.13}$$

Since knowing $h^*(x)$ would defeat the point of using simulation, the lesson is to find an importance function that is close to (3.13). This is very similar to the case of the zero-variance estimator of importance sampling $\hat{J}[g^*]_n$ (see (3.11)).

In a setting of a stationary process with a noise term scaled by a rarity parameter ε, Dean and Dupuis [21] derive sufficient conditions for an asymptotically optimal performance of a given importance function $h(x)$. The probability $\gamma^\varepsilon = \mathbb{P}(\tau_B < \tau_A | X_0 = x)$ is assumed to satisfy a large deviations scaling. Under appropriate conditions, the asymptotic decay rate of the second moment of the splitting estimator $\hat{\gamma}^\varepsilon$ is optimal if

$$\lim_{\varepsilon \downarrow 0} -\varepsilon \log \mathbb{E}[(\hat{\gamma}^\varepsilon)^2] = 2W(x),$$

with $W(x)$ the minimum good rate function of the stochastic process starting in x. In fact, Dean and Dupuis [21] show that to achieve this asymptotic optimality, $1 - h(x)$ should be proportional to a subsolution of the Hamilton-Jacobi-Bellman equations associated with the calculus of variations problem that $W(x)$ solves. This relation to optimal control theory goes beyond the scope of this thesis, but we will use the results to justify a choice for an importance function in Chapter 7.

Other results on the efficiency of splitting can be derived under the assumption that the probability of hitting the next level does not depend on the current entrance state

$$\mathbb{P}\left(D_k | D_{k-1}, (T_{k-1}, X_{T_{k-1}})\right) = p_k \quad \forall\ (T_{k-1}, X_{T_{k-1}}). \tag{3.14}$$

Under this assumption and using the optimal choice for m and p_k Amrein and Künsch [2] show that the squared relative error of the FNS and FE estimators is proportional to

$$\frac{\mathrm{Var}(\hat{\gamma})}{\gamma^2} \propto \frac{(\log \gamma)^2}{n}. \tag{3.15}$$

Here n denotes the total number of sample paths. Hence the relative variance (3.15) diverges as $\mathcal{O}((\log \gamma)^2)$ when $\gamma \to 0$. This squared logarithmic divergence is slower than the divergence of the CMC squared relative error in (3.8). This illustrates the potential computational gain of a splitting estimator compared to a CMC estimator.

3.4 CONCLUSION AND OUTLINE OF NEXT CHAPTERS

In this chapter, we discussed modeling uncertain power injections as stochastic processes to assess power grid reliability. Direct derivation of various grid reliability indices is not possible if we assume the AC power flow equations. We described how to estimate these indices using a Crude Monte Carlo simulation.

Chapter 4 contains the material of Wadman et al. [90]. We present a probabilistic power flow model subject to the connection temperature constraints as in (2.24). We model wind power as a modified ARMA model and we employ CMC simulation to simulate constraint violations. In contrast to conventional models that enforce connection current constraints as in (2.25), short-term current overloading is allowed. Temperature constraints are weaker than current constraints, and hence the proposed model quantifies the overload risk more realistically. Using such a constraint is justified the more by the intermittent nature of the renewable power source. Several IEEE test case examples illustrate

the more realistic reliability analysis. An example shows that a current constraint model may overestimate these risks, potentially leading to unnecessary over-investments by grid operators in grid connections.

We illustrated the computational inefficiency of CMC simulation in Section 3.2. We introduced two categories of rare event simulation in Section 3.3. For complex models it will often be challenging to decide a priori whether importance sampling or splitting is computationally more efficient. In this research, using importance sampling to estimate grid reliability indices immediately poses the problem of choosing a suitable alternative distribution. Instead, we applied a splitting technique by heuristically choosing an importance function, as described in Chapter 5 (based on Wadman et al. [91]). For simple power grids, grid reliability indices (1.2) – (1.5) are estimated. For a fixed squared relative error of index estimators, orders of magnitude less workload is required than when using an equivalent Crude Monte Carlo method. We show further that a bad choice for the time step size or for the importance function can increase this squared relative error.

For a grid with many stochastic power injections this heuristic importance function replicates sample paths that are in fact relatively unlikely to hit the rare event set. To overcome this issue, we focus on the splitting technique itself. This resulted in Wadman et al. [92] which is described in Chapter 6: we address the issue of using splitting for rare event sets consisting of separated subsets. We propose to mitigate the problem of replicating unpromising sample paths by estimating the probabilities of the subsets separately using a modified splitting technique. We compare the proposed separated splitting technique with a standard splitting technique by estimating the probability of entering either of two separated intervals on the real line. The squared relative error of the estimator is shown to be significantly higher when using standard splitting than when using separated splitting. We show that this difference increases if the rare event probability becomes smaller, illustrating the advantage of the separated splitting technique.

Chapter 7 is based on Wadman et al. [93]. We derive asymptotic results on overload probabilities of connections assuming the linear DC power flow equations. We develop an importance function based on large

deviations results to compute splitting estimates for connection overload probabilities. Using typical IEEE-14 test cases with up to eleven stochastic power injections we show that for a fixed accuracy Crude Monte Carlo requires tens to millions as many samples as the proposed splitting technique. We investigate the balance between accuracy and workload of three numerical approximations of the importance function. We justify the workload increase of large deviations based splitting compared to a naive one based on merely the Euclidean distance to the rare event set: for a fixed accuracy naive splitting requires over 60 times as much CPU time as large deviation based splitting. In these examples naive splitting — unlike large deviations based splitting — requires even more CPU time than CMC simulation, demonstrating its pitfall. This illustrates the relevance of the choice for the importance function.

Part III

SIMULATION METHODS TO ASSESS POWER GRID RELIABILITY

4

PROBABILISTIC POWER FLOW SIMULATION ALLOWING TEMPORARY CURRENT OVERLOADING

4.1 INTRODUCTION

Renewable energy generation is increasingly integrated, but high penetration of renewable generators is expected to strain the power grid. The limited predictability of distributed renewable sources implies that substantial implementation in the grid will result in a significantly increased risk of power imbalances. Uses of storage, trade or unit commitment may mitigate these risks. Above all, a quantitative uncertainty analysis of the power flow has to be performed, which is the topic of this chapter.

An electricity network should fulfill the following constraints:

- The absolute voltage should be between acceptable bounds at all nodes. Formally stated,

  $$V_{\min} < |V(t)| < V_{\max}$$

  should hold at all nodes for all times t.

- The reactive power should be between acceptable bounds at all generation nodes:

  $$Q_{\min} < Q(t) < Q_{\max},$$

  should hold at all nodes for all times t.

- The temperature of each connection should be bounded:

  $$T(t) < T_{\max}, \tag{4.1}$$

This chapter is based on the conference paper Wadman et al. [90].

should hold at all node connections for all times t. $T_{\max}$ is assumed to be the critical temperature of the connection above which operation failure or degradation over time may occur.

A straightforward method to satisfy the latter constraint, is to ensure that the current never exceeds a certain maximum. That is,

$$|I(t)| < I_{\max}, \tag{4.2}$$

should hold at all node connections for all times t. In this chapter, we assume that $I_{\max}$ corresponds to $T_{\max}$ in the sense that if $I(t) = I_{\max}$ for all times t, then

$$\lim_{t\to\infty} T(t) = T_{\max}.$$

These maxima depend on the material and thickness of the connection. Tables displaying this correspondence for cables can be found in XLPE [99], for example.

However, the transient temperature adjustment incurs some lag time, so a mild violation of a given current maximum—with a short duration—may not lead to violation of the temperature constraint. Hence, directly imposing the current constraint may be too restrictive. In fact, the grid dimensioning should anticipate the most extreme event, which may very well be accidental and of short duration. Underestimating the connection capacities in this way, may lead to over-investments in grid connections. Therefore, this chapter will treat an improved "soft" current constraint, which basically demands that the current be not too high for too long, by focusing on constraint (4.1) instead of (4.2).

To include renewable generation units, one must model their uncertain nature. The choice of model should be consistent with available data. Often, and especially when considering investments in new infrastructure, power generation data are scarce, and data of their meteorological sources (e.g. wind speed, solar radiation) are preferred because of their wide availability. Further, the power generation and therefore the connection currents exhibit time correlation. This means that checking for short-term current overloading necessitates the inclusion of chronology in our

model, which discourages the choice for frequency domain approaches [64, 55, 26]. Instead, we prefer a model which involves time correlation of the meteorological sources.

A second reason for proposing a time domain based model is the possible inclusion of storage devices. In order to know the storage capacity and maximum power at some time step, the state of charge information is required. This information will depend on the device behavior at the previous time step, again necessitating the introduction of chronology into the model. Since storage is one of the main solutions proposed to mitigate the very problem of highly variable renewable power generation, the possibility to extend the model with storage is a welcome feature. Furthermore, we will show that the theoretical benefit of mitigation will be underestimated by use of the current constraint, which implies that storage mitigation is even more promising.

Monte Carlo techniques are one way to quantify the risk of violating the three mentioned constraints. In a straightforward approach, one would first sample the meteorological source. Then the corresponding power injection would be used in a steady state power flow problem. In this way, many power flow solution samples are drawn, after which the risk of constraint violation can be estimated statistically.

This chapter elaborates on this approach, using wind power as the straining renewable resource. First, Section 4.2.1 presents a time integration scheme for the dynamic connection temperature. Section 4.2.2 describes a stochastic wind power simulation method. In Section 4.2.3, we investigate an efficient solver for the steady state power flow problem. Simulation results are presented in Section 4.3. After proposing possible extensions in Section 4.4, we conclude this chapter in Section 4.5.

4.2 METHODOLOGY

4.2.1 *Short-term overloading*

Short-term overloading may warm up a connection insufficiently to increase the temperature to dangerous levels. In fact, the actual quantity

to be controlled is the connection temperature $T(t)$, and not the current itself. Fortunately, as is well-known [65], the transient temperature of the connection is described by a first order ordinary differential equation:

$$\tau \frac{\mathrm{d}\Theta(t)}{\mathrm{d}t} + \Theta(t) = \frac{|I(t)|^2}{I_{\max}^2}, \tag{4.3}$$

with

$$\Theta(t) = \frac{T(t) - T_0}{T_{\max} - T_0}.$$

Here, T_0 denotes the ambient temperature and $I(t)$ the current. The other three coefficients are determined by the connection properties: τ denotes the thermal time constant for the heating of the conductor, whereas $T_{\max}$ and $I_{\max}$ are as defined in Section 4.1.

The solution of (4.3) is obtained by direct integration:

$$T(t) = T_0 + \frac{T_{\max} - T_0}{\tau I_{\max}^2} \int_0^t |I(s)|^2 e^{(s-t)/\tau} ds.$$

To qualitatively demonstrate to what sense a temperature constraint weakens the current constraint, let us first assume a constant current $I(t) \equiv I$. In this case, the formula above simplifies to

$$T(t) = T_0 + \frac{|I|^2}{I_{\max}^2} \left(T_{\max} - T_0\right) \left(1 - e^{-t/\tau}\right).$$

Practically, this equation states that in order to satisfy constraint (4.1), one requires

$$1 - \frac{I_{\max}^2}{|I|^2} < e^{-t/\tau} \quad \forall\, t.$$

This inequality naturally shows that no excessive temperature can occur as long as $|I| < I_{\max}$. Otherwise, I is allowed to take on some (constant) value higher than $I_{\max}$ for a maximum duration of

$$-\tau \ln\left(1 - \frac{I_{\max}^2}{|I|^2}\right),$$

as long as the current subsequently drops below $I_{\max}$.

In reality, $I(t)$ is neither constant in time nor known analytically, so we cannot find the analytic solution of (4.3). However, suppose that we obtain a numeric sample path for $I(t)$. Then we can construct a corresponding sample path for the temperature, by discretizing (4.3):

$$\tau \frac{\Theta_t - \Theta_{t-\Delta}}{\Delta} + \Theta_{t-\Delta} = \frac{|I_t|^2}{I_{\max}^2}. \tag{4.4}$$

Here, Θ_t and I_t denote the numerical approximation for $\Theta(t)$ and $I(t)$, respectively, and Δ is the time step. Solving this equation for Θ_t yields a numerical scheme for the relative temperature $\Theta(t)$, and thus for the absolute temperature $T(t)$. In order to fulfill the temperature constraint, $\Theta_t < 1$ should hold for all t.

4.2.2 *ARMA based wind power model*

In this chapter, we will choose wind power as the intermittent power resource. First, to check the time dependent temperature constraint (4.1), we require a time domain for the wind speed model. Second, the model should capture the wind speed distribution as observed in nature, which is assumed to be the Weibull distribution. Furthermore, to reflect inertia and recurrence of meteorological systems, spatial correlation between meteorological sources as well as temporal periodicity should be incorporated. The autoregressive moving-average (ARMA) model is a well-known technique to fulfill these requirements.

The authors of [53] elaborate on an ARMA-GARCH wind speed time series model and demonstrate that the simulated times series realistically represent wind speed observations. For simplicity, we use an ARMA model, thus using the model in Lojowska et al. [53] except that homoscedasticity is assumed. The autoregressive moving-average model captures the time correlations naturally. The Weibull distributed nature of the wind speed is preserved: the input wind data are first transformed from Weibull realizations to standard normal realizations. On these transformed data, an ARMA(1,1) model is fitted. Parameters are

estimated using a standard statistical tool in MATLAB (armax). New time series samples, simulated from this model, are then transformed back to Weibull samples. The daily periodicity is automatically attained by fitting different Weibull cdfs to each hour of the day. The yearly periodicity can be incorporated as well, but is neglected as we consider time series of no longer than one month.

Spatial dependency of wind speeds at different nodes is estimated from the residuals as fitted to the transformed data. The model in turn imposes this dependency by simulating correlated white noise terms: consider the vector of white noise terms $Y \in \mathbb{R}^m$ at a specific time step of a specific Monte Carlo sample, where its elements correspond to all m wind farm locations in the network. Suppose first that we desire the white noise to be a multivariate normally distributed random variable with zero mean:

$$Y \sim \sigma\mathcal{N}(0, \Sigma),$$

with $\sigma > 0$ the desired standard deviation and $\Sigma \in \mathbb{R}^{m\times m}$ the correlation matrix exhibiting the spatial wind speed dependence. Then we sample the multivariate standard normal random variable $Z \sim \mathcal{N}(0, I)$ with independent elements, and perform a Cholesky decomposition $\Sigma = LL^\top$, with $L \in \mathbb{R}^{m\times m}$ lower triangular. By setting

$$Y := \sigma LZ,$$

Y will indeed be a multivariate normally distributed with mean zero, standard deviations σ and correlation matrix Σ. Alternatively, we may desire a multivariate Student's t-distributed random variable Y_k as white noise, where all elements have k degrees of freedom and where the same dependence structure is assumed. In this case, we extend the above procedure by independently sampling a chi-squared distributed random variable v with k degrees of freedom, and set

$$Y_k := Y\sqrt{k/v}.$$

Then Y_k is as desired (more details can be found in Torrent-Gironella and Fortiana [85]).

One month of hourly wind speed measurements from the KNMI [49] are used as data. The wind speed at each hour is estimated by the last 10 minutes mean wind speed of the previous hour, in open landscape at 10 meters height. For a specific wind turbine, the relation between the wind speed and the wind power is known, as illustrated in Fig. 4.1. We

Figure 4.1: Wind power as function of the wind speed.

transform wind speed time series by use of this function, thus obtaining wind power time series.

4.2.3 *Accelerated power flow method*

In order to achieve a satisfactory accuracy level for a connection reliability analysis, one should use a realistic time frame as well as a sufficient number of Monte Carlo samples. Then, for each time step and each Monte Carlo sample, a steady state power flow problem has to be solved. This means that each power flow problem should be solved reasonably fast. This requirement will drive the choice of power flow method.

A steady state power flow problem involves the solution of the power balance equations:

$$P_i = \sum_j^N |V_i||Y_{ij}||V_j| \cos(\psi_{ij} + \delta_j - \delta_i), \tag{4.5}$$

$$Q_i = -\sum_j^N |V_i||Y_{ij}||V_j| \sin(\psi_{ij} + \delta_j - \delta_i). \tag{4.6}$$

Here, $P_i, Q_i \in \mathbb{R}$ denote the active and reactive power, respectively, injected at node i. $|V_i|, \delta_i \in \mathbb{R}$ denote the voltage magnitude and angle, respectively, in grid node i. $|Y_{ij}|, \psi_{ij} \in \mathbb{R}$ denote the absolute value and angle, respectively, of the connection admittance between nodes i and j. N is the number of grid nodes. This nonlinear system of equations has to be solved for the state vectors $|\boldsymbol{V}|$ and $|\boldsymbol{\delta}|$, which is normally done using a Newton-Raphson method [39].

The Fast Decoupled Power Flow (FDPF) method [81] speeds up the conventional method, mainly by assuming approximations which ensure that the Jacobian depends on the admittance matrix Y only. This implies that the Jacobian will be constant in the Newton-Raphson iteration number, and it thus has to be inverted only once. This feature is particularly beneficial in our proposed Monte Carlo method, since the inverse can be reused for all samples.

Elements of the admittance matrix Y are zero precisely when there is no edge between the corresponding nodes. The number of edges in a typical power grid is on the order of the number of grid nodes. This means that Y is typically sparse, which can be used to accelerate computations, as explained in Section 2.1.3.

Another acceleration for the power flow method involves the power flow solution from the previous time step. Since the amount of renewable power is a piecewise continuous function of time, one may expect that two subsequent solutions will be close. Therefore, the previous solution will be a reasonable first guess for the current problem.

The three acceleration techniques discussed above (i.e. use of FDPF method, sparse computations, and smart initialization) significantly speed up the Newton-Raphson iteration loop. Table 4.1 gives an impression of

the CPU times of some standard IEEE-test cases: the test case number corresponds to the number of grid nodes. Experiments were performed in MATLAB R2011a on an Intel Core i7 CPU M 640 2.80GHz, 2.79 GHz, 3.24 GB of RAM. All average CPU times are based on 1000 trials, and a

IEEE-test case (# nodes)	14	30	57	118	300
Conventional power flow	0.96	1.66	4.0	12.3	116.4
FDPF, sparse	0.72	0.78	1.5	2.2	11.6
FDPF, sparse, smart initialization	0.57	0.76	1.5	1.1	11.9

Table 4.1: The average CPU time (ms) of a sophisticated power flow method is on the order of milliseconds.

Newton-Raphson tolerance error of 10^{-5} is used. The table clearly shows that a sparse FDPF method is accelerating the conventional power flow method, especially for large grids. Smart initialization may yield some further acceleration, depending on the test case.

We conclude that the computational time for a steady state power flow is on the order of milliseconds. This order of magnitude is desirable, since an accurate uncertainty analysis requires a large number of Monte Carlo samples, each of which involves as many steady state power flow problems as the number of time steps.

4.3 RESULTS

4.3.1 *Comparison between current and temperature constraint*

To demonstrate the use of the temperature constraint in a time domain based model, we consider the IEEE-14 test case [87]. The conventional generators at nodes 3 and 6 are replaced by wind farms with comparable rated power (4 base MVA). Wind power time series samples are generated 1000 times, on an interval of one month, on an hourly basis, using spatially correlated KNMI wind speed measurements during August 2011 at Valkenburg and IJmuiden, the Netherlands. Consumption is assumed constant in time. For simplicity, we choose $I_{\max} = 3.7 I_{base}$ uniformly at all

connections. Precisely this value is used since then the current exceeds this maximum at some connection approximately once a year. We choose $\tau = 3$ hours (see [25] for realistic values of the thermal time constant).

	ϕ	θ
Valkenburg	0.94	-0.34
IJmuiden	0.93	-0.15

Table 4.2: The autoregressive coefficient ϕ and moving-average coefficient θ of the ARMA(1,1) models at the two wind nodes.

In our results, the current overloading occurs most of the times at the same connection during periods of high values of wind power generation. Fig. 4.2 shows an example of temporary overloading at this critical connection, when the temperature constraint is not violated. One can

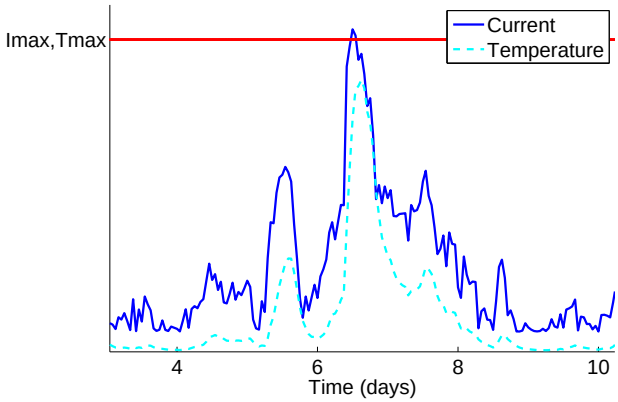


Figure 4.2: Example: temporary current overloading, which is allowed since the temperature constraint is fulfilled. τ = 3.

see that the temperature time series is indeed following the current time series. However, local temperature peaks are lower, less frequent and smoother than local current peaks, and slopes are more gradual. This illustrates the "softness" of temperature constraint (4.1) compared to current constraint (4.2).

In the upper graph of Fig. 4.3, all 1000 current time series samples at the critical connection are displayed. In the lower graph of the same figure, the corresponding temperature time series are displayed. One

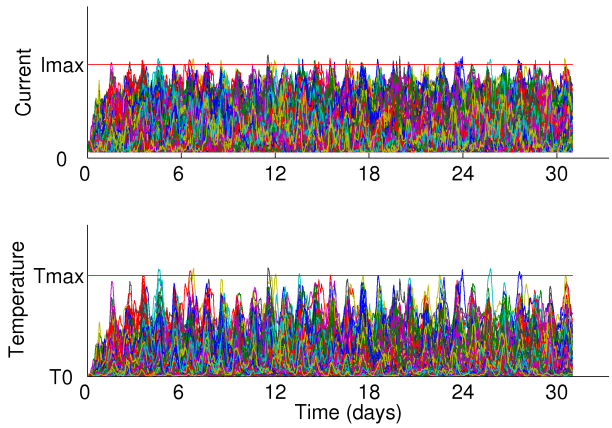


Figure 4.3: The current and temperature at the critical connection, 1000 time series samples, $\tau = 3$.

can see from this figure that the current and temperature indeed exceed their maximum only rarely. The graph magnification in Fig. 4.4 clearly illustrates that a current overload does not necessarily imply excess temperature at this connection. This result can be extended to the other connections. In fact, in total 88 current violations were incurred over all samples, which indeed corresponds to approximately once a year. In contrast, the temperature exceeds $T_{\max}$ only 6 times. Other IEEE test cases yield similar results, as can be seen in Table 4.3.

Test cases	IEEE-14	IEEE-30	IEEE-57	IEEE-118
Current Violations	88	69	152	101
Temperature Violations	6	16	20	16

Table 4.3: Number of constraint violations for different IEEE test cases. 1000 time series samples of 1 month, $\tau = 3$

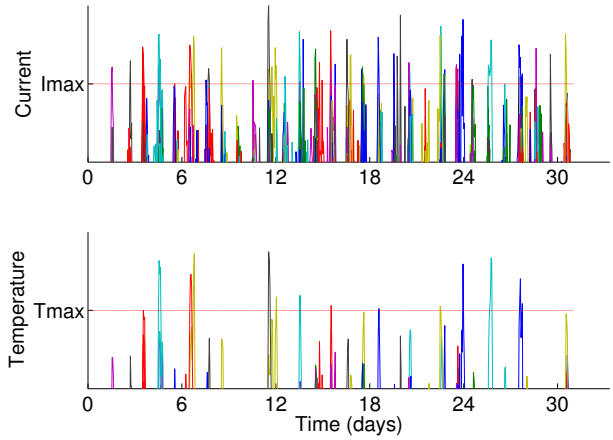


Figure 4.4: Magnification of Fig. 4.3: the temperature constraint is violated less frequently than the current constraint.

4.3.2 *Sensitivity to the thermal time constant*

It is clear that the higher the thermal time constant τ, the more the grid capacity will be underestimated when checked by use of current constraints. Table 4.4 shows a quantification estimate of this sensitivity. We repeated the simulation of the previous subsection for different values

τ (hours)	1	1.5	2	3	4	5	6
Current Violations	119	80	100	88	95	100	101
Temperature Violations	119	50	25	6	3	0	0

Table 4.4: Number of constraint violations in the IEEE-14 test case as function of τ. 1000 time series samples of 1 month.

of τ. The table suggests that our proposed model will yield a significantly more accurate reliability estimate for $\tau > 1.5$. For values of τ close to

the time step $\Delta = 1$, our discrete model loses its ability to detect any differences. To explain this, note that for $\tau \to \Delta$, equation (4.4) goes to:

$$\Theta_t = \frac{|I_t|^2}{I_{\max}^2},$$

and thus becomes independent of the previous time step. This model phenomenon is partly realistic. On the one hand, the decreasing difference between the two constraints indeed corresponds to reality: τ reflects the time the temperature requires to reach $1 - 1/e = 63.2\%$ of its asymptotic value, in case of constant current. So for small values of τ, the temperature will be close to its asymptotic value, which will cause the two constraints to agree. On the other hand, the total agreement between current and temperature constraints is an overestimation. Current peaks with a duration less than the time step size do not necessarily violate the temperature constraint in reality, in contrast to our model which regards the current as constant during one time step Δ. Therefore, the number of temperature violations in Table 4.4 is overestimated. Since hourly based data limit us to a time step of one hour, this overestimation cannot be reduced by choosing a smaller time step size. It therefore makes no sense to choose $\tau < \Delta$ in the model, whereas the overestimation can be reduced by acquiring data with a smaller time step.

4.4 FURTHER RESEARCH

We aim to extend the model with distributed storage devices, in order to investigate their potential mitigating effect on variable power flows. In fact, Fig. 4.3 and Fig. 4.4 suggest that the theoretical benefit of grid mitigation is expected to be substantially higher when estimated using the temperature constraint rather than the current constraint. Specifically, the mean current of all time series is 24% of $I_{\max}$, whereas the mean temperature is only 8% of $T_{\max}$. In other words, the peaks that can be mitigated by use of decentralized storage are relative to the mean even more extreme than conventionally estimated. This implies that mitigation can theoretically increase the connection ampacity by an even higher factor than estimated using the current constraint. Note that mitigation

during one time step increases this ampacity even more (although to an extent which is not estimable by our model).

Since storage devices produce and consume, and both to varying degrees, the uncertain nature of their strategies makes such an extension challenging. Further, we aim to increase the efficiency of the Monte Carlo technique, to achieve higher accuracy with the same number of simulations. We already explained the computational intensity of the proposed model, so an extension with storage devices will definitely necessitate an increase of computational efficiency.

A time frame of one year may simply be incorporated in the model by iterating the work of this chapter twelve times. In this way, the model will automatically exhibit approximate yearly periodicities, since each month model will be calibrated separately.

Other forms of renewable generation may be included in the model as well. Suppose that the characteristics of the considered meteorological source are known, data are available and the relation between the source and power parameters is known. Then one may try to fit an ARMA model and simulate power generation as done in Section 4.2.2. Note that the proposed model can be applied to transportation networks as well as to distribution networks. Finally, stochastic, time-varying consumption can be analogously included.

4.5 CONCLUSION

Due to the implementation of uncertain energy generators in power grids, grid operators require quantitative uncertainty analysis of power flow. Grids should satisfy certain constraints in order to match the demand while controlling overload risks. Using a conventional current constraint for grid connections, Monte Carlo simulations underestimate the grid capacity. Instead, a temperature constraint quantifies the risk more accurately. Especially for connections with a high thermal time coefficient, the temperature constraint estimate for overloading frequency may be many times smaller than the current constraint estimate. Therefore, using

a model allowing for temporary overloading may save costs by avoiding over-investments.

5

APPLYING A SPLITTING TECHNIQUE TO ESTIMATE ELECTRICAL GRID RELIABILITY

5.1 INTRODUCTION

Contemporary societies have grown accustomed to a very reliable supply by electrical power grids. However, substantial implementation of intermittent renewable generation, such as photovoltaic power or wind power, may form a challenge for grid reliability. Power imbalances caused by generation intermittency may force grid operators to curtail power to preserve grid stability. To assess long-term grid investments or decide on short-term operational strategies, the responsible grid operator should be able to estimate grid reliability.

For this purpose, various grid reliability indices exist [7], and many depend on the probability $\mathbb{P}(C)$, where

$$C = \{\text{A power curtailment occurs during } [0, T]\} \tag{5.1}$$

denotes the event of a power curtailment during the time interval $[0, T]$ of interest. We model the uncertain energy sources as stochastic processes, discretized in time. At each time step, the mapping from the state of these processes to the outcome of C (true or false) requires solving a nonlinear algebraic system. As this mapping is defined implicitly, we can not derive $\mathbb{P}(C)$ directly, and we estimate it by a Monte Carlo (MC) simulation.

However, as power curtailments are undesirable, we may expect C to be rare. In case of a time interval T equal to one week, values for $\mathbb{P}(C) \approx 10^{-4}$ or even much smaller are not uncommon [11, 12]. Crude Monte Carlo (CMC) estimators for rare event probabilities require a large number of samples to achieve a fixed accuracy [71]. Since one CMC sample already involves solving a large number of high dimensional

This chapter is based on the conference paper Wadman et al. [91].

nonlinear systems, CMC estimation is computationally too intensive for grid reliability analyses in general.

Rare event simulation techniques have been developed to accurately estimate very small probabilities, of which importance sampling and (importance) splitting are two well-known variants. In importance sampling, one samples from an alternative distribution, whereafter the estimator is multiplied by an appropriate likelihood ratio to correct for the induced bias [71]. Crucial for variance reduction is to find a distribution that increases rare event occurrences. Adaptive importance sampling techniques [46] have been developed to recursively learn this distribution. For example, the Cross-Entropy Method (CEM) iteratively changes the distribution parameters of random variables responsible for approaching the rare event in a pilot run [73]. However, in general power grids various typical paths may lead to rare event C, especially when considering a large number of stochastic sources and a large time domain. In this case CEM changes the distribution of all corresponding random variables, wherefore the resulting alternative distribution may (counterproductively) *increase* the variance by pushing too much in the direction of C.

Splitting techniques do not change the distribution, but resample trajectories as soon as they are presumed substantially closer to the rare event [71, 31, 50]. In this way, variance reduction is achieved using an increased occurrence of rare events. In the literature splitting techniques have rarely been applied to power systems. Wang et al. [95] estimated small probabilities of instantaneous, unforeseen failures of grid components. Our work though considers the rare event of power curtailments over a certain time domain due to (and given) the uncertain nature of generation. Further, Schlapfer and Mancarella [75] estimated the probability of a transmission line temperature exceeding a critical value, using Markov processes with a discrete state space. Our model considers Markov processes with a continuous state space, and allows for the assessment of more general reliability indices.

In this chapter, we speed up an MC method for grid reliability estimation with an existing splitting technique called Fixed Number of Successes (FNS). In Section 5.2, we introduce common reliability indices for transmission power grids. In Section 5.3, we describe a CMC method that

estimates these indices. We specify a stochastic model for the intermittent energy sources, define the mapping from these sources to the outcome of a power curtailment and explain the computational intensity of this brute force approach. In Section 5.4 we extend the CMC method with the FNS technique. We investigate the computational performance of the FNS technique in Section 5.5 on an example power grid, and demonstrate how to choose the time step size and the importance function such that the estimate accuracy can be controlled.

5.2 GRID RELIABILITY INDICES

A reliability assessment of a power grid during a time interval $[0, T]$ of interest (e.g. day/week/year) involves estimating to what extent electrical constraints are violated. The power grid topology may be regarded as an undirected graph, with M edges representing connections (lines or cables), and with N nodes representing buses where power is possibly injected or extracted. Two important types of constraints [96] are absolute voltage constraints at all grid nodes

$$\mathbf{V}_{\min} < |\mathbf{V}(t)| < \mathbf{V}_{\max}, \quad \text{for all } t \in [0, T], \tag{5.2}$$

and absolute current constraints at all grid connections

$$|\mathbf{I}(t)| < \mathbf{I}_{\max}, \quad \text{for all } t \in [0, T]. \tag{5.3}$$

Here $\mathbf{V}(t)$ and $\mathbf{I}(t)$ are complex-valued vector functions of nodal voltages and connection currents, respectively, at time t, and $\mathbf{V}_{\min}, \mathbf{V}_{\max}, \mathbf{I}_{\max}$ real-valued vectors of the allowed extrema. In practice, when constraints are violated, grid operators perform corrective actions to restore stability. These actions include rescheduling of generation and curtailing power. For simplicity, we assume that a violation of (5.2) at some node or a violation of (5.3) at some connection is a sufficient (and necessary) condition for a power curtailment. That is, such constraint violation immediately affects power delivery somewhere in the grid.

Various indices exist [17, 7] to indicate the extent of load curtailments, where load refers to power consumption. These can broadly be divided

in probability, duration, frequency and severity of curtailments. The indices of interest may differ per consumer, depending on the behavior of their connected devices during and shortly after a power outage. Most conventional index definitions measure the extent of *load* curtailments only. However, in contemporary privatized energy markets grid operators must ensure power *supply* by electricity producers as well. Therefore, load-based definitions are easily generalized to power curtailments (where power corresponds to both load and supply) in general. We describe four of these indices below.

1. The Probability of Power Curtailments during $[0, T]$

$$\mathrm{PPC}(T) := \mathbb{P}(C), \tag{5.4}$$

with C as defined in (5.1).

2. The Expected Duration of Power Curtailments

$$\mathrm{EDPC}(T) := \mathbb{E}\left[D(T)\right], \tag{5.5}$$

with $D(T)$ the total duration of curtailments during $[0, T]$.

3. The Expected Number of Power Curtailments

$$\mathrm{ENPC}(T) := \mathbb{E}\left[N_C(T)\right], \tag{5.6}$$

with $N_C(T)$ the number of distinct power curtailments during $[0, T]$.

4. The Expected Energy Not Supplied accounts for the severity of the curtailment

$$\mathrm{EENS}(T) := \mathbb{E}\left[\int_0^T S_C(t)dt\right], \tag{5.7}$$

with $S_C(t)$ the size of power curtailment (in MW) at time t.

Other well-known indices account for duration, frequency and unsupplied energy *per curtailment* or *per customer*. These include SAIDI, CAIFI, SAIFI, CAIDI, denoting the System Average Interruption Duration Index, the Customer Average Interruption Frequency Index, and so on [17], and they can easily be derived from the above indices.

5.3 A CONVENTIONAL RELIABILITY ESTIMATION METHOD

To assess the risk of curtailment caused by intermittent generation, one should properly model the uncertain power generation. Typically, the meteorological source (e.g. wind speed, solar radiation) is modeled instead of the amount of power generation itself, as data of the latter are often scarce, especially when considering investments in infrastructure by estimating reliability of a future grid. A source can be modeled as a Markov process, of which the distribution should be realistic and exhibit temporal periodicities due to daily and seasonal cycles. Further, one should take into account spatial correlations between sources at different locations, as they may affect reliability indices significantly [97]. In case of wind power, Lojowska et al. [53] and Lujano-Rojas et al. [56] proposed wind speed ARMA models, whereas Wadman et al. [90] and Wangdee and Billinton [97] extended such models to the multivariate case, imposing spatial dependency.

Suppose we are interested in the value of one of the indices (5.4) – (5.7), and write it as $\mathbb{E}[I]$ for an appropriately defined random variable I (note that $\mathbb{P}(C) = \mathbb{E}[\mathbf{1}_C]$, with $\mathbf{1}$ the indicator function). Then at each time step, the function $f : S \mapsto I$ from the stochastic sources S to I requires solving the *power flow equations*

$$P_i = \sum_{j}^{N} |V_i||Y_{ij}||V_j| \cos(\theta_{ij} + \delta_j - \delta_i), \tag{5.8}$$

$$Q_i = -\sum_{j}^{N} |V_i||Y_{ij}||V_j| \sin(\theta_{ij} + \delta_j - \delta_i), \tag{5.9}$$

for each grid node $i = 1, \ldots, N$. Here, $P_i, Q_i \in \mathbb{R}$ denote the active and reactive power, respectively, injected at node i. $P_i, Q_i > 0$ indicates generation, whereas $P_i, Q_i < 0$ indicates extraction at node i. $|V_i|, \delta_i \in \mathbb{R}$ denote the voltage magnitude and voltage angle, respectively, at grid node i. $|Y|, \theta \in \mathbb{R}^{N \times N}$ denote the absolute value and angle, respectively, of the complex admittance matrix, containing the grid topology and electrical admittances of all grid connections. We desire the solution of this nonlinear algebraic system for the state vectors $|\mathbf{V}|, \mathbf{ffi}$, given vectors

P, Q. We can not solve system (5.8) – (5.9) directly, so we approximate the solution numerically using a Newton-Raphson method [39]. Using Ohm's law we can then immediately derive all other electrical quantities required to check curtailment constraints (5.2) and (5.3).

Note that P_i, Q_i are random variables if an intermittent generator is connected at node i. By nonlinearity of (5.8) – (5.9), the function f is only implicitly defined. Therefore, we can not derive $\mathbb{E}[I]$ directly, and instead we estimate it using an MC simulation. That is, we sample a realization of the discrete time Markov processes corresponding to all intermittent generators. Then at each time step, the power flow equations are solved and curtailment constraints are checked. Repeating this for all time steps yields one realization of I. The average over many such realizations constitutes a CMC estimate for the index.

To estimate power grid reliability accurately, the time interval of the MC simulation should cover the significant temporal periodicities of generation and consumption. On the other hand, the step size should be sufficiently small to address sudden changes of consumption and generation. For example, Frunt [30] states that less than 10 minutes is required in case of wind power. Hence, one MC sample may already require thousands of solutions of (5.8) – (5.9), each of which will take on the order of milliseconds (see Table 5.1). However, as power curtailments

Number of grid nodes	14	30	57	118	300
Average CPU time (ms)	0.72	0.78	1.5	2.2	11.6

Table 5.1: Average CPU time requires on the order of milliseconds to solve the power flow equations at one time step. Well-known power flow test cases are considered, and simulations are performed using MATLAB R2011a on an Intel Core 2.80GHz.

are undesirable, we may expect their occurrence to be rare. For the unbiased CMC estimator for $\mathbb{P}(C)$

$$\hat{P}_n := \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}\{C \text{ in sample } j\},$$

the *squared relative error*

$$\mathrm{SRE}\left(\hat{P}_n\right) := \frac{\mathrm{Var}\left(\hat{P}_n\right)}{\mathbb{P}(C)^2} = \frac{1-\mathbb{P}(C)}{n\mathbb{P}(C)} \tag{5.10}$$

goes to infinity as $\mathbb{P}(C)$ goes to 0. If accuracy is considered sufficient when the squared relative error is less than, say, $1/2$, we require $n > 160\,000$ CMC samples for probabilities smaller than 10^{-4}. Estimating all other indices requires a similar sample size: $n = 1/\mathbb{P}(C)$ CMC samples will on average yield one nonzero realization for $\mathbb{E}[I]$, which obviously is the bare minimum for a magnitude indication. Recalling the high computational intensity of one MC sample, we conclude that CMC estimation is not feasible in practice for general grid reliability analyses.

5.4 SPLITTING THE RELIABILITY ESTIMATION METHOD

To reduce the computational burden of the conventional reliability estimation method, we write all indices (5.4) – (5.7) as a function of a rare event probability. That is, the law of total expectation gives

$$\mathbb{E}[I] = \mathbb{P}(C)\mathbb{E}[I|C] + \mathbb{P}(C^c)\mathbb{E}[I|C^c] = \mathbb{P}(C)\mathbb{E}[I|C].$$

Here C^c denotes the complement of C, i.e. no curtailment has occurred. The last equality holds since all the mentioned indices are zero given that no curtailment has occurred. Hence, $\mathbb{E}[I]$ can be written as the product of the curtailment probability and the *conditional index*. This representation suggests estimation of $\mathbb{E}[I]$ by $\hat{I} := \hat{P}\hat{I}^C$, with $\hat{P}$ and $\hat{I}^C$ independent unbiased estimators for $\mathbb{P}(C)$ and $\mathbb{E}[I|C]$, respectively. As $\mathbb{P}(C)$ and $\mathbb{E}[I|C]$ are independent, $\hat{I}$ is obviously an unbiased estimator for $\mathbb{E}[I]$, and its variance is

$$\begin{aligned}\mathrm{Var}\left(\hat{I}\right) &= \mathrm{Var}\left(\hat{P}\hat{I}^C\right) \\ &= \mathbb{E}^2[I|C]\,\mathrm{Var}\left(\hat{P}\right) + \mathbb{P}(C)^2\,\mathrm{Var}\left(\hat{I}^C\right) + \mathrm{Var}\left(\hat{P}\right)\mathrm{Var}\left(\hat{I}^C\right).\end{aligned}$$

Dividing by $\mathbb{P}(C)^2\mathbb{E}^2[I|C]$ results in a decomposition of the squared relative error of $\hat{I}$

$$\mathrm{SRE}(\hat{I}) = \mathrm{SRE}\left(\hat{P}\right) + \mathrm{SRE}\left(\hat{I}^C\right) + \mathrm{SRE}\left(\hat{P}\right)\mathrm{SRE}\left(\hat{I}^C\right). \tag{5.11}$$

Expression (5.11) basically states that to control the precision of the index estimator, one should control the precision of both the probability estimator $\hat{P}$ and the conditional estimator $\hat{I}^C$.

5.4.1 *Controlling the probability estimator precision*

As $\hat{P}$ is the estimator of a rare event probability, a splitting technique may control $\mathrm{SRE}(\hat{P})$ using significantly less workload compared to a CMC method. The basic idea of a splitting technique is to decompose the probability into several conditional probabilities that are separately estimated using less total computational effort. This is done by splitting each sample path into multiple paths whenever the process is substantially closer to the rare event set. This subsection we adapt a splitting technique for the described MC reliability estimation method.

By defining the vector of discrete time Markov processes $\mathbf{X}(t) = (-|\mathbf{V}(t)| \; |\mathbf{V}(t)| \; |\mathbf{I}(t)|)$ with state space $\mathcal{E}$, we can concatenate the three vector inequalities of curtailment constraints (5.2) – (5.3):

$$\mathbf{X}(t) \leq \mathbf{U}, \quad \text{with} \quad \mathbf{U} = \begin{pmatrix} -\mathbf{V}_{\min} & \mathbf{V}_{\max} & \mathbf{I}_{\max} \end{pmatrix}.$$

In accordance with a general splitting procedure, we should choose an appropriate importance function $h : \mathcal{E} \mapsto \mathbb{R}$, assigning importance values to the states of $\mathbf{X}(t)$. Increasing values of h should correspond to approaching the rare event. We propose an intuitive importance function that takes the maximum over all ratios between the state variables and their allowed extrema:

$$h(\mathbf{X}(t)) = \max_i \left(\frac{\mathbf{X}_i(t) - \mathbf{L}_i}{\mathbf{U}_i - \mathbf{L}_i} \right), \tag{5.12}$$

with

$$\mathbf{L} = \begin{pmatrix} -\frac{\mathbf{V}_{\max}+\mathbf{V}_{\min}}{2} & \frac{\mathbf{V}_{\max}+\mathbf{V}_{\min}}{2} & \mathbf{0} \end{pmatrix}.$$

Here the subscript $i = 1, \ldots, 2N + M$ denotes the index of the vector. For $h(\mathbf{X}(t)) < 1$, h indicates how relatively close we are to a constraint

violation at time t. If for some constraint $\mathbf{X}_i(t) > \mathbf{U}_i$, then $h(\mathbf{X}(t)) > 1$, signifying that the rare event is hit, i.e. a curtailment occurs at time t. The linear transformation of each $\mathbf{X}_i$ in (5.12) ensures that the ratio of each constraint type has the same codomain $[0, 1]$ as long as the rare event is not hit, and $h \nearrow 1$ corresponds to approaching the rare event. In this sense, each constraint is assumed equally important when evaluating $h(\mathbf{X}(t))$.

Using importance function (5.12), we set up a Fixed Number of Successes (FNS) splitting technique as described in Section 3.3.2. That is, we decompose the rare event probability

$$\mathbb{P}(C) = \prod_{k=1}^{m} \mathbb{P}(D_k | D_{k-1})$$

with number of levels m, event $D_k := \{T_k < T\}$, stopping time $T_k := \inf\{t > 0 : h(\mathbf{X}(t)) \geq l_k\}$ and thresholds $0 = l_0 < l_1 < \cdots < l_m = 1$. Then the FNS estimator

$$\hat{P} := \prod_{k=1}^{m} \frac{r_k - 1}{n_{k-1} - 1},$$

with r_k, n_k the number of hits and trials, respectively, at level k, is an unbiased estimator for $\mathbb{P}(C)$.

The optimal level heights and number of stages are not known beforehand, and Amrein and Künsch [2] recommend a pilot run to determine these parameters. This pilot run uses a large number of equidistant levels and $r := r_k$ of moderate size, say $r = 20$, yielding first estimates $\hat{p}_k$. The optimal value for p_k in terms of variance reduction is $p_{\text{opt}} \approx 0.2032$ [2]. A pilot run estimate for $\hat{p}_k$ close to one suggests to merge stage k with a neighboring stage, whereas an estimate close to zero suggests to divide the stage into multiple stages. More precisely, one finds an improved stage partition for the final run by interpolating the pilot stage partition on the log scale.

An unbiased estimator for the variance of the FNS (and FE) estimate is not known. However, Amrein and Künsch [2] showed that under the

assumption that the conditional hitting probability does not depend on the entrance state of the previous stage,

$$\mathbb{P}\left(D_k | D_{k-1}, (T_{k-1}, \mathbf{X}(T_{k-1}))\right) = \mathbb{P}(D_k | D_{k-1}) \tag{5.13}$$

for all $(T_{k-1}, \mathbf{X}(T_{k-1}))$ for all k, one can bound the squared relative error of $\hat{P}$ by choosing r_k appropriately:

$$\mathrm{SRE}(\hat{P}) \leq q := -1 + \prod_{k=1}^{m} \left(\frac{1}{r_k - 2} + 1 \right). \tag{5.14}$$

In this way, we are able to control the precision of $\hat{P}$.

5.4.2 *Controlling the conditional estimator precision*

We can reuse the sample paths that hit the rare event set to estimate $\mathbb{E}[I|C]$, by performing one additional splitting stage. We randomly choose one of the r_m realizations of the last stage entrance state $(T_m, \mathbf{X}(T_m))$ and continue to generate the path from this point until time T, obtaining an index realization $\hat{I}_i^C$. Repeating this n times yields the estimator

$$\hat{I}^C := \frac{1}{n} \sum_{i=1}^{n} \hat{I}_i^C \tag{5.15}$$

for $\mathbb{E}[I|C]$. One might question whether I_i^C is unbiased as it depends on a randomly chosen realization of $(T_m, \mathbf{X}(T_m))$, instead of directly on $(T_m, \mathbf{X}(T_m))$. However, randomly choosing this realization is equivalent to sampling it from the empirical distribution $\hat{G}_m$ of the distribution G_m of $(T_m, \mathbf{X}(T_m))$. Therefore, as $\hat{G}_m$ is an unbiased estimator for G_m, we conclude that I_i^C (and thus I^C) is indeed an unbiased estimator for $\mathbb{E}[I|C]$. More rigorously: note that $\hat{I}_i^C = \hat{I}_i^C(Z)$ is a function of $Z \sim \hat{G}_m$, the entrance state randomly chosen from all simulated entrance states $Z_1, \ldots Z_{r_m} \sim G_m$ into the rare event. So $\hat{G}_m$ is the empirical distribution of G_m and we can write

$$\mathbb{E}_{\hat{G}_m}\left[\hat{I}_i^C(Z) \middle| \text{ all but } Z \right] = \frac{1}{r_m} \sum_{j=1}^{r_m} \hat{I}_i^C(Z_j).$$

Here we conditioned on all random variables that I_i^C depends on except Z, that is: $Z_1, \ldots, Z_{r_m}$ and the sample path from Z on. Unbiasedness of $\hat{I}^C$ immediately follows since

$$\begin{aligned}
\mathbb{E}\left[\hat{I}^C\right] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\hat{I}_i^C\right] \\
&= \mathbb{E}\left[\hat{I}_1^C(Z)\right] \\
&= \mathbb{E}\left[\mathbb{E}_{\hat{G}_m}\left[\hat{I}_1^C(Z)\middle|\ \text{all but } Z\right]\right] \\
&= \frac{1}{r_m}\sum_{j=1}^{r_m}\mathbb{E}\left[\hat{I}_1^C(Z_j)\right] \\
&= \mathbb{E}\left[\hat{I}_1^C(Z_1)\right] \\
&= \mathbb{E}\left[I|C\right],
\end{aligned}$$

where we used the law of total expectation in the third equality.

Estimators $\hat{I}^C$ and $\hat{P}$ are indeed independent, a fact that we used at the start of Section 5.4, before we explicitly defined these estimators. Further, the construction of FNS (in contrast to FS and FE) ensures the existence of realizations $(T_m, \mathbf{X}(T_m))$ and thus $\hat{I}^C$ as long as the rare event probability $\mathbb{P}(C)$ is not zero. We can estimate all indices using only one simulation run, just as in a similar CMC run. Furthermore, we attain an arbitrarily small $\mathrm{SRE}(\hat{I}^C)$ by choosing n sufficiently large, thereby controlling the conditional estimate precision. As we do not estimate a rare event probability here, we expect the computational intensity to be negligible compared to the estimation of $\mathbb{P}(C)$.

5.5 PERFORMANCE RESULTS ON AN EXAMPLE GRID

We will investigate the computational intensity of the proposed FNS technique for grid reliability estimation on a very simple transmission grid. As shown in Figure 5.1, the grid exists of one wind farm node, one consumption node and one so-called slack node, where the total surplus or shortage of power is absorbed or emitted, respectively. All

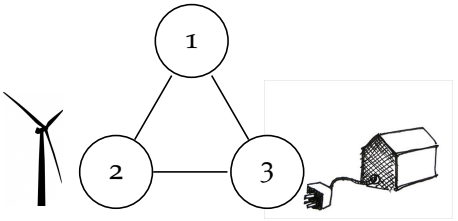


Figure 5.1: A small power grid with one slack node (1), one intermittent source node (2), and one consumption node (3).

nodes are connected by identical transmission lines. The discrete time domain is one week. For simplicity, we omit periodicities and model the active power $P_2(t)$ of the wind farm in node 2 as an Ornstein-Uhlenbeck (OU) process

$$dP_2(t) = \theta(\mu - P_2(t))dt + \sigma dW(t).$$

Here $W(t)$ denotes a Brownian motion and the long-term mean $\mu = P_3 = 15$ base MVA (using the scalable per-unit system [39]) equals the constant active power consumption at node 3. The process initiates under normal conditions by setting $P_2(0) = \mu$, and realistic values for the mean-reverting rate $\theta = 0.13$ per hour and volatility $\sigma = 1.3$ are chosen by fitting the model to historical wind power measurements [49]. We choose the extrema $\mathbf{V}_{\min}, \mathbf{V}_{\max}$ and $\mathbf{I}_{\max}$ such that $\mathbb{P}(C)$ is indeed small (around 10^{-4}), using a preliminary CMC simulation with 100 000 MC samples. The injected reactive wind power is equal to one third of the active power: $Q_2(t) = P_2(t)/3$. Similarly, for the consumed reactive power in node 3: $Q_3 = P_3/3$. Admittance matrix Y consists of elements $Y_{ii} = -200\imath$ and $Y_{ij} = 100\imath$ for $i \neq j$, where $\imath$ is the imaginary unit. To estimate the curtailment size $S_C(t)$ in (5.7) at an unfeasible time step (that is, when a constraint is violated), we assume a simple strategy: the grid operator instantaneously reschedules generation as it was during the last feasible time step. The absolute differences between all nodal powers at

the feasible and unfeasible situations sum up to the power curtailment size at time t:

$$S_C(t) = \sum_{\text{nodes } i} \left| P_i^{\text{feasible}} - P_i^{\text{unfeasible}} \right|. \tag{5.16}$$

After obtaining $S_C(t)$ at all discrete times steps, numerical time integration yields a realization of $EENS(T)$. More realistic strategies would include a linear optimization problem where the curtailment size is minimized such that all curtailment constraints are satisfied. As we expect a curtailment to be rare, we do not expect significant increase of the total simulation workload if we replace the curtailment strategy (5.16) by a more realistic one.

The results of the pilot run as described in Section 5.4.1 recommend to use $m = 6$ stages in this case. Then according to the squared relative error bound in (5.14), choosing $r_k = r = 100$ will yield a squared relative error smaller than $q = 0.063$. Assuming asymptotic normality of $\hat{P}$, we will obtain the conservative 95% confidence interval

$$[\hat{P} - 1.96\sqrt{q}\hat{P}, \hat{P} + 1.96\sqrt{q}\hat{P}] \approx [\hat{P}/2, 3\hat{P}/2].$$

Similarly, $r_k = r = 40$ will yield $q = 0.17$ and thus a conservative confidence interval of the form $[c, 10c]$ for some c, rather indicating the order of magnitude. For the final run, we choose $r = 100$, a time step size of 6 minutes and the importance function h as in (5.12). We further assume that (5.13) and thus (5.14) hold. The resulting index estimates are displayed in the first column of Table 5.2. The second column lists estimates $\widehat{\text{SRE}}\left(\hat{I}^C\right)$ of $\text{SRE}(\hat{I}^C)$. Finally, (5.11) suggests to estimate the bound for $\text{SRE}(\hat{I})$ by

$$q_{\hat{I}} := q + \widehat{\text{SRE}}\left(\hat{I}^C\right) + q\widehat{\text{SRE}}\left(\hat{I}^C\right).$$

Values for $q_{\hat{I}}$ are displayed in the third column. Since the estimates for $\text{SRE}(\hat{I}^C)$ will contain an error, the bound $\text{SRE}(\hat{I}) < q_{\hat{I}}$ for the relative error of the index estimate is not guaranteed. However, as explained in Section 5.4.2, we can have $\text{SRE}(\hat{I}^C)$ arbitrarily small by choosing sufficiently large n in (5.15). The estimates for $\text{SRE}(\hat{I}^C)$ are indeed significantly smaller than

q (using only $n = 5r = 500$ samples, with a workload comparable to that of one splitting stage). Therefore, the conservative estimate $\mathrm{SRE}(\hat{I}) \approx q_{\hat{I}}$ will be satisfactory for most practical reliability assessments. In total 3075

	$\hat{I}$	$\widehat{\mathrm{SRE}}\left(\hat{I}^C\right)$	$q_{\hat{I}}$
$PPC(T)$	6.35e-5	0	0.0628
$EDPC(T)$ (hour)	2.89e-5	2.4e-3 (1.5e-4)	0.0710
$ENPC(T)$	1.01e-4	6.1e-4 (3.9e-5)	0.0666
$EENS(T)$ (per-unit MWh)	2.18e-5	5.7e-3 (3.6e-4)	0.0819

Table 5.2: Estimates $\hat{I}$ for indices (5.4) – (5.7) of the three node grid reliability over one week, the squared relative error estimate $\widehat{\mathrm{SRE}}\left(\hat{I}^C\right)$ of the conditional index (standard error in parentheses), and the corresponding bound estimate $q_{\hat{I}}$ for $\mathrm{SRE}(\hat{I})$. The time step size is 6 minutes, and $q = 0.063$.

MC samples where required using the FNS technique. To obtain a CMC estimate for $PPC(T)$ with a comparable squared relative error, equation (5.10) suggests to use as many as 250 000 MC samples. This workload decrease of a factor 79 illustrates the computational gain of the FNS technique compared to the CMC method. For smaller values of $PPC(T)$, the gain will be even larger.

5.5.1 *Choice for the time step size*

The approximation $\mathrm{SRE}(\hat{I}) \approx q_l$ relies on assumption (5.13), which holds as long as the level hitting probability does not depend on the entrance state $(T_{k-1}, \mathbf{X}(T_{k-1}))$ of the previous stage. We will explain that this is not necessarily the case in the proposed power grid model. Repeating the FNS index estimation would yield an unbiased estimate of the variance under general circumstances. However, this will heavily increase the computational effort, so we are interested under which circumstances (5.13) and thus (5.14) hold approximately.

Note that a sample path of $h(\mathbf{X}(t))$ may skip stage k entirely by a very large increment in a single time step. Then any sample path started from

this entrance state has immediately hit level k, as shown in Figure 5.2. This example shows the dependence of the hitting probability on $\mathbf{X}(T_{k-1})$. In general, large increments are undesirable as they may increase the squared relative error beyond bound q. As we can not bound an OU

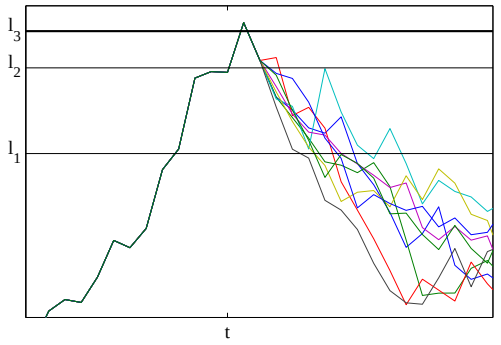


Figure 5.2: The problem of too large increments in an extreme example: the only survived path skips level 3 entirely at time t, wherefore $\hat{p}_3 = 1$ will probably overestimate p_3.

increment, there is no step size in general that bounds the increments of $h(\mathbf{X}(t))$. However, decreasing the time step size Δ may reduce the variance of an OU increment sufficiently for (5.13) to hold approximately. Using the exact solution of the OU process [33], one can derive the OU increment variance

$$\mathrm{Var}(P_2(t+\Delta) - P_2(t)) = \frac{\sigma^2}{2\theta}\left(2\left(1 - e^{-\theta\Delta}\right) - e^{-2\theta t}\left(1 - e^{-\theta\Delta}\right)^2\right),$$

which indeed approaches zero as $\Delta \to 0$. By computing FNS estimates for $PPC(T)$ 50 times, we estimate $\mathrm{SRE}(\hat{P})$ and investigate for which time step sizes the squared relative error bound (5.14) holds. Table 5.3 shows that a step size of one hour is too large (as $\mathrm{SRE}(\hat{P}) \not< q$), whereas runs with a step size of 6 minutes or smaller are consistent with (5.14). This is in agreement with the statement in Frunt [30] that one will not fully capture the typical variability of wind power generation when using step sizes larger than 10 minutes. Although the optimal time step size

is unknown a priori, such knowledge on the variability of the stochastic sources may hint a proper choice for Δ. Further, the typical increment size of h in the pilot run described in Section 5.5 may test as well whether the time step is sufficiently small.

Δ	$q = 0.169,\ (r = 40)$	$q = 0.063,\ (r = 100)$
1 hour	0.847	0.482
6 minutes	0.168	0.051
1 minute	0.164	0.034

Table 5.3: Estimates for $\mathrm{SRE}(\hat{P})$, using importance function h in (5.12). Estimates are consistent with bound (5.14) for sufficiently small time step sizes Δ. We used 50 estimates for $\hat{P}$.

5.5.2 *Choice for the importance function*

We would like to address the relevance of choosing a suitable importance function. For example, one may have chosen the importance function

$$h_1(\mathbf{X}(t)) = \max_i \left((\mathbf{X}_i(t) - \mathbf{U}_i)/\mathbf{U}_i \right) \tag{5.17}$$

instead of (5.12), where $h_1(\mathbf{X}(t)) \nearrow 0$ corresponds to approaching the rare event. Since now the codomain of arguments $(\mathbf{X}_i(t) - \mathbf{U}_i)/\mathbf{U}_i$ differ per constraint i, some constraint ratios may be much more volatile than others, which may significantly increase the probability of large increments. This can be seen in the three node power grid if we reestimate $\mathrm{SRE}(\hat{P})$ using h_1. Table 5.4 shows that $\mathrm{SRE}(\hat{P})$ is larger than q, even for small step sizes. This illustrates the relevance of a choosing a suitable importance function. A priori known (or recursively learned) relations between certain system states and the conditional probabilities may help to improve the choice for the importance function. Finding and using these relations is part of further research.

Δ	$q = 0.169,\ (r = 40)$	$q = 0.063,\ (r = 100)$
1 hour	2.18	1.26
6 minutes	1.28	0.38
1 minute	1.18	0.32

Table 5.4: Estimates for $\mathrm{SRE}(\hat{P})$, using importance function h_1 in (5.17). Estimates exceed bound q, even for small step sizes Δ. We used 100 estimates for $\hat{P}$.

5.6 CONCLUSION AND RECOMMENDATIONS

We demonstrated the high computational intensity of a typical Crude Monte Carlo method for reliability estimation of electrical power grids. We showed in Table 5.2 that a splitting technique may decrease the workload of estimating various reliability indices by orders of magnitude while controlling the squared relative error of the estimators. The reliability indices may either be the rare event probability of a power curtailment or an expectation depending on this probability. An implementation on a small transmission network shows that the proposed method requires 79 times less workload to estimate four common reliability indices than an equivalent Crude Monte Carlo simulation would require. To control the squared relative error of the estimator, an appropriate choice for the time step size and the importance function is crucial. The time step size should be sufficiently small to capture the typical variability of the stochastic power sources. Furthermore, the importance function should assign equal importance to all curtailment risks, like importance function (5.12) does.

6

A SEPARATED SPLITTING TECHNIQUE FOR DISCONNECTED RARE EVENT SETS

6.1 INTRODUCTION

Efficient estimation of very low probabilities is desired in many application areas like telecommunication networks and power grid reliability [19, 95]. Monte Carlo simulation is a robust and popular technique in case the model does not allow an exact derivation, but Crude Monte Carlo may require prohibitively many samples to be practically feasible. To reduce the computational burden, rare event simulation techniques have been developed [71], and (importance) splitting is a well-known variant. A splitting technique aims to reduce the squared relative error by replicating (splitting) trajectories as soon as the rare event is presumed more likely to occur [50, 34, 89].

The efficiency of a splitting technique depends strongly on the choice of the importance function. The optimal importance function strictly increases with the probability of reaching the rare event set from the current state. In this chapter we assume no knowledge of this probability except that the underlying Markov process is (the discretization of) a continuous stochastic process. In a simple case with a one-dimensional state space and an interval as the rare event set, any importance function that measures the proximity of the current state to the interval is optimal. In many other cases however, the choice for the importance function is much more difficult.

For example, a challenge arises when the rare event set is disconnected; that is, it consists of multiple nonempty, separated subsets. Such a rare event set is plausible in a reliability assessment of a multiple component system for which various combinations of component behavior may lead

This chapter is based on the conference paper Wadman et al. [92].

to a system failure [39, 7]. One example involves the reliability estimation of an electrical power grid with nodal power injections modeled as stochastic processes [91]. Then if at any grid node the injection is very different from that at an adjacent node, the current through the connection in between may exceed its allowed maximum. Hence there could be various combinations of power injections leading to the rare event of a grid fault, and they could involve opposite extremes of power injections. In case of a disconnected rare event set, an importance function based on the proximity to the complete rare event set may lead to poor results: the algorithm may replicate sample paths near a subset that has only little or even negligible contribution to the rare event probability. Therefore the importance function may be far from optimal as the simulation focuses on areas where the rare event probability is hardly increased or even decreased.

In this chapter we assume that we have a partition of the rare event set into subsets that are (or at least might be) separated from each other. We propose to estimate the rare event subset probabilities separately. We use a modified splitting technique to estimate each subset probability, taking care that sample paths hitting multiple sets are not counted multiple times. Standard splitting framework and notation are introduced in Section 6.2. We discuss some properties of the optimal importance function in Section 6.3. In Section 6.3.1 we show a one-dimensional problem where the choice for the importance function is clear, and we describe a technique to calibrate the levels using a pilot run. In Section 6.3.2, we discuss the challenge of a disconnected rare event set, and we describe how to estimate each subset probability separately. We present the numerical results of an experiment in Section 6.4, after which conclusions are given in Section 6.5.

6.2 RARE EVENT SIMULATION

We consider the class of rare events where a discrete-time Markov chain $\{X_t, t \geq 0\}$ with state space $\mathcal{X}$ enters the rare event set $B \subset \mathcal{X}$ before stopping time ζ. We define

$$\tau_B = \inf\{t > 0 : X_t \in B\}$$

as the first entrance time in B and $x_0 \notin B$ as the initial state. We want to estimate the rare event probability

$$\gamma := \mathbb{P}(\tau_B < \zeta | X_0 = x_0).$$

Two typical choices for ζ are a fixed finite time $\zeta = T$ and stopping time $\zeta = \tau_A := \inf\{t > 0 : \exists\ 0 \leq s < t \text{ s.t. } X_s \notin A \text{ and } X_t \in A\}$, the first time the chain reenters A, where $A \subset \mathcal{X}$ is some set that contains x_0. The Crude Monte Carlo (CMC) estimator

$$\tilde{\gamma}_n := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{\tau_B < \zeta \text{ in sample } i\}}$$

for γ is unbiased but its squared relative error

$$\mathrm{SRE}(\tilde{\gamma}_n) := \frac{\mathrm{Var}\, \tilde{\gamma}_n}{\gamma^2} = \frac{\gamma(1-\gamma)}{\gamma^2 n} = \frac{1-\gamma}{\gamma n} \tag{6.1}$$

diverges to infinity as $O(1/\gamma)$ when $\gamma \to 0$. Therefore, to estimate a very small probability using CMC simulation, one will need a number of samples that may be computationally too intensive in practice.

6.2.1 *Importance splitting*

In this chapter, we will use the Fixed Number of Successes (FNS) splitting estimator as described in detail in Section 3.3.2. That is, we start by defining an importance function $h : \mathcal{X} \mapsto \mathbb{R}$. We decompose the rare event probability

$$\gamma = \prod_{k=1}^{m} \mathbb{P}(D_k | D_{k-1})$$

with number of levels m, event $D_k := \{T_k < \zeta\}$, stopping time $T_k := \inf\{t > 0 : h(X_t) \geq l_k\}$ and thresholds $0 = l_0 < l_1 < \cdots < l_m = 1$. Then the FNS estimator

$$\hat{\gamma} := \prod_{k=1}^{m} \frac{r_k - 1}{n_{k-1} - 1},$$

with r_k, n_k the number of hits and trials, respectively, at level k, is an unbiased estimator for γ.

6.3 FINDING A SUITABLE IMPORTANCE FUNCTION

Finding a suitable importance function is crucial to reduce the variance of the splitting estimator [31, 50]. The importance function should 'reward good behavior' by splitting trajectories that are more likely to hit the rare event set. Glasserman et al. [36] show that the levels should be chosen in a way consistent with the most likely path to the rare event set. Ideally, this path should coincide with the most likely path to any intermediate level. In case $\zeta = \tau_A$ is a stopping time as defined in Section 6.2, Garvels et al. [32] propose to use the importance function

$$g(h^*(x)), \tag{6.2}$$

with $g : [0, 1] \mapsto \mathbb{R}$ some strictly increasing function and

$$h^*(x) := \mathbb{P}(\tau_B < \zeta | X_0 = x). \tag{6.3}$$

As knowing $h^*(x)$ would defeat the point of using simulation, the lesson is to find an importance function that is close to (6.2). Some results on the efficiency of splitting can be derived under the assumption that the probability of hitting the next level does not depend on the current entrance state

$$\mathbb{P}\left(D_k | D_{k-1}, (T_{k-1}, X_{T_{k-1}})\right) = p_k \ \ \forall\ (T_{k-1}, X_{T_{k-1}}). \tag{6.4}$$

Under this assumption and using the optimal choice for m and p_k Amrein and Künsch [2] show that the squared relative error of the FNS estimator is proportional to

$$\mathrm{SRE}(\hat{\gamma}) \propto \frac{(\log \gamma)^2}{n}.$$

Hence the $\mathrm{SRE}(\hat{\gamma})$ diverges as $O((\log \gamma)^2)$ when $\gamma \to 0$. This squared logarithmic divergence rate is slower than the divergence rate of the CMC squared relative error in (6.1). Splitting estimators outperfom the CMC estimator in this sense. Furthermore, under assumption (6.4) the authors show that $\hat{p}_1, \ldots, \hat{p}_m$ are independent and derive the bound

$$\mathrm{SRE}(\hat{\gamma}) \leq -1 + \prod_{k=1}^{m} \frac{r_k - 1}{r_k - 2} \tag{6.5}$$

for the squared relative error of the FNS splitting estimator. That is, the accuracy of the estimator can be controlled by choosing the number of hits r_k sufficiently large for all k. In Section 6.4 we will check if the simulation results of an experiment fulfill bound (6.5). As the validity of this bound is a necessary condition for (6.4) to hold, this check will serve as a test for the quality of the chosen importance function.

6.3.1 *A simple case with a connected rare event set*

Consider the one-dimensional case $\mathcal{X} = \mathbb{R}$, X_t the discretization of a continuous stochastic process and $B = [b, \infty)$ an interval for some $b > x_0$. Then $\gamma = \mathbb{P}(\tau_B < \zeta | X_0 = x_0)$ is a barrier crossing probability and it is fair to assume that $\mathbb{P}(\tau_B < \zeta | X_0 = x)$ increases in x as long as $x < b$, since for the continuous counterpart of X_t this is certainly true. Therefore, we may assume that the normalized proximity function

$$h_p(x) := 1 - (b - x)/(b - x_0) \tag{6.6}$$

is of the form (6.2) even though g and h^* are not explicitly known. Stated in terms of Glasserman et al. [36], the most likely path to the rare event set

will automatically coincide with the most likely path to any intermediate level in this simple case.

The remaining challenge is to find a suitable partition of the interval $[0,1]$ into level thresholds. Amrein and Künsch [2] propose a pilot run to obtain a rough idea of the optimal number and height of the thresholds for a subsequent final run. That is, they suggest a splitting simulation with a large number of levels and a moderate number of level hits such that the computational intensity is small compared to that of the final run. The authors showed that the optimal value for p_k in terms of variance reduction is $p_{\text{opt}} := 0.2032$. Therefore, a pilot estimate of p_k close to one suggests to merge level k with a adjacent level, whereas a pilot estimate close to zero suggests to divide the level into multiple levels.

Along these lines we now describe such a calibration procedure in more detail. That is, we search for m, and for all $k = 1, \ldots, m-1$ we search for the threshold l_k such that

$$p(l_k) = p_{\text{opt}}^{m-k},$$

where $p(l)$ is the conditional probability of the rare event given that the importance function has exceeded l. Assuming no knowledge of h^*, we perform a pilot run using many levels, say 20, and equidistant thresholds $l_k^{\text{pilot}} = k/20$. As this will yield information on $p(l)$ only at the pilot threshold values $l = l_i^{\text{pilot}}$, we assume a power law relationship for $p(l)$ between each pair of subsequent pilot thresholds:

$$p(\alpha \Delta l) = p(\Delta l)^{\alpha}, \tag{6.7}$$

for $\alpha > 0$. For example, if the top level probability $\hat{p}_m^{\text{pilot}} = p_{\text{opt}}^2$, then (6.7) suggests to set the penultimate threshold l_{m-1} twice as close to the top as the penultimate pilot threshold l_{19}^{pilot}. Hence, we approximate $p(l)$ by the log-linear interpolation $p^*(l)$ of the estimators

$$\prod_{i=k+1}^{20} \hat{p}_i^{\text{pilot}}$$

for $p\left(l_k^{\text{pilot}}\right)$, as shown in Figure 6.1. Then we simply solve the obtained piecewise linear equation

$$\log p^*(l_k) = (m-k)\log p_{\text{opt}},$$

for l_k. The number of levels is automatically found by computing l_{m-1}, $l_{m-2}, \ldots, l_i$ where i is the last index such that $l_i > 0$. After re indexing we have found the levels $0 = l_0 < \cdots < l_m = 1$ that we use for the final run.

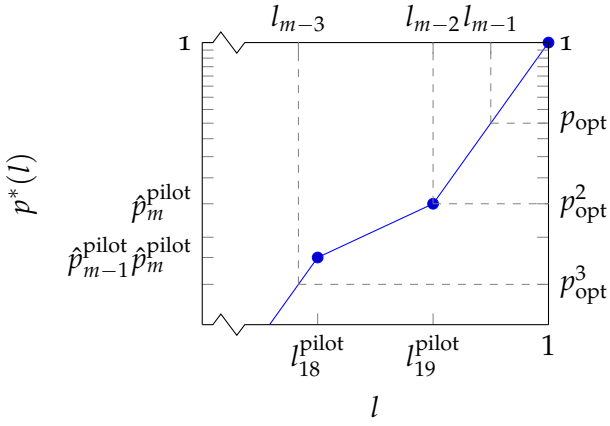


Figure 6.1: An updated set of thresholds $l_0, \ldots, l_m$ is found by a log-linear interpolation of the estimates of reaching the rare event set from the pilot thresholds.

6.3.2 *A disconnected rare event set*

Now we consider a possibly multidimensional state space $\mathcal{X}$, and a rare event set B that can be partitioned into two nonempty, separated subsets B_1 and B_2. By an induction argument the rest of this section can be generalized to a partition into more than two subsets. As an extension of (6.6), an intuitive choice for the importance function is the maximum

$$h_{12}(x) := \max_{i=1,2} h_i(x) \tag{6.8}$$

of the normalized proximities of x to each subset:

$$h_i(x) := 1 - \frac{d(x, B_i)}{d(x_0, B_i)}. \tag{6.9}$$

Here $d(v, V)$ denotes the distance between point $v \in \mathcal{X}$ and set $V \subset \mathcal{X}$. However, as we assume no prior knowledge of $h^*(x)$, it could be that the most likely path to an intermediate level approaches B_2, whereas the most likely path to B does not and instead goes directly to B_1. If X_t is the discrete counterpart of a continuous stochastic process, we may expect that this deviation between the most likely paths to different level sets becomes smaller as the distance between B_1 and B_2 goes to zero.

Therefore, we propose not to estimate the rare event probability $\gamma = \mathbb{P}(\tau_B < \zeta | X_0 = x_0)$ using h_{12} in (6.8) for disconnected B. Instead, we write $\gamma = \gamma_1 + \gamma_2$ with

$$\begin{aligned}
\gamma_1 &:= \mathbb{P}(\tau_{B_1} < \zeta | X_0 = x_0), \\
\gamma_2 &:= \mathbb{P}(\tau_{B_2} < \zeta \cup \tau_{B_1} \geq \zeta | X_0 = x_0).
\end{aligned}$$

We can interpret this as the rare event being partitioned into a part where B_1 is hit in due time (and maybe B_2 as well), and a part where B_2 is hit in due time while B_1 is never hit during $[0, \zeta]$. We propose to separately estimate γ_1 and γ_2 using two independent splitting simulations. To obtain an unbiased estimator for γ_1 we simply choose h_1 in (6.9) as importance function (see Figure 6.2). To obtain an unbiased estimator for γ_2 we choose h_2 in (6.9) as importance function and a modified splitting framework rewarding sample paths approaching B_2 while killing those that hit B_1 (see Figure 6.3). We call these two runs together the *separated splitting technique*, as opposed to the single *standard splitting technique* using importance function h_{12} in (6.8).

More precisely, to estimate γ_2 we start from a standard splitting setup for $\mathbb{P}(\tau_{B_2} < \zeta | X_0 = x_0)$ as described in subsection 6.2.1. We set l_k, T_k and D_k for all k as before, but we modify the levels by defining

$$\begin{aligned}
\bar{D}_k &:= D_k \cap \{\tau_{B_1} \geq T_k\} \quad \forall\, k = 0, \ldots, m, \\
\bar{D}_{m+1} &:= D_m \cap \{\tau_{B_1} \geq \zeta\}.
\end{aligned}$$

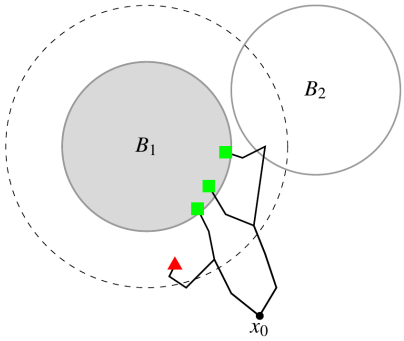


Figure 6.2: Illustration of splitting sample paths to estimate $\gamma_1 = \mathbb{P}(\tau_{B_1} < \zeta | X_0 = x_0)$, ignoring B_2. Squares correspond to subpaths for which the rare event does occur, triangles correspond to subpaths for which it does not.

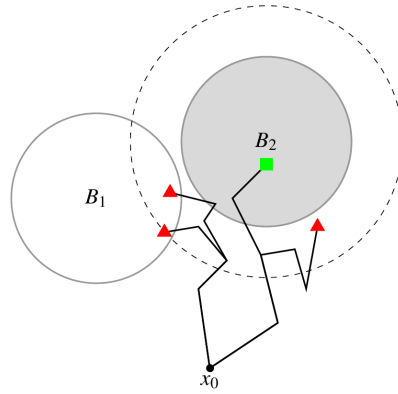


Figure 6.3: Illustration of splitting sample paths to estimate $\gamma_2 = \mathbb{P}(\tau_{B_2} < \zeta \cup \tau_{B_1} \geq \zeta | X_0 = x_0)$, killing paths that enter B_1. Squares correspond to subpaths for which the rare event does occur, triangles correspond to subpaths for which it does not.

So to enter level k, we additionally require the chain not to have entered B_1 in the mean time. Further, the occurrence of the additional level $\bar{D}_{m+1}$ requires that after entering B_2 in due time, the chain should not have entered B_1 during the remaining time domain $[T_m, \zeta]$. Note that $\bar{D}_{m+1}$ is exactly the rare event, and since again $\bar{D}_0 \supset \bar{D}_1 \supset \cdots \supset \bar{D}_m \supset \bar{D}_{m+1}$ we can write

$$\gamma_2 = \mathbb{P}(\bar{D}_{m+1}) = \prod_{k=1}^{m+1} \mathbb{P}(\bar{D}_k | \bar{D}_{k-1}).$$

We may now resume the standard splitting technique as described in subsection 6.2.1 using $D_k = \bar{D}_k$ for all k. Therefore, an unbiased estimator for γ_2 is given by

$$\hat{\gamma}_2 := \prod_{k=1}^{m+1} \hat{p}_k,$$

where for all levels k $\hat{p}_k$ is the proportion of sample paths for which $\bar{D}_k$ occurs. We conclude that $\hat{\gamma}_1 + \hat{\gamma}_2$ is an unbiased estimator for γ. Furthermore, if assumption (6.4) holds we can compute bounds for the squared relative errors $\mathrm{SRE}(\hat{\gamma}_1)$ and $\mathrm{SRE}(\hat{\gamma}_2)$ using (6.5). Then a bound for $\hat{\gamma}_1 + \hat{\gamma}_2$ follows immediately, since

$$\begin{aligned} \mathrm{SRE}(\hat{\gamma}_1 + \hat{\gamma}_2) &= \frac{\operatorname{Var} \hat{\gamma}_1 + \operatorname{Var} \hat{\gamma}_2}{(\gamma_1 + \gamma_2)^2} \\ &\leq \frac{\operatorname{Var} \hat{\gamma}_1}{\gamma_1^2} + \frac{\operatorname{Var} \hat{\gamma}_2}{\gamma_2^2} \\ &= \mathrm{SRE}(\hat{\gamma}_1) + \mathrm{SRE}(\hat{\gamma}_2). \end{aligned} \tag{6.10}$$

We used independence of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in the first equality.

6.4 EXPERIMENT

We compare the proposed *separated splitting technique* using importance functions h_1 and h_2 in (6.9) with the *standard splitting technique* using importance function h_{12} in (6.8). The state space $\mathcal{X} = \mathbb{R}$ is one-dimensional and the rare event set $B = B_1 \cup B_2$ consists of two separate intervals $B_1 = [b_1, \infty)$ and $B_2 = (-\infty, b_2]$, $b_1 > b_2$. We assume that $\{X_t, t \geq 0\}$ is a modification of an Ornstein-Uhlenbeck process, where we change the sign of the long-term mean whenever the process passes zero:

$$dX(t) = \theta(\mu \operatorname{sgn}(X_t) - X_t)dt + \sigma dW(t). \tag{6.11}$$

The drift of this stochastic differential equation corresponds to the potential

$$V(x) = -\int \theta(\mu \operatorname{sgn}(x) - x)dx = \frac{\theta}{2}(x - \mu \operatorname{sgn}(x))^2. \tag{6.12}$$

Figure 6.4 shows that the double well potential (6.12) is just the original Ornstein-Uhlenbeck potential, mirrored in the y-axis. We substitute $\mu \operatorname{sgn}(X_t)$ for μ in the exact solution of the Ornstein-Uhlenbeck process [33] to obtain a discretization of (6.11):

$$X_{i+1} = X_i e^{-\theta\Delta} + \mu \operatorname{sgn}(X_i)\left(1 - e^{-\theta\Delta}\right) + \sigma\sqrt{\frac{1 - e^{-2\theta\Delta}}{2\theta}} Z_i,$$

with $X_i = X(i\Delta)$ with time step size Δ, and $Z_i \sim N(0,1)$ i.i.d. standard normal for all i. We choose parameter values $\theta = \sigma = 1$, $\mu = -x_0 = 2$, $b_1 = 4$, $b_2 = -6$, $\Delta = 0.001$ and a deterministic stopping time $\zeta = T = 10$. We will estimate the rare event probability $\gamma := \mathbb{P}(\tau_B < T | X_0 = x_0)$. Obviously $\gamma = \gamma_1 + \gamma_2$, with $\gamma_1 := \mathbb{P}(\tau_{B_1} < T | X_0 = x_0)$ and $\gamma_2 = \mathbb{P}(\tau_{B_2} < T \cup \tau_{B_1} \geq T | X_0 = x_0)$. Figure 6.4 shows that the process starts in the left well and initially has equal probability of going right or left. To arrive at B_1 the process should overcome the potential barrier $V(0) = 2$ at $x = 0$. In spite of this barrier, CMC estimates in Table 6.1 show that $\gamma_1 \gg \gamma_2$, relating to the fact that $V(b_1) < V(b_2)$. We regard these CMC estimates as reference values for splitting estimates.

Probability	CMC estimate $\tilde{\gamma}_n$	$\mathrm{SRE}(\tilde{\gamma}_n)$
$\gamma = \gamma_1 + \gamma_2$	3.50×10^{-3}	2.85×10^{-5}
γ_1	3.50×10^{-3}	2.85×10^{-5}
γ_2	2.50×10^{-6}	4.00×10^{-2}

Table 6.1: CMC estimates $\tilde{\gamma}_n$ for the subset probabilities, using $n = 10^7$ samples.

We use both the standard splitting technique and the separated splitting technique to estimate γ. We choose the levels as described in Section 6.3.1 and use $r_k = r = 100$ hits per level. The standard splitting estimator for γ is denoted by $\hat{\gamma}_{12}$, whereas the separated splitting technique computes estimators $\hat{\gamma}_1$ for γ_1 and $\hat{\gamma}_2$ for γ_2, yielding $\hat{\gamma}_1 + \hat{\gamma}_2$ for γ. We investigate both techniques by comparing their estimates with the CMC estimates. After repeating both simulations 1000 times we check if the squared relative errors $\mathrm{SRE}(\hat{\gamma}_{12})$, $\mathrm{SRE}(\hat{\gamma}_1)$ and $\mathrm{SRE}(\hat{\gamma}_2)$ are indeed bounded by (6.5), which should hold under assumption (6.4).

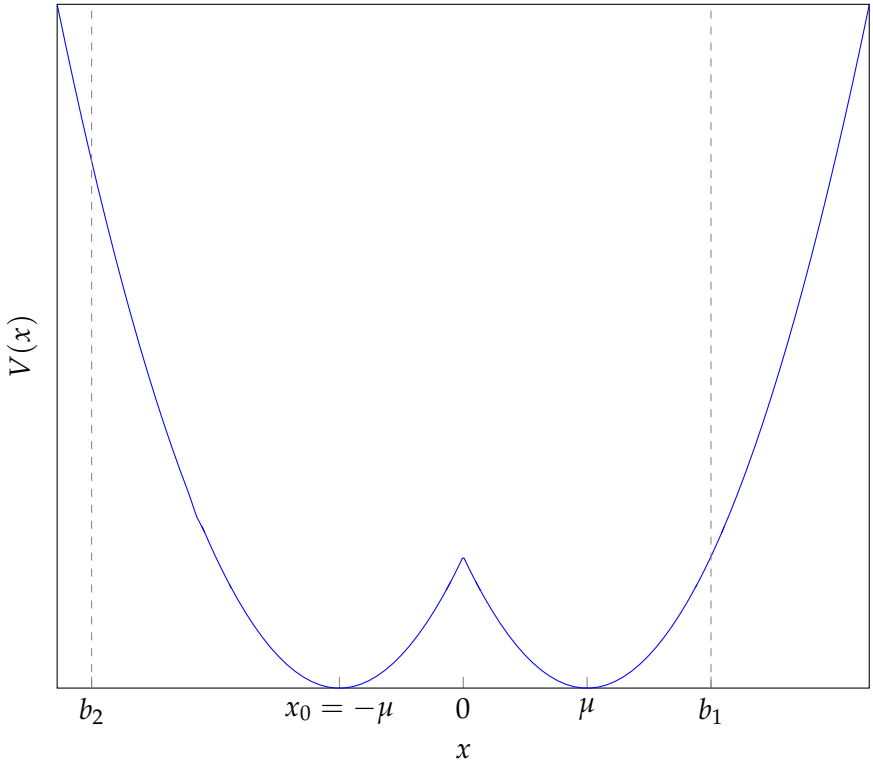


Figure 6.4: The double well potential (6.12) of the modified OU process in (6.11).

Table 6.2 shows that the squared relative error $\mathrm{SRE}(\hat{\gamma}_{12})$ of the standard splitting estimator is more than twice as high as the bound suggested by (6.5). This suggests that condition (6.4) under which this bound is derived does not hold, that is, hitting the next level does depend on the entrance state. Table 6.3 confirms this: more than two third of the first level hits occurred at the left side, whereas hits at the third and fourth level only occurred by exceeding the right side threshold. So hitting the next level clearly depends on the side at which paths crossed the previous threshold, i.e. on the entrance state, whereby (6.4) cannot hold when using the standard splitting technique.

	Value	SRE$(\cdot)$	Bound for SRE$(\cdot)$
$\hat{\gamma}_{12}$	2.48×10^{-3}	0.069	0.031
$\hat{\gamma}_1$	3.58×10^{-3}	0.034	0.041
$\hat{\gamma}_2$	1.52×10^{-6}	0.066	0.096

Table 6.2: Standard splitting estimator $\hat{\gamma}_{12}$ for γ and separated splitting estimator $\hat{\gamma}_1 + \hat{\gamma}_2$ for γ. The squared relative error SRE$(\hat{\gamma}_{12})$ exceeds bound (6.5), whereas SRE$(\hat{\gamma}_1)$ and SRE$(\hat{\gamma}_2)$ do not.

Level	Hits on left side	Hits on right side
1	67 (at $x = -3.92$)	33 (at $x = 0.88$)
2	8 (at $x = -4.78$)	92 (at $x = 2.17$)
3	0 (at $x = -5.63$)	100 (at $x = 3.44$)
4	0 (at $x = -6$)	100 (at $x = 4$)

Table 6.3: Efficiency loss using standard splitting: at the first level a significant number of successful sample paths approached the left interval 'in vain', since no sample path entered the left interval at the final level.

Separated splitting is not affected by this problem, and indeed the squared relative errors SRE$(\hat{\gamma}_1)$ and SRE$(\hat{\gamma}_2)$ do fulfill bound (6.5) in Table 6.2. Using (6.10) we find the bound $0.041 + 0.096 = 0.137$ for SRE$(\hat{\gamma}_1 + \hat{\gamma}_2)$. Assuming normality one can show that the implied conservative 95% confidence interval for γ overlaps with the narrow CMC confidence interval. However, one can show that the conservative confidence interval implied by the standard splitting technique is a significant underestimation.

We repeat the experiment for higher barriers by increasing μ, i.e. for smaller rare event probabilities. Table 6.4 shows that for barrier height $V(0) = 2.42$, SRE$(\hat{\gamma}_{12})$ is 3 times as high as bound (6.5) suggests. Table 6.5 and 6.6 show that SRE$(\hat{\gamma}_{12})$ exceeds the theoretical bound (6.5) by a factor up to 15 for even higher barriers. This illustrates that the accuracy of the splitting estimate may be much lower than expected when one does not assess separated rare event subsets in separated splitting simulations.

	Value	SRE$(\cdot)$	Bound for SRE$(\cdot)$
$\hat{\gamma}_{12}$	5.28×10^{-4}	0.120	0.041
$\hat{\gamma}_1$	8.75×10^{-4}	0.043	0.052
$\hat{\gamma}_2$	6.18×10^{-8}	0.095	0.118

Table 6.4: Experiment as in Table 6.2, but now $\mu = x_0 = 2.2$, thus raising the barrier height to $V(0) = 2.42$, and $b_1 = 4.4$, $b_2 = -6.6$. The squared relative error SRE$(\hat{\gamma}_{12})$ exceeds bound (6.5), whereas SRE$(\hat{\gamma}_1)$ and SRE$(\hat{\gamma}_2)$ do not.

	Value	SRE$(\cdot)$	Bound for SRE$(\cdot)$
$\hat{\gamma}_{12}$	2.56×10^{-5}	0.43	0.085
$\hat{\gamma}_1$	3.91×10^{-5}	0.076	0.074
$\hat{\gamma}_2$	1.79×10^{-6}	0.071	0.096

Table 6.5: Experiment as in Table 6.2, but now $\mu = x_0 = 2.5$, further raising the barrier height to $V(0) = 3.125$, and $b_1 = 5$, $b_2 = -6.5$. The squared relative error SRE$(\hat{\gamma}_{12})$ exceeds bound (6.5) by a factor 5.

Finally, we should note that SRE$(\hat{\gamma}_1)$ is exceeding its bound as well in these last two experiments, although only to a moderate extent. To explain this, we point out that bound (6.5) is based on the product of the inequality

$$\mathrm{SRE}(\hat{p}_k) \leq \frac{1}{r_k - 2} \tag{6.13}$$

over all levels k, where r_k is the number of hits at level k. The simulation results show that every transgression of the bound by SRE$(\hat{\gamma}_1)$ corresponds to violation of (6.13) for the level for which X_t has to traverse the interval $[0, \mu]$. For large μ, the drift $-\nabla V(X_t)$ of X_t rapidly changes at left side of this interval (see Figure 6.4). Hence even for a small time step the probability that X_t enters the next level depends sensitively on the entrance state of the previous level, whereby we cannot assume (6.4) anymore. Nevertheless, the overshoot of SRE$(\hat{\gamma}_1)$ is a factor 1.03 and 1.14,

	Value	SRE$(\cdot)$	Bound for SRE$(\cdot)$
$\hat{\gamma}_{12}$	3.89×10^{-7}	1.42	0.096
$\hat{\gamma}_1$	2.16×10^{-7}	0.121	0.107
$\hat{\gamma}_2$	3.01×10^{-7}	0.078	0.107

Table 6.6: Experiment as in Table 6.2, but now $\mu = x_0 = 3$, further raising the barrier height to $V(0) = 4.5$, and $b_1 = 6$, $b_2 = -7.2$. The squared relative error SRE$(\hat{\gamma}_{12})$ exceeds bound (6.5) by a factor 15.

respectively, which is small compared to that of SRE$(\hat{\gamma}_{12})$ of a factor 5 and 15, respectively.

6.5 CONCLUSION

Estimating the probability of a rare event set consisting of multiple separated subsets by use of a splitting technique is challenging since the most likely path to the rare event set may be very different from the most likely path to an intermediate level. We proposed a separated splitting technique to mitigate the difficulties posed by this path deviation by estimating the probability of hitting each subset individually. For each subset, a modified splitting simulation replicates sample paths that approach this subset, while killing those that enter a subset that is already addressed in a previous simulation.

We compared the separated splitting technique with a standard splitting technique by estimating the probability of entering either of two separated intervals on the real line. Choosing a stochastic process with a potential barrier close to one of the intervals, we show that the standard splitting technique replicates sample paths that approach the subset that is not reached in the end. This efficiency loss — quantified as the squared relative error of the estimator — is observed to be significantly higher when using standard splitting than when using separated splitting. This difference is shown to become more severe for smaller rare event probabilities, illustrating the advantage of the separated splitting technique.

7

A LARGE DEVIATION BASED SPLITTING ESTIMATION OF POWER FLOW RELIABILITY

7.1 INTRODUCTION

Many modern societies have grown accustomed to a very reliable electricity supply by electrical power grids. However, substantial implementation of intermittent renewable generation, such as photovoltaic power or wind power, will form a challenge for grid reliability. Power imbalances caused by generation intermittency may cause grid stability constraints to be violated. Grid operators may even have to curtail power to avoid grid instability, whereby some demanded power consumption (or generation) is not delivered (or produced) in the end. As grid operators are obliged to keep reliability at a prescribed level, they must assess the probability of constraint violations.

A connection overload is an example of such a constraint violation. If an excessive amount of power flows through the connection – e.g., a transmission line or cable —, the connections temperature will eventually exceed its allowed maximum. As a result the connection may get damaged or it may sag and loose tensile strength [94].

Various indices exist to assess grid reliability [7]. Many of these quantify the occurrence of constraint violations that lead to a power curtailment. Examples are the probability, expected duration or expected frequency of violations during a fixed time interval, or the expected induced energy curtailed due to these violations. This paper focuses on the probability that specific connections overload, given the distribution of the multidimensional stochastic process of all uncertain power injections. Overload probabilities are important indices for many long-term investment questions of grid operators. For example, the power grid may

This chapter is based on the submitted article Wadman et al. [93].

not meet reliability standards after a significant amount of renewable generators are integrated in the power grid. Fast and accurate estimation of overload risks will enable otherwise computationally too intensive optimizations of power grid investments. Overload probability estimates will also improve short-term operational strategies as a grid operator can act on these statistics during the next day, hour or even minutes.

Monte Carlo simulation can be used to estimate connection overload probabilities. However, constraint violations causing power curtailments are rare events in modern power grids. In case of a time interval of one week, probabilities of 10^{-5} or even much smaller are not uncommon [11, 13]. Crude Monte Carlo (CMC) estimators for rare event probabilities may require a prohibitively large number of samples to achieve a fixed accuracy. Since one CMC sample already involves solving a large number of high dimensional nonlinear systems, CMC estimation is computationally too intensive for grid reliability analyses in general.

Rare event simulation techniques have been developed for accurate and efficient estimation of very small probabilities. Importance sampling and (multilevel) splitting are two well-known variants. In importance sampling, one samples from an alternative distribution, whereafter the estimator is multiplied by a factor to correct for the induced bias [71]. Crucial for variance reduction is to find a distribution that increases rare event occurrences. Splitting techniques do not change the distribution, but replicate sample paths whenever the rare event is presumed substantially more likely given the current chain state [31, 50]. Crucial for variance reduction is a suitable importance function. Ideally, this function maps the system states to the probability of hitting the rare event set starting from that system state. This is very similar to importance sampling in the sense that information on typical paths to the rare event is desired.

In a simple case with a one-dimensional state space and an interval as the rare event set, the distance to the rare event set will serve as a suitable importance function [50, 92]. In many other cases however, the choice for the importance function is more difficult. In particular, in a high dimensional state space – in this paper: multiple power injections modeled by stochastic processes — it is in general not immediately clear which typical path towards the rare event should be stimulated.

Glasserman et al. [36] show the importance of choosing the levels in a way consistent with the most likely path to a rare set.

In this paper we derive the most likely path towards the rare event of a connection overload using results from large deviations theory. We model power injections as a vector of correlated Ornstein-Uhlenbeck processes in Section 7.2. Section 7.3 introduces the framework of a splitting simulation. We derive an expression for the most likely path towards a connection overload in Section 7.4. We show that the derivation of this path is exact up to a numerical optimization problem with one equality constraint. We show that this optimization problem becomes one-dimensional if we assume a linear mapping between power injections and the connection power flow. We use the corresponding decay rate to construct a suitable importance function for a splitting technique in Section 7.5. We show that this time-dependent importance function is analogous to a time-independent importance function for which Dean and Dupuis [21] show asymptotically optimal performance. The performance of the latter importance function is described in detail in [60]. To reduce the workload we propose three approximations of the decay rate based on different numerical solvers of the optimization problem. We investigate the accuracy and workload of the three corresponding splitting schemes on different stochastic extensions of the IEEE-14 test case in Section 7.6. We compare the performance of the large deviations based splitting simulation to a naive splitting simulation based on merely the proximity of the rare event set in the injection space. We conclude in Section 7.7.

In the literature, rare event simulation techniques have been based on large deviations theory for many applications including finance, engineering, molecular biology and power systems [40, 88, 23, 62]. Hult and Nykvist [43] designs an asymptotically optimal importance sampling scheme to estimate voltage collapse probabilities in a power grid by constructing subsolutions of the Hamilton-Jacobi equations associated with the large deviations of the system. Also splitting techniques have been applied to power systems. Wang et al. [95] estimated small probabilities of instantaneous, unforeseen failures of grid components. Our work though considers the rare event of a connection overload over a certain time domain due to (and given) the uncertain nature of generation. Second,

Schlapfer and Mancarella [75] estimated the probability of a transmission line temperature exceeding a critical value. Markov processes with a discrete state space are used whereas we use a continuous state space. Furthermore, our importance function is based on the asymptotic overload probability, and not on the proximity of the constraint state variable to its allowed maximum. In Wadman et al. [91] a splitting technique is applied to estimate various grid reliability indices. Although current paper focuses on connection overload probabilities only, the large deviations approach enables accurate estimates even for high-dimensional state spaces (see Section 7.6.2) and in cases where the shortest path to the rare event is much less likely than the most likely path (see Section 7.6.4).

7.2 THE POWER FLOW MODEL

Let the following vector-valued stochastic process $\{X^\varepsilon(t), t \geq 0\}$ denote n uncertain power injections of the power grid as function of time t, defined as a multidimensional Ornstein-Uhlenbeck (OU) process:

$$dX^\varepsilon(t) = D(\mu - X^\varepsilon(t))dt + \sqrt{\varepsilon}LdW(t), X^\varepsilon(0) = x_0. \tag{7.1}$$

Here $D := \text{diag}(\theta_1, \ldots, \theta_n) \in \mathbb{R}^{n\times n}$ is a diagonal matrix with mean-reverting terms $\theta_1, \ldots, \theta_n > 0$ on the diagonal. The vector of long-term means is denoted by $\mu \in \mathbb{R}^n$, $\varepsilon > 0$ is a scalar, $L \in \mathbb{R}^{n\times n}$ is a lower triangular matrix with $\Sigma = LL^\top \in \mathbb{R}^{n\times n}$ the covariance matrix of $LW(1)$, and $W(t)$ is a vector of i.i.d. Brownian motions. Then $X_i^\varepsilon(t)$ is clearly a one-dimensional OU process with mean-reverting term θ_i, long-term mean μ_i, volatility $\sqrt{\varepsilon\Sigma_{ii}}$ and initial value $x_{0,i}$. The uncertain pattern of power injection i will therefore deviate according to $\sqrt{\varepsilon\Sigma_{ii}}$ but will be reverted back to mean μ_i with force θ_i. For example, wind power can — especially on short time scales — be assumed to deviate from but also attracted to some historical average. Furthermore, the model incorporates dependencies through L between different power injections, reflecting the correlation between the meteorological sources of renewable energy or between consumption at different nodes.

We should note that the power injections state space is unbounded in this model, whereas in fact this state space is bounded for existing

generators and consumers. On a larger scale however, many sources of generation and consumption are aggregated per node, so using the central limit theorem their net power injection at time t can be modeled realistically by an unbounded (normal) distribution.

We define the function $p : \mathbb{R}^n \to \mathbb{R}$ that maps the power injections to the power flowing through a specific grid connection. A common choice for p involves the AC power flow equations, which are described in detail in Grainger and Stevenson [39, Chapter 9]. In short, a nonlinear algebraic system of steady state equations relates the power injections at each grid node to the voltages at all nodes. To compute a connection power flow at some time t given the power injections at that time, this nonlinear algebraic system has to be solved numerically for the nodal voltages. Then Ohm's law and the definition of power will immediately yield the power flow through a connection.

Another choice for p is a linear function of the power injections to the power flow through the connection of interest:

$$p(x) = v^\top x, \tag{7.2}$$

for some constant vector $v \in \mathbb{R}^n$ depending on the grid connection admittances. The DC power flow equations form a well-known example [39], but also for radial AC networks linear functions have been derived [54]. Most results in this paper are derived and experimentally tested assuming p to be linear. However, as some results also hold for nonlinear p, we assume for now only that p is a deterministic and continuous function of x that solves a system of steady state equations, and we will mention it in later sections whenever we further assume linearity as in (7.2).

We are interested in the overload probability

$$\gamma := \mathbb{P}\left\{ \sup_{\tau \in (0,T]} p(X^\varepsilon(\tau)) \geq P_{\max} \right\}$$

Actually $p(x) = v^\top x + b$ is the correct form, but we as we will only consider probabilities of exceeding a constant barrier, we can set $b = 0$ without loss of generality.

before some time $T > 0$, where $P_{\max} > 0$ is the maximum allowed value of power flowing through the connection. As X^ε and p are continuous, we have

$$\gamma = \mathbb{P}\left\{\exists\ \tau \in (0, T] : p(X^\varepsilon(\tau)) = P_{\max}\right\}. \tag{7.3}$$

The scalar ε is known as the rarity parameter, since for fixed T this probability goes to zero for vanishing ε.

7.3 THE SPLITTING TECHNIQUE

To estimate (7.3) using CMC simulation, we sample trajectories from the discretization of (7.1) and check if $p(X^\varepsilon(t)) > P_{\max}$ at any time step. In this chapter we assume that the time step is sufficiently small to neglect potential rare event hits between subsequent discrete times. Then the CMC estimator

$$\hat{\gamma}_{\text{CMC}} := \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{\exists\ \tau\in(0,T]:p(X^\varepsilon(\tau))\geq P_{\max} \text{ in sample } i\}}$$

for γ is unbiased. However, its squared relative error

$$\frac{\operatorname{Var}\hat{\gamma}_{\text{CMC}}}{\gamma^2} = \frac{\gamma(1-\gamma)}{\gamma^2 N} = \frac{1-\gamma}{\gamma N} \tag{7.4}$$

diverges to infinity as $\mathcal{O}(1/\gamma)$ when N is fixed and $\gamma \to 0$. Therefore, to estimate a very small probability using CMC simulation, one may need a prohibitively large number of samples. Multilevel splitting, or (importance) splitting, is a rare event simulation technique developed to decrease this computational burden.

In this chapter, we will use the Fixed Number of Successes (FNS) splitting variant as described in Section 3.3.2. That is, we start by defining an importance function $h : [0, T] \times \mathbb{R}^n \mapsto \mathbb{R}$. It is constructed such that

The probability $\mathbb{P}\left\{\inf_{\tau\leq T} p(X^\varepsilon(\tau)) \leq -P_{\max}\right\}$ is equally important for the grid operator, but is omitted here as this is a problem of completely similar complexity.

$h(t,x) \geq 1$ precisely when (t,x) corresponds to a rare event occurrence and $h(0,x_0) = 0$. We decompose the rare event probability

$$\gamma = \prod_{k=1}^{m} \mathbb{P}(D_k | D_{k-1})$$

with m levels, event $D_k := \{T_k < T\}$, stopping time $T_k := \inf\{t > 0 : h(X^\varepsilon(t)) \geq l_k\}$ and thresholds $0 = l_0 < l_1 < \cdots < l_m = 1$. Then the FNS estimator

$$\hat{\gamma} := \prod_{k=1}^{m} \frac{r_k - 1}{n_{k-1} - 1},$$

with r_k, n_k the number of hits and trials, respectively, at level k, is an unbiased estimator for γ [2].

The choice for the importance function is crucial for variance reduction of the splitting estimator [31, 50]. Intuitively, the importance function should 'reward good behavior' by splitting sample paths that are more likely to hit the rare event set. The levels should be chosen in a way consistent with the most likely path to the rare event set [36]. Garvels et al. [32] propose to use the importance function equal to (an increasing function of) the rare event probability given that one starts at the considered system state. As knowing this probability would defeat the point of using simulation, the lesson is to find an importance function that is close to this probability. We will use a result from large deviations theory to find an asymptotic probability of the rare event, in the limit of rarity parameter ε.

7.4 RESULTS FROM LARGE DEVIATIONS THEORY

The following derivation is inspired by Bosman et al. [8]. Using the Freidlin–Wentzell theorem [79], the authors derive the decay rate and most likely path of the limiting barrier crossing probability of a one-dimensional OU process before a fixed end time. We generalize this work to a function of multiple correlated OU processes (7.1), and use the result to construct a suitable importance function in Section 7.5.

The multidimensional stochastic process $\{X^\varepsilon(t)\}$ satisfies a sample path large deviation principle with good rate function $I_{0,x_0}(x)$ defined by

$$I_{s,x_s}(x) = \begin{cases} \frac{1}{2}\int_s^T \mathcal{L}(x,x')dt & \text{if } x \in H_{s,x_s}, \\ \infty & \text{if } x \notin H_{s,x_s}, \end{cases} \quad \text{for } s \in [0,T), x_s \in \mathbb{R}^n$$

with Lagrangian $\mathcal{L}(x,x') = u^\top u$, $u = u(x,x') := L^{-1}(x' + D(x-\mu))$, $x' := \frac{dx}{dt}$, and

$$H_{s,x_s} = \{x : [s,T] \mapsto \mathbb{R}^n \text{ such that for all } i, x_i \in C[s,T], \ldots \\ x_i(t) = x_s + \int_s^t \phi_i(z)dz, \phi_i \in L_2[s,T]i\Big\}.$$

$$I^*(s,x_s) := \inf_{\tau\in(s,T], x\in H_{s,x_s}: p(x(\tau))=P_{\max}} I_{s,x_s}(x)$$

of the good rate functions over all paths x that start in x_s at time $s \in [0,T)$ and that exhibit an overload $p(x(\tau)) = P_{\max}$ at some time τ in the remaining time interval $(s,T]$. Then a combination of the Freidlin–Wentzell theorem and the contraction principle [79] implies

$$\lim_{\varepsilon\downarrow 0} -\varepsilon \log \mathbb{P}\left\{\exists\, \tau \in (s,T] : p(X^\varepsilon(\tau)) = P_{\max} \middle| X^\varepsilon(s) = x_s\right\} = I^*(s,x_s).$$

So in specific, the minimum good rate function of the probability in (7.3) converges $I^*(0,x_0)$, advocating the following approximation for small ε:

$$\gamma \approx e^{-I^*(0,x_0)/\varepsilon}.$$

Therefore, the minimum good rate function is also called the decay rate. We should note that a subexponential factor times $e^{-I^*(0,x_0)/\varepsilon}$ would be a more accurate approximation, but as the exponential function dominates this factor as ε vanishes, we neglect it here [86]. This approximation may serve as a first rough guess to distinguish connections with a significant overload probability. Furthermore, the most likely path will shed light on the typical combination of power injection paths that leads to an overload. We will use this approximation to construct a suitable importance function for a splitting simulation in Section 7.5.

7.4.1 *Minimizing the good rate function*

We will write the decay rate $I^*(s, x_s)$ as an minimization over $\tau \in (s, T]$ of an infimum $g_{s,x_s}(\tau)$ for general τ:

$$I^*(s, x_s) = \inf_{\tau \in (s,T]} g_{s,x_s}(\tau), \tag{7.5}$$

with

$$g_{s,x_s}(\tau) := \inf_{x \in H_{s,x_s} : p(x(\tau)) = P_{\max}} I_{s,x_s}(x).$$

We use this formulation since we will first derive $g_{0,x_0}(\tau)$ for general τ and then show that derivative $dg_{0,\mu}/d\tau(\tau) < 0$ for all $\tau \in (s, T]$, implying that $g_{0,\mu}(\tau)$ is smallest in $\tau = T$. This means that the most likely path from the mean to the rare event set enters the rare event set at the latest possible time.

Since the event $p(x(\tau)) = P_{\max}$ only depends on the path x up until $t = \tau$, we have

$$\begin{aligned} g_{0,x_0}(\tau) &= \inf_{x \in H_{0,x_0} : p(x(\tau)) = P_{\max}} \frac{1}{2} \left(\int_0^\tau \mathcal{L}(x, x')dt + \int_\tau^T \mathcal{L}(x, x')dt \right) \\ &= \frac{1}{2} \inf_{x \in H_{0,x_0} : p(x(\tau)) = P_{\max}} \int_0^\tau \mathcal{L}(x, x')dt. \end{aligned} \tag{7.6}$$

Necessary conditions for x are the Euler-Lagrange equations, which are in vector form as follows:

$$\nabla_x \mathcal{L} - \frac{d}{dt} (\nabla_{x'} \mathcal{L}) = 0, \tag{7.7}$$

where $\nabla_x \mathcal{L}$ and $\nabla_{x'} \mathcal{L}$ are the gradients of $\mathcal{L}(x, x')$ w.r.t. x and x', respectively. Elementary calculus yields

$$\begin{aligned} \nabla_x \mathcal{L} &= 2(\nabla_x u)^\top u \\ &= 2(L^{-1}D)^\top L^{-1}(x' + Dx - D\mu) \\ &= 2D\Sigma^{-1}(x' + Dx - D\mu), \end{aligned}$$

$$\begin{aligned}\nabla_{x'}\mathcal{L} &= 2(\nabla_{x'}u)^\top u \\ &= 2L^{-\top}L^{-1}(x' + Dx - D\mu) \\ &= 2\Sigma^{-1}(x' + Dx - D\mu), \\ \frac{d}{dt}(\nabla_{x'}\mathcal{L}) &= 2\Sigma^{-1}(x'' + Dx').\end{aligned}$$

Therefore, (7.7) becomes after some rearrangements

$$x'' = (\Sigma D\Sigma^{-1} - D)x' + \Sigma D\Sigma^{-1}Dx - \Sigma D\Sigma^{-1}D\mu. \tag{7.8}$$

In the one-dimensional case, the first order derivative term vanishes and the ODE in the proof of Proposition 2.2 in [42] arises. Equation (7.8) is a system of second order nonhomogeneous linear differential equations, and its homogeneous counterpart can be written as a first order system:

$$y' = My, \tag{7.9}$$

with

$$\begin{aligned}M &= \begin{pmatrix} O & I \\ \Sigma D\Sigma^{-1}D & \Sigma D\Sigma^{-1} - D \end{pmatrix} \in \mathbb{R}^{2n\times 2n} \\ y &= \begin{pmatrix} x_h \\ x_h' \end{pmatrix},\end{aligned} \tag{7.10}$$

zero matrix $O \in \mathbb{R}^{n\times n}$ and identity matrix $I \in \mathbb{R}^{n\times n}$. To find the eigenvalues and eigenvectors of M, we have to solve

$$\begin{pmatrix} O & I \\ \Sigma D\Sigma^{-1}D & \Sigma D\Sigma^{-1} - D \end{pmatrix} \begin{pmatrix} w \\ \bar{w} \end{pmatrix} = \lambda \begin{pmatrix} w \\ \bar{w} \end{pmatrix}$$

for $\lambda \in \mathbb{R}, w, \bar{w} \in \mathbb{R}^n$. The upper block equation reads $\bar{w} = \lambda w$, so each eigenvector will be of the form $(w,\ \lambda_i w)^\top$. Substituting $\bar{w} = \lambda w$ in the lower block equation yields the characteristic polynomial

$$(\Sigma D\Sigma^{-1} - \lambda I)(D + \lambda I)w = 0. \tag{7.11}$$

Any eigenvalue of $-D$ for λ together with a corresponding eigenvector for w would obviously solve this equation. As $-D$ is diagonal, these eigenvalues are $-D_{11}, \ldots, -D_{nn}$ with the standard unit vectors $e_1, \ldots, e_n$ as corresponding eigenvectors. Therefore, for $i = 1, \ldots, n$, $\lambda_i := -D_{ii}$ is an eigenvalue of M and $(w_i, \bar{w}_i)^\top := (e_i, -D_{ii}e_i)^\top$ is the corresponding eigenvector. Note that here $w_i \in \mathbb{R}^n$ denotes the i-th eigenvector, and not to the i-th element of a vector. Likewise, any eigenvalue of $\Sigma D \Sigma^{-1}$ for λ with the corresponding eigenvector for $(D + \lambda I)w$ would also solve characteristic polynomial (7.11). As $\Sigma D \Sigma^{-1}$ and D are similar matrices, $\Sigma D \Sigma^{-1}$ has eigenvalue D_{ii} with eigenvector Σe_i for $i = 1, \ldots, n$. Therefore, for $i = 1, \ldots, n$, $\lambda_{n+i} := D_{ii}$ is an eigenvalue of M and $(w_{n+i}, \bar{w}_{n+i})^\top := ((D+D_{ii}I)^{-1}\Sigma e_i, (D+D_{ii}I)^{-1}D_{ii}\Sigma e_i)^\top$ is the corresponding eigenvector. So now we have specified the general solution

$$y(t) = \sum_{i=1}^{2n} c_i e^{\lambda_i t} \begin{pmatrix} w_i \\ \bar{w}_i \end{pmatrix}$$

of (7.9) up to constants $c_1, \ldots, c_{2n}$, since for all $i = 1, \ldots, n$

$$\begin{aligned} \lambda_i = -D_{ii}, \quad \begin{pmatrix} w_i \\ \bar{w}_i \end{pmatrix} &= \begin{pmatrix} e_i \\ -D_{ii}e_i \end{pmatrix}, \\ \lambda_{n+i} = \quad D_{ii}, \quad \begin{pmatrix} w_{n+i} \\ \bar{w}_{n+i} \end{pmatrix} &= \begin{pmatrix} (D + D_{ii}I)^{-1}\Sigma e_i \\ D_{ii}(D + D_{ii}I)^{-1}\Sigma e_i \end{pmatrix}. \end{aligned}$$

The homogeneous solution x_h of (7.8) is by definition (7.10) the subvector with the first n elements of y, or in matrix-vector form

$$\begin{aligned} x_h(t) &= \begin{pmatrix} w_1 & \ldots & w_{2n} \end{pmatrix} \operatorname{diag}\left(e^{\lambda_1}, \ldots, e^{\lambda_{2n}}\right) \begin{pmatrix} \bar{c} \\ c \end{pmatrix} \\ &= \begin{pmatrix} I & V \end{pmatrix} \begin{pmatrix} e^{-Dt} & O \\ O & e^{Dt} \end{pmatrix} \begin{pmatrix} \bar{c} \\ c \end{pmatrix} \end{aligned}$$

with $\bar{c} = (c_1 \ldots c_n)^\top$, $c = (c_{n+1} \ldots c_{2n})^\top$ and matrix $V \in \mathbb{R}^{n \times n}$ given by

$$V_{ij} = \frac{\Sigma_{ij}}{\theta_i + \theta_j}. \tag{7.12}$$

A particular solution of (7.8) is $x_p = \mu$. Therefore, the general solution of (7.8) is

$$x(t) = x_h(t) + x_p = e^{-Dt}\bar{c} + Ve^{Dt}c + \mu.$$

One can verify that indeed $x \in H_{0,x_0}$. The initial condition $x(0) = x_0$ determines $\bar{c}$:

$$x(t) = (Ve^{Dt} - e^{-Dt}V)c + e^{-Dt}(x_0 - \mu) + \mu. \tag{7.13}$$

We conclude that the optimal path is of the above form, with $c \in \mathbb{R}$ such that $p(x(\tau)) = P_{\max}$. However, since $p(x(\tau)) = P_{\max}$ is only one equation whereas the degrees of freedom for c is n, the Euler-Lagrange equations are in general not sufficient to find infimum (7.6). Therefore, we substitute (7.13) in (7.6) and further minimize the resulting objective function under the constraint $p(x(\tau)) = P_{\max}$. That is, since $x'(t) = (VDe^{Dt} + De^{-Dt}V)c - De^{-Dt}(x_0 - \mu)$, we have

$$\begin{aligned} u(x, x') &= L^{-1}\left(VDe^{-Dt} + De^{-Dt}V + DVe^{Dt} - De^{Dt}V\right)c \\ &= L^{-1}\left(VD + DV\right)e^{Dt}c \\ &= L^{-1}\Sigma e^{Dt}c \\ &= L^{\top}e^{Dt}c. \end{aligned}$$

In the third equality, $VD + DV = \Sigma$ follows directly from the definitions of V and D. The objective function in (7.6) becomes

$$\begin{aligned} \int_0^\tau \mathcal{L}(x, x')dt &= \int_0^\tau u(x, x')^{\top}u(x, x')dt \\ &= \int_0^\tau c^{\top}e^{Dt}LL^{\top}e^{Dt}c\ dt \\ &= c^{\top}\int_0^\tau e^{Dt}\Sigma e^{Dt}dt\ c \\ &= c^{\top}\left(e^{D\tau}Ve^{D\tau} - V\right)c \\ &= c^{\top}e^{D\tau}\left(V - e^{-D\tau}Ve^{-D\tau}\right)e^{D\tau}c \\ &= c^{\top}e^{D\tau}\,\mathrm{Cov}(X^1(\tau))e^{D\tau}c. \end{aligned}$$

We used symmetry of the diagonal matrix exponential in the second equality and the integral in the fourth equality is easily derived elementwise. The covariance identity in the last equality, where X^1 denotes X^ε for $\varepsilon = 1$ is shown elementwise in (7.32) in the Appendix. After substitution in (7.6) we obtain the minimization problem

$$g_{0,x_0}(\tau) = \inf_{c\in\mathbb{R}^n \,:\, p(x_c(\tau))=P_{\max}} \frac{1}{2} c^\top e^{D\tau} \operatorname{Cov}(X^1(\tau)) e^{D\tau} c \tag{7.14}$$

over c, where we introduced $x_c(t) := x(t)$ as defined in (7.13) to emphasize the dependence on c. Note that the objective function is quadratic in c: if the constraint is linear, this optimization problem would be a quadratic programming problem.

7.4.2 *Starting from the mean the rare event is most likely at the end time*

Now assume $x_0 = \mu$ — i.e., all processes start at their long-term mean — and define

$$b := \left(V e^{D\tau} - e^{-D\tau} V \right) c = \operatorname{Cov}(X^1(\tau)) e^{D\tau} c.$$

In this case minimization problem (7.14) becomes

$$g_{0,\mu}(\tau) = \frac{1}{2} \inf_{b\in\mathbb{R}^n : p(b+\mu)=P_{\max}} b^\top \operatorname{Cov}(X^1(\tau))^{-1} b.$$

The optimal value b_{opt} for b clearly solves $p(b_{\text{opt}} + \mu) = P_{\max}$ so it does not depend on τ. The corresponding decay rate becomes

$$I^*(0,\mu) = \frac{1}{2} \inf_{\tau\in(0,T]} b_{\text{opt}}^\top \operatorname{Cov}(X^1(\tau))^{-1} b_{\text{opt}}. \tag{7.15}$$

We differentiate the objective function to τ:

$$\begin{aligned}
&\frac{d}{d\tau} b_{\text{opt}}^\top \operatorname{Cov}(X^1(\tau))^{-1} b_{\text{opt}} \\
&\quad = b_{\text{opt}}^\top \frac{d \operatorname{Cov}(X^1(\tau))^{-1}}{d\tau} b_{\text{opt}} \\
&\quad = -b_{\text{opt}}^\top \operatorname{Cov}(X^1(\tau))^{-1} \frac{d \operatorname{Cov}(X^1(\tau))}{d\tau} \operatorname{Cov}(X^1(\tau))^{-1} b_{\text{opt}} \\
&\quad = -b_{\text{opt}}^\top \operatorname{Cov}(X^1(\tau))^{-1} \frac{d(V - e^{-D\tau} V e^{-D\tau})}{d\tau} \operatorname{Cov}(X^1(\tau))^{-1} b_{\text{opt}} \\
&\quad = -b_{\text{opt}}^\top \operatorname{Cov}(X^1(\tau))^{-1} e^{-D\tau}(DV + VD) e^{-D\tau} \operatorname{Cov}(X^1(\tau))^{-1} b_{\text{opt}} \\
&\quad = -b_{\text{opt}}^\top \operatorname{Cov}(X^1(\tau))^{-1} e^{-D\tau} \Sigma e^{-D\tau} \operatorname{Cov}(X^1(\tau))^{-1} b_{\text{opt}}. \qquad (7.16)
\end{aligned}$$

We used the identity for the derivative of a matrix inverse in the second equality, expression (7.32) in the third one, and the fact that diagonal matrices commute in the fourth one. We continue using the property that for an invertible matrix A and positive definite matrix B, $A^\top B A$ is positive definite too. Since Σ is a covariance matrix it is positive semi-definite, and without loss of generality we can assume it is positive definite. As $e^{-D\tau} \operatorname{Cov}(X^1(\tau))^{-1}$ is clearly invertible, the matrix

$$\operatorname{Cov}(X^1(\tau))^{-1} e^{-D\tau} \Sigma e^{-D\tau} \operatorname{Cov}(X^1(\tau))^{-1}$$

is positive definite, so (7.16) is strictly negative for general b_{opt}. Hence, the objective function in (7.15) strictly decreases in τ, so the minimizer $\tau^* = T$ is the latest possible time, and thus we conclude:

$$I^*(0, \mu) = g_{0,\mu}(T).$$

7.4.3 *Quadratic programming assuming linear power flow equations*

If the constraint function p is linear — implying linear power flow equations —, the minimization problem (7.14) has a closed-form solution. Assuming $p(x) = v^\top x$ as in (7.2), optimization program (7.14) becomes

$$g_{0,x_0}(\tau) = \inf_{c \in \mathbb{R}^n \,:\, v^\top \operatorname{Cov}(X^1(\tau)) e^{D\tau} c = a} \frac{1}{2} c^\top e^{D\tau} \operatorname{Cov}(X^1(\tau)) e^{D\tau} c,$$

with $a = a(\tau) := P_{\max} - v^\top \left(\mu + e^{-D\tau}(x_0 - \mu)\right)$. The minimizer c^* of this convex quadratic programming problem with one linear constraint is the solution of

$$\begin{pmatrix} e^{D\tau} \operatorname{Cov} X^1(\tau) e^{D\tau} & e^{D\tau} \operatorname{Cov} X^1(\tau) v \\ v^\top \operatorname{Cov} X^1(\tau) e^{D\tau} & 0 \end{pmatrix} \begin{pmatrix} c^* \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ a \end{pmatrix},$$

with $\lambda \in \mathbb{R}$ a (redundant) Lagrange multiplier [61]. Using an identity for the inverse of a block matrix, we obtain the expression

$$c^* = a \frac{e^{-D\tau} v}{v^\top \operatorname{Cov} X^1(\tau) v} \tag{7.17}$$

for the minimizer with corresponding minimum

$$g_{0,x_0}(\tau) = \frac{1}{2} \frac{a^2}{v^\top \operatorname{Cov} X^1(\tau) v}. \tag{7.18}$$

By differentiating g_{0,x_0} to τ,

$$\frac{dg_{0,x_0}}{d\tau} = \frac{1}{2} \frac{2a v^\top \operatorname{Cov} X^1(\tau) v v^\top D e^{-D\tau}(x_0 - \mu) - a^2 v^\top e^{-D\tau} \Sigma e^{-D\tau} v}{(v^\top \operatorname{Cov} X^1(\tau) v)^2},$$

we confirm the result of Section 7.4.2 that if $x_0 = \mu$ the end time is the most likely time for the rare event to occur: again using properties of positive definite matrices, it is readily checked that $g_{0,\mu}(\tau)$ indeed decreases in τ, so the decay rate becomes

$$I^*(0, \mu) = g_{0,\mu}(T) = \frac{1}{2} \frac{(P_{\max} - v^\top \mu)^2}{v^\top \operatorname{Cov} X^1(T) v}. \tag{7.19}$$

For general $x_0 \in \mathbb{R}^n$ however, we do not have this guarantee. Even in the one-dimensional case $n = 1$, one can easily derive that the minimum of (7.18) is attained at

$$\tau^* = -\theta_1^{-1} \log \frac{x_0 - \mu}{P_{\max}/v - \mu}, \tag{7.20}$$

with $\theta_1 = D_{11}$. This root is not defined if $x_0 = \mu$, but for any fixed T, sufficiently close values of x_0 to $P_{\max}/v$ this minimizer τ^* will be

smaller than T. Figure 7.1 illustrates such counterexamples of x_0 for $n = \theta = \mu = \Sigma = v = T = 1$ and $P_{\max} = 5$: indeed $\tau^* < T$ for $x_0 \in [2.4715, 5]$, where the critical value 2.4715 is derived by solving (7.20) for x_0 assuming $\tau^* = T$ and all other parameters as given.

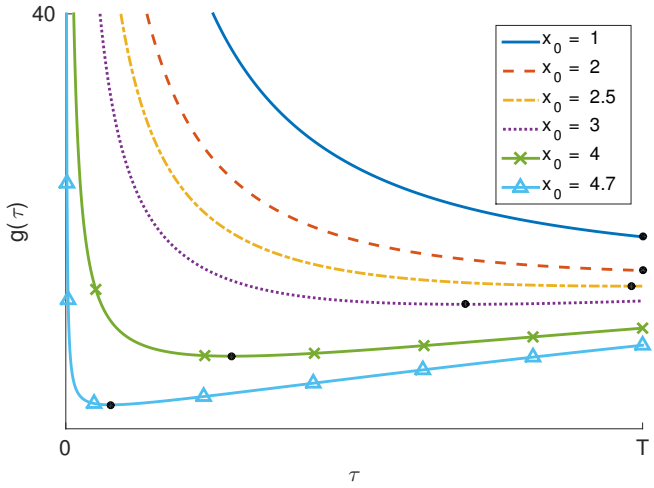


Figure 7.1: For values of x_0 relatively close to rare event border $P_{\max}/v = 5$ compared to $\mu = 1$, the hitting time τ that minimizes decay rate $g(\tau)$ given a hit at τ is not the end time T. Black dots indicate minima $\tau^* = \arg\inf g(\tau)$ over the interval $[0, T]$.

7.5 A LARGE DEVIATIONS BASED IMPORTANCE FUNCTION

In this section we generalize the results of Section 7.4 by conditioning on $X^\epsilon(s) = x_s$ for general $s \in (0, T]$ instead of on $X^\epsilon(0) = x_0$. We are interested in the decay rate $I^*(s, x_s)$ of the limiting probability that, given $X^\epsilon(s) = x_s$ at time s, the rare event $p(X^\epsilon(\tau)) \geq P_{\max}$ will occur at some time $\tau \in (s, T]$ in the remaining time domain. We will use $I^*(s, x_s)$ to compute an approximate probability to hit the rare event given a realized chain state. In turn, we will use this proxy as importance function in a

splitting technique — i.e., to decide whether or not to split the sample path at the corresponding time step.

One can easily derive — i.e., completely analogous to (7.5) - (7.13) —, that the most likely path from $X^\varepsilon(s) = x_s$ to the rare event is of the form

$$x(t) = (Ve^{Dt} - e^{-D(t-s)}Ve^{Ds})c + e^{-D(t-s)}(x_s - \mu) + \mu \tag{7.21}$$

for $t \in [s, T]$. Likewise, for general $s \in (0, T]$ the decay rate

$$I^*(s, x_s) = \inf_{\tau \in (s,T]} g_{s,x_s}(\tau), \tag{7.22}$$

with

$$g_{s,x_s}(\tau) = \frac{1}{2} \inf_{c \in \mathbb{R}^n : p(x(\tau)) = P_{\max}} c^\top e^{D\tau} \operatorname{Cov}(X^1(\tau - s)) e^{D\tau} c \tag{7.23}$$

can be derived analogously to (7.5) – (7.14). If we would assume linear power flow equations $p(x) = v^\top x$ as in Section 7.4.3, the latter minimization has closed-form solution

$$g_{s,x_s}(\tau) = \frac{1}{2} \frac{\left(P_{\max} - v^\top \mu - v^\top e^{-D(\tau - s)}(x_s - \mu)\right)^2}{v^\top \operatorname{Cov} X^1(\tau - s) v}. \tag{7.24}$$

Note that we can write the rare event probability of interest as

$$\begin{aligned}
\gamma &= \mathbb{P}\left\{\exists s \in (0, T] : I^*(s, X^\varepsilon(s)) = 0\right\} \\
&= \mathbb{P}\left\{\exists s \in (0, T] : I^*(s, X^\varepsilon(s)) = 0 \;\middle|\; \exists s \in (0, T] : \ \ldots \right. \\
&\qquad \left. I^*(s, X^\varepsilon(s)) < \alpha I^*(0, x_0)\right\} \mathbb{P}\left\{\exists s \in (0, T] : I^*(s, X^\varepsilon(s)) < \alpha I^*(0, x_0)\right\},
\end{aligned}$$

for some threshold $\alpha \in (0, 1)$. The first equality basically means that for the rare event to occur one should arrive at a system state $(s, X^\varepsilon(s))$ from where it takes 'zero Brownian effort' to arrive in the rare event set. The second equality obviously holds since the conditional event $\{\exists s \in (0, T] : I^*(s, X^\varepsilon(s)) < lI^*(0, x_0)\}$ is a subset of $\{\exists s \in (0, T] : I^*(s, X^\varepsilon(s)) = 0\}$. We iterate this decomposition by first defining thresholds $0 =: l_0 < l_1 < \cdots < l_m := 1$ and events

$$D_k := \left\{\exists s \in (0, T] : 1 - \frac{I^*(s, X^\varepsilon(s))}{I^*(0, x_0)} \geq l_k\right\},$$

for $k = 0, \ldots, m$. Then since $\mathbb{P}\{D_0\} = 1$, $\mathbb{P}\{D_m\} = \gamma$ and $D_0 \supset \ldots \supset D_m$, the following decomposition holds:

$$\gamma = \prod_{k=1}^{m} \mathbb{P}\{D_k | D_{k-1}\}.$$

This decomposition naturally suggests the large deviations based importance function $h : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$h_{\mathrm{LD}}(t, x) := 1 - \frac{I^*(t, x)}{I^*(0, x_0)}. \tag{7.25}$$

In Dean and Dupuis [21], sufficient conditions are derived for an asymptotically optimal performance of a given importance function in a slightly different setting. They consider a stationary stochastic process with values in D and a probability of hitting a rare event set $B \subset D$ before entering another set $A \subset D$. Since both sets are time-independent, the decay rate $I^*(x_s)$ is (and thus important function $h(x)$ may be) time-independent too. The authors show that under appropriate conditions, the asymptotic decay rate of the second moment of the splitting estimator $\hat{\gamma}^\varepsilon$ is optimally

$$\lim_{\varepsilon \downarrow 0} -\varepsilon \log \mathbb{E}[(\hat{\gamma}^\varepsilon)^2] = 2I^*(x_0). \tag{7.26}$$

One condition for this optimality is that the function

$$\bar{W}(x) := \frac{\mathbb{E}[r]}{\Delta}(1 - h(x)),$$

with r the number of splitting particles, Δ the level size, satisfies

$$\bar{W}(x) \leq 0 \tag{7.27}$$

for all $x \in B$, and

$$\bar{W}(x) - \bar{W}(y) \leq \inf_{f,t:f(0)=x,f(t)=y} \int_0^t \mathcal{L}(f, f')du, \tag{7.28}$$

for all $x, y \in D \setminus (A \cup B)$. Dean and Dupuis [21] call such a function a subsolution as it is the subsolution of the related Hamilton-Jacobi-Bellman equations. For $\bar{W}(x)$ to satisfy both inequalities one requires an

importance function $h(x)$ of the form $1 - h(x) \propto I^*(x)$. In this case the first inequality holds as $h(x) > 1$ for all rare event set elements $x \in B$. The second inequality can be written as a triangle inequality: the minimum good rate function of a path from point x to B is not larger when going directly than when traversing via point y. Along these lines Miretskiy et al. [60] chooses an importance function equal to the exponential decay rate to estimate a probability of first entrance into a rare set, and they prove asymptotic efficiency of their proposed Fixed Splitting scheme. The importance function (7.25) in our nonstationary setting is equivalent to that in Miretskiy et al. [60] but depends directly on time as well. This heuristic argument suggests that the large deviations based importance function (7.25) has an asymptotic decay rate close to (7.26).

7.5.1 *Approximation of the decay rate: 3 algorithms*

Assuming linear power flow equations $p(x) = v^\top x$, computing $I^*(t, X^\varepsilon(t))$ requires finding the optimal τ in (7.22) – (7.23). Although the search space of the optimization is one-dimensional, the optimization is required at every time step of every sample path in the splitting simulation. The associated workload will therefore form a challenge for the computational advantage of rare event simulation as compared to CMC. To reduce the computational burden we define importance function

$$h_{\mathrm{LD},i} := 1 - \frac{I_i^*(t,x)}{I_i^*(0,x_0)}$$

using the following three approximations $I_1^*(t,x)$, $I_2^*(t,x)$, $I_3^*(t,x)$ of decay rate $I^*(t,x)$:

Approximation 1. *We assume that the most likely time τ^* to enter the rare event set given the current state $(s, X^\varepsilon(s))$ is one of the discrete time steps $s + \Delta, s + 2\Delta \ldots, T$ of the discretization of (7.1). I_1^* denotes the corresponding decay rate approximation and $\hat{\gamma}_1$ the corresponding splitting estimator.*

The assumption in this approximation is reasonable for small step size Δ, and practical for any discrete time implementation of the involved

stochastic processes. It reduces the optimization to computing $g_{s,x_s}(\tau)$ for all $\tau = s+\Delta, s+2\Delta, \ldots, T$. As this assumption is weaker than those in the two subsequent approximations, the relative error of the accompanying splitting estimator will serve as a benchmark and thus it will be compared to that of a CMC estimator in Section 7.6.

Approximation 2. *For any $s < T$, we assume that the most likely time τ^* to enter the rare event set given the current state $(s, X^\varepsilon(s))$ is end time T. I_2^* denotes the corresponding decay rate approximation and $\hat{\gamma}_2$ the corresponding splitting estimator.*

The assumption $\tau^* = T$ avoids the optimization problem (7.22) – (7.23) at each time step s. We expect that this assumption is reasonable for states 'relatively close to μ' (see Figure 7.1). However, for sample paths that approach the rare event set relatively soon in the simulation, the end time T may be a suboptimal hitting time as then the mean-reverting force of the OU processes will require a very unlikely Brownian motion for a relatively long time.

Approximation 3. *We assume that the most likely state $(\tau^*, X^\varepsilon(\tau^*))$ to enter the rare event set given the current state $(s, X^\varepsilon(s))$ is independent of $(s, X^\varepsilon(s))$. So for example if $x_0 = \mu$, for all $(s, X^\varepsilon(s))$ the optimal rare event entrance state is (T, x_T) for some constant $x_T \in \mathbb{R}^n$. I_3^* denotes the corresponding decay rate approximation and $\hat{\gamma}_3$ the corresponding splitting estimator.*

This assumption can be justified as the splitting algorithm will stimulate the paths advancing to the entrance state (τ^*, x_{τ^*}) that is optimal given $(0, X^\varepsilon(0))$. Therefore, the entrance state that is optimal given a later state $(s, X^\varepsilon(s))$ is expected to be close to this (τ^*, x_{τ^*}).

We now derive the optimal hitting point x_T assuming $x_0 = \mu$. As shown in Section 7.4.2, $\tau^* = T$ is the most likely hitting time at $(0, X^\varepsilon(0))$,

explaining the notation x_T instead of x_{τ^*}. Expressions (7.13) and (7.17) then yield the most likely chain state

$$\begin{aligned} x_T := x(T) &= (Ve^{DT} - e^{-DT}V)c^* + \mu \\ &= \mathrm{Cov}(X^1(T))e^{DT}\frac{e^{-DT}v}{v^\top \mathrm{Cov}(X^1(T))v}a + \mu \\ &= \frac{\mathrm{Cov}(X^1(T))v}{v^\top \mathrm{Cov}(X^1(T))v}\left(P_{\max} - v^\top \mu\right) + \mu. \end{aligned}$$

According to Approximation 3, the most likely path from x_s at any time $s \in (0, T]$ will hit the rare event set at x_T. By imposing the boundary condition $x(T) = x_T$ to (7.21), we derive the corresponding approximation $\tilde{x}(t)$ for the most likely path

$$\tilde{x}(t) = A(t)A^{-1}(T)\left(x_T - B(T)\right) + B(t),$$

with

$$\begin{aligned} A(t) &:= Ve^{Dt} - e^{-D(t-s)}Ve^{Ds}, \\ B(t) &:= e^{-D(t-s)}(x_s - \mu) + \mu. \end{aligned}$$

This approximation avoids the optimization problem (7.22) – (7.23) too as c is completely determined by the entrance point x_T. The corresponding approximate decay rate is

$$I_3^*(s, x_s) = w^\top \mathrm{Cov}(X^1(T-s))^{-1}w, \tag{7.29}$$

with $w := x_T - \mu - e^{-D(T-s)}(x_s - \mu)$. The covariance matrix inverses are independent of the chain state x_s so they can be computed for each s before the simulation starts. $\hat{\gamma}_3$ denotes the corresponding splitting estimator.

7.6 EXPERIMENTS

We perform experiments on the IEEE-14 test network [87], which has deterministic power injections at its 14 nodes. At the first n nodes that

have nonzero power injections, we replace these n deterministic power injections P_{det} by OU model (7.1) with $\mu = x_0 = P_{\text{det}}$ — i.e., the processes tend to revert to the original deterministic power injection values. Further parameters are $\varepsilon = 0.1$ and the mean-reverting terms $\theta_i = 1 + (i - 1)/(n - 1)$ increase from 1 to 2. The lower Cholesky factor L is such that covariance matrix $\Sigma = LL^\top = \text{diag}(\Sigma)(\rho \mathbf{1}\mathbf{1}^\top + (1-\rho)I)\text{diag}(\Sigma)$. Here $\rho = .5$ reflects the typically positive correlation of power injections, $\mathbf{1} \in \mathbb{R}^n$ is a vector of ones and the volatilities $\text{diag}(\Sigma)_{ii} = 1 + (i-1)/(n-1)$ of the marginal OU processes also increase from 1 to 2.

We assume DC power flow equations implying $p(x) = v^\top x$ where v depends on the connection under consideration. We use the MATPOWER package in MATLAB to extract the values for v [100]. For each connection, we set the maximum allowed power flow $P_{\max} = C|v^\top \mu|$ equal to a factor $C > 1$ times the average absolute power flow through that connection. All experiments are performed on an Intel Core 2 Quad CPU Q9550 2.83GHz computer in MATLAB R2014b.

7.6.1 *Two nodes with stochastic power injections*

We choose power injections at nodes 2 and 3 to be stochastic, so $n = 2$, and we choose $C = 1.5$. The approximate decay rate and results of a Crude Monte Carlo (CMC) simulation and splitting simulations are displayed in Table 7.1. Each row corresponds to a connection denoted by $i \to j$, and each column contains different probability estimates of connection overloads. To distinguish between excessive power flow in opposite directions, this probability is defined as

$$\gamma = \begin{cases} \mathbb{P}\{\sup_{\tau \in (0,T]} v^\top X^\varepsilon(\tau) \geq C|v^\top \mu|\} & \text{if } i < j, \\ \mathbb{P}\{\inf_{\tau \in (0,T]} v^\top X^\varepsilon(\tau) \leq -C|v^\top \mu|\} & \text{if } i > j. \end{cases} \tag{7.30}$$

The second column in Table 7.1 contains the largest approximate overload probabilities $\tilde{\gamma}_{\text{LD}} := e^{-g_{0,\mu}(T)/\varepsilon}$ of all connections, with $g_{0,\mu}(T)$ as in (7.18). The third column contains CMC estimates $\hat{\gamma}_{\text{CMC}}$ (with the relative error between parentheses) obtained from 10^6 samples. We omitted CMC estimates for which $\tilde{\gamma}_{\text{LD}} < 10^{-6}$ for computational reasons. The

CMC estimates show that $\hat{\gamma}_{LD}$ is reasonably accurate for those overload probabilities we can compare to a CMC estimate. As the computation of $\hat{\gamma}_{LD}$ required the evaluation of (7.19) only, it serves as a suitable first guess to distinguish grid connections that are exposed to significant overloading risks.

However, the accuracy of the single-point approximations $\hat{\gamma}_{LD}$ are unknown. Fortunately, relative errors of multiple splitting estimates give this insight. The fourth column contains the means of 100 splitting estimates $\hat{\gamma}_1$ using the FNS scheme [2] with 100 hits at each level. We omitted splitting estimates for which $\tilde{\gamma}_{\mathrm{LD}} < 10^{-25}$ as knowing such small estimates will have no practical purpose. We chose the number of equidistant thresholds to be the closest integer to $-0.6275 \log \tilde{\gamma}_{\mathrm{LD}}$, following the reasoning in Amrein and Künsch [2]. We estimated the relative errors of the splitting estimators by repeating the simulation 100 times. For all but the last CMC estimate, the corresponding splitting estimates are accurate in the sense that 95% confidence intervals contains the CMC estimates. Although the last CMC estimate does not lie in the corresponding splitting confidence interval, its relative error is quite large — in fact, the confidence intervals of CMC and splitting overlap —, so CMC does not yield a good benchmark for such a small probability.

To illustrate the computational gain of a splitting technique over CMC simulation, we use equation (7.4) to compute the expected number of CMC samples required to obtain an accuracy comparable to that of the splitting estimates. For example, the squared relative error of $\hat{\gamma}_1$ for connection $3 \to 4$ is $0.021^2 \approx 0.00044$. Equation (7.4) suggests that 9.1×10^6 CMC samples will be required to achieve a squared relative error of similar size, whereas the splitting estimator required only 2.8×10^5 samples. This difference of a factor 32 becomes as much as 9.8×10^6 for connection $2 \to 4$, as CMC simulation would require the prohibitively large number of 6.8×10^{12} samples. The computational gain of splitting will be even larger for smaller probabilities.

	$\tilde{\gamma}_{\text{LD}}$	$\hat{\gamma}_{\text{CMC}}$	$\hat{\gamma}_1$
$4 \to 3$	0.72	0.69 (0.000 68)	0.69 (0.0056)
$1 \to 2$	0.11	0.090 (0.0032)	0.091 (0.0097)
$2 \to 3$	0.10	0.10 (0.0030)	0.10 (0.0093)
$5 \to 4$	0.013	0.012 (0.0090)	0.013 (0.016)
$1 \to 5$	0.0018	0.0014 (0.027)	0.0015 (0.022)
$3 \to 4$	2.8E−4	2.6E−4 (0.062)	2.5E−4 (0.021)
$11 \to 10$	7.6E−5	5.3E−5 (0.14)	6.1E−5 (0.026)
$2 \to 4$	8.5E−11	-	6.1E−11 (0.049)
$9 \to 10$	6.7E−14	-	3.3E−14 (0.051)
$6 \to 11$	1.2E−18	-	5.1E−19 (0.083)
$2 \to 5$	4.0E−23	-	1.0E−23 (0.11)
$2 \to 1$	1.6E−24	-	3.7E−25 (0.13)
$3 \to 2$	1.8E−25	-	7.5E−28 (0.13)
$13 \to 14$	8.8E−26	-	-
⋮	⋮	⋮	⋮

Table 7.1: Estimates of highest overloading probabilities γ as in (7.30) for connections $i \to j$ in the DC IEEE-14 test case with $n = 2$ stochastic power injections, and $\varepsilon = 0.1$. Columns contain large deviations approximations $\tilde{\gamma}_{\text{LD}}$, CMC estimates $\hat{\gamma}_{\text{CMC}}$ and the means of 100 large deviations based splitting estimates $\hat{\gamma}_1$, with relative errors of the mean between parentheses.

7.6.2 *Eleven nodes with stochastic power injections*

We increase the number of stochastic nodes to $n = 11$ and repeat the experiment. We choose $C = 20$ in (7.30) to again achieve a wide range of overload probabilities, see Table 7.2. Again $\hat{\gamma}_{LD}$ is reasonably accurate for those probabilities we can compare to a CMC estimate, confirming that it may serve as a reasonable first guess even when a high number of stochastic power injections are involved. For all seven CMC estimates

the 95% confidence intervals obtained by the corresponding splitting estimates contain the CMC estimates. Again achieving the relative error of displayed splitting estimates using a CMC simulation will often require a prohibitively large number of samples.

	$\tilde{\gamma}_{\text{LD}}$	$\hat{\gamma}_{\text{CMC}}$	$\hat{\gamma}_1$
$12 \to 13$	0.049	0.043 (0.0047)	0.043 (0.011)
$13 \to 12$	0.025	0.021 (0.0068)	0.022 (0.012)
$9 \to 10$	0.0060	0.0045 (0.015)	0.0046 (0.019)
$10 \to 9$	0.0019	0.0014 (0.027)	0.0014 (0.021)
$11 \to 10$	$6.4\text{E}{-4}$	$4.9\text{E}{-4}$ (0.045	$5.0\text{E}{-4}$ (0.019)
$10 \to 11$	$1.2\text{E}{-4}$	$9.3\text{E}{-5}$ (0.10)	$9.1\text{E}{-5}$ (0.026)
$9 \to 14$	$3.1\text{E}{-11}$	-	$1.9\text{E}{-11}$ (0.041)
$6 \to 12$	$1.3\text{E}{-11}$	-	$6.3\text{E}{-12}$ (0.043)
$6 \to 11$	$5.2\text{E}{-12}$	-	$2.2\text{E}{-12}$ (0.048)
$14 \to 9$	$1.5\text{E}{-13}$	-	$8.3\text{E}{-14}$ (0.069)
$12 \to 6$	$4.9\text{E}{-14}$	-	$2.1\text{E}{-14}$ (0.056)
$11 \to 6$	$1.6\text{E}{-14}$	-	$6.0\text{E}{-15}$ (0.063)
$8 \to 7$	$1.1\text{E}{-15}$	-	$3.1\text{E}{-16}$ (0.060)
$13 \to 14$	$5.6\text{E}{-17}$	-	$2.7\text{E}{-17}$ (0.085)
$7 \to 8$	$5.4\text{E}{-19}$	-	$1.3\text{E}{-19}$ (0.062)
$14 \to 13$	$1.4\text{E}{-20}$	-	$5.5\text{E}{-21}$ (0.086)
$5 \to 6$	$8.7\text{E}{-25}$	-	$2.5\text{E}{-25}$ (0.11)
$6 \to 13$	$3.2\text{E}{-25}$	-	$1.2\text{E}{-25}$ (0.17)
$7 \to 9$	$3.9\text{E}{-29}$	-	-
⋮	⋮	⋮	⋮

Table 7.2: As in Table 7.1, but now with $n = 11$ stochastic power injections, and $C = 20$.

7.6.3 *Comparison of the three decay rate approximations*

We investigate the performance of three splitting techniques, each using one of the three decay rate approximations I_1^*, I_2^* and I_3^*. As performance measures we will use the relative error and CPU time of the respective splitting estimators $\hat{\gamma}_1, \hat{\gamma}_2$ and $\hat{\gamma}_3$ of overload probability γ of connection $3 \to 4$ in the two-dimensional model as in Section 7.6.1. The sample mean and relative error of 100 realizations of $\hat{\gamma}_i$ are displayed in Table 7.3, using 100 hits at every level in all splitting runs. All 95% confidence intervals implied by the splitting estimates contain the corresponding CMC estimate — which is 2.6×10^{-4}, see Table 7.1. The relative error of $\hat{\gamma}_1$ is smaller than that of $\hat{\gamma}_2$, which is as expected as Approximation 1 is based on a weaker assumption than that of Approximation 2 (see Section 7.5.1). For a similar reason, the difference in relative error between $\hat{\gamma}_2$ and $\hat{\gamma}_3$ is as expected. Fortunately, the increase of the relative error is at most a factor 2.

The higher workload of $\hat{\gamma}_1$ is obviously due to the necessary computation of good rate function (7.24) for all discrete candidates for τ^*. The workload of $\hat{\gamma}_2$ is smaller than that of $\hat{\gamma}_3$: the most demanding step to compute $\hat{\gamma}_3$ is the quadratic product $w^\top \operatorname{Cov}(X^1(T-s))^{-1} w$ in (7.29), whereas to compute $\hat{\gamma}_2$, the quadratic product $v^\top \operatorname{Cov}(X^1(T)) v$ in (7.19) is most demanding. The matrix inverse of the former vector matrix-vector computation depends on the time step, explaining the slightly higher workload of $\hat{\gamma}_3$. As $\hat{\gamma}_3$ is inferior to $\hat{\gamma}_2$ in both accuracy and workload, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are better choices for our proposed large deviations based splitting technique.

We will illustrate the accuracy of Approximation 3 of a constant optimal hitting time $\tau^* = T$ and constant optimal endpoint $x_{\tau^*} = x_T$. We consider connection $2 \to 4$ and again assume the model as in Section 7.6.1. For each entrance state at each level in the I_1^*-based splitting run, we numerically compute the hitting time τ^* and endpoint x_{τ^*} that are most likely given that entrance state. We use fifteen levels and 1000 hits per level. Figure 7.2 displays the histograms of τ^* for four different levels. The optimal hitting time τ^* is relatively close to T for second level entrance states. This can be interpreted as many entrance states still being close to μ —

	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
Sample mean	2.47E−4	2.37E−4	2.40E−4
Relative error	0.019	0.033	0.037
CPU time per run (sec)	2.71	0.564	0.709

Table 7.3: Sample mean and its relative error of 100 splitting estimates $\hat{\gamma}_i$ using corresponding decay rate proxy I_i^*, and the average time to compute one estimate.

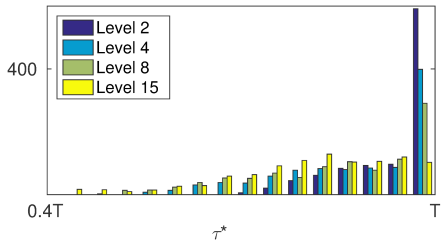


Figure 7.2: The optimal hitting time τ^* given an entrance state at a certain level typically decreases in this level.

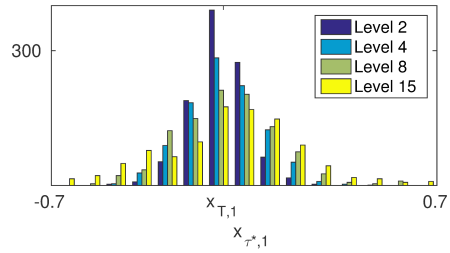


Figure 7.3: The endpoint x_{τ^*} that is most likely given an entrance state at a certain level typically diverges from the endpoint x_T that was initially the most likely.

we have proven in Section 7.4.2 that starting from μ the most likely hitting time is exactly T. For higher levels, typical values for τ^* decrease. This is intuitive since at higher levels samples are more likely to be so close to the rare event set that a rare event occurrence is more likely before the end time than at the end time. This intuition — at least partly — explains the increased relative error of $\hat{\gamma}_2$ and $\hat{\gamma}_3$ in Table 7.3 compared to the relative error of $\hat{\gamma}_1$.

We perform a similar analysis on the optimal hitting point x_{τ^*}. Since the two elements of x_{τ^*} fulfill the linear equation of the rare event, we will only investigate the first element $x_{\tau^*,1}$. Figure 7.3 displays the histogram of $x_{\tau^*,1}$ again for four different levels. As expected, all four empirical

distributions are centered around the endpoint $x_{T,1}$ that was initially the most likely. The variance increases in the considered level; in fact, the sample variance increases monotonically over all 15 levels. We can attribute this observation to the fact that the variance of $X_i^\varepsilon(t)$ increases in time (see equation (7.32) in the Appendix). Therefore, chain states from which the most likely endpoint is far away from $x_{T,1}$ become more likely over time.

7.6.4 *Performance comparison with a naive importance function*

Instead of employing large deviation theory, one could base an importance function on the Euclidean distance of the constraint state variable to its allowed maximum. For example, the importance function

$$h_d(x) = \frac{v^\top x - v^\top x_0}{P_{\max} - v^\top x_0} \tag{7.31}$$

is zero at $x = x_0$ and larger than one if the rare event set is entered. Although this choice for the importance function is intuitive, we will show in an experiment that the choice is naive since it replicates relatively unpromising sample paths. We call function (7.31) the naive importance function and compare it with importance function (7.25) based on decay rate approximation I_2^*. We choose the model with 2 stochastic power injections as in Section 7.6.1, but now it is nodes 3 and 5 that are stochastic and we set $C = 1.3$, $\rho = 0.95$, $D_{22} = \theta_2 = 5$. Figure 7.4 shows the most likely path from $x_0 = \mu$ to the rare event. Because now correlation ρ is very high, path increments diagonally to the upper right and left down are much more likely than diagonally to the upper left or right down. Second, since $\theta_2 \gg \theta_1$, mean reversion of horizontal increments is significantly less powerful than that of vertical increments. For these two reasons, the most likely path differs significantly from the shortest path to the rare event set.

The sample paths hitting a next level are displayed in Figures 7.5 and 7.6 for the two splitting simulations, respectively. We chose only 10 hits per level for clarity reasons. Paths of the large deviations based splitting run stay around the path that is initially the most likely, whereas

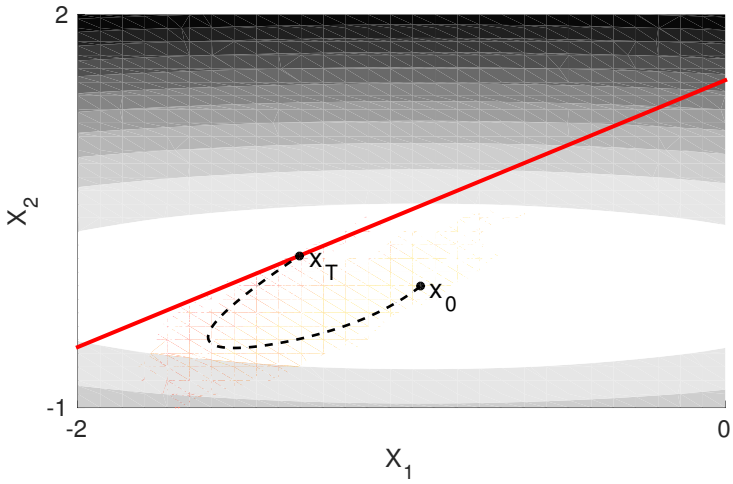


Figure 7.4: The most likely path of a 2-D OU process from $x_0 = \mu$ (dashed line) towards the rare event set (boundary given by solid line). A contour plot of the OU potential $D(\mu - x)^2/2$ is in the background. The OU model parameters are such that the most likely path differs significantly from the shortest path to the rare event.

paths of the naive splitting run deviate to the upper right. This suggests that naive splitting is replicating many paths that are not necessarily promising to hit the rare event set. A CMC simulation with 10^7 samples yielded the estimate 3.83×10^{-5} and 95% confidence interval $[3.45 \times 10^{-5}, 4.21 \times 10^{-5}]$. The large deviation based splitting estimates in Table 7.4 are relatively close to the CMC estimate compared to the naive splitting estimates. Second, the relative error of 100 large deviations based estimates is lower than that of 100 naive splitting estimates, confirming that a significant number of samples are replicated in vain in the naive splitting run. This can be explained by the next statistic in Table 7.4: on average much more samples are required to observe a next level hit, and this difference increases for smaller r. The CPU times exhibit a similar difference. Both differences are intuitive since for a small number of next level hits in the naive run chances are higher that none of them is actually promising from a large deviation perspective. In contrast, the

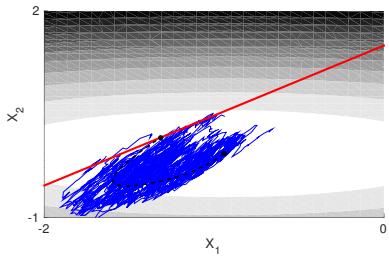


Figure 7.5: Using importance function (7.25), with decay rate proxy I_2^*, paths of a splitting simulation stay around the most likely path (see also Figure 7.4)

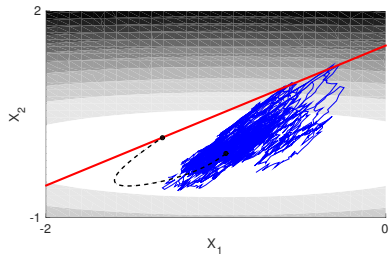


Figure 7.6: Using importance function (7.31), based on the proximity to the rare event set, paths of a splitting simulation deviate significantly from the most likely path (see also Figure 7.4).

large deviation based splitting simulation requires around 6 samples on average to hit the next level, even for a small number of hits per level. In this sense, the workload of $\hat{\gamma}_2$ per level hit is robust in the number of hits per level.

	h_{LD}				h_d			
r	Est	RE	#paths	CPU	Est	RE	#paths	CPU
250	3.94E−5	0.021	5.5	1.7	3.94E−5	0.13	13	2.7
100	3.85E−5	0.040	5.5	0.70	3.70E−5	0.26	20	1.6
25	3.75E−5	0.064	5.6	0.21	1.53E−5	0.36	85	1.1
10	3.81E−5	0.12	5.7	0.11	3.28E−6	0.40	707	1.75

Table 7.4: Performance statistics of 100 splitting estimates using either large deviations based importance function (7.25) or naive importance function (7.31), for different numbers r of hits per level: sample mean, relative error, average number of samples required to hit the next level and average CPU time in seconds for one estimate.

To give a quantification of the workload gain, first note that for $r = 250$ the relative error is a factor $0.13/0.021 = 6.19$ smaller using h_{LD} instead of h_d. Again using expression (7.4), we expect the naive splitting simulation to require a factor $6.19^2 \approx 38$ as many estimates to achieve a relative error similar to that of the large deviation based splitting simulation. This would translate in a total CPU time of $2.7 \times 100 \times 38$ seconds, which is 61 times as much as that of the CPU time (1.7×100 seconds) of the large deviations based splitting simulation. For $r = 100, 25, 10$, this factor of increased CPU time becomes $96.5, 166, 177$, respectively. We conclude that for a fixed accuracy the large deviation based splitting technique computationally more efficient than the naive splitting technique, especially for a relatively small number of hits per level.

Using expression (7.4) and an estimated CPU time of 25×10^{-6} seconds per CMC sample, a similar comparison of naive splitting with CMC can be performed. In fact, for $r = 250, 100, 25, 10$, the naive splitting technique required $11, 25, 33, 67$, respectively, times as much CPU time to obtain the relative errors $0.13, 0.26, 0.36, 0.40$ in Table 7.4, respectively, than CMC simulation would require. So in this case CMC simulation is computationally more efficient than naive splitting. In contrast, large deviation based splitting outperforms CMC for $r = 250, 100, 25, 10$ with factors $5.7, 3.8, 5.0, 2.7$, respectively, in computational efficiency. This efficiency gain will be even more for smaller probabilities.

7.7 CONCLUSION AND OUTLOOK

Based on results from large deviations theory, we developed an importance function for a splitting technique to efficiently estimate overload probabilities of power grid connections. The large deviations approximation serves as a suitable first guess to distinguish connections with significant overload probabilities. For both 2 and 11 stochastic power injections and a fixed accuracy, Crude Monte Carlo would require tens to millions as many samples than the proposed splitting technique required. The assumption that the rare event is most likely at the end

time significantly accelerates the computation with only a modest loss of accuracy.

We showed an example (see Table 7.4) where a naive importance function based on the Euclidean distance to the rare event set replicates many unpromising sample paths. When using a naive importance function the required CPU time to achieve a fixed relative error is over 60 times larger than when using our proposed importance function, justifying the use of a large deviations based splitting technique. In fact, naive splitting — unlike large deviations based splitting — required more (over 11 times as much) CPU time than CMC simulation, illustrating its pitfall.

Interesting further research would be to compare performances of large deviation based importance functions for nonlinear power flow equations. In that case optimization problem (7.22) – (7.23) is multidimensional and has a nonlinear constraint, so solving it each time step will be computationally too intensive for a high dimensional state space. Assuming a fixed end point (see Approximation 3 in Section 7.5.1) avoids the optimization problem and may be sufficiently accurate to save a substantial amount of workload for a fixed relative error. Another aim is to replace the OU process by a diffusion process that incorporates periodicities or an alternative stationary distribution that is typical for generation patterns of renewable energy.

APPENDIX

In this appendix we will derive the first two moments of the multidimensional OU process. Consider the i-th element of the multidimensional OU process in (7.1):

$$dX_i^\varepsilon(t) = \theta_i(\mu_i - X_i^\varepsilon(t))dt + \sqrt{\varepsilon}\sum_{k=1}^{i} L_{ik}dW_k(t), \quad X_i^\varepsilon(0) = x_i^0.$$

Then marginal $X_i^\varepsilon(t)$ is a one dimensional OU process with volatility parameter σ_i given by $\sigma_i^2 = \varepsilon \sum_{k=1}^{i} L_{ik}^2 = \varepsilon \Sigma_{ii}$, and its solution is well-known:

$$X_i^\varepsilon(t) = x_i^0 e^{-\theta_i t} + \mu_i(1 - e^{-\theta_i t}) + \varepsilon \sum_{k=1}^{i} L_{ik} \int_0^t e^{\theta_i(s-t)} dW_k(s).$$

The first two RHS terms are deterministic and the third is a weighted sum of independent Itō integrals with a deterministic integrand. Therefore, all Itō integrals are normally distributed with zero mean and a variance equal to the time integral of the squared integrand, implying

$$\begin{aligned}
\mathbb{E}[X_i^\varepsilon(t)] &= x_i^0 e^{-\theta_i t} + \mu_i(1 - e^{-\theta_i t}), \\
\operatorname{Var}(X_i^\varepsilon(t)) &= \varepsilon \sum_{k=1}^{i} L_{ik}^2 \operatorname{Var} \int_0^t e^{\theta_i(s-t)} dW_k(s) \\
&= \varepsilon \sum_{k=1}^{i} L_{ik}^2 \int_0^t e^{2\theta_i(s-t)} ds \\
&= \varepsilon \frac{\sigma_i^2}{2\theta_i t}(1 - e^{-2\theta_i}).
\end{aligned}$$

As every linear combination of components of $X^\varepsilon(t)$ is univariate normally distributed, $X^\varepsilon(t)$ is multivariate normal. Its expectation is the vector of above marginal expectations, so it remains to find the covariance matrix of $X^\varepsilon(t)$. Assuming $i \leq j$ without loss of generality, the elements of this covariance matrix are

$$\begin{aligned}
&\operatorname{Cov}(X_i^\varepsilon(t), X_j^\varepsilon(t)) \\
&\quad = \mathbb{E}[(X_i^\varepsilon(t) - \mathbb{E}[X_i^\varepsilon(t)])(X_j^\varepsilon(t) - \mathbb{E}[X_j^\varepsilon(t)])] \\
&\quad = \varepsilon \sum_{k=1}^{i} \sum_{l=1}^{j} L_{ik} L_{jl} e^{-(\theta_i + \theta_j)t} \mathbb{E}\left[\int_0^t e^{\theta_i s} dW_k(s) \int_0^t e^{\theta_j s} dW_l(s) \right] \\
&\quad = \varepsilon \sum_{k=1}^{i} L_{ik} L_{jk} e^{-(\theta_i + \theta_j)t} \mathbb{E}\left[\int_0^t e^{\theta_i s} dW_k(s) \int_0^t e^{\theta_j s} dW_k(s) \right],
\end{aligned}$$

where the last equality holds as for $k \neq l$ the two Itō integrals are independent and have a deterministic integrand, so the expectation of their product is zero. The argument of the remaining expectations are quadratic covariations, of which each can be written in terms of three quadratic variations using the polarization identity:

$$\begin{aligned}
&\mathbb{E}\left[\int_0^t e^{\theta_i s} dW_k(s) \int_0^t e^{\theta_j s} dW_k(s)\right] \\
&\quad = \frac{1}{2}\mathbb{E}\left[\left(\int_0^t (e^{\theta_i s} + e^{\theta_j s}) dW_k(s)\right)^2\right. \\
&\qquad \left. - \left(\int_0^t e^{\theta_i s} dW_k(s)\right)^2 - \left(\int_0^t e^{\theta_j s} dW_k(s)\right)^2\right] \\
&\quad = \frac{1}{2}\left(\int_0^t (e^{\theta_i s} + e^{\theta_j s})^2 ds - \int_0^t e^{2\theta_i s} ds - \int_0^t e^{2\theta_j s} ds\right) \\
&\quad = \int_0^t e^{(\theta_i+\theta_j)s} ds = \frac{e^{(\theta_i+\theta_j)t} - 1}{\theta_i + \theta_j}.
\end{aligned}$$

After substitution we conclude that element (i, j) of the covariance matrix of $X^\varepsilon(t)$ is given by

$$\begin{aligned}
\mathrm{Cov}(X_i^\varepsilon(t), X_j^\varepsilon(t)) &= \varepsilon \sum_{k=1}^{i} L_{ik} L_{jk} e^{-(\theta_i+\theta_j)t} \frac{e^{(\theta_i+\theta_j)t} - 1}{\theta_i + \theta_j} \\
&= \varepsilon \Sigma_{ij} \frac{1 - e^{-(\theta_i+\theta_j)t}}{\theta_i + \theta_j}.
\end{aligned}$$

With V as defined in (7.12), we conclude in vector notation:

$$\mathrm{Cov}(X^\varepsilon(t)) = \varepsilon(V - e^{-Dt} V e^{-Dt}). \tag{7.32}$$

BIBLIOGRAPHY

[1] Jamshid Aghaei and Mohammad-Iman Alizadeh. Demand response in smart electricity grids equipped with renewable energy sources: A review. *Renewable and Sustainable Energy Reviews*, 18:64–72, 2013.

[2] Michael Amrein and Hans R Künsch. A variant of importance splitting for rare event estimation: Fixed number of successes. *ACM Transactions on Modeling and Computer Simulation*, 21(2):13:1–13:20, 2011.

[3] Søren Asmussen and Peter W Glynn. *Stochastic simulation: algorithms and analysis*. Springer Science & Business Media, 2007.

[4] Raul Banos, Francisco Manzano-Agugliaro, FG Montoya, Consolacion Gil, Alfredo Alcayde, and Julio Gómez. Optimization methods applied to renewable and sustainable energy: a review. *Renewable and Sustainable Energy Reviews*, 15(4):1753–1766, 2011.

[5] Marc Beaudin, Hamidreza Zareipour, Anthony Schellenberglabe, and William Rosehart. Energy storage for mitigating the variability of renewable electricity sources: An updated review. *Energy for Sustainable Development*, 14(4):302–314, 2010.

[6] Florence Belmudes, Damien Ernst, and Louis Wehenkel. Cross-entropy based rare-event simulation for the identification of dangerous events in power systems. In *Probabilistic Methods Applied to Power Systems, 2008. PMAPS'08. Proceedings of the 10th International Conference on*. IEEE, 2008.

[7] Roy Billinton and Wenyuan Li. *Reliability assessment of electrical power systems using Monte Carlo methods*. Springer, 1994.

[8] Joost Bosman, Jayakrishnan Nair, and Bert Zwart. On the probability of current and temperature overloading in power grids: a large

deviations approach. *ACM SIGMETRICS Performance Evaluation Review*, 42(2):33–35, 2014.

[9] Zdravko I. Botev and Dirk P. Kroese. Efficient Monte Carlo simulation via the generalized splitting method. *Statistics and Computing*, 22(1): 1–16, 2012.

[10] HR Cai, CY Chung, and KP Wong. Application of differential evolution algorithm for transient stability constrained optimal power flow. *Power Systems, IEEE Transactions on*, 23(2):719–728, 2008.

[11] Kevin Carden and Nick Wintermantel. The economic ramifications of resource adequacy. Technical report, Astrape Consulting, 2013.

[12] CEER. On the quality of electricity supply. Technical report, Council of European Energy Regulators, 2011. Available at `www.ceer.eu/`.

[13] CEER. On the continuity of electricity supply. Technical report, Council of European Energy Regulators, 2014. Available at `www.ceer.eu/`.

[14] Michael Chertkov, Feng Pan, and Mikhail G Stepanov. Predicting failures in power grids: The case of static overloads. *Smart Grid, IEEE Transactions on*, 2(1):162–172, 2011.

[15] Michael Chertkov, Mikhail Stepanov, Feng Pan, and Ross Baldick. Exact and efficient algorithm to discover extreme stochastic events in wind generation over transmission power grids. In *50th Conference on Decision and Control and European Control Conference (CDC-ECC)*, pages 2174–2180. IEEE, 2011.

[16] Guo Chun-lin, Wu Li, Wang Dan, Qi Wen-bo, and Xiao Xiang-ning. Impact of electric vehicle charging on power grid. In *Electrical and Control Engineering (ICECE), 2011 International Conference on*, pages 2270–2274. IEEE, 2011.

[17] IEEE PES T&D Committee. IEEE guide for electric power distribution reliability indices (IEEE Std 1366-2003). Technical report, The IEEE Inc., USA, 2004.

[18] JC Das. *Power system analysis: short-circuit load flow and harmonics*. CRC press, 2011.

[19] Pieter-Tjerk de Boer. *Analysis and efficient simulation of queueing models of telecommunication systems*. PhD thesis, University of Twente, Enschede, Netherlands, 2000.

[20] Ailson P de Moura and Adriano Aron F de Moura. Newton–Raphson power flow with constant matrices: a comparison with decoupled power flow methods. *International Journal of Electrical Power & Energy Systems*, 46:108–114, 2013.

[21] Thomas Dean and Paul Dupuis. Splitting for rare event simulation: a large deviation approach to design and analysis. *Stochastic processes and their applications*, 119(2):562–587, 2009.

[22] Pierre Del Moral and Josselin Garnier. Genealogical particle analysis of rare events. *The Annals of Applied Probability*, 15(4):2496–2534, 2005.

[23] Paul Dupuis, Konstantinos Spiliopoulos, and Hui Wang. Importance sampling for multiscale diffusions. *Multiscale Modeling & Simulation*, 10(1):1–27, 2012.

[24] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. *Phys. Rev. B*, 66, 2002. doi: 10.1103/PhysRevB.66.052301. URL http://link.aps.org/doi/10.1103/PhysRevB.66.052301.

[25] Siemens PTD EA. Thermal overload protection of cables. Technical report, Applications for SIPROTEC Protection Relays, 2005. http://sm-industry.ru/titan_img/ecatalog/Appl_07_Thermal_Overload_Protection_en.pdf.

[26] J.S.G. Ehnberg and M.H.J. Bollen. Generation reliability for small isolated power systems entirely based on renewable sources. In *Power Engineering Society General Meeting*, pages 2322–2327. IEEE, 2004.

[27] ENTSOE. 10-year network development plan 2014. Technical report, European Network of Transmission System Operators for Electricity, 2014.

[28] Regional Group Continental Europe and Synchronous Area Great Britain. Solar eclipse 2015 - Impact analysis. Technical report, European Network of Transmission System Operators for Electricity, 2015.

[29] Ingvar B Fridleifsson, Ruggero Bertani, Ernst Huenges, John W Lund, Arni Ragnarsson, and Ladislaus Rybach. The possible role and contribution of geothermal energy to the mitigation of climate change. In *IPCC scoping meeting on renewable energy sources, proceedings, Luebeck, Germany*, pages 59–80. Intergovernmental Panel on Climate Change, 2008.

[30] J. Frunt. *Analysis of balancing requirements in future sustainable and reliable power systems*. PhD thesis, Eindhoven University of Technology, Eindhoven, Netherlands, 2011.

[31] Marnix J.J. Garvels. *The splitting mthod in rare event simulation*. PhD thesis, University of Twente, Enschede, Netherlands, 2000.

[32] Marnix J.J. Garvels, Jan-Kees C.W. Van Ommeren, and Dirk P. Kroese. On the importance function in splitting simulation. *European Transactions on Telecommunications*, 13(4):363–371, 2002.

[33] Daniel T Gillespie. Exact numerical simulation of the ornstein-uhlenbeck process and its integral. *Physical review E*, 54(2):2084–2091, 1996.

[34] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic. Multilevel splitting for estimating rare event probabilities. *Operations Research*, 47(4):585–600, 1999.

[35] Paul Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer Science & Business Media, 2003.

[36] Paul Glasserman, Philip Heidelberger, Perwez Shahabuddin, and Tim Zajic. A large deviations perspective on the efficiency of multilevel splitting. *Automatic Control, IEEE Transactions on*, 43(12): 1666–1679, 1998.

[37] J Duncan Glover, Mulukutla Sarma, and Thomas Overbye. *Power System Analysis & Design, SI Version*. Cengage Learning, 2011.

[38] Reinaldo A González-Fernández and Armando M Leite da Silva. Reliability assessment of time-dependent systems via sequential cross-entropy monte carlo simulation. *Power Systems, IEEE Transactions on*, 26(4):2381–2389, 2011.

[39] John J. Grainger and William D. Stevenson. *Power system analysis*, volume 621. McGraw-Hill, 1994.

[40] Paolo Guasoni and Scott Robertson. Optimal importance sampling with explicit formulas in continuous time. *Finance and Stochastics*, 12 (1):1–19, 2008.

[41] Paul Hines, Seth Blumsack, E Cotilla Sanchez, and Clayton Barrows. The topological and electrical structure of power grids. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, pages 1–10. IEEE, 2010.

[42] Gang Huang, Michel Mandjes, and Peter Spreij. Limit theorems for reflected Ornstein–Uhlenbeck processes. *Statistica Neerlandica*, 68(1): 25–42, 2014.

[43] Henrik Hult and Johan Nykvist. Efficient importance sampling to assess the risk of voltage collapse in power systems. *Preprint*, 2015.

[44] Reijer Idema, Domenico JP Lahaye, Cornelis Vuik, and Lou van der Sluis. Scalable newton-krylov solver for very large power flow problems. *Power Systems, IEEE Transactions on*, 27(1):390–396, 2012.

[45] Reijer Idema, Georgios Papaefthymiou, Domenico Lahaye, Cornelis Vuik, and Lou van der Sluis. Towards faster solution of large power

flow problems. *Power Systems, IEEE Transactions on*, 28(4):4918–4925, 2013.

[46] S Juneja and P Shahabuddin. Rare-event simulation techniques: An introduction and recent advances. *Handbooks in Operations Research and Management Science*, 13:291–350, 2006.

[47] Janghoon Kim, James A Bucklew, and Ian Dobson. Splitting method for speedy simulation of cascading blackouts. *Power Systems, IEEE Transactions on*, 28(3):3010–3017, 2013.

[48] Gerd Kjølle. State of the art on reliability assessment in power systems. Technical report, GARPUR, 2014.

[49] KNMI. Hourly data of the weather in the Netherlands. `www.knmi.nl/klimatologie/uurgegevens/`, 2013.

[50] Pierre. L'Ecuyer, Valérie Demers, and Bruno Tuffin. Splitting for rare-event simulation. In *Proceedings of the Winter Simulation Conference 2006*, pages 137–148. IEEE, 2006.

[51] Armando M Leite da Silva, Reinaldo AG Fernandez, and Chanan Singh. Generating capacity reliability evaluation based on Monte Carlo simulation and cross-entropy methods. *Power Systems, IEEE Transactions on*, 25(1):129–137, 2010.

[52] Whei-Min Lin and Jen-Hao Teng. Three-phase distribution network fast-decoupled power flow solutions. *International Journal of Electrical Power & Energy Systems*, 22(5):375–380, 2000.

[53] A. Lojowska, D. Kurowicka, G. Papaefthymiou, and L. van der Sluis. Advantages of ARMA-GARCH wind speed time series modeling. In *11th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 83–88. IEEE, 2010.

[54] Steven H. Low. Convex relaxation of optimal power flow, part I: formulations and equivalence. *IEEE Transactions on Control of Network Systems*, 1(1):15–27, 2014.

[55] P Luickx, W Vandamme, P Souto Pérez, J Driesen, and W D'haeseleer. Applying Markov chains for the determination of the capacity credit of wind power. In *6th International Conference on the European Energy Market (EEM)*. IEEE, 2009.

[56] Juan M Lujano-Rojas, José L Bernal-Agustín, Rodolfo Dufo-López, and José A Domínguez-Navarro. Forecast of hourly average wind speed using ARMA model with discrete probability transformation. *Electrical Engineering and Control*, 98:1003–1010, 2011.

[57] David MacKay. *Sustainable energy - without the hot air*. UIT Cambridge, 2008.

[58] Neal Noah Madras. *Lectures on Monte Carlo methods*. Springer Science & Business, 2002.

[59] Arthur Mazer. *Electric power planning for regulated and deregulated markets*. John Wiley & Sons, 2007.

[60] Denis I. Miretskiy, Werner R.W. Scheinhardt, and Michel R.H. Mandjes. On efficiency of multilevel splitting. *Communications in Statistics-Simulation and Computation*, 41(6):890–904, 2012.

[61] Katta G. Murty. *Optimization for decision making*. Springer, 2009.

[62] Johan Nykvist. *Topics in importance sampling and derivatives pricing*. PhD thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2015.

[63] James Oswald, Mike Raine, and Hezlin Ashraf-Ball. Will British weather provide reliable electricity? *Energy Policy*, 36(8):3212–3225, 2008.

[64] G. Papaefthymiou and B. Klöckl. MCMC for wind power simulation. *IEEE Transactions on Energy Conversion*, 23(1):234–240, 2008.

[65] H. Pender and W.A. Del Mar. *Electrical engineers' handbook*, volume 2. Wiley, 1949.

[66] Magnus Perninge, Filip Lindskog, and Lennart Soder. Importance sampling of injected powers for electric power system security analysis. *Power Systems, IEEE Transactions on*, 27(1):3–11, 2012.

[67] Ram Rajagopal, Eilyan Bitar, Pravin Varaiya, and Felix Wu. Risk-limiting dispatch for integrating renewable power. *International Journal of Electrical Power & Energy Systems*, 44(1):615–628, 2013.

[68] Anubhav Ratha, Emil Iggland, and Göran Andersson. Value of Lost Load: how much is supply security worth? In *Power and Energy Society General Meeting (PES), 2013 IEEE*. IEEE, 2013.

[69] REN21. Renewables 2010 global status report. Technical report, Renewable Energy Policy Network for the 21st Century, 2010.

[70] REN21. Renewables 2014 global status report. Technical report, Renewable Energy Policy Network for the 21st Century, 2014.

[71] Gerardo Rubino and Bruno Tuffin. *Rare event simulation using Monte Carlo methods*. John Wiley & Sons, 2009.

[72] Reuven Y Rubinstein and Dirk P Kroese. *Simulation and the Monte Carlo method*, volume 707. John Wiley & Sons, 2011.

[73] R.Y. Rubinstein and D.P. Kroese. *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning*. Springer Verlag, 2004.

[74] S. Rusck. The simultaneous demand in distribution networks supplying domestic consumers. *ASEA Journal*, 10:59–61, 1956.

[75] Markus Schlapfer and Pierluigi Mancarella. Probabilistic modeling and simulation of transmission line temperatures under fluctuating power flows. *Power Delivery, IEEE Transactions on*, 26(4):2235–2243, 2011.

[76] John Shortle, Steffen Rebennack, and Fred W Glover. Transmission-capacity expansion for minimizing blackout probabilities. *Power Systems, IEEE Transactions on*, 29(1):43–52, 2014.

[77] John F Shortle. Efficient simulation of blackout probabilities using splitting. *International Journal of Electrical Power & Energy Systems*, 44 (1):743–751, 2013.

[78] KN Shubhanga and AM Kulkarni. Stability-constrained generation rescheduling using energy margin sensitivities. *Power Systems, IEEE Transactions on*, 19(3):1402–1413, 2004.

[79] Adam Shwartz and Alan Weiss. *Large deviations for performance analysis: queues, communication and computing*, volume 5. CRC Press, 1995.

[80] Soliman Abdel-Hady Soliman and Abdel-Aal Hassan Mantawy. *Modern optimization techniques with applications in electric power systems*. Springer Science & Business Media, 2011.

[81] B. Stott and O. Alsac. Fast decoupled load flow. *Power Apparatus and Systems, IEEE Transactions on*, 1(3):859–869, 1974.

[82] S. Strand. The simultaneous demand in distribution networks supplying domestic consumers. *ASEA Journal*, 10:59–61, 1956.

[83] Muthu Kumar Subramanian. Application of holomorphic embedding to the power-flow problem. Master's thesis, Arizona State University, Tempe, Arizona, USA, 2014.

[84] TenneT. Quality and capacity plan 2010-2016, part I. Technical report, TenneT TSO, 2009. Available at `www.tennet.eu/nl/about-tennet/news-press-publications/publications/technical-publications.html`.

[85] G. Torrent-Gironella and J. Fortiana. Simulation of high-dimensional t-Student copulas with a given block correlation matrix. Unpublished, 2005.

[86] Hugo Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1):1–69, 2009.

[87] Electrical Engineering University of Washington. Power systems test case archive. www.ee.washington.edu/research/pstca/, 2006.

[88] Eric Vanden-Eijnden and Jonathan Weare. Rare event simulation of small noise diffusions. *Communications on Pure and Applied Mathematics*, 65(12):1770–1803, 2012.

[89] Manuel Villén-Altamirano and José Villén-Altamirano. On the efficiency of RESTART for multidimensional state systems. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 16(3): 251–279, 2006.

[90] Wander S. Wadman, Gabriël Bloemhof, Daan T. Crommelin, and Jason E. Frank. Probabilistic power flow simulation allowing temporary current overloading (volume to appear). In *Conference proceedings on Probabilistic Methods Applied to Power Systems*, pages 494–499, 2012.

[91] Wander S. Wadman, Daan T. Crommelin, and Jason E. Frank. Applying a splitting technique to estimate electrical grid reliability. In *Winter Simulation Conference proceedings*, pages 577–588, 2013.

[92] Wander S. Wadman, Daan T. Crommelin, and Jason E. Frank. A separated splitting technique for disconnected rare event sets. In *Winter Simulation Conference proceedings*, pages 522–532, 2014.

[93] Wander S. Wadman, Daan T. Crommelin, and Bert P. Zwart. A large deviation based splitting estimation of power flow reliability. *Submitted (23 pages)*, 2015.

[94] Hua Wan, James D. McCalley, and Vijay Vittal. Increasing thermal rating by risk analysis. *Power Systems, IEEE Transactions on*, 14(3): 815–828, 1999.

[95] Sing-Po Wang, Argon Chen, Chih-Wen Liu, Chun-Hung Chen, and John Shortle. Rare-event splitting simulation for analysis of power system blackouts. In *Power and Energy Society General Meeting, 2011 IEEE*, pages 1–7. IEEE, 2011.

[96] W. Wangdee. *Bulk electric system reliability simulation and application*. PhD thesis, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, 2005.

[97] Wijarn Wangdee and Roy Billinton. Considering load-carrying capability and wind speed correlation of WECS in generation adequacy assessment. *Energy Conversion, IEEE Transactions on*, 21 (3):734–741, 2006.

[98] R. J. R. Waumans. *Openbare netten voor elektriciteitsdistributie*. Kluwer Technische Boeken, 1986. ISBN 90 201 1887 0.

[99] XLPE. XLPE land cable systems - user's guide ABB Group. `http://www02.abb.com/global/gad/gad02181.nsf/0/a8a42f36692365dcc1257a62003101ce/$file/XLPE+Land+Cable+Systems+-+User%C2%B4s+Guide.pdf`, 2010.

[100] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. MATPOWER: steady-state operations, planning, and analysis tools for power systems research and education. *Power Systems, IEEE Transactions on*, 26(1):12–19, 2011.

PUBLICATIONS

The work in this thesis is based on the articles mentioned below. Chapters 4, 5, 6 and 7 contain the material of [1], [2], [3] and [4], respectively. Wander Wadman carried out the research and wrote the articles, under supervision of Daan Crommelin (promotor) and Jason Frank (copromotor). Gabriël Bloemhof helped with the development of a relevant research topic for article [1]. For article [4] Bert Zwart suggested applying large deviations theory and helped with finding useful references. All co-authors gave valuable feedback on the drafts of each corresponding article before it was submitted.

[1] Wander S. Wadman, Gabriël Bloemhof, Daan T. Crommelin, and Jason E. Frank. Probabilistic power flow simulation allowing temporary current overloading (volume to appear). In *Conference proceedings on Probabilistic Methods Applied to Power Systems*, pages 494–499, 2012.

[2] Wander S. Wadman, Daan T. Crommelin, and Jason E. Frank. Applying a splitting technique to estimate electrical grid reliability. In *Winter Simulation Conference proceedings*, pages 577–588, 2013.

[3] Wander S. Wadman, Daan T. Crommelin, and Jason E. Frank. A separated splitting technique for disconnected rare event sets. In *Winter Simulation Conference proceedings*, pages 522–532, 2014.

[4] Wander S. Wadman, Daan T. Crommelin, and Bert P. Zwart. A large deviation based splitting estimation of power flow reliability. 2015. Submitted (23 pages).