Computational Techniques for Assessing Power Grids with Wind Energy and Storage

by

Debarati Bhaumik

This work is part of the Industrial Partnership Program (IPP) "Computational sciences for energy research" of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organization for Scientific Research (NWO). This research program is co-financed by Shell Global Solutions International B.V. The research work was carried out at Centrum Wiskunde & Informatica (CWI), the Dutch national institute for mathematics and computer science.

Computational Techniques for Assessing Power Grids with Wind Energy and Storage

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. K.I.J. Maex

ten overstaan van een door het College voor Promoties ingestelde

commissie, in het openbaar te verdedigen in de Agnietenkapel

op woensdag 30 mei 2018, te 10:00 uur door

Debarati Bhaumik

geboren te Surya Sen Nagar W Bengal, India

Promotiecommissie

Promotores:

prof. dr. D.T. Crommelin	Universiteit van Amsterdam
prof. dr. A.P. Zwart	Technische Universiteit Eindhoven

Overige leden:

prof. dr. ir. J.E. Frank	Universiteit Utrecht
prof. dr. ir. B. Koren	Technische Universiteit Eindhoven
prof. dr. M.R.H. Mandjes	Universiteit van Amsterdam
prof. dr. R. Núñez Queija	Universiteit van Amsterdam
prof. dr. J.H. van Zanten	Universiteit van Amsterdam

Faculteit:

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

To my parents

Ma, Baba.. this is for you!

Acknowledgements

To begin with, I would like to thank my supervisors Bert Zwart and Daan Crommelin who have provided me with their incessant ideas and guidance throughout this PhD journey. Without Daan's patience, when I made mistakes and his motivations, when I felt directionless this thesis would not have seen the light of day. I still vividly remember my first day in Amsterdam when he came to pick me up at the airport.

I would like to acknowledge Stella Kapodistria and Barry Koren for their collaboration and support which lead to the last two chapters of this dissertation. I am grateful to Benjamin Sanderse for helping me out whenever I had questions regarding solving PDEs.

Starting this PhD gave me a new life in Amsterdam with many friends and some dear friends. From initial days of my PhD life, Wander, Keith, Nick, and Jesse thank you for tolerating all my questions and for being there for me. You were the first friends I made in this city. I miss the (in)appropriate lunch discussions we used to have. I would like to remember Zsolt, my first officemate who was always there to help me out. Bart, Anne, Ki Wai, Prashant, Laurent, Nada and Duda thanks for being very helpful colleagues.

Emma, Teresa, and Sangeetika, you girls have a very special place in my life. Thank you for tolerating all my rants and spending hours on my life problems.

One person who has constantly helped and supported me during my PhD will be Willem Jan. Be it guiding me in writing codes, solving technical issues or just being there as a friend, he was always there and

still is.

Alex, thanks for working with me from cafes at any time of the day and week during my thesis-writing phase. Your help for making the cover page of this thesis is much appreciated.

The best officemate one can have, Krzysztof Bisewski. Well you are much more than an officemate to me but had it not been this PhD I would have never met you. You have always been there teaching me probability, Markov processes, how to properly write math notations and what not. You are one of the closest people I have in this city.

Finally, I would like to remember my life-lines who kept me sane during my PhD. Tirthankar Banerjee, words are not enough to describe you. Without you on the other side of Skype, I can't even imagine how I could have managed my life and PhD in Amsterdam. My mother, without her I wouldn't even have known about this PhD offer. *Maa*, this thesis is as much yours as it is mine. My father, *Baba* knowing that you are there eased me during stressful times.

Summary

Computational techniques for assessing power grids with wind energy and storage

The electricity sector is going through a rapid transition in the energy landscape. The global share of distributed renewable energy sources like solar panels and wind power for energy generation is on the rise. Even though this transition to renewable energy sources is beneficial for reducing global CO_2 emissions, the intermittent nature of solar and wind power makes their integration into the electricity grid a highly challenging task.

For reliable operation of the electricity grid, the balance between power generation and demand should be maintained continuously. The intermittent nature of wind and solar power disrupts the power balance of the grid. Such power imbalances may lead to frequency fluctuations, voltage limit violation, line current limit violations and power losses in the grid; hence straining the grid. Therefore to ensure reliable operation, grid operators occasionally need to curtail either renewable power generation or power demand. Curtailment of renewable power generation incurs economic losses to the producers, whereas curtailment of power demand is undesirable as our society has become highly dependent on a reliable power supply.

Various measures are deployed for mitigating the adverse effects of the unpredictable renewable power injections into the grid. These measures include energy storage, demand response for end-users and optimal dispatch of distributed generators. Computational methods and tools can assist in making decisions and taking measures for mitigating the adverse effects of renewables into the grid. These tools include quantitative probabilistic reliability assessment methodologies for the grid, accurate forecasting methods for predicting potential stability threats posed by renewable energy

to the grid, and formulating stochastic models for wind and solar power for estimating grid reliability indices.

In this thesis we formulate mathematical models and develop various computational techniques to assess reliability of power grid integrated with wind power and energy storage devices. The adverse effects of unpredictable wind power can be mitigated by incorporating energy storage devices into the grid. Hence, we focus on operation and placement of batteries into the distribution grid for reliable operation.

For assessing reliability of power grids various grid reliability indices exist. In this thesis the aim is to assess reliability of power grids integrated with wind power generation and energy storage by estimating and minimizing the probability of occurrence of events which can strain the grid. Hence, we need to estimate probabilities of events whose occurrence is *rare*. The conventional Monte Carlo estimator becomes computationally inefficient for estimating such small probabilities. To reduce the computational workload of the conventional Monte Carlo estimator, we use the *splitting technique* for rare-event simulation in this thesis.

In this thesis we have developed various stochastic models which can represent the intermittency and unpredictability of wind power. Such models can be used to generate surrogate time-series of wind power injections to estimate power grid reliability indices and to study statistical properties of power grids integrated with wind power. We have developed a wind speed model using *Ornstein-Uhlenbeck processes* and have formulated models for wind farm power output using *Hidden Markov Models*.

Micro-grids powered by wind energy and supported with energy storage devices can be modeled by *Markov-modulated fluid queues*. The cumulative distribution function of the first passage (hitting) time of a Markov-modulated fluid queue follows a first order advection-reaction hyperbolic partial differential equation. We propose to solve this hyperbolic partial differential equation using different numerical integration schemes in this thesis.

Samenvatting

Computationele technieken voor het beoordelen van elektriciteitsnetten met windenergie en opslag

De energiesector maakt momenteel een snelle transitie door. Decentrale hernieuwbare energiebronnen zoals wind en zon vormen een steeds groter deel van de globale energievoorziening. Hoewel deze transitie naar duurzame bronnen bijdraagt aan het verlagen van de wereldwijde CO_2-uitstoot, leidt de onvoorspelbaarheid van hun beschikbaarheid tot grote uitdagingen bij de integratie in het elektriciteitsnet.

Voor de betrouwbaarheid van het elektriciteitsnet moeten productie en verbruik voortdurend in balans gehouden worden, maar de veranderlijke natuur van wind- en zonne-energie verstoort deze balans. Dergelijke verstoringen kunnen leiden tot spanningsfrequentieschommelingen, overschrijding van spannings- of stroomlimieten, of stroomuitval. Om betrouwbaarheid te garanderen, moeten beheerders daarom soms de opwekking of het verbruik van energie beperken. Echter, beperking van energieopwekking leidt tot economisch verlies bij de producenten, en beperking van verbruik is onwenselijk in onze samenleving die sterk afhankelijk is geworden van betrouwbaar beschikbare elektriciteit.

Er worden verschillende middelen ingezet voor het beperken van de nadelige kanten van onvoorspelbare vermogensinjecties, zoals energie-opslag, vraagregulatie bij eindgebruikers, en optimale inzet van decentrale generatoren. Computationele technieken kunnen besluitvorming ondersteunen rond het beperken van de nadelige effecten van duurzame energie in het energienet. Deze technieken omvatten kwantitatieve stochastische betrouwbaarheidsbeoordeling van het energienet, nauwkeurige voorspellingen van potentiële instabiliteit veroorzaakt door hernieuwbare ener-

gieproductie, en formulering van stochastische modellen voor wind- en zonne-energie om betrouwbaarheidsindices te schatten.

In dit proefschrift formuleren we wiskundige modellen en ontwikkelen we computationele technieken om de betrouwbaarheid van elektriciteitsnetten met geïntegreerde windenergie en energie-opslag te beoordelen. De negatieve effecten van de onvoorspelbaarheid van windenergie kunnen beperkt worden door energie-opslag in het net op te nemen. Daarom richten we ons op gebruik en plaatsing van accu's in het distributienet voor een betrouwbare energievoorziening.

Er bestaan verscheidene maten voor de betrouwbaarheid van een energienet. In dit proefschrift is het doel om de betrouwbaarheid van energienetten met geïntegreerde windenergie en energie-opslag te beoordelen door de kans op gebeurtenissen die het net overbelasten te schatten, en die vervolgens te minimaliseren. We moeten hiervoor kansen schatten van zeldzame gebeurtenissen ('rare events'). Conventionele Monte Carlo-simulaties zijn computationeel inefficiënt bij het schatten van dergelijk kleine kansen. Om de rekentijd van de Monte Carlo-simulatie te beperken, gebruiken we de zogeheten 'splitting'-techniek voor 'rare event simulation' in dit proefschrift.

In dit proefschrift hebben we stochastische modellen ontwikkeld die het variabele en onvoorspelbare karakter van wind-energie kunnen beschrijven. Zulke modellen kunnen gebruikt worden om surrogaat-tijdreeksen van windenergieproductie te genereren, om aan de hand daarvan betrouwbaarheidsmaten te schatten en statistieken te bestuderen van energienetten met geïntegreerde windenergie. We hebben een windsnelheidsmodel ontwikkeld gebruikmakend van Ornstein-Uhlenbeck-processen, en windenergieproductiemodellen geformuleerd aan de hand van Hidden Markov-modellen.

Micro-energienetten met windturbines ondersteund door accu's kunnen gemodelleerd worden aan de hand van Markov-gemoduleerde vloeistofmodellen. De cumulatieve verdelingsfunctie van de 'first passage' van

een Markov-gemoduleerd vloeistofmodel volgt een eerste orde advectie-reactie hyperbolische partiële differentiaalvergelijking. In dit proefschrift stellen we een oplossing van deze hyperbolische partiële differentiaalvergelijking voor gebruikmakend van verschillende numerieke integratiemethoden.

Contents

Chapter 1

Introduction

The necessity for reducing the carbon footprint to fight global climate change got strengthened by the 2015 Paris Agreement of the United Nations Framework Convention on Climate Change (UNFCCC). 195 UNFCCC countries have signed the agreement and 165 among them have ratified it to keep the global average temperature rise well below 2 degree Celsius and to pursue efforts to limit the temperature increase to 1.5 degree Celsius above pre-industrial levels [8, 9]. Many governments have implemented new policies and targets for renewable energy. In particular, the electricity (power) sector is experiencing a predominant transition in the energy landscape.

1.1 The advent of renewables into the power grid

Biomass, hydropower, geothermal energy, wind and solar energy are examples of renewable energy sources. Among these, hydropower and geothermal energy have been well-established as sources of energy. Figure 1.1 shows the global share of renewable energy production at the end of 2016. A quarter of the total global electricity consumption was met by renewable energy sources [7]. It can be seen from Figure 1.1 that hydropower is by far the leading contributor of renewable power.

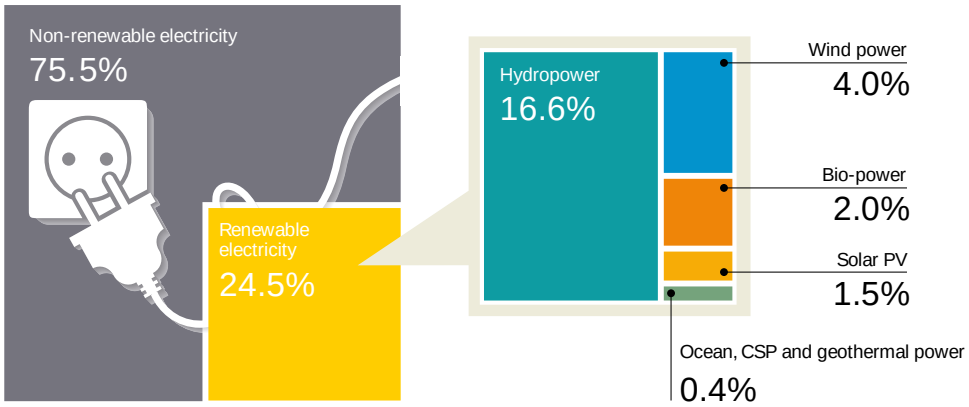


Figure 1.1: Global estimated share of renewable energy production at end of 2016. Source [7].

Current government policies such as subsidy on PV panels and net metering, steep fall in prices of solar PV and wind power technologies have accelerated the integration of solar PV panels and wind power into the power grid. For example, 161 GW of renewable power was added to the total global renewable power generating capacity in 2016. Solar PV accounted for almost half of the share of 161 GW, whereas wind power constituted one third of the share. By the end of the same year, the total global renewable power generating capacity went up to 2017 GW [7].

1.2 Power grid reliability and renewables

In an electricity grid the power demand and production should be balanced at all times. The frequency of the oscillations of the alternating current in the electricity grid serves as the indicator of this power balance between production and demand and its nominal value is 50 Hz in Europe, see Policy 1 of [2]. The value of the frequency falls below 50 Hz when power production is greater than power demand and vice-versa. Under normal conditions this frequency can fluctuate between 49.9 and 50.1 Hz. However, the maximum possible dynamic frequency deviation from the nominal value is 0.8 Hz (see Policy 1 of [2]). Hence, it is the work of the grid operators to maintain this frequency in the acceptable range by balancing the power consumption and demand in the grid. Along with the frequency regulations, for maintaining grid stability the nodal voltages of the grid should also be kept within acceptable limits. Voltages mostly are controlled by balancing the reactive power of the grid [61]. Therefore, the active and reactive power generation and demand need to be balanced for reliable and secure grid operations.

Traditionally, power is balanced in the grid using dispatchable power sources. Generators such as coal or nuclear power plants provide a steady and reliable power supply to the base-load power demand of the grid. For peak-hours power demand (morning and evening peaks) fast-ramping power sources like natural gas and hydro power are used. Thus, the grid operators maintain grid stability and security by performing various ancillary services like scheduling (day ahead) and dispatch (real time) of power generators, voltage and frequency control [132].

Grid reliability is jeopardized in various ways. Examples are *natural calamities* like storms, lightning, earthquake and so on; *technical reasons* like component failure, short circuit, component aging, etc and *human errors* [36, 61]. Such events can cause cascading failures in the grid leading to blackouts (see Chapter 13 of [61]).

The uncertain nature of power consumption can also disrupt the power

balance of the grid. Typically the aggregated power consumption profile at one node of the grid can be predicted to a certain extent. However, developments like the increase in electric vehicles (EVs) makes power consumption stochastic in nature by creating large power demand at uncertain locations, durations and times. Such large power demands may lead to frequency fluctuations, nodal voltage limit violations and power losses in the grid [51, 57, 78].

Like uncertainty in power consumption, uncertainty in power generation due to variable renewable energy sources like wind and solar challenges the secure and reliable operation of the grid [111, 145]. In order to balance the power demand and generation in the grid, the operators curtail renewable energy sources when there is over-production of renewable power [75]. Such curtailments lead to undesirable economic losses to the renewable energy producers and the grid operators. For example, according to a news article in January 2015, a sudden influx of wind power cost the grid operator 13 million euros in order to keep the grid stable in Germany [17]. Hence, it is important to mitigate over-production or curtailment of renewable power both for reliable operation of the grid and for minimizing economic losses of power producers.

1.3 Mitigating grid reliability risks

The adverse consequences of variable power injection by renewables into the grid can be mitigated by various measures. We discuss some of these measures in this section. *Demand response* is one measure where the end-user is encouraged to change their power consumption pattern in response to electricity pricing initiated by the grid operator in order to achieve monetary profit for both parties or when grid reliability is jeopardized [148, 162]. Another measure is *optimal dispatch* of distributed generators in the distribution grid [52, 112]. For the planning of dispatch of conventional generators, incorporating *chance constraints* into the conventional

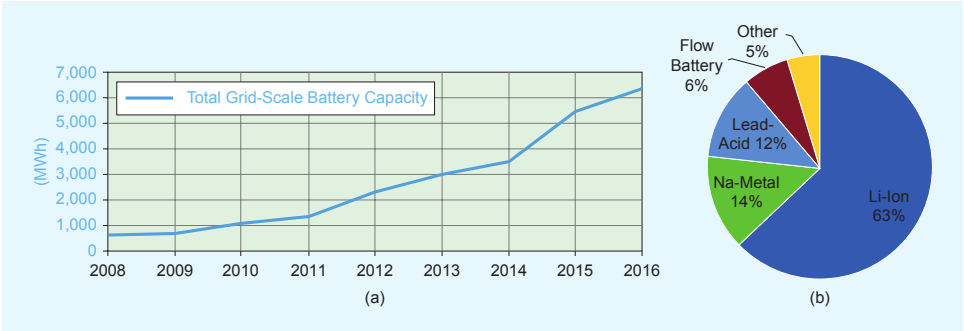


Figure 1.2: Global battery storage integration in the grid by (a) capacity and (b) chemistry. Source [153].

optimal power flow method mitigates risks due to increased penetration of renewables in the grid [31]. Given the stochastic nature of renewable power it is of great importance to develop quantitative *probabilistic reliability assessment* methodologies for the grid. Such methods help in planning, expansion, and operation of the grid integrated with renewable generation [36, 109, 163].

Energy storage plays an important role in mitigating variability in the grid by shaving peaks of intermittent renewable power injections [56, 154, 180]. Battery energy storage will likely play an important role for accommodating further integration of renewables in the distribution grid [153]. Figure 1.2 shows the increase in the global battery storage capacity in the grid over 2008 - 2016. Integration, placement, and operation of battery energy storage to enhance the reliability of the distribution grid is one of the main topics of this thesis.

Along with the above mentioned measures, accurate *probabilistic forecasting* methods can help in predicting potential threats to stability posed by the intermittent renewable power injection into the grid [179]. However, for planning and expansion of the grid integrated with renewable energy sources, formulation of *stochastic models* for wind and solar power is of great relevance. Such models can be used for e.g. generating surrogate time series of power injections to estimate power grid reliability indices.

To this end, several stochastic models are presented in this thesis.

1.4 Aim and plan of the thesis

The aim of this thesis is to formulate mathematical models and to develop computational techniques which facilitate reliability analysis of power grids integrated with wind power and energy storage devices. This thesis focuses on the reliable operation of the power grid which can in general be disrupted by the intermittent nature of wind power. Adverse effects of wind power variability can be mitigated by incorporating energy storage devices into the grid. Hence we investigate how to operate and place energy storage devices into the power grid for reliable operation of the grid. We also develop simple stochastic models for wind power which can be used to study statistical properties of power grids and to estimate grid reliability indices.

Chapter 2 briefly summarizes various methodologies and concepts which are used in this thesis to develop different computational techniques and models for assessing reliability of power systems integrated with wind energy and storage. In this chapter we present the power flow equations and discuss grid constraints, grid reliability indices, Monte-Carlo simulation, as well as the splitting technique for rare-event simulation. The basic concept of hidden Markov models which are used for modeling wind park power output is presented in this chapter. Finally, we briefly discuss Markov-modulated fluid queues and hyperbolic partial differential equations which are used to estimate the probability distribution of the first passage (or hitting) time of a storage device getting empty.

In Chapter 3 we perform an overall assessment of a single domestic household powered by a wind turbine and supported by an energy storage device. The aim is to investigate the best operation mode of the storage device such that the occurrence of large power spills (excess or overproduction) can be minimized. A sudden influx of large power excess into the

grid jeopardizes the stability of the grid by creating voltage imbalances and current overloads, hence it is important to minimize their occurrence. For estimating the small probability of large power spills, we use the *splitting technique* for rare event simulation for reducing the workload of the conventional Crude Monte Carlo probability estimator. Simulation results show that the *ramp constraints* imposed on the charging/discharging rate of the storage device play a pivotal role in mitigating large power spills. It is observed that by employing a new charging strategy for the storage device large power spills can be minimized further. We show that there exists a trade-off between reducing the large power spills versus reducing the average power spills. This chapter is based on [29].

In Chapter 4 we develop a computational technique to optimize the placement of storage devices in the network with stochastic generation and demand such that the power network can be made more reliable. We use the probability of a line current violation as the reliability index of the network and find the optimal storage position so that this probability is minimal. We use the *simulated annealing* algorithm to minimize this probability under the variation of storage locations and capacities in the network, keeping the total storage capacity constant. For estimating the small probabilities of line current violations we use splitting. As an illustration, we apply our method to the IEEE-14 bus network. This chapter is based on [28].

In Chapter 5, which is based on [27], we investigate to what extent the power output of wind farms can be modeled with discrete stochastic models (Markov chains). Such models have a rather simple structure, thereby facilitating theoretical analysis and simulations of power grids integrated with wind farms. We investigate the merits of modeling power output with straightforward Markov chains as well as *Hidden Markov Models* (HMMs). The parameters of these models are inferred from measurement data from multiple turbines in a wind farm. We use these models both for individual turbine output and for total aggregate power output of multiple

turbines. When modeling individual turbine output, the *hidden process* in the HMM is instrumental for incorporating the correlations between output of different turbines. Accounting for these correlations is important for correctly capturing the upper quantiles (90%, 95%, 99%) of the distribution of the aggregate power output of the entire wind farm. We show that despite their simple structure, discrete stochastic models are able to reproduce important features of the power output of wind farms. This opens up possibilities to use and analyze these models with methods and techniques from the field of queueing theory which is the topic of Chapter 6.

In Chapter 6 we study different numerical methods to estimate the cumulative distribution of the first passage (or hitting) time of a buffer being empty. The rate at which the buffer is filled/drained is modulated by a continuous time Markov chain (CTMC). Such Markov-modulated fluid queue models can be used to model micro-grids integrated with wind park and energy storage devices. The distribution function of the first passage time follows a hyperbolic partial differential equation (PDE). For systems with large state spaces (of the CTMC), we explore different numerical integration schemes to solve the PDE efficiently and compare this method with the existing Laplace Stieltjes transform inversion method. This chapter is based on [30].

Chapter 2

Overview of methodologies

In this chapter we give an overview of existing methodologies that have been used in this thesis to develop computational techniques and models for assessing reliability of power systems integrated with wind energy and storage.

2.1 Power flow equations

For planning of future expansions of power systems and to determine in which operational state (normal, emergency or restorative) the grid is operating, grid operators need to compute the line power flows (or current flows) and the nodal voltages for a specific generation and load profile of power injections. These computations are done by solving the so-called *power flow equations*. These equations are steady-state solutions of a dynamic system [77]. The power flow equations can be modeled using the

non-linear Alternating Current (AC) model and the linear Direct Current (DC) model. Extensive analysis of the power flow equations can be found in the literature [53, 77]. We shall briefly discuss both the AC and DC models in this section. We provide a detailed derivation of the power flow equations in Chapter 4.

AC power flow equations: The topology of a power grid can be defined by an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N} := \{1, 2, \ldots, N\}$ is the set of nodes (also known as buses) and $\mathcal{E}$ is the set of edges (also called lines). The line from node i to j is denoted by (i, j). The presence of energy storage components such as inductors and capacitors in the power grid gives rise to sinusoidal (or complex) line currents and nodal voltages. For computing these line currents and nodal voltages the AC power flow equations needs to be solved. The AC power flow equations in polar form $\forall i \in \mathcal{N}$ are given by [77],

$$P_i = \sum_{k=1}^{N} |V_i||V_k|(G_{i,k}\cos(\theta_i - \theta_k) + B_{i,k}\sin(\theta_i - \theta_k)) \tag{2.1}$$

and

$$Q_i = \sum_{k=1}^{N} |V_i||V_k|(G_{i,k}\sin(\theta_i - \theta_k) - B_{i,k}\cos(\theta_i - \theta_k)). \tag{2.2}$$

In the above equations $P_i \in \mathbb{R}$ is the active power injected (or extracted) at the i-th bus, $Q_i \in \mathbb{R}$ is the reactive power injected (or extracted) at the i-th bus, $|V_i| \in [0, \infty)$ is the voltage magnitude and $\theta_i \in (-\pi, \pi]$ is the voltage angle of bus i. $G_{i,k}$ and $B_{i,k}$ are the conductance and susceptance of line $(i, k) \in \mathcal{E}$, respectively. We discuss the AC power flow equations in more detail in Chapter 4.

DC power flow equations: The DC power flow equations can be seen as a linear approximation to the AC power flow equations [77] and are

given by

$$P_i = \sum_k B_{i,k}(\theta_i - \theta_k) \ \forall i \in \mathcal{N}. \tag{2.3}$$

In this thesis we use the DC power flow equations to calculate line current flows in Chapter 4. We give a more detailed derivation of the DC power flow equations and line current flows in that chapter.

2.2 Operational constraints of the grid

The power grid can operate in three states, namely the normal, emergency and restorative state [11]. The grid operates in the normal state when the generators in the grid supply sufficient power to satisfy all demand without violating any *operational constraints*. These constraints are [36]:

1. *Transmission line constraints*
   For the normal state operation of the grid the modulus of active power flowing through all the transmission lines should not exceed its given (line-specific) maximum at any time t, i.e.,

   $$|P_{i,j}(t)| \leq P_{i,j}^{\max} \quad \forall (i.j) \in \mathcal{E}. \tag{2.4}$$

   Similarly, the line current flowing through the lines should not exceed its given maximum at any time t,

   $$|I_{i,j}(t)| \leq I_{i,j}^{\max} \quad \forall (i.j) \in \mathcal{E}. \tag{2.5}$$

   Violation of the active power or line-current constraints leads to generation of excess heat in the transmission lines, depending on the material and thickness of the lines. Hence these violations can lead to physical damage of the transmission line due to overheating. The grid operators need to keep the temperature of each of the transmission lines $T_{i,j}(t)$ bounded all the time:

$$T_{i,j}(t) \leq T_{i,j}^{\max}. \tag{2.6}$$

However, the constraints posed in (2.4) or (2.5) are stricter than (2.6) as there exists a time lag between the change in temperature in the lines and change in line current [121, 164]. We will focus on line-current constraints (2.5) in Chapter 4.

The frequency regulations require the voltage angle difference between bus i and bus j to stay within their limits [46],

$$\theta_{i,j}^{\min} \leq \theta_i - \theta_j \leq \theta_{i,j}^{\max}, \ \forall (i,j) \in \mathcal{E}. \tag{2.7}$$

2. *Nodal constraints*
   Along with the transmission line constraints the voltage (magnitude), and the reactive power at each of the i nodes (or buses) must be maintained within their prescribed limits for all time t [102],

$$V_i^{\min} \leq |V_i(t)| \leq V_i^{\max}, \tag{2.8}$$

and

$$Q_i^{\min} \leq Q_i(t) \leq Q_i^{\max}, \tag{2.9}$$

where $|V_i(t)|$ is the voltage magnitude and $Q_i(t)$ is the reactive power at the i-th node of the grid. A violation of the nodal voltage constraint can lead to damaging of electrical devices connected to the node.

Hence, grid operators must ensure these constraints are satisfied at all times for normal operation of the grid. In this thesis we focus on assessing the probability of maintaining the normal operational state of the grid.

2.3 Grid reliability indices

Reliability of the power grid is defined as the ability of the grid to supply power to all the loads without violating any operational constraints discussed in Section 2.2 [6, 32]. Reliability of the power grid comprises of two aspects - *adequacy* and *security*. Adequacy is *"the ability of the electric system to supply the aggregate electrical demand and energy requirements of the customers at all times"* and security refers to *"the ability of the electric system to withstand sudden disturbances such as electric short circuits or unanticipated loss of system elements"*. These definitions are taken from [6].

The reliability of the power grid is estimated by *grid reliability indices* which can be categorized as relating to the frequency, duration, probability and extent of grid constraint violations [125, 163]. The traditional grid reliability indices are based on consumer satisfaction [3, 125]:

- System Average Interruption Frequency Index (SAIFI): average number of interruptions experienced per customer over a prescribed interval of time.
- System Average Interruption Duration Index (SAIDI): average duration of interruptions for the customers who experienced interruptions over a prescribed interval of time.
- Customer Average Interruption Frequency Index(CAIFI): average frequency of interruptions for those customers who experienced interruptions.
- Customer Average Interruption Duration Index (CAIDI): average time required to restore the service.
- Loss of Load Probability (LOLP): probability that the peak demand will exceed the available power generation in a given interval of time.

- Loss of Load Expectation (LOLE): expected number of days in a year when the daily peak demand exceeds the available power generation. It is closely related to LOLP.

Various other reliability indices are defined in [3]. Note that the indices mentioned above are based on curtailment of load (or power demand). However, given the new energy landscape of distributed renewable generation, part of the consumers have become *prosumers*. Hence, it is also important to study the curtailment of power generation as producers incur economic loss by generation curtailment.

This thesis is concerned with assessing reliability of power grids integrated with wind power generation and energy storage. Hence we aim to estimate and minimize the probability of occurrence of events (depending on the problem we are solving in Chapters 3 and 4) which can lead to grid operational constraint violations (see Section 2.2). Let A be the event leading to grid constraint violation in time interval $[0, T]$. In this thesis we wish to efficiently estimate the probability of occurrence of A, i.e, $\mathbb{P}(A)$.

2.4 Rare event simulation

One of the main topics of this thesis is to estimate probabilities of events which lead to grid constraint violations with an aim to minimize such probabilities. Hence, we need to calculate the probability of events whose occurrence is unlikely. Such events are called *rare events*. In this section we briefly discuss the computational inefficiency of crude Monte Carlo estimation for rare events and the splitting technique for rare event simulations.

2.4.1 Monte Carlo Simulation

Monte Carlo (MC) simulation is a very robust method for estimating probabilities when these probabilities are not available analytically [136]. For

Crude Monte Carlo (CMC) simulation, the stochastic system is sampled M times and for each sample it is verified whether the event of interest A has occurred in that sample or not.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A \in \mathcal{F}$ be the rare event of interest. The probability of occurrence of A is defined as $\gamma := \mathbb{P}(A)$. If A is a rare event then the value of γ is small. The CMC estimator of $\mathbb{P}(A)$ is given by [135]

$$\tilde{\gamma} := \frac{1}{M} \sum_{j=1}^{M} \mathbb{1}_{\{A \text{ occurs in sample } j\}}, \tag{2.10}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function which takes value 1 if the expression in the parenthesis is realized else it takes value zero. The variance of the CMC estimator using M samples is given by

$$\mathrm{Var}(\tilde{\gamma}) = \frac{\gamma(1-\gamma)}{M}, \tag{2.11}$$

and hence the *squared relative error* of the CMC estimator (2.10) is given by

$$\mathrm{SRE}(\tilde{\gamma}) := \frac{\mathrm{Var}(\tilde{\gamma})}{\gamma^2} = \frac{\gamma(1-\gamma)/M}{\gamma^2} = \frac{1-\gamma}{\gamma M}. \tag{2.12}$$

Hence for fixed M, the SRE diverges as $\gamma \to 0$. Therefore the CMC estimate becomes unreliable when γ is small unless one uses very large values of M to obtain an acceptable SRE. For example, to estimate probabilities smaller than 10^{-4} one needs $M \gtrsim 10^6$ CMC samples for achieving SRE ≈ 0.01.

2.4.2 Splitting technique

To reduce the computational workload for estimating γ in case $\gamma << 1$, rare-event simulation techniques are used. The main two types of rare-event simulation techniques are *importance sampling* and *splitting* [135]. We use the splitting technique in Chapters 3 and 4 and we shall discuss

it briefly in this section. For details on importance sampling please see [135].

In splitting, the sample paths of the stochastic processes involved are split into multiple copies at various levels of the *importance function* till the rare event set is reached [69]. The importance function measures the distance to the rare-event set. The probability γ is decomposed into the product of several conditional probabilities which are less computationally intensive to estimate than γ itself. To make this feasible, the splitting technique is primarily used for Markov processes.

The most important challenge for splitting is to find an appropriate importance function, ϕ, which assigns importance values to the state of a n-dimensional vector-valued Markov process $\boldsymbol{X}(t)$ with state space $\mathcal{X} \subseteq \mathbb{R}^n$, i.e., $\phi : \mathcal{X} \longrightarrow \mathbb{R}$. We choose the importance function such that $\phi(\boldsymbol{X}(0)) = 0$. Let $S_{\phi,L}$ be the rare event set defined in terms of ϕ, as $S_{\phi,L} := \{\boldsymbol{x} \in \mathcal{X} : \phi(\boldsymbol{x}) \geq L\}$. We are interested in the rare event

$$A = \{\exists t \leq T : \boldsymbol{X}(t) \in S_{\phi,L}\}. \tag{2.13}$$

For the splitting technique we split the interval $[0, L]$ into m sub-intervals with boundaries $0 = l_0 < l_1 < \cdots < l_m = L$. We define $T_k = \inf\{t > 0 : \phi(\boldsymbol{X}(t)) \geq l_k\}$ as the time of hitting the k-th level and $H_k = \{T_k < T\}$ as the event that the k-th level is hit during $[0, T]$. Therefore, $\gamma = \mathbb{P}(H_m)$ and $\mathbb{P}(H_0) = 1$. As $H_m \subset H_{m-1} \subset \cdots \subset H_1 \subset H_0$, we have

$$\gamma = \mathbb{P}(A) = \prod_{k=1}^{m} \mathbb{P}(H_k|H_{k-1}) = \prod_{k=1}^{m} p_k,$$

where $p_k := \mathbb{P}(H_k|H_{k-1}) = \mathbb{P}(H_k)/\mathbb{P}(H_{k-1})$. Each p_k is estimated separately by generating independent sample paths from the distribution of the entrance state $G_{k-1} := (T_{k-1}, \boldsymbol{X}(T_{k-1}))$ conditioned on H_{k-1} at the threshold level l_{k-1}. The empirical distribution $\hat{G}_k$ is an estimate of the entrance distribution G_k. Thus we can proceed recursively, replacing $\hat{G}_{k-1}$ for G_{k-1} and estimate p_k at each level k, by the proportion of level hits

$$\hat{p}_k = R_k/N_{k-1} \text{ for all } R_k > 0, \tag{2.14}$$

in which R_k is the number of sample paths where H_k occurs and N_k is the total number of sample paths started at level k. How N_k is chosen depends on which variant of splitting is deployed. Hence we get the estimate of γ as the product

$$\hat{\gamma} = \prod_{k=1}^{m} \hat{p}_k. \tag{2.15}$$

Computing each of the non-rare conditional probabilities $\hat{p}_k$ for $k = 1, \ldots, m$ is computationally less intensive than estimating $\hat{\gamma}$ directly as in CMC. We will use the splitting technique in Chapters 3 and 4. We discuss different variants of splitting in Chapter 3.

2.5 Wind power modeling

For maintaining a reliable power supply, as well as for future power system planning and daily system operation it is of great importance to study the statistical properties of power systems integrated with intermittent wind power generation. In order to do so, it is important to develop stochastic methods for modeling power output, as stochastic models enable one to represent the intermittency and unpredictability of wind power. These models can be used for e.g. generating time series of power injections to estimate power grid reliability indices [93].

Stochastic models to simulate wind power generation can be developed in two ways: (i) model the wind speed and employ a separate model (e.g. a wind power curve (WPC)) to obtain power output from wind speed and (ii) model the power output directly. We present a literature review of these approaches in Chapter 5.

In Chapter 3 we first model wind velocities in the zonal (east-west) direction and the meridional (north-south) direction, as two 1-d Ornstein-Uhlenbeck (OU) processes. The wind speed is obtained by taking the

Euclidean-norm of these two OU processes. The wind speed then is translated to wind power by a WPC.

An OU process $\{X_t, t \geq 0\}$ is a mean reverting, stationary, Gaussian, Markovian stochastic process which satisfies the following stochastic differential equation [91],

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t. \tag{2.16}$$

In the above equation $\mu > 0$ is the long term mean of the process, $\theta > 0$ is the mean reverting rate, $\sigma > 0$ is the volatility term and $\{W_t, t \geq 0\}$ is a Brownian motion.

In Chapter 5 we directly model the wind power output instead of first modeling wind speed and then wind power using a WPC. In that chapter, we model wind farm power output with simple Markov chain models. In order to do so we use a *Hidden Markov Model* (HMM) where the wind turbine power output is modeled as a stochastic process (Markov chain) that depends on the state of a so-called *hidden process*. The hidden process is a discrete-time Markov chain and is not observed; hence the name hidden process. Interpreting this hidden process depends on the application. We speculate that in the case of wind power generation, the hidden process may include the meteorological conditions to which the wind turbines are subjected.

A standard HMM consists of a collection of random variables $\{X_1, \ldots, X_T, Y_1, \ldots, Y_T\}$. The stochastic process $\{Y_t, t = 1, \ldots, T\}$ is the observed (continuous or discrete) process and $\{X_t, t = 1, \ldots, T\}$ is the hidden process which is a discrete-time Markov chain with discrete state space $\mathcal{M}$. Under the standard HMM formalism for a given set of observations $(y_1, \ldots, y_T)$, the stochastic processes $\{X_t, t = 1, \ldots, T\}$ and $\{Y_t, t = 1, \ldots, T\}$ are governed by the following stochastic matrix $\boldsymbol{A}$ and set of emission distributions $\boldsymbol{B}$ respectively [37, 130],

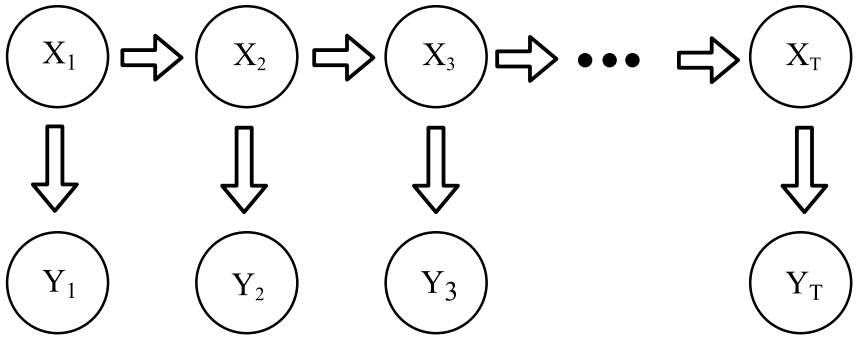


Figure 2.1: Schematic diagram of the standard HMM described in (2.17)

$$\begin{cases} A_{i,j} = \mathbb{P}(X_{t+1} = j | X_t = i), \\ B_i(y_t) = \mathbb{P}(Y_t = y_t | X_t = i), \end{cases} \tag{2.17}$$

$\forall i, j \in \mathcal{M}$. For $t = 1$, the initial distribution of X_1 is given by $\pi_i = \mathbb{P}(X_1 = i)$. A schematic diagram showing the dependence of the random variables $\{X_t, Y_t, t = 1, \ldots, T\}$ of the HMM described by (2.17) is presented in Figure 2.1.

In Chapter 5 we discuss in more detail the hidden Markov model we use to model power output of a wind farm and parameter estimation of the model. Figure 2.2 shows the schematic diagram of the HMM model we present in Chapter 5.

2.6 Markov-modulated fluid queue

A Markov-modulated fluid model (or queue) is a mathematical model that describes the fluid level in a buffer (or storage device) which is being filled and emptied for exponentially distributed random periods. These periods are generally related to the state of a continuous-time Markov chain (CTMC) [16, 115]. Such models find application in various fields

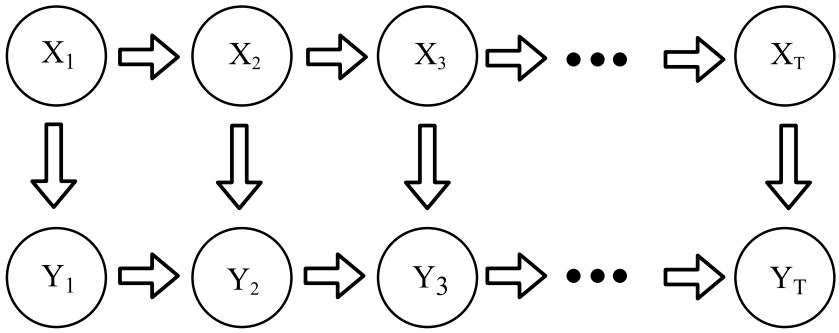


Figure 2.2: Schematic diagram of the HMM model for modeling wind power as used in Chapter 5.

of research, for example in dam theory [117], telecommunication networks [16, 60, 142], transportation systems [122], modeling forest fires [152], ruin probability [19, 23], video streaming [41, 170], and other fields of queuing theory [141].

In Chapter 6 we apply the Markov-modulated fluid queue model to micro-grids powered by a wind park and supported with energy storage devices. A micro-grid is a localized group of interconnected distributed energy sources and loads which is connected to the centralized grid. It can operate in two modes, namely the *grid-connected* mode and the autonomous *island-mode* [157]. For autonomous operation of such micro-grids in island-mode it is important to study the probability of the energy storage device (buffer) running empty.

2.6.1 Hyperbolic partial differential equations

The cumulative distribution function of the first passage (or hitting) time of a buffer (modulated by a CTMC) being empty follows a first order *hyperbolic partial differential equation*, more specifically, an *advection-reaction* partial differential equation (PDE) [119]. We discuss the model in detail in Chapter 6. In this section we will briefly describe hyperbolic PDEs.

Hyperbolic PDEs find application in various physical systems like wave motion, and advective transport of chemical species [85, 106]. A simple time-dependent linear first order hyperbolic PDE with one space dimension is given by

$$\frac{\partial \boldsymbol{U}(x,t)}{\partial t} + \boldsymbol{A}\frac{\partial \boldsymbol{U}(x,t)}{\partial x} = 0, \tag{2.18}$$

where $\boldsymbol{U} \in \mathbb{R}^n$ is an n-dimensional vector. Each component of the vector generally represents some unknown physical quantity like temperature, pressure, velocity etc. $\boldsymbol{A}$ is a $n \times n$ real matrix. The above equation is *hyperbolic* if the matrix $\boldsymbol{A}$ has real eigenvalues and n linearly independent eigenvectors.

2.6.1.1 Advection-reaction PDE

We will see in Chapter 6 that the cumulative distribution function of the first passage time of the Markov modulated buffer being empty follows an advection-reaction PDE. Such PDEs are commonly used to model physical systems with different chemically reacting species being advected by a fluid flow. They have the following form (see Chapter 17 of [106])

$$\frac{\partial \boldsymbol{U}(x,t)}{\partial t} + \boldsymbol{A}\frac{\partial \boldsymbol{U}(x,t)}{\partial x} = \psi(\boldsymbol{U}(x,t)). \tag{2.19}$$

In the above equation $\boldsymbol{U}(x,t)$ may represent concentrations of various chemically reacting species which needs to solved for, $\boldsymbol{A}$ is the coefficient matrix (determined by the advective velocity) and $\psi(\boldsymbol{U}(x,t))$ is called the source or reaction term. In the problem we consider in Chapter 6, $\boldsymbol{U}(x,t)$ denotes a vector of cumulative distribution functions of the first passage time of a Markov modulated buffer being empty. This is discussed in more details in Chapter 6.

In Chapter 6 we deploy various numerical integration schemes for solving the advection-reaction PDE. These schemes are well established in the scientific computing discipline, however their use in the field of queueing systems is new.

Chapter 3

Mitigation of large power spills by an energy storage device in a stand-alone energy system

3.1 Introduction

Integration of intermittent renewable sources of energy like solar and wind power into the electric grid has increased in recent times. The depletion of the exhaustible resources of energy and the strive for a carbon-free future will enhance the usage of these renewable sources more. The unpredictable nature of the renewable energy sources lead to intermittent power generation. This makes the integration of renewable energy sources into the power grid a highly challenging task.

For stable or reliable operation of power systems the voltage and frequency of the grid should be maintained within acceptable limits. The

stability should be maintained within the timescale of seconds [4]. Large scale integration of distributed variable energy sources like PV arrays and micro-generator wind turbines can jeopardize the demand and supply balance of the grid due to sudden injection or deficit of power [4, 105]. An instantaneous or sudden influx of power challenges the reliability of the grid, and grid operators frequently resort to curtailment of renewable energy sources in order to maintain grid stability [13, 55, 64, 75]. Such curtailments lead to economic losses for the involved energy producers. Therefore, it is important to minimize such sudden influx of renewable power into the grid.

To improve the practical efficiency of intermittent renewable energy and to minimize the need of drastic actions (like using expensive fast ramping generators) for ensuring reliable operations of the power grid, local storage of excess power can be an important tool. The energy storage device acts as buffer energy source. It stores energy when there is over-generation of power, and delivers the stored energy to the system when there is under-generation of power.

Stand-alone systems with renewable generations like solar photovoltaic (PV) and wind supported with battery storage have been investigated in great detail with respect to the PV-wind generation sizing, performance, battery storage sizing, efficiency, optimization, system cost and reliability indices in [21, 24, 59, 107, 113, 128, 146, 167, 174, 175, 181]. [172] investigated the feasibility of replacing diesel generation entirely with solar PV and wind turbines supplemented with energy storage by characterizing the load-shedding probabilities. [143] recommended a model to optimize the sizes of battery capacity and PV generator for stand-alone PV system using two optimization criteria, the loss of power supply probability (LPSP) reliability criterion and energetic cost for economic evaluation. [45] analyzed loss of power supply (LPS) and LPSP for sizing the PV generators of stand-alone PV systems. [34] presented a simulation model for the reliability evaluation (loss of load expectation and loss of energy expectation)

of small stand-alone wind energy conversion systems with respect to battery size, charging (discharging characteristics), wind speed, wind turbine characteristics etc. [48] devised a mathematical methodology to predict loss of load probability (LLP) for sizing stand-alone photovoltaic systems.

In this chapter we focus on large power spills caused by wind power, i.e. on instances where the power generation is substantially larger than the locally consumed power so that there is a large excess. This excess of power either is lost or must be fed to the grid. The former situation leads to economic losses to the wind energy producers and the latter situation may cause problems in the grid such as voltage imbalances or current overloads. In this study we investigate a stand-alone single domestic energy system with a local micro-generator wind turbine supplemented with a battery. We aim to answer: How can we efficiently quantify the *Probability of Large Power Spills* (PLPS) for a stand-alone energy system model? We present a computational methodology to do so. Using this methodology, we compare different battery switching strategies and their impact on PLPS. We analyze a single domestic household because many such households make up a part of a distribution grid and an instantaneous influx of large power from geographically correlated households powered by wind can lead to severe damage to the grid. Also, this analysis of a single household serves as a prototype for bigger systems such an energy island or a region of the grid with high penetration of renewable energy sources.

We use a simple energy balance method for the switching (charging/discharging) of the battery, i.e., when there is excess power generation the battery is charged and when there is deficit of power it is discharged. Such a simple switching strategy of the storage device has been considered in many previous studies on energy systems with renewable generations supplemented with storage devices, e.g. [34, 39, 45, 48, 65, 128, 143, 167, 172].

To this end, we devise models for simulating the wind speeds and

power demand such that the invariant probability densities of the data generated by the models are comparable to the data from measurements. With these models for power generation and demand, we analyze how the *ramp constraints*, the imposed maximal charging/discharging rates on the storage device affects the probability of large power spills. We define a strategy for charging the storage device to reduce the probability of large power spills further. Finally, we study the effect of the ramp constrains and the new charging strategy on the average power spilled in a given time interval of interest. It is expected that the new scheme for charging the storage device will increase the average power spill. We find that there exists a trade-off between reducing the probability of large power spills and reducing the average power spilled by the system.

The probability of occurrence of large power spill is small. The Crude Monte Carlo (CMC) probability estimator is robust but becomes computationally expensive for small probabilities. To reduce the workload of CMC we use the *splitting technique* for rare-event simulations in our study [135]. We use a variant of the splitting technique called the Fixed Number of Successes (FNS) proposed by [15] for calculating the probability of large power spills. [165] used FNS to estimate electrical grid reliability in. It is of great relevance to find an appropriate *importance function* for the splitting technique, as it plays the most significant role in the efficiency of splitting [68]. We formulate an appropriate IF for our hybrid stochastic power system described above.

In Section 3.2 we describe the system, the storage model, problem description and the stochastic models for power generation and demand. Section 3.3 provides details of the splitting technique and the appropriate importance function for the problem. Section 3.4 presents the simulation results showing how the probability of large power spills vary with the battery parameters and charging strategy. In this section we also compare the CMC and FNS computation time. Section 3.5 concludes the chapter.

3.2 System setup

In the single domestic power system with stochastic wind power generation and demand, a battery is incorporated as a storage device in order to reduce large power spills. Let $P(t)$ be the power mismatch between the wind power generation and demand (load) defined as

$$P(t) := W(t) - D(t), \tag{3.1}$$

where $W(t)$ is the wind power generated and $D(t)$ is the power demand at time t. $P(t) > 0$ implies there is excess of power in the system and can be used to charge the energy storage device and $P(t) < 0$ denotes paucity of power in the system and the storage device needs to be discharged.

3.2.1 The storage model

Let us consider a battery as the energy storage device in the power system. The state of the battery at time t is given by $B(t)$ and it has a maximum storage capacity B_{max}. For any storage device there will be bounds on the rate at which it can be charged or discharged known as *ramp constraints* [26]. The ramp constraints are denoted as δ and β such that $\delta < 0$ and $\beta > 0$. Losses occur during charging and discharging the battery which depends on the *efficiency parameters*, α_c and α_d of the battery, where $\alpha_c, \alpha_d \in [0, 1]$. The battery is modeled according to

$$\frac{dB}{dt} = \dot{B}(t) := (\alpha_c \mathbb{1}_{\{P(t)>0\}} + \frac{1}{\alpha_d}\mathbb{1}_{\{P(t)<0\}})P(t), \text{ for } t \in [0, T] \tag{3.2}$$

,

with the *battery constraints*, namely the ramp and capacity constraints imposed on it

$$\delta \leq \dot{B}(t) \leq \beta \quad \text{where } \delta < 0 < \beta,$$
$$0 \leq B(t) \leq B_{max} \quad \forall\, t \in [0, T].$$

T is the time length of 24 hours. $\mathbb{1}_{\{.\}}$ is the indicator function which takes value 1 if the expression in the parenthesis is realized else it takes value zero. Thus, in principle, the battery is charged when $P(t) > 0$ and discharged if $P(t) < 0$ unless the battery constraints are met.

In our computational experiments, time is discretized into $N = \frac{T}{\Delta t}$ time steps, where Δt is the time step of integration. The battery state is updated according to the Euler scheme

$$B(t+1) = \min(B_{max},\ \max(0, B(t) + \Delta B(t))), \tag{3.3}$$

where

$$\Delta B(t) := \min(\beta,\ \max(\delta, \alpha P(t)))\Delta t, \tag{3.4}$$

for $t = 0, \ldots, N-1$ and $\alpha = \alpha_c \mathbb{1}_{\{P(t)>0\}} + \frac{1}{\alpha_d}\mathbb{1}_{\{P(t)<0\}}$. $B(0)$ is the initial state of the battery. If the battery is fully charged, $B(t) = B_{max}$, it will only discharge if $P(t) < 0$. Otherwise if $P(t) > 0$ it remains at B_{max} and vice-versa for the empty state of the battery, i.e., when $B(t) = 0$.

3.2.2 Power Spill

Let $\tilde{P}(t)$ be the amount of power getting absorbed or delivered by the battery,

$$\tilde{P}(t) = \begin{cases} \hat{P}(t) & \text{if } 0 < B(t) < B_{max}, \\ \hat{P}(t) & \text{if } B(t) = B_{max} \text{ and } P(t) < 0, \\ \hat{P}(t) & \text{if } B(t) = 0 \text{ and } P(t) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{3.5}$$

where $\hat{P}(t) = \min(\beta, \max(\delta, \alpha P(t)))$ and $\alpha = \alpha_c \mathbb{1}_{\{P(t)>0\}} + \frac{1}{\alpha_d}\mathbb{1}_{\{P(t)<0\}}$. Let us define the *residual power* as

$$F(t) := P(t) - \tilde{P}(t). \tag{3.6}$$

When $F(t) > 0$, *power spill* occurs: there is more power production than

demand and the battery cannot absorb all the excess power because of the *battery constraints*, i.e., either the battery is completely charged or it cannot charge fast enough due to the ramp constraints. Note that power spill can also occur when $P(t)$ is large and $\alpha_c < 1$.

3.2.3 Problem description

We are interested in calculating the probability of large power spills in the system over a time length T, i.e.,

$$\mathbb{P}(\sup_{t\in[0,T]}\{F(t)\} \geq F^*), \tag{3.7}$$

where $F^* > 0$ is the large *power spill threshold*.

To make an overall assessment of the integrated system we compare how this probability varies for various values of B_{max}, the ramp constraints and for different battery charging schemes.

3.2.4 Modeling power mismatch $P(t)$

As discussed earlier we model the stochastic power mismatch $P(t)$ from (3.1) by modeling the wind power $W(t)$ and the power demand $D(t)$.

3.2.4.1 Wind power generation $W(t)$

Earlier studies have shown that the Rayleigh distribution (a special case of Weibull distribution with shape parameter equal to 2) arises from the wind speeds if the wind vector components along the zonal (east-west) direction, u and the meridional (north-south) direction, v are taken to be individually independently Gaussian distributed with zero means and equal standard deviations [47]. As mentioned in [116], the probability density function for the wind speed obtained from these assumptions can be suitable globally but in general they are not true locally. We generalize the assumptions by taking u and v as independent (no cross-correlation)

and Gaussian with *non-zero* means (μ_u and μ_v respectively) and different standard deviations (σ_u and σ_v respectively). μ_u and μ_v are the average wind velocities in the zonal and the meridional direction respectively. σ_u and σ_v are the standard deviations along the wind component directions. To get realistic values for the means and standard deviations of the wind velocity components, we use hourly wind data from KNMI [5]. The KNMI data consists of hourly measurements throughout 2013 of wind velocity at a height of 10 m at the Schiphol airport in the Netherlands. The Gaussian processes for the wind velocities are modeled as two 1-d Ornstein-Uhlenbeck processes,

$$dU(t) = \Theta_U(M_U - U(t))dt + \Sigma_U dW_U(t), \tag{3.8}$$

and

$$dV(t) = \Theta_V(M_V - V(t))dt + \Sigma_V dW_V(t), \tag{3.9}$$

where $M_U = \mu_u$ and $M_V = \mu_v$; $\Sigma_U = \sqrt{2\Theta_U}\sigma_u$, and $\Sigma_V = \sqrt{2\Theta_V}\sigma_v$. $W_U(t)$ and $W_V(t)$ are Wiener processes. Note that the above described model is similar to the model for wind velocities in [116], however we neglect the surface drag force terms here. The values of Θ_u and Θ_v are to be determined later (see section 3.2.4.3). The simulated wind speed is given by

$$W_s(t) = \sqrt{U^2(t) + V^2(t)}. \tag{3.10}$$

From Figure 3.1 we observe probability density function of the wind speed data generated from the model is in reasonable agreement with the KNMI measurement data. Note from Table 3.1, the mean and the variance of the simulated data are reproduced well compared to the KNMI data, however the skewness is underestimated by the simulated data compared to the KNMI data.

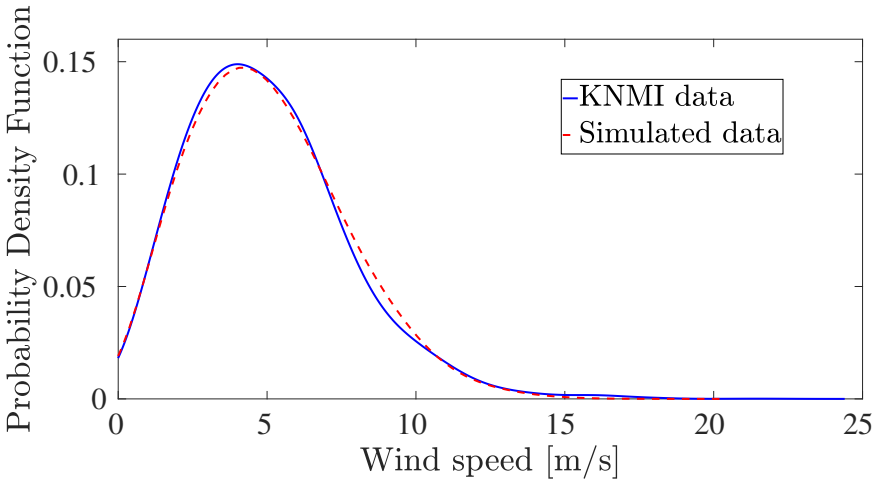


Figure 3.1: Comparing the probability density functions of KNMI measurement data and simulated wind speed data.

	Mean	Variance	Skewness
KNMI data	4.97	7.09	0.92
Simulated data	5.04	6.86	0.61

Table 3.1: Comparing the first three moments of KNMI data and simulated data for wind speed.

The power produced by a wind turbine is a function of wind speed and can be modeled by the following *wind power curve* [110, 118],

$$W(t) = \begin{cases} W_{RP}\left[\frac{W_s(t)-W_{CI}}{W_{RS}-W_{CI}}\right] & \text{if } W_{CI} \leq W_s(t) \leq W_{RS}, \\ W_{RP} & \text{if } W_{RS} < W_s(t) < W_{CO} \\ 0 & \text{otherwise.} \end{cases} \tag{3.11}$$

Here, W_{CI} is the cut-in wind speed at which the turbine starts to generate power. Wind speeds below this threshold are not sufficient enough to exert torque for rotating the blades of the wind turbine. W_{RS} is the rated wind speed at which the turbine produces its rated power W_{RP}. The rated power is the maximum power a wind turbine can produce. For the above

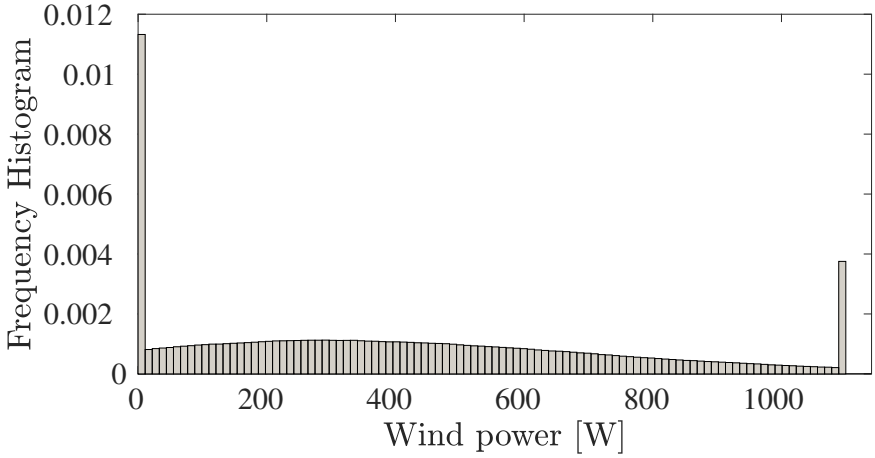


Figure 3.2: Frequency histogram plot of the wind power produced for W_{CI}=2 m/s, W_{CO}=18 m/s, W_{RS}=10 m/s and W_{RP}= 1100 W.

prescribed wind power curve, the wind power is a linear function of wind speed between the wind speeds W_{CI} and W_{RS}. W_{CO} is the cut-out wind speed at which the turbine shuts down for safety reasons. Above the cut-out wind speed the turbine needs to be shut down to avoid physical damage to the turbine rotor. Between wind speeds W_{RS} and W_{CO} the wind turbine produces its rated power W_{RP}. In our study we consider a micro-generator domestic wind turbine with W_{RP} = 1100 W, W_{CI} = 2 m/s, W_{RS} = 10 m/s and W_{CO} = 18 m/s.

From the frequency histogram plot of wind power based on simulated wind speed data in Figure 3.2 we observe that there exist an upper and a lower bound on the power produced because of the physical restrictions on the wind turbine. The peak at zero is due to the turbine output being zero and the higher wind power peak is due to the turbine output being W_{RP}.

3.2.4.2 Power demand $D(t)$

In this section we will discuss the modeling of the power demand $D(t)$. The bar plot in Figure 3.3 shows the frequency histogram of the power

demand (measurement data) of a typical household for one year [1]. We observe that there exist two peaks in the histogram plot, one peak is at 230 W and the other is peak at around 500 W. We model the power demand by the following SDE which mimics these peaks,

$$dD(t) = -\frac{V'(D(t))}{c}dt + \frac{1}{\sqrt{c}}dW_D(t), \tag{3.12}$$

where $V(D(t))$ is the so-called potential function for $D(t)$, $W_D(t)$ is the Wiener process and c is a constant to be determined later (see Section 3.2.4.3). The relation between the time invariant probability density $\psi(D)$ and the potential function $V(D(t))$ for (3.12) is given by

$$\psi(D) = \psi_0 \exp(-V(D)), \tag{3.13}$$

where ψ_0 is a normalization constant. The potential function $V(D)$ was obtained from the density distribution of the measurement power demand data by inverting (3.13):

$$V(D) = -\log \psi(D) + \log \psi_0. \tag{3.14}$$

In order to obtain an expression for the potential, a 6^{th} order polynomial was fitted on the values of $V(D)$ calculated from (3.14). In order to avoid ill-conditioned values for the polynomial constants, the demand data was re-scaled by a factor of 100 and re-centered by 300 W, i.e., $D_{\text{re-scaled}} = \frac{D-300}{100}$.

Figure 3.3 shows the comparison of the probability density function of the simulated data and frequency histogram of the measurement data for power demand. Note to construct the frequency histogram plot of the measurement data in Figure 3.3, we have scaled the data back again. We can see that our power demand model captures the 230 W and 500 W peaks reasonably well. However, the demand model does not capture values greater than 600 W. This is because of the order of the polynomial chosen for $V(D(t))$. In this model we neglect the periodicity in the switch-

i	a_i
1	0.1977
2	-1.1643
3	1.9195
4	0.6014
5	-3.2804
6	0.5133
7	-5.3003

Table 3.2: Polynomial constants for $V(D) \approx \sum_i^7 a_i D^{7-i}$.

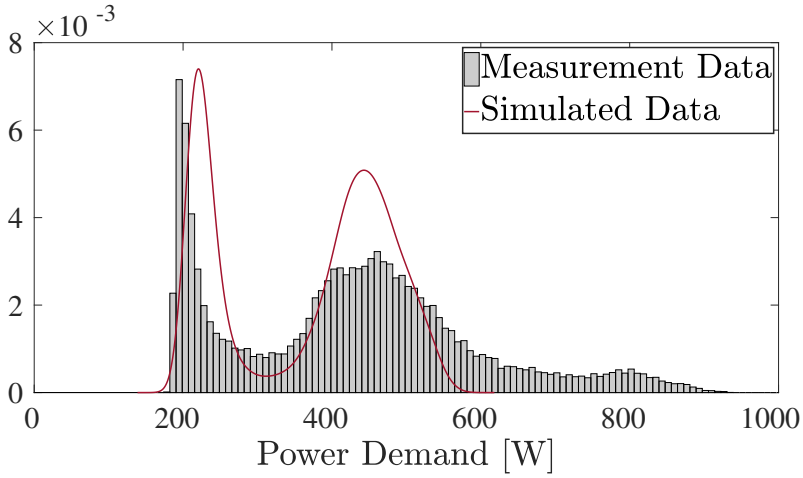


Figure 3.3: Comparison of the measurement data (frequency histogram bar plot) and the simulated data (probability density function in red line) for the yearly electricity consumption of a typical household.

ing related to power demand consumption pattern and consider random switching between these peaks.

3.2.4.3 Auto-correlation function (ACF)

The choice of Θ_U, Θ_V and c for (3.8), (3.9) and (3.12) is done in such a way that all the simulated data decorrelate at a comparable time with the KNMI wind data (see Figure 3.4). To achieve this, we take $\Theta_U = \Theta_V = 0.025$ and $c = 2$.

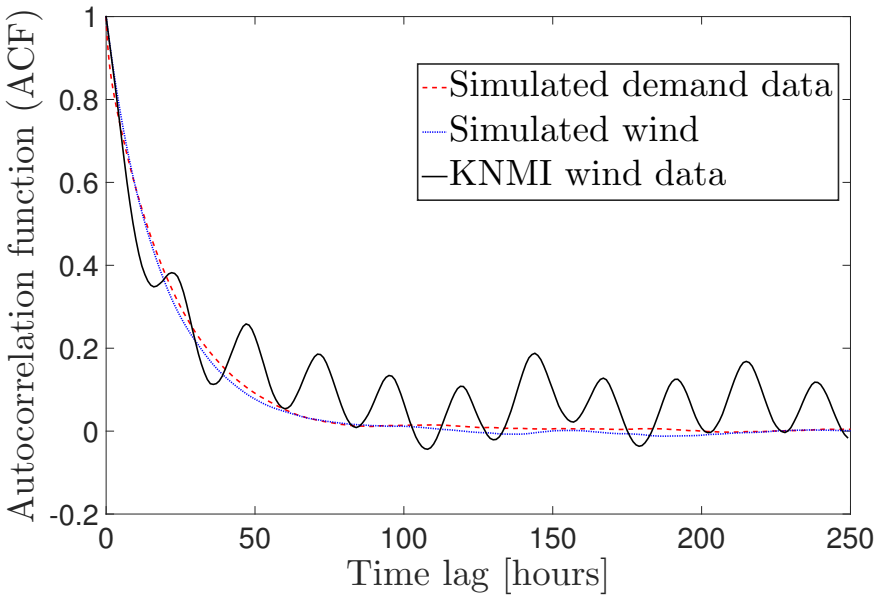


Figure 3.4: Comparing the ACF of the simulated wind and demand data with KNMI wind speed data.

3.2.4.4 Power mismatch

The power mismatch between the generation and demand is given by $P(t) = W(t) - D(t)$. We observe four maxima in the frequency histogram plot of power mismatch in Figure 3.5. These maxima occur when the power generated $W(t)$ is minimum or maximum and the 230 W or 500 W peaks of power demand $D(t)$ occur.

From the wind and demand models discussed in this section we calculate the probability of large power spills (3.7) when a battery is incorporated into the system.

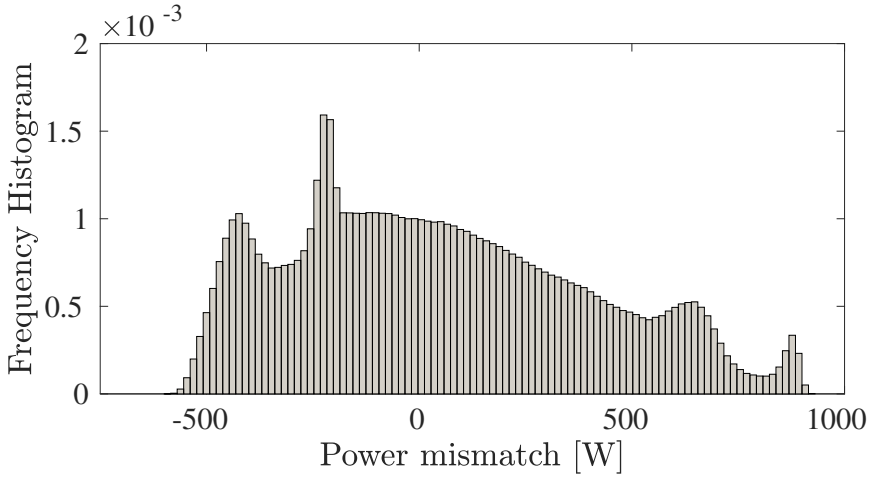


Figure 3.5: Frequency histogram plot of power mismatch between power generation and power demand.

3.3 Splitting technique for rare event simulations

In our model we are interested in estimating probabilities of large power spills when a battery is incorporated in a power system with stochastic power generation and demand. It is expected that the probability of large power spills will be small when a battery is incorporated in the system. The Crude Monte Carlo (CMC) estimations of these small probabilities will become computationally very expensive.

3.3.1 A review of the splitting technique

We now give a brief review of the rare event simulation technique [135]. As discussed in Section 2.4.1 of Chapter 2, the CMC estimation becomes computationally expensive when the value of the probability under consideration, γ is small. To reduce the computational workload for estimating small values of γ we use the *splitting technique* for rare event simulation [69]. In splitting, the sample paths of the stochastic processes involved

are split into multiple copies at various levels of the *importance function* till the rare event set is reached [69]. The importance function measures the distance to the rare-event set. As discussed in Section 2.4.2 of Chapter 2 in splitting, probability γ is decomposed into the product of several conditional probabilities which occur more easily and are hence less computationally intensive to calculate. Hence we get the estimate of γ to be (see (2.15) of Chapter 2),

$$\hat{\gamma} = \prod_{k=1}^{m} \hat{p}_k, \tag{3.15}$$

where m is the number of sub-intervals the importance function valued between $[0, L]$ is divided into, and

$$\hat{p}_k = R_k / N_{k-1} \text{ for all } R_k > 0. \tag{3.16}$$

In the above equation, R_k is the number of sample paths that have hit the k-th level before the end time T and N_k is the total number of sample paths launched at level k. N_k depends on which variant of splitting is deployed. Please see Section 2.4.2 of Chapter 2 for more details on splitting technique.

3.3.1.1 Variants of splitting

We briefly discuss the three main variants of the splitting technique:

- **Fixed Splitting**: Each of the sample paths that has entered level k is resampled exactly $C_k \geq 1$ times where C_k is fixed [69]. Hence, the total number of sample paths launched from level k becomes $N_k = C_k R_k$. Note that both N_k and R_k are random now. Using (3.16) we get,

$$\hat{p}_k = \frac{R_k}{C_{k-1} R_{k-1}} \text{ for } k = 1, \ldots, m, \tag{3.17}$$

  where $C_0 = 1$ and $R_0 = N_0$. The disadvantage of this method is

that paths might not hit the next level leading to a wrong estimate $\hat{\gamma} = 0$. This can happen if the levels are not chosen optimally.

- **Fixed Effort**: N_k, the number of samples path launched from level k is kept fixed. Resampling for the next level $k+1$ is done N_k times by random sampling with replacement of the R_k paths that have reached level k. The conditional probability estimates are given by,

$$\hat{p}_k = \frac{R_k}{N_{k-1}} \text{ for } k = 1, \ldots, m. \tag{3.18}$$

  This method too suffers from path explosion and extinction problem.

- **Fixed Number of Successes (FNS)**: This variant of splitting has been developed by [15]. For this case the number of sample paths that reaches level k, i.e, R_k is kept fixed and the resampling with replacement from $k-1$ level is done till R_k sample paths have reached (or crossed) level k.

  The unbiased estimator of the rare-event probability is given by,

$$\hat{\gamma} = \hat{p}_k = \prod_{k=1}^{m} \frac{R_k - 1}{N_{k-1} - 1} \text{ for } k = 1, \ldots, m. \tag{3.19}$$

  As R_k is kept fixed, this method does not suffer from path explosion or extinction. We will use this variant of splitting technique in this chapter and in Chapter 4.

  The unbiased estimator for the variance $\mathrm{Var}(\hat{\gamma})$ is not known for the FNS method. However, under the assumption that the conditional hitting probability does not depend on the entrance states of the previous stage, i.e., $\mathbb{P}(H_k|H_{k-1}, (T_{k-1}, \boldsymbol{X}(T_{k-1}))) = \mathbb{P}(H_k|H_{k-1})$, $(\forall(T_{k-1}, \boldsymbol{X}(T_{k-1})), \forall k)$, the squared relative error of $\hat{\gamma}$ can be bounded by

$$\mathrm{SRE}(\hat{\gamma}) \leq \prod_{k=1}^{m} \left(\frac{1}{R_k - 2} + 1\right) - 1. \tag{3.20}$$

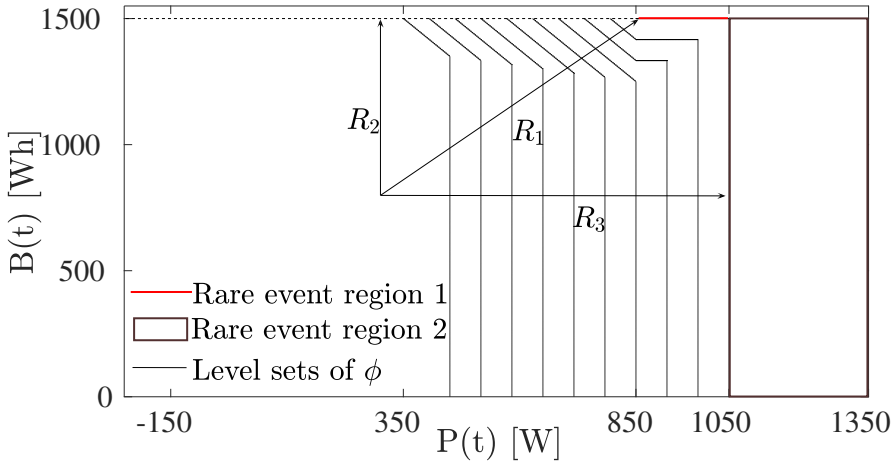


Figure 3.6: The phase space of $B(t)$ and $P(t)$, where $B_{max} = 1500$ Wh, $F^* = 850$ W and $\beta = 200$ W. Rare event region 1 : $B(t) \geq 1500$ W and $P(t) \geq 850$ W. Rare event region 2: $0 \leq B(t) \leq 1500$ and $P(t) \geq 1050$ W.

In the description above $\boldsymbol{X}(t)$ in a n-dimensional Markov process, T_k is the time of hitting the k-th level and H_k is the event that the k-th level is hit during the time interval $[0, T]$ (see Section 2.4.2 of Chapter 2 for more detailed definitions).

3.3.2 Importance function for the stochastic domestic power system

The importance function plays a pivotal role in the efficiency of splitting [68]. In this section we present an appropriate importance function for the stochastic power system with a battery incorporated. We define the importance function, ϕ as the distance of the system from the rare event sets in the phase space of the state of battery $B(t)$ and the power mismatch $P(t)$.

For formulating the importance function, we define three ℓ_1 (or Manhattan) distances from the rare event sets (see Figure 3.6) as

$$\begin{cases} R_1(P(t), B(t)) & = a_1(F^* - P(t)) + a_2(B_{max} - B(t)), \\ R_2(P(t), B(t)) & = a_2(B_{max} - B(t)), \\ R_3(P(t), B(t)) & = a_3(F^* + \beta - P(t)), \end{cases}$$

where, $a_1 = 1/F^*$, $a_2 = 1/B_{max}$ and $a_3 = 1/(F^* + \beta)$. We define the importance function as

$$\phi(B(t), P(t)) = \begin{cases} -\min(R_1(P(t), B(t)), R_3(P(t), B(t))) \\ \qquad\qquad \text{if } P(t) < F^*, \\ -\min(R_2(P(t), B(t)), R_3(P(t), B(t))) \\ \qquad\qquad \text{if } P(t) \geq F^*. \end{cases} \tag{3.21}$$

The negative sign makes ϕ an increasing function in its arguments.

Figure 3.6 depicts the level sets of the importance function ϕ and the rare-event sets in the phase space of the battery state $B(t)$ and the power mismatch $P(t)$. Power spill occurs when $P(t) > 0$ and the battery cannot absorb all the excess power because of the *battery constraints*, i.e., either the battery is completely charged or it cannot charge fast enough due to the ramp constraints. As discussed in Section 3.2.3, our problem is to find the probability of large power spills, i.e., when $F(t) \geq F^*$, where $F^* > 0$. *Rare event region 1* occurs when the battery is charged to its maximum capacity B_{max} and $P(t) \geq F^*$. *Rare event region 2* occurs when the battery cannot absorb all the power available to it because of the ramp constraints imposed on it and the residual power left is spilled. This happens when $P(t) \geq F^* + \beta$.

For splitting, we construct the levels sets of the importance function ϕ such that it depicts the distance of the system state from the rare event

sets. From (3.7) and (3.21) we have

$$\hat{\gamma} := \mathbb{P}(\sup_{t\in[0,T]}\{F(t)\} \geq F^*) = \mathbb{P}(\sup_{t\in[0,T]}\{\phi(B(t), P(t))\} \geq 0).$$

3.4 Results and Discussion

In this section, we discuss how the probability of power spills (beyond a threshold) varies with the battery size, ramp constraint imposed on the battery and for dimportance functionferent charging scheme of the battery. We integrate (3.2), (3.8), (3.9), and (3.12) using forward Euler integration scheme with time step $\Delta t = 0.01$ hours and $T = 24$ hours. For all the calculations we take the initial state of the battery $B(0) = B_{max}/2$ and $P(0) =$ -65 W. For simplicity we use $\delta = -\beta$ and the *efficiency parameters* $\alpha_c = \alpha_d = 1$, as it does not add any new character to the structure of the model [26]. We calculate the probability of power spills greater than 850 W, i.e., $F^* = 850$.

For performing FNS we first calculate the number of levels m by the pilot run such that $\tilde{p}_k$ is nearly equal to the optimal value of $\mathrm{p_{opt}} \approx 0.2032$ [15]. We follow the procedure prescribed in Chapter 6 of [163]. For the pilot run we use $R_k = 50\ \forall k$. For the final run we calculate R_k from (3.20) such that the SRE $(\hat{\gamma}) \leq 0.01$. In order to obtain an accurate estimate of the probabilities, FNS is repeated $n \geq 30$ times (suggested by [67]) to calculate the mean of the estimator

$$\hat{\bar{\gamma}}_n := \frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_i. \tag{3.22}$$

We are interested in the squared relative error of the mean given by

$$\mathrm{SRE}(\hat{\bar{\gamma}}_n) := \frac{1}{n}\mathrm{SRE}(\hat{\gamma}_i). \tag{3.23}$$

To obtain the value of n we first repeat FNS 30 times, and it is repeated further until the $\mathrm{SRE}(\hat{\bar{\gamma}}_n) \leq 0.005$.

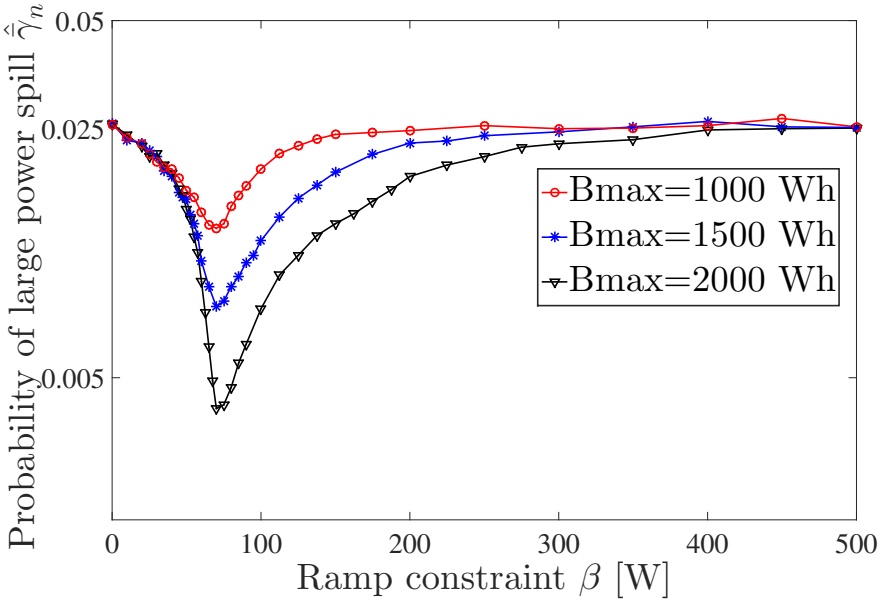


Figure 3.7: Probability of large power spill versus ramp constraint with F^*=850 W and T= 24 hours for various storage capacities.

3.4.1 PLPS for different values of B_{max} and β

We study the effect of the ramp constraint β on the probability of large power spills $\hat{\bar{\gamma}}_n$ for various battery capacities B_{max}. Figure 3.7 shows how $\hat{\bar{\gamma}}_n$ varies with β for various values of B_{max}. We observe that $\hat{\bar{\gamma}}_n$ reduces with β till an optimal value β^* where it is minimal, then again increases and becomes constant. This means that, either a very fast or a very slow charging/discharging battery accounts for more large power spills. A very fast charging/discharging battery will get to its maximum capacity very soon; hence won't be able to store any excess power generated which will account for the large power spills. On the other hand, a very tightly constrained battery (small β) cannot absorb all the excess power in the system and power is spilled.

In terms of the phase space (see Figure 3.6), for small values of β the system hits the *rare event region 2* and for $\beta > \beta^*$ the system hits *rare event region 1.*

We observe that for very small values of β there is no effect of the battery size B_{max} on the $\hat{\bar{\gamma}}_n$. This is because when the battery is very restricted it never gets charged to its maximum capacity, hence B_{max} does not affect the probabilities. As β approaches β^* we see the effects of B_{max}: it reduces $\hat{\bar{\gamma}}_n$, which is expected. The larger the value of B_{max}, the longer it will take the battery to reach region 1. When β is very large the effect of B_{max} becomes negligible because the battery reaches it maximum capacity very fast and any excess power in the system greater than F^* lead to large power spills.

3.4.2 PLPS for different battery charging strategies

In order to reduce the large power spills further, we employ a different charging strategy for the battery. A fraction of the battery $1-\epsilon$ is reserved for absorbing only those values of excess power which are greater than a threshold, where $0 \leq \epsilon \leq 1$. In our case we take the threshold same as the *power spill threshold* F^*. The battery is charged till :

1. ϵB_{max} if $P(t) < F^*$
2. B_{max} if $P(t) \geq F^*$.

The amount of power absorbed/ delivered by the battery is

$$\tilde{P}(t) = \begin{cases} 0 & \text{if } B(t) = \epsilon B_{max} \text{ and } 0 < P(t) < F^*, \\ 0 & \text{if } B(t) = B_{max} \text{ and } P(t) > 0, \\ 0 & \text{if } B(t) = 0 \text{ and } P(t) < 0, \\ \hat{P}(t) & \text{otherwise,} \end{cases}$$

where $\hat{P}(t) = \min(\beta, \max(\gamma, \alpha P(t)))$ and $\alpha = \alpha_c \mathbb{1}_{\{P(t)>0\}} + \frac{1}{\alpha_d}\mathbb{1}_{\{P(t)<0\}}$. We compare the probability of large power spills $\hat{\bar{\gamma}}_n$ for different ϵ.

Figure 3.8 shows $\hat{\bar{\gamma}}_n$ versus β with $B_{max} = 1500$ Wh, $F^* = 850$ W and $T = 24$ hours for various values of ϵ. We observe that as the value of ϵ

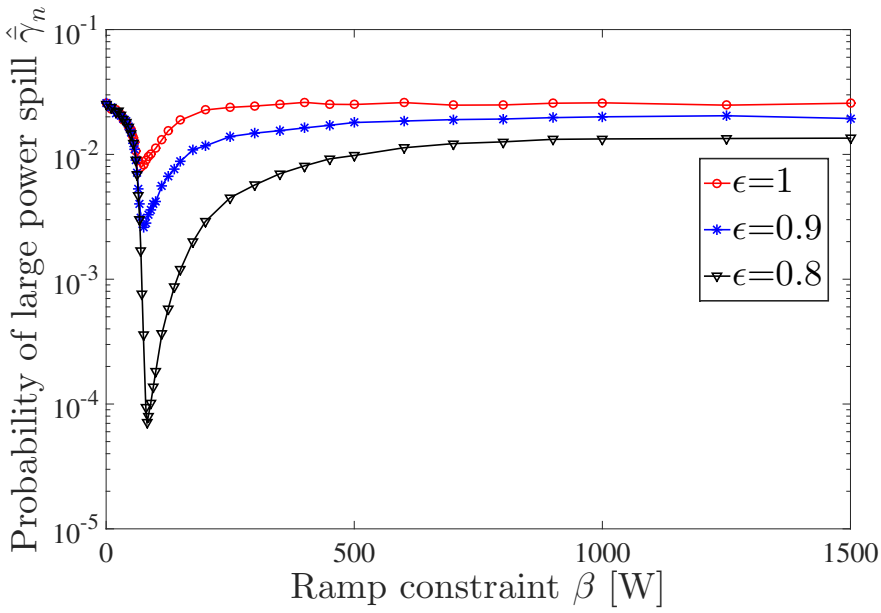


Figure 3.8: Probability of large power spill versus ramp constraint β with B_{max}=1500 Wh, F^*=850 W and T= 24 hours for various values of ϵ.

reduces, that is, as we reserve the battery more for larger fluctuations in $P(t)$, $\hat{\bar{\gamma}}_n$ decreases. This happens because of the reserved space accessible to the battery only for absorbing the large values of $P(t)$. We observe similar response of the system to β as Figure 3.7. From Figure 3.8 we observe that the minimum value of $\hat{\bar{\gamma}}_n$ drops by a factor of 1000 from $\epsilon = 1$ to $\epsilon = 0.9$. So, the more we reserve the battery for large power mismatch, the lower will the probability of large power spills $\hat{\bar{\gamma}}_n$ become.

3.4.3 Average power spill

In the previous section 3.4.2 we observe that, if we reserve the battery for larger fluctuations in the net power mismatch, i.e., as ϵ decreases, the probability of large power spills $\hat{\bar{\gamma}}_n$ goes down drastically. But this increases the probability of small power spills. In this section we examine the time average of power spilled by the system over 24 hours versus β for

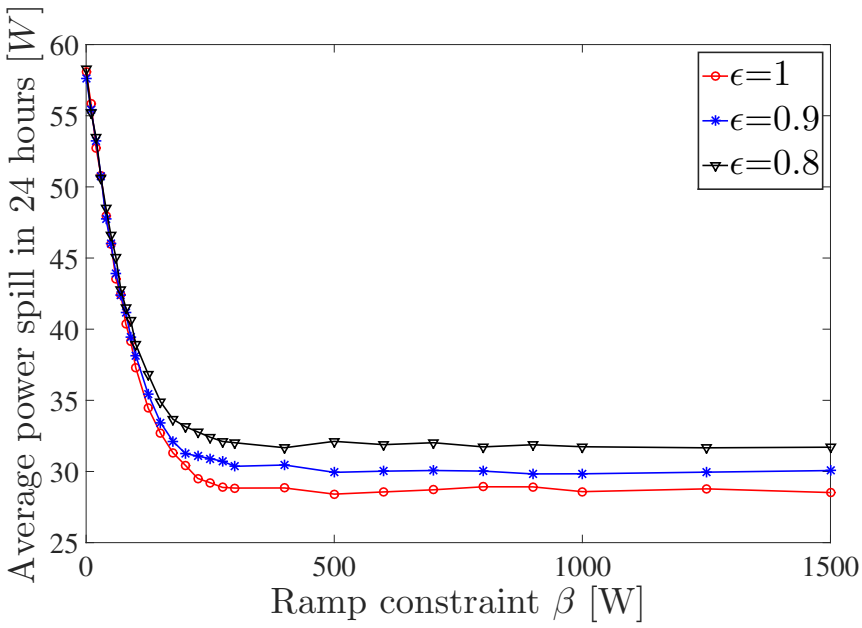


Figure 3.9: Average power spill power spill versus ramp constraint β with $B_{max} = 1500$ Wh, $F^* = 850$ W and $T = 24$ hours for various values of ϵ.

different values of ϵ

$$\langle [F(t)]^+ \rangle_T = \frac{1}{T} \int_0^T [F(t)]^+ dt, \tag{3.24}$$

where $[x]^+ := \max(0, x)$.

Figure 3.9 shows the average power spill over a 24 hours of time interval versus β with $B_{max} = 1500$ Wh and $F^* = 850$ W for different ϵ. We observe that, the more we reserve the battery for the large values of $P(t)$, the higher the average power spill becomes. This happens because, when the battery state reaches ϵB_{max} it can further charge to its maximum capacity only if $P(t) \geq F^*$ and net power generated $P(t) < F^*$ is spilled. Hence, we observe increased average power spill.

β		CMC	FNS
40	$\bar{\gamma}_n$	1.81×10^{-2}	1.81×10^{-2}
	CPU time(s)	1×10^{3}	1.24×10^{2}
75	$\bar{\gamma}_n$	3.61×10^{-4}	3.60×10^{-4}
	CPU time(s)	2.55×10^{4}	4.33×10^{2}
85	$\bar{\gamma}_n$	7.11×10^{-5}	7.12×10^{-5}
	CPU time(s)	1.51×10^{5}	5.52×10^{2}

Table 3.3: Comparing the computation time for CMC and FNS for B_{max}= 1500 W, ϵ= 0.8, F^*= 850 W and n=30. $\bar{\gamma}_n$ is the mean of the probability estimator for the estimation methods used.

3.4.4 CMC versus FNS

We compare the computation time for CMC and FNS probability estimator such that the squared relative error for both the methods are comparable, i.e, $\text{SRE}(\tilde{\gamma}) \approx \text{SRE}(\hat{\gamma}_i)$. Table 3.3 compares the computation time of CMC and FNS for few values of the probability of large power spill. The CPU time of FNS includes the pilot runs. As the value of the probability of large power spills goes down the time gain of FNS over CMC becomes more profound. The simulations are performed using MATLAB 2012b on an Intel Core 2.50 GHz.

3.5 Conclusion

In our assessment for finding the best design to operate the storage device in a single domestic power system with wind generation for the simple switching strategy of the battery, we observe that the ramp constraints imposed on the battery play a major role in mitigating the large power spills. We find there exists an optimal value of the ramp constrains for which occurrence of large power spills is minimal. We also find that a fast charging/discharging battery increases the probability of large power

spills. Increasing the storage capacity B_{max} reduces the probability of large power spill only when the battery operates around the optimal values of the ramp constraint β.

We employed a different charging scheme to the battery where a certain part of it is reserved only for absorbing large values of excess power in the system. We found that the probability of large power spills goes down significantly with the reserve level (see Figure 3.8). But this charging scheme comes with a trade-off. The more we reserve the battery for larger fluctuations in the excess power generated, the higher the average power spill becomes, as depicted in Figure 3.9. It should be noted that there is a nominal increase in the average power spill with ϵ, whereas the probability of the large power spills decreases drastically with ϵ.

We formulated the importance function for the FNS splitting technique used to calculate the probability of large power spills for our system. Table 3.3 shows the time efficiency of FNS over CMC. The time gain of FNS over CMC becomes more evident as the probability of interest becomes smaller.

Chapter 4

A computational method for optimizing storage placement to maximize power grid reliability

The strive for reducing carbon footprints and a carbon free future has rapidly increased the usage of renewable energies in power grid. Renewable energy sources like photo-voltaic (PV) arrays and wind turbines are unpredictable in nature, which lead to intermittent power generation. The integration of intermittent renewable energy sources into the electrical power grid challenges the grid reliability.

However, grid reliability can be enhanced by incorporating energy storage devices (batteries) in the grid. The energy storage device acts as buffer by storing excess energy generated and delivering power when there is energy deficiency. The peak-shaving benefit of batteries have been studied

long back in 1981 by [176]. [24] developed a probabilistic method to study the ability of energy storage to increase penetration of intermittent energy sources in power grids.

Recent studies have investigated storage placement in the power grid under the framework of optimal power flow in [12, 40, 49, 70, 72, 124, 150, 161]. [72] minimized the hourly social cost using a market-based probabilistic optimal power flow with energy storage integration and wind generation. [70] proposed a solution strategy to solve the optimal control problem to investigate the effects of different energy storage capacities on generation costs and peak-shaving. However, they have neglected uncertainties due to fluctuations in demand and intermittency in generation. To study the energy storage dispatch and placement problem in power network with wind generation, [150] proposed a risk-mitigated optimal power flow framework. [40] studied optimal placement of large-scale energy storage in power grids using semidefinite relaxation of AC optimal power flow. [124] proposed a method to model the storage devices under the framework of DC optimal power flow.

In this chapter we focus on the optimal storage placement in a power grid for reliable operation of the grid. We model the line currents in the grid according to the DC power flow equations and consider the Probability of Line Current Violation (PLCV) as the *reliability index* of the grid. PLCV calculates the probability that one of the lines in the grid has been overloaded, i.e., one of the line currents has exceeded its allowed maximum. These line current violations lead to physical damage to the lines because of the eventual temperature overload (see Section 2.2 of Chapter 2).

Given the distribution of the stochastic processes of the uncertain power injections and the total storage installation size, we aim to find the optimal placement of the storage devices in the grid such that the PLCV is minimal. In order to do so, we use the *Simulated Annealing* algorithm to minimize the PLCV in the configuration space of different

storage sizes and locations. We resort to simulated annealing because the configuration space of different storage locations and capacities is very large and the quantity we wish to minimize (the PLCV) is not guaranteed to be convex. [159] have briefly discussed using simulated annealing for optimal storage placement in power grid to minimize generation cost. Our study focuses on minimizing a cost function which is a rare-event probability.

For reliable operation of the power grid, PLCV should be small. The conventional Crude Monte Carlo (CMC) method is robust but becomes very inefficient for estimating small probabilities. To increase the efficiency we use the *splitting technique* for rare-event simulations [135]. We use a variant of splitting called the Fixed Number of Successes (FNS) proposed by [15]. [165] used FNS to estimate electrical grid reliability. The efficiency of splitting is highly dependent on the *importance function* used [68]. We develop an appropriate importance function for our problem. We verify numerically that our importance function works well. A theoretical analysis would be interesting but is beyond the scope of this research work. Note that this is non-trivial since the power injections in the grid are not diffusion processes due to the buffers (storage devices), and as such, our setting does not fit in the framework of [166].

To the best of our knowledge, the combination of rare-event simulation with simulated annealing has not been carried out before. [147] used splitting to estimate the probability of large-scale blackout in power network, which is embedded within a higher level optimization technique to minimize the probability subjected to a budget constraint. We apply our method to the IEEE-14 bus test case grid. The uncertain power injections are modeled as Ornstein-Uhlenbeck processes. Given a fixed total storage instalment capacity, we find the optimal storage locations and capacities at each nodes of the grid such that the PLCV is minimal. The total fixed storage installation capacity reflects financial/physical constraints. We start from different initial configurations of the storage locations and

capacities to check the convergence of simulated annealing to the final storage configuration.

In Section 4.1 we discuss the power flow equations and the stochastic processes used to model the net power generation and the storage model. Section 4.2 defines the problem. Section 4.3 introduces the simulated annealing algorithm and its various aspects used in the problem. Section 4.4 provides details of the importance function used for the problem. Section 4.5 presents the simulation results showing how SA algorithm along with FNS minimizes PLCV efficiently for the given IEEE-14 bus test case grid. Finally we conclude in Section 4.6.

4.1 System setup

The topology of the power grid can be defined by a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N} := \{1, 2, \ldots, N\}$ is the set of nodes (also known as buses) and $\mathcal{E}$ is the set of edges (also called lines). In this chapter we solve the direct current (DC) power flow equations for calculating the line currents.

4.1.1 Power flow equations

In this section we discuss both the alternating current (AC) power flow equations and the DC power flow equations. The DC power flow equations can be seen as a linear approximation to the AC power flow equations [77]. In order to derive the power flow equations we first define the following quantities:

- **Admittance matrix:** Let $y_{i,j} \in \mathbb{C}$ be the admittance of the line $(i, j) \in \mathcal{E}$, and $y_{i,i}$ is the admittance-to-ground at bus $i \in \mathcal{N}$. If $(i, j) \notin \mathcal{E}$, then $y_{i,j} = 0$. The admittance matrix $\mathbf{Y} \in \mathbb{C}^{N \times N}$ is given by

$$Y_{i,j} = |Y_{i,j}| e^{\mathbf{i}\alpha_{i,j}} = \begin{cases} -y_{i,j} & \text{if } i \neq j, \\ \sum_k y_{i,k} & \text{if } i = j. \end{cases} \tag{4.1}$$

The row sums of $\mathbf{Y}$ are zero and it is a symmetric matrix. The conductance matrix $\mathbf{G}$ and the susceptance matrix $\mathbf{B}$ are the real and imaginary parts of the admittance matrix $\boldsymbol{Y}$, respectively, such that $Y_{i,j} = G_{i,j} + \boldsymbol{i}B_{i,j}$. In order to avoid confusion we will denote the imaginary unit as $\boldsymbol{i}$.

- **Bus voltages:** Let $\boldsymbol{V} = (V_i,\ \forall i \in \mathcal{N})$ be the vector of bus voltages. In polar form it is given by

$$V_i = |V_i|e^{\boldsymbol{i}\theta_i}, \tag{4.2}$$

where $|V_i| \in [0, \infty)$ is the voltage magnitude and $\theta_i \in (-\pi, \pi]$ is the voltage angle of bus i.

- **Line current flow:** Let $I_{i,j} \in \mathbb{C}^{N\times N},\ \forall (i,j) \in \mathcal{E}$ be the current flowing in line (i,j) from node i to j and is given by Ohm's law,

$$I_{i,j} = y_{i,j}(V_i - V_j). \tag{4.3}$$

- **Bus current injections:** Let $\boldsymbol{I} = (I_i \in \mathbb{C},\ \forall i \in \mathcal{N})$ be the vector of bus current injections which are given by Ohm's law and Kirchoff's current law,

$$\boldsymbol{I} = \boldsymbol{YV}. \tag{4.4}$$

- **Complex nodal power:** Let $S_i \in \mathbb{C}$ be the complex power injected at node i and is given by

$$S_i = P_i + \boldsymbol{i}Q_i, \tag{4.5}$$

where the real part $P_i \in \mathbb{R}$ is called the active power. The imaginary part $Q_i \in \mathbb{R}$ is called the reactive power.

4.1.1.1 AC power flow equations

Let $\boldsymbol{S} = \boldsymbol{P} + \boldsymbol{iQ} \in \mathbb{C}^N$ be the vector of nodal complex power injections. The AC power flow equations are given by

$$S_i = V_i I_i^* \quad \forall i \in \mathcal{N}. \tag{4.6}$$

Using Ohm's law (4.4) the above equation (4.6) can be written as $\forall i \in \mathcal{N}$

$$S_i = V_i \sum_{k=1}^{N} Y_{i,k}^* V_k^*. \tag{4.7}$$

Substituting (4.1) and (4.2) in (4.7) we get

$$S_i = \sum_{k=1}^{N} |V_i||Y_{i,k}||V_k|(\cos(\theta_i - \theta_l - \alpha_{i,k}) + \boldsymbol{i}\sin(\theta_i - \theta_k - \alpha_{i,k})). \tag{4.8}$$

Comparing the real and the imaginary part we obtain the *AC power flow equations* in polar form $\forall i \in \mathcal{N}$

$$P_i = \sum_{k=1}^{N} |V_i||V_k|(G_{i,k}\cos(\theta_i - \theta_k) + B_{i,k}\sin(\theta_i - \theta_k)) \tag{4.9}$$

and

$$Q_i = \sum_{k=1}^{N} |V_i||V_k|(G_{i,k}\sin(\theta_i - \theta_k) - B_{i,k}\cos(\theta_i - \theta_k)). \tag{4.10}$$

4.1.1.2 DC approximations

Under DC approximation the following assumptions are made:

1. $|V_i| = 1, \ \forall i \in \mathcal{N}$.

2. The voltage phase angle differences across a transmission line are

small, i.e, $\forall(i,j) \in \mathcal{E}$, we have $\sin(\theta_i - \theta_j) \approx (\theta_i - \theta_j)$ and $\cos(\theta_i - \theta_j) \approx 1$.

3. The resistive components of each line are ignored, i.e., $Y_{i,j} = \boldsymbol{i}B_{i,j}$.

4. The real power flow across lines are significantly larger than the reactive power flow, i.e., $P_{i,j} >> Q_{i,j} \ \forall(i,j) \in \mathcal{E}$. Hence the reactive power is ignored.

These assumptions reduces the AC power flow equations into

$$P_i = \sum_k B_{i,k}(\theta_i - \theta_k) \ \forall i \in \mathcal{N}, \tag{4.11}$$

known as the *DC power flow equations*. For any line $(i,j) \in \mathcal{E}$, the line current $I_{i,j}$ flowing from $i \to j$ is given by the Ohm's Law

$$I_{i,j} = Y_{i,j}(V_i - V_j) = B_{i,j}(\theta_i - \theta_j). \tag{4.12}$$

Conventionally, one bus in the grid is taken to be reference bus and is called the *slack bus*. Typically it is designated to node 1 in the grid and its voltage angle is set to be zero, $\theta_1 = 0$ and $P_1 = -\sum_{j=2}^{N} P_j$.

4.1.2 Power Generation

We interpret every non-slack node as a single household which has stochastic generation $G(t)$ and demand $D(t)$ at time t and produces net power $P(t) := G(t) - D(t)$. We model the net power generation at the i-th non-slack node $P_i(t)$ as discretized Ornstein-Uhlenbeck (OU) processes which are in fact AR(1) processes,

$$\Delta P_i(t) = \beta_i(\mu_i - P_i(t))\Delta t + \sigma_i \Delta W_i(t) \text{ for } i = 1, \dots N-1. \tag{4.13}$$

where μ_i is the long term mean, β_i is the mean reverting term, σ_i is the volatility term and $W_i(t)$ denotes the Wiener process of the i^{th} OU process.

The values of these terms are determined later in section 3.4. Modeling power injections as OU processes have been suggested by [165, 166].

4.1.3 Storage Model

We consider the storage devices (batteries) to be co-located with the stochastic non-slack nodes and are charged/discharged locally by the net power produced at each node. Let $B_i(t)$ be the level of energy stored in the battery at time t at the i-th non-slack node, and it has a maximum capacity $B_i^{\max}$. The batteries are updated according to

$$B_i(t + \Delta t) = B_i(t) + p_i^{\mathrm{B}}(t)\Delta t \quad \forall t \in [0, T], \tag{4.14}$$

where $p_i^{\mathrm{B}}(t)$ is the power flowing in/out of i-th the battery, Δt is the length of the time step and T is the time horizon of interest. The batteries are bounded by their corresponding capacity and total installation capacity constraints

$$0 \leq B_i(t) \leq B_i^{\max} \quad \text{and} \quad \sum_i B_i^{\max} = B^{\mathrm{tot}} \quad \forall t \in [0, T]. \tag{4.15}$$

Let $P_i^{\mathrm{B}}(t)$ be the power generated by the i-th battery and is given by $P_i^{\mathrm{B}}(t) = -p_i^{\mathrm{B}}(t)$. This is because, $p_i^{\mathrm{B}}(t) > 0$ implies the battery is getting charged and it is consuming power and $p_i^{\mathrm{B}}(t) < 0$ implies the battery is discharging and is generating power.

4.1.3.1 Switching of the Battery

The battery charging/discharging depends on the power flowing in/out of the battery $p_i^{\mathrm{B}}(t)$ which is given by

$$p_i^{\mathrm{B}}(t) = \begin{cases} P_i(t) & \text{if } 0 \leq X_i(t) \leq B_i^{\max} \\ (B_i^{\max} - B_i(t))/\Delta t & \text{if } X_i(t) > B_i^{\max} \\ -B_i(t)/\Delta t & \text{if } X_i(t) < 0. \end{cases} \tag{4.16}$$

where $X_i(t) = P_i(t)\Delta t + B_i(t)$ for $i = 1, \ldots, N-1$. $X_i(t)$ is the energy level of the battery at time t without any constraints imposed. Hence the battery charging/discharging is dependent on the net power generated by the stochastic non-slack buses and the state of the battery. The above equation (4.16) ensures that (4.15) is true. To keep the storage model simple we neglect the *ramp constraints*, the imposed maximal charging/discharging rate on the storage device in this study (see Section 3.2.1 of Chapter 3).

4.2 Problem description

Our aim is to find the optimal battery locations and capacities at each node to ensure a reliable operation of the grid. We consider the *Probability of Line Current Violation* (PLCV) as the reliability index of the power grid. PLCV is defined as the probability that any one of the line currents violate its given line constraint maximum at any time $t \in [0, T]$, i.e,

$$\gamma := \mathbb{P}\{\exists (i,j) \in \mathcal{E} : \sup_{t \in [0,T]} |I_{i,j}(t)| \geq I_{i,j}^{\max}\}. \tag{4.17}$$

In the above equation $I_{i,j}$ is the current flowing between nodes i and j and $I_{i,j}^{\max} > 0$ is the maximum current carrying capacity the edge connecting nodes i and j.

To solve the optimal storage (battery) placement problem and calculating PLCV we use a novel combination of two algorithms namely the *simulated annealing* algorithm and the *splitting* technique for rare-event simulations, respectively, discussed in the subsequent sections.

4.3 Simulated annealing algorithm

We wish to find the optimal location and capacities of the battery in the grid such that PLCV is minimal. We do not expect PLCV to be convex. Also, the configuration of space of battery locations and capacities to grow

exponentially with the number of nodes in the grid. To overcome these problems we use the simulated annealing (SA) [160] algorithm to minimize PLCV.

Simulated annealing is a metaheuristic algorithm designed to approximate the global optimum of a given function. The main aim of the algorithm is to perform a local search in the solution space $\boldsymbol{X}$ of the problem to minimize a desired *cost function* $f(\boldsymbol{X})$. Annealing is a physical process of heating a solid to very high temperature, then it is cooled by slowly lowering the temperature of the solid for eliminating point defects. It should be noted that there is no hard guarantee to find the global optimum using SA. The SA algorithm is based on the following search principle :

1. Start with an initial solution and consider it as the best solution $\boldsymbol{X}_{\text{best}}$. Initialize T_c, the temperature of the *acceptance probability* of a *bad* solution .

2. Randomly select a new solution $\boldsymbol{X}^*$ in the neighborhood of the previously obtained best solution.

3. If the new solution is better than the previously found best solution, i.e, if $\Delta E = f(\boldsymbol{X}^*) - f(\boldsymbol{X}_{\text{best}}) < 0$, then consider the new solution as the best solution $\boldsymbol{X}_{\text{best}} = \boldsymbol{X}^*$.

4. If not, i.e, $\Delta E = f(\boldsymbol{X}^*) - f(\boldsymbol{X}_{\text{best}}) > 0$, then accept the new *bad* solution as the best solution with a probability $\exp(-\Delta E/T_c)$.

5. Slowly cool the temperature of the acceptance probability (decrease T_c).

6. Repeat from 2 until the *stopping criterion* is reached.

The cooling law has to be chosen carefully to allow the algorithm sufficiently explore the region around the initial guess. If the cooling is too fast, the system will get stuck in the nearest local minimum and the algorithm may not converge. If the cooling is too slow, the algorithm spends a

lot of time in useless explorations which slows down the process. Usually, an exponential decrease is considered for the cooling by multiplying the current temperature with a constant, i.e. $T_c^{\text{new}} = \kappa T_c^{\text{old}}$ where $0 < \kappa < 1$.

4.3.1 Cost Function: $\log(\gamma)$

In this study we want to minimize γ (PLCV (4.17)) in the battery locations and capacities configuration space. The values of γ's are typically small and while minimizing can go down to $\sim 10^{-5} - 10^{-7}$ or smaller depending on the total installation capacity of the battery. The *acceptance probability* of the bad solution not only depends on T_c but also on the difference of the function values for the new solution and the previously found best solution $\Delta E = \gamma(\boldsymbol{X}^*) - \gamma(\boldsymbol{X}_{\text{best}})$. As the γ's are very small, their differences are also small hence the acceptance probability becomes large and the algorithm accepts too many bad solutions and might never converge. So, instead of minimizing γ we minimize $\log(\gamma)$ such that $\Delta E = \log(\gamma(\boldsymbol{X}^*)) - \log(\gamma(\boldsymbol{X}_{\text{best}}))$ does not take very small values and the algorithm does not accept too many bad solutions.

4.3.2 Random Moves in Battery Configuration Space

We evaluate γ for different battery locations and capacities in the SA algorithm and then minimize $\log(\gamma)$. To move randomly in the solution configuration space, we randomly select two non-slack nodes (i, j) $\forall$ $i \in \mathcal{N}/\{1\}$ and $\forall$ $j(\neq i) \in \mathcal{N}/\{1\}$. Then exchange m_B blocks of battery unit ΔB between the two chosen nodes such that conditions in (4.15) are satisfied. Initially we start the algorithm by exchanging $m_B = N_B$ blocks of the battery unit and gradually reduce m_B till it is equal to 1. The gradual reduction of m_B depends on the decrease of γ. As the value of γ reduces by a factor of 10, m_B is decreased by Δm. This ensures that when the desired minimum is reached the system does not jump out of the minimum. The values of ΔB, N_B and Δm will be discussed in subsequent

sections.

4.3.3 Stopping Criterion for SA

We enforce three simultaneous *stopping criteria* for SA algorithm:

1. The number of iterations, n_{iter}, exceeds a pre-defined threshold value, $n_{\max}$, i.e., $n_{\text{iter}} \geq n_{\max}$.

2. The difference between n_{iter} and the number of solutions the algorithm has accepted, n_a, exceeds a maximum value, n_d, i.e., $n_{\text{iter}} - n_a \geq n_d$.

3. The improvements in γ have reached a desired minimum, ϵ, i.e., $\gamma_{\min} = \max_{n_m} |\gamma^{n_a} - \gamma^{n_a - n_m}| \leq \epsilon$, where γ^{n_a} is γ for the accepted iteration n_a.

If any of the stopping criteria is true the algorithm stops.

4.4 Importance function for splitting technique

In our model we are interested in estimating PLCV and then minimize the logarithm of PLCV. It is expected that PLCV will take very small values during minimization. The Crude Monte Carlo (CMC) estimations of these small probabilities will become computationally very expensive and thus we use the *splitting technique* for rare-event simulations. In particular we use the FNS variant of splitting proposed by [15]. For details on splitting technique see Section 3.3 of Chapter 3.

The efficiency of splitting is significantly determined by the importance function [68]. We take the maximum of the ratio of the absolute value of line currents and their respective maximum line current capacity as the importance function ϕ. This makes ϕ an increasing function in [0,1] and it is given by

$$\phi(I_{i,j}(t)) := \max_{(i,j)\in\mathcal{E}} \frac{|I_{i,j}(t)|}{I_{i,j}^{\max}}. \tag{4.18}$$

At time t, for any $(i,j) \in \mathcal{E}$ if $|I_{i,j}(t)| \geq I_{i,j}^{\max}$ implies $\phi(I_{i,j}(t)) \geq 1$, signifying that the rare event is hit, i.e., one of the line currents has exceeded its line capacity. $\phi \to 1$ corresponds to approaching the rare event set. A similar importance function was used by [165].

4.5 Simulation results

In this section we apply the SA algorithm and FNS to find the optimal storage position in a power grid by minimizing PLCV.

4.5.1 Simulation Parameters

We first discuss the different parameters chosen for our simulations. For the OU processes (4.13), we consider all the long term mean terms μ_i to be zero, which implies on an average at each non-slack node the power demand is compensated by the local power generation. The mean reverting terms $\theta_i = 1 + (i-1)/(N-2)$ for $i = 1, \dots N-1$ increases from 1 to 2 with i. The volatility terms σ_i are calculated from the long-term standard deviations of the OU process $std(P_i(t)) = \sigma_i/\sqrt{2\theta_i}$. The values of $std(P_i(t))$ will be discussed in the subsequent sections. Currently, we only consider uncorrelated OU processes for our assessment.

We perform our simulations for $T = 24$ hours and $\Delta t = 0.01$ hours. The initial state of the batteries are taken as $B_i(0) = B_i^{\max}/2 \; \forall i \in \mathcal{N}/\{1\}$, i.e, they are half-filled.

For performing FNS we first calculate the number of levels m by the pilot run such that $\tilde{p}_k$ is nearly equal to the optimal value of $\mathrm{p_{opt}} \approx 0.2032$ [15]. For the pilot run we use $R_k = 50$ for all k. For the final run we calculate R_k from (3.20) such that the $\mathrm{SRE}(\hat{\gamma}) \leq 0.03$. In order to obtain an accurate estimate of the probabilities, FNS is repeated $n = 30$ times (suggested by [67]) to calculate the mean of the estimator, $\hat{\bar{\gamma}}_n := \frac{1}{n}\sum_{i=1}^{n} \hat{\gamma}_i$. The squared relative error of the mean computed from the n samples is $\mathrm{SRE}(\hat{\bar{\gamma}}_n) := \frac{1}{n}\mathrm{SRE}(\hat{\gamma}_i)$.

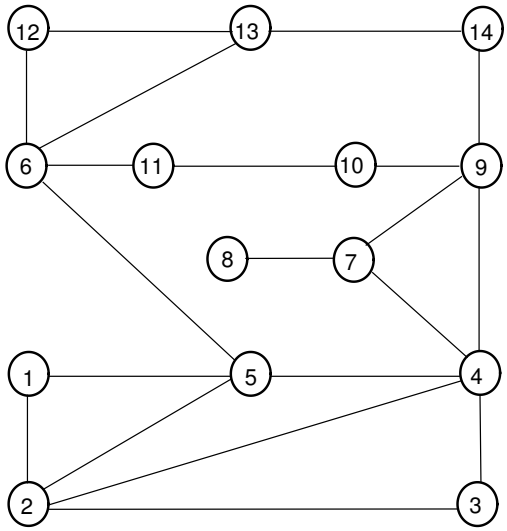


Figure 4.1: IEEE-14 bus line diagram showing the bus numbers and line connections. Bus 1 is the slack bus.

For SA we take initial temperature $T_c = 1$ and $\kappa = 0.99$. For the stopping criterion we take $n^{\max} = 1000, n_d = 300$ and $\epsilon = 10^{-7}$.

4.5.2 IEEE-14 Bus Test Case

In Figure 4.1 the line diagram of the IEEE-14 bus test case is shown. We test our algorithm for different scenarios for the IEEE-14 bus test case. The different scenarios being different maximum current carrying capacity of the lines $I_{i,j}^{max}\ \forall\ (i,j) \in \mathcal{E}$ and different standard deviations $std(P_i(t))$ of the OU processes of the non-slack nodes and total installation capacity of storage B^{tot}. We use the MATPOWER package [182] of MATLAB for the topological details (admittance matrix) of the IEEE-14 bus test case grid. However, the maximum line current carrying capacity is not set by the test case. We will discuss the values of the maximum line current carrying capacities in the subsequent sections.

4.5.2.1 Example 1

For this case the values of $I_{i,j}^{max}\ \forall\ (i,j) \in \mathcal{E}$ were obtained by simulating a long time-series, T=10^4 hours, for the system. The maximum value of line currents attained from the time-series run was taken to be the allowed maximum of each lines. For the $std(P_i(t))$ we use the net power injections at each non-slack node from MATPOWER, the values range from 1 to 95 p.u. (where p.u. is per unit value of the quantity of interest).

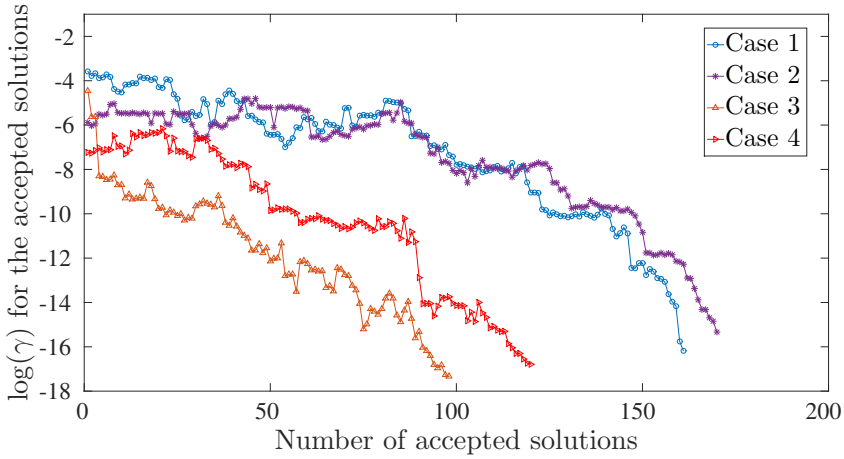
We start from different random initial configurations (cases 1-4) of the battery locations and capacity to minimize $\log(\gamma)$. The total installation capacity of storage $B^{\text{tot}} = 13000$ p.u. For the random movement of the algorithm in the battery configuration space (section 4.3.2) we take the battery unit $\Delta B = 100$ p.u., the initial number of blocks exchanged $N_B = 5$ and $\Delta m = 1$.

From Figure 4.2a we observe that γ has reduced by roughly a factor of e^{10}. In Figure 4.2b we compare the initial and final configurations of the battery position and capacities. It is observed that for all the cases in the final configurations about 35 percent of B^{tot} is placed at bus 3. One plausible reason for this final configuration can be the fact that $\frac{std(P_3(t))}{\sum_{i=2}^{14} std(P_i(t))} = 0.365$. We note that the final battery sizes at all the other nodes (apart from 3) are different for all the four cases and are not as consistent as node 3. Notwithstanding, for all four cases γ is reduced to very small values, see Figure 4.2a.

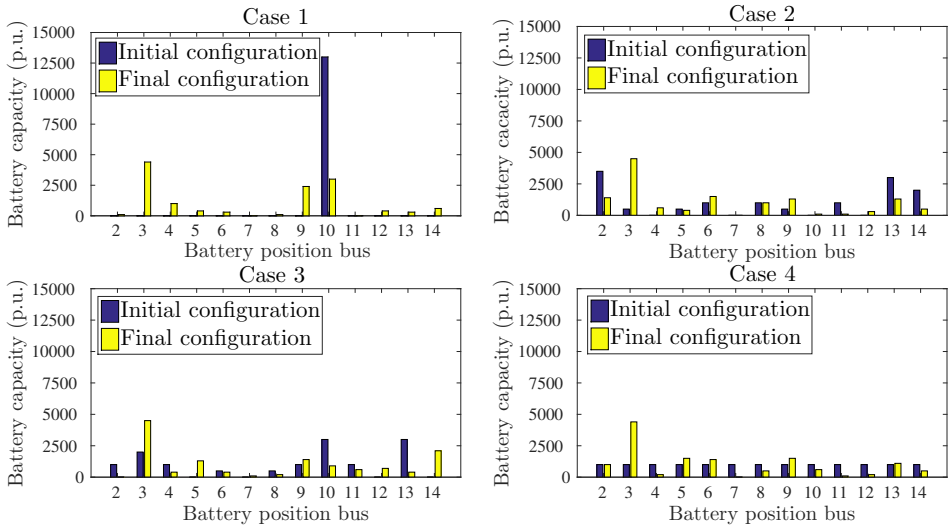
In Figure 4.3 we compare the accepted configuration solutions with the total configuration solutions the algorithm has searched for (from case 3).

4.5.2.2 Example 2

In this example, we take all the non-slack nodes to be similar , i.e., all the OU processes have same standard deviations $std(P_i(t)) = 10$ p.u. $\forall\ i \in \mathcal{N}/\{1\}$. Unlike Example 1 we randomize the $I_{i,j}^{max}$. To do this, we perform a long time series run, T=10^4 hours, for the system to calculate the maximum current flown through each lines, and multiply them with



(a) log(γ) versus the number of accepted solutions.



(b) Initial and the final configuration of the battery.

Figure 4.2: Figure 4.2a plots log(γ) versus the number of accepted solutions for different starting configuration of battery locations and capacities with $B^{\text{tot}} = 13000$ p.u. (Example 1). Figure 4.2b Compares the initial and the final configuration of the battery locations and capacities for four different initial states from Figure 4.2a.

uniform random numbers between $[0.5, 1]$ to obtain $I_{i,j}^{max}$. We take the battery unit $\Delta B = 12.5$ p.u., the initial number of blocks exchanged

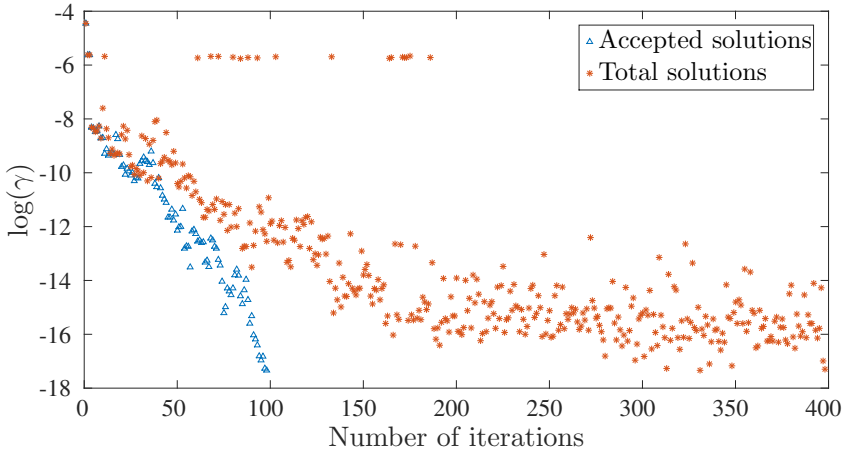


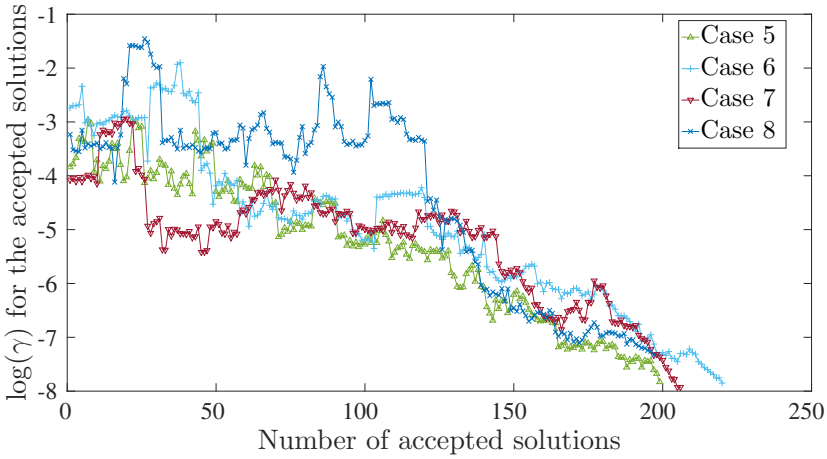
Figure 4.3: $\log(\gamma)$ versus the number of accepted solutions and total number to solutions the SA algorithm searched for (Case 3 of Example 1).

$N_B = 8$ and $\Delta m = m_B/2$ (see Section 4.3.2).

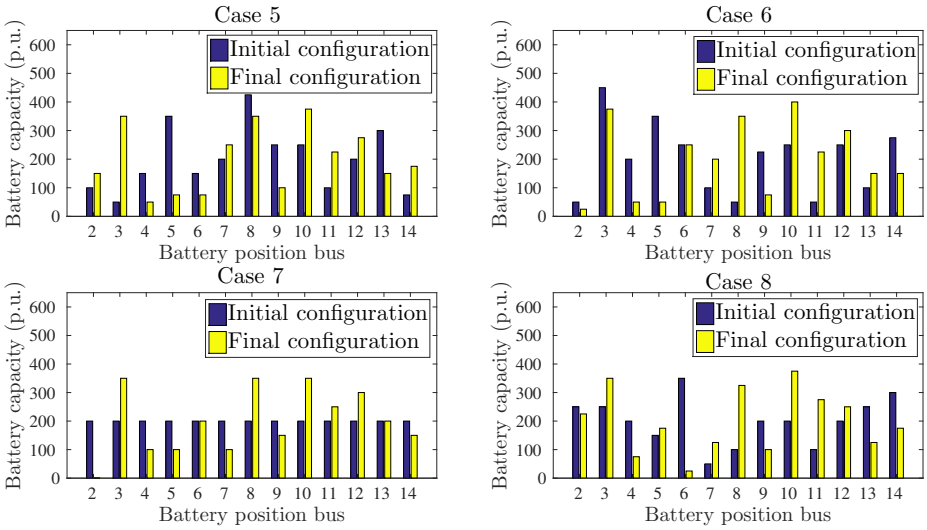
Figure 4.4a shows the minimization of $\log(\gamma)$ for four different initial configurations (Case 5-8). We find that γ is reduced by a factor of e^4. By comparing the initial and final configurations of the battery locations in Figure 4.4b, we find that buses 3, 8 and 10 require each around 15 percent of B^{tot} for all the cases for this minimization.

4.5.2.3 Example 3

In the example we make all the non-slack nodes and connection lines equivalent to study the effect of number of connections at nodes on the optimization. To do so, we take $std(P_i(t)) = 10$ p.u. $\forall\ i \in \mathcal{N}/\{1\}$ and $I_{i,j}^{max} = 50$ p.u. $\forall\ (i,j) \in \mathcal{E}$. We take the battery unit $\Delta B = 12.5$ p.u., the initial number of blocks exchanged $N_B = 8$ and $\Delta m = m_B/2$ (section 4.3.2). Figure 4.5a shows the optimization of $\log(\gamma)$ for various cases (Case 9 - 12), where each case represents a different starting configuration. The initial and final configurations of the battery placement is shown in Figure 4.5b. Notice for Case 11 when the batteries are placed equally at the non-slack buses the SA algorithm is not able to minimize $\log(\gamma)$ further. This



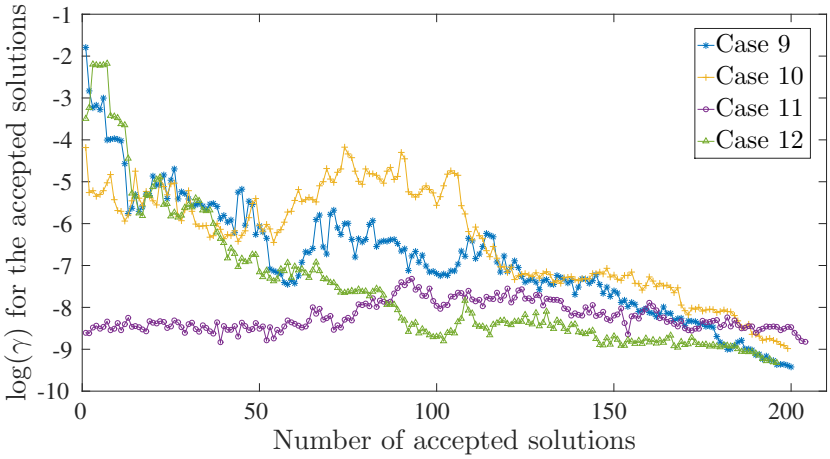
(a) log(γ) versus the number of accepted solutions.



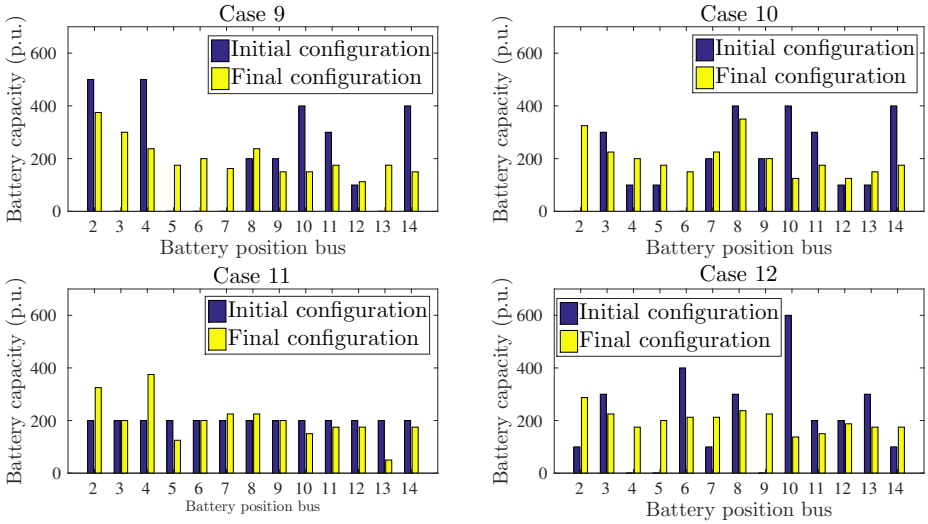
(b) Initial and the final configuration of the battery.

Figure 4.4: Figure 4.4a plots log(γ) versus the number of accepted solutions for different starting configuration of battery locations and capacities with $B^{\text{tot}} = 2600$ p.u. (Example 2). Figure 4.4b Compares the initial and the final configuration of the battery locations and capacities for four different initial states from Figure 4.4a.

hints towards the fact that equally placing the batteries at the non-slack buses is near optimal solution to the problem. Equal battery placement



(a) log(γ) versus the number of accepted solutions.



(b) Initial and the final configuration of the battery.

Figure 4.5: Figure 4.5a plots log(γ) versus the number of accepted solutions for different starting configuration of battery locations and capacities with $B^{\text{tot}} = 2600$ p.u. (Example 3). Figure 4.5b Compares the initial and the final configuration of the battery locations and capacities for four different initial states from Figure 4.5a.

being the near optimal solution shows that number of connections at nodes is not important for minimizing log(γ).

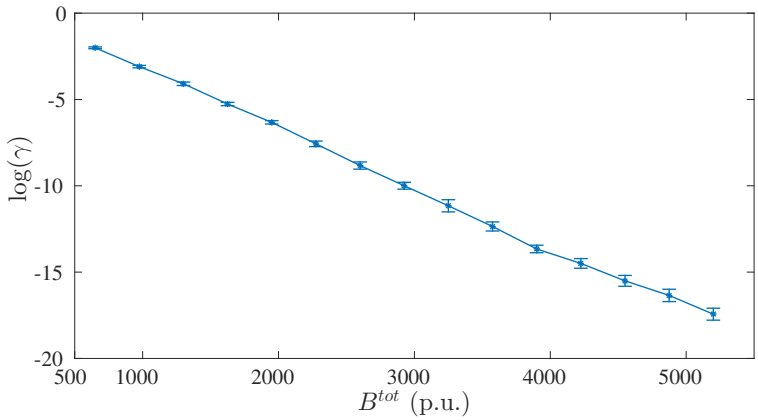


Figure 4.6: $\log(\gamma)$ versus B^{tot}. The error bars represents 95 percent confidence interval of $\log(\gamma)$.

We now investigate how $\log(\gamma)$ varies with B^{tot}. In order to do that we place the batteries equally at the non-slack buses. We do this because, Figure 4.5a suggests that equally placing the batteries at the non-slack buses is near the optimal solution of the problem. Figure 4.6 shows a linear relation between $\log(\gamma)$ and B^{tot}.

Finally we study the effect of B^{tot} on the final configuration of battery locations and capacities. We compare the final configurations for $B^{\text{tot}} =$ 1300 p.u. and 2600 p.u. For both the cases we start from equally placing the batteries at the non-slack buses and start the optimization from a lower temperature $T_c = 0.05$. We do this because equal battery placement is already near the optimal solution. We repeat the optimization 10 times for different B^{tot} and take an average for presenting the final configurations in Figure 4.7. Doubling B^{tot} does not affect the final configuration of the batteries, it almost doubles the size of batteries at the final configuration as shown in Figure 4.7.

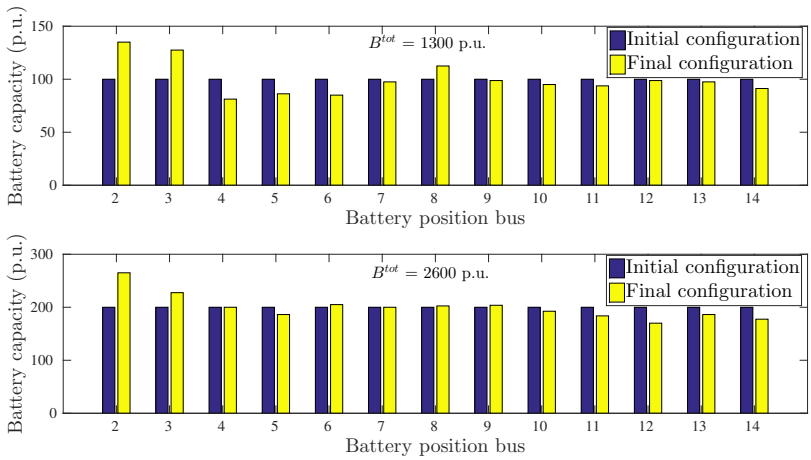


Figure 4.7: Comparing the initial and the final configuration of the battery locations and capacities for $B^{\text{tot}} = 1300$ p.u. and 2600 p.u.

4.6 Conclusion

For finding the optimal storage placement to enhance power grid reliability we use a novel combination of two computational techniques namely Simulated Annealing and the Splitting technique for rare-event simulation. To best of our knowledge this combination has not been used before. We use simulated annealing to minimize the reliability index, PLCV (γ), of the grid. We find that for very small values of γ, SA might not converge, however this problem disappears if we use $\log(\gamma)$ instead of γ as the cost function in SA. In order to calculate the small values of γ's we use FNS splitting technique for rare-event simulation.

We apply our method to the IEEE-14 bus grid for three different examples. In example 1 we have different nodal power injections and line current maxima. For this example we find that at the final configuration (after the minimization), bus 3 has 35 percent of B^{tot}. In example 2 we keep the nodal power injections to be similar and find that at the final configuration buses 3, 8 and 10 each get 15 percent of B^{tot}.

For example 3 we keep all the nodal power injections and the line

current maxima to be similar. For this we find that equal placement of the batteries at the non-slack nodes is the near optimal solution, which suggests that the number of connections to a node does affect the reliability of the power grid. We also find that the $\log(\gamma)$ decreases linearly with B^{tot}. Finally we examine if B^{tot} has an effect on final configuration of the battery locations and capacities. We observe that it does not affect the final configuration.

Chapter 5

Hidden Markov models for wind farm power output

The necessity to reduce carbon footprints has led to a need for integrating renewable energy sources like wind turbines and photo-voltaic arrays into the power grid. Being unpredictable in nature, these renewable energy sources inject power intermittently into the grid, thereby challenging the reliability of the power grid. The reliability of power systems that incorporate wind farms has been studied extensively (e.g. [58, 83, 92]). For maintaining a reliable power supply, as well as for future power system planning and daily system operation it is of great importance to study the statistical properties of power systems with intermittent, renewable power generation. In order to do so, it is important to develop stochastic methods for modeling power output, as stochastic models enable one to represent the intermittency and unpredictability of wind and solar power. These

models can be used for e.g. generating time series of power injections to estimate power grid reliability indices such as loss of load probability [95].

In this chapter, we focus on the power output of wind farms, in particular on the occurrence of high levels (corresponding to high quantiles) of wind farm power output. Wind farms are the largest source of intermittent renewable energy worldwide, and assessing the upper quantiles of their power output is relevant for e.g., planning, management, operation and grid reliability assessment [144]. Various approaches have been proposed to model wind farm power output, summarized briefly below. The stochastic models resulting from these approaches are useful for simulation, however their theoretical properties are complex and often partly unknown, making them less suitable for theoretical analysis. This was also reported in [120], in which the authors discuss the use of discrete-time Markov chains (DTMCs) to model the wind speed. For the above mentioned reasons, we propose, in this study, a class of stochastic models (discrete Markov chains) that are able to accurately capture the distribution of the power output, and are also simple enough to be amenable to analysis and to Monte Carlo simulation. In particular, the theoretical methods and techniques from the field of queuing systems (such as communication networks) provide a rich set of tools that can be employed for analyzing power systems when using these simple stochastic models [119, 149]. Such tools are useful for model-based risk and reliability studies.

For developing stochastic models to simulate wind power generation, two choices that must be faced are (i) whether to model the aggregate wind farm output, or model output from individual turbines, and (ii) model the power output directly, or model the wind speed and employ a separate model (e.g. a wind power curve (WPC)) to obtain power output from wind speed. The latter approach, mapping wind speed to the power output, is employed and investigated in e.g. [10, 33, 62, 74, 82, 87, 93, 137, 138, 178]. Such indirect models do not effectively capture the stochasticity of the

power output for a given wind speed [96]. In the context of our study, it must be noted that [76] and [120] applied discrete time Markov chains to model the wind speed.

The direct modeling of power output is considered by e.g., [50]. In that study, a limited autoregressive integrated moving average (LARIMA) model is used for the modeling of the individual wind turbine power, and a multivariate-ARIMA model is used for the modeling of the cross-correlations between the produced power for different parts of the wind farm. Furthermore, [126] and [139] use discrete-time Markov chains for modeling power output of an individual turbine, similar to what we propose in this study. These studies do not consider the aggregate power of a wind farm however.

Regarding the first choice, in order to model the joint or the aggregated power output of a wind farm, the output of individual wind turbines in a farm cannot be considered independent as the wind speed driving the turbines is correlated. This correlation has been taken into consideration in [35, 104, 108, 127, 171], where models for correlated wind speeds are combined with WPC models. These approaches, are oftentimes too complicated to be used for simulations, and for theoretical studies concerning the design and the planning of power grids. Another interesting approach that can be deployed for the short-term forecasting of the joint or the aggregated power output is that of deep neural networks (DNNs). In [94], DNNs are used to model wind speed and, in [129], DNN based on meta regression is used for wind power forecasting. DNNs require a trove of available data for training and an concrete understanding of the system at hand so as to carefully choose the type of the neural network. Moreover, DNNs have a very complicated structure which is not suitable for theoretical analysis. For all aforementioned reasons, in our study the focus is on the modeling of the power output of wind farms using simple stochastic models that can be used for simulations, and for theoretical studies concerning the design and planning of power grids.

In this study, we directly model the wind farm power output (cf. choice (ii)) using discrete Markov chains. In particular, we assess the accuracy of such simple stochastic models in fitting the upper quantiles $(90\%, 95\%, 99\%)$ of the observed power output of a wind farm. These models are directly developed from power output measurement data of individual turbines from an existing on-shore wind farm. We utilize time series data of the measured power output from six individual turbines in the same wind farm. Furthermore, we obtain our models through statistical inference, i.e. in a data-driven modeling. One advantage of inferring models directly from the measured power output is that they then reflect various factors that have an impact on the power generation: not only the variability of the wind speed, but also e.g., curtailing (deration) of wind turbines, shut-down of turbines for maintenance or other operational purposes, wear and tear of equipment, etc.

Large power injections from a wind farm into the transmission grid through a single grid connection point may lead to voltage or current overloads in the grid, jeopardizing its reliability. Simple tractable wind power generation models facilitate the reliability analysis of transmission grids [166]. Hence we focus on accurately modeling the upper quantiles $(90\%, 95\%, 99\%)$ of the power output of a wind farm.

Regarding the issue whether to model individual turbine output or aggregated power output (cf. choice (i)), we consider both approaches in this study. With the *microscopic approach* we model the individual wind turbine output while aiming to account for the dependencies (e.g. correlations) of output of different turbines. By summing their outputs at equal times, we can study the aggregate power output and its upper quantiles. This approach is flexible, as it is easy to consider the impact of adding more turbines or shutting down turbines (e.g. for maintenance). By contrast, with the *macroscopic approach* we directly model the aggregate wind power produced by multiple turbines (e.g. all turbines in the wind farm). It lacks the flexibility of the microscopic approach, however

there is only one scalar quantity (the aggregate power output) to model with the macroscopic approach instead of multiple quantities. In this approach we infer the model from the summed, measured output of the individual turbines. Clearly, this summed output automatically reflects the correlation between different turbines.

The stochastic models we use in our study are discrete-time Markov chains with finite state spaces. Wind power output is generally a continuous variable, therefore we have to discretize the state space to be able to use finite-state Markov chains. We do so by defining N discrete levels (or states) of power output and assigning a measurement in continuous space to its nearest-by discrete level. We investigate the effect of this discretization (with varying N) on the probability distribution of the aggregate power output, in particular on the upper quantiles of this distribution.

A Markov chain model is characterized by the matrix of transition probabilities between the N states (the stochastic matrix). The basic model we consider is a straightforward Markov chain with a single $N \times N$ stochastic matrix. We refer to this model as the *discrete-time Markov chain* (DTMC). In the microscopic approach, each turbine has its own DTMC. In the macroscopic approach, there is a single DTMC for the aggregate output. We also consider a more general model than the DTMC, a model which alternates over time between different stochastic matrices (in random fashion). This is the so-called *hidden Markov model* (HMM). The HMM has a hidden process (whose state is not observed) that is itself a finite-state Markov chain (say with M states). The observed process depends on the hidden process, in the sense that the transition probabilities of the N-state observed process depends on the state of the hidden process (thus, the observed process has M different stochastic matrices, each of size $N \times N$). One can view the HMM as a Markov-modulated Markov chain.

How to interpret the hidden process of an HMM (e.g. in physical terms) depends on the application. We hypothesize that in our case of

wind power generation, the hidden process represents primarily the meteorological conditions to which the wind turbines are subjected (see [66, 79] for an example of HMMs used in meteorology). These conditions are correlated in space and time, resulting in correlations between power output of geographically close turbines (e.g. turbines in the same wind farm). Modeling with an HMM, in which different turbines experience the same hidden state, enables one to account for these correlations. To the best of our knowledge, modeling with a single hidden process for multiple output processes in the HMM formalism has not been used before. In [86], one of the implemented approaches uses the Viterbi algorithm to estimate the parameters of a simple HMM from pseudo-measurements (with deterministic wind power generation and load patterns) and concludes based on simulations that this approach increases prediction accuracy. In contrast to the models under consideration in this chapter, the authors, in [86], mention that simple DTMC yields the least reliable results compared to the HMM.

The chapter is organized as follows. In Section 5.1, we discuss post processing of the original measurement data and the inherent correlation between the wind turbines in the farm. Section 5.2 describes discretization of the measurement data. In Section 5.3, we discuss the DTMC model and compare results with the measurement data. We discuss the HMM model and the expectation-maximization (EM) algorithm for estimating HMM parameters, and we compare results obtained from the HMM model with the measurement data. In Section 5.4 we present result of the macroscopic approach. We present some conclusions in Section 5.5. Finally, in the appendix, details of the EM algorithm are given.

5.1 Description of the data

For our analysis we use data from wind turbines from an on-shore wind farm in the Netherlands. The farm consists of 16 wind turbines, and we

have data from 6 of them for the period of April 1, 2016 to September 12, 2016. The wind turbines are of the type ETW Directwind 900/52 model. The data is obtained through the Supervisory Control and Data Acquisition (SCADA) of the wind turbines. For this wind farm, the measurements are stored as 20-minutes averages.

5.1.1 Post processing data

The data includes several variables besides wind power (turbine state, rotor speed, wind speed, wind direction, rotations per minutes, etc). However, we only use the wind power data. Furthermore, since we are interested in the aggregate output of all turbines together, we use only complete cases, i.e. only the time instances for which we have simultaneously data from all six wind turbines. This results in 8673 data points with aligned time stamps. We call this post-processed data the *measurement data*. We denote by P_t^w the t-th (time-ordered) datapoint of the power output of turbine w, with $w \in \mathcal{W} = \{1, \ldots, W\}$ and $W = 6$. The measurement data for turbine w is denoted $\boldsymbol{P}^w = [P_1^w, \ldots, P_T^w]$ with $T = 8673$. Note that T represents the number of complete cases.

5.1.2 Analysis of measurement data

We first study the individual wind turbine power data. Figure 5.1 shows histograms of the measurement data for all the turbines. Notice that $\boldsymbol{P}^6$ has the highest probability mass at the right tail. This is because turbine 6 is located at a corner of the farm next to the water, hence it experiences uninterrupted wind and produces more power compared to the other turbines.

The peak at 200 kW (see Figure 5.1) is caused by operational deration, i.e. a purposeful reduction of the power output of a wind turbine with the objective to prolong the remaining useful lifetime of the turbine. We include the data corresponding to operational deration instances, because

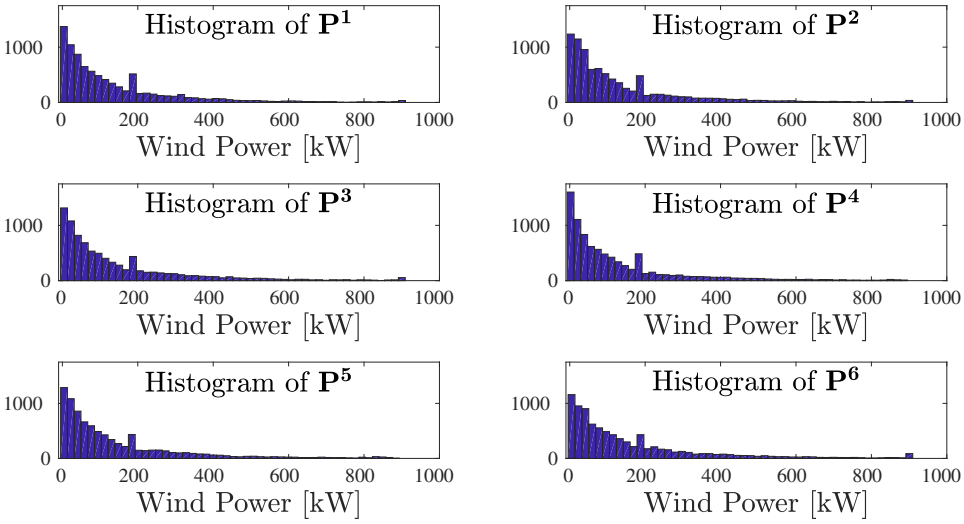


Figure 5.1: Histogram of wind power data for the six wind turbines.

w	1	2	3	4	5
Correlation coefficient	0.886	0.909	0.929	0.916	0.954

Table 5.1: Correlation between P_t^6 (power of turbine 6) and P_t^w, $w = \{1, \ldots, 5\}$ (power of turbines 1-5).

their inclusion makes the measured data more realistic (as deration is a common practice in the wind power industry) and renders the modeling more challenging.

It is expected that the power output of the different turbines (at equal times, i.e. zero time lag) is highly dependent. This is due to the geographic proximity of the turbines, resulting in the generation of power outputs being dependent due to similar environmental conditions. In Table 5.1, we present the correlation coefficients (linear dependency) for the power output between turbine 6 and the other turbines. Note that, as expected, these correlation coefficients are very high, but also that they are smaller than 1 and thus not fully correlated.

5.2 Discretized measurement data

We denote by $G_t = \sum_{w=1}^{6} P_t^w$ the total (aggregated) power produced by the six turbines together at time t, $t = 1, \ldots, T$. Let $\boldsymbol{G} = [G_1, \ldots, G_T]$ be the time series of G_t from the measurement data. The values in $\boldsymbol{G}$ lie in the interval $[-26.3, 5342.89]$ kW. Note that the negative values of $\boldsymbol{G}$ correspond to consumption of power by the wind turbines for performing internal functions like blade-pitch control, sensors, hydraulic breaks, etc.

In Section 5.2.1, we discuss the discretization of the support of $\boldsymbol{P}^w$. In order to differentiate the variables with the discretized (binned) support from the original variables with the continuous support, all notation will be adapted to have a tilde when referring to the discretized measurements, e.g., we denote by $\tilde{P}^w$ the discretized t-th measurement of power output of the w-th turbine. In the sequel, we sketch the discretization procedure and in Section 5.2.2, we compare the distributions as well as the 90%, 95% and 99% quantiles of the original data and the discretized.

5.2.1 Equidistant power levels and thresholds

To bin the support of $\boldsymbol{P}^w$ we define N levels, whose values are denoted by B_k, $k = \{1, \ldots, N\}$. In order to produce these values, we first consider the entire power support $[0, P_R]$, where P_R is the rated power of the turbines (the theoretical maximum power the turbines can produce). For the wind farm under consideration $P_R = 900$ kW. Secondly, we equidistantly divide $[0, P_R]$ in N levels. More concretely, this procedure produces $N-1$ thresholds, say $B_k^{\text{thres}} = \frac{B_k + B_{k+1}}{2}$, with $B_k = \left[\frac{k-1}{N-1}\right] P_R$, $k = \{1, \ldots, N\}$. Based on the above procedure, the measurement data are binned according to

$$\tilde{P}_t^w = \begin{cases} B_1, & \text{if } P_t^w < B_1^{\text{thres}}, \\ B_{k+1}, & \text{if } B_k^{\text{thres}} \leq P_t^w \leq B_{k+1}^{\text{thres}}, \\ & \text{for } k = 1, \ldots, N-2, \quad \text{and} \\ B_N, & \text{if } P_t^w > B_{N-1}^{\text{thres}}, \end{cases} \tag{5.1}$$

for $t = 1, \ldots, T$ and $w \in \mathcal{W}$. We use the same $\{B_k\}$ and $\{B_k^{\text{thres}}\}$ values for all turbines. We would like to note that depending on the modeling objective (here it is mainly to capture the tail distribution of the wind power output), one could consider other approaches for the discretization of the data besides equal width, see [98] for a recent survey in discretization techniques.

5.2.2 Comparing measurement power with binned measurement power

In this section, we analyze the effect of the discretization on the total power output distribution. We do so by comparing the distribution and the 90%, 95% and 99% quantiles of the continuous, $\boldsymbol{G}$, and discretized data, $\tilde{\boldsymbol{G}}$, varying the numbers of the bins, N.

5.2.2.1 Empirical distribution function (EDF) comparison

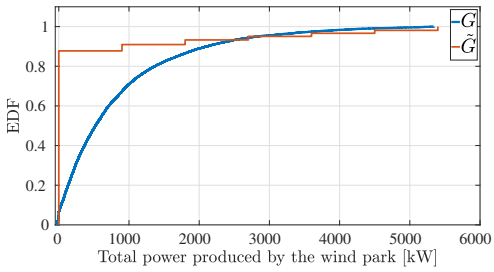
We compare the EDFs of $\boldsymbol{G}$ and $\tilde{\boldsymbol{G}}$. For a vector of data $\boldsymbol{Z} = [Z_1, \ldots, Z_T]$, the EDF is defined as $F(z) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}(Z_t \leq z)$. The EDF of $\tilde{\boldsymbol{G}}$ with $N = 2, 5, 100$ is shown in Fig. 5.2a–5.2c, together with the EDF of $\boldsymbol{G}$. Note that, for $N = 5$ (Fig. 5.2b), the right tails match well. It is to be noted that for $N = 5$ the EDFs do not match for smaller values of the support. However for large N, the EDFs match over the entire support (see Fig. 5.2c).

5.2.2.2 Quantile comparison

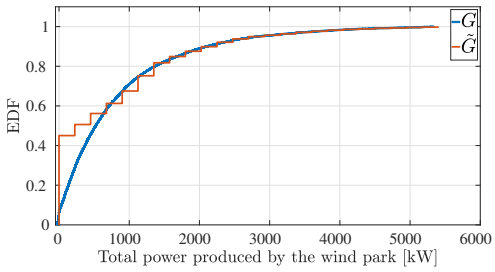
To compare the 90%, 95% and 99% quantiles of $\boldsymbol{G}$ and $\tilde{\boldsymbol{G}}$, we first compute the corresponding threshold values (denoted G^*) for the continuous data $\boldsymbol{G}$. We define

$$\gamma_P := \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}(G_t > G^*), \tag{5.2}$$

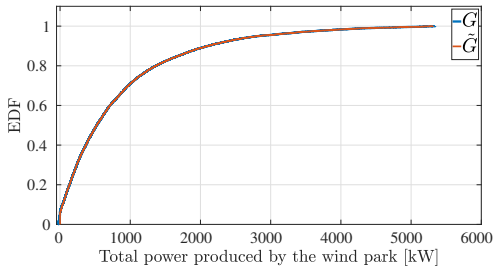
where $\mathbb{1}(\cdot)$ is the indicator function, and compute the value of G^* given



(a) $N = 2$



(b) $N = 5$



(c) $N = 100$

Figure 5.2: Comparing EDFs of the total power $\boldsymbol{G}$ and its binned counterpart $\tilde{\boldsymbol{G}}$, for different numbers of bins N.

γ_P. We find $G^* = 2210.6$ kW for $\gamma_P = 0.1$, $G^* = 2819.8$ kW for $\gamma_P = 0.05$ (95% quantile) and $G^* = 4411.3$ for $\gamma_P = 0.01$ (99% quantile).

Using the same values for G^*, we compute the fractions from the binned data $\tilde{\boldsymbol{G}}$ with N bins, i.e. we compute $\gamma_{\tilde{P}}^{(N)} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}(\tilde{G}_t > G^*)$. As expected, $\gamma_P = \lim_{N\to\infty} \gamma_{\tilde{P}}^{(N)}$, this is illustrated in Fig. 5.3. Furthermore, we find that for $N = 5$ the relative error, $RE = \frac{|\gamma_P - \gamma_{\tilde{P}}^{(N)}|}{\gamma_P} \times 100$ (in %), equals $1.3\%, 0.69\%$ and 11%, for the $90\%, 95\%$ and 99% quantiles, respectively. Hence, for studying the tail distribution (represented by the $90\%, 95\%$ and 99% quantiles) it is sufficient to choose $N = 5$, however, if the objective is to fit the entire distribution N should be chosen larger or the binning approach implemented in Section 5.2.1 should be appropriately changed.

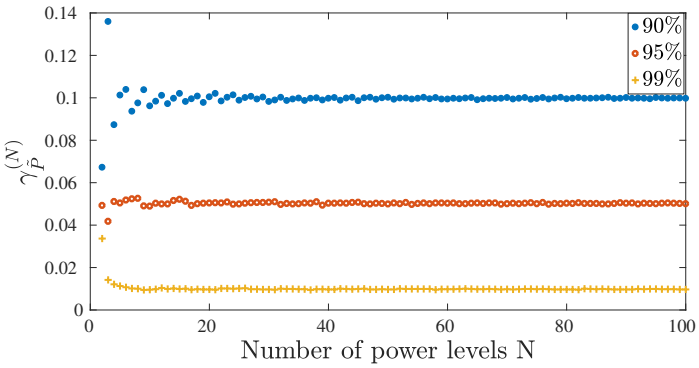


Figure 5.3: Comparing $\gamma_{\tilde{P}}^{(N)}$ for different quantile thresholds, G^*, with number of power levels N.

5.3 Discrete Markov models for the microscopic approach

In this section we model the individual wind turbine power output with simple Markov models that are discrete in both time and space. We consider two models, the Discrete Time Markov Chain (DTMC) model and the more general Hidden Markov Model (HMM). The model parameters are inferred from the binned measurement data for individual turbines, $\tilde{\boldsymbol{P}}^w$. We assess how well these models are capable of reproducing the distribution (in particular, the 90%, 95% and 99% quantiles) of $\tilde{\boldsymbol{G}}$.

We model the binned wind power of individual wind turbines as stochastic processes $\{Y_t^w, t = 1, \ldots, T\}$ with state space $\mathcal{N} = \{B_1, \ldots, B_N\}$ and $w \in \mathcal{W}$. We assume all $\mu, \nu \in \mathcal{N}$ and $w \in \mathcal{W}$ in the subsequent text unless otherwise mentioned. Also note that in this chapter all vectors are assumed to be row vectors.

5.3.1 Discrete time Markov chain model

In the microscopic DTMC framework, we model each wind turbine with an independent DTMC. This model has been used for modeling the wind

N	Quantile	$\bar{\gamma}^{(5)}_{\hat{P}_{\mathrm{MC}}} \pm \sigma(\gamma^{(5)}_{\hat{P}_{\mathrm{MC}}})$	RE (%)	$-\log \widehat{\mathcal{L}} \times 10^{-4}$	AIC $\times 10^{-4}$	BIC $\times 10^{-4}$
5	90%	$6.67 \times 10^{-3} \pm 4.23 \times 10^{-5}$	93.33	3.4465	6.9218	7.0236
	95%	$3.01 \times 10^{-4} \pm 7.12 \times 10^{-6}$	99.39			
	99%	$1.70 \times 10^{-8} \pm 2.14 \times 10^{-8}$	99.99			

Table 5.2: Comparing $\gamma^{(N)}_{\tilde{P}}$ and $\gamma^{(N)}_{\hat{P}_{\mathrm{MC}}}$ for different quantile thresholds, G^*, for $N = 5$.

power of an individual turbine in [126]. Let $L^w_{\mu\nu}$ be the one-step transition probability from B_μ to B_ν for wind turbine w. We estimate the elements of the stochastic matrix L^w for all the wind turbines $(w = 1, ..., W)$.

5.3.1.1 Maximum Likelihood Estimation for Markov chain

To estimate L^w we use the maximum likelihood estimator [123]. Given a series of observation Y^w_t for $t = 1, \ldots, T$, the elements of this estimator are

$$\widehat{L}^w_{\mu\nu} = \frac{\sum_{t=1}^{T-1} \mathbb{1}(Y^w_t = \mu)\mathbb{1}(Y^w_{t+1} = \nu)}{\sum_{t=1}^{T-1} \mathbb{1}(Y^w_t = \mu)}. \tag{5.3}$$

5.3.1.2 Quantile fraction and EDF comparison

After estimating the values of (L^w) for all six wind turbines we generate surrogate wind power timeseries, denoted $\hat{\boldsymbol{P}}^w_{\mathrm{MC}}$, from the DTMC model. Let $\hat{G}^{\mathrm{MC}}_t$ be the total power obtained by summing the surrogate timeseries, i.e. $\hat{G}^{\mathrm{MC}}_t = \sum_{w=1}^{6} \hat{P}^w_{t,\mathrm{MC}}$.

The fraction of time $\hat{G}^{\mathrm{MC}}_t$ is greater than the quantile thresholds, G^* (see Section 5.2.2.2) is given by $\gamma^{(N)}_{\hat{P}_{\mathrm{MC}}} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}(\hat{G}^{\mathrm{MC}}_t > G^*)$. We generate 100 realizations of $\hat{P}^w_{MC}$ with $T = 10^7$. From these realizations we compute the mean $\bar{\gamma}^{(N)}_{\hat{P}_{\mathrm{MC}}}$ and standard deviation $\sigma(\gamma^{(N)}_{\hat{P}_{\mathrm{MC}}})$.
In Table 5.2 we present results for $N = 5$. We find that $\gamma^{(5)}_{\hat{P}_{\mathrm{MC}}}$ highly underestimates $\gamma^{(5)}_{\tilde{P}}$. Even though the standard deviations are of the same order

as the means, the values are highly underestimated. We recall (see Table 5.1) that the wind power from the turbines are positively correlated to each other. The microscopic DTMC model is unable to capture this correlation, leading to strong underestimation of $\gamma^{(N)}_{\hat{P}_{\mathrm{MC}}}$ for different quantiles. We also tabulate the relative error of $\gamma^{(N)}_{\hat{P}_{\mathrm{MC}}}$ and γ_P, $\mathrm{RE} = \frac{|\gamma_P - \bar{\gamma}^{(N)}_{\hat{P}_{\mathrm{MC}}}|}{\gamma_P} \times 100$ (in %). Note that the RE is almost 100% for all the three quantiles.

In order to comment on the best model fit for the given set of measurement data we compare the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) statistic values. The AIC and the BIC value of a model is given by [44]

$$\begin{cases} \mathrm{AIC} = 2p - 2\log\widehat{\mathcal{L}}, \\ \mathrm{BIC} = p\log T - 2\log\widehat{\mathcal{L}}, \end{cases} \tag{5.4}$$

where p is the number of free parameters in the model, $\widehat{\mathcal{L}}$ is the maximized value of the likelihood function for the parameter values $\widehat{\lambda}$ for the data set $\boldsymbol{y}$, i.e., $\widehat{\mathcal{L}} = \mathbb{P}(\widehat{\lambda}|\boldsymbol{y})$, and T is the number of observations in data set $\boldsymbol{y}$. We will compare the AIC and BIC values of the DTMC model with the Hidden Markov model described in next section.

We also notice from Figure 5.4 that the EDFs of the total power produced by the wind farm from measurement data and binned measurement data are not comparable to DTMC model. The DTMC model is unable to capture the shape of the right tail of the EDFs.

5.3.2 Hidden Markov model

In the previous section, we found that a summation of independent DTMCs does not accurately capture the high quantiles of the total power produced by the wind farm. A main reason is that it neglects the dependency between the wind turbine power outputs (see, Table 5.1, for the linear dependency measurement). In this section we present a more general model, where the individual wind turbine power output is still discrete

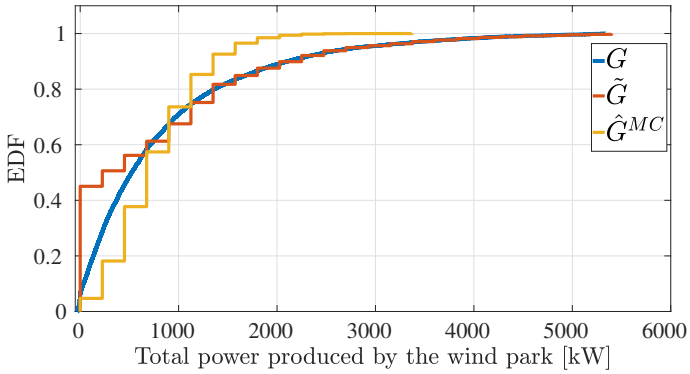


Figure 5.4: Comparing EDFs of the total power produced by the wind farm for data sets P^w, $\tilde{P}^w$ and $\hat{P}^w_{\text{MC}}$, $\forall w \in \mathcal{W}$ for $N = 5$.

and Markovian in nature but is also dependent on a hidden Markov process. Such a hidden Markov process captures the information regarding, e.g., meteorological conditions, that affect the individual wind turbines in the wind farm and in turn affect the total power output of the wind farm. This process is not fully known or observed, therefore it is referred to as hidden [131]. Note that such a model offers a flexible framework that can integrate different layers of complexity (by accounting for the joint state description of all the turbines in the park) and of information for the spatio-temporal dynamics (environmental and geographic conditions) captured in the hidden process, this surpasses the capabilities of copula modeling [127].

5.3.2.1 Individual stochastic modeling (matrix) per turbine

Let $\{Y_t^w, t = 1, 2, \ldots\}$, $w \in \mathcal{W}$, denote the binned wind power output of the individual wind turbines at time t, with state-space $\mathcal{N} = \{B_1, \ldots, B_N\}$. In the HMM framework we assume that the stochastic process $\{Y_t^w, t = 1, \ldots, T\}$ depends on a sequence of *hidden states*, $\{X_t, t = 1, \ldots, T\}$ with state space $\mathcal{M} = \{1, \ldots, M\}$. More specifically, the processes $\{Y_t^w, , t = 1, \ldots, T\}$ $\forall w \in \mathcal{W}$, are governed by the stochastic matrices $\boldsymbol{L}^{w,i} = \left(L_{\mu\nu}^{w,i}\right)_{\mu,\nu\in\mathcal{N}}$ whenever $\{X_t, t = 1, \ldots, T\}$ is in state i. The

transition probabilities of $\{X_t, t = 1, \dots, T\}$ are given by the stochastic matrix $\boldsymbol{A} = (A_{ij})_{i,j\in\mathcal{M}}$. Thus,

$$\begin{cases} L_{\mu\nu}^{w,i} = \mathbb{P}\big(Y_{t+1}^w = \nu | Y_t^w = \mu, X_{t+1} = i\big) \\ A_{ij} = \mathbb{P}\big(X_{t+1} = j | X_t = i\big). \end{cases} \tag{5.5}$$

For $t = 1$, the distribution of X_1 is given by $\pi_i = \mathbb{P}(X_1 = i)$ and those of Y_1^w by $\rho_\mu^{w,i} = \mathbb{P}(Y_1^w = \mu | X_1 = i)$. Note that if $M = 1$ the model simplifies to the DTMC model described in Section 5.3.1.

Our goal is to estimate the parameters of the above mentioned model, $\lambda = (\boldsymbol{\pi}, \boldsymbol{\rho}, \boldsymbol{A}, \boldsymbol{L})$ where $\boldsymbol{\pi} = (\pi_i)_{i\in\mathcal{M}}$, $\boldsymbol{A} = (A_{ij})_{i,j\in\mathcal{M}}$, $\boldsymbol{\rho} = \big(\rho_\mu^{w,i}\big)_{\mu\in\mathcal{N}}^{w\in\mathcal{W},i\in\mathcal{M}}$ and $\boldsymbol{L} = \big(L_{\mu\nu}^{w,i}\big)_{\mu,\nu\in\mathcal{N}}^{w\in\mathcal{W},i\in\mathcal{M}}$. In order to do so we will use the expectation-maximization algorithm for HMMs [54], see also [80].

5.3.2.2 Expectation-Maximization algorithm

The Expectation-Maximization (EM) algorithm is a general method for finding the maximum-likelihood parameter estimates of an underlying distribution given a data set which has incomplete values [38, 54, 71, 89]. Given the sequence of observations for all the wind turbine power outputs, $\boldsymbol{y} = [\boldsymbol{y}^1, \dots, \boldsymbol{y}^W]$ where $\boldsymbol{y}^w = [y_1^w, \dots, y_T^w] = [\tilde{P}_1^w, \dots, \tilde{P}_T^w]$, our goal is to estimate the parameters of the model, λ. In order to do so we will maximize the joint likelihood function for the model given the complete data set of the output processes and the hidden process.

Likelihood function For a given model λ, the joint likelihood function for the model given the complete data set is the joint probability distribution for the observation and the hidden state sequences

$$\mathcal{L}(\lambda | \boldsymbol{y}, \boldsymbol{x}) = \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X} = \boldsymbol{x} | \lambda). \tag{5.6}$$

An expression for $\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X} = \boldsymbol{x}|\lambda)$ is given in Appendix 5.A, see (5.16).

The EM algorithm for parameter estimation of HMM given a set of observations is known as the Baum-Welch (BW) algorithm [25]. The EM algorithm first finds the expected value of the complete data set log-likelihood with respect to the hidden data set $\boldsymbol{X} = \boldsymbol{x}$ given the observed data $\boldsymbol{Y} = \boldsymbol{y}$ and the current parameter estimates, in the *expectation-step*, i.e.,

$$\begin{aligned}\mathcal{Q}(\lambda, \lambda_k) &= \boldsymbol{E}\big[\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X} = \boldsymbol{x}|\lambda)|\boldsymbol{Y} = \boldsymbol{y}, \lambda_k\big] \\ &= \sum_{\boldsymbol{x}\in\mathcal{M}^T} \log(\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X} = \boldsymbol{x}|\lambda))\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X} = \boldsymbol{x}|\lambda_k), \quad (5.7)\end{aligned}$$

where λ_k is the current set of parameters estimates used to calculate the expectation $\mathcal{Q}$ and λ is a new set of parameters. A key element of the EM algorithm is to optimize λ in order to increase $\mathcal{Q}$. A detailed discussion of expression (5.7) is given in [37]. The *maximization-step* determines the next iterate λ_{k+1} by maximizing the expectation $\mathcal{Q}$, i.e.

$$\lambda_{k+1} = \underset{\lambda}{\operatorname{argmax}}\ \mathcal{Q}(\lambda, \lambda_k). \quad (5.8)$$

The maximization step guarantees that $\mathcal{L}(\lambda_{k+1}|\boldsymbol{y}, \boldsymbol{x}) \geq \mathcal{L}(\lambda_k|\boldsymbol{y}, \boldsymbol{x})$. The expectation and maximization steps are repeated until the desired convergence is reached. For literature on the convergence of the EM algorithm, see ([90, 169, 173]).

5.3.2.3 Parameter estimation

In this section we present expressions for the parameter estimates λ given the observed data $\boldsymbol{y}$ and the hidden sequence $\boldsymbol{x} = [x_1, \ldots, x_T]$. For convenience we will denote λ_k, the old parameter set, as λ' and the next iterate

of the parameter set λ_{k+1} as λ. The parameter estimates are given by

$$\begin{cases} \widehat{\pi}_i = \frac{\mathbb{P}(\boldsymbol{Y}=\boldsymbol{y}, X_1=i|\lambda')}{\mathbb{P}(\boldsymbol{Y}=\boldsymbol{y}|\lambda')}, \\ \widehat{A}_{ij} = \frac{\sum_{t=2}^{T} \mathbb{P}(\boldsymbol{Y}=\boldsymbol{y}, X_{t-1}=i, X_t=j|\lambda')}{\sum_{t=2}^{T} \mathbb{P}(\boldsymbol{Y}=\boldsymbol{y}, X_{t-1}=i|\lambda')}, \\ \widehat{\rho}_{\mu}^{w,i} = \mathbb{1}(Y_1^w = \mu), \\ \widehat{L}_{\mu\nu}^{w,i} = \frac{\sum_{t=2}^{T} \mathbb{1}(Y_{t-1}^w=\mu)\mathbb{1}(Y_t^w=\nu)\mathbb{P}(\boldsymbol{Y}=\boldsymbol{y}, X_t=i|\lambda')}{\sum_{t=2}^{T} \mathbb{1}(Y_{t-1}^w=\mu)\mathbb{P}(\boldsymbol{Y}=\boldsymbol{y}, X_t=i|\lambda')}. \end{cases} \tag{5.9}$$

A detailed derivation of the expressions of the expectation function and the parameter estimates can be found in Appendix 5.B.

5.3.2.4 Parameter estimates in terms of forward-backward equation

For calculating the estimates in (5.9) it is convenient to introduce the so-called forward backward variables $\alpha_i(t), \beta_i(t),\ \forall i \in \mathcal{M}$ ([37], [131]). For lack of space we have dropped the random variable Y terms in front of the y terms,

$$\begin{cases} \alpha_i(t) = \mathbb{P}(y_1^1, \ldots, y_t^1, \ldots, y_1^W, \ldots, y_t^W, X_t = i|\lambda'), \\ \beta_i(t) = \mathbb{P}(y_{t+1}^1, \ldots, y_T^1, \ldots, y_{t+1}^W, \ldots, y_T^W | X_t = i, y_t^1, \ldots, y_t^w, \lambda'). \end{cases} \tag{5.10}$$

These variables are computed recursively (see Appendix 5.C) and the numerical effort grows linearly in T. Using (5.24) and (5.25), the expressions in (5.9) in terms of forward-backward equations becomes,

$$\begin{cases} \widehat{\pi}_i = \frac{\alpha_i(1)\beta_i(1)}{\sum_{j=1}^{M}\alpha_i(1)\beta_i(1)}, \\ \widehat{A}_{ij} = \frac{\sum_{t=2}^{T}\alpha_i(t-1)\widehat{A}'_{ij}\Big[\prod_{w=1}^{W}\widehat{L}'^{w}_{j,y^w_{t-1}y^w_t}\Big]\beta_j(t)}{\sum_{t=2}^{T}\alpha_i(t-1)\beta_i(t-1)}, \\ \widehat{L}^{w,i}_{\mu\nu} = \frac{\sum_{t=2}^{T}\alpha_i(t)\beta_i(t)\mathbb{1}(Y^w_{t-1}=\mu)\mathbb{1}(Y^w_t=\nu)}{\sum_{t=2}^{T}\alpha_i(t)\beta_i(t)\mathbb{1}(Y^w_{t-1}=\mu)}. \end{cases} \tag{5.11}$$

The detailed derivations of the terms are given in Appendix 5.D.

Re-scaling forward backward equations From (5.10) we see that as t increases the values of $\alpha_i(t)$ and $\beta_i(t)$ become very small. Hence, the terms in (5.11) diverge when computed numerically. To avoid this numerical problem we normalize the values of the forward and backward equations

$$\begin{cases} \bar{\alpha}_i(t) = \frac{\alpha_i(t)}{\sum_{i'=1}^{M}\alpha_{i'}(t)}, \\ \bar{\beta}_i(t) = \frac{\beta_i(t)\cdot\sum_{i'=1}^{M}\alpha_{i'}(t)}{\sum_{i'=1}^{M}\alpha_{i'}(T)}. \end{cases} \tag{5.12}$$

5.3.2.5 Stopping criterion for the EM algorithm

We enforce two simultaneous *stopping criteria* for the EM algorithm:

1. The number of iterations, n_i, exceeds a predefined threshold value, $n_{\max}$, i.e., $n_i \geq n_{\max}$.
2. The improvements in λ have reached a desired minimum, ϵ, i.e., $\Delta\lambda_{\min} = \max_{n_m} |\lambda^{n_i} - \lambda^{n_i - n_m}| \leq \epsilon$, where λ^{n_i} is λ for the iteration n_i.

If any one of the stopping criteria is true the algorithm stops.

5.3.2.6 Quantile and EDF comparison

Given the binned measurement data ($\tilde{\boldsymbol{P}}^w$) and a choice for the number of hidden states (M), we estimate the parameters of the HMM using

M	Quantile	$\bar{\gamma}^{(5,M)}_{\hat{P}_{\text{HMM}}} \pm \sigma(\gamma^{(5,M)}_{\hat{P}_{\text{HMM}}})$	RE (%)	$-\log\widehat{\mathcal{L}} \times 10^{-4}$	AIC $\times 10^{-4}$	BIC $\times 10^{-4}$	p
7	90%	0.0993 ± 0.0030	0.6857	2.4534	5.1179	5.643	1056
	95%	0.0468 ± 0.0020	6.5061				
	99%	0.0045 ± 0.0004	55.421				
8	90%	0.1013 ± 0.0036	1.3336	2.4193	5.8016	5.9404	1215
	95%	0.0547 ± 0.0023	9.3667				
	99%	0.0082 ± 0.0005	18.232				
9	90%	0.1011 ± 0.0032	1.1512	2.4011	5.0776	6.0500	1376
	95%	0.0517 ± 0.0020	3.2605				
	99%	0.0114 ± 0.0060	13.127				
10	90%	0.1268 ± 0.0050	26.887	2.4117	5.1312	6.2190	1539
	95%	0.0633 ± 0.0028	26.496				
	99%	0.0076 ± 0.0005	24.654				

Table 5.3: Comparing $\gamma^{(N,M)}_{\hat{P}_{\text{HMM}}}$ for different quantile thresholds, G^*, with γ_P for $N = 5$ for different number of hidden states M.

(5.11). We generate surrogate wind power output, denoted $\hat{\boldsymbol{P}}^w_{\text{HMM}}$, from the HMM. The corresponding total power surrogate timeseries is $\hat{G}^{\text{HMM}}_t := \sum_{w=1}^{6} \hat{P}^w_{t,\text{HMM}}$. We calculate the fraction of time $\hat{G}^{\text{HMM}}_t$ is greater than the quantile thresholds, G^* (see Section 5.2.2),

$$\gamma^{(N,M)}_{\hat{P}_{\text{HMM}}} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}(\hat{G}^{\text{HMM}}_t > G^*). \tag{5.13}$$

After estimating the HMM parameters we generate 100 independent realizations of $\hat{P}^w_{\text{HMM}}$ of length $T = 10^5$, and compute the fraction $\gamma^{(N,M)}_{\hat{P}_{\text{HMM}}}$ for each of these realizations. The mean and standard deviation of these 100 fractions are denoted $\bar{\gamma}^{(N,M)}_{\hat{P}_{\text{HMM}}}$ and $\sigma(\gamma^{(N,M)}_{\hat{P}_{\text{HMM}}})$ respectively. For the stopping criterion of the EM algorithm (see Section 5.3.2.5) we take $n_{\max} = 10^4, n_m = 100$ and $\epsilon = 10^{-6}$.
In Table 5.3 we compare $\bar{\gamma}^{(N,M)}_{\hat{P}_{\text{HMM}}}$ for different quantile thresholds, G^*, with γ_P for $N = 5$ and different number of hidden states M. We tabulate the relative error of $\bar{\gamma}^{(N,M)}_{\hat{P}_{\text{HMM}}}$ and γ_P, RE $= \frac{|\gamma_P - \bar{\gamma}^{(N,M)}_{\hat{P}_{\text{HMM}}}|}{\gamma_P} \times 100$ (in %). Note that for $M = 9$ the RE is smallest for all the three quantiles.

From Table 5.3 we observe that the smallest value of $-\log\widehat{\mathcal{L}}$ and AIC is for $M = 9$. However the BIC value grows with M as the penalty term for the BIC test is higher than the AIC test (see (5.4)). However by

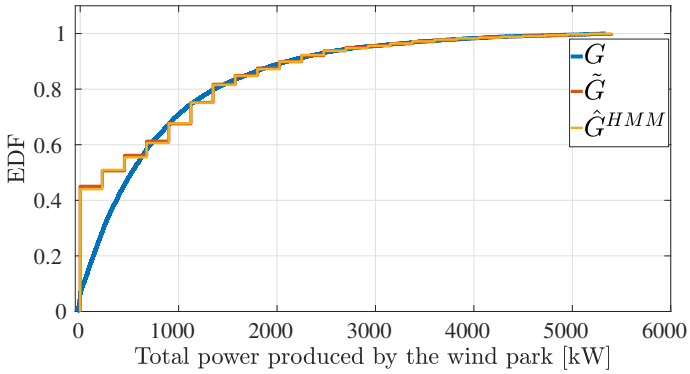


Figure 5.5: Comparing EDFs of the total power produced by the wind farm for data sets P^w, $\tilde{P}^w$ and $\hat{P}^w_{\text{HMM}}$, $\forall w \in \mathcal{W}$ for $N = 5, M = 9$.

comparing the quantile fraction values, $-\log \widehat{\mathcal{L}}$ and the AIC test we can conclude that for $N = 5$, the optimal number of hidden states is $M = 9$.

Also note from Table 5.2 that $-\log \widehat{\mathcal{L}}$, AIC and BIC values are much higher for the DTMC model compared to the HMM model. The DTMC model is a special case of the HMM model with $M = 1$.

In Figure 5.5 we compare the EDFs of the total power from the measurement data $\boldsymbol{G}$, the discretized measurement data $\tilde{\boldsymbol{G}}$ and the HMM timeseries $\hat{\boldsymbol{G}}^{\text{HMM}}$, for $N = 5$ and $M = 9$. It can be seen that the HMM model is able to reproduce the right tail of the EDF. This is consistent with table 5.3, which shows that the $90\%, 95\%$ and 99% quantiles are correctly reproduced by the HMM with $N = 5$ and $M = 9$.
We would like to note, that if the objective would be to fit the entire distribution without drastically increasing the values of N and M, then we propose to use a moment matching based fitting method, see, e.g., [81] and the references therein.

5.3.3 Same stochastic modeling (matrix) per turbine

In (5.5) we define separate stochastic matrices for each wind turbine (indexed by w). For this model set up the number of parameters is

$p = M^2 + MWN^2 - MN - 1$. For example, for $N = 5, M = 9$ and $W = 6$, the number of parameters to be estimated is $p = 1376$; which is a rather high number.

In this section we consider a model set up in which the same (set of) stochastic matrices is used for each of the wind turbines. Thus, the transition probability matrix of each power output process $\{Y_t{}^w\}$ depends on the hidden state but not on the turbine index w, i.e., the matrix $L^{w,i}$ in (5.5) is now independent of w. This reduces the number of parameters of the model. We study the trade-off between reducing the number of parameters of the model with $-\log\widehat{\mathcal{L}}$, AIC and BIC values and compare them with values in Table 5.3. Following the definition of the HMM model in Section 5.3.2.1 we assume that the individual wind turbine power output processes $\{Y_t{}^w\}$, are governed by matrices $(\boldsymbol{L}^i)$ and $\boldsymbol{A}$ as follows:

$$\begin{cases} L^i_{\mu\nu} = \mathbb{P}\big(Y^w_{t+1} = \nu | Y^w_t = \mu, X_{t+1} = i\big) \\ A_{ij} = \mathbb{P}\big(X_{t+1} = j | X_t = i\big). \end{cases} \tag{5.14}$$

The initial distribution of Y^w_1 is given by $\rho^i_\mu = \mathbb{P}(Y^w_1 = \mu | X_1 = i)$. The details of the parameter estimates are given in Appendix 5.E.

The number of parameters to be estimated for this model are $p = M^2 + MN^2 - M - 1$. We follow the same procedure as Section 5.3.2.6 to estimate the quantile fractions, denoted $\gamma'^{(N,M)}_{\hat{P}_{\text{HMM}}}$ for this model. The mean and the standard deviations of the 100 estimates of $\gamma'^{(N,M)}_{\hat{P}_{\text{HMM}}}$ are given by $\bar{\gamma}'^{(N,M)}_{\hat{P}_{\text{HMM}}}$ and $\sigma(\gamma'^{(N,M)}_{\hat{P}_{\text{HMM}}})$ respectively. From Table 5.4 we find that for $M = 9$ we obtain the smallest value of $-\log\widehat{L}$ and AIC. We also get good estimates of the quantile fractions for $M = 9$.

Comparing the values of $-\log\widehat{\mathcal{L}}$, AIC and BIC in Table 5.3 and Table 5.4 we find that $-\log\widehat{\mathcal{L}}$ is slightly lower for the former. This implies having different stochastic matrices (as in (5.5)) for different wind turbines fits the given data better than having the same stochastic matrices (as in (5.14)). However for the model described in (5.5) we have more parameters to

M	Quantile	$\bar{\gamma'}^{(5,M)}_{\hat{P}_{\text{HMM}}} \pm \sigma(\gamma'^{(5,M)}_{\hat{P}_{\text{HMM}}})$	RE (%)	$-\log\widehat{\mathcal{L}} \times 10^{-4}$	AIC $\times 10^{-4}$	BIC $\times 10^{-4}$	p
7	90%	0.1009 ± 0.0031	0.940	2.4936	5.0304	5.1830	216
	95%	0.0497 ± 0.0018	0.689				
	99%	0.0075 ± 0.0006	25.24				
8	90%	0.1025 ± 0.0029	2.532	2.4766	5.0041	5.1844	255
	95%	0.0532 ± 0.0019	6.369				
	99%	0.0126 ± 0.0009	25.45				
9	90%	0.0987 ± 0.0032	1.243	2.4724	5.0040	5.2132	296
	95%	0.0517 ± 0.0020	3.391				
	99%	0.0108 ± 0.0060	8.116				
10	90%	0.0984 ± 0.0043	1.562	2.4729	5.0038	5.2087	339
	95%	0.0522 ± 0.0027	4.313				
	99%	0.0076 ± 0.0008	1.728				

Table 5.4: Comparing $\gamma'^{(N,M)}_{\hat{P}_{\text{HMM}}}$ for different quantile thresholds, G^*, with γ_P for $N = 5$ for different number of hidden states M along with $-\log\widehat{\mathcal{L}}$, AIC and BIC values.

estimate compared to the model in (5.14). For $N = 5, M = 9$ we have 1376 parameters for (5.5) and only 296 parameters to estimate for (5.14). We find that the AIC and the BIC values in Table 5.4 for model (5.14) are lower than in Table 5.3 for the model described in (5.5).

5.4 Results of the Macroscopic approach

In the microscopic approach studied in Section 5.3, we modeled the individual wind turbine power output processes with DTMC and HMM models, with the aim to model the high quantiles of the distribution of the total power produced by all turbines together ($\boldsymbol{G}$). As discussed in the introduction, a different approach is to model $\boldsymbol{G}$ directly. Here, the measurement data of $\boldsymbol{G}$ is discretized (see first paragraph of Section 5.2), and modeled with DTMC and HMM models. We refer to it as the macroscopic approach, and study it in this section.

We bin $\boldsymbol{G}$ in a similar way as we bin the individual wind turbine power output (see Section 5.2.1). Let N be the total number of power bins the total power output of the wind farm can produce power. We divide the bins in an equidistant way between $[0, 6 \times R_P]$, we do this because we have

data from 6 wind turbines in the wind farm. We follow the same procedure as (5.1) for G_t for $t = 1, \ldots, T$ and obtain the binned measurement total power produced by the wind farm data series $\hat{\boldsymbol{G}}$. Note that $\hat{\boldsymbol{G}}$ is different from $\tilde{\boldsymbol{G}}$ in Section 5.2 because the value of the power bins are different in this case.

For notational clarity we define new quantile fractions given the quantile thresholds G^*,

$$\begin{cases} \gamma_G = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}(G_t > G^*), \\ \gamma_{\hat{G}}^{(N)} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}(\hat{G}_t > G^*). \end{cases} \tag{5.15}$$

G^* are the same quantile thresholds set at 2210.6, 2819.8 and 4411.3 kW for 90%, 95% and 99% quantile respectively as in Section 5.2.2.2. Note that $\gamma_G = \gamma_P$ from Section 5.2.2.2. We find that for $N = 15$ the relative error (RE) for the quantile estimates between γ_G and $\gamma_{\hat{G}}^{(15)}$ are 0.9227%, 5.7604% and 4.5977% respectively for the quantiles 90%, 95% and 99% respectively. Hence we will set the number of bins required to capture the high quantiles of the distribution of the total power produced by the binned measurement data at 15 bins, i.e., $N = 15$.

5.4.1 Quantile fraction comparison

Given the observation of the binned measurement total power produced by the wind farm data $\hat{\boldsymbol{G}} = [\hat{G}_1, \ldots, \hat{G}_T]$ we first estimate the DTMC and the HMM model parameters. We generate surrogate data series of the total power produced by the wind farm, $\bar{\boldsymbol{G}}$ from the model parameters. Let $\gamma_{\bar{G}}^{(N,M)} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}(\bar{G}_t > G^*)$ be the fraction of time the total power produced by the wind farm obtained from the Markov models be greater than G^* threshold. Note that when $M = 1$ we have the DTMC model and for $M \geq 2$ we have the HMM model. We generate 100 realizations of $\bar{\boldsymbol{G}}$ of length $T = 10^4$. Then we calculate the mean and the standard deviations of $\gamma_{\bar{G}}^{(N,M)}$ from the 100 realizations, $\bar{\gamma}_{\bar{G}}^{(N,M)}$ and $\sigma(\gamma_{\bar{G}}^{(N,M)})$ respectively.

In Table 5.5 we compare the quantile fraction values for different number of hidden states M. We also tabulate the relative error $RE = \frac{|\bar{\gamma}_G^{(N,M)} - \gamma_G|}{\gamma_G} \times 100$ (in %). In order to find the best model fit we also compare the values of $-\log \widehat{\mathcal{L}}$, AIC and BIC. Note that the DTMC model, i.e. $M = 1$, can capture the quantile fractions very well, however the values of $-\log \widehat{\mathcal{L}}$, AIC and BIC are much higher compared to the $M \geq 2$ cases. DTMC is a very robust model for capturing the invariant distribution however it fails to capture the time correlation in the data set which is captured by the HMM model ($M \geq 2$). For $M = 5$ we have the lowest value for $-\log \widehat{\mathcal{L}}$. However we find that the AIC and the BIC values increase with M for $M \geq 2$. We do not see a minimum in the AIC value for this this case as seen in Table 5.3 and Table 5.4 for $M = 9$.

Recall from the model in Section 5.3.3 for the best model fit there were 296 parameters to estimate, whereas here the number of parameters is larger when $M \geq 2$, see Table 5.5. Furthermore, note that we cannot compare the $-\log \widehat{\mathcal{L}}$, AIC and BIC values of the model in Table 5.5 with the models in Table 5.3 and Table 5.4 because the former uses the observation data set $\hat{\boldsymbol{G}}$ (directly discretizing $\boldsymbol{G}$) and the later two uses $\tilde{\boldsymbol{P}}^w$ (obtained from discretizing the individual wind turbine power output).

5.5 Conclusions

In this chapter we studied how discrete Markov chain models can be used to model the power output of a wind farm. Such models have a rather simple structure, making them suitable for theoretical analysis. Despite their simplicity, they are able to reproduce the distribution of the wind farm power output, as demonstrated in this study. Of particular interest are the upper quantiles $(90\%, 95\%, 99\%)$, which can be accurately captured with the Markov chain models studied here.

We considered two approaches. In the microscopic approach, described in Section 5.3, the power output of individual turbines was modeled with

M	Quantile	$\bar{\gamma}_{\vec{G}}^{(15,M)} \pm \sigma(\gamma_{\vec{G}}^{(N,M)})$	RE (%)	$-\log \widehat{\mathcal{L}} \times 10^{-4}$	AIC $\times 10^{-5}$	BIC $\times 10^{-5}$	p
1	90%	0.0984 ± 0.0042	1.569	6.2738	1.2590	1.2738	210
	95%	0.0469 ± 0.0025	6.302				
	99%	0.0096 ± 0.0008	4.216				
2	90%	0.1067 ± 0.0101	6.717	1.0129	0.2116	0.2435	451
	95%	0.0509 ± 0.0060	1.788				
	99%	0.0110 ± 0.0024	9.918				
3	90%	0.1055 ± 0.0111	5.572	0.9983	0.2132	0.2613	680
	95%	0.0512 ± 0.0073	2.409				
	99%	0.0107 ± 0.0021	6.488				
4	90%	0.1173 ± 0.0128	17.37	0.9852	0.2152	0.2796	911
	95%	0.0567 ± 0.0076	13.34				
	99%	0.0101 ± 0.0021	0.487				
5	90%	0.0988 ± 0.0102	1.154	0.9761	0.2181	0.2989	1144
	95%	0.0475 ± 0.0057	5.095				
	99%	0.0101 ± 0.0022	0.507				
6	90%	0.1004 ± 0.0113	0.387	0.9793	0.2234	0.3209	1379
	95%	0.0471 ± 0.0065	5.850				
	99%	0.0096 ± 0.0020	4.597				
7	90%	0.1026 ± 0.0099	2.594	0.9708	0.2265	0.3407	1616
	95%	0.0505 ± 0.0063	0.866				
	99%	0.0113 ± 0.0025	13.07				

Table 5.5: Comparing $\gamma_{\vec{G}}^{(N,M)}$ for different quantile thresholds, G^*, with γ_P for $N = 15$ for different number of hidden states M along with $-\log \widehat{\mathcal{L}}$, AIC, BIC and number of parameters of the model, p.

Markov chains, and subsequently summed to obtain the total power output. With the most straightforward type of Markov chain model (the DTMC model, see Section 5.3.1), the correlation between different turbines could not be taken into account, leading to strong underestimation of the upper quantiles. With the more general HMM (Section 5.3.2) we were able to account for these correlations, resulting in a faithful reproduction of the upper quantiles. Both the HMM and DTMC models were inferred from timeseries data of the power output of wind turbines in the same wind farm.

With the macroscopic approach, we directly modeled the aggregate (summed) output of all turbines, see Section 5.4. In this approach, both the DTMC model and the HMM were able to capture the upper quantiles well. However, the HMM performed much better in terms of the log-likelihood of the estimated models.

Besides the log-likelihood of the estimated models, we also computed the Akaike and Bayesian information criteria (AIC and BIC), to take into account the number of model parameters (and, for BIC, amount of data) when judging the quality of fitted models. In the microscopic approach, the HMM in which all turbines had the same set of stochastic matrices (Section 5.3.3) performed most favorable in terms of AIC and BIC. This is due to the smaller number of parameters needed in these models compared to the HMM with different sets of stochastic matrices for the different turbines. The HMM from the macroscopic approach also required more parameters than the HMM from the microscopic approach with the same set of stochastic matrices (as described in Section 5.4).

The microscopic approach has the advantage of flexibility, as it is easy to add or leave out individual turbines in the analysis of the total wind farm power output (e.g. to account for temporarily shutting down turbines for maintenance). The macroscopic approach is more straightforward to apply, because there is only a single scalar quantity to model, rather than multiple quantities (one for each considered turbine) as in the microscopic approach. However, the macroscopic approach required more parameters (as discussed above), and lacks the flexibility of the microscopic approach.

In the microscopic approach, the DTMC model was not able to account for correlations in the output of different turbines. However, we were able to capture these correlations with the HMM due to the inclusion of the hidden process in these models. The hidden process can reflect, for example, the meteorological conditions that the turbines experience, conditions that are spatially correlated and that can lead to correlations between power output of different turbines.

In this study we have shown that simple discrete Markov chain models are able to reproduce the distribution of wind farm power output, in particular its upper quantiles. Due to their simple structure, these models are amenable for theoretical analysis using methods from e.g. the field of queueing theory. We will report on such analysis in the following chapter.

Appendix

5.A EM algorithm

Likelihood function For a given model $\lambda = (\boldsymbol{\pi}, \boldsymbol{\rho}, \boldsymbol{A}, \boldsymbol{L})$, the joint probability distribution for the observation and the hidden state sequences is given by

$$\begin{aligned}
\mathbb{P}(\mathbf{Y} = \mathbf{y}, \mathbf{X} = \mathbf{x}) &= \mathbb{P}(Y_1^1 = y_1^1, \ldots, Y_T^1 = y_T^1, \ldots, Y_1^W = y_1^W, \ldots, Y_T^W = y_T^W, \\
&\qquad X_1 = x_1, \ldots, X_T = x_T) \\
&= \mathbb{P}(Y_1^1 = y_1^1, \ldots, Y_1^W = y_1^W, X_1 = x_1) \cdot \prod_{t=2}^{T} \mathbb{P}(Y_t^1 = y_t^1, \ldots, Y_t^W = y_t^W, \\
&\qquad X_t = x_t | Y_{t-1}^1 = y_{t-1}^1, \ldots, Y_{t-1}^W = y_{t-1}^W, X_{t-1} = x_{t-1}) \\
&= \mathbb{P}(X_1 = x_1) \cdot \prod_{w=1}^{W} \mathbb{P}(Y_1^w = y_1^w | X_1 = x_1) \cdot \prod_{t=2}^{T} \big[\mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}) \cdot \\
&\qquad \prod_{w=1}^{W} \mathbb{P}(Y_t^w = y_t^w | Y_{t-1}^w = y_{t-1}^w, X_t = x_t) \big]
\end{aligned}$$

$$= \pi_{x_1} \prod_{w=1}^{W} \rho_{y_1^w}^{w,x_1} \prod_{t=2}^{T} \big[A_{x_{t-1}x_t} \prod_{w=1}^{W} L_{y_{t-1}y_t}^{w,x_t}\big]. \tag{5.16}$$

The above equation can be factorized because of the Markovian properties in (5.5).

5.B Parameter estimation derivations

Expectation function for HMM model Using (5.6) and (5.16), the expectation function (5.7) can be expanded as

$$\begin{aligned}
\mathcal{Q}(\lambda, \lambda') = &\sum_{\boldsymbol{x} \in \mathcal{M}^T} \big(\log \pi_{x_1}\big) \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X} = \boldsymbol{x} | \lambda') \\
&+ \sum_{w=1}^{W} \big[\sum_{\boldsymbol{x} \in \mathcal{M}^T} \big(\log \rho_{y_1^w}^{w,x_1}\big) \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X} = \boldsymbol{x} | \lambda')\big] \\
&+ \sum_{\boldsymbol{x} \in \mathcal{M}^T} \big(\sum_{t=2}^{T} \log A_{x_{t-1}x_t}\big) \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X} = \boldsymbol{x} | \lambda') \\
&+ \sum_{w=1}^{W} \big[\sum_{\boldsymbol{x} \in \mathcal{M}^T} \big(\sum_{t=2}^{T} \log L_{y_{t-1}^w y_t^w}^{w,x_t}\big) \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X} = \boldsymbol{x} | \lambda')\big].
\end{aligned} \tag{5.17}$$

Since the parameters we wish to optimize can be independently factorized into explicit terms as shown in the equation above, we can optimize each term individually using Lagrange multipliers.

First term, π_i The first term of (5.17) can be written as

$$\begin{aligned}
&\sum_{\boldsymbol{x} \in \mathcal{M}^T} \big(\log \pi_{x_1}\big) \cdot \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X} = \boldsymbol{x} | \lambda') \\
&= \sum_{i=1}^{M} \sum_{(x_2, \ldots, x_T) \in \mathcal{M}^{T-1}} (\log \pi_i) \cdot \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, X_1 = i, X_2 = x_2, \ldots X_T = x_T | \lambda')
\end{aligned}$$

$$= \sum_{i=1}^{M} (\log \pi_i) \cdot \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, X_1 = i|\lambda').$$

Using Lagrange multiplier γ and the constraint that $\sum_{i=1}^{M} \pi_i = 1$, we set the derivative equal to zero, i.e.,

$$\frac{\partial}{\partial \pi_i} \Big(\sum_{i=1}^{M} (\log \pi_i) \cdot \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, X_1 = i|\lambda') - \gamma (\sum_{i=1}^{M} \pi_i - 1) \Big) = 0.$$

Taking the derivative and using $\sum_{i=1}^{M} \pi_i = 1$ we get $\gamma = \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}|\lambda')$ and for π_i we get,

$$\pi_i = \frac{\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, X_1 = i|\lambda')}{\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}|\lambda')}. \tag{5.18}$$

Second term, $\rho^{w,i}$ The second term of (5.17) has a sum in w. We solve for the the w^{th} term inside the sum,

$$\sum_{\boldsymbol{x} \in \mathcal{M}^T} \big(\log \rho_{y_1^w}^{w,x_1} \big) \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X} = \boldsymbol{x}|\lambda') = \sum_{i=1}^{M} (\log \rho_{y_1^w}^{w,i}) \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, X_1 = i|\lambda').$$

We have M constraint equations $\sum_{\mu=1}^{N} \rho_\mu^{w,i} = 1$, as $i \in \mathcal{M}$. Hence we have M Lagrange multipliers. Setting the derivative to zero and using $\sum_{\mu=1}^{N} \rho_\mu^{w,i} = 1$, we get

$$\rho_\mu^{w,i} = \frac{\mathbb{1}(Y_1^w = \mu) \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, X_1 = i|\lambda')}{\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, X_1 = i|\lambda')} = \mathbb{1}(Y_1^w = \mu). \tag{5.19}$$

Third term, $\boldsymbol{A}$ The third term of (5.17) can be written as

$$\sum_{\boldsymbol{x} \in \mathcal{M}^T} \big(\sum_{t=2}^{T} \log A_{x_{t-1} x_t} \big) \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X} = \boldsymbol{x}|\lambda')$$

$$= \sum_{t=2}^{T} \sum_{i=1}^{M} \sum_{j=1}^{M} (\log A_{ij}) \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, X_{t-1} = i, X_t = j | \lambda').$$

We have M constraint equations $\sum_{k=1}^{M} a_{lk} = 1$, as $l \in \mathcal{M}$. Hence we need M Lagrange multipliers. Setting the derivative to zero and using $\sum_{k=1}^{M} a_{lk} = 1$, we get

$$A_{ij} = \frac{\sum_{t=2}^{T} \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, X_{t-1} = i, X_t = j | \lambda')}{\sum_{t=2}^{T} \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, X_{t-1} = i | \lambda')}. \tag{5.20}$$

Fourth term, $L^{w,i}$ We solve for the w^{th} term inside the sum of the fourth term of (5.17),

$$\sum_{\boldsymbol{x} \in \mathcal{M}^T} \Big(\sum_{t=2}^{T} \log L^{w,x_t}_{y^w_{t-1} y^w_t} \Big) \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X} = \boldsymbol{x} | \lambda')$$
$$= \sum_{t=2}^{T} \sum_{i=1}^{M} (\log L^{w,i}_{y^w_{t-1} y^w_t}) \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, X_t = i | \lambda').$$

Solving for $L^{w,i}_{\mu\nu}$ using the Lagrange multipliers we get,

$$L^{w,i}_{\mu\nu} = \frac{\sum_{t=2}^{T} \mathbb{1}(Y^w_{t-1} = \mu) \mathbb{1}(Y^w_t = \nu) \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, X_t = i | \lambda')}{\sum_{t=2}^{T} \mathbb{1}(Y^w_{t-1} = \mu) \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}, X_t = i | \lambda')}. \tag{5.21}$$

5.C Forward Backward variable recursion derivation

Forward equation Here we will derive the recursion for the forward variables. For simplicity we drop the λ' term for now. From (5.10) we

have for $t = 1$,

$$\begin{aligned}\alpha_i(1) &= \mathbb{P}(y_1^1, \ldots, y_1^W, X_1 = i) \\ &= \mathbb{P}(y_1^1 | y_1^2, \ldots, y_1^W, X_1 = i) \cdot \mathbb{P}(y_1^2, \ldots, y_1^W, X_1 = i) \\ &= \rho_{y_1^1}^{1,i} \cdots \rho_{y_1^W}^{W,i} \cdot \pi_i.\end{aligned}$$

For deriving the recursion relation, we have

$$\begin{aligned}&\alpha_i(t) = \mathbb{P}(y_1^1, \ldots, y_t^1, \ldots, y_1^W, \ldots, y_t^W, X_t = i) \\ &= \sum_{j=1}^{M} \mathbb{P}(y_1^1, \ldots, y_t^1, \ldots, y_1^W, \ldots, y_t^W, X_{t-1} = j, X_t = i) \\ &= \sum_{j=1}^{M} \mathbb{P}(y_t^1 | y_1^1, \ldots, y_{t-1}^1, \ldots, y_1^W, \ldots, y_t^W, X_{t-1} = j, X_t = i) \cdot \\ &\qquad \mathbb{P}(y_1^1, \ldots, y_{t-1}^1, \ldots, y_1^W, \ldots, y_t^W, X_{t-1} = j, X_t = i) \\ &= L_{y_{t-1}^1 y_t^1}^{1,i} \cdots L_{y_{t-1}^W y_t^W}^{W,i} \cdot \mathbb{P}(y_1^1, \ldots, y_{t-1}^1, \ldots, y_1^W, \ldots, y_{t-1}^W, X_{t-1} = j) \cdot A_{ji} \\ &= \sum_{j=1}^{M} \alpha_j(t-1) A_{ji} [\prod_{w=1}^{W} L_{y_{t-1}^w y_t^w}^{w,i}]. \end{aligned} \tag{5.22}$$

Backward equation From (5.10) we have, $\beta_i(T) = 1$. For the backward recursion we have,

$$\begin{aligned}&\beta_i(t) = \mathbb{P}(y_{t+1}^1, \ldots, y_T^1, \ldots, y_{t+1}^W, \ldots, y_T^W | X_t = i, y_t^1, \ldots, y_t^W) \\ &= \sum_{\{x_{t+1}, \ldots, x_T\} \in \mathcal{M}^{T-t}} \mathbb{P}(y_{t+1}^1, \ldots, y_T^1, \ldots, y_{t+1}^W, \ldots, y_T^W, X_{t+1} = x_{t+1}, \ldots, \\ &\qquad X_T = x_T | X_t = i, y_t^1, \ldots, y_t^W) \\ &= \sum_{\{x_{t+1}, \ldots, x_T\} \in \mathcal{M}^{T-t}} \mathbb{P}(y_{t+2}^1, \ldots, y_T^1, \ldots, y_{t+2}^W, \ldots, y_T^W, X_{t+2} = x_{t+2} \ldots, \\ &\qquad X_T = x_T | X_{t+1} = x_{t+1} X_t = i, y_t^1, \ldots, y_t^W, y_{t+1}^1, \ldots, y_{t+1}^W) \cdot\end{aligned}$$

$$\begin{aligned}
&\mathbb{P}(y_{t+1}^1|X_{t+1}=x_{t+1},X_t=i,y_t^1,\ldots,y_t^W)\cdots\\
&\mathbb{P}(y_{t+1}^W|X_{t+1}=x_{t+1},X_t=i,y_t^1,\ldots,y_t^W,y_{t+1}^1,\ldots,y_{t+1}^{W-1})\cdot\\
&\mathbb{P}(X_{t+1}=x_{t+1}|X_t=i)\\
=\sum_{j=1}^{M}&\mathbb{P}(y_{t+2}^1,\ldots,y_T^1,\ldots,y_{t+2}^W,\ldots,y_T^W,|X_{t+1}=j,y_{t+1}^1,\ldots,y_{t+1}^W)\cdot\\
&\mathbb{P}(y_{t+1}^1|X_{t+1}=j,X_t=i,y_t^1,\ldots,y_t^W)\cdots\mathbb{P}(y_{t+1}^W|X_{t+1}=j,X_t=i,\\
&y_t^1,\ldots,y_t^W,y_{t+1}^1,\ldots,y_{t+1}^{W-1})\cdot\mathbb{P}(X_{t+1}=j|X_t=i)\\
=\sum_{j=1}^{M}&A_{ij}[\prod_{w=1}^{W}L^w_{j,y_{t-1}^w y_t^w}]\beta_j(t+1).
\end{aligned}\tag{5.23}$$

5.D Expressions in terms of Forward-Backward equations

In order to re-write (5.9) in terms of forward-backward equations, we first write the following in terms of the forward-backwards equations. Note for lack of space we are dropping the random variable Y term in front of the y terms.

$$\begin{aligned}
\mathbb{P}(\boldsymbol{Y}=\boldsymbol{y},X_t=i)&=\mathbb{P}(y_1^1,\ldots,y_T^1,\ldots,y_1^W,\ldots,y_T^W,X_t=i)\\
&=\alpha_i(t)\beta_i(t).
\end{aligned}\tag{5.24}$$

$$\begin{aligned}
&\mathbb{P}(\boldsymbol{Y}=\boldsymbol{y},X_{t-1}=i,X_t=j)=\mathbb{P}(y_1^1,\ldots,y_T^1,\ldots,y_1^W,\ldots,y_T^W,X_{t-1}=i,X_t=j)\\
&=\mathbb{P}(y_t^1,\ldots,y_T^1,\ldots,y_t^W,\ldots,y_T^W,X_t=j|y_1^1,\ldots,y_{t-1}^1,\ldots,y_1^W,\ldots,y_{t-1}^W,\\
&\quad X_{t-1}=i)\cdot\mathbb{P}(y_1^1,\ldots,y_{t-1}^1,\ldots,y_1^W,\ldots,y_{t-1}^W,X_{t-1}=i)\\
&=\alpha_i(t-1)\cdot\mathbb{P}(y_t^1,\ldots,y_T^1,\ldots,y_t^W,\ldots,y_T^W|y_1^1,\ldots,y_{t-1}^1,\ldots,y_1^W,\ldots,y_{t-1}^W,\\
&\quad X_{t-1}=i,X_t=j)\cdot\mathbb{P}(X_t=j|X_{t-1}=i)
\end{aligned}$$

$$= \alpha_i(t-1)A_{ij} \cdot \mathbb{P}(y_{t+1}^1, \ldots, y_T^1, \ldots, y_t^W, \ldots, y_T^W | y_1^1, \ldots, y_t^1, \ldots, y_1^W, \ldots, y_{t-1}^W,$$
$$X_{t-1} = i, X_t = j) \cdot \mathbb{P}(y_t^1 | y_{t-1}^1, X_t = j)$$
$$= \alpha_i(t-1)A_{ij}[\prod_{w=1}^{W} L^w_{j,y_{t-1}^w y_t^w}]\beta_j(t). \tag{5.25}$$

5.E Estimates for same one-step transition matrix model

The parameter estimates for this model are given by

$$\begin{cases} \widehat{\pi}_i = \frac{\alpha_i(1)\beta_i(1)}{\sum_{j=1}^{M}\alpha_i(1)\beta_i(1)}, \\ \widehat{A}_{ij} = \frac{\sum_{t=2}^{T}\alpha_i(t-1)\widehat{A}'_{ij}\left[\prod_{w=1}^{W}\widehat{L}'^{j}_{y_{t-1}^w y_t^w}\right]\beta_j(t)}{\sum_{t=2}^{T}\alpha_i(t-1)\beta_i(t-1)} \\ \widehat{\rho}^i_\mu = \frac{1}{W}\sum_{w=1}^{W}\mathbb{1}(Y_1^w = \mu), \text{and} \\ \widehat{L}^i_{\mu\nu} = \frac{\sum_{t=2}^{T}[\alpha_i(t)\beta_i(t)\sum_{w=1}^{W}\mathbb{1}(Y_{t-1}^w=\mu)\mathbb{1}(Y_t^w=\nu)]}{\sum_{t=2}^{T}[\alpha_i(t)\beta_i(t)\sum_{w=1}^{W}\mathbb{1}(Y_{t-1}^w=\mu)]}. \end{cases} \tag{5.26}$$

The recursions of the forward-backward equations are given by $\forall w \in \mathcal{W},\ \forall i \in \mathcal{M}$

$$\begin{cases} \alpha_i(t) = \sum_{j=1}^{M}\alpha_j(t-1)A_{ji}[\prod_{w=1}^{W} L^i_{y_{t-1}^w y_t^w}], \\ \beta_i(t) = \sum_{j=1}^{M} A_{ij}[\prod_{w=1}^{W} L^j_{y_{t-1}^w y_t^w}]\beta_j(t+1), \end{cases} \tag{5.27}$$

with initialization $\alpha_i(1) = \pi_i \cdot \prod_{w=1}^{W}\rho^i_{y_1^w}$ and $\beta_i(T) = 1$.

Chapter 6

Computing first passage times for Markov-modulated fluid models using PDE schemes

6.1 Introduction

A Markov-modulated fluid model (or queue) is a mathematical model which describes the fluid level in a buffer (or storage device) which is being filled or emptied for exponentially distributed random periods [16, 115]. The rate at which the buffer is filled or emptied depends generally on the state of a continuous-time Markov chain (CTMC) known as the *background process* [140] or *environment process* [99, 119]. Given this state, the rate is deterministic. A CTMC as environment process is the most common, however other environment processes like white noise and Ornstein-Uhlenbeck process have also been studied [99].

Fluid queues are used to approximate discrete queueing models for dam theory [117], telecommunication networks [16, 60, 114, 142], transportation systems [122], modeling forest fires [152], ruin probability [19, 23], video streaming [41, 101, 170], and more [141].

Steady state analysis of fluid queues is covered in detail in the literature, see e.g. [18, 99, 134]. In this chapter we are interested in the transient behavior of the Markov-modulated fluid queue, in particular the distribution of the first passage time defined as the first time the buffer empties. Mean first passage times for fluid queues have been studied by [20, 100]. In the literature the common approach involves derivation of first passage times from Laplace-Stieltjes transforms (LST) [43, 63, 119, 133].

The motivating application for this chapter is that of power grids and renewable energy. For energy sufficiency and reliability of micro-grids powered by wind farms, incorporation of energy storage devices is of great importance. Hence, for reliable power delivery, as well as for scheduling and planning purposes in micro-grids it is relevant to study the probability of energy storage devices running empty in a given time interval. The storage device running empty implies there is insufficient power in the micro-grid and it needs to be fed by different energy sources or loads needs to be shed. In order to quantify the probability of the micro-grid being no longer self sufficient, we wish to apply the Markov-modulated fluid queue to micro-grids powered by wind energy and supported with energy storage devices. To this end we consider a continuous-time version of the hidden Markov model for wind farm power output discussed in Chapter 5.

In [119] it was proven that the cumulative distribution function of the first passage time of a Markov-modulated fluid queue model follows a *hyperbolic partial differential equation* (PDE). In [119] the PDE is solved using LST and eigenvalue theory. However, for practical purposes and applications it is also important to invert the LST solution back to the time domain. However, this involves spectral decomposition of a matrix of size of the state space of the CTMC and numerical inversion of the LST,

leading to cumbersome and expensive computations, thereby limiting the applicability of the method. Hence, in this chapter we propose methods to numerically solve the PDE directly, without using the LST. Our focus is on cases with large cardinality of the state space of the CTMC, as this is relevant in the context of micro-grids (see Chapter 5).

Numerical integration of hyperbolic PDEs has been studied extensively in the literature [85, 106, 151]. The PDE central to this chapter is a 1-dimensional (1-D) coupled advection-reaction PDE [85, 106]. We propose to use first and second order upwind space discretization schemes and different (explicit and implicit) time integration schemes in this chapter. These schemes are well established in scientific computing disciplines but have not been used in the field of queueing theory. We compare results from numerical integration of the PDE with Monte Carlo simulations to assess the accuracy. We find that the PDE schemes can efficiently and accurately compute the first passage time distribution and they can be applied to large state space systems. We also compare our PDE approach with the LST method proposed in [119].

In Section 6.2 we describe the system setup, buffer model, first passage time distribution and the PDE for the distribution function. In Section 6.3 we propose different integration schemes for solving the PDE under consideration. In Section 6.4 we briefly present the LST method proposed by [119]. In Section 6.5 we compare results of the different PDE integration schemes. We compare the PDE schemes with Monte Carlo simulation results. For systems with a small state space of the CTMC, we also compare the PDE schemes with the LST method. The chapter finishes with a brief conclusion in Section 6.6.

6.2 System setup

Let $\{Z(t), t \geq 0\}$ be a CTMC on a finite state space $\mathcal{S}$ with generator matrix $\boldsymbol{Q} = (q_{\alpha\beta})_{\alpha,\beta\in\mathcal{S}}$. The steady state probability distribution of $\{Z(t)\}$

is denoted by $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_S)$, where $S = |\mathcal{S}|$. Let r_α be the net input rate with which the level of fluid in the buffer increases or decreases when the CTMC is in state α. Let $B(t)$ be the fluid level in the buffer at time t. The dynamics of the fluid level is given by

$$\frac{dB(t)}{dt} = \begin{cases} r_\alpha & \text{if } 0 \leq B(t) \leq \mathrm{B}^{\max} \text{ and } Z(t) = \alpha, \\ \max(r_\alpha, 0) & \text{if } B(t) = 0 \text{ and } Z(t) = \alpha, \\ \min(r_\alpha, 0) & \text{if } B(t) = \mathrm{B}^{\max} \text{ and } Z(t) = \alpha. \end{cases} \tag{6.1}$$

$\mathrm{B}^{\max}$ is the maximum fluid storage capacity of the buffer. Clearly, the fluid level increases when $r_\alpha > 0$ and $B(t) < \mathrm{B}^{\max}$, and decreases when $r_\alpha < 0$ and $B(t) > 0$. Furthermore, $B(t) \in [0, \mathrm{B}^{\max}]$ $\forall t$. When $r_\alpha = 0$ the fluid level in the buffer remains unchanged.

6.2.1 First passage time distribution

We are interested in computing the distribution of the first passage (or hitting) time, i.e, the first time the buffer empties given that it started with some initial fluid level x at time $t = 0$. In order to do so, we define a random variable T which denotes the first time the buffer empties,

$$T := \inf\{t > 0 : B(t) = 0\}. \tag{6.2}$$

The distribution function of T given the initial fluid level in the buffer x, i.e., the probability that the buffer empties before time t given that it started with initial fluid level x is defined as,

$$J(x, t) := \mathbb{P}(T \leq t | B(0) = x). \tag{6.3}$$

We are also interested in finding the minimum value of the initial fluid level in the buffer x_{min} such that the buffer empty probability is smaller than p_{empty}, i.e.,

$$x_{min} = \min\{x : J(x,t) < p_{empty}\}. \tag{6.4}$$

A similar problem of finding x_{min} has been studied in [42] for a Markov-modulated video streaming model. However, the analysis in [42] is focused on the case with only two states ($S = 2$).

In the literature, slightly different variants of the above mentioned distribution function (6.3) are used depending on the initial conditions. A joint probability distribution of the buffer being empty before time t and the state of the CTMC at time T conditioned on the initial states of the buffer and the CTMC is defined in [119, 177]:

$$H_{\alpha\beta}(x,t) := \mathbb{P}(T \leq t, Z(T) = \beta | B(0) = x, Z(0) = \alpha), \tag{6.5}$$

$\forall \alpha, \beta \in \mathcal{S}, x > 0$ and $t \geq 0$. In [22, 88], a probability distribution of T conditioned on the initial state of the buffer fluid content and the CTMC is proposed:

$$K_\alpha(x,t) := \mathbb{P}(T \leq t | B(0) = x, Z(0) = \alpha) = \sum_{\beta=1}^{S} H_{\alpha\beta}(x,t), \tag{6.6}$$

$\forall \alpha \in \mathcal{S}, x > 0$ and $t \geq 0$. If we assume that the CTMC starts in stationarity, i.e. $\mathbb{P}(Z(0) = \alpha) = \pi_\alpha, \ \forall \alpha \in \mathcal{S}$, we can compute the original probability distribution of T given in (6.3) from (6.6) by

$$J(x,t) = \sum_{\alpha=1}^{S} \pi_\alpha K_\alpha(x,t). \tag{6.7}$$

Hence, in this chapter we focus on solving for $K_\alpha(x,t)$ which follows the following backward equation (see [119] for the proof) $\forall \alpha \in \mathcal{S}$ and $\forall x, t > 0$,

$$\frac{\partial K_\alpha(x,t)}{\partial t} - r_\alpha \frac{\partial K_\alpha(x,t)}{\partial x} = \sum_{\beta=1}^{S} q_{\alpha\beta} K_\beta(x,t). \tag{6.8}$$

The initial and boundary conditions are given by:

- $K_\alpha(x,0) = 0 \ \forall x > 0, \quad \forall \alpha \in \mathcal{S}.$
- $K_\alpha(0,0) = \begin{cases} 1 & \forall \{\alpha : r_\alpha \leq 0\}, \\ 0 & \forall \{\alpha : r_\alpha > 0\}. \end{cases}$
- $K_\alpha(0,t) = 1 \ \ \forall \{\alpha : r_\alpha \leq 0\} \quad \text{and} \quad \forall t > 0.$
- $K_\alpha(B^{\max}, t)$ follows an ordinary differential equation (ODE) at this boundary (when the buffer has reached $B^{\max}$), i.e.,
  $\frac{dK_\alpha(B^{\max},t)}{dt} = \sum_{\beta=1}^{S} q_{\alpha\beta} K_\beta(B^{\max},t) \quad \forall \{\alpha : r_\alpha > 0\} \quad \text{and} \quad \forall t > 0.$

6.3 PDE schemes

In this section we summarize different PDE schemes for numerically integrating the PDE in (6.8). Direct numerical integration for solving this PDE has not been deployed before in the field of queueing systems. The PDE in (6.8) is a hyperbolic PDE, more specifically an *advection-reaction* PDE with one space dimension [85, 106]. In general, advection-reaction PDEs represent different chemically reacting species that are being advected by a fluid flow.

For numerical integration of the PDE, the continuous space and time domains are discretized. The error of the numerically integrated solution of the PDE depends on the discretization scheme used. In this section we discuss different space and time integration schemes. We present two *upwind* space discretization schemes, namely the first order upwind and the second order upwind scheme. We present different time integration schemes for integrating the PDE, including one that is appropriate if the PDE is stiff (when the time step length gets highly restrictive leading to high computational cost for numerically integrating the PDE).

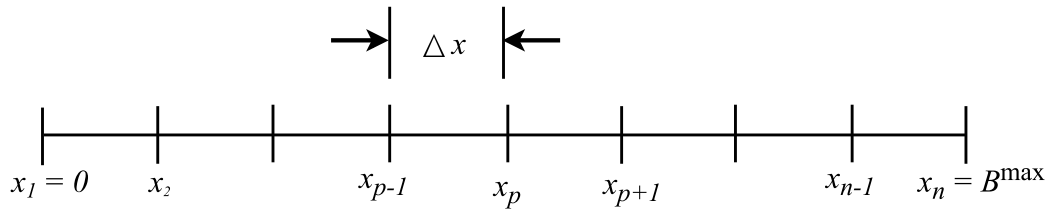


Figure 6.1: Spatial discretization for first order upwind scheme.

6.3.1 First order upwind scheme for spatial discretization

In this section we discuss the space discretization scheme for the PDE in (6.8). As the PDE is an advection-reaction equation we use the simplest first order upwind *finite difference* scheme [85].

In our problem the space variable corresponds to the initial fluid level in the buffer. We discretize the interval $[0, B^{\max}]$ using n equidistant grid points with spacing $\Delta x = B^{\max}/(n-1)$. The grid points are given by $\{x_1 = 0, x_2, x_3, \ldots, x_n = B^{\max}\}$ with $x_p = (p-1)\Delta x$ for $p = 1, \ldots, n$. The schematic representation of the grid point spacing and positions is shown in Figure 6.1.

We discretize the space first keeping the time constant. This will give us a set of coupled ordinary differential equations (ODEs). The first order upwind discretization scheme for the p-th space point is given by,

$$\frac{dK_\alpha^p}{dt} = \begin{cases} \frac{r_\alpha}{\Delta x}(K_\alpha^p - K_\alpha^{p-1}) + \sum_{\beta=1}^{S} q_{\alpha\beta} K_\beta^p & \forall\{\alpha : r_\alpha \leq 0\}, \\ \frac{r_\alpha}{\Delta x}(K_\alpha^{p+1} - K_\alpha^p) + \sum_{\beta=1}^{S} q_{\alpha\beta} K_\beta^p & \forall\{\alpha : r_\alpha > 0\}, \end{cases} \tag{6.9}$$

for $p = 2, \ldots, n-1$. In the above equation, $K_\alpha^p := K_\alpha(x_p)$ for $p = 1, \ldots, n$.

The above equation can be written in a much more compact vectorial form. Let $\boldsymbol{K} = [K_1^1, \ldots, K_1^n, \ldots, K_S^1, \ldots, K_S^n]^T$, then (6.9) can be written as

$$\frac{d\boldsymbol{K}}{dt} = \boldsymbol{A}\boldsymbol{K}, \tag{6.10}$$

where $\boldsymbol{A}$ is a matrix whose elements are determined by the space discretization and the boundary conditions. The structure of matrix $\boldsymbol{A}$ is given in Appendix 6.A. Note that (6.10) is a set of linear ODEs and can in principle be solved by

$$\boldsymbol{K}^{q+1} = e^{(\boldsymbol{A}\Delta t)}\boldsymbol{K}^q. \tag{6.11}$$

In the above equation $\boldsymbol{K}^q := \boldsymbol{K}(t_q)$ and $t_{q+1} - t_q = \Delta t$. In this chapter we divide the time interval $[0, t]$ using m time discretization points with spacing $\Delta t = t/(m-1)$. The q-th time point is given by $t_q = (q-1)\Delta t$ for $q = 1, \ldots, m$.

Computing the exponential of the matrix $\boldsymbol{A}$ in (6.11) will be computationally very expensive as the size of the matrix is $(nS) \times (nS)$, hence it can be large. In the subsequent sections we describe efficient numerical time integration schemes.

6.3.1.1 Time integration schemes

In this section we discuss two different time integration schemes. The first is the explicit 4th order *Runge-Kutta* (RK4) scheme and the second is the implicit 2nd order *backward differentiation formula* (BDF2) scheme that is more suitable for stiff problems.

Explicit RK4 scheme: For numerical time integration, the integration time step Δt needs to be specified. In case of a hyperbolic PDE discretized with a finite difference scheme, for numerical stability of explicit time integration schemes, it is necessary that the *Courant-Friedrichs-Lewy* (CFL)

condition is satisfied [85, 106]. For our problem the CFL condition yields

$$\Delta t \leq \min_{\alpha} \left| \frac{\Delta x}{r_\alpha} \right| \quad \forall \alpha \in \mathcal{S}. \tag{6.12}$$

Given Δx must choose Δt such that it satisfies the above condition. We deploy the explicit RK4 scheme:

$$\boldsymbol{K}^{q+1} = \boldsymbol{K}^q + \frac{\Delta t}{6}[\boldsymbol{p}_1 + 2\boldsymbol{p}_1 + 2\boldsymbol{p}_3 + \boldsymbol{p}_4], \tag{6.13}$$

where

$$\begin{aligned}
\boldsymbol{p}_1 &= \boldsymbol{A}\boldsymbol{K}^q, \\
\boldsymbol{p}_2 &= \boldsymbol{A}[\boldsymbol{K}^q + \frac{\Delta t}{2}\boldsymbol{p}_1], \\
\boldsymbol{p}_3 &= \boldsymbol{A}[\boldsymbol{K}^q + \frac{\Delta t}{2}\boldsymbol{p}_2], \\
\boldsymbol{p}_4 &= \boldsymbol{A}[\boldsymbol{K}^q + \Delta t\boldsymbol{p}_3].
\end{aligned}$$

The RK4 scheme is 4th order accurate, i.e. the local error is $\mathcal{O}(\Delta t^4)$. The computational complexity of the RK4 scheme along with the first order upwind scheme is $\mathcal{O}(mnS^2)$. Note that $\boldsymbol{A}$ is a sparse matrix with $\mathcal{O}(nS^2)$ nonzero elements.

The computational cost of an explicit scheme such as RK4 can become high in case of stiffness. In that case, using an implicit time integration scheme is beneficial. A system of ODEs is said to be stiff if the time step length for explicit schemes becomes very restrictive (i.e., very small) due to stability requirements rather than accuracy requirements [103]. For the system under consideration here, stiffness can occur if some of the rates r_α (or $q_{\alpha\beta}$) have different orders of magnitude. To overcome this problem we resort to the implicit time integration scheme discussed in the following paragraph.

Implicit BDF2 scheme: The backward differentiation formula (BDF) is a family of implicit schemes for integrating ordinary differential equations. The k-th order BDF scheme uses k previous time step approximations of the solution for computing the present approximation of the solution. We use the 2nd order BDF (BDF2) scheme which is given by [155]

$$\boldsymbol{K}^{q+1} = \frac{4}{3}\boldsymbol{K}^{q} - \frac{1}{3}\boldsymbol{K}^{q-1} + \frac{2}{3}\Delta t[\boldsymbol{A}\boldsymbol{K}^{q+1}]. \tag{6.14}$$

The method is implicit, and to compute $\boldsymbol{K}^{q+1}$ given $\boldsymbol{K}^{q}, \boldsymbol{K}^{q-1}$ a system of linear equations needs to be solved. This makes the method computationally expensive. However, the method being implicit in nature there is less restriction on Δt for stability [155]. In order to solve the system of linear equations in (6.14) we use the LU decomposition method (worst case complexity $\mathcal{O}((nS)^3)$ [73]) implemented in MATLAB . In practice the sparsity of the matrix will reduce the computational complexity of the LU decomposition, however the degree of reduction is complicated to assess because it depends on the sparsity pattern. The complexity of BDF2 with first order upwind scheme becomes $\mathcal{O}(mn^2S^2 + (nS)^3)$, where m is the number of time discretization points and n is the number of space discretization points.

6.3.2 Second-order upwind scheme using flux limiters

In Section 6.3.1 we described the method of solving the PDE (6.8) using the first order finite difference upwind space discretization scheme. The drawback of the this scheme is that the error in the solution is $\mathcal{O}(\Delta x)$. In order to achieve higher order accuracy in the numerically integrated solution of the PDE one can use higher order schemes. Hence, in this section we discuss the second-order upwind scheme for space discretization which has a global error $\mathcal{O}(\Delta x^2)$.

The second order upwind scheme leads to oscillations around discon-

tinuities in the solution domain. In our problem the distribution function in (6.6) has jumps, leading to discontinuities in the solution. These jumps occur because there is a positive probability that the CTMC starts and remains in a state with negative rate ($r_\alpha < 0$) for any finite amount of time and hence there is a positive probability that the buffer hits zero exactly at time $|xr_\alpha|$, where x is the initial fluid level in the buffer. To limit the values of the spatial derivatives to realistic values around discontinuity of the solution, the concept of *flux limiters* is used.

6.3.2.1 Second order upwind

The PDE in (6.8) is spatially approximated on cell-centered grid points x_p (see Figure 6.2) by the semi-discrete conservation form [84]

$$\frac{dK_\alpha^p(t)}{dt} - \frac{1}{\Delta x}[F_\alpha^{p+1/2}(t) - F_\alpha^{p-1/2}(t)] = \sum_{\beta=1}^{S} q_{\alpha\beta} K_\beta^p(t), \quad \forall \alpha \in \mathcal{S}. \tag{6.15}$$

In the above equation, $K_\alpha^p(t)$ is the continuous-time approximation of $K_\alpha(x_p, t)$ at the grid point $x_p = (p - \frac{1}{2})\Delta x$ for $p = 1, \ldots, n$, where n is the number of cell centered grid points. $F_\alpha^{p+1/2}$ is the numerical flux at the right vertex of the p^{th} cell and it is the approximation to the analytical flux at cell vertex $x_{p+1/2} = \frac{1}{2}(x_{p+1} + x_p)$ for $p = 1, \ldots, n-1$. The boundary vertices are $x_{1/2} = 0$ and $x_{n+1/2} = B^{\max}$.

For the second order upwind scheme the numerical flux is given by [84]

$$F_\alpha^{p+1/2} = \begin{cases} r_\alpha[\frac{3}{2}K_\alpha^p - \frac{1}{2}K_\alpha^{p-1}] & \forall\{\alpha : r_\alpha \leq 0\}, \\ r_\alpha[\frac{3}{2}K_\alpha^{p+1} - \frac{1}{2}K_\alpha^{p+2}] & \forall\{\alpha : r_\alpha > 0\}. \end{cases} \tag{6.16}$$

For other higher order (for example third and fourth order) discretization schemes, see [84].

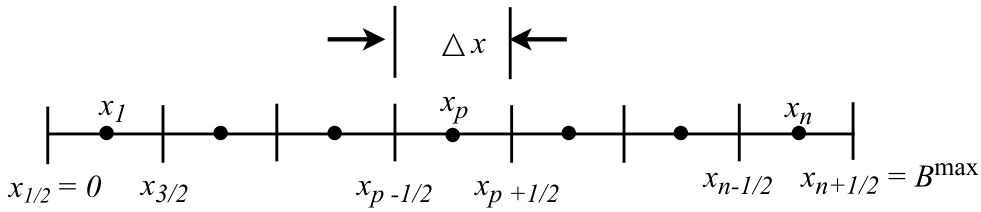


Figure 6.2: Cell-centered grid discretization for second order upwind scheme.

6.3.2.2 Positivity of the second order upwind scheme

As mentioned earlier the second order upwind scheme suffers from oscillations around discontinuities in the solution domain. It implies that the scheme can lead to negative values in the numerical solution. This is unwanted as we are approximating a distribution function whose values must lie in $[0, 1]$. In this section we give an explanation for this behavior of the scheme. We follow [84] closely in this section.

The positivity rule says that for any non-negative initial solution ($k_\alpha^p(t_0) \geq 0,\ \forall p$) the solution $K_\alpha^p(t)$ which evolves in time remains non-negative $\forall t \geq t_0$. A scheme is positive if and only if $\forall p$ and $\forall t \geq t_0$ [84],

$$K_\alpha^p(t) = 0, \quad K_\alpha^{p'}(t) \geq 0, \quad \forall p' \neq p \implies \frac{dK_\alpha^p(t)}{dt} \geq 0. \tag{6.17}$$

The above (6.17) means that if the solution at any cell $p' \neq p$ increases, then the solution in cell p should also increase or remain the same. It should be noted that the first order upwind scheme (see Section 6.3.1) satisfies this rule. For the second order upwind scheme we plug in the values of the numerical fluxes from (6.16) into (6.15) and we get $\forall\{\alpha$:

$r_\alpha < 0\}$

$$\frac{dK_\alpha^p(t)}{dt} = \frac{|r_\alpha|}{2\Delta x}[-K_\alpha^{p-2}(t) + 4K_\alpha^{p-1}(t) - 3K_\alpha^p(t)] + \sum_{\beta=1}^{S} q_{\alpha\beta}K_\beta^p(t). \quad (6.18)$$

From above we can see that the coefficient of $K_\alpha^{p-2}(t)$ is negative and hence when $K_\alpha^{p-2}(t)$ increases $\frac{dK_\alpha^p(t)}{dt}$ decreases. Therefore, the positivity rule in (6.17) is not satisfied by the second order upwind scheme. The q-terms in the above equation pose no problems for the positivity rule because of the structure of the $\boldsymbol{Q}$ matrix, i.e the non-diagonal terms of the matrix are non-negative. Analogous arguments show that the positivity rule is also not satisfied for $\frac{dK_\alpha^p(t)}{dt} \quad \forall\{\alpha : r_\alpha \geq 0\}$.

6.3.2.3 Flux limiters

We have seen in the previous section that the second order upwind scheme does not preserve non-negativity of the time evolving solution. To ensure non-negativity of the solution we apply the *flux limiter* method. The expression for the flux with the limiter is given by

$$F_\alpha^{p+1/2} = \begin{cases} r_\alpha[K_\alpha^p + \frac{1}{2}\phi_{p+1/2}(K_\alpha^p - K_\alpha^{p-1})] & \forall\{\alpha : r_\alpha \leq 0\}, \\ r_\alpha[K_\alpha^{p+1} + \frac{1}{2}\phi_{p+1/2}(K_\alpha^{p+1} - K_\alpha^{p+2})] & \forall\{\alpha : r_\alpha > 0\}, \end{cases} \quad (6.19)$$

where $\phi_{p+1/2} := \phi(f_{p+1/2})$. Here ϕ is a non-linear function called the *limiter function*, and $f_{p+1/2}$ is related to the adjacent fluxes. Both ϕ and f will be specified later in this section. The limiter is applied such that the solution remains non-oscillatory in the discontinuous part of the solution and in the smooth part the higher order scheme is applied [84, 106]. Note that with $\phi = 0$, we get the first order upwind scheme. We show that the flux limited expression in (6.19) preserves positivity of the semi-discrete scheme in (6.15). For $\{\alpha : r_\alpha < 0\}$ we get from (6.19) and (6.15)

$$\frac{dK_\alpha^p}{dt} = -\frac{|r_\alpha|}{\Delta x}[(K_\alpha^p - K_\alpha^{p-1}) + \frac{1}{2}\phi_{p+1/2}(K_\alpha^p - K_\alpha^{p-1}) - \frac{1}{2}\phi_{p-1/2}(K_\alpha^{p-1} - K_\alpha^{p-2})] + \sum_{\beta=1}^{S} q_{\alpha\beta}K_\beta^p. \tag{6.20}$$

Let $f_{p-1/2} = \dfrac{K_\alpha^p - K_\alpha^{p-1}}{K_\alpha^{p-1} - K_\alpha^{p-2}}$ and assuming $f_{p-1/2} \neq 0$, i.e., $K_\alpha^p - K_\alpha^{p-1} \neq 0$ we get from (6.20),

$$\frac{dK_\alpha^p}{dt} = -\frac{|r_\alpha|}{\Delta x}[(1 + \frac{1}{2}\phi_{p+1/2}) - \frac{1}{2}\frac{\phi_{p-1/2}}{f_{p-1/2}}](K_\alpha^p - K_\alpha^{p-1}) + \sum_{\beta=1}^{S} q_{\alpha\beta}K_\beta^p. \tag{6.21}$$

Applying the positivity rule in (6.17) to (6.21), we find that the term in the square bracket in the above equation should be non-negative and this leads to

$$\frac{\phi_{p-1/2}}{f_{p-1/2}} - \phi_{p+1/2} \leq 2. \tag{6.22}$$

If $f_{p-1/2} = 0$, i.e., $K_\alpha^p - K_\alpha^{p-1} = 0$, then (6.20) leads to $\frac{K_\alpha^p(t)}{dt} > 0$ if $\phi_{p-1/2} = 0$. Hence, we assume that $\phi_{p-1/2} = 0$ if $f_{p-1/2} \leq 0$. Also assuming $0 \leq \phi_{p-1/2}, \phi_{p+1/2} \leq \delta$, for any constant $\delta > 0$, then (6.22) is satisfied if $\phi_{p-1/2} \leq 2f_{p-1/2}$. Hence, $\phi_{p-1/2}$ and $\phi_{p+1/2}$ can be expressed as functions of $f_{p-1/2}$ and $f_{p+1/2}$. Thus we have,

$$\begin{cases} \phi(f) = 0 & \text{for } f \leq 0, \\ \phi(f) \leq 2f & \text{for } f > 0, \quad \text{and} \\ \phi(f) \leq 2 & \text{for } f > 0. \end{cases} \tag{6.23}$$

From the conditions in (6.23) the limiter must be within or at the boundary of the so-called Sweby's TVD (total variation diminishing) domain (see Fig. 1a of [156]). One gets the same results $\forall\{\alpha : r_\alpha \geq 0\}$.

A wide range of limiters that fall in this TVD region has been proposed in the literature, see [14, 168] for an overview and comparisons of different limiters. For solving our problem we use the Koren limiter [97], given by

$$\phi(f) = \max(0, \min(2f, \frac{2+f}{3}, 2)). \tag{6.24}$$

6.3.2.4 Time integration scheme

For time integration we cannot use RK4 scheme (see Section 6.3.1.1) with the flux limited second order upwind scheme. This is because for RK4 scheme to satisfy the TVD conditions, the step size must be zero (see equation (33) of [84]). Hence, we use the explicit 3rd order Runge Kutta 3b (RK3b) scheme described and suggested in [84]. We describe the scheme in this section. Let the space discretized $K_\alpha(x)$ is given by $K_\alpha^p := K_\alpha(x_p)$, where p is the index of the space discretization point (see Section 6.3.2.1). Let $\boldsymbol{K} = [K_1^1, \ldots, K_1^n, \ldots, K_S^1, \ldots, K_S^n]^T$. Then we can rewrite the system of equations (6.15) by expanding $F_\alpha^{p+1/2}$ using (6.19) in the form

$$\frac{d\boldsymbol{K}}{dt} = \boldsymbol{A}(\boldsymbol{K})\boldsymbol{K}, \tag{6.25}$$

where the matrix $\boldsymbol{A}(\boldsymbol{K})$ is a function of $\boldsymbol{K}$ because of the flux-limiter (which is a function of $\boldsymbol{K}$). As can be seen in (6.15) and (6.19), the elements of $\boldsymbol{A}(\boldsymbol{K})$ are determined by the second order space discretization, the boundary conditions and the source terms. For details of the space discretization see Appendix 6.B. Note that unlike the upwind scheme described in Section 6.3.1 and (6.10) the matrix $\boldsymbol{A}$ now has to be evaluated at each time step. This is because the flux limiter function ϕ needs to be evaluated at every time step which changes the matrix $\boldsymbol{A}$. The explicit RK3b is given by

$$\boldsymbol{K}^{q+1} = \boldsymbol{K}^q + \frac{\Delta t}{6}[\boldsymbol{p}_1 + \boldsymbol{p}_2 + 4\boldsymbol{p}_3], \tag{6.26}$$

where

$$\begin{aligned}\boldsymbol{p}_1 &= \boldsymbol{A}(\boldsymbol{K}^q)\boldsymbol{K}^q,\\ \boldsymbol{p}_2 &= \boldsymbol{A}(\boldsymbol{K}^q + \Delta t\boldsymbol{p}_1)[\boldsymbol{K}^q + \Delta t\boldsymbol{p}_1],\\ \boldsymbol{p}_3 &= \boldsymbol{A}(\boldsymbol{K}^q + \frac{\Delta t}{4}[\boldsymbol{p}_1 + \boldsymbol{p}_2])[\boldsymbol{K}^q + \frac{\Delta t}{4}[\boldsymbol{p}_1 + \boldsymbol{p}_2]],\end{aligned}$$

where q is the discretized time index and the time discretization step size Δt is chosen in such a way that the CFL condition in (6.12) is satisfied. We take $\Delta t = 0.25 \times \min_{\alpha\in\mathcal{S}} \left|\frac{\Delta x}{r_\alpha}\right|$. RK3b gives a total error $\mathcal{O}(\Delta t^3)$. The computation complexity of the scheme is $\mathcal{O}(nmS^2)$, where as before, n is the number of space discretization points, m the number of time discretization points and S is the cardinality of the state space of the CTMC.

6.4 Laplace-Stieltjes transform method

In queueing theory the common approach to compute first passage times for fluid queues is by the Laplace-Stieltjes transform (LST) [43, 63, 119, 133]. In this section we discuss briefly the LST method presented in [119] for calculating (6.6). The starting point in [119] is the distribution function defined in (6.5). The PDE satisfied by $H_{\alpha\beta}(x,t)$, $\forall\{\beta : r_\beta < 0\}, \forall\alpha \in \mathcal{S}$ and $\forall x, t > 0$ is

$$\frac{\partial H_{\alpha\beta}(x,t)}{\partial t} - r_\alpha \frac{\partial H_{\alpha\beta}(x,t)}{\partial x} = \sum_{\gamma=1}^{S} q_{\gamma\beta} H_{\gamma\beta}(x,t), \tag{6.27}$$

with appropriate boundary and initial conditions discussed in [119]. Note that $H_{\alpha\beta}(x,t) = 0$, $\forall\{\beta : r_\beta > 0\}$, and $\forall\alpha \in \mathcal{S}$. Taking the LST, $\tilde{H}_{\alpha\beta}(x,w) := \int_0^\infty e^{-wt} dH_{\alpha\beta}(x,t)$, the vector $\tilde{\boldsymbol{H}}_\beta = [\tilde{H}_{1\beta}, \ldots \tilde{H}_{S\beta}]$ satisfies the following ODE

$$\boldsymbol{R}\frac{d\tilde{\boldsymbol{H}}_\beta(x,w)}{dx} = (w\boldsymbol{I} - \boldsymbol{Q})\tilde{\boldsymbol{H}}_\beta(x,w), \tag{6.28}$$

where $\boldsymbol{R} = diag(r_1, \dots r_S)$. The following functional form is assumed

$$\tilde{\boldsymbol{H}}_\beta(x, w) = e^{s(w)x}\boldsymbol{\Phi}(w). \tag{6.29}$$

Substituting in (6.28) we obtain

$$\boldsymbol{R}s(w)\boldsymbol{\Phi}(w) = (w\boldsymbol{I} - \boldsymbol{Q})\boldsymbol{\Phi}(w), \tag{6.30}$$

where $s(w)$ is a scalar and $\boldsymbol{\Phi}(w)$ is a vector that are both to be determined. The above equation can be re-written as

$$(\boldsymbol{Q} + s(w)\boldsymbol{R} - w\boldsymbol{I})\boldsymbol{\Phi}(w) = 0. \tag{6.31}$$

Hence, the problem boils down to solving an eigenvalue problem, i.e., we need to find the roots of $\Delta(s, w) = \det(\boldsymbol{Q} + s(w)\boldsymbol{R} - w\boldsymbol{I}) = 0$. Assuming that the diagonal elements of $\boldsymbol{R}$ has no zeros, let $s_k(w)$ for $k = 1, \dots, S$ be the roots of $\Delta(s, w)$ and $\boldsymbol{\Phi}^k(w)$ the corresponding eigenvectors. We have

$$\tilde{H}_{\alpha\beta}(x, w) = \sum_{k=1}^{S} a_{k\beta} e^{s_k(w)x} \phi_\alpha^k(w), \tag{6.32}$$

where $\phi_\alpha^k(w)$ are the elements of $\boldsymbol{\Phi}^k(w)$. The coefficients $a_{k\beta}$ are obtained from the following initial and boundary conditions $\forall\{\beta : r_\beta < 0\}$:

- $\tilde{H}_{\alpha\beta}(0, w) = \begin{cases} 1 & \text{if} \quad \alpha = \beta, \\ 0 & \text{if} \quad \alpha \neq \beta \quad \text{and} \quad r_\alpha < 0. \end{cases}$
- $\tilde{H}_{\alpha\beta}(B^{\max}, w) = \sum\limits_{m \neq \alpha} \frac{q_{\alpha m}}{-q_{\alpha\alpha} + w} \tilde{H}_{m\beta}(B^{\max}, w), \quad \forall\{\alpha : r_\alpha > 0\}.$

From the above boundary conditions we can obtain the full expression for $\tilde{H}_{\alpha\beta}(x, w)$ from which we can obtain

$$\tilde{K}_\alpha(x, w) = \sum_{\beta=1}^{S} \tilde{H}_{\alpha\beta}(x, w), \quad \forall \alpha \in \mathcal{S}. \tag{6.33}$$

Finally to obtain $K_\alpha(x,t)$ as in (6.6) we need to numerically invert the transformed solution $\tilde{K}_\alpha(x,w)$.

Summarizing the LST method, first one needs to solve the eigenvalue problem in (6.31), symbolically or numerically. Then one solves the system of linear equation (6.32) for obtaining the coefficients $a_{k\beta}$. Finally, for applying the result to practical purposes, one must numerically invert the LST to obtain the solution $K_\alpha(x,t)$ in the time domain.

6.5 Application and Results

The Markov-modulated fluid queues described in Section 6.2 can be applied to micro-grids powered with wind energy. In Chapter 5 we modeled wind farm power output with discrete-time hidden Markov models (HMM). We consider a continuous-time version (because time is a continuous variable) of the HMM for the application of Markov-modulated fluid queues to micro-grids. We found in Chapter 5 that for modeling wind power accurately we need a very large state space, i.e., large value of S. Hence, in this section we focus on approximating the first passage time distribution of the buffer being empty as defined in (6.3) for large state space systems (large values of S) using the PDE schemes discussed in the previous sections.

We present results of various cases to compare the numerical integration schemes presented in Sections 6.3.1 and 6.3.2. We also compare the numerical integration schemes with Monte Carlo (MC) simulations. Finally, we compare results with the LST method described in Section 6.4 (albeit for small S).

6.5.1 Results for first order upwind

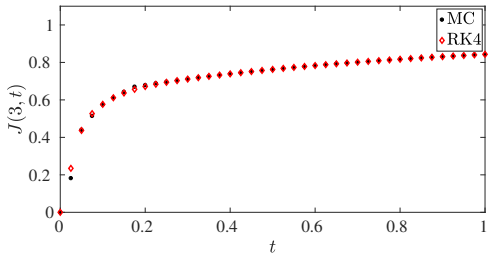
In this section we show results obtained from the PDE scheme discussed in Section 6.3.1 and compare with MC simulations. We perform the simulation on a Intel(R) Core(TM) i7-4500U CPU @ 1.80GHz computer. In

this section we denote the first order upwind scheme coupled with RK4 time integration as 1UP-RK4 and first order upwind scheme coupled with BDF2 time integration as 1UP-BDF2.

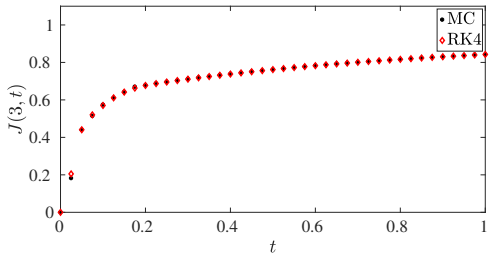
6.5.1.1 $S = 100$

We start with results for number of states $S = 100$. We generate a $\boldsymbol{Q}$ matrix for a birth-death process with $q_{\alpha\alpha} = -5,\ \forall\alpha \in \mathcal{S}$, $q_{\alpha,\alpha\pm1} = 2.5,\ \forall\alpha \in \mathcal{S}/\{1, 100\}$ and periodic boundary. The rates $\{r_\alpha\},\ \forall\alpha \in \mathcal{S}$ are randomly sampled from a normal distribution, i.e. $r_\alpha \sim \mathcal{N}(-50, 100^2), \forall\alpha \in \mathcal{S}$. We take $B^{\max} = 5$. First we compare $J(3, t)$ calculated from the PDE scheme with MC simulations. For this case we use the RK4 time integration scheme described in Section 6.3.1.1. We also compare how the space and time discretization steps Δx and Δt affect the solution. Note that from the CFL conditions in (6.12) Δx and Δt have to be modified in tandem. We take $\Delta t = 0.25 \times \min_{\alpha\in\mathcal{S}} \left|\frac{\Delta x}{r_\alpha}\right|$ such that the CFL condition is satisfied. In Figure 6.1 we plot the values of $J(3, t)$ for the RK4 method and compare with MC simulation results. For MC simulation we simulate the buffer model described in (6.1) starting the buffer at $B(0) = x = 3$ until $t = 1$ for 1000 samples and then compute $J(3, t)$ using the MC probability estimator described in (2.10) of Chapter 2. The relative error on the MC estimates is of the order 0.03 for 1000 samples.

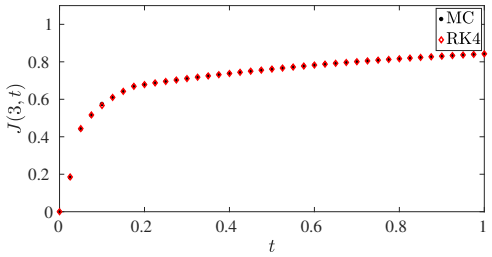
Notice from Figure 6.1 as Δx becomes smaller the 1UP-RK4 solution gets more aligned with the MC solution. The smaller the space and time discretization steps are, the better the solution becomes however it comes with a computational effort as shown in Table 6.1 for $t \in [0, 1]$. Comparing the CPU time of the MC simulation with the PDE scheme is not straightforward. This is because the PDE scheme gives us results for $J(x, t)$ for different values of initial buffer fluid content $x \in [0, B^{\max}]$ at intervals of Δx, on the other hand MC simulation computes $J(x, t)$ for a single value of x. MC simulation took 52 seconds for computing $J(3, t)$ for the same values of t as of 1UP-RK4 scheme.



(a) $\Delta x = 0.5$



(b) $\Delta x = 0.1$



(c) $\Delta x = 0.01$

Figure 6.1: Comparing 1UP-RK4 scheme with MC simulations for different Δx values for $S = 100$.

Δx	Δt	CPU time (s)
0.5	1.8×10^{-4}	0.07
0.1	8.9×10^{-5}	0.67
0.01	8.3×10^{-6}	160

Table 6.1: Comparing CPU time for different Δx for the system described in Section 6.5.1.1 for $t \in [0, 1]$ and $S = 100$.

6.5.1.2 $S = 200$

In this section we present results for $S = 200$; all other parameters are kept the same as in Section 6.5.1.1. We compare the values of $J(3, t)$ computed from the 1UP-RK4 scheme with MC simulation estimates. From Figure 6.2 we can see that for $\Delta x = 0.1$ the 1UP-RK4 results match closely with MC simulation results. As seen in Section 6.5.1.1 smaller values of Δx will lead to better alignment of results obtained from the PDE scheme with

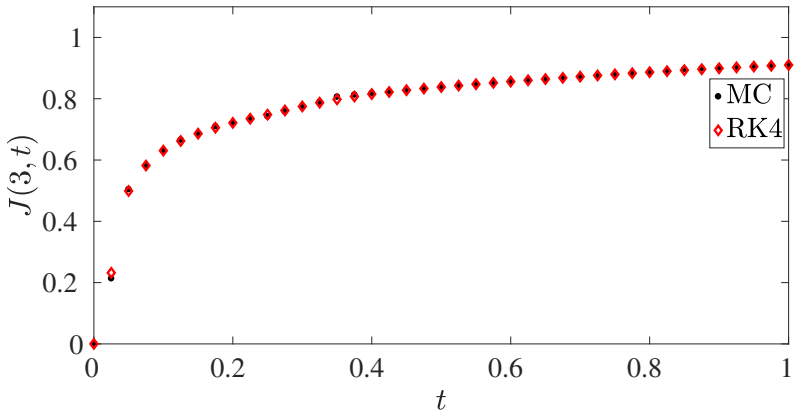


Figure 6.2: Comparing $J(3,t)$ computed from the 1UP-RK4 scheme (with $\Delta x = 0.1$) and MC simulations for $S = 200$.

S	Δt	CPU time (s)
200	6.8×10^{-5}	3.3
1000	1.3×10^{-4}	50

Table 6.2: Comparing CPU time for different S for the systems described in Sections 6.5.1.2 and 6.5.1.3 with $t \in [0,1]$ and $\Delta x = 0.1$.

MC simulation estimates.

Table 6.2 tabulates the CPU time for estimating $J(x,t)$ for $t \in [0,1]$ and $x \in [0,5]$ for the 1UP-RK4 method. The MC simulation took 112 seconds for computing $J(3,t)$ for the same values of t as of the 1UP-RK4 scheme.

As we are also interested in finding x_{min}, i.e., the initial starting level of the buffer so that the probability of it running empty before time t is less than p_{empty} (see (6.4)), using a numerical integration scheme readily gives us the numerical solution to the problem. In Figure 6.3 we plot the values of $J(x,t)$ (color bar) for different x and t. From the plot x_{min} can be estimated given p_{empty} and t.

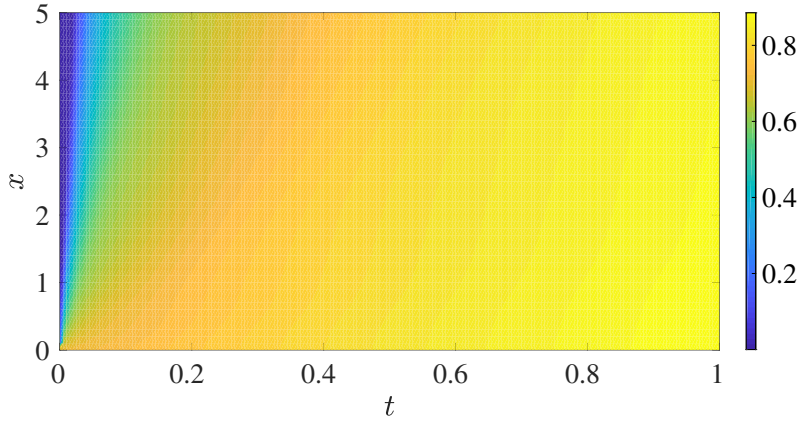


Figure 6.3: Values of $J(x,t)$ calculated using 1UP-RK4 scheme with $\Delta x = 0.1$ and $\Delta t = 6.8 \times 10^{-5}$ for $S = 200$.

6.5.1.3 $S = 1000$

In this section we present results for $S = 1000$. We take $\Delta x = 0.1$ and keep the other variables the same as Section 6.5.1.1. In Figure 6.4 we compare the values of $J(3,t)$ with $t \in [0,1]$ obtained from 1UP-RK4 scheme with MC simulation estimates and find that the solutions computed from both the methods are in good agreement with each other. In Figure 6.5 we plot the values of $J(x,t)$ (color bar) for different x and t. Table 6.2 tabulates the CPU time for estimating $J(x,t)$ for $t \in [0,1]$ and $x \in [0,5]$ for the RK4 scheme. MC simulation took 630 seconds to compute $J(3,t)$ for the same values of $t \in [0,1]$ as of the 1UP-RK4 scheme.

6.5.1.4 Solving stiff problems

In the previous cases, the values of the rates were such that $\Delta t \sim 10^{-4}$. We create stiffness in the PDE by setting one of the rates r_α to -10^4 and keeping all the other rates similar as in Section 6.5.1.1. Furthermore, we take $S = 100$, $t \in [0,1]$, $\Delta x = 0.1$ and $\Delta t = 5 \times 10^{-6}$, satisfying the CFL condition, for the RK4 scheme. This time step is much smaller than was used in Section 6.5.1.1, due to stiffness. It makes the RK4

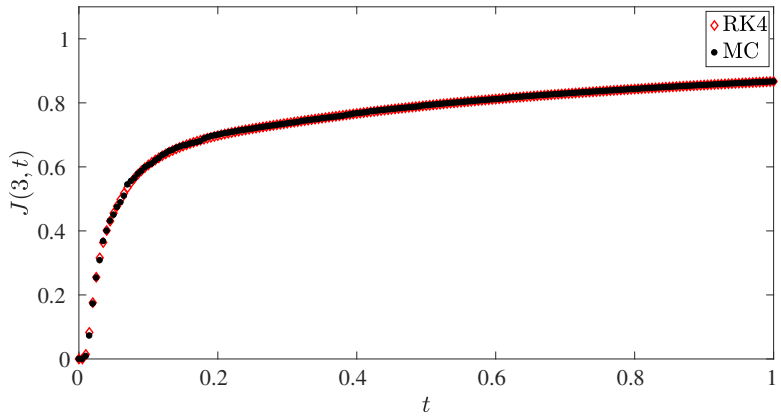


Figure 6.4: Comparing $J(3,t)$ estimated from the 1UP-RK4 scheme (with $\Delta x = 0.1$) and MC simulations for $S = 1000$.

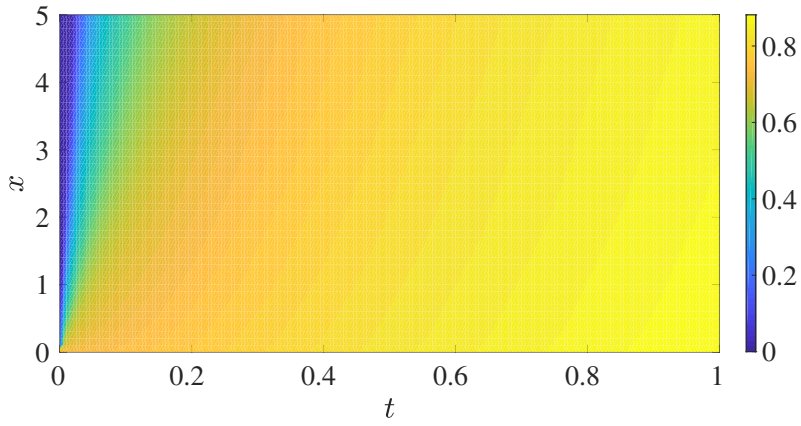


Figure 6.5: Values of $J(x,t)$ computed using the 1UP-RK4 scheme with $\Delta x = 0.1$ and $\Delta t = 6.8 \times 10^{-5}$ for $S = 1000$.

scheme computationally intensive, since the time interval $t \in [0,1]$ of interest remains unchanged. We use the implicit BDF2 scheme (discussed in Section 6.3.1.1), which allows us to take larger time steps.

We compare RK4 and BDF2 schemes in this section. We compare the results for $J(3,t)$ for both the schemes and with MC simulation results. For the choice of Δt for BDF2 scheme, we compare the sum squares error of RK4 scheme with $\Delta x = 0.1$ and MC simulation estimates for $J(3,t)$

Time integration scheme	Δt	Sum of squares error	CPU time (s)
BDF2	1×10^{-3}	0.0034	15
RK4	5×10^{-6}	0.0035	48

Table 6.3: Comparing CPU time for BDF2 and 1UP-RK4 time integration schemes with $S = 100$, $\Delta x = 0.1$, $B^{\max} = 5$ and $t \in [0, 1]$ for the stiff problem.

with the sum of squares error of BDF2 scheme for different Δt and MC simulation estimates for $J(3, t)$. We define the sum of squares error for both the schemes as

$$E^{RK4} := \sum_{t_q^{RK4}=1}^{m^{RK4}} (J^{MC}(3, t_q^{RK4}) - J^{RK4}(3, t_q^{RK4}))^2, \quad \text{and} \tag{6.34}$$

$$E^{BDF2} := \sum_{t_q^{BDF2}=1}^{m^{BDF2}} (J^{MC}(3, t_q^{BDF2}) - J^{BDF2}(3, t_q^{BDF2}))^2, \tag{6.35}$$

respectively. In the above equation t_q^{RK4} is the q-th time integration point of the RK4 scheme, m^{RK4} is the total number of time integration points of the RK4 scheme, $J^{MC}(3, t_q^{RK4})$ is the MC estimate for $x = 3$ and $t = t_q^{RK4}$. $J^{RK4}(3, t_q^{RK4})$ is computed by RK4 scheme for $x = 3$ and $t = t_q^{RK4}$. The variables with superscript BDF2 are computed by the BDF2 scheme.

In Table 6.3 we compare CPU time needed to compute $J(x, t)$ for both time integration schemes for similar sum of squares error, and in Figure 6.6 we compare $J(3, t)$ for these schemes. In Figure 6.6 we see that the BDF2 and RK4 schemes give accurate results when compared to the benchmark MC simulations. However, RK4 is slower than BDF2, as can be seen in Table 6.3.

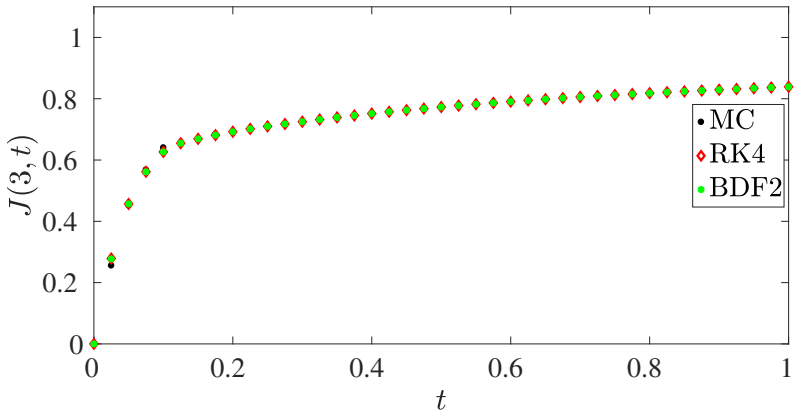


Figure 6.6: Comparing BDF2, 1UP-RK4 and MC results for $S = 100$, $\Delta x = 0.1$, $B^{\max} = 5$ and $t \in [0, 1]$ for the stiff problem.

6.5.2 Results for second order upwind

Here, we compare $J(x, t)$ computed using the flux-limited second order upwind scheme (see Section 6.3.2) with MC simulation. Note that in this section we will denote flux-limited second order upwind scheme coupled with RK3b time integration scheme as 2FLUP-RK3b. We keep all the parameters same as in Section 6.5.1.1. We take $\Delta x = 0.1$ and $S = 100$. In Figure 6.7 we compare values of $J(3.05, t)$ obtained from 2FLUP-RK3b with MC simulation results. We present results for $x = 3.05$ in this section due to the cell-centered grid discretization (see Figure 6.2). From Figure 6.7 we can see that the results computed from 2FLUP-RK3b are in good agreement with MC simulation results.

6.5.2.1 Different parameter settings

We presented results of $J(x, t)$ in the previous sections using the same parameters as described in Section 6.5.1.1. We used a $\boldsymbol{Q}$ matrix for a birth-death process which is sparse in nature. Therefore in this section and the next we present results for non-sparse $\boldsymbol{Q}$ matrix. We generate a random $\boldsymbol{Q}$ matrix (instead of a birth-death process) for the CTMC.

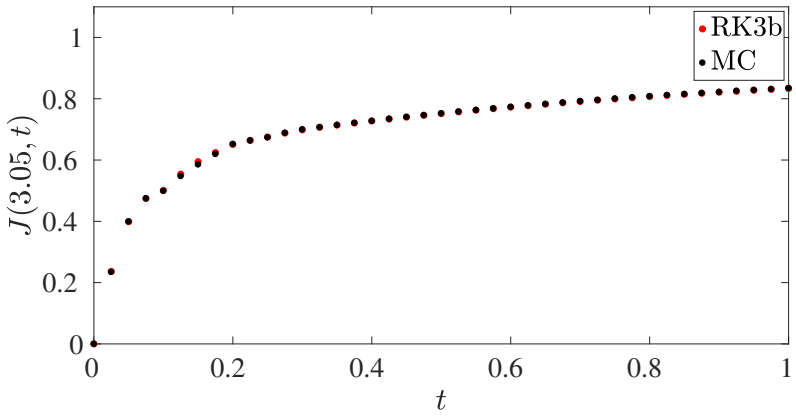


Figure 6.7: Comparing values of $J(3.05, t)$ obtained from 2FLUP-RK3b for $\Delta x = 0.1$ and $S = 100$ with MC simulation results.

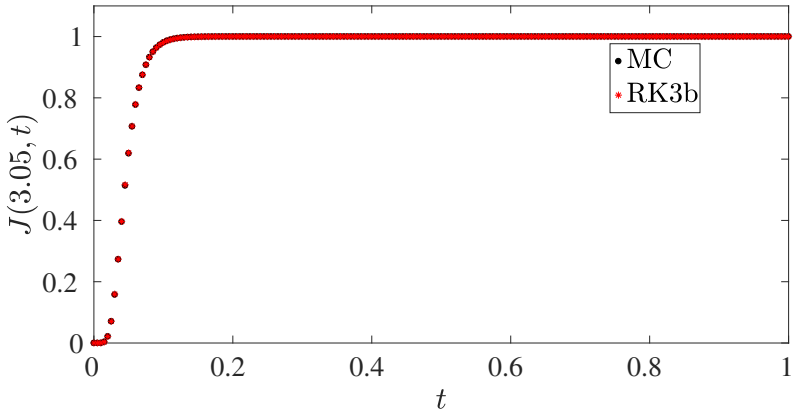


Figure 6.8: Comparing values of obtained from 2FLUP-RK3b for $\Delta x = 0.1$, $\Delta t = 6 \times 10^{-4}$ and $S = 100$ with MC simulation results for a non-sparse $\boldsymbol{Q}$ matrix.

The non-diagonal elements of the $\boldsymbol{Q}$ matrix are uniformly sampled from random integers between $[0, 10]$. We keep the other parameter the same as in Section 6.5.1.1. In Figure 6.8 we take $\Delta x = 0.1$ units and present results for $J(3.05, t)$. From Figure 6.8 we can see that the results computed from 2FLUP-RK3b are well aligned with MC simulation results.

S	Scheme	Δx	Sum of squares error	CPU time (s)
2	1UP-RK4	0.05	0.092	0.15
	2FLUP-RK3b	0.5	0.087	0.29
50	1UP-RK4	0.05	0.001	20
	2FLUP-RK3b	0.5	6.6×10^{-4}	22
100	1UP-RK4	0.05	0.0041	118
	2FLUP-RK3b	0.5	0.0044	47
200	1UP-RK4	0.05	0.0075	583
	2FLUP-RK3b	0.5	0.0079	94

Table 6.4: Comparing CPU time for different S for first order upwind scheme and second order upwind scheme.

6.5.2.2 Comparing 1UP-RK4 and 2FLUP-RK3b schemes

In this section we compare CPU time of 1UP-RK4 scheme described in Section 6.3.1 with 2FLUP-RK3b scheme described in Section 6.3.2.

The error for first order upwind space discretization scheme is $\mathcal{O}(\Delta x)$ and for second order upwind space discretization scheme is $\mathcal{O}(\Delta x^2)$. In order to compare the CPU time of first order scheme with the second order upwind scheme, we fix Δx for the first order scheme and vary Δx for the second order scheme such that the sum of squares error (see equation (6.34)) of both the schemes compared to MC simulation are close to each other. We calculate the sum of squares error at $x = 3.25$ for this comparison.

The time discretization step size Δt for both the schemes is computed by $\Delta t = 0.25 \times \min_{\alpha \in \mathcal{S}} \left| \frac{\Delta x}{r_\alpha} \right|$, which satisfies the CFL condition in (6.12). We keep the other parameters same as described Section 6.5.2.1 with non-sparse $\boldsymbol{Q}$ matrix. We solve numerically for $J(x,t)$ (see (6.3)) using both the discretization schemes for increasing number of states S.
It can be seen from Table 6.4 that as the number of states of the system

S increases the first order upwind scheme becomes more expensive and implementing second order upwind scheme for achieving the same order of error in the solution pays off. Since Δx for the first order scheme is 10 times smaller than for the second order upwind scheme, we have 10 times finer space and time discretization grid for the first order upwind scheme compared to the second order upwind scheme.

6.5.3 Comparing results with LST method

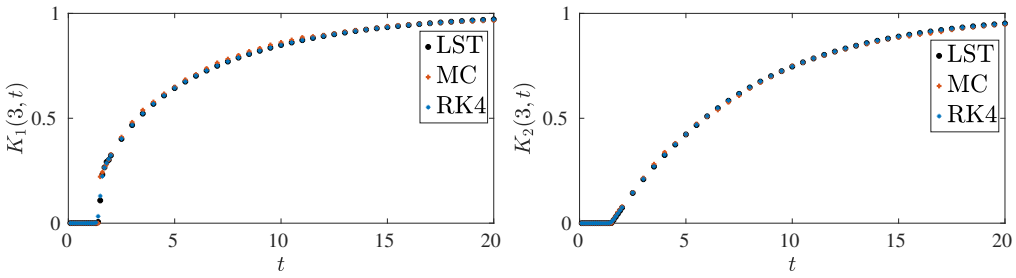
In this section we compare values of $K_\alpha(x,t)\ \forall \alpha \in \mathcal{S}$ (see (6.6)) obtained from LST method described in Section 6.4 with 1UP-RK4 (described in Section 6.3.1) as well as with MC simulations. For the LST method we first obtain the analytical expressions for $\tilde{K}_\alpha(x,w)$ (see (6.33)) using the symbolic toolbox of MATLAB and then numerically invert the transforms using the method described in [158] to compute $K_\alpha(x,t)$.

6.5.3.1 Two state system, $S = 2$

We take $S = 2$, $\boldsymbol{R} = \texttt{diag}(-2, 1)$ and $B^{\max} = 5$. We take the generator matrix $\boldsymbol{Q}$ for a birth-death process as explained in Section 6.5.1.1. The space discretization step is set to $\Delta x = 0.01$ and the time discretization step to $\Delta t = 0.0025$. In Figure 6.9 we present results for $B(0) = x = 3$ and find that all three methods are in good agreement with one another.

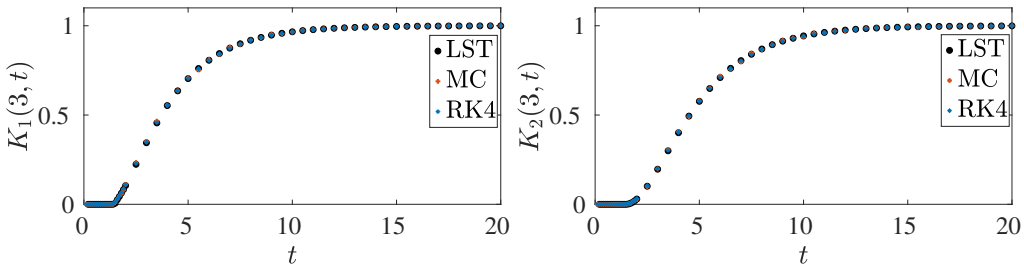
6.5.3.2 Three state system, $S = 3$

In this section, we take $S = 3$, $\boldsymbol{R} = \texttt{diag}(-2, 1, -0.5)$, and $B^{\max} = 5$. Following the previous section, we use $\Delta x = 0.01$ and $\Delta t = 0.0025$. From Figure 6.10 we find the results obtained from the three methods are in good agreement to each other.
We find that for a small system with only two states, i.e., $S = 2$ the LST method is computationally not very expensive. However for $S = 3$ our implementation of the LST method already becomes computationally

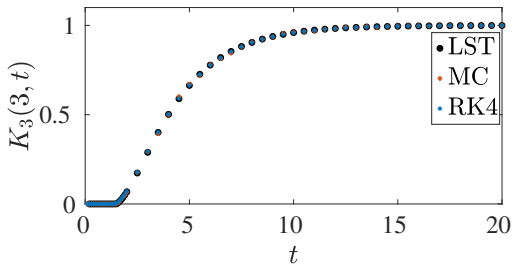


(a) $K_1(3,t)$ (b) $K_2(3,t)$

Figure 6.9: Comparing the LST and PDE integration methods with MC simulations for $S = 2$ and $x = 3$.



(a) $K_1(3,t)$ (b) $K_2(3,t)$



(c) $K_3(3,t)$

Figure 6.10: Comparing the LST and PDE integration methods with MC simulations for $S = 3$ and $x = 3$.

very intensive. It should be noted that we implemented a straightforward method for solving the problem using the LST method. First we solve the eigenvalue problem in (6.31) symbolically and then solve the system

of linear equations for obtaining the coefficients in (6.32) symbolically. Finally, we numerically invert the transforms. It is worthwhile to explore whether the computations needed for the LST method (i.e., solving the generalized eigenvalue problem in (6.31) and numerically inverting the LST) can be carried out more efficiently. This exploration is however beyond the scope of this chapter.

6.6 Conclusion

In this chapter we proposed PDE schemes to numerically compute the first passage time distribution of a Markov-modulated fluid queue model. We solved the one-dimensional coupled advection-reaction PDEs associated with this problem using upwind space discretization schemes. To achieve higher order accuracy in the space discretization we deployed the flux limited second order upwind scheme. We compare the results of these numerical schemes with MC simulation results.

For time integration of the PDE we used both explicit and implicit schemes. We deployed explicit Runge-Kutta methods as well as the implicit backward difference formula method for stiff problems. We studied the computational efficiencies of these schemes. We also found that the larger the system becomes (i.e., the larger S becomes), the second order upwind scheme becomes computationally more efficient compared to the first order upwind scheme. We find that results obtained with the PDE schemes are in close agreement with the benchmark MC simulation results, as well as with the results obtained (for small S) with the LST method.

Appendix

6.A Space discretization matrix for first order upwind scheme

The matrix $\boldsymbol{A}$ in (6.10) can be written as $\boldsymbol{A} = \tilde{\boldsymbol{R}} + \tilde{\boldsymbol{Q}} + \bar{\boldsymbol{Q}}$, in which $\tilde{\boldsymbol{R}}$ is a block diagonal matrix given by $\tilde{\boldsymbol{R}} = \texttt{blkdiag}(\tilde{\boldsymbol{R}}_1, \ldots, \tilde{\boldsymbol{R}}_S)$, where

$$\tilde{\boldsymbol{R}}_\alpha = \frac{r_\alpha}{\Delta x} \begin{bmatrix} 0 & 0 & 0 & 0 & \ldots \\ -1 & 1 & 0 & 0 & \ldots \\ 0 & -1 & 1 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad \forall \{\alpha : r_\alpha < 0\} \qquad \text{and,}$$

$$\tilde{\boldsymbol{R}}_\alpha = \frac{r_\alpha}{\Delta x} \begin{bmatrix} -1 & 1 & 0 & 0 & \ldots \\ 0 & -1 & 1 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & \ldots & 0 \end{bmatrix} \quad \forall \{\alpha : r_\alpha > 0\}.$$

Furthermore, $\tilde{\boldsymbol{Q}} = \boldsymbol{Q} \otimes \boldsymbol{I}_{q\times q}$, where $\boldsymbol{I}_{q\times q}$ is the identity matrix of size $q \times q$. The matrix $\bar{\boldsymbol{Q}}$ contains the discretization of the boundary conditions for $K_\alpha^{n,q}$ for $\{\alpha : r_\alpha > 0\}$ and has dimensions equal to $\tilde{\boldsymbol{Q}}$. The elements of the matrix are given by $\bar{Q}_{n\alpha,n\beta} = \tilde{Q}_{n\alpha,n\beta}$ for $\{\alpha : r_\alpha > 0\}$ and $\beta = 1, \ldots, S$. All the other elements of matrix $\bar{\boldsymbol{Q}}$ are zero.

6.B Flux limited second order upwind scheme

We present the details of the flux limited second order upwind scheme here. From (6.19) it can be seen that flux values for faces $F_\alpha^{3/2}$ and $F_\alpha^{n+1/2}$, $\forall\{\alpha : r_\alpha \leq 0\}$ cannot be defined by (6.19). Similarly, flux values for faces $F_\alpha^{1/2}$ and $F_\alpha^{n-1/2}$, $\forall\{\alpha : r_\alpha > 0\}$ cannot be defined by (6.19). Hence we define the fluxes for all the faces as follows:

- $\boldsymbol{\forall\{\alpha : r_\alpha \leq 0\}}$
  1. $F_\alpha^{1/2}$ is given by the boundary condition in Section 6.2.1.
  2. $F_\alpha^{3/2} = \frac{1}{2}(K_\alpha^1 + K_\alpha^2)$, i.e, the central scheme.
  3. $F_\alpha^{p+1/2}$ for $p = 2, \ldots, n-1$ is given by (6.19).
  4. $F_\alpha^{n+1/2} = (\frac{3}{2}K_\alpha^n - \frac{1}{2}K_\alpha^{n-1})$, i.e., the second order upwind scheme without limiters.

- $\boldsymbol{\forall\{\alpha : r_\alpha > 0\}}$
  1. $F_\alpha^{1/2} = (\frac{3}{2}K_\alpha^1 - \frac{1}{2}K_\alpha^2)$.
  2. $F_\alpha^{p+1/2}$ for $p = 1, \ldots, n-2$ is given by (6.19).
  3. $F_\alpha^{n-1/2} = \frac{1}{2}(K_\alpha^n + K_\alpha^{n-1})$, i.e, the central scheme.
  4. $F_\alpha^{n+1/2}$ is given by the boundary condition in Section 6.2.1.

Note that in (6.19),the expression $\phi_{p+1/2} = \phi(f_{p+1/2})$ was still left unspecified. As a convenience we repeat here how we specified them later in section 6.3.2.3. As limiter ϕ we chose the Koren limited (6.24), and $f_{p+1/2}$

is defined as:

$$f_{p+1/2} = \begin{cases} \dfrac{K_\alpha^{p+1} - K_\alpha^p}{K_\alpha^p - K_\alpha^{p-1}} & \forall\{\alpha : r_\alpha \leq 0\}, \\ \dfrac{K_\alpha^{p-1} - K_\alpha^p}{K_\alpha^p - K_\alpha^{p+1}} & \forall\{\alpha : r_\alpha > 0\}. \end{cases} \tag{6.36}$$

Bibliography

[1] EDSN. (2014). Hourly comsumption profile of a typical household. http://www.edsn.nl/verbruiksprofielen/. [Online; accessed 12-February-2015].

[2] European Network of Transmission System Operators for Electricity (ENTSOE). (2004). Continental Europe Operation Handbook. Technical report.

[3] IEEE Guide for Electric Power Distribution Reliability Indices. (2012). *IEEE Std 1366-2012 (Revision of IEEE Std 1366-2003)*, pages 1–43.

[4] International Energy Agency (IEA) and Organisation for Economic Co-operation and Development (OECD). (2011). *Harnessing Variable Renewables: A Guide to the Balancing Challenge*. ISBN 9789264111387.

[5] KNMI.(2013). Hourly data of the weather in the Netherlands. http://www.knmi.nl/klimatologie/uurgegevens/selectie.cgi. [Online; accessed 25-September-2014].

[6] North American Electric Reliability Council (NERC). (1996). Glossary of Terms. https://naesb.org//pdf/weq_glossary072804w10.pdf. [Online; accessed 15-November-2017].

[7] Renewable Energy Policy Network for the 21st Century. (2017). Renewables 2017 Global Status Report. Technical report. Online access: http://www.ren21.net/wp-content/uploads/2017/06/17-8399_GSR_2017_Full_Report_0621_Opt.pdf,.

[8] United Nations. Adoption of the Paris Agreement. (2015). http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf. [Online; accessed 21-September-2017].

[9] United Nations Treaty Collection. (2015). Paris Agreement. https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-d&chapter=27&clang=_en. [Online; accessed 21-September-2017].

[10] Abdullah, M. A., Agalgaonkar, A., and Muttaqi, K. M. (2013). Probabilistic load flow incorporating correlation between time-varying electricity demand and renewable power generation. *Renewable energy*, 55:532–543.

[11] Abur, A. and Expósito, A. (2004). *Power System State Estimation: Theory and Implementation*. Power Engineering (Willis). Taylor & Francis. ISBN 9780824755706.

[12] Aburub, H. and Jewell, W. T. (2014). Optimal generation planning to improve storage cost and system conditions. In *2014 IEEE Power and Energy Society General Meeting Conference & Exposition*, pages 1–5. IEEE.

[13] Albadi, M. and El-Saadany, E. (2010). Overview of wind power intermittency impacts on power systems. *Electric Power Systems Research*, 80(6):627–632.

[14] Alhumaizi, K. (2007). Flux-limiting solution techniques for simulation of reaction–diffusion–convection system. *Communications in Nonlinear Science and Numerical Simulation*, 12(6):953–965.

[15] Amrein, M. and Künsch, H. R. (2011). A variant of importance splitting for rare event estimation: Fixed number of successes. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 21(2):13.

[16] Anick, D., Mitra, D., and Sondhi, M. M. (1982). Stochastic theory of a data-handling system with multiple sources. *Bell Labs Technical Journal*, 61(8):1871–1894.

[17] Appunn, K. Connecting up the energiewende. *Clean Energy Wire*. https://www.cleanenergywire.org/dossiers/energy-transition-and-germanys-power-grid. [Online; accessed 10-November-2017].

[18] Asmussen, S. (1995). Stationary distributions for fluid flow models with or without Brownian noise. *Communications in statistics. Stochastic models*, 11(1):21–49.

[19] Asmussen, S., Avram, F., and Usabel, M. (2002). Erlangian approximations for finite-horizon ruin probabilities. *ASTIN Bulletin: The Journal of the IAA*, 32(2):267–281.

[20] Asmussen, S. and Bladt, M. (1994). A sample path approach to mean busy periods for Markov-modulated queues and fluids. *Advances in applied probability*, 26(4):1117–1121.

[21] Badwawi, R. A., Abusara, M., and Mallick, T. (2015). A review of hybrid solar PV and wind energy system. *Smart Science*, 3(3):127–138.

[22] Barbot, N., Sericola, B., and Telek, M. (2001). Distribution of busy period in stochastic fluid models. *Stochastic Models*, 17(4):407–427.

[23] Bargès, M., Loisel, S., and Venel, X. (2013). On finite-time ruin probabilities with reinsurance cycles influenced by large claims. *Scandinavian Actuarial Journal*, 2013(3):163–185.

[24] Barton, J. P. and Infield, D. G. (2004). Energy storage and its use with intermittent renewable energy. *IEEE Transactions on Energy Conversion*, 19(2):441–448.

[25] Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171.

[26] Bejan, A. I., Gibbens, R. J., and Kelly, F. P. (2012). Statistical aspects of storage systems modelling in energy networks. In *46th Annual Conference on Information Sciences and Systems*, pages 1–6. IEEE.

[27] Bhaumik, D., Crommelin, D., Kapodistria, S., and Zwart, B. (2017). Hidden Markov models for wind farm power output. Submitted for publication.

[28] Bhaumik, D., Crommelin, D., and Zwart, B. (2016). A computational method for optimizing storage placement to maximize power network reliability. In *Proceedings of the 2016 Winter Simulation Conference*, pages 883–894. IEEE Press.

[29] Bhaumik, D., Crommelin, D., and Zwart, B. (2018a). Mitigation of large power spills by an energy storage device in a stand alone energy system. *Journal of Energy Storage*, 16:76–83.

[30] Bhaumik, D., Crommelin, D., Zwart, B., Boon, M., and Koren, B. (2018b). Computing first passage times for Markov-moduated fluid model using PDE solvers. Manuscript to be submitted for publication.

[31] Bienstock, D., Chertkov, M., and Harnett, S. (2014). Chance-constrained optimal power flow: risk-aware network control under uncertainty. *SIAM Review*, 56(3):461–495.

[32] Billinton, R. and Allan, R. N. (1984). Power-system reliability in perspective. *Electronics and Power*, 30(3):231–236.

[33] Billinton, R., Chen, H., and Ghajar, R. (1996). Time-series models for reliability evaluation of power systems including wind energy. *Microelectronics Reliability*, 36(9):1253–1261.

[34] Billinton, R., Cui, Y., et al. (2003). Reliability evaluation of small stand-alone wind energy conversion systems using a time series simulation model. *IEE Proceedings-Generation, Transmission and Distribution*, 150(1):96–100.

[35] Billinton, R. and Wangdee, W. (2007). Reliability-based transmission reinforcement planning associated with large-scale wind farms. *IEEE Transactions on Power Systems*, 22(1):34–41.

[36] Billinton, R. and Wenyuan, L. (2013). *Reliability assessment of electric power systems using Monte Carlo methods*. Language of science. Springer Science & Business Media. ISBN 9781489913463.

[37] Bilmes, J. A. (1998). A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute*, 4(510):126.

[38] Bishop, C. (1995). *Neural Networks for Pattern Recognition*. Advanced Texts in Econometrics. Clarendon Press. 9780198538646.

[39] Blöchle, M., Zucker, G., Xypolytou, E., Leber, T., Prüggler, W., and Rührlinger, T. (2015). Optimizing neighborhood consumption of renewables through clustering and H2 storage: An economic assessment of an Austrian community. In *12th International Conference on the European Energy Market*, pages 1–5. IEEE.

[40] Bose, S., Gayme, D. F., Topcu, U., and Chandy, K. M. (2012). Optimal placement of energy storage in the grid. In *51stIEEE Annual Conference on Decision and Control*, pages 5605–5612. IEEE.

[41] Bosman, J. W. (2014). *Optimal Quality of Service Control in Communication Systems*. PhD thesis, Vrije Universiteit Amsterdam.

[42] Bosman, J. W., van der Mei, R. D., and Núñez-Queija, R. (2012). A fluid model analysis of streaming media in the presence of time-varying bandwidth. In *Proceedings of the 24th International Teletraffic Congress*, page 31. International Teletraffic Congress.

[43] Boxma, O. J. and Dumas, V. (1998). The busy period in the fluid queue. In *ACM SIGMETRICS Performance Evaluation Review*, volume 26, pages 100–110. ACM.

[44] Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2):261–304.

[45] Cabral, C. V. T., Oliveira Filho, D., Diniz, A. S. A. C., Martins, J. H., Toledo, O. M., and Lauro de Vilhena, B. (2010). A stochastic method for stand-alone photovoltaic system sizing. *Solar Energy*, 84(9):1628–1636.

[46] Cain, M. B., OâĂŹneill, R. P., and Castillo, A. (2012). History of optimal power flow and formulations. *Federal Energy Regulatory Commission*, pages 1–36.

[47] Cakmur, R., Miller, R., and Torres, O. (2004). Incorporating the effect of small-scale circulations upon dust emission in an atmospheric general circulation model. *Journal of Geophysical Research: Atmospheres*, 109(D7).

[48] Casares, F., Lopez-Luque, R., Posadillo, R., and Varo-Martinez, M. (2014). Mathematical approach to the characterization of daily energy balance in autonomous photovoltaic solar systems. *Energy*, 72:393–404.

[49] Chandy, K. M., Low, S. H., Topcu, U., and Xu, H. (2010). A simple optimal power flow model with energy storage. In *49th IEEE Conference on Decision and Control*, pages 1051–1057. IEEE.

[50] Chen, P., Siano, P., Bak-Jensen, B., and Chen, Z. (2010). Stochastic optimization of wind turbine power factor using stochastic model of wind power. *IEEE Transactions on Sustainable Energy*, 1(1):19–29.

[51] Clement-Nyns, K., Haesen, E., and Driesen, J. (2010). The impact of charging plug-in hybrid electric vehicles on a residential distribution grid. *IEEE Transactions on Power Systems*, 25(1):371–380.

[52] Dall'Anese, E., Dhople, S. V., and Giannakis, G. B. (2014). Optimal dispatch of photovoltaic inverters in residential distribution systems. *IEEE Transactions on Sustainable Energy*, 5(2):487–497.

[53] Das, J. (2002). *Power System Analysis: Short-Circuit Load Flow and Harmonics*. Power Engineering (Willis). CRC Press. ISBN 9780203908952.

[54] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, pages 1–38.

[55] Denholm, P. (2012). Energy storage to reduce renewable energy curtailment. In *IEEE Power and Energy Society General Meeting*, pages 1–4. IEEE.

[56] Denholm, P., Ela, E., Kirby, B., and Milligan, M. (2010). The role of energy storage with renewable electricity generation. Technical Report TP-6A2-47187, National Renewable Energy Laboratory.

[57] Dharmakeerthi, C., Mithulananthan, N., and Saha, T. (2014). Impact of electric vehicle fast charging on power system voltage stability. *International Journal of Electrical Power & Energy Systems*, 57:241–249.

[58] Dobakhshari, A. S. and Fotuhi-Firuzabad, M. (2009). A reliability model of large wind farms for power system adequacy studies. *IEEE Transactions on Energy Conversion*, 24(3):792–801.

[59] Ekren, O. and Ekren, B. Y. (2010). Size optimization of a PV/wind hybrid energy conversion system with battery storage using simulated annealing. *Applied Energy*, 87(2):592–598.

[60] Elwalid, A. I. and Mitra, D. (1991). Analysis and design of rate-based congestion control of high speed networks, i: stochastic fluid models, access regulation. *Queueing Systems*, 9(1):29–63.

[61] Eremia, M. and Shahidehpour, M. (2013). *Handbook of Electrical Power System Dynamics: Modeling, Stability, and Control*. IEEE Press Series on Power Engineering. John Wiley & Sons. ISBN 9781118516065.

[62] Feijoo, A. and Villanueva, D. (2017). Wind farm power distribution function considering wake effects. *IEEE Transactions on Power Systems*, 32(4):3313–3314.

[63] Field, A. and Harrison, P. G. (2010). Busy periods in fluid queues with multiple emptying input states. *Journal of Applied Probability*, 47(2):474–497.

[64] Fink, S., Mudd, C., Porter, K., and Morgenstern, B. (2009). Wind energy curtailment case studies. *NREL subcontract report, NREL/SR-550*, 46716.

[65] Fragaki, A. and Markvart, T. (2013). System memory effects in the sizing of stand-alone PV systems. *Progress in Photovoltaics: Research and Applications*, 21(4):724–735.

[66] Franzke, C., Crommelin, D., Fischer, A., and Majda, A. J. (2008). A hidden Markov model perspective on regimes and metastability in atmospheric flows. *Journal of Climate*, 21(8):1740–1757.

[67] Freund, J. and Perles, B. (2007). *Modern Elementary Statistics*. Pearson Prentice Hall. ISBN 9780131874398.

[68] Garvels, M. J., Van Ommeren, J.-K. C., and Kroese, D. P. (2002). On the importance function in splitting simulation. *Transactions on Emerging Telecommunications Technologies*, 13(4):363–371.

[69] Garvels, M. J. J. (2000). *The splitting method in rare event simulation*. PhD thesis, Universiteit Twente.

[70] Gayme, D. and Topcu, U. (2013). Optimal power flow with large-scale storage integration. *IEEE Transactions on Power Systems*, 28(2):709–717.

[71] Ghahramani, Z. and Jordan, M. I. (1994). Supervised learning from incomplete data via an EM approach. *Advances in Neural Information Processing Systems*, pages 120–120.

[72] Ghofrani, M., Arabali, A., Etezadi-Amoli, M., and Fadali, M. S. (2013). A framework for optimal placement of energy storage units within a power system with high wind penetration. *IEEE Transactions on Sustainable Energy*, 4(2):434–442.

[73] Gilbert, J. R. and Peierls, T. (1988). Sparse partial pivoting in time proportional to arithmetic operations. *SIAM journal on scientific and statistical computing*, 9(5):862–874.

[74] Giorsetto, P. and Utsurogi, K. F. (1983). Development of a new procedure for reliability modeling of wind turbine generators. *IEEE Transactions on Power Apparatus and Systems*, (1):134–143.

[75] Golden, R. and Paulos, B. (2015). Curtailment of renewable energy in California and beyond. *The Electricity Journal*, 28(6):36–50.

[76] Gong, B., Wang, X., Dang, C., Xiao, Y., Wu, Z., Xie, X., and Cui, Q. (2016). The wind model based on Markov process to determine the capacity for wind farm. In *IEEE PES Asia-Pacific Power and Energy Engineering Conference*, pages 2096–2100. IEEE.

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

[78] Green, R. C., Wang, L., and Alam, M. (2011). The impact of plug-in hybrid electric vehicles on distribution networks: A review and outlook. *Renewable and Sustainable Energy Reviews*, 15(1):544–553.

[79] Hocaoğlu, F. O., Gerek, Ö. N., and Kurban, M. (2010). A novel wind speed modeling approach using atmospheric pressure observations and hidden Markov models. *Journal of Wind Engineering and Industrial Aerodynamics*, 98(8):472–481.

[80] Horenko, I., Dittmer, E., Fischer, A., and Schütte, C. (2006). Automated model reduction for complex systems exhibiting metastability. *Multiscale Modeling & Simulation*, 5(3):802–827.

[81] Horváth, G. and Telek, M. (2016). Fitting methods based on distance measures of marked markov arrival processes. In *Seminal Contributions to Modelling and Simulation*, pages 159–183. Springer.

[82] Hosseini, S. H., Tang, C. Y., and Jiang, J. N. (2014). Calibration of a wind farm wind speed model with incomplete wind data. *IEEE Transactions on Sustainable Energy*, 5(1):343–350.

[83] Hu, P., Karki, R., and Billinton, R. (2009). Reliability evaluation of generating systems containing wind power and energy storage. *IET Generation, Transmission & Distribution*, 3(8):783–791.

[84] Hundsdorfer, W., Koren, B., Verwer, J., et al. (1995). A positive finite-difference advection scheme. *Journal of Computational Physics*, 117(1):35–46.

[85] Hundsdorfer, W. and Verwer, J. (2003). *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*. Springer Series in Computational Mathematics. Springer Science & Business Media. ISBN 9783662090176.

[86] Jafarzadeh, S., Fadali, S., Evrenosoglu, C. Y., and Livani, H. (2010). Hour-ahead wind power prediction for power systems using hidden Markov models and Viterbi algorithm. In *IEEE Power and Energy Society General Meeting*, pages 1–6. IEEE.

[87] Johnson, G. L. (2006). *Wind energy systems*. Kansas State University. Electronic edition of the book is available online https://www.rpc.com.au/information/faq/wind-power/wind-energy-systems.html.

[88] Jones, G. L., Harrison, P. G., Harder, U., and Field, T. (2011). Fluid queue models of battery life. In *19th IEEE International Symposium on Modeling, Analysis & Simulation of Computer and Telecommunication Systems*, pages 278–285. IEEE.

[89] Jordan, M. I. and Jacobs, R. A. (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, 6(2):181–214.

[90] Jordan, M. I. and Xu, L. (1995). Convergence results for the EM approach to mixtures of experts architectures. *Neural Networks*, 8(9):1409–1431.

[91] Karatzas, I. and Shreve, S. (2012). *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer Science & Business Media. ISBN 9781468403022.

[92] Karki, R. and Billinton, R. (2001). Reliability/cost implications of PV and wind energy utilization in small isolated power systems. *IEEE Transactions on Energy Conversion*, 16(4):368–373.

[93] Karki, R., Hu, P., and Billinton, R. (2006). A simplified wind power generation model for reliability evaluation. *IEEE Transactions on Energy Conversion*, 21(2):533–540.

[94] Khodayar, M., Kaynak, O., and Khodayar, M. E. (2017). Rough deep neural architecture for short-term wind speed forecasting. *IEEE Transactions on Industrial Informatics*.

[95] Klein, S. and Beckman, W. (1987). Loss-of-load probabilities for stand-alone photovoltaic systems. *Solar Energy*, 39(6):499–512.

[96] Kolumbán, S., Kapodistria, S., and Nooraee, N. (2017). Short and long-term wind turbine power output prediction. *arXiv preprint arXiv:1707.06497*.

[97] Koren, B. (1993). Numerical methods for advection-diffusion problems. chapter A robust upwind discretization method for advection, diffusion and source terms, pages 117–138. Vieweg.

[98] Kotsiantis, S. and Kanellopoulos, D. (2006). Discretization techniques: A recent survey. *GESTS International Transactions on Computer Science and Engineering*, 32(1):47–58.

[99] Kulkarni, V. G. (1997). Frontiers in queueing. chapter Fluid Models for Single Buffer Systems, pages 321–338. CRC Press, Inc. ISBN 0-8493-8076-6.

[100] Kulkarni, V. G. and Tzenova, E. (2002). Mean first passage times in fluid queues. *Operations Research Letters*, 30(5):308–318.

[101] Kumar, R., Liu, Y., and Ross, K. (2007). Stochastic fluid theory for P2P streaming systems. In *IEEE INFOCOM 2007 - 26th IEEE*

International Conference on Computer Communications, pages 919–927. IEEE.

[102] Kundur, P. (1994). *Power System Stability and Control*. McGraw-Hill, Inc. ISBN 007035958X.

[103] Lambert, J. (1991). *Numerical Methods for Ordinary Differential Systems: The Initial Value Problem*. Wiley. ISBN 9780471929901.

[104] Le, D. D., Gross, G., and Berizzi, A. (2015). Probabilistic modeling of multisite wind farm production for scenario-based applications. *IEEE Transactions on Sustainable Energy*, 6(3):748–758.

[105] Lee, D. and Baldick, R. (2012). Limiting ramp rate of wind power output using a battery based on the variance gamma process. In *Proceedings International Conference on Renewable Energies and Power Quality (ICREPQ 2012)*.

[106] LeVeque, R. (2002). *Finite Volume Methods for Hyperbolic Problems*. Cambridge Texts in Applied Mathematics. Cambridge University Press. ISBN 9781139434188.

[107] Li, C.-H., Zhu, X.-J., Cao, G.-Y., Sui, S., and Hu, M.-R. (2009). Dynamic modeling and sizing optimization of stand-alone photovoltaic power systems using hybrid energy storage technology. *Renewable Energy*, 34(3):815–826.

[108] Li, P., Guan, X., Wu, J., and Zhou, X. (2015). Modeling dynamic spatial correlations of geographically distributed wind farms and constructing ellipsoidal uncertainty sets for optimization-based generation scheduling. *IEEE Transactions on Sustainable Energy*, 6(4):1594–1605.

[109] Lin, J., Cheng, L., Chang, Y., Zhang, K., Shu, B., and Liu, G. (2014). Reliability based power systems planning and operation with wind power integration: A review to models, algorithms and applications. *Renewable and Sustainable Energy Reviews*, 31:921–934.

[110] Lydia, M., Kumar, S. S., Selvakumar, A. I., and Kumar, G. E. P. (2014). A comprehensive review on wind turbine power curve modeling techniques. *Renewable and Sustainable Energy Reviews*, 30:452–460.

[111] Mahela, O. P. and Shaik, A. G. (2016). Comprehensive overview of grid interfaced wind energy generation systems. *Renewable and Sustainable Energy Reviews*, 57:260–281.

[112] Mahmud, N. and Zahedi, A. (2016). Review of control strategies for voltage regulation of the smart distribution network with high penetration of renewable distributed generation. *Renewable and Sustainable Energy Reviews*, 64:582–595.

[113] Markvart, T. (1996). Sizing of hybrid photovoltaic-wind energy systems. *Solar Energy*, 57(4):277–281.

[114] Mitra, D. (1987). Stochastic fluid models. In *Proceedings of the 12th IFIP WG 7.3 International Symposium on Computer Performance Modelling, Measurement and Evaluation*, pages 39–51. North-Holland Publishing Co.

[115] Mitra, D. (1988). Stochastic theory of a fluid model of producers and consumers coupled by a buffer. *Advances in Applied Probability*, 20(3):646–676.

[116] Monahan, A. H. (2006). The probability distribution of sea surface wind speeds. Part I: Theory and SeaWinds observations. *Journal of Climate*, 19(4):497–520.

[117] Moran, P. (1956). A probability theory of a dam with a continuous release. *The Quarterly Journal of Mathematics*, 7(1):130–137.

[118] Munteanu, I., Bratcu, A., Cutululis, N., and Ceanga, E. (2008). *Optimal Control of Wind Energy Systems: Towards a Global Approach*. Advances in Industrial Control. Springer Science & Business Media. ISBN 9781848000803.

[119] Narayanan, A. and Kulkarni, V. (1996). First passage times in fluid models with an application to two priority fluid systems. In *Proceedings of IEEE International Computer Performance and Dependability Symposium*, pages 166–175. IEEE.

[120] Negra, N. B., Holmstrøm, O., Bak-Jensen, B., and Sørensen, P. (2008). Model of a synthetic wind speed time series generator. *Wind Energy*, 11(2):193–209.

[121] Nesti, T., Nair, J., and Zwart, B. (2016). Reliability of dc power grids under uncertainty: a large deviations approach. *arXiv preprint arXiv:1606.02986*.

[122] Newell, C. (2013). *Applications of Queueing Theory*. Ettore Majorana International Science Series. Springer Science & Business Media. ISBN 9789400959705.

[123] Norris, J. (1998). *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. SBN 9780521633963.

[124] Oh, H. (2011). Optimal planning to include storage devices in power systems. *IEEE Transactions on Power Systems*, 26(3):1118–1128.

[125] Osborn, J. and Kawann, C. (2001). Reliability of the U.S. electricity system: Recent trends and current issues. Technical Report LBNL-47043, Energy Analysis Department, Ernest Orlando Lawrence Berkeley National Laboratory, CA.

[126] Papaefthymiou, G. and Klockl, B. (2008). MCMC for wind power simulation. *IEEE Transactions on Energy Conversion*, 23(1):234–240.

[127] Papaefthymiou, G. and Kurowicka, D. (2009). Using copulas for modeling stochastic dependence in power system uncertainty analysis. *IEEE Transactions on Power Systems*, 24(1):40–49.

[128] Prasad, A. R. and Natarajan, E. (2006). Optimization of integrated photovoltaic–wind power generation systems with battery storage. *Energy*, 31(12):1943–1954.

[129] Qureshi, A. S., Khan, A., Zameer, A., and Usman, A. (2017). Wind power prediction using deep neural network based meta regression and transfer learning. *Applied Soft Computing*.

[130] Rabiner, L. and Juang, B. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3(1):4–16.

[131] Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

[132] Rebours, Y. G., Kirschen, D. S., Trotignon, M., and Rossignol, S. (2007). A survey of frequency and voltage control ancillary services - part i: Technical features. *IEEE Transactions on power systems*, 22(1):350–357.

[133] Ren, Q. and Kobayashi, H. (1995). Transient solutions for the buffer behavior in statistical multiplexing. *Performance evaluation*, 23(1):65–87.

[134] Rogers, L. (1994). Fluid models in queueing theory and Wiener-Hopf factorization of Markov chains. *The Annals of Applied Probability*, pages 390–413.

[135] Rubino, G. and Tuffin, B. (2009). *Rare Event Simulation using Monte Carlo Methods*. Wiley. ISBN 9780470745410.

[136] Rubinstein, R. and Kroese, D. (2016). *Simulation and the Monte Carlo Method*. Wiley Series in Probability and Statistics. Wiley. ISBN 9781118632208.

[137] Safari, B. and Gasore, J. (2010). A statistical investigation of wind characteristics and wind energy potential based on the Weibull and Rayleigh models in Rwanda. *Renewable Energy*, 35(12):2874–2880.

[138] Sansavini, G., Piccinelli, R., Golea, L., and Zio, E. (2014). A stochastic framework for uncertainty analysis in electric power transmission systems with wind generation. *Renewable Energy*, 64:71–81.

[139] Sayas, F. C. and Allan, R. (1996). Generation availability assessment of wind farms. *IEE Proceedings-Generation, Transmission and Distribution*, 143(5):507–518.

[140] Scheinhardt, W., Van Foreest, N., and Mandjes, M. (2005). Continuous feedback fluid queues. *Operations Research Letters*, 33(6):551–559.

[141] Scheinhardt, W. R. W. (1998). *Markov-modulated and feedback fluid queues*. PhD thesis, Universiteit Twente.

[142] Schwartz, M. (1996). *Broadband Integrated Networks*. Prentice Hall PTR. ISBN 9780135192405.

[143] Semaoui, S., Arab, A. H., Bacha, S., and Azoui, B. (2013). Optimal sizing of a stand-alone photovoltaic system with energy management in isolated areas. *Energy Procedia*, 36:358–368.

[144] Shafiee, M. and Sørensen, J. D. (2017). Maintenance optimization and inspection planning of wind energy assets: Models, methods and strategies. *Reliability Engineering & System Safety*.

[145] Shafiullah, G., Oo, A. M., Ali, A. S., and Wolfs, P. (2013). Potential challenges of integrating large-scale wind energy into the power grid–a review. *Renewable and Sustainable Energy Reviews*, 20:306–321.

[146] Shakib, A. D. and Balzer, G. (2010). Energy storage design and optimization for power system with wind feeding. In *11th IEEE Interna-*

tional Conference on Probabilistic Methods Applied to Power Systems, pages 54–59. IEEE.

[147] Shortle, J., Rebennack, S., and Glover, F. W. (2014). Transmission-capacity expansion for minimizing blackout probabilities. *IEEE Transactions on Power Systems*, 29(1):43–52.

[148] Siano, P. (2014). Demand response and smart grids - a survey. *Renewable and Sustainable Energy Reviews*, 30:461–478.

[149] Simonian, A. and Guibert, J. (1995). Large deviations approximation for fluid queues fed by a large number of on/off sources. *IEEE Journal on Selected Areas in Communications*, 13(6):1017–1027.

[150] Sjödin, E., Gayme, D. F., and Topcu, U. (2012). Risk-mitigated optimal power flow for wind powered grids. In *American Control Conference*, pages 4431–4437. IEEE.

[151] Smith, G. (1985). *Numerical solution of partial differential equations: finite difference methods*. Oxford Applied Mathematics and. Clarendon Press. ISBN 9780198596509.

[152] Stanford, D. A., Latouche, G., Woolford, D. G., Boychuk, D., and Hunchak, A. (2005). Erlangized fluid queues with application to uncontrolled fire perimeter. *Stochastic models*, 21(2-3):631–642.

[153] Stenclik, D., Denholm, P., and Chalamala, B. (2017). Maintaining balance: The increasing role of energy storage for renewable integration. *IEEE Power and Energy Magazine*, 15(6):31–39.

[154] Suberu, M. Y., Mustafa, M. W., and Bashir, N. (2014). Energy storage systems for renewable energy power sector integration and mitigation of intermittency. *Renewable and Sustainable Energy Reviews*, 35:499–514.

[155] Süli, E. and Mayers, D. (2003). *An Introduction to Numerical Analysis*. Cambridge University Press. ISBN 9780521810265.

[156] Sweby, P. K. (1984). High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM Journal on Numerical Analysis*, 21(5):995–1011.

[157] Ton, D. T. and Smith, M. A. (2012). The US Department of Energy's microgrid initiative. *The Electricity Journal*, 25(8):84–94.

[158] Valsa, J. and Brančik, L. (1998). Approximate formulae for numerical inversion of Laplace transforms. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 11(3):153–166.

[159] van den Akker, J., Leemhuis, S., and Bloemhof, G. (2014). Optimizing storage placement in electricity distribution networks. In *Operations Research Proceedings 2012*, pages 183–188. Springer.

[160] van Laarhoven, P. and Aarts, E. (1987). *Simulated Annealing: Theory and Applications*. Mathematics and Its Applications. Springer Science & Business Media. ISBN 9789027725134.

[161] Varaiya, P. P., Wu, F. F., and Bialek, J. W. (2011). Smart operation of smart grid: Risk-limiting dispatch. *Proceedings of the IEEE*, 99(1):40–57.

[162] Vardakas, J. S., Zorba, N., and Verikoukis, C. V. (2015). A survey on demand response programs in smart grids: Pricing methods and optimization algorithms. *IEEE Communications Surveys & Tutorials*, 17(1):152–178.

[163] Wadman, W. (2015). *Assessing power grid reliability using rare event simulation*. PhD thesis, University van Amsterdam.

[164] Wadman, W., Bloemhof, G., Crommelin, D., and Frank, J. (2012). Probabilistic power flow simulation allowing temporary current overloading. In *Proceedings of the International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 494–499.

[165] Wadman, W., Crommelin, D., and Frank, J. (2013). Applying a splitting technique to estimate electrical grid reliability. In *Winter Simulation Conference*, pages 577–588. IEEE.

[166] Wadman, W. S., Crommelin, D. T., and Zwart, B. P. (2016). A large-deviation-based splitting estimation of power flow reliability. *ACM Transactions on Modeling and Computer Simulation*, 26(4):23.

[167] Wang, C. and Nehrir, M. H. (2008). Power management of a standalone wind/photovoltaic/fuel cell energy system. *IEEE Transactions on Energy Conversion*, 23(3):957–967.

[168] Waterson, N. P. and Deconinck, H. (2007). Design principles for bounded higher-order convection schemes–a unified approach. *Journal of Computational Physics*, 224(1):182–207.

[169] Wu, C. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, pages 95–103.

[170] Wu, D., Liu, Y., and Ross, K. (2009). Queuing network models for multi-channel p2p live streaming systems. In *IEEE INFOCOM 2009*, pages 73–81. IEEE.

[171] Xie, K., Miao, S., Xia, Y., Ma, Y., and Li, Y. (2016). A two-stage wind speed model for multiple wind farms considering autocorrelations and cross-correlations. In *International Conference on Probabilistic Methods Applied to Power Systems*, pages 1–6. IEEE.

[172] Xu, H., Topcu, U., Low, S. H., Clarke, C. R., and Chandy, K. M. (2010). Load-shedding probabilities with hybrid renewable power gen-

eration and energy storage. In *48th Annual Allerton Conference on Communication, Control, and Computing*, pages 233–239. IEEE.

[173] Xu, L. and Jordan, M. I. (1996). On convergence properties of the EM algorithm for gaussian mixtures. *Neural Computation*, 8(1):129–151.

[174] Yang, H., Lu, L., and Zhou, W. (2007). A novel optimization sizing model for hybrid solar-wind power generation system. *Solar Energy*, 81(1):76–84.

[175] Yang, H., Zhou, W., Lu, L., and Fang, Z. (2008). Optimal sizing method for stand-alone hybrid solar–wind system with lpsp technology by using genetic algorithm. *Solar Energy*, 82(4):354–367.

[176] Yau, T., Walker, L. N., Graham, H. L., Gupta, A., and Raithel, R. (1981). Effects of battery storage devices on power system dispatch. *IEEE Transactions on Power Apparatus and Systems*, (1):375–383.

[177] Ye, Z., EL-Azouzi, R., Jimenez, T., and Xu, Y. (2014). Computing Quality of Experience of Video Streaming in Network with Long-Range-Dependent Traffic. *arXiv preprint arXiv:1412.2600*.

[178] Zeng, J. and Qiao, W. (2012). Short-term wind power prediction using a wavelet support vector machine. *IEEE Transactions on Sustainable Energy*, 3(2):255–264.

[179] Zhang, Y., Wang, J., and Wang, X. (2014). Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews*, 32:255–270.

[180] Zhao, H., Wu, Q., Hu, S., Xu, H., and Rasmussen, C. N. (2015). Review of energy storage system for wind power integration support. *Applied Energy*, 137:545–553.

[181] Zhou, W., Lou, C., Li, Z., Lu, L., and Yang, H. (2010). Current status of research on optimum sizing of stand-alone hybrid solar–wind power generation systems. *Applied Energy*, 87(2):380–389.

[182] Zimmerman, R. D., Murillo-Sánchez, C. E., and Thomas, R. J. (2011). MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education. *IEEE Transactions on Power Systems*, 26(1):12–19.

List of Publications

- *Mitigation of large power spills by an energy storage device in a stand alone energy system.* **D. Bhaumik**, D. Crommelin and B. Zwart. Journal of Energy Storage, 16:76-83, Elsevier, 2018

  Chapter 3 is based on this research article. The research work, numerical simulations and writing the article were performed by D. Bhaumik. The research work was carried out under the supervision of B. Zwart and D. Crommelin.

- *A computational method for optimizing storage placement to maximize power network reliability.* **D. Bhaumik**, D. Crommelin and B. Zwart. In Proceedings of the 2016 Winter Simulation Conference, IEEE Press, Piscataway, NJ, USA, 883-894, 2016.

  Chapter 4 is based on this research article. The research work, numerical simulations and writing the article were performed by D. Bhaumik. The research work was carried out under the supervision of B. Zwart and D. Crommelin.

- *Hidden Markov Models for Wind Farm Power Output.* **D. Bhaumik**, D. Crommelin, S. Kapodistria and B. Zwart. Submitted for publication, 2017.

  Chapter 5 is based on this research article. The research work,

numerical computations and writing the article were performed by D. Bhaumik. The wind power data was provided by S. Kapodistria (Eindhoven University of Technology), who also contributed to research discussions and editing of the article. The research work was carried out under the supervision of D. Crommelin and B. Zwart.

- *Computing first passage times for Markov-modulated fluid model using PDE schemes.* **D. Bhaumik**, D. Crommelin, B. Zwart, M. Boon, and B. Koren. Manuscript in progress, 2018.

  Chapter 6 is based on this research article in progress. The research work, numerical computations for the PDE schemes and writing the manuscript have been carried out by D. Bhaumik. Further computations for the LST transforms will be done by M. Boon (Eindhoven University of Technology). B. Koren (Eindhoven University of Technology and Centrum Wiskunde & Informatica) provided us with insights on the second-order upwind scheme. The research work was carried out under the supervision of D. Crommelin and B. Zwart.