Skoltech
Skolkovo Institute of Science and Technology

Skolkovo Institute of Science and Technology

Data-Driven Stochastic AC-OPF using Gaussian Processes

Doctoral Thesis
by
Mile Mitrovic

DOCTORAL PROGRAM IN ENGINEERING SYSTEMS

Supervisor:
Assistant Professor Elena Gryazina

Co-advisor:
Assistant Professor Petr Vorobev

Moscow - 2023

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology (Moscow) and has not been submitted for any other degree.

Candidate: Mile Mitrovic
Supervisors: Assistant Professor Elena Gryazina

Abstract

The thesis focuses on developing a data-driven algorithm, based on machine learning, to solve the stochastic alternating current (AC) chance-constrained (CC) Optimal Power Flow (OPF) problem. Although the AC CC-OPF problem has been successful in academic circles, it is highly nonlinear and computationally demanding, which limits its practical impact. The proposed approach aims to address this limitation and demonstrate its empirical efficiency through applications to multiple IEEE test cases.

To solve the non-convex and computationally challenging CC AC-OPF problem, the proposed approach relies on a machine learning Gaussian process regression (GPR) model. The full Gaussian process (GP) approach is capable of learning a simple yet non-convex data-driven approximation to the AC power flow equations that can incorporate uncertain inputs. The proposed approach uses various approximations for GP-uncertainty propagation. The full GP CC-OPF approach exhibits highly competitive and promising results, outperforming the state-of-the-art sample-based chance constraint approaches.

To further improve the robustness and complexity/accuracy trade-off of the full GP CC-OPF, a fast data-driven setup is proposed. This setup relies on the sparse and hybrid Gaussian processes (GP) framework to model the power flow equations with input uncertainty.

Personal References

[1] M. Mitrovic, A. Lukashevich, P. Vorobev, V. Terzija, Y. Maximov, and D. Deka. Fast data-driven chance constrained ac-opf using hybrid sparse gaussian processes. In *2023 IEEE Belgrade PowerTech*, pages 1–7. IEEE, 2023.

[2] M. Mitrovic, A. Lukashevich, P. Vorobev, V. Terzija, S. Budennyy, Y. Maximov, and D. Deka. Data-driven stochastic ac-opf using gaussian process regression. *International Journal of Electrical Power & Energy Systems*, 152:109249, 2023.

[3] M. Mitrovic, O. Kundacina, A. Lukashevich, S. Budennyy, P. Vorobev, V. Terzija, Y. Maximov, and D. Deka. Gp cc-opf: Gaussian process based optimization tool for chance-constrained optimal power flow. *Software Impacts*, 16:100489, 2023.

Non-thesis Personal References

[1] S. Asefi, M. Mitrovic, D. Cetenovic, V. Levi, E. Gryazina, V. Terzija. Anomaly detection and classification in power system state estimation: Combining model-based and data-driven methods. *Sustainable Energy, Grids and Networks*, 2023.

[2] S. Asefi, M. Mitrovic, D. Cetenovic, V. Levi, E. Gryazina, V. Terzija. Anomaly Detection, Classification and Identification Tool (ADCIT). *Software Impact*, 2023.

[3] M. Mitrovic, D. Titov, K. Volkhov, I. Lukicheva, A. Kudryavzev, P. Vorobev, V. Terzija. ML-based Method for Diagnostics of Overhead Transmission Line Insulator State by Leakage Current. *IEEE Transactions on Power Delivery*, 2023 - **Submitted**.

[4] R. Vorobyev, M. Mitrovic, I. Kremnev. Computer Vision & Machine Learning Techniques for Non-Destructive Testing of Composites. ECCM 2022 - Proceedings of the 20th European Conference on Composite Materials: Composites Meet Sustainability 4, pp. 520-525.

[5] O. Kundacina, G. Gojic, D. Miskovic, M. Mitrovic, and D. Vukobratovic. Supporting future electrical utilities: Using deep learning methods in EMS and DMS algorithms. 2023, 22nd International Symposium INFOTEH-JAHORINA (INFOTEH). IEEE, 2023, pp. 1–6.

Acknowledgments

I would like to convey my heartfelt appreciation to the members of my thesis jury, supervisors, colleagues, as well as my friends and family, for their invaluable support, for sharing ideas, and for unwavering encouragement throughout my PhD journey.

I would like to express my deepest gratitude to the members of my thesis jury for their invaluable time and dedication in reviewing and enhancing this dissertation: Henni Ouerdane, Andrei Osiptsov, Federico Martin Ibanez, Haoran Zhao, Ashok Kumar Pradhan and Alexander Nazin. Their commitment and insightful feedback have greatly contributed to the improvement of this work and have provided me with valuable ideas for future research.

I am especially thankful to my advisors, Elena Gryazina and Petr Vorobev, for their guidance and support throughout my PhD studies. Their expertise and mentorship have been instrumental in shaping the direction of my research.

Special thanks to Deepjyoty Deka and Yury Maximov for introducing me to the thesis topic and for their continuous encouragement to explore new fields and develop new skills.

I would like to express my profound appreciation to Prof. Vladimir Terzija for his support of my ideas and research directions. Our engaging discussions on research, science, and life have been both enlightening and inspiring. Additionally, I am grateful to Prof. Terzija for enabling me to connect with scholars from various universities worldwide.

I cannot thank Dmitry Titov and Dragan Cetenovic enough for their collaboration on a separate project during my PhD. Their assistance in writing papers and their unwavering technical support throughout my studies have been invaluable.

I would like to extend my heartfelt appreciation to my friends and colleagues from Skoltech and around the world who have been an integral part of my PhD journey. Their unwavering support has been instrumental in making this academic endeavor a fulfilling and memorable experience. I am immensely grateful to Strahinja and Kruna Markovic, Zeljko and Anja Tekic, Julijana Cvjetinovic, Dejan Dzunja, Ognjen Kundacina, Aleksandr Lukashevich, Sajjad Asefi, Nikolay Ivanov, Oleg Khamisov, Irina Lukicheva, Aleksandra Burashnikova, Sahar Moghimian, Anastasia Cumik, Artem Zabolotnyi, Roman Yusupov, Akshay Vishwanathan, Vito Michele Leli and many others who have provided their invaluable support and encouragement throughout my journey. Although I may not have listed everyone individually, I am deeply thankful to each and every person who has played a role in my PhD journey. Their presence, friendship, and shared experiences have created a nurturing and inspiring environment,

Contents

Nomenclature

Abbreviations

AC   Alternating Current

AGC   Automatic generation control

CC   Chance-constrained

DC   Direct Current

DS   Distribution System

DSOs Distribution System Operators

EM   Exact Moment Matching

ERM   Empirical Risk Minimization

GP   Gaussian Process

GPR   Gaussian Process Regression

KKT   Karush-Kuhn-Tucker conditions

MAE   Mean Absolute Error

MLE   Maximum Likelihood Estimate

MSE   Mean Squared Error

MSLE Mean Squared Logarithmic Error

OPF   Optimal Power Flow

P-OPF Probabilistic Optimal Power Flow

PF   Power Flow

QP   Quadratic programming

RES   Renewable energy sources

RMSE Root Mean Squared Error

RO   Robust optimization

SDP   Semidefinite programming problem

SE   Squared Exponential

SEard Squared Exponential Automatic Relevance Determination

SLSQP Sequential Least Squares Programming

SOCP Second-order cone programming

SRM   Structural Risk Minimization

TA   Taylor Approximation

TSOs Transmission System Operators

Notation

α Participation factor

$\bar{i}$ Complex current

$\bar{v}$ Complex voltage

$\bigtriangledown_{xx}L$ Hessian of Lagrangian system equations

ℓ_c Loss function

ϵ Violation probability

$\frac{\partial \cdot (x)}{\partial x}$ the partial derivative of function $\cdot$ evaluated at x

γ Power ratio

$\mathbb{D}$ Vector dimension $\mathcal{S}$

$\mathbb{E}$ Expectations

$\mathbb{P}$ Probabilities

$\mathcal{B}$ Set of all buses in power grid

$\mathcal{D}$ Population distribution

$\mathcal{E}$ Set of transmission lines

$\mathcal{F}$ Set of learning functions

$\mathcal{G}$ Set of conventional generator buses

$\mathcal{L}$	Set of load and RES buses
$\mathcal{N}(\cdot)$	Normal distribution
$\mathcal{R}$	Set of real numbers $\mathcal{S}$
$\mathcal{S}$	Sample distribution
Ω	Total power mismatch
ω	Uncertanty realizations
θ	Voltage angle
B	Susceptance matrix
c	Scalar cost coefficients
G	Conductance matrix
I	Identity matrix
J_{KKT}	Jacobian matrix of the KKT system
L	Lagrangian system equations
m	Overall number of buses
m_s	Number of samples in traing set $\mathcal{S}$
n	Overall number of transmission lines
p	Active power
q	Reactive power
R	Resistance matrix
$rank(\cdot)$	Rank of a matrix
s	Apparent power
s_{sl}	Slack variables
$tr(\cdot)$	Trace of a matrix
v	Voltage magnitude
X	Reactance matrix
x	Decision variables
x^{in}	Input sample vector of dataset

Y Admittance matrix

y Line admittance

y^{out} Output sample vector of dataset - label

z Line impendance

List of Tables

List of Figures

Chapter 1

General Introduction

The power grid is widely regarded as one of the most remarkable engineering accomplishments of the 20th century. It has played a key role in enabling economic prosperity and advancing social progress for billions of people worldwide. However, the task of managing and controlling the power grid is becoming increasingly complex for Transmission System Operators (TSOs). Despite the stabilization of the average annual total demand, new patterns of energy consumption and generation have emerged, posing new challenges for TSOs [Liu et al., 2012].

TSOs rely on Optimal Power Flow (OPF) as a fundamental tool to ensure secure and cost-effective power system operation, commonly used in electricity markets [Ng et al., 2018] and system security assessments [Capitanescu et al., 2011]. OPF is a mathematical optimization problem that determines the optimal settings for power generators, transformers, and other devices in the power system [Cain et al., 2012]. In this context, OPF can be seen as an economic dispatch problem, with the goal of determining how the grid should set generator outputs in real-time. This involves dispatching generators at regular intervals, usually every fifteen minutes to an hour (depending on the power grid), to balance demand and generator output at the lowest possible cost while respecting the operational limitations of the generators and transmission lines. By using OPF to optimize the generator output, TSOs can ensure a reliable and efficient power supply while minimizing costs.

In recent years, we have witnessed that conventional power grids have been undergoing a transformation into modern smart grids. This transition has resulted in the integration of various new facilities, including renewable energy sources (RES), electric cars (EC), and Internet of Things (IoT) devices. The use of RES like wind and solar plants introduces large uncertainty into the power grids, which can have unexpected consequences. Neglecting or underestimating the impact of these uncertainties can result in conservative operations, driving up operating costs. Moreover, aggressive operations will probably lead to constraint violations and thus jeopardize the security of the grid. For example, renewable outputs have the potential to produce power flows (PFs) that significantly exceed the ratings of power lines. Exceeded line ratings can cause grid instability and cascading failures that may ultimately result in a blackout. Moreover, there is a regulation that RES should cover 20% of all generations in the US and Europe by 2030 [CIGRE, 2009, EERE, 2008, DENA, 2005, Gonzalez et al.,

2006]. As a result, TSOs must strike a balance between operating costs and security. This fact leads to solving the OPF problem in a stochastic context.

The thesis aims to tackle the challenges outlined in the problem background above by formulating a more faster and robust solution method for addressing stochastic OPF problems. The primary research objectives of this thesis include:

1. developing an approach that strikes a trade-off between computational complexity and optimal solutions in solving stochastic OPF problems;

2. enhancing the overall efficiency of the solutions derived from the stochastic OPF approach.

Accordingly, the main research questions of this thesis are:

1. how can a trade-off between computational complexity and optimal solutions be achieved in the stochastic OPF problem?

2. how proposed approach can improve the efficiency of the stochastic OPF solution?

Stochastic OPF involves solving the OPF problem while taking uncertainties into account. Various approaches have been proposed in the literature to address this issue. Among them, the most popular are robust optimization (RO), probabilistic OPF (P-OPF), and chance-constrained (CC) approach. While RO ensures secure operations against all possible uncertainty realizations within a given set, it often leads to conservative solutions [Ben-Tal et al., 2009, Warrington et al., 2013]. P-OPF, on the other hand, is challenging to put into practice as it results in probabilistic distributions of control variables that can not lead to deterministic scheduling strategies [Ullah et al., 2022, Schellenberg et al., 2005, Zhang and Li, 2010]. The chance-constrained approach, however, ensures that chance constraints are satisfied within an acceptable violation probability [Du et al., 2021, Xiao et al., 2001, Zhang and Li, 2011, Sjödin et al., 2012, Vrakopoulou et al., 2013a, Roald et al., 2015, Vrakopoulou et al., 2013d, Bienstock et al., 2014, Roald et al., 2015, Morillo et al., 2022, Wu et al., 2019]. In general, the chance-constrained approach is a mathematical optimization technique used in decision-making under uncertainty. It means that this approach helps in making decisions that account for uncertain parameters or variables while maintaining control over the allowable level of risk or the probability of constraint violations. These uncertainties often stem from external factors like market fluctuations, varying weather conditions, or unpredictable changes in demand. Since the chance-constrained problem is typically applied to constrained optimization problems, decision-making is specified with the risk tolerance or acceptable level of constraint violation. Thus, the chance-constrained (CC) OPF approach enables TSOs to balance security and operating costs in an intuitive and transparent manner. Therefore, this work focuses on the CC-OPF approach.

While chance constraints offer a way to address uncertainty in a quantitative manner, solving the Alternating Current Chance-Constrained Optimal Power Flow (AC CC-OPF) is notoriously challenging [Roald and Andersson, 2017, Mühlpfordt et al., 2019]. Alternating current (AC) is a type of electric current that periodically

reverses direction within a circuit. Unlike direct current (DC), which flows steadily in one direction, AC changes direction periodically, typically in a sinusoidal waveform. To overcome the challenge of solving AC CC-OPF, many studies convert the stochastic optimization problem to a deterministic one. However, this reduces the confidence region of AC CC-OPF compared to the feasible region of AC OPF. Additionally, both regions are non-convex and difficult to deal with. To make the problem more manageable, researchers try to properly eliminate the nonlinear aspects and reformulate the stochastic optimization problem as a deterministic one, using convexification techniques such as linear approximation or convex relaxation. Thus, they provide the resulting convex optimization problem numerically tractable.

Although convexification techniques have been used to make the AC CC-OPF problem tractable to solve, the resulting solutions may not be optimal in terms of achieving the lowest possible cost. To address these challenges, we decided to investigate a new approach. Therefore, this study proposes a data-driven approach to replace the AC power flow (AC-PF) balance equations with a probabilistic approximation based on a supervised machine learning (ML) model. The proposed approach employs a Gaussian Process Regression (GPR) as the probabilistic supervised ML model. The GPR model is directly integrated into the CC-OPF formulation to create a new data-driven approach called Gaussian Process Chance-Constrained Optimal Power Flow (GP CC-OPF). This approach offers a novel way to model and solve stochastic OPF problems using Gaussian Processes (GP) and has the potential to provide more accurate solutions with reduced computational effort.

1.1 Contributions

This thesis aims to develop a robust and effective approach for solving stochastic CC-OPF problems that strikes a trade-off between solution accuracy and computational complexity. By achieving this, the proposed approach can enable the optimization of operating costs in the presence of uncertainty while maintaining the security of the power system. Therefore, we propose a novel GP-based data-driven approximation of the AC-PF balance equations, which is integrated into the CC-OPF formulation. Moreover, we consider two cases:

1. **full** GP CC-OPF: where GPR fully replaces AC-PF balance equations;

2. **hybrid** GP CC-OPF: that combines linear direct current (DC) PF balance equations with the data-driven estimation of the residuals between DC-PF and AC-PF based on GPR.

We account for both fluctuating loads and RES as input uncertainty variables. To additionally propagate input uncertainty to output variables we consider and compared different approximation techniques such as first and second-order Taylor Approximation (TA) and Exact Moment Matching (EM).

To ensure the scalability of the proposed GP-based approach, we use sparse GPR, which employs a few selective data samples for estimation. The practical efficiency

of the proposed approach is validated and illustrated using a number of standard IEEE test cases. Additionally, we compare the proposed data-driven GP CC-OPF approach with state-of-the-art sample-based CC-OPF approaches. Results show that the proposed GP-based reformulation of the CC-OPF is competitive and outperforms conventional sample-based formulations.

One of the key advantages of the GP-based approach is reflected in the fact that the proposed approach does not require knowledge of the grid configuration and parameters. This thesis can have significant importance for TSOs who seek to make informed decisions regarding cost-effective and secure system operation under uncertain conditions.

1.2 Thesis Structure

This thesis is divided into two main parts. The first part focuses on presenting the theory of optimal power flow and the Gaussian processes relevant to our research topic. In this section, we also provide an overview of the current state-of-the-art in the field.

The second part of the thesis will be dedicated to presenting our contributions to the field. Here, we will showcase our unique perspective and original research findings. The first part of this thesis consists of two chapters.

* In **chapter 2**, we introduce the fundamental concept of power systems and provide a detailed explanation of the OPF framework. Furthermore, we delve into recent approaches and methods that have been developed to solve OPF problems. Additionally, in this chapter, we describe how a synthetic dataset is generated using a simulated power system model.

* In **chapter 3**, we introduce the fundamental concepts of machine learning focusing on supervised learning and present the theoretical background of the Gaussian process regression model.

The second part of this thesis consists of three chapters.

* In **chapter 4**, we present a novel data-driven CC-OPF approach based on GPR. The GPR model is trained on a synthetic dataset and used to replace the full AC-PF balance equations in the CC-OPF approach. This approach allows for the propagation of input uncertainties to output variables, and we compare different approximation techniques for this. Additionally, we compare the results of the full GP CC-OPF approach with sample-based stochastic approaches.

* In **chapter 5**, we propose a more accurate, scalable and robust approach than in chapter 4. Specifically, we utilize a hybrid approach that combines the linear DC-PF and additive GP part. The GP part is learned on residuals between AC-PF and DC-PF, allowing us to better capture the complex relationships between the input and output variables of the power system. To further improve the computational efficiency of our approach, we incorporate a sparse GP. This involves using a small set of points to better approximate the marginal likelihood

and reduce the computational complexity, while still preserving important information from the original sample span.

∗ In **chapter 6**, we provide a brief explanation and documentation of the software design for this research.

Finally, in **chapter 7** we conclude our work and present directions for future work.

1.3 Corresponding papers

The contributions presented in this manuscript are built upon the following papers that were developed as part of the research conducted during this Ph.D. It is important to acknowledge that these papers were collaborative efforts involving numerous co-authors. However, it should be noted that the personal contribution to these papers encompasses all experimental aspects, with the exception of the scenario-based chance-constrained method for comparison that was performed by Aleksandr Lukashevich. Additionally, there was a partial contribution made to the theoretical parts, guided by the supervision of Elena Gryazina, Petr Vorobev, Yury Maximov and Deepjyoti Deka.

Chapter 4 is based on the paper [2] published at the *International Journal of Electrical Power & Energy Systems* (IJEPES 2023).

Chapter 5 is based on the paper [1] published at the *IEEE Belgrade PowerTech 2023 Conference* (PowerTech 2023).

Chapter 6 is based on the paper [3] published at the *Software Impact* journal (SIMPA 2023).

PART I

BACKGROUND THEORY

&

STATE-OF-THE-ART

Chapter 2

Power System Description and Optimal Power Flow

2.1 Introduction

Chapter 2 aims to provide readers with a comprehensive understanding of power systems and optimal power flow (OPF).

The first part introduces the fundamental concepts of power grids and the equations used to describe their behavior. The goal is to help readers develop intuition about these objects and concepts that are critical for modeling and simulating power grids.

The second part focuses on OPF and its mathematical concepts, highlighting its importance in power systems. Additionally, we discuss state-of-the-art solutions for solving OPF problems, along with some of the challenges involved.

In the third part, we provide readers with a brief introduction to sampling and generating a synthetic dataset used in the contribution part.

In summary, this chapter is structured into five sections. Section 2.2 introduces the main concepts of power systems, while section 2.3 presents an overview of OPF and its significance. Section 2.4 talks about generating a synthetic dataset from a simulated power system. Finally, section 2.5 provides a summary of the key points covered in the chapter. Overall, this chapter will provide readers with a solid foundation in power systems and OPF, setting the stage for more advanced discussions later in the contribution part (part II).

2.2 Power System Description

Generally, power systems can be divided into three main components: production (generation, supply), transmission, and distribution (load, consumption). A visual representation of this concept is presented in Fig. 2-1.
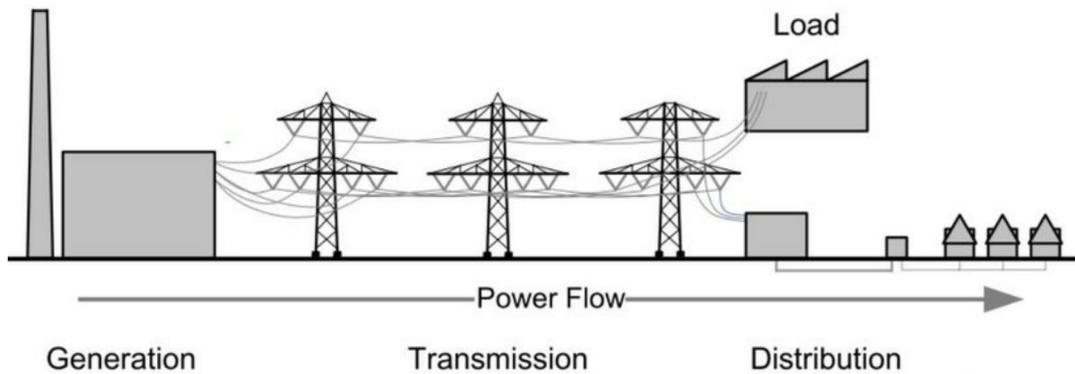
Figure 2-1: Fundamental concept of the power systems; the figure is taken from https://engineeringnotesonline.com.

Generation involves facilities that produce electricity and sends it to the power system. There are various sources of electricity production, including thermal power plants (coal, fuel, gas, or nuclear), hydropower plants, and renewable sources (wind or solar). In our discussion, we will consider thermal and hydropower plants as controlled conventional generation, while renewables will be treated as uncontrolled, uncertain generation.

Load refers to all facilities that consume electrical power. It is essential to note that the load is not just a single household, but rather a group of consumers, such as a small town or a large industrial firm. This assumption is made from the perspective of Transmission System Operators (TSOs), as this group of consumers is directly connected to the high-voltage transmission system. Typically, this group of consumers works on a low-voltage called the Distribution System (DS) operated by the Distribution System Operators (DSOs).

Transmission system contains high-voltage lines that serve as a connection between power generation and consumption. It is operated by TSOs whose responsibility is to ensure that consumers can access the required amount of power at any time and from any location. Additionally, TSOs are charged with maintaining the system's reliability and ensuring that consumers have a secure supply of electricity. This thesis will primarily focus on the high-voltage transmission system, which will be further described in this section.

In most developed countries, the energy market is structured so that there are three distinct entities: producers, TSOs, and DSOs. However, the transmission system is usually a state monopoly. This means that TSOs are responsible for ensuring that producers and consumers with their specific behavior have efficient access to the grid, while also controlling the entire power grid to maintain its reliability and security.

Power grids are often represented as graphs to help explain the terminology and notation used in the field. To provide an example of this, we examine a graph of the IEEE 5-bus high-voltage power system in Fig. 2-2.

24
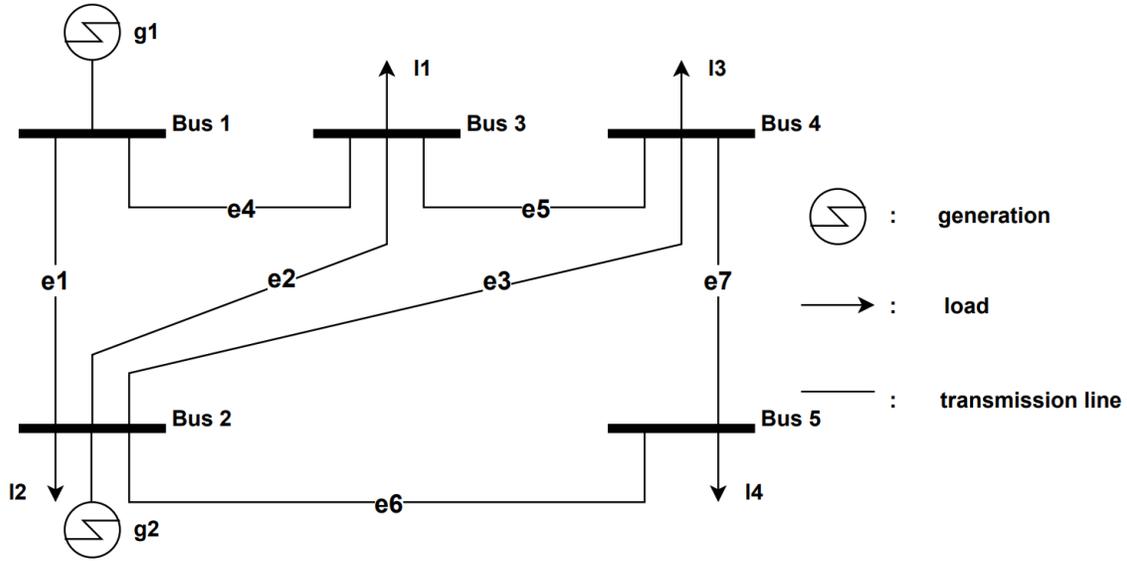
Figure 2-2: Graph representation of the power system (IEEE5 bus example).

The power system graph shown in the figure comprises of four essential components: buses, generators, loads, and transmission lines. A bus refers to a physical substation that connects various elements via switching devices, allowing for changes in connectivity between components. Generators are represented by g and form a set $\mathcal{G}$ ($g \in \mathcal{G}$), whereas loads are represented by l and belong to set $\mathcal{L}$ ($l \in \mathcal{L}$). In this work, for brevity, we will use the common word **injection** which refers to both generation and loads. Transmission lines, denoted by e ($e \in \mathcal{E}$), connect two or more buses and enable power flow (PF) in the power grid. All buses in the power grid, we will denote with set $\mathcal{B}$

While the power grid also includes other elements such as transformers, phase shifter transformers, HVDC, AC/DC converters, and capacitor banks, this work focuses solely on the four primary components mentioned above. Therefore, we will not delve into the description of these additional elements.

2.2.1 Power Flow

In this subsection, we will discuss the power flow analysis used in power systems. This analysis is crucial for determining the steady-state operating conditions of the power grid. It involves calculating the voltage magnitudes, phase angles, and power flows in all components of the power system, including generators, transmission lines, and loads.

During power flow analysis, the electrical characteristics of the power system components, such as their impedances, and the real and reactive power demands of the loads are considered. By solving the first principle equations that describe the power system's behavior, the power flow analysis calculates the steady-state voltages and power flows throughout the grid.

The results obtained from the power flow analysis are essential for power system

planning, design, and operation. They enable us to determine if the system is operating within its capacity limits, identify potential bottlenecks, and assess the impact of new loads or generators on the power system. Moreover, the power flow analysis is vital for developing control and protection schemes that ensure the stability and reliability of the power system.

To provide a better understanding of the power flow analysis, we will discuss the first principal equation. Since the majority of the grid operates on alternating current (AC), we will focus on AC power flow analysis. However, we will also briefly explain direct current (DC) power flow analysis for the reader's benefit.

In AC power transmission systems, voltage, and current intensity are the key variables that describe the system [Kundur et al., 1994]. These variables are represented by complex numbers consisting of a magnitude and a phase (angle). The objective of power flow analysis is to determine these magnitudes and angles at every bus in the grid using the available data and to utilize them to calculate power line flows.

This thesis focuses on power flow equations that apply to a fixed voltage frequency and a quasi-stationary operating condition, where power generation and loads are balanced. This condition ensures that the total power generated is equal to the total power consumed, including losses. Although this work does not delve into transient phenomena that occur over shorter periods, it is worth noting that such phenomena can have a significant impact on power grid operations in certain situations.

Based on Fig. 2-2, we can infer that the grid under consideration comprises $m = 5$ buses and $n = 7$ lines. Each bus $k \in (1, ...m)$ is characterized by voltage magnitude v_k and voltage angle θ_k, while each line e connecting two buses (j and k) is described by an impedance ($z_{j,k}$) which represents the physical properties of the line [Kundur et al., 1994]. This impedance is a complex number that includes both resistance and reactance. In power flow equations, impedance is often expressed as admittance, where $y_{j,k} = \frac{1}{z_{j,k}}$. A power line with admittance 0 indicates that there is no connection between the respective buses.

Following [Kundur, 2012], the current in lines between busses k and j is denoted with $\bar{i}_{k,j}$, and represents:

$$\bar{i}_{k,j} = y_{k,j}(\bar{v}_j - \bar{v}_k) \tag{2.1}$$

where $\bar{v}_j$ and $\bar{v}_k$ are voltages (complex number) in buses and $y_{k,j}$ is an admittance of the considered line. The power system is based on Kirchhoff's law where the sum of the currents entering and leaving a bus must be equal. Therefore, any current injected at bus k must also leave bus k and can be expressed as:

$$\bar{i}_k + \sum_{j=1, j \neq k}^{m} \bar{i}_{k,j} = 0 \tag{2.2}$$

The equation 2.2 can easily be derived from Kirchhoff's bus law presented in Fig. 2-3.
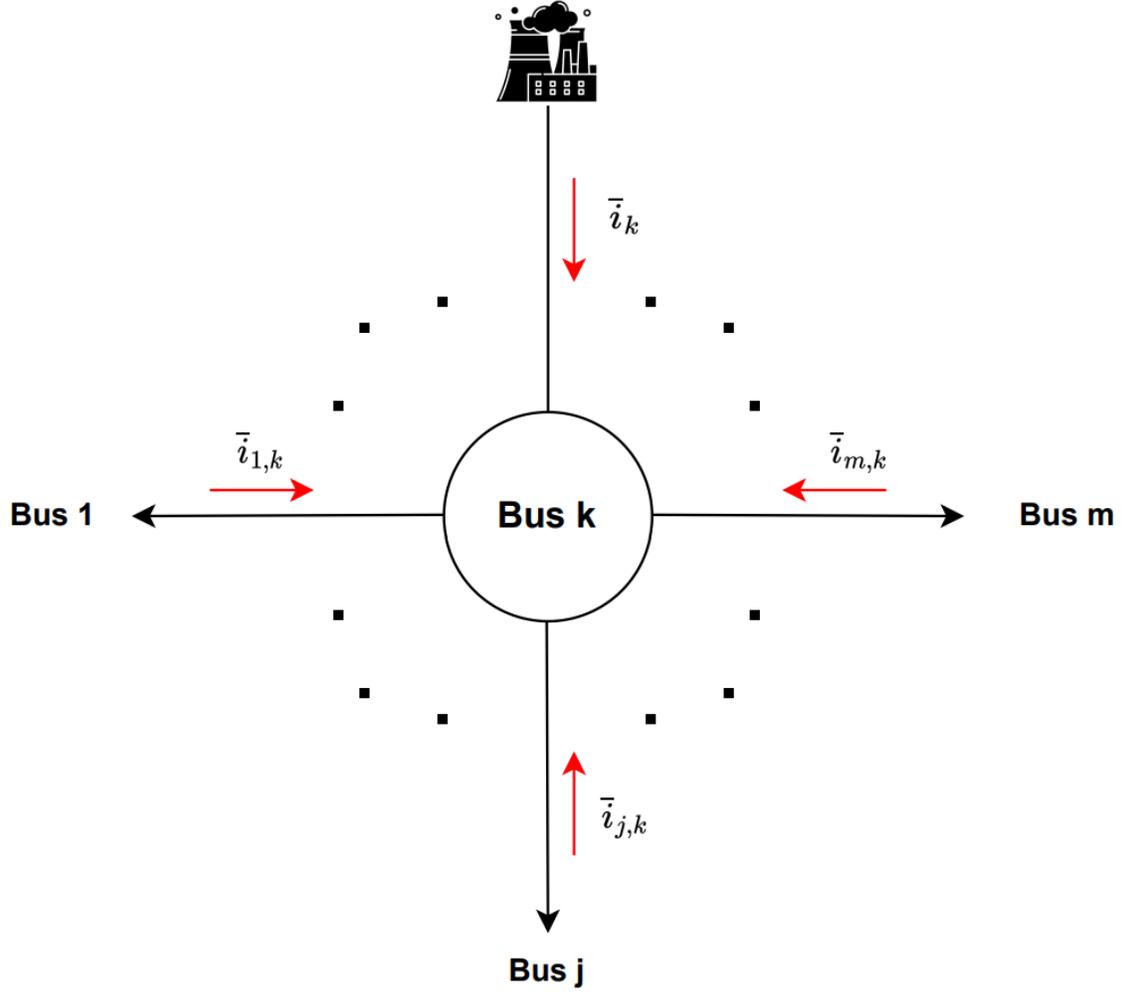
$$\bar{i}_k + \bar{i}_{1,k} + \ldots + \bar{i}_{j,k} + \ldots + \bar{i}_{m,k} = 0$$

Figure 2-3: Illustration of Kirchoff's bus law.

Further, combining equations 2.1 and 2.2 we get:

$$\bar{i}_k + \sum_{j=1, j \neq k}^{m} \bar{i}_{k,j} = 0 \tag{2.3}$$

By vectorizing equation 2.3, the full power grid can be described in matrix form as follows:

$$\begin{bmatrix} \bar{i}_1 \\ .. \\ \bar{i}_j \\ .. \\ \bar{i}_m \end{bmatrix} = Y \begin{bmatrix} \bar{v}_1 \\ .. \\ \bar{v}_j \\ .. \\ \bar{v}_m \end{bmatrix} \tag{2.4}$$

where Y is the m x m admittance matrix, defined as:

$$Y = \begin{bmatrix} -\sum_{j\neq 1} y_{1,j} & y_{1,2} & \cdots & y_{1,j} & \cdots & y_{1,m} \\ y_{2,1} & -\sum_{j\neq 2} y_{2,j} & \cdots & y_{2,j} & \cdots & y_{2,m} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ y_{m,1} & \cdots & \cdots & y_{m,j} & \cdots & -\sum_{j\neq m} y_{m,j} \end{bmatrix} \quad (2.5)$$

where the non-diagonal elements are equal as e.g $y_{1,2} = y_{2,1}$.

The injected current can be represented as the injected complex power $\bar{s}_k$:

$$\bar{s}_k = \bar{v}_k \bar{i}_k^* \quad (2.6)$$

or as a complex number of injected active power (p_k) and reactive power (q_k):

$$\bar{s}_k = p_k + \mathbf{i}q_k \quad (2.7)$$

where in this context $\mathbf{i}$ is the complex number $(\mathbf{i}^2 = -1)$, and p_k and q_k are a real and imaginary part of complex power.

Similarly, the active power $(p_{k,j})$ and reactive power $(q_{k,j})$ in transmission lines can be expressed as power flow as follows:

$$\bar{s}_{k,j} = p_{k,j} + \mathbf{i}q_{k,j} \quad (2.8)$$

By replacing Eq. 2.4 with equations 2.6 and 2.7 we get the AC power flow equations as:

$$p_{k,j} = v_k v_j \left(G_{k,j} \cos(\theta_k - \theta_j) + B_{k,j} \sin(\theta_k - \theta_j) \right) \quad (2.9a)$$
$$q_{k,j} = v_k v_j \left(G_{k,j} \sin(\theta_k - \theta_j) - B_{k,j} \cos(\theta_k - \theta_j) \right) \quad (2.9b)$$

where Eq. 2.9a describes the relationship between active power flow and voltage magnitude and angle (v and θ), while Eq. 2.9b expresses reactive power flow. The admittance $y_{k,j}$ between buses k and j in a power grid is a complex parameter, which can be expressed as a sum of its real and imaginary parts. Therefore, the real part is denoted as $G_{k,j}$ and represents the conductance of the line between buses k and j, while the imaginary part is denoted as $B_{k,j}$ and represents the susceptance of the line.

The system is balanced when the power flows leaving each bus are equal to the grid injection at that bus:

$$p_k = v_k^2 G_{kk} + \sum_{j=1}^{m} p_{k,j} \quad (2.10a)$$

$$q_k = -v_k^2 B_{kk} + \sum_{j=1}^{m} q_{k,j}, \quad (2.10b)$$

where p_k and q_k are active and reactive power injections at bus k, while the first term ($v_k^2 G_{kk}$ and $v_k^2 B_{kk}$) represents bus shunt elements.

The power flow equation serves to determine the unknown variables in the power grid. These variables can be either inputs or outputs of the power flow equations, depending on the type of bus they are associated with. Known variables represent inputs in the power flow equation system, while unknown ones (outputs) should be calculated. There are three types of buses: Slack, PV bus and PQ bus (Table. 2.1).

Bus Type	Bus properties	θ_k	v_k	p_k	q_k
Slack	at least one per grid	**known**	**known**	unknown	unknown
PV	connected generation	unknown	**known**	**known**	unknown
PQ	no connected generation	unknown	unknown	**known**	**known**

Table 2.1: Bus description and variables that are calculated in the power flow calculation process depending on the type of bus.

A **Slack** bus is a bus where the voltage angle θ_k and magnitude v_k are known and considered as inputs. At least one bus in the network must be a slack bus, where the voltage angle is set to 0 degrees ($\theta_k = 0$). The remaining buses in the network can be either PV or PQ buses.

A **PV** bus is a bus where the active power p_k injection and voltage magnitude v_k are known and considered as inputs. This type of bus is usually connected to a conventional generator unit such as thermal or hydropower, which can control the bus voltage.

A **PQ** bus is a bus where the active p_k and reactive power q_k injections are known and considered as inputs. This type of bus is usually connected to a renewable energy source (solar or wind) or has no generation unit connected to it. These buses are not voltage controlled and, therefore, the voltage magnitude is an output of the power flow equation.

In summary, the power flow equation is a crucial tool in determining the unknown variables of an electrical power system. These variables include voltage magnitudes v_k and angles θ_k, as well as power flow ($p_{k,j}$ and $q_{k,j}$) in the transmission lines. To obtain these values, a process called power flow computation is required. Power flow computation involves solving a sequence of nonlinear equations using a simulation engine. This engine takes in the known inputs, such as generator and load information, and uses them to calculate the unknown variables. The process is iterative, and the equations are solved until convergence is achieved, meaning that the computed values satisfy the set of equations.

2.2.2 DC Approximation

In this section, we will provide a brief overview of the direct current (DC) approximation, which is a common approximation technique in power flow modeling. Understanding this approximation will be helpful for readers to grasp the concepts discussed in the next section 2.3 and chapter 5.

The DC approximation involves linearizing the power flow equations to make the problem convex. Although it has certain limitations, particularly in voltage-prone power grids, it offers two primary benefits. Firstly, the DC approximation

always provides a solution and cannot diverge, ensuring its reliability. Secondly, it is computationally fast, making it an attractive option for real-time operation in the power grid, which is why transmission system operators TSOs often utilize it. Additionally, the DC approximation can also be applied in other areas such as grid development planning.

Comparing the DC power flow approximation and the AC power flow model from 2.9, there are three significant simplifications made in the former:

1. $R \ll X$, resistance R is much less than reactance X in a line;

2. $\theta_k - \theta_j \approx 0$, the difference of angles between two connected busses is small;

3. $v_k \approx v_{nom}$, the voltage amplitude is close to the nominal voltage at each bus.

The first simplification $R \ll X$ neglects the losses in the system solved by the DC approximation. This simplification can be justified by the fact that in high-voltage grids reactance X is higher than resistance R in line. It directly affects the admittance of the transmission line y. As previously mentioned, admittance y is the reciprocal of impedance z where the impedance is the complex number of resistance R as the real part and reactance X as the imaginary part. Accordingly, admittance y is equal:

$$y_{k,j} = \frac{1}{z_{k,j}} = \frac{1}{R_{k,j} + \mathbf{i}X_{k,j}} = \frac{R_{k,j}}{R_{k,j}^2 + X_{k,j}^2} - \mathbf{i}\frac{X_{k,j}}{R_{k,j}^2 + X_{k,j}^2} \tag{2.11}$$

where conductance $G_{k,j}$ and susceptance $B_{k,j}$ are defined as:

$$G_{k,j} = \frac{R_{k,j}}{R_{k,j}^2 + X_{k,j}^2} \tag{2.12a}$$

$$B_{k,j} = \frac{-X_{k,j}}{R_{k,j}^2 + X_{k,j}^2} \tag{2.12b}$$

Since simplification implies that $R \ll X$ we get:

$$G_{k,j} \to 0 \tag{2.13a}$$

$$B_{k,j} \to \frac{-1}{X_{k,j}} \tag{2.13b}$$

Accordingly, the AC power flow equations from 2.9 are simplified into:

$$p_{k,j} = v_k v_j \left(B_{k,j} \sin(\theta_k - \theta_j) \right) \tag{2.14a}$$

$$q_{k,j} = v_k v_j \left(B_{k,j} \cos(\theta_k - \theta_j) \right) \tag{2.14b}$$

The second simplification $\theta_k - \theta_j \approx 0$ linearized the trigonometric functions sin and cos, giving:

$$p_{k,j} = v_k v_j \left(B_{k,j}(\theta_k - \theta_j) \right) \tag{2.15a}$$

$$q_{k,j} = v_k v_j B_{k,j} \tag{2.15b}$$

Finally, the last simplification $v_k \approx v_{nom}$ made the voltage amplitude v constant and therefore the power flow model completely linear resulting in:

$$p_{k,j} = B_{k,j}(\theta_k - \theta_j) \tag{2.16a}$$

$$q_{k,j} = B_{k,j} \tag{2.16b}$$

Eq. 2.16b can be ignored since the parameter B is known for each line. Accordingly, the first equation (Eq. 2.16a) plays a key role in solving the DC power flow as a function of unknown θ. Following the vectorized equation 2.4 for the AC power flow, in a similar way, we can represent it for the DC power flow as:

$$\begin{bmatrix} p_1 \\ .. \\ p_j \\ .. \\ p_m \end{bmatrix} = Y \begin{bmatrix} \theta_1 \\ .. \\ \theta_j \\ .. \\ \theta_m \end{bmatrix} \tag{2.17}$$

This system of linear equations can be easily solved using any linear solver.

2.3   Optimal Power Flow

In this section, we will talk about optimal power flow (OPF) which consists of the power flow as the key component of the problem. OPF is an essential tool for power markets and power security, which involves solving optimization problems based on power flow equations. It is performed annually for system planning or daily, hourly, or even every five minutes for day-ahead planning. OPF was first introduced by Carpentier in 1962, where he dealt with the economic dispatch problem [Carpentier, 1962]. Since then, the formulation of OPF has evolved to solve various types of problems.

In general, OPF involves constrained optimization problems that consist of variables optimizing an objective function, equality constraints such as power balance and power flow equations, and inequality constraints including variable bounds. Depending on the type of OPF, the variables, constraints, and objective function may change.

The AC-OPF is the key method for solving Transmission System Operator (TSO) problems, but it remains challenging. It is complex economically, electrically, and computationally due to the need for multi-part nonlinear pricing in an efficient market, additional non-linearities caused by AC-PF, and its non-convexity, including binary and integer variables, that increases computational complexity in optimization. Despite several decades since OPF is formulated, fast and robust solution techniques for solving the full AC-OPF are still lacking.

OPF can solve various types of problems such as economic dispatch, unit commitment, optimal topology, and long-term planning. In this section, we focus on the economic dispatch problem with continuous variables since it is discussed further in the contribution part. We do not discuss other problems involving binary and integer variables. Moreover, in this section, we review deterministic and stochastic OPF and

state-of-the-art solutions in this field.

2.3.1 Deterministic OPF

The deterministic or conventional OPF problem can be formulated in an abstract form as follows:

$$\min_{p_g, q_g, v, \theta} f(p_g) \tag{2.18a}$$

$$\text{s.t. } F(p, q, v, \theta) = 0, \tag{2.18b}$$

$$p_{g,i}^{\min} \le p_{g,i} \le p_{g,i}^{\max}, \quad \forall g \in \mathcal{G} \tag{2.18c}$$

$$q_{g,i}^{\min} \le q_{g,i} \le q_{g,i}^{\max}, \quad \forall i \in \mathcal{G} \tag{2.18d}$$

$$v_j^{\min} \le v_j \le v_j^{\max}, \quad \forall j \in \mathcal{B} \tag{2.18e}$$

$$s_{k,j} \le s_{k,j}^{\max}, \quad \quad \forall k \sim j \tag{2.18f}$$

$$\theta_{slack} = 0 \tag{2.18g}$$

where f is the objective (cost) function; F is the AC power flow balance equations from 2.10; p_g and q_g are active and reactive power injections of the controlable generation; θ_{slack} is voltage angel in Slack bus; $s_{k,j}$ is a value of apparent power flow along the line $e_{k,j}$ defined as $s_{k,j} = \sqrt{p_{k,j}^2 + q_{k,j}^2}$.

The corresponding OPF formulation 2.18 has been adopted for the economic dispatch problem. The economic dispatch problem aims to determine the most cost-effective output levels for power generators while satisfying the physical and load-related limitations of the power system. To achieve this, an optimization problem is formulated with a cost function 2.18a, which is minimized to find the optimal solution for the economic dispatch OPF problem. Usually, the cost function of the economic dispatch OPF is the function of the total real power generation p_g. This cost function is typically represented by a quadratic or piecewise linear function. For the purpose of this discussion, we will focus on the quadratic cost function, which is expressed as follows.

$$f(p_g) = \sum_{i \in \mathcal{G}} \left\{ c_{2,i} p_{g,i}^2 + c_{1,i} p_{g,i} + c_{0,i} \right\} \tag{2.19}$$

where $\{c_{2,i}, c_{1,i}, c_{0,i}\}_{i \in \mathcal{G}} \ge 0$ are scalar cost coefficients.

Equality constraints 2.18b in an optimization problem are mathematical expressions that represent the relationship between the optimization variables, usually in the form of equations. In the context of an AC-OPF, the power flow equations 2.9 and power balance equations 2.10 are fundamental physical laws that must be incorporated as equality constraints. The voltage angle differences are of primary importance than individual angles in the power flow equations. Accordingly, the voltage angle of the slack bus 2.18g can be selected as an additional equality constraint by setting its value to zero. This simplifies the calculation of voltage angle differences and facilitates the solution of the OPF problem.

Inequality constraints, which represent the physical limitations of the system elements, are also included in the OPF problem. These constraints also represent the function of the optimization variables by restricting the function value within a specified range. For example, in the context of the economic dispatch problem, the most common inequality constraints reflect the physical limitations of the system, such as minimum and maximum power output constraints 2.18c - 2.18d, voltage level constraints 2.18e and transmission line flow constraints 2.18f. These constraints ensure that the solution to the optimization problem is feasible and physically realizable.

During steady-state operations at PV buses, the AC-OPF in 2.18 returns the generation set points p_g and q_g, while voltage magnitudes v need to be fixed. On the other hand, active and reactive power loads are typically fixed for PQ buses.

Interior-point method

Solving the economic dispatch optimization problem in 2.18 is challenging, as it involves non-linear and non-convex functions. However, one of the most effective techniques for solving large-scale optimization problems, and in this case, a full AC-OPF economic dispatch problem like this in 2.18 is the interior-point method [Nocedal and Wright, 2006, Wächter and Biegler, 2006]. In this section, we will give a brief explanation of the main aspects of this method.

Mathematical optimization involves finding the optimal values of decision variables that minimize the objective function, subject to a set of constraints, including equality and inequality constraints. Accordingly, to explain how the interior point method works, we introduce simplified expressions where x is the decision variable, c_{eq} represents all equality constraints, and c_{in} symbolizes all inequality constraints. Therefore, the problem of mathematical optimization can be written in a simplified form as follows:

$$\min_{x} f(x) \tag{2.20a}$$

$$\text{s.t. } c_{eq}(x) = 0 \tag{2.20b}$$

$$c_{in}(x) \geq 0 \tag{2.20c}$$

Barrier methods and interior-point methods are often used interchangeably because interior-point methods create a surrogate model of the optimization problem. This model is created by transforming inequality constraints into equality constraints and substituting the objective function with a logarithmic barrier function. Accordingly, formulation 2.20 can be rewritten as follows:

$$\min_{x,s_{sl}} f(x) - \psi \sum_{i=1}^{|c_{in}|} log(s_{sl_i}) \tag{2.21a}$$

$$\text{s.t. } c_{eq}(x) = 0 \tag{2.21b}$$

$$c_{in}(x) - s_{sl} = 0 \tag{2.21c}$$

where the objective function includes a barrier term with a barrier parameter ψ ($\psi > 0$) and the slack variables s_{sl} represented as a vector. From formulation 2.21, it is

obvious that the barrier term enforces an inequality constraint on the slack variables $s_{sl_i} \geq 0$ through the logarithmic function. While the barrier problem and the original non-linear program from 2.20 are not equivalent, the solution of the barrier problem approaches the solution of the original optimization problem as the barrier parameter approaches zero. In other words, the barrier problem provides an approximate solution to the original problem by gradually relaxing the constraint until it is no longer a factor.

The Karush-Kuhn-Tucker (KKT) conditions [Gordon and Tibshirani, 2012] are fundamental to the basic interior-point algorithm, which is used to solve optimization problems. These conditions involve four sets of equations, where the first two originate from the system Lagrangian's first-order condition. Thus, the first two sets of equations require the derivatives with respect to the decision and slack variables to be equal to zero. The third set of equations relates to the equality constraints 2.21b, while the fourth set refers to the inequality constraintc 2.21c. KKT conditions are formulated as follows:

$$
\bigtriangledown f(x) - J_{eq}^T(x)r - J_{in}^T(x)q \tag{2.22a}
$$

$$
\bigtriangledown Sq - \psi e_{ind} = 0 \tag{2.22b}
$$

$$
c_{eq}(x) = 0 \tag{2.22c}
$$

$$
c_{in}(x) - s_{sl} = 0 \tag{2.22d}
$$

In KKT formulation, $J_{eq}^T(x)$ and $J_{in}^T(x)$ are transposed Jacobian matrices of the equality and inequality functions, respectively; r and q are corresponding Lagrange multipliers; S is a diagonal matrix of the slack variables ($S = diag(s_{sl})$); e_{ind} is a vector of diagonal elements of identity matrix I ($e_{ind} = [1, ..., 1]^T$).

The KKT conditions create a non-linear system by realizing a relationship between the primal (x, s_{sl}) and dual variables (r, q). Concisely, a non-linear system can be expressed by a vector-valued function $F_{KKT} = (x, s_{sl}, r, q) = 0$. This system needs to be solved to obtain the optimal solution. One common approach to solving the non-linear system is to use the Newton-Raphson method. This method can be applied by looking for a search direction based on the following linear problem:

$$
J_{KKT}(x, s_{sl}, r, q)p_{dir} = -F_{KKT}(x, s_{sl}, r, q) \tag{2.23}
$$

where J_{KKT} is the Jacobian matrix of the KKT system; p_{dir} is the search direction vector. The full form of the J_{KKT} is:

$$
J_{KKT}(x, s_{sl}, r, q) = \begin{bmatrix} \bigtriangledown_{xx}L & 0 & -J_{eq}^T(x) & -J_{eq}^T(x) \\ 0 & Q & 0 & S \\ J_{eq}^T(x) & 0 & 0 & 0 \\ J_{eq}^T(x) & -I & 0 & 0 \end{bmatrix} \tag{2.24}
$$

where $Q = diag(q)$; $\bigtriangledown_{xx}L$ is the Hessian of Lagrangian system L defined as:

$$
L(x, s_{sl}, r, q) = f(x) - r^T c_{eq}(x) - q^T(c_{in}(x) - s_{sl}) \tag{2.25}
$$

The search direction vector p_{dir} is:

$$p_{dir} = \begin{bmatrix} p_{dir,x} \\ p_{dir,s_{sl}} \\ p_{dir,r} \\ p_{dir,q} \end{bmatrix} \tag{2.26}$$

Iteratively applying the Newton-Raphson method to a formulation 2.25, also called a primal-dual system, allows converging to the optimal solution of the optimization problem. It means that in iteration $n + 1$ we obtain p_{dir}^{n+1} and accordingly update variables (x, s_{sl}, r, q) and barrier parameter ψ until convergence criteria is not satisfied, as follows:

$$x^{n+1} = x^n + \alpha_{dir,x}^{n+1} p_{dir,x}^{n+1} \tag{2.27a}$$

$$s_{sl}^{n+1} = s_{sl}^n + \alpha_{dir,s_{sl}}^{n+1} p_{dir,s_{sl}}^{n+1} \tag{2.27b}$$

$$r^{n+1} = r^n + \alpha_{dir,r}^{n+1} p_{dir,r}^{n+1} \tag{2.27c}$$

$$q^{n+1} = q^n + \alpha_{dir,q}^{n+1} p_{dir,q}^{n+1} \tag{2.27d}$$

$$\psi^{n+1} = f_\psi(\psi^n, x^{n+1}, s_{sl}^{n+1}, r^{n+1}, q^{n+1}) \tag{2.27e}$$

where α^{dir} is the line search parameter for different variables; f_ψ is a function that computes the updated value of ψ from the previous value and the updated variables. Various methods have been studied to enhance the convergence rates of optimization algorithms. These methods utilize different heuristic functions f_ψ for calculating line search and barrier parameters.

The interior-point method is a common and reliable approach for solving AC-OPF problems. However, solving AC-OPF with the interior-point method can be costly due to the need for computing the Hessian of the Lagrangian $\nabla_{xx} L$ at each iteration step. Solving large-scale problems with the interior-point method can be particularly challenging, as the computational time scales superlinearly with system size. To overcome these challenges, there are several strategies available. One approach is to simplify the original OPF problem by using DC power flow equations, while another is to compute approximate solutions. These strategies can make the problem more manageable and reduce the computational burden.

Convex relaxations

Convex relaxations are a set of approximations used to tackle non-convex optimization problems. The underlying concept is to approximate the original problem with a convex optimization problem, which is easier to solve. In contrast to non-convex problems, where local minima may not be the global minimum, in convex problems, any local minimum is also the global minimum. Thus, there are several efficient algorithms available with excellent convergence rates to solve convex optimization problems [Boyd and Vandenberghe, 2004].

Convex relaxations can be illustrated through the economic dispatch problem,

which can be formulated using the basic BIM approach. With some slight adjustments, the economic dispatch problem can be reformulated into a quadratic programming (QP) problem using the complex voltages of the buses. In the QP optimization problem utilized for convex relaxations, the objective function and all inequality constraints have a quadratic form in the complex voltages [Low, 2014a, Low, 2014b]. Thus QP form of the convex relaxations is:

$$\min_{\bar{v} \in \mathbb{C}} \bar{v}^* C \bar{v} \tag{2.28a}$$

$$\text{s.t. } \bar{v}^* M_{herm} \bar{v} \leq m_{herm} \tag{2.28b}$$

where M_{herm} is Hermitian matrix; m_{herm} is real-valued vectors of the corresponding inequality constraints; sunbscript $*$ denotes conjugate transpose. However, QP formulation 2.28 is still a non-convex. To make the QP problem convex, the expression $\bar{w} = \bar{v}\bar{v}^*$ is included in formulation 2.28. $\bar{w}$ is positive semidefinite matrix with rank 1 ($\bar{w} \succeq 0$, rank($\bar{w}$) = 1). By utilizing the identity of the trace function for any Hermitian matrix, formulation 2.28 can be reformulated as:

$$\min_{\bar{w} \in \mathbb{S}} tr(C\bar{w}) \tag{2.29a}$$

$$\text{s.t. } tr(M_{herm}\bar{w}) \leq m_{herm} \tag{2.29b}$$

$$\bar{w} \succeq 0 \tag{2.29c}$$

$$rank(\bar{w}) = 1 \tag{2.29d}$$

where $\mathbb{S}$ is the space of m x m symmetric matrices. Finally, the formulation 2.29 is convex in $\bar{w}$ and represents a semidefinite programming problem (SDP), known as the SDP relaxation of the OPF. Convex relaxation approaches possess an advantage over other approximation techniques in that the infeasibility of the relaxed problem directly implies the infeasibility of the original problem. Despite the recent success of regularized semidefinite programming [Krechetov et al., 2018], which allows to produce tighter bounds, their scalability does not allow applications to large-scale power systems problems.

Linearized DC-OPF

The DC-OPF approach is a commonly used approximation in power systems. This approach implies replacing the AC power flow balance equation with the DC power flow balance equation described in 2.2.2. The DC approximation linearized the original OPF problem by removing several variables and constraints (see Eq. 2.16). Accordingly, DC-OPF creates a linear programming problem that can be solved very efficiently by interior-point methods or simplex methods. However, one major limitation is that the solution obtained from DC-OPF may not be feasible in AC-OPF [Low, 2019]. This leads to restarting the DC-OPF calculation and tightening some constraints.

2.3.2 Stochastic OPF

This section discusses stochastic OPF techniques that handle the uncertainty of power injections at the bus. This uncertainty primarily arises from the increasing use of RES like wind and solar, as well as the implementation of smarter grids that allow for the deferral of load and storage devices. Stochastic optimal power flow is a mathematical optimization technique used in power systems to determine the optimal operating conditions of the system under uncertainty. It considers the stochastic nature of the power grid variables, such as the fluctuating load and the uncertain renewable energy sources (RES). Accordingly, firstly we need to understand how these variables are modeled under uncertainty.

The deterministic AC OPF assumes that the power injections are precisely known at the time of scheduling. In practice, this is not true, as both load and renewable generation might vary from their forecasted value. Deviations in forecasted values are usually caused by forecast errors, external fluctuations, or intra-day electricity trading. For this reason, it is important to account for the impact of injection uncertainties on system operation to ensure secure operation. Accordingly, we assume that active power uncertainties are modeled as the sum of the deterministic forecasted value p and a real-time fluctuation ω as:

$$p_l(\omega) = p_l + \omega \tag{2.30}$$
$$p_r(\omega) = p_r + \omega \tag{2.31}$$

where ω is an independent random variable with zero mean and known standard deviation σ_ω; p_l and p_r are load and RES active power injections. In this thesis, the reactive power injections of load and RES correspond to the same uncertainty realization while maintaining a given constant power factor $cos(\phi)$ as:

$$q_l(\omega) = \gamma p_l(\omega) = \gamma p_l + \gamma \omega \tag{2.32}$$
$$q_r(\omega) = \gamma p_r(\omega) = \gamma p_r + \gamma \omega \tag{2.33}$$

where $\gamma = \sqrt{\frac{1-cos^2(\phi)}{cos^2(\phi)}}$ is the power ratio. Thus, the ratio of active and reactive power injections remains unchanged during fluctuations.

Following uncertainty realizations ω, the controllable generators adapt their generation to maintain the total power balance and feasibility. We use an affine policy representing the automatic generation control (AGC) [Roald and Andersson, 2017] to balance active power generation. Therefore, the total power mismatch $\Omega = \sum_{j\in\mathcal{G}} \omega_j$ is divided among the generators according to the participation factors α based on the following generation control policy:

$$p_{g,j}(\omega) = p_{g,j} + \alpha_j \Omega, \qquad \forall j \in \mathcal{G} \tag{2.34a}$$
$$\sum_{j\in\mathcal{G}} \alpha_j = 1 \tag{2.34b}$$

The reactive power is controlled locally at buses $\mathcal{G}$, keeping their outputs constant

at buses $\mathcal{L}$. Generators at buses $\mathcal{G}$ change reactive power outputs by $\delta q(\omega)$ to keep voltage magnitudes constant on these buses,

$$q_j(\omega) = q_j + \delta q_j(\omega), \qquad \forall j \in \mathcal{G} \qquad (2.35a)$$

$$q_j(\omega) = q_j, \qquad \forall j \in \mathcal{L} \qquad (2.35b)$$

In contrast, the voltage magnitudes is fixed at buses $\mathcal{G}$ and fluctuates at buses $\mathcal{L}$,

$$v_j(\omega) = v_j + \delta v_j(\omega), \qquad \forall j \in \mathcal{L} \qquad (2.36a)$$

$$v_j(\omega) = v_j, \qquad \forall j \in \mathcal{G} \qquad (2.36b)$$

Let $p_{k,j}$ and $q_{k,j}$ denote the active and reactive power flows from bus k to bus j along line $(k, j) \in \mathcal{E}$. In the AC power flow formulation, the active and reactive power flows on each transmission line depend non-linearly on the voltage magnitudes v and voltage angles θ. Consider all possible fluctuations ω within the uncertainty set W, the power flow equations from 2.9 are given by:

$$p_{k,j}(\omega) = v_k(\omega)v_j(\omega)\left(G_{k,j}\cos\left(\theta_{k,j}(\omega)\right) + B_{k,j}\sin\left(\theta_{k,j}(\omega)\right)\right) \qquad (2.37a)$$

$$q_{k,j}(\omega) = v_k(\omega)v_j(\omega)\left(G_{k,j}\sin\left(\theta_{k,j}(\omega)\right) - B_{k,j}\cos\left(\theta_{k,j}(\omega)\right)\right) \qquad (2.37b)$$

where $\theta_{k,j}(\omega) = \theta_k(\omega) - \theta_j(\omega)$. According to 2.10, power balanced equation under uncertainty is given as:

$$p_k(\omega) = v_k^2(\omega)G_{k,k} + \sum_j^m p_{k,j}(\omega) \qquad (2.38a)$$

$$q_k(\omega) = -v_k^2(\omega)B_{k,k} + \sum_j^m q_{k,j}(\omega) \qquad (2.38b)$$

Deviations in power injections cause changes in power flows throughout the system. Accordingly, the apparent power flow fluctuates in the lines as:

$$s_{k,j}(\omega) = s_{k,j} + \delta s_{k,j}(\omega), \qquad \forall k \sim j \in \mathcal{E} \qquad (2.39)$$

Finally, similar to Eg. 2.18 and taking into account Eg. 2.30 - 2.39, the stochastic

OPF can be define as:

$$\min_{p_g,q_g,v,\theta,\alpha} f(p_g(\omega)) \tag{2.40a}$$

$$\text{s.t. } F(p(\omega),q(\omega),v(\omega),\theta(\omega)) = 0, \quad \forall \omega \in W \tag{2.40b}$$

$$p_{g,i}^{\min} \leq p_{g,i}(\omega) \leq p_{g,i}^{\max}, \qquad \forall g \in \mathcal{G} \tag{2.40c}$$

$$q_{g,i}^{\min} \leq q_{g,i}(\omega) \leq q_{g,i}^{\max}, \qquad \forall i \in \mathcal{G} \tag{2.40d}$$

$$v_j^{\min} \leq v_j(\omega) \leq v_j^{\max}, \qquad \forall j \in \mathcal{B} \tag{2.40e}$$

$$s_{k,j}(\omega) \leq s_{k,j}^{\max}, \qquad \forall k \sim j \tag{2.40f}$$

$$\theta_{slack} = 0 \tag{2.40g}$$

This section will focus on the chance-constrained (CC) approach for modeling and solving Stochastic OPF while acknowledging that there are other approaches such as robust optimization and probabilistic optimal power flow (P-OPF).

Chance-constrained OPF

Chance constrained approach uses probability-based constraints that rely on a specific probability distribution of uncertainty to limit the chances of constraint violations below a desired threshold. Essentially, these constraints ensure that the probability of violating any constraints remains under a set limit. The full AC chance-constrained OPF is formulated as follows:

$$\min_{p_g,q_q,v,\theta} \sum_{i \in \mathcal{G}} \mathbb{E}[c_i(p_{g,i}(\omega))] \tag{2.41a}$$

$$\text{s.t. } F(\theta(\omega),v(\omega),p(\omega),q(\omega)) = 0, \qquad \forall \omega \in W \tag{2.41b}$$

$$\mathbb{P}(p_{g_i}(\omega) \leq p_{g_i}^{max}) \geq 1 - \epsilon_p, \qquad \forall i \in \mathcal{G} \tag{2.41c}$$

$$\mathbb{P}(p_{g_i}(\omega) \geq p_{g_i}^{min}) \geq 1 - \epsilon_p, \qquad \forall i \in \mathcal{G} \tag{2.41d}$$

$$\mathbb{P}(q_{g_i}(\omega) \leq q_{g_i}^{max}) \geq 1 - \epsilon_q, \qquad \forall i \in \mathcal{G} \tag{2.41e}$$

$$\mathbb{P}(q_{g_i}(\omega) \geq q_{g_i}^{min}) \geq 1 - \epsilon_q, \qquad \forall i \in \mathcal{G} \tag{2.41f}$$

$$\mathbb{P}(v_j(\omega) \leq v_j^{max}) \geq 1 - \epsilon_v, \qquad \forall i \in \mathcal{N} \tag{2.41g}$$

$$\mathbb{P}(v_j(\omega) \geq v_j^{min}) \geq 1 - \epsilon_v, \qquad \forall i \in \mathcal{N} \tag{2.41h}$$

$$\mathbb{P}(s_{kj}(\omega) \leq s_{kj}^{max}) \geq 1 - \epsilon_s, \qquad \forall k \sim j \in \mathcal{E} \tag{2.41i}$$

where expectations $\mathbb{E}$ and probabilities $\mathbb{P}$ are defined over the distribution of ω.

The power outputs of conventional generators, voltage magnitudes at buses and apparent power flow in the lines are constrained using the separate chance constraints 2.41c–2.41i. Moreover, chance constraints can be modeled as joint chance constraints, which ensure that all constraints hold jointly with a pre-described probability [Vrakopoulou et al., 2013c, Vrakopoulou et al., 2013b]. The inequality constraints are formulated probabilistically, i.e., must be satisfied with the prescribed probability. The prescribed probability is controlled by a choice of acceptable violation probabilities

ϵ_p, ϵ_q, ϵ_v and ϵ_s. Accordingly, the concept of chance constraint is used to define an ϵ-reliability set, which allows TSOs to safely execute additional control actions within that set, with a low probability of violating a constraint. A lower acceptable violation probability ϵ ensures more reliable system operation, but at a higher cost. In contrast, a higher acceptable violation probability ϵ is riskier and provides no guarantee that such controls will be available.

CC-OPF formulation 2.41 is intractable and causes challenging computation as it involves computing multi-dimensional probability integrals. However, recent advancements in chance-constrained optimization have inspired the creation of state-of-the-art CC-OPF applications [Calafiore and Campi, 2013, Margellos et al., 2014].

The scenario-based approach is a commonly used method for solving the CC-OPF [Calafiore and Campi, 2013]. This approach involves replacing the chance constraint with a finite set of constraints that correspond to different potential outcomes of the uncertain parameters. The key advantage of this approach is that it assumes that all relevant functions are convex with respect to the decision variables, making the problem easier to solve. The scenario-based approach provides probabilistic guarantees by using a set of scenarios that increases linearly with the number of decision variables. In simpler terms, this method uses a finite set of constraints to represent different potential outcomes of uncertain parameters, and it assumes that the problem is solvable through convex functions. There are also some alternative approaches to the scenario-based approach. One of them is given in [Margellos et al., 2014]. This approach solves a robust problem with bounded uncertainty, where bounded uncertainties are computed using a scenario approach. In addition, convexity is not required in this approach, and the number of scenarios does not depend on the number of decision variables. Accordingly, this alternative approach reduces computational costs or provides less conservative guarantees. In [Lukashevich and Maximov, 2021], the authors used internal approximations to the feasibility set later on providing more efficient algorithms to generate samples (scenarios) for reducing the complexity of the optimization methods [Owen et al., 2019, Lukashevich et al., 2021, Lukashevich and Maximov, 2021]. Furthermore, the resulting deterministic problem may have a special structure allowing for more efficient numerical methods [Anikin et al., 2022, Krechetov et al., 2018]

Since CC-OPF with full AC power flow equation is very difficult to solve, relaxation and linear approximation are often used to make problems easier. Full AC power flow equation in CC-OPF with a polynomial chaos expansion is proposed in [Mühlpfordt et al., 2019]. Analytical reformulation of the CC-ACOPF based on linear sensitivity factors is proposed in [Schmidli et al., 2016]. In [Roald and Andersson, 2017] and [Roald et al., 2017], a deterministic AC-OPF problem was proposed that is solved iteratively and estimates the uncertainty limits by a linear approximation at the operating point. However, this method cannot guarantee the convergence of the algorithm and finding the global optimum.

In recent years, convex relaxation has been used to solve the AC CC-OPF problem [Halilbašić et al., 2018, Venzke et al., 2017, Baker and Toomey, 2017, Lubin et al., 2019, Xie and Ahmed, 2017]. Semi-definite programming (SDP) relaxation and sample-based methods are used in [Venzke et al., 2017]. However, it is still difficult

to solve the problem reformulated in this way, since all the relaxed solutions in the set of uncertainties have a physical meaning. Since SDP relaxation is computationally demanding, the second-order cone programming (SOCP) relaxation is applied in [Halilbašić et al., 2018] to achieve faster computations than SDP in [Venzke et al., 2017]. In addition, in [Baker and Toomey, 2017] the convex optimization problem is reformulated from a non-convex joint CC-OPF using the improved Boole's inequality. The SOCP problem is also addressed in [Lubin et al., 2019] by linearizing the power balance equation around the forecasted operation point. Similarly, an exact SOCP is proposed in [Xie and Ahmed, 2017] to reformulate a distributionally robust CC-OPF with known first and second moments. However, convex relaxation methods have not yet been developed for practical application in industry. Furthermore, CC-OPF problem can be combined with the contingency analysis and detection [Anikin et al., 2022, Burashnikova et al., 2022, Mikhalev et al., 2020, Stulov et al., 2020].

The literature review indicates that DC CC-OPF based on a linearized DC power flow equation is widely used due to its impressive efficiency and scalability [Bienstock et al., 2014, Roald et al., 2016, Lubin et al., 2015, Hou and Roald, 2020, Pena-Ordieres et al., 2020]. However, these methods oversimplify the power system by neglecting voltage and reactive power, which can lead to solutions that deviate significantly and compromise the secure operation of the system. This is especially depicted in systems with a strong relationship between active and reactive power [Castillo et al., 2015]. To address this issue, a more accurate linearization method for power flow equations is needed, particularly in stochastic scenarios. This requires retaining the essential characteristics of the AC power flow model as much as possible while making the model deterministic and linear. Such an approach would provide a more reliable and accurate solution, especially in complex power systems with significant uncertainties. We also should mention a series of papers that potentially allow us to reduce the computational time in the above-mentioned problems [Grinis, 2022, Kadilenko and Grinis, 2023, Grinis, 2016] based on efficient approximations of complicated numerical problems.

In summary, the CC-OPF approach offers potential cost savings by allowing violations of operational constraints with low probability, while still meeting forecasted operating conditions. This approach is particularly useful for intraday operations where forecast uncertainty is low, but may not be as suitable for day-ahead operations where uncertainty is higher [Capitanescu, 2016]. However, the disadvantage of CC-OPF is that still solving the full AC CC-OPF is challenging and computationally intensive, while approximation methods can compromise the safe operation of the system and are still not available in practice. To overcome these issues, the contribution of this thesis is a developing novel data-driven based approach (part II) to address these drawbacks.

2.4 Synthetic dataset generation

In this section, we describe how we created a synthetic dataset by simulating a power system. This generated dataset is essential for the contribution part of our research.

One significant advantage of using a synthetic dataset is that it allows us to conduct a controlled experiment where we have complete knowledge and control over all the variables involved.

Motivated by [Donnot, 2019], we generated the synthetic dataset using three different parts of injection sampling:

1. active power loads and RES;

2. reactive power loads and RES

3. active power (controllable) generation

2.4.1   Sampling active power loads and RES

To simulate real active powers for i-th load and j-th renewable source, we define stochasticity as:

$$p_l^i = \eta^i p_{l_{ref}}^i, \ i \in \mathcal{L} \tag{2.42a}$$

$$p_r^j = \nu^j p_{r_{ref}}^j, \ j \in \mathcal{L} \tag{2.42b}$$

where $p_{l_{ref}}^i$ and $p_{r_{ref}}^j$ are fixed reference values of loads and RES; η^i and ν^j are a random variables that follow a required distribution. Further, we will consider log-normal distribution. As in [Donnot, 2019], we separate random variables η^i and ν^j into two components:

$$\eta^i = \eta_{corr} \eta_{uncorr}^i, \ i \in \mathcal{L} \tag{2.43a}$$

$$\nu^j = \nu_{corr} \nu_{uncorr}^j, \ j \in \mathcal{L} \tag{2.43b}$$

where η_{corr} and ν_{corr} denotes spatio-temporal correlations, while η_{uncorr}^j and ν_{uncorr}^j are local variations. We propose that both variables have log-normal distribution. We consider correlated variables η_{corr} at all loads and ν_{corr} at all RES to take the same values with mean and standard deviation, $(\mu(\eta_{corr}) = -1, \mu(\nu_{corr}) = 0.2)$ and $(\sigma(\eta_{corr}) = 0.1, \sigma(\nu_{corr}) = 0.4)$. In contrast, uncorrelated variables are different for each load and renewable source, distributed with means $(\mu(\eta_{corr}) = 1, \mu(\nu_{corr}) = 1)$ and standard deviations $(\sigma(\eta_{corr}) = 0.05, \sigma(\nu_{corr}) = 0.3)$.

2.4.2   Sampling reactive power loads and renewable sources

Reactive powers are sampled as a fixed power ratio from the active injections as follows:

$$q_l^i = \gamma^i p_l^i, \ i \in \mathcal{L} \tag{2.44a}$$

$$q_{rs}^j = \gamma^j p_{rs}^j, \ j \in \mathcal{L} \tag{2.44b}$$

The power ratio is derived form the reference values $\gamma^i = \frac{q_{l_{ref}}^i}{p_{l_{ref}}^i}, \ \gamma^j = \frac{q_{rs_{ref}}^j}{q_{rs_{ref}}^j}.$

2.4.3 Sampling active power generation

Since the voltages of the generation at PV buses are fixed to their nominal values, we only considered the sampling of the active power generation p_g. To sample generation correctly, we have to ensure that the power balance condition is satisfied:

$$\sum_{k=1}^{n_u} p_g^k = \sum_{i=1}^{n_L} p_l^i - \sum_{j=1}^{n_R} p_{rs}^j + losses \tag{2.45}$$

Consequently, the total power generation is approximately some percentage ϱ higher than the total demand, covering losses in the power system. In our case, this percentage ϱ is calculated as the ratio of the referent values between the total generation and the total load for our test systems[1]. Following [Donnot, 2019], we first formulate the part in which each production sample has to target the total demand increased by losses percentage as:

$$\breve{p}_g^k = \psi^k \cdot \varrho \cdot \frac{\sum_{i=1}^{n_L} p_l^i}{\sum_{k=1}^{n_u} p_{g_{ref}}^k} \cdot p_{g_{ref}}^k$$

where $\psi^k \sim \mathcal{U}[0.8, 1.2]$ is the uniform distribution to ensure no deterministic behavior for the productions.

We applied the second step to ensure that the relation $\sum_{k=1}^{n_u} p_g^k + \sum_{j=1}^{n_R} p_{rs}^j = \varrho \cdot \sum_{i=1}^{n_L} p_l^i$ holds perfectly, as:

$$p_g^k = \breve{p}_g^k \cdot \varrho \cdot \frac{\sum_{i=1}^{n_L} p_l^i}{\sum_{k=1}^{n_u} \breve{p}_g^k}$$

2.5 Conclusion

This chapter is dedicated to introducing the fundamentals of the power grid, with a specific focus on the transmission power system. The state of the power system is described by power system variables that can be calculated using the power flow equations. The AC power flow equation is explained in detail, and its simplification and linearization using DC power flow equations are demonstrated.

The chapter also explores the purpose of OPF and its solutions, which are limited to economic dispatch problems since other tasks are beyond the scope of this thesis. The discussion includes both deterministic and stochastic OPF, with an emphasis on the chance-constrained approach for solving stochastic OPF. This approach can save costs by allowing violations of operational constraints with low probability, but it remains challenging to solve.

Additionally, the chapter outlines a process for generating the synthetic dataset required for the contribution part of this thesis. This dataset will aid in the conducting

---

[1]For tested IEEE 9 bus system ϱ=1.0139 that means the total losses are 1.39% of the total output. Similarly, for IEEE 39 bus system ϱ=1.0086 and IEEE 118 ϱ=1.0322

of the proposed contribution, and the chapter provides a brief explanation of its generation.

Overall, this chapter provides a thorough introduction to the fundamental concepts of power systems, with a particular emphasis on the transmission power system. It is a valuable resource for readers to develop a comprehensive understanding of this complex and vital field.

Chapter 3

Gaussian Process

3.1 Introduction

Chapter 3 serves to provide the reader with the fundamentals of machine learning, with a specific emphasis on supervised learning. The concept of supervised learning will be clearly explained, along with a thorough explanation of key terms within the field.

The second part of the chapter focuses on the Gaussian Process, as a model for solving regression problems within the context of supervised learning. The Gaussian Process model will be explored in detail, specifically within the domain of regression.

Overall, the chapter is divided into four sections, with section 3.2 serving as an introduction to the main concepts of machine learning, emphasizing supervised learning. Section 3.3 dives deeper into the specifics of the Gaussian Process model, while section 3.4 provides a concise summary of the chapter's key points.

3.2 Supervised Learning

In this section, we will provide a brief overview of the fundamental concepts of machine learning (ML). Subsequently, we will delve into the task of supervised learning in machine learning and explore the associated terminology in greater depth. By providing a detailed explanation of these key concepts, we aim to facilitate a better understanding of supervised learning tasks.

3.2.1 Fundamentals of machine learning

Machine learning is a rapidly evolving field that has seen significant advancements in recent years. It is a field of artificial intelligence that focuses on developing algorithms and models that can learn from data and make predictions or decisions based on that learning. It has become increasingly important in recent years due to its ability to work with large amounts of data, so-called big data, and has the potential to revolutionize the way of solving problems and automating tasks. Therefore, one of the key advantages of machine learning is that it can handle large amounts of data

much more efficiently than traditional methods. This is because machine learning algorithms can extract patterns and structure from the data, allowing them to make more accurate predictions and decisions. As a result, machine learning is increasingly being used in applications such as image and speech recognition, natural language processing, and fraud detection [Vinoth and Datta, 2022, Zhou, 2021].

There are three primary types of machine learning tasks:

1. supervised learning;

2. unsupervised learning;

3. reinforcement learning.

Each type of task is suited to different types of problems and data sets. The types of machine learning are visually illustrated in Fig. 3-1.
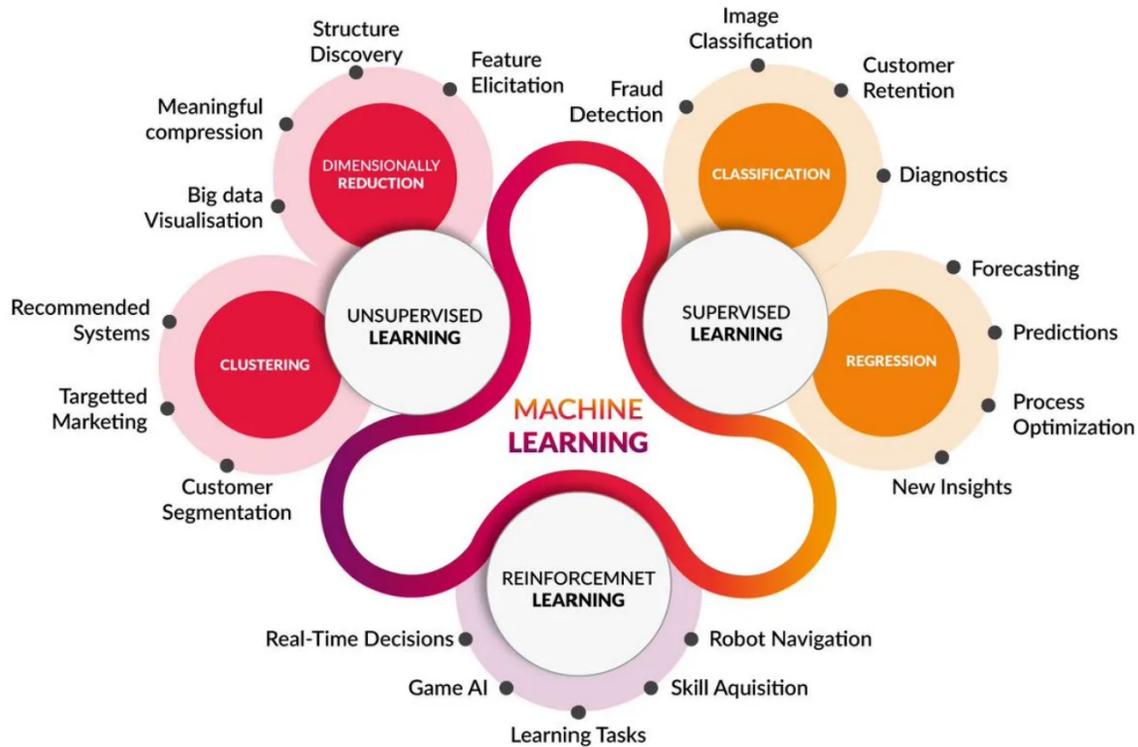
Figure 3-1: Types of machine learning tasks; the figure is taken from https://medium.com.

Supervised learning is one of the most widely used types of machine learning tasks. It involves training a model on labeled data, where the desired output is already known. This allows the model to learn to predict the output for new input data. In classification tasks, the output is a label or category, while in regression tasks, the output is a continuous value. Supervised learning has many applications, including spam detection, medical diagnosis, weather forecasting, etc [Cunningham et al., 2008, Hastie et al., 2009].

Unsupervised learning is another important type of machine learning task. It involves training a model on unlabeled data, where the desired output is not known. The objective of unsupervised learning is to discover patterns or structures in the data. This can be useful for tasks such as clustering, where the goal is to group similar data points together, or for dimensionality reduction, where the goal is to reduce the number of features in the data [Ghahramani, 2004, Usama et al., 2019].

Reinforcement learning is a type of machine learning that involves training a model to make decisions in an environment to maximize a reward signal. This type of learning is commonly used in robotics and game playing, where the model learns to take actions that will lead to a desired outcome based on feedback from the environment. Reinforcement learning is a powerful tool for creating intelligent agents that can interact with the world in a sophisticated way [Kaelbling et al., 1996, Li, 2017].

In addition to these primary machine learning tasks, there are also related concepts such as deep learning, semi-supervised learning, transfer learning, and adversarial learning. Deep learning is a subfield of machine learning that involves training models with multiple layers of neural networks [LeCun et al., 2015]. This allows the models to learn complex representations of the input data, making them highly effective for tasks such as image and speech recognition. Deep learning has seen significant advancements in recent years, leading to breakthroughs in areas such as computer vision and natural language processing. Semi-supervised learning involves training a model on a combination of labeled and unlabeled data [Van Engelen and Hoos, 2020], while transfer learning involves using a pre-trained model as a starting point for a new task [Weiss et al., 2016]. Adversarial learning involves training models to be robust against attacks or adversarial inputs [Lowd and Meek, 2005].

Overall, machine learning is a powerful tool that has the potential to transform the way we solve problems and automate tasks. It requires expertise in mathematics, statistics, and computer science, but can lead to significant advancements in a wide range of fields. As data continues to become more abundant, the importance of machine learning is only expected to grow.

3.2.2 Supervised learning

Supervised learning is a widely used machine learning technique that involves learning the probabilistic relationship between input data $x^{in} \in \mathcal{X}$ and corresponding output data $y^{out} \in \mathcal{Y}$. This is accomplished through the use of a labeled training set $\{(x_i^{in}, y_i^{out}) | i = 1, ..., m_s\}$, where each sample is a pair consisting of an input observation and its associated output. The ultimate objective is to develop a prediction function f that can accurately predict the output for new, previously unseen samples.

In this section, we will focus specifically on supervised learning for the regression task, which involves predicting a continuous output variable. It is important to note that supervised learning also includes the classification task, which involves predicting a discrete output variable. However, this topic will not be covered here. The main difference between these two frameworks lies in the nature of the output set $\mathcal{Y}$.

To train a supervised learning algorithm, a loss function ℓ_c is utilized to measure the degree of difference between the predicted output and the true output, also known

as the label. The goal of the algorithm is to minimize the average error on the training set, which is referred to as the empirical error. This process of minimizing error is achieved through the principle of Empirical Risk Minimization (ERM) [Burashnikova, 2022]. The aim is to minimize empirical error, and thus reduce the generalization error. Generalization error implies an error of the prediction function made on new unseen samples. The relationship between empirical error and generalization error, aiming of finding a balance between a low empirical error and a high complexity of the prediction function by choosing the appropriate learning function, is a key area of study in statistical learning theory [Vapnik, 1999]. This process of choosing the right learning function is known as the Structural Risk Minimization (SRM) principle [Burashnikova, 2022].

Fundamental assumptions

In statistics, a crucial assumption is that all samples are independently and identically distributed (i.i.d.) from an unknown population distribution $\mathcal{D}$. This means that a sample set denoted as $\int$ and consisting of pairs of input samples x^{in} and their corresponding outputs y^{out}, is generated randomly and without any dependence between samples. This assumption is essential for determining the representativeness of a training or test dataset, where the input samples and their corresponding outputs are generated from the same source.

Another essential notation in statistic learning is an error (risk or loss). In the machine learning community, this error is measured through the loss function. The loss function evaluates the error between prediction $f(x^{in})$ and actual (desired) output y^{out}. Therefore, the loss function implies a distance over predicted and desired output and can be expressed in general form as:

$$\ell_c(f(x^{in}), y^{out})$$

Generalization error and empirical error

According to ℓ_c, the error on all examples $(f(x^{in}), y^{out})$ can be defined. This error is called a generalization error and is given as:

$$\mathcal{L}(f) = \mathbb{E}_{\mathcal{D}} \left[\ell_c(x^{in}, y^{out}) \right] \tag{3.1}$$

The prediction function f is responsible for minimizing the error on new, unseen data, which is crucial for achieving good performance. However, it is impossible to directly estimate this error because the distribution of the population is unknown. To address this, we can optimize the average error on the training set by searching for the most suitable function f in set of functions $\mathcal{F}$. By doing so, we can estimate the generalization error on $\mathcal{S}$, also called the empirical error, as:

$$\hat{\mathcal{L}}(f, S) = \frac{1}{m_s} \sum_{i=1}^{m_s} \ell_c(f(x_i^{in}), y_i^{out}) \tag{3.2}$$

In regression problems, the choice of an appropriate loss function is crucial in accurately measuring the difference between predicted and actual values. The most commonly used loss functions for error measuring in regression problems are Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Squared Logarithmic Error (MSLE).

MAE is a simple and robust loss function that measures the average absolute difference between the predicted and actual values. It is particularly useful in cases where outliers are present and variables do not have a strict Gaussian distribution. The MAE is defined as:

$$\mathcal{L} = \frac{1}{m_s} \sum_{i=1}^{m_s} \| f(x_i^{in}) - y_i^{out} \| \tag{3.3}$$

MSE measures the average squared difference between the predicted and actual values. It is sensitive to outliers and larger errors are penalized more heavily. Accordingly, MSE is suitable for problems where minimizing large errors is important. MSE is defined as:

$$\mathcal{L} = \frac{1}{m_s} \sum_{i=1}^{m_s} (f(x_i^{in}) - y_i^{out})^2 \tag{3.4}$$

RMSE is the square root of MSE. It is more useful in practice since it is measured in the same units as the actual values. Therefore, the results are easier to interpret. RMSE is defined as:

$$\mathcal{L} = \sqrt{\frac{1}{m_s} \sum_{i=1}^{m_s} (f(x_i^{in}) - y_i^{out})^2} \tag{3.5}$$

MSLE measures the average squared logarithmic difference between the predicted and actual values. It is particularly beneficial in scenarios where the target variable has a wide range of values, and the objective is to minimize the relative error instead of the absolute error. MSLE is defined as:

$$\mathcal{L} = \frac{1}{m_s} \sum_{i=1}^{m_s} (log(f(x_i^{in})) - log(y_i^{out}))^2 \tag{3.6}$$

ERM principle

To learn the prediction function f, we use the input and output data from the training dataset to find the best fit between them. The ERM principle suggests that if the training samples from set S are representative of the distribution D, then the empirical error is a reliable estimate of the generalization error. By minimizing the empirical error on a given training set, we can also minimize the generalization error. In summary, the ERM principle returns as a way to optimize the function f by minimizing the empirical error on the training set as follows:

$$f_S = arg \min_{f \in \mathcal{F}} \frac{1}{m_s} \sum_{i=1}^{m_s} \ell_c(f(x_i^{in}), y_i^{out}) \tag{3.7}$$

49

Learning algorithm complexity

In supervised learning, the ability of a learning algorithm to generalize to new, unseen data is reflected in the complexity of the learned function. To evaluate function complexity, we compare the empirical error (the error on the training data) and the generalization error (the error on the unseen data). If the empirical error is small, but the generalization error is high, this indicates *overfitting*, meaning that the learned function is too complex for the training data. To avoid *overfitting*, simpler functions should be learned. Apart from identifying an appropriate learning function f, overfitting can be prevented by implementing various techniques such as early stopping, pruning, regularization, etc.

However, if the learning function is too simple, it may not have sufficient expressiveness to capture the underlying patterns in the data, leading to *underfitting*. In this case, both the empirical and generalization errors will be high. Therefore, it is important to strike a balance between function complexity and performance. This trade-off between low empirical error and complex function is known as the *bias-variance* trade-off. Achieving a balance between function complexity and generalization is crucial to prevent *overfitting* or *underfitting* in machine learning. This involves finding the right level of learning function complexity that can effectively capture the underlying patterns in the data, while also being simple enough to avoid *overfitting*. In Fig. 3-2, all properties of ERM (*overfitting*, *underfitting*, and *bias-variance* trade-off) are illustrated with a toy example for a regression problem.
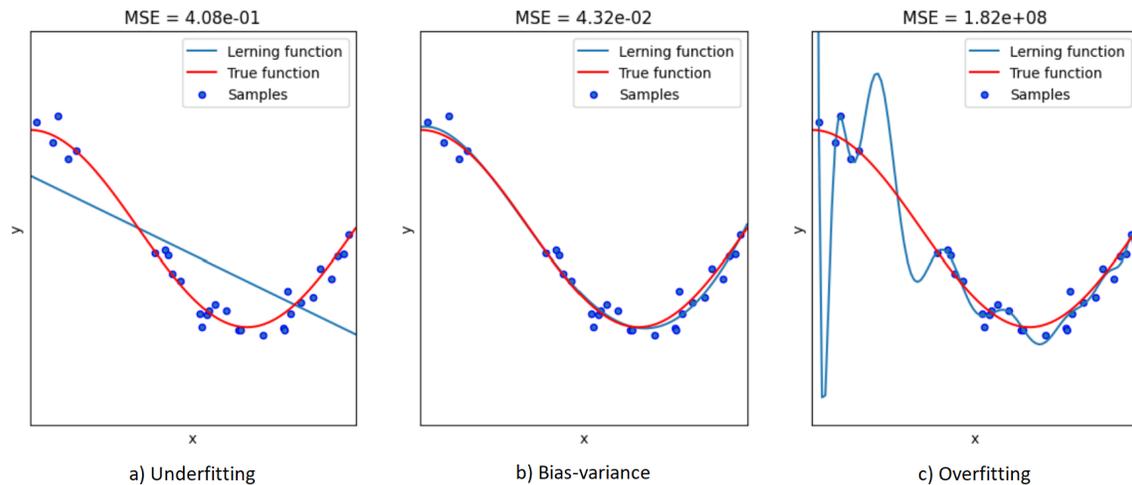

Figure 3-2: Illustration of the ERM properties on a regression problem using a toy function.

SRM principle

The SRM principle aims to provide a bias-variance trade-off in supervised learning by finding the most suitable learning function f from set $\mathcal{F}$. However, selecting an optimal learning function can be a challenging task. To simplify this process, one can use a simple strategy that involves applying the ERM algorithm to find a function and

then calculating the bound on the generalization error. The learning function that minimizes this bound provides the best theoretical guarantee on the generalization error compared to other function candidates. Therefore, this function can be selected as the appropriate learning function f. By following this approach, one can effectively minimize the risk of overfitting and improve the generalization performance of a supervised learning model.

3.3 Gaussian Process

In recent years, deep neural networks have gained popularity in the supervised machine learning field, while the Gaussian Process remains a widely used framework in probabilistic machine learning. The key difference between these models lies in how they make predictions. Neural networks optimize weights to fit functions and return deterministic outputs, whereas the Gaussian process uses kernel functions to interpolate training data and returns probabilistic outputs distributed as a Gaussian. As a result, the Gaussian Process is a nonparametric model that can provide measures of uncertainty for each point prediction.

A comprehensive introduction to the Gaussian Process is given in [Rasmussen et al., 2006]. The research has shown that GP can be successfully established to propagate uncertainty in multi-step predictions, as demonstrated in [Girard et al., 2002]. Consequently, the Gaussian Process has been effectively used as a predictive model in model predictive control [Murray-Smith et al., 2003, Kocijan et al., 2004]. However, this thesis focuses on one-step prediction by propagating input uncertainties to the predicted output. The idea of using the Gaussian Process in the stochastic OPF problem was inspired by [Girard et al., 2002] and [Hewing et al., 2019], who found that cautious control and uncertainty propagation in model predictive control produced effective results.

3.3.1 Gaussian Process Framework

A Gaussian Process is a statistical model based on Bayesian inference that defines a distribution over functions. In a Gaussian process, every point in a function is modeled as a random variable, and the joint distribution of all these variables is a multivariate Gaussian distribution. The GP function $f(x^{in})$ can be represented as an infinite vector, where each input x^{in} in the vector corresponds to the function value at a specific input. However, working with an infinite object is not practical. The advantage of a Gaussian process is that it allows us to perform inference using a finite subset of points and still obtain the same results as we would with the entire infinite vector. According to [Rasmussen et al., 2006], the Gaussian Process is a collection of random variables, any finite number of which have a joint Gaussian distribution.

For the given input x^{in}, the output y^{out} is given by:

$$y^{out} = f(x^{in}) + \zeta \tag{3.8}$$

where $\zeta \sim \mathcal{N}(0, \sigma_\zeta^2)$ is Gaussian measurement noise and $f : \mathcal{R}^{\mathbb{D}} \to \mathcal{R}$. The GP function $f(x^{in})$ is defined by a mean function $\mu(x_p^{in})$ and a covariance function $k(x_p^{in}, x_q^{in})$:

$$\mu(x_p^{in}) = \mathbb{E}[f(x_p^{in})] \tag{3.9a}$$

$$k(x_p^{in}, x_q^{in}) = \mathbb{E}[(f(x_p^{in}) - \mu(x_p^{in}))(f(x_q^{in}) - \mu(x_q^{in}))] \tag{3.9b}$$

where the mean function represents the expected value of the function at any point, while the covariance function measures the correlation between the values of the function at different points. Accordingly, the GP function can be defined as:

$$f(x^{in}) = GP(\mu(x_p^{in}), k(x_p^{in}, x_q^{in})) \tag{3.10}$$

3.3.2 Bayesian inference

The Gaussian Process is a statistical model that relies on Bayesian inference. According to [Gelman et al., 1995], Bayesian inference involves building a probability model based on a given set of training data, where a set of training data is defined as:

$$D = \{(x_i^{in}, y_i^{out}), \quad i = 1, ..., m_s\}$$

In Bayesian inference, two terms are used: prior inference and posterior inference. Prior inference implies inference without data, while posterior inference is inference with data.

Prior inference

The Gaussian process uses training data to interpolate between them via kernel function. However, the quality of the interpolation depends on good prior inference, which involves guessing the function without relying solely on the available training data. This is particularly important for regions of the function where no training data are available.

To make an accurate prior inference, it is important to define the mean and covariance functions from 3.9. The most commonly used prior mean function is zero, but if the behavior of the physical system is known, choosing a different type of prior mean function may be more appropriate (such as a linear function). However, the choice of covariance function, which takes into account the features of the data, is the most critical aspect of prior inference. There are many different types of covariance functions [Rasmussen et al., 2006]. Moreover, combining these functions can often result in a better fit for the model. The most popular and used covariance functions, depending on the type of problem, are:

1. Squared exponential (SE): $k(x_p^{in}, x_q^{in}) = exp(-\frac{1}{2}\frac{|x_p^{in} - x_q^{in}|^2}{l^2})$

2. Matern: $k(x_p^{in}, x_q^{in}) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\frac{\sqrt{2\nu}|x_p^{in} - x_q^{in}|}{l})^\nu K_\nu(\frac{\sqrt{2\nu}|x_p^{in} - x_q^{in}|}{l})$

3. Rational quadratic (RQ): $(1 + |x_p^{in} - x_q^{in}|)^{-\iota}$

4. Periodic: $exp(-\frac{2sin^2(\frac{(x_p^{in}-x_q^{in})}{2})}{l^2})$

where ν, l, ι are hyperparameters; $K_\nu(\cdot)$ is a modified Bessel function; $\Gamma(\cdot)$ is the gamma function. Enumerated covariance functions are illustrated in Fig. 3-3.
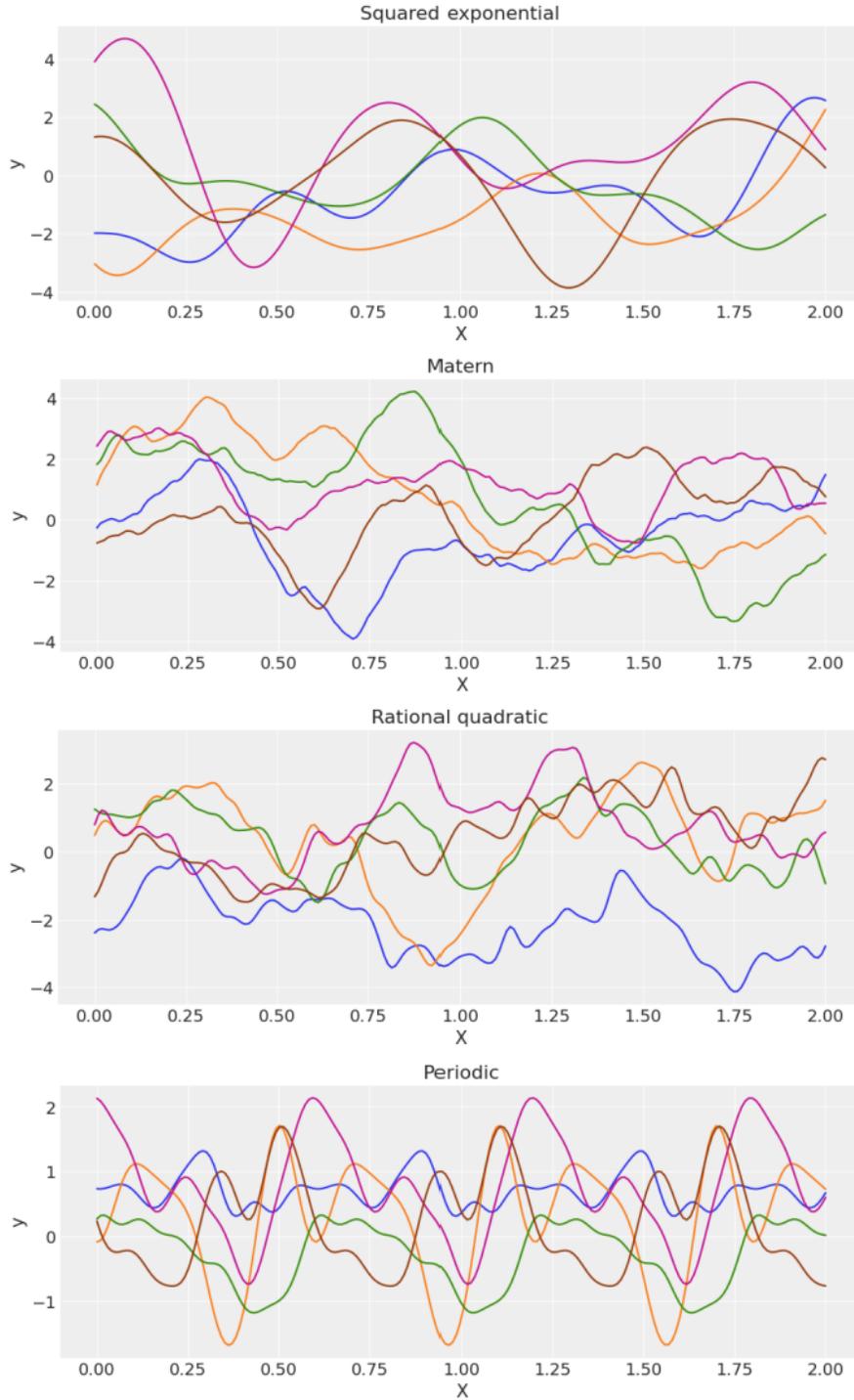

Figure 3-3: Illustration of the most popular covariance functions.

In this thesis, we will utilize a squared exponential covariance function. This

function is well-suited for our problem due to its inherent smoothness, which makes it capable of effectively fitting non-linear functions while avoiding superfluous local minima. Specifically, we will be employing a variant of the squared exponential function known as Squared Exponential Automatic Relevance Determination (SEard). In SEard, a unique length scale l hyperparameter is assigned for each input, resulting in a more specialized and effective model. SEard covariance function is given as:

$$k(x_p^{in}, x_q^{in}) = \sigma_f^2 exp(-\frac{1}{2}(x_p^{in} - x_q^{in})\Lambda^{-1}(x_p^{in} - x_q^{in})) \tag{3.11}$$

where σ_f^2 is a signal variance and $\Lambda = diag([l_1^2, l_2^2, ..., l_{\mathbb{D}}^2])$ is diagonal matrix of length-scales.

According to the SE function, covariance k is close to one when input points are close indicating that the points are similar. On the other hand, as the distance between the points increases, the covariance decreases exponentially, indicating that the points are not similar.

Posterior inference

As previously mentioned, prior inference evaluates the model without data, based on hyperparameters. While this is an important part of GP, it has limitations compared to the posterior part, which takes data into account. To address the differences between the prior and posterior, as well as to insight into how the Gaussian Process Regression (GPR) model works, we will consider a toy sine function. Assuming a noise signal y^{out}, then the SE function can be expressed as:

$$k(x_p^{in}, x_q^{in}) = \sigma_f^2 exp(-\frac{1}{2}(x_p^{in} - x_q^{in})\Lambda^{-1}(x_p^{in} - x_q^{in})) + \sigma_\zeta^2 \delta_{pq} \tag{3.12}$$

where σ_ζ^2 is noise variance that provides avoiding over-fitting the model on the noise; δ_{pq} is an element of the identity matrix I. Accordingly, the vector $\Theta = [\Lambda, \sigma_f^2, \sigma_n^2]$ consists of hyperparameters that need to be optimized during GPR training. Here we will illustrate with a toy sine function how each of these hyperparameters affects model fit.

In general, for the given input data matrix $X^{in} = [x_i^{in}]_{1 \leq i \leq m_s}^T \in \mathbb{R}^{m_s \times \mathcal{D}}$, each output dimension $y^{out} \in \mathbb{R}^{m_s}$ is learnt independently. A GP with prior mean function $\mu(\cdot)$ and chosen kernel $k(\cdot, \cdot)$ in each output dimension, models the output y^{out} as normally distributed data:

$$y^{out} \sim \mathcal{N}(\mu(X^{in}), K + \sigma_\zeta^2 I) \tag{3.13}$$

where σ_ζ^2 is the variance of measurement noise ζ in (3.8); identity matrix I is $I = diag(\delta_{pq})$; $\mu(X^{in}) = [\mu(x_1^{in}), ..., \mu(x_{m_s}^{in})]^T$, and $K_{pq} = k(x_{pg}^{in}, x_{pg}^{in})$ is the Gram matrix of the sampled data points. This work is based on the common zero prior mean $\mu(x^{in}) = 0$.

In this comparison, we will analyze the prior, with zero mean and variance $\sigma_f^2 = 0.5$, and the posterior distribution of the sine curve. The prior distribution indicates that

the sine curve will usually be within 95% (percentile) of the zero mean, as illustrated by the 10 random functions generated from the prior distribution in Fig 3-4. Fig 3-4 depicts a zero mean function as a red dashed line, while 95% of the signal variance is represented by the red area.
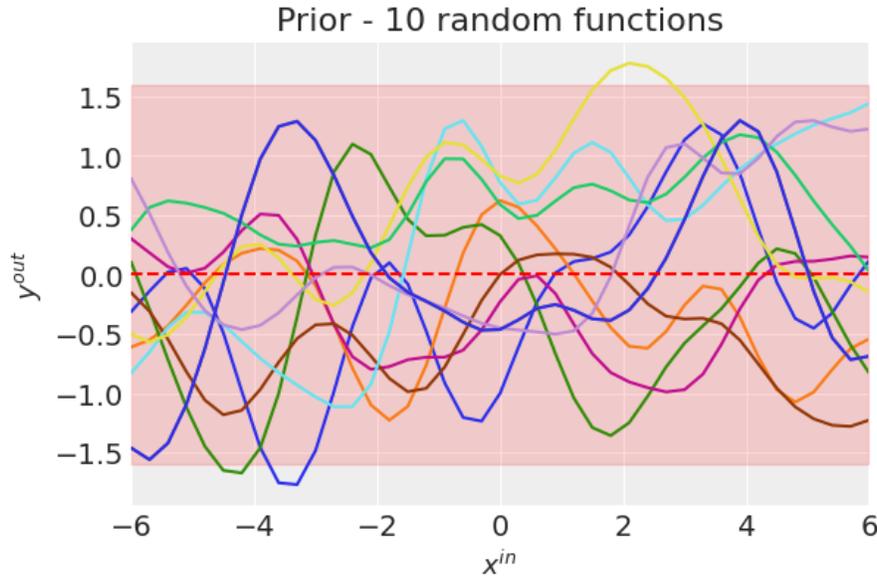

Figure 3-4: Illustration of 10 random functions with the prior distribution.

However, when training points are available, the posterior distribution changes. Fig. 3-5 shows the same 10 random functions, but with the posterior distribution after 8 training points have been added. We can see that the functions are closer to the training points, indicating that if a new unseen point is close to the training points, it is highly probable that it will be close to the spline between the known points. Nevertheless, if the unseen point is too far from the training points, the posterior distribution will revert to the prior distribution. This implies that the prior distribution still holds for points that are far from the training data.

To verify the aforementioned claim, Fig. 3-6 illustrates examples of the sine curve with and without some training points. It is evident that uncertainty increases when the training data points are widely spaced, and the predicted distribution approaches the prior distribution. In other words, the prior distribution becomes more significant when we have fewer data or data that is not well distributed in the space, and the predicted distribution becomes less confident.
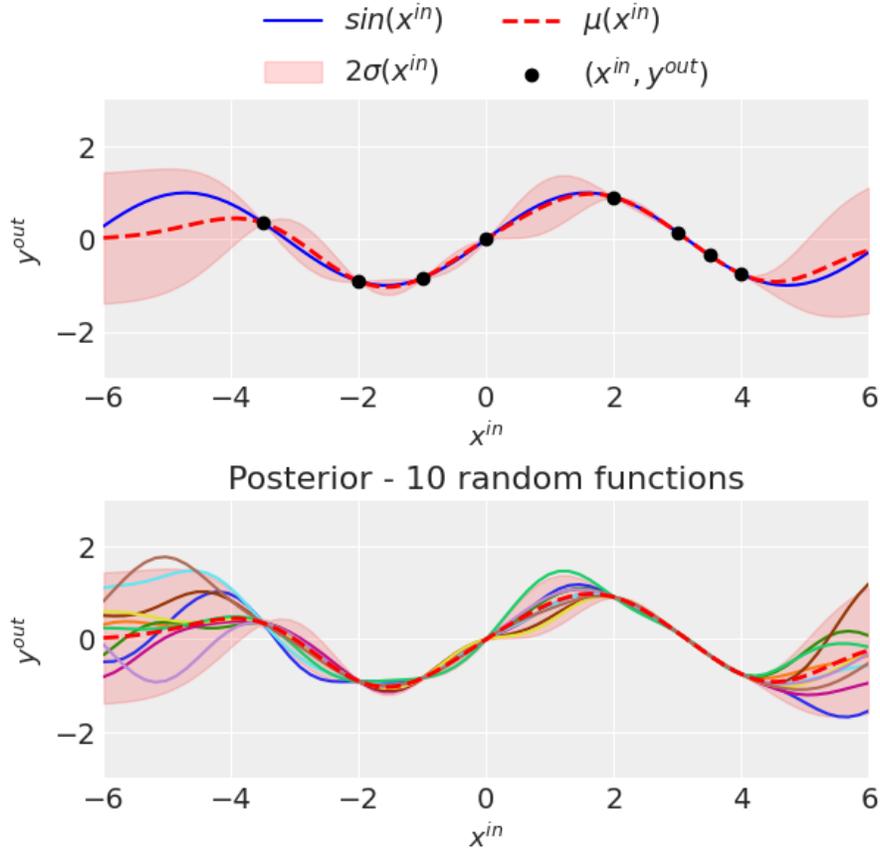
Figure 3-5: Illustration of 10 random functions with the posterior distribution distributions based on 8 training data.

In this context, we can examine how different hyperparameters impact the GP function. One hyperparameter of interest is the noise variance σ_ζ^2, which can disrupt the signal. To better understand how noise variance affects GP prediction, in Fig. 3-7 we visualize two scenarios with noise variance $\sigma_\zeta^2 = 0.2$ and zero. In the absence of noise, the GP function can exactly fit all the training data points, and the posterior variance at these points is zero. However, when we consider the possibility of noise, we become uncertain about the true location of the signal, and this uncertainty leads to higher posterior variance at the training points. The GP mean, which represents the best estimate of the underlying function, tries to fit the data points while accounting for this uncertainty. In the presence of noise, the GP mean may not pass through all the training points, trying to find the simplest path that fits the data points while accounting for the noise.

Unlike noise variance, signal variance σ_f^2 is a key factor in predicting the certainty of a model's output, as it is independent of the training data and reflects the underlying variability of the signal being modeled. While it does not affect the mean prediction, it does have a significant impact on the variance of the predictions. When the signal variance is low, the model's uncertainty is naive when predicting data far from training data, which can result in overly optimistic predictions. Conversely, when the signal

variance is high, the model's uncertainty is more conservative and it becomes more difficult to determine the underlying signal from the noise. This is illustrated in Fig. 3-8 where we compare signal variances of 0.5 and 3. The appropriate signal variance will depend on the specific application and the level of uncertainty precision required in the model's predictions.
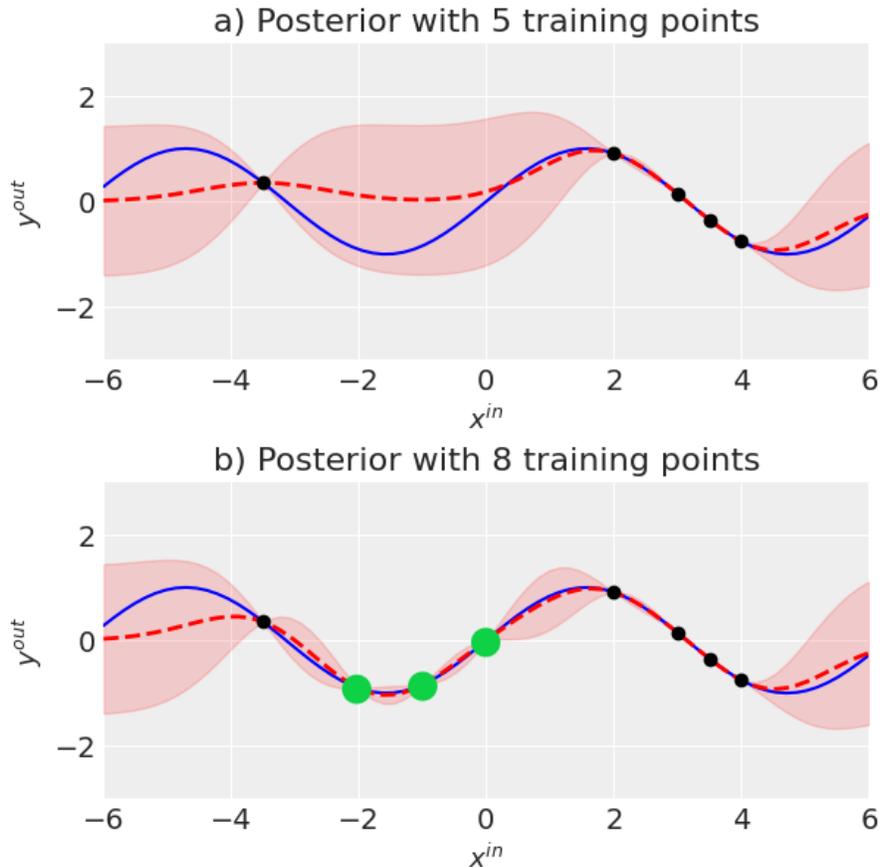

Figure 3-6: Comparing the posterior distribution with different numbers of training points. a) 5 training points are used to fit GP; b) 3 new training points are added (green dots).

The length-scale l^2 is a crucial hyperparameter in GP modeling. It determines the shape of the spline that links the training points in the GP function. A low length-scale will result in a steep and aggressive function, while a higher length-scale will produce a smoother one. Therefore, the length scale affects the flexibility of the GP function, and choosing the appropriate value is crucial for achieving good performance in finding new data. The impact of the length-scale hyperparameter is demonstrated in Fig. 3-9 for the $l^2 = 1$ and $l^2 = 0.2$.
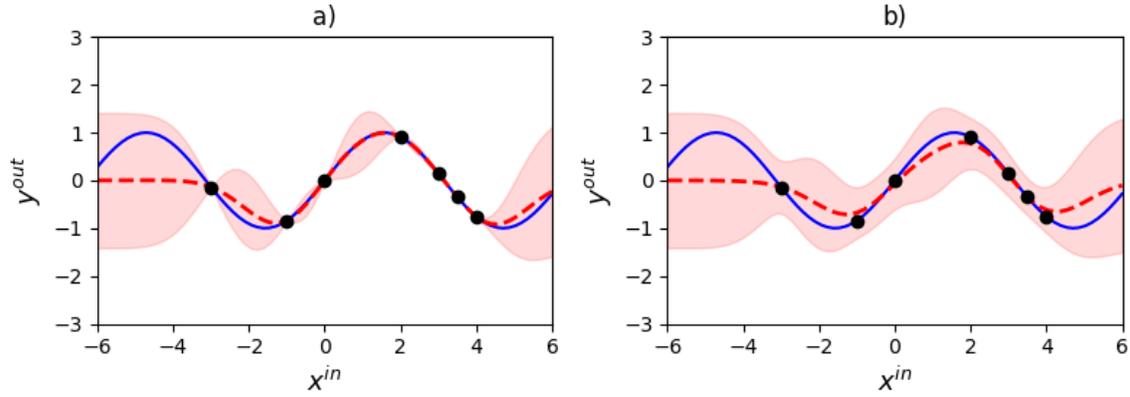
Figure 3-7: Posterior distribution obtained by adjusting noise variance σ_ζ^2. a) signal with $\sigma_\zeta^2 = 0$; b) signal with $\sigma_\zeta^2 = 0.2$.
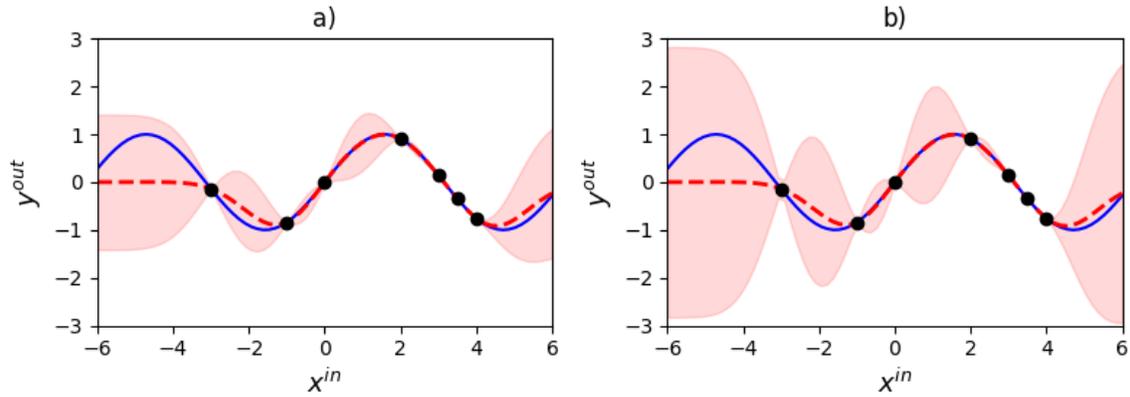

Figure 3-8: Posterior distribution obtained by adjusting signal variance σ_f^2. a) signal with $\sigma_f^2 = 0.5$; b) signal with $\sigma_f^2 = 3$.

In this exploration, we investigate how the distribution and spread of training data in space can impact the fitting of the GP function. The ideal distribution of training data points, where the sin function is perfectly fitted by the GP function, is demonstrated in Fig. 3-9 a). However, this ideal distribution is rare or even impossible in practice. To ensure a good spread of data, a Latin hypercube design can be used to distribute the training data, as shown in Fig. 3-9 b). Fig. 3-9 c) displays the worst-case distribution where the GP function fails to fit the sin function at all. In practice, it is common to encounter examples where the distribution fits some parts of the function well while leaving other parts equal to the prior, as illustrated in Fig. 3-9 d). Therefore, achieving a high-quality distribution of training data points is one of the crucial factors in obtaining a good GP model.
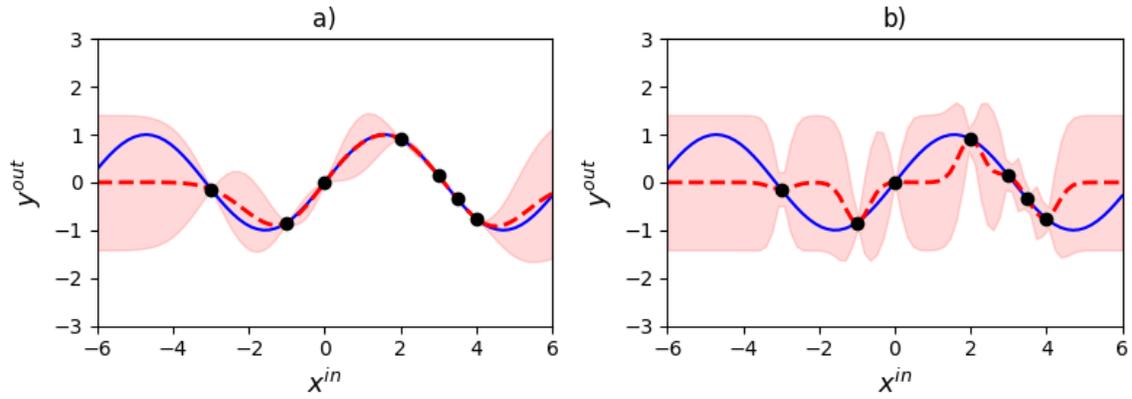
Figure 3-9: Posterior distribution obtained by adjusting length-scales l^2. a) signal with $l^2 = 1$; b) signal with $l^2 = 0.2$.
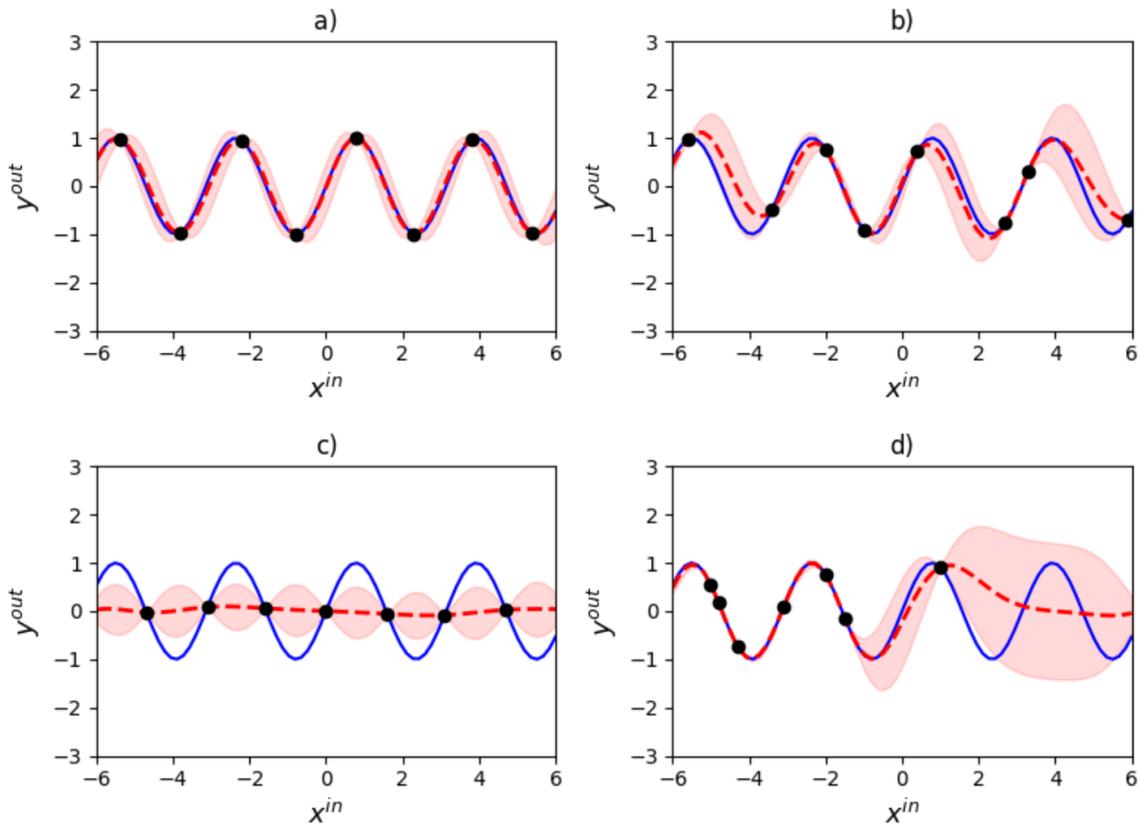
Figure 3-10: Illustrations of the GP results using different sampling distributions with the same hyperparameters. All cases have 10 points in the distribution. a) data points are distributed manually to cover the minimum/maximum of input and output data; b) data points are distributed using the Latin hypercube design; c) data points are uniformly distributed with the worst case distribution; d) data points are uniformly distributed with a random distribution.

3.3.3 Optimization of hyperparameters

In the previous section, it was demonstrated that hyperparameters $\Theta = [\Lambda, \sigma_f^2, \sigma_\zeta^2]$ play a crucial role in determining the behavior of a GP function. However, it is not feasible to manually set hyperparameter values during the training process. Therefore, it is essential to optimize hyperparameters automatically. To achieve this, we use the Maximum Likelihood Estimate (MLE) method with the log marginal likelihood. In this method, we treat the hyperparameters as unknown parameters and try to find their optimal values by maximizing the likelihood function. To simplify the optimization process, we assume that the length-scales of the GP function are independent of each other. This means that the hyperparameter matrix, denoted by Λ, is a diagonal matrix.

The likelihood function is the probability of the observed data given the hyperparameters. Following [Rasmussen et al., 2006], the MLE with log marginal likelihood is given as:

$$
\begin{aligned}
\log\ p(y^{out}|X^{in}, \Theta) &= \int p(y^{out}|\mathrm{w}, X^{in}, \Theta)\mathrm{dw} \\
&= \frac{1}{2}(y^{out} - \mu(X^{in}))^T(K + \sigma_\zeta^2 I)(y^{out} - \mu(X^{in})) \\
&\quad - \frac{1}{2}\log|K + \sigma_\zeta^2 I| - \frac{N}{2}\log 2\pi
\end{aligned}
\tag{3.14}
$$

where w are the weights of all parameters, and $\mu(X^{in})$ represents a prior mean value. Since we are considering a GP with zero prior mean, Eq. 3.14 can be simplified as:

$$
\log\ p(y^{out}|X^{in}, \Theta) = \frac{1}{2}(y^{out})^T(K + \sigma_\zeta^2 I)y^{out} - \frac{1}{2}\log|K + \sigma_\zeta^2 I| - \frac{m_s}{2}\log 2\pi
\tag{3.15}
$$

where each term of the Eq. 3.15 has interpretative roles. For example, the first term $\frac{1}{2}(y^{out})^T(K + \sigma_\zeta^2 I)y^{out}$ represents the data-fit term, which measures how well the GP function fits the observed data using only the available samples. This term decreases with the length-scales of the covariance function, which means that the GP function becomes less flexible. The second term $\frac{1}{2}\log|K + \sigma_\zeta^2 I|$ represents the complexity term, which penalizes overly complex models and encourages simplicity. This term increases as the length-scales of the covariance function increase, which means that the GP function becomes less complex. Finally, the last term $\frac{m_s}{2}\log 2\pi$ defines a normalization constant.

By maximizing the log marginal likelihood, we can find the optimal hyperparameters for the GP function. However, by using the negative log marginal likelihood as the objective function, any non-linear solver can be utilized. Therefore, the objective function will be minimized according to hyperparameters. The negative log marginal likelihood is a non-convex function, meaning that it may have multiple local minima. To overcome this, it is common practice to use a multi-start process with different initial guesses for the hyperparameters. This involves running the optimization algorithm multiple times with different initial values for the hyperparameters and selecting the

solution with the lowest negative log marginal likelihood as the optimal solution. Accordingly, the optimal hyperparameters are found as:

$$\Theta^* = argmin(-\log \ p(y^{out}|X^{in}, \Theta)) \tag{3.16}$$

By optimizing the hyperparameters using the negative log marginal likelihood, we can ensure that the GP function fits the observed data well and achieves good generalization performance on new data. However, the optimization process can be computationally expensive, especially for large datasets. Therefore, it is important to use efficient optimization algorithms. In this thesis, we utilize the Sequential Least Squares Programming (SLSQP) [Kraft, 1988] solver to determine the optimal solution.

3.3.4 Prediction with deterministic inputs

Predictive distribution of noise-free model y_*^{out} for a given unseen sample data x_*^{in} can be represent as marginal distribution with zero mean:

$$y_*^{out} \sim \mathcal{N}(0, K(x_*^{in}, x_*^{in})) \tag{3.17}$$

By combining the marginal distribution with the distribution of the GP using the training samples, we obtain a joint distribution:

$$p(y_*^{out}, y^{out}) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K(X^{in}, X^{in}) & k(X^{in}, x_*^{in}) \\ k(x_*^{in}, X^{in}) & k(x_*^{in}, x_*^{in}) \end{bmatrix}\right) \tag{3.18}$$

In the case of the nosy signal, the joint distribution can be generalized:

$$p(y_*^{out}|y^{out}) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K(X^{in}, X^{in}) + \sigma_\zeta^2 I & k(X^{in}, x_*^{in}) \\ k(x_*^{in}, X^{in}) & k(x_*^{in}, x_*^{in}) \end{bmatrix}\right) \tag{3.19}$$

Since the joint distribution of the variables is Gaussian, the resulting distribution is conditioned on the training data $p(y_*^{out}|y^{out}) = \mathcal{N}(\mu(x_*^{in}), \sigma^2(x_*^{in}))$ is also Gaussian [Rasmussen, 2003] with:

$$\mu(x_*^{in}) = k_*^T(K + \sigma_\zeta^2 I)^{-1} y^{out} \tag{3.20a}$$

$$\sigma^2(x_*^{in}) = k(x_*^{in}, x_*^{in}) - k_*^T(K + \sigma_\zeta^2 I)^{-1} k_* \tag{3.20b}$$

where $k_* = k(X^{in}, x_*^{in})$ is the vector of covariance functions and $K(X^{in}, X^{in})$ is a Gram matrix.

In general, the predictive distribution of multivariate output dimension y_*^{out} can be written as follows:

$$\mu(x_*^{in}) = [\mu_1(x_*^{in}), \mu_2(x_*^{in}), ..., \mu_{n_y}(x_*^{in})]^T \tag{3.21a}$$

$$\Sigma(x_*^{in}) = diag([\sigma_1^2(x_*^{in}), \sigma_2^2(x_*^{in}), ..., \sigma_{n_y}^2(x_*^{in})]) \tag{3.21b}$$

3.3.5 Prediction with uncertain inputs

Up until now, we have assumed that the input to the GP is deterministic, while the output is Gaussian with errors caused by measurement or modeling inaccuracies. However, if we are dealing with random input, where the uncertainties in the input need to be taken into account when propagating through the GP model, we can assume that the input follows a Gaussian distribution, denoted as $x_*^{in} = \mathcal{N}(\mu_{x_*^{in}}, \Sigma_{x_*^{in}})$. To obtain the distribution of the predicted output, which is also uncertain, we need to integrate over the input distribution using the following approach:

$$p(y_*^{out}|\mu_{x_*^{in}}, \Sigma_{x_*^{in}}) = \int p(y_*^{out}|x_*^{in})\, p(x_*^{in}|\mu_{x_*^{in}}, \Sigma_{x_*^{in}})\, dx_*^{in} \tag{3.22}$$

In the case of the general kernel functions, it is not possible to compute the posterior distribution analytically because the Gaussian input is mapped through a non-linear function. To address this issue, the predictive distribution can be approximated using a Gaussian distribution and utilize various methods to calculate its statistics, allowing propagation of the uncertainty through the model. In our context, we consider two techniques to propagate input uncertainties: Taylor Approximation (TA) and Exact Moment Matching (EM). These techniques are used to approximate complex functions and probability distributions, respectively. For a more comprehensive understanding of these techniques, we suggest referring to the works of Girard and Deisenroth [Girard et al., 2003, Deisenroth, 2010].

Taylor Approximation

By using TA approximation, the mean output $\mu(x_*^{in})$ is approximated by its first-order Taylor expansion. For the output variance $\sigma^2(x_*^{in})$, we use the first and second-order Taylor expansions around the mean output $\mu_{x_*}^{in}$ to approximate it.

When using a first-order Taylor expansion, the mean prediction at a random input x_*^{in} does not offer any correction beyond what is already provided by the zero-order, as in (3.20a):

$$\mu_{TA}(x_*^{in}) = k_*^T(\mu_{x_*^{in}})(K + \sigma_\zeta^2 I)^{-1} y^{out} \tag{3.23}$$

The first-order Taylor Approximation (TA1) of the output variance is given as follows:

$$\sigma_{TA1}^2(x_*^{in}) = \sigma^2(\mu_{x_*^{in}}) + \frac{\partial \mu_{TA}(\mu_{x_*^{in}})}{\partial \mu_{x_*^{in}}}^T \Sigma_{x_*^{in}} \frac{\partial \mu_{TA}(\mu_{x_*^{in}})}{\partial \mu_{x_*^{in}}} \tag{3.24}$$

as well as the second-order Taylor Approximation (TA2):

$$\sigma_{TA2}^2(x_*^{in}) = \sigma_{TA1}^2(x_*^{in}) + \frac{1}{2} tr\left(\frac{\partial^2 \sigma^2(\mu_{x_*^{in}})}{\partial \mu_{x_*^{in}} \partial \mu_{x_*^{in}}^T} \Sigma_{x_*^{in}} \right) \tag{3.25}$$

where $\sigma^2(\mu_{x_*^{in}})$ corresponds to 3.20b. In contrast to the deterministic input case in 3.20, using the first and second-order Taylor expansion leads to correction terms for the posterior variance and covariance. These correction terms consider both the

gradient of the posterior mean and the variance of the input of the GP. As a result, according to [Hewing et al., 2019], the complexity of prediction with TA1 is $O(n_y n_x^2 m_s^2)$. However, TA2 approximation requires additional computational effort as it involves solving for the second derivative.

Exact Moment Matching

For a zero prior mean function and SEard kernel, the predicted mean and variance can be computed analytically using the law of iterated expectations (Fubini's theorem) [Deisenroth, 2010]. Therefore, the predicted mean can be expressed as:

$$\mu_{EM}(x_*^{in}) = k_*^T(\mu_{x_*^{in}}, \Sigma_{x_*^{in}})(K + \sigma_n^2 I)^{-1} y \tag{3.26}$$

where the covariance function k_* is simultaneously a function of input mean and covariance:

$$\begin{aligned} k_*^T(\mu_{x_*^{in}}, \Sigma_{x_*^{in}}) &= \sigma_f^2 |\Sigma_{x_*^{in}} \Lambda^{-1} + I|^{-\frac{1}{2}} \cdot \\ &exp(-\frac{1}{2}(x^{in} - \mu_{x_*^{in}})^T (\Sigma_{x_*^{in}} + \Lambda)^{-1}(x^{in} - \mu_{x_*^{in}})) \end{aligned} \tag{3.27}$$

The variance of the predictive distribution is defined as follows:

$$\sigma_{EM}^2(x_*^{in}) = \sigma_f^2 - tr((K + \sigma_\zeta^2 I)^{-1} Q) + \beta^T Q \beta - \mu_{x_*^{in}}^2 \tag{3.28}$$

in which $\beta = (K + \sigma_\zeta^2 I)^{-1} y^{out}$ and elements of $Q \in \mathbb{R}^{m_s \times m_s}$ are computed as:

$$\begin{aligned} [Q]_{ij} &= \frac{k_*(x_i^{in}, \mu_{x_*^{in}}) k_*(x_j^{in}, \mu_{x_*^{in}})}{|2\Sigma_{x_*^{in}} \Lambda^{-1} + I|^{\frac{1}{2}}} \cdot \\ &exp((z_{ij} - \mu_{x_*^{in}})^T (\Sigma_{x_*^{in}} + \frac{1}{2}\Lambda)^{-1} \Sigma_{x_*^{in}} \Lambda^{-1}(z_{ij} - \mu_{x_*^{in}})) \end{aligned} \tag{3.29}$$

where $z_{ij} = \frac{1}{2}(x_i^{in} + x_j^{in})$. According to to [Hewing et al., 2019], the computational complexity in prediction with EM approximation is given as $O(n_y^2 n_x m_s^2)$

GP is limited in its applicability for relatively small systems or slow systems in dynamic contexts due to the fact that their computational complexity is directly influenced by the input and output dimensions, as well as the number of training points. This means that as these dimensions and the number of training points increase, the computational complexity also increases, which can make GP computationally difficult for larger systems. Therefore, the use of GP may be limited in such scenarios.

3.3.6 Computational complexity

As emphasized in the previous subsection, GP is a limited model in the context of scalability. Specifically, as the size of the system and the number of samples increases, the computational complexity of the GP model also increases. One of the primary reasons for this computational complexity is the computation of the inverse covariance

matrix $(K + \sigma_\zeta^2 I)^{-1}$. This matrix is an essential component of GP and its calculation becomes increasingly difficult as the size of the data set grows. This can make the use of GP challenging and computationally expensive in large-scale systems.

To minimize the computational effort in inverting $(K + \sigma_\zeta^2 I)^{-1}$ during prediction, it is possible to apply the Cholesky decomposition. Cholesky decomposition decompose matrix into $(K + \sigma_\zeta^2 I)^{-1} = L_{ch} L_{ch}^T$, where L_{ch} is the lower triangular matrix with all positive elements. Reducing the computational complexity is possible by pre-computing both the Cholesky L_{ch} and the linear β_{ch} terms as:

$$L_{ch} = chol(K + \sigma_\zeta^2 I) \tag{3.30a}$$

$$\beta_{ch} = L_{ch}^T \setminus (L_{ch} \setminus y^{out}) \tag{3.30b}$$

where $\setminus$ is the left matrix division.

Hence the mean and variance can be calculated as:

$$\mu(x_*^{in}) = k_*^T \beta_{ch} \tag{3.31a}$$

$$\sigma^2(x_*^{in}) = k(x_*^{in}, x_*^{in}) - \tau_{ch}^T \tau_{ch} \tag{3.31b}$$

where $\tau_{ch} = L_{ch} \setminus k_*$.

3.3.7 Sparse approximation

To handle the increasing computational complexity that can arise from large training data, one approach is to use inducing points to create a sparse approximation of the GP. This approach can significantly reduce the computational cost while still maintaining a high level of accuracy. However, selecting the appropriate set of inducing points is critical to obtaining a good sparse approximation. The inducing points should be representative of the entire input space to ensure that the resulting approximation captures the full complexity of the original GP. If the selection of inducing points is suboptimal, the accuracy of the approximation may suffer. Therefore, while sparse approximation can be a useful tool, the process of selecting inducing points requires careful consideration to achieve a successful approximation.

The literature has proposed several sparse methods that use inducing points, such as those presented by [Seeger et al., 2003, Smola and Bartlett, 2001, Quinonero-Candela and Rasmussen, 2005]. In this thesis, we adopt the variational learning approach, which jointly infers the inducing inputs and the kernel hyperparameters by maximizing a lower bound of the actual log marginal likelihood [Titsias, 2009]. The idea behind sparse Gaussian Processes is selecting a small set of points, known as inducing points, denoted by, $X_m^{in} = [x_i^{in}]_{1 \leq i \leq m_m}^T \in \mathbb{R}^{m_m \times n_x}$, $m_m \leq m_s$, to improve the approximation of the marginal likelihood. By using this subsample, the approximate posterior GP mean and covariance function depends on a mean vector μ_m and a covariance matrix

A_m as:

$$\widehat{\mu}(x_*^{in}) = k_{m*}^T (K_{mm} + \sigma_m^2 I)^{-1} \mu_m \tag{3.32a}$$

$$\widehat{\sigma}^2(x_*^{in}) = k(x_*^{in}, x_*^{in}) - k_{m*}^T (K_{mm} + \sigma_m^2 I)^{-1} k_{m*} + k_{m*}^T K_{mm} A_m K_{mm} k_{m*} \tag{3.32b}$$

where

$$\mu_m = \sigma_m^{-2} K_{mm} \Gamma K_{mn} y^{out}, \quad A_m = K_{mm} \Gamma K_{mm}$$
$$\Gamma = (K_{mm} + \sigma_m^{-2} K_{mn} K_{nm})^{-1}$$

where K_{mm} is $m_m \times m_m$ covariance matrix of $\{X_i^{in}\}_{i=1}^{m_m}$; K_{mn} is $m_m \times m_s$ covariance matrix between $\{X_i^{in}\}_{i=1}^{m_m}$ and the training set X^{in}; k_{m*} is the cross-covariance vector.

In general, sparse approximation is a technique that reduces the time complexity of GP prediction from $O(m_s^3)$ to $O(m_s m_m^2)$, while still preserving important information from the original data. This means that instead of having to compute all pairwise distances between m_s data points, we only need to compute m_s distances using a subset of informative points $m_m \leq m_s$. The resulting prediction accuracy is nearly as good as the full GP but with significantly less computational cost.

3.4   Conclusion

This chapter has presented a concise introduction to the fundamental concepts of machine learning, with a focus on supervised learning and regression tasks. The Gaussian process framework has been explained in detail as a powerful tool for supervised regression modeling.

While unsupervised and semi-supervised learning are two additional important frameworks in machine learning, they fall outside the scope of this thesis and have not been covered.

The purpose of this chapter is to provide a foundation for the subsequent introduction and application of the Gaussian process framework, which is the primary focus of this thesis.

Part II

Contribution

Chapter 4

Data Driven CC-OPF using Gaussian Process

4.1 Introduction

This chapter presents our first contribution in the form of a full GP CC-OPF approach that replaces the traditional AC-PF balance equations with GPR. The GPR is learned from the input and output data of the full AC-PF and is embedded in the CC-OPF problem. To propagate input uncertainties, various approximation techniques such as Taylor Approximation (TA) and Exact Moment Matching (EM) are used. This allows an analytical reformulation of the CC-OPF to determine the optimal solution by linearizing around the expected operating point and assuming normally distributed deviations. The proposed approach has been evaluated on various system buses such as IEEE9, 39 and 118. The approach and results of this chapter have been published in the *International Journal of Electrical Power & Energy Systems (IJEPES)* [2].

The rest of this chapter is organized as follows: Section 4.2 introduces the problem statement, research question, and hypothesis, while Section 4.3 presents the proposed approach. Section 4.4 discusses the experimental results that support the approach, and finally, Section 4.5 provides a summary of the outcomes of the study and suggests areas for further research.

4.2 Problem Statement

The AC-OPF problem is crucial for efficient and reliable power systems and markets. However, it is complex in many ways: economically, electrically and computationally. In this research, the computational problem of ACOPF will be considered which is a consequence of the lack of a fast and robust solution technique for the full AC-OPF. Usually, approximation techniques are employed to obtain reasonably acceptable solutions for the AC-OPF. While these techniques may reduce computational costs, they can also compromise the reliability of solutions, resulting in billions of dollars in unnecessary costs and environmental damage from emissions and energy waste [Cain et al., 2012].

Improving the solving of AC-OPF is crucial, as seen from Energy Information Administration (EIA) data on wholesale electricity prices and US and World energy production [EIA,]. Improvements can increase market efficiency by reducing costs by 5%, resulting in billions of dollars in savings per year. [ONeill et al., 2011] states that installing new software with an improved solution can cost less than 10 million dollars, leading to potential benefit/cost ratios in the range of 10 to 1000. Therefore, small increases in dispatch efficiency can have significant economic benefits. Thus, efficient and reliable AC-OPF solutions are crucial for the optimal operation of power systems and markets. Investing in improving the computational methods for solving the AC-OPF problem can offer significant economic and environmental benefits.

The integration of renewable energy sources (RES) into modern power grids provides significant advantages in terms of reducing energy production costs and carbon emissions, thereby promoting environmental protection. However, the rise in RES generation, such as wind and solar, poses significant challenges for Transmission System Operators (TSOs) in solving AC-OPF. The intermittent nature of RES generation is the most significant challenge as it poses a significant threat to grid security and can result in power outages.

The traditional approach to power grid operation and control, which relies on a fixed generation profile, is not feasible for grids with a substantial share of RES. To address this challenge, there is a need for a new class of optimization and control algorithms that account for generation uncertainty. These algorithms should be capable of handling the challenges that come with integrating intermittent RES generation into the grid.

This thesis addresses the stochastic AC-OPF problem, which is made more challenging due to the inclusion of uncertain generation. Various approaches have been proposed to tackle stochastic OPF problems, including robust optimization (RO), probabilistic OPF (P-OPF), and chance-constrained (CC) optimization. In this work, we focus on the Chance-Constrained Optimal Power Flow (CC-OPF) approach [Morillo et al., 2022, Wu et al., 2019], which aims to minimize power generation costs while providing optimal power dispatch in the face of changing demand and operational conditions of a power grid. The CC-OPF problem ensures that the operational constraints, such as nodal voltage limits and transmission capacity limits, are met with a high probability [Lubin et al., 2019, Roald and Andersson, 2017]. This provides a guarantee of grid resilience against sudden changes in RES generation. However, the chance-constrained (CC) OPF remains a challenging optimization problem both in terms of solution accuracy and computational complexity [Venzke et al., 2020].

Researchers have been actively exploring ways to solve the CC-OPF problem, which remains a challenging optimization problem. Previous studies, such as [Calafiore and Campi, 2006] and [Vrakopoulou et al., 2013a], have employed Monte Carlo (MC) scenario-based CC-OPF with nonlinear AC-PF equations. However, this approach is not scalable when the system size or the number of scenarios increases [Mezghani et al., 2020]. To address this issue, advanced sampling policies have been proposed, such as importance sampling [Mezghani et al., 2020, Lukashevich et al., 2021, Lukashevich and Maximov, 2021], and active sampling [Capitanescu et al., 2012, Owen et al., 2019]. These methods aim to reduce the number of required samples while maintaining

solution accuracy, making it possible to tackle larger and more complex CC-OPF problems. Unlike the scenario-based approach for solving CC-OPF, this study focuses on the analytical approach, which offers better computational performance for large-scale power systems by using certain distribution functions to model uncertainty. Although many papers have proposed convex/linear approximations to the chance-constrained problem [Lubin et al., 2015, Roald et al., 2016, Du et al., 2021], most of them are conservative and cannot handle large variations in power generation and/or demand. Another line of research employs the nonlinear Polynomial chaos expansion to model uncertainties in power generation and demand, but this approach often lacks good computational performance [Mühlpfordt et al., 2019, Mühlpfordt et al., 2016]. Additionally, most of the prior work relies on accurate knowledge of power system parameters, which may not be available due to insufficient and irregular equipment calibration. This challenge becomes even more critical in micro-grids and low-voltage grids where real-time metering capabilities are reduced, and parametric uncertainties are prevalent.

According to the aforementioned problems, this thesis is driven by the need to address the challenges posed by the computational complexity of solving stochastic CC-OPF problems without compromising the quality of optimal solutions. To this end, the research question is focused on how to strike a trade-off between the computational complexity and optimal solutions of the problem. The hypothesis put forward is that a data-driven approach, leveraging machine learning techniques, can enhance the solution of CC-OPF, leading to a trade-off between computational complexity and optimal solutions.

This chapter introduces a novel approach to solving the CC-OPF problem that balances computational complexity and solution accuracy. Instead of traditional approaches that rely on power balance equations, we propose using Gaussian process regression (GPR) to learn an approximation to power flow equations based on empirical samples of RES and load fluctuations. GPR is a model-free machine learning algorithm that can fit any smooth nonlinear function and offers a reasonable complexity-accuracy tradeoff [Dudley, 2010, Schulz et al., 2018, Liu et al., 2020]. Our approach differs from prior work, which has focused on deterministic OPF formulations using GPR [Pareek and Nguyen, 2020, Xu et al., 2020] or neural network-based power-flow models [Pan et al., 2020, Singh et al., 2021]. Such models do not consider CC-OPF with input uncertainty. Neural network-based stochastic AC-OPF has been proposed in [Gupta et al., 2021], but considers scenario-based CC-OPF. Instead, our GP-based formulation accounts for both parametric and input uncertainties, including analytic CC-OPF constraints. This approach exploits uncertainty propagation in Gaussian processes, resulting in a more accurate and efficient solution to the CC-OPF problem.

4.3 Proposed Approach

The problem given in 2.41 is not tractable to solve. There are two main reasons for this. Firstly, 2.41b is a semi-infinite problem, meaning that the set of fluctuations W is uncountable. Secondly, the chance constraints 2.41c - 2.41i are not tractable

for solving by optimization solver. Accordingly, we reformulated 2.41 as a nonlinear programming (NLP) problem with learned GPR in the constraints, such that an optimization solver can handle it.

Following [Bienstock et al., 2014], the objective function is reformulated as:

$$\mathbb{E}[c_i(p_{g,i}(\omega))] = \sum_{i \in \mathcal{G}} \left\{ c_{2,i}(p_{g,i}^2 + tr(\Sigma_\omega)\alpha_i^2) + c_{1,i}p_{g,i} + c_{0,i} \right\} \qquad (4.1)$$

where $\{c_{2,i}, c_{1,i}, c_{0,i}\}_{i=1}^{\mathcal{G}} \geq 0$ are scalar cost coefficients and α is the participation factor. Note that α (participation factor) is not a standard (fixed) value, but variables to be optimized with constraints $\sum_{i=1}^{\mathcal{G}} \alpha_i = 1$ and $\alpha_k \geq 0$. Accordingly, the objective function is a convex quadratic function of p_g and α.

Reformulating chance constraints to the tractable form implies reformulating both output and decision input constraints. According to [Hewing et al., 2019], a GPR individual output probabilities could be expressed in terms of the mean μ and variance σ^2:

$$\begin{cases} \texttt{Prob}(\mu_i + \sqrt{\sigma_i^2} \leq y_i^{out\max}) \leq 1 - \epsilon_i \\ \texttt{Prob}(\mu_i - \sqrt{\sigma_i^2} \geq y_i^{out\min}) \leq 1 - \epsilon_i \end{cases} \qquad (4.2)$$

that is equivalent to:

$$y_i^{out\min} + r_i\sqrt{\sigma_i^2} \leq \mu_i \leq y_i^{out\max} - r_i\sqrt{\sigma_i^2} \qquad (4.3)$$

where r_i represents the quantile function $\Phi^{-1}(1-\epsilon_i)$ of the standard normal distribution and $i = 1, ..., n_y$; n_y is number of outputs.

Similarly, the decision input constraints are [Bienstock et al., 2014]:

$$\begin{cases} \texttt{Prob}(p_{g,i} + \alpha_i \, tr(\Sigma_\omega) \leq p_{g,i}^{max}) \leq 1 - \epsilon_{p_{g,i}} \\ \texttt{Prob}(p_{g,i} - \alpha_i \, tr(\Sigma_\omega) \geq p_{g,i}^{min}) \leq 1 - \epsilon_{p_{g,i}} \end{cases} \qquad (4.4)$$

As in (4.2), this formulation is equivalent to:

$$p_{g,i}^{min} + r_{p_{g,i}}\alpha_i \, tr(\Sigma_\omega) \leq p_{g,i} \leq p_{g,i}^{max} - r_{p_{g,i}}\alpha_i \, tr(\Sigma_\omega) \qquad (4.5)$$

where $r_{p_{g,i}} = \Phi^{-1}(1 - \epsilon_{p_{g,i}})$ and $i \in \mathcal{G}$.

By taking into account of the formulations in (4.1), (4.3), and (4.5), the following

full GP CC-OPF is derived:

$$\min_{p_g, \alpha, \mu, \sigma} \sum_{i \in \mathcal{G}} \{c_{2,i}(p_{g,i}^2 + tr(\Sigma_\omega)\alpha_i^2) + c_{1,i}p_{g,i} + c_{0,i}\} \tag{4.6a}$$

$$\text{s.t.} \sum_{i \in \mathcal{G}} \alpha_i = 1, \ \alpha_i \geq 0 \tag{4.6b}$$

$$\sum_{i \in \mathcal{G}} p_{g,i} = \sum_{i \in \mathcal{L}} p_{l_i} - \sum_{j \in \mathcal{L}} p_{rs_j} \tag{4.6c}$$

$$\mu_i = \mu_i(x_*^{in}) \tag{4.6d}$$

$$\sigma_i^2 = \sigma_i^2(x_*^{in}) \tag{4.6e}$$

$$y_i^{out_{min}} + \lambda_i \leq \mu_i \leq y_i^{out_{max}} - \lambda_i \tag{4.6f}$$

$$p_{g,i}^{min} + \lambda_{p_{g,i}} \leq p_{g,i} \leq p_{g,i}^{max} - \lambda_{p_{g,i}} \tag{4.6g}$$

where $\lambda_i = r_i \sqrt{\sigma_i^2}$ for $i \in n_y$ and $\lambda_{p_{g,i}} = r_{p_{g,i}} \alpha_i \sqrt{tr(\Sigma_\omega)}$ are uncertainty margins. By $\sqrt{tr(\Sigma_\omega)}$ the standard deviation of the total active power imbalance Ω is represented. Equation (4.6c) describes the balance equation in which decision variables of controllable generators have to be satisfied in the power system keeping the balance between generation and consumption, while (4.6d) and (4.6e) represent GPR output mean and variance formulation. Thus, we replace the standard AC power flow equation with a data-driven method using these equations.

The proposed OPF optimization method using the GP-based model results in a non-convex optimization problem that can be challenging to solve. However, for the chosen twice differentiable SE kernel and zero prior means, the second-order derivative information of all quantities is available. Thus, the problem can be solved using sequential quadratic programming or the nonlinear interior-point method. We used a primal-dual interior-point algorithm with a filter line-search approach, and we have used the IPOPT [Wächter and Biegler, 2006] solver in the non-linear optimization framework CasADi [Andersson et al., 2019]. To solve this optimization problem, we assume simultaneous approaches that simultaneously treat p_g, α, μ, and σ^2 as optimization variables.

4.3.1 Full GP CC-OPF Pipeline

The proposed full GP CC-OPF problem is solved using a three-stage procedure. In the first stage, historical data or a simulator is used to collect fluctuations of power demand and RES power generation, along with dependent variables. In the second stage, Gaussian process regression is employed to provide a data-driven probabilistic approximation for the dependent variables in the power balance equations, which are much more tractable than the PF equations. This approximation simplifies the optimization problem significantly. A synthetic example illustrating this procedure is shown in Figure 4-1. In the third stage, the probability for constraint violations, referred to as uncertainty margins, is computed using Eq. 4.3 and 4.5. This probability has a straightforward expression for Gaussian process-based modeling with input

uncertainty and enables the CC-OPF problem to be transformed into a deterministic optimization problem. Solving this problem gives us a minimal cost of power dispatch that meets the probabilistic security constraints.
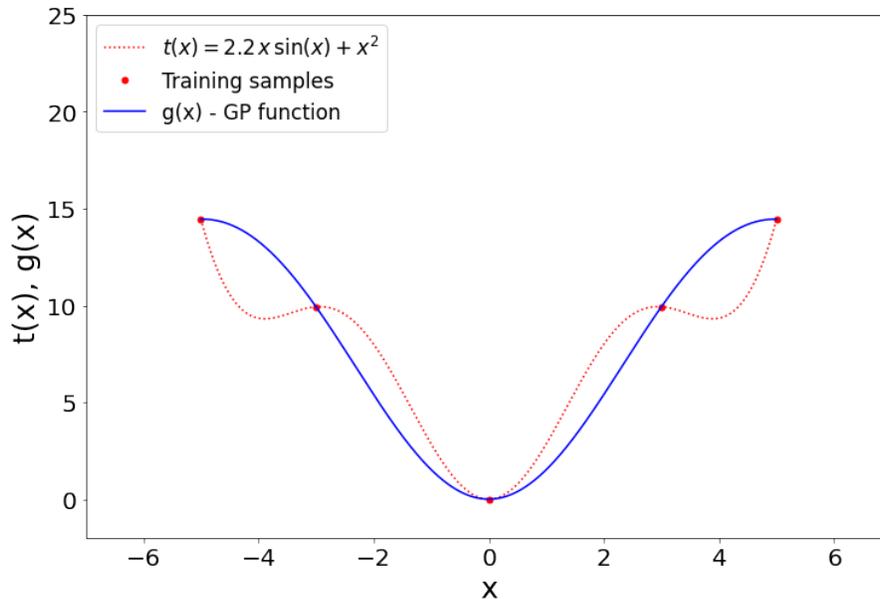

Figure 4-1: A synthetic example illustrates the power of the Gaussian process regression. Function $t(x) = 2.2x \sin x + x^2$ stands for the initial equation and $g(x)$ is its data-driven approximation via the Gaussian process with the squared exponential kernel. We refer to the approximation as a GP-function.

4.3.2 Input Covariance

As discussed in Section 3.3.5, we make an assumption that the unseen input points x_*^{in} for the GPR that follow a normal distribution with a vector mean denoted by $\mu_{x_*} \in \mathbb{R}^{n_x}$ and a covariance matrix denoted by $\Sigma_{x_*} \in \mathbb{R}^{n_x \times n_x}$. The mean values correspond to the optimized control variables, while the uncertain parameters are forecasted as:

$$\mu_{x_*^{in}} = [p_g^\top, p_l^\top, p_{rs}^\top]^\top$$

where $p_g = [p_g^1, p_g^2, ..., p_g^{n_\mathcal{G}}]^\top$, $p_l = [p_l^1, p_l^2, ..., p_l^{n_\mathcal{L}}]^\top$ and $p_{rs} = [p_{rs}^1, p_{rs}^2, ..., p_{rs}^{n_\mathcal{L}}]^\top$. Due to the linear feedback, the input covariance matrix Σ_{x_*} needs to be correctly defined as p_g is correlated with uncertainty ω in p_l and p_{rs}. This matrix is:

$$\Sigma_{x_*^{in}} = \begin{bmatrix} \Sigma_g & \Sigma_{g\omega} \\ \Sigma_{g\omega}^\top & \Sigma_\omega \end{bmatrix} \tag{4.7}$$

so that $\Sigma_g \in \mathbb{R}^{n_\mathcal{G} \times n_\mathcal{G}}$, $\Sigma_\omega \in \mathbb{R}^{n_\mathcal{L} \times n_\mathcal{L}}$ and $\Sigma_{g\omega} \in \mathbb{R}^{n_\mathcal{G} \times n_\mathcal{L}}$ are covariance sub-matrices.

The covariance sub-matrix Σ_ω represents uncertain variations in loads and renew-

able sources. We consider this matrix as diagonal:

$$\Sigma_\omega = \begin{bmatrix} \sigma_{l_1}^2 & ... & 0 & 0 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & ... & \sigma_{l_\mathcal{L}}^2 & 0 & ... & 0 \\ 0 & ... & 0 & \sigma_{rs_1}^2 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & ... & 0 & 0 & ... & \sigma_{rs_\mathcal{L}}^2 \end{bmatrix} \tag{4.8}$$

and assume that fluctuations between stochastic power injections are independent, consistent with prior work [Bienstock et al., 2014].

Sub-matrix Σ_g refers to controllable generator deviations caused by loads and renewable sources fluctuations keeping the balance of the system. Since controllable variables p_g are dependent on forecasted uncertain injections p_l and p_{rs}, the deviations of p_g are also dependent on fluctuations of ω as follows:

$$\Sigma_g = \begin{bmatrix} \alpha_1^2 tr(\Sigma_w) & ... & \alpha_1 tr(\Sigma_w)\alpha_{n_\mathcal{G}} \\ ... & ... & ... \\ \alpha_{n_\mathcal{G}} tr(\Sigma_\omega)\alpha_1 & ... & \alpha_{n_\mathcal{G}}^2 tr(\Sigma_\omega) \end{bmatrix} \tag{4.9}$$

The covariance elements between dependent decision variables and stochastic parameters are expressed in sub-matrix $\Sigma_{g\omega}$ as:

$$\Sigma_{g\omega} = \begin{bmatrix} \alpha_1 \sigma_{l_1}^2 & ... & \alpha_1 \sigma_{l_{n_\mathcal{L}}}^2 & \alpha_1 \sigma_{rs_1}^2 & ... & \alpha_1 \sigma_{rs_{n_\mathcal{L}}}^2 \\ ... & ... & ... & ... & ... & ... \\ \alpha_{n_\mathcal{G}} \sigma_{l_1}^2 & ... & \alpha_{n_\mathcal{G}} \sigma_{l_{n_\mathcal{L}}}^2 & \alpha_{n_\mathcal{G}} \sigma_{rs_1}^2 & ... & \alpha_{n_\mathcal{G}} \sigma_{rs_{n_\mathcal{L}}}^2 \end{bmatrix} \tag{4.10}$$

4.4 Experimental Setup and Results

The results presented in this chapter are obtained on an Intel Core i7-5500U CPU @ 2.40GHz and 8GB of RAM. To simulate the experiment, we use our python gp-ccopf[1] framework developed for this article and pandapower [Thurner et al., 2018] package to validate results.

4.4.1 Case Study

We use the 9, 39, and 118-bus IEEE Test Systems to evaluate the performance and scalability of the proposed GP CC-OPF method. Simulations are conducted using pandapower [Thurner et al., 2018] package for Python.

IEEE 9-bus Test Systems consists of 3 generators and 3 loads on the high-voltage level of $V_n = 345kV$. We introduce two renewable source generators at buses 3 and 5 with a total forecast power output of 80 MW, which is approximately 25.4% of the

---

[1]https://github.com/mile888/gp_cc-opf

total active power demand. We assume that the reactive to real power ratio is 0.3. All loads and renewable sources are uncertain.

IEEE 39-bus Test Systems consists of 10 generators and 21 loads with $V_n = 345kV$. Renewable source generators are placed at buses 1, 11, 14, 21, 23, and 28 with a total forecast power output of 1260 MW, which is approximately 20.2% of the total active power demand. The assumed power ratio is 0.3. Also, we consider that all loads and renewable sources are assumed to be uncertain.

IEEE 118-bus Test Systems consists of 54 generators and 99 loads with $V_n = 345kV$. The six renewable source generators are placed at buses 10, 27, 47, 51, 78, and 92 with a total forecast power output of 1260 MW, which is approximately 29.7% of the total active power demand. All loads and RES are considered as uncertain elements with an assumed reactive to real power ratio of 0.3.

The forecast errors are modeled as zero mean, multivariate Gaussian random variables with a standard deviation corresponding to 15% of the forecasted loads ($\sigma_l = 0.15\, p_l$) and 30% of the forecasted renewable sources ($\sigma_{rs} = 0.3\, p_{rs}$). The acceptable violation probabilities ϵ are set to $\epsilon_{p_g} = 0.1\%$ and $\epsilon = 2.5\%$. Since generation constraints are physically impossible to violate, we assume a very small percentage of violations (ϵ_{p_g}), as proposed in Swissgrid (the swiss system operator) [Abbaspourtorbati and Zima, 2015] for their reserve procurement process. In contrast, output constraints are soft constraints where some violations can be tolerated (ϵ) if the magnitude and duration are not too large or removed through additional control actions such as generation re-dispatch [Roald and Andersson, 2017].

4.4.2 Evaluation Procedure and Metrics

To evaluate the prediction quality of the GPR optimized using training data, we use the root mean squared error (RMSE) metric for each output k as:

$$RMSE_k = \sqrt{\frac{1}{m_*}\sum_{i=1}^{m_*}(y_{i,k}^{out} - \hat{y}_{i,k}^{out})^2}$$

where y^{out} and $\hat{y}^{out}$ are the actual (AC-PF) and predicted (GPR) mean values for testing data, and m_* is the number of test samples. The average evaluation score of all output variables is calculated as follows:

$$RMSE = \frac{\sum_{k=1}^{n_y} RMSE_k}{n_y} \tag{4.11}$$

Solving the data-driven CC-OPF with GP-based constraints for different test data, we compute the mean of output variables $\hat{y}^{out}$ (Eq. (4.6d)) and report the RMSE against the variables computed by AC-PF. To analyze chance constraints feasibility, we compare the distribution of output variables, computed analytically in GP CC-OPF, with the true empirical spread computed via AC-PF with Monte-Carlo samples of uncertainty ω around the mean inputs (loads, RES, and optimized generator

set-points).

To measure the performance of the solution, we also compute the computation time, generation cost (Eq. (4.6a)), and the empirical spread of output variables for (a) model-based scenario CC-OPF [Vrakopoulou et al., 2013a], [Mezghani et al., 2020], (b) AC-OPF on individual test samples (i.e., full recourse), and (c) AC-OPF for base case (mean loads and RES). Note that AC-OPF on the base case does not optimize for input uncertainty and presents the worst-case violations. On the other hand, the full-recourse solves AC-OPF for each sample separately and is the benchmark for feasibility as it is not restricted to linear generator feedback.

4.4.3   Models Performance

Proposed GPR models for IEEE 9, 39, and 118-bus systems are trained on randomly sampled data, generated as discussed in Section 2.4. The GPR model of the IEEE 9-bus system consists of 8 inputs features and 15 outputs, while the IEEE 39 bus has 37 inputs and 74 outputs. Likewise, the GPR model of the IEEE 118-bus system has 159 inputs and 291 outputs. In the case of the IEEE 9 bus case, we applied 75 random samples for training and 25 for validation. For the IEEE 39 bus system, training is conducted on 200 samples, and 65 samples are used for validation. Similarly, for the IEEE 118 system, 400 training samples and 100 validation samples are used. The average RMSE metric in Eq. (4.11) for IEEE 9, IEEE 39, and IEEE 118 bus are RMSE $= 7.72 \cdot 10^{-5}$ $p.u.$, RMSE $= 8.91 \cdot 10^{-4}$ $p.u.$, and RMSE $= 2.83 \cdot 10^{-2}$ $p.u.$, respectively. The small errors suggest that the models for both systems have good generalization for deterministic inputs and that the hyperparameters are well optimized.

Next, we use the GP CC-OPF method with three approximation methods: first-order Taylor approximation (TA1), second-order Taylor approximation (TA2), and Exact moment matching (EM). The iterative solution interior-point algorithm is applied to find a feasible solution in k iterations. In our simulations, the algorithm converged within a relatively small number of iterations for the convergence tolerance of $\epsilon_{tol} = 10^{-5}$, as shown in Table 4.1. Table 4.1 also presents each approximation method's RMSE for GP-output against AC-PF based outputs, for the same inputs. The RMSE results are the same for TA1 and TA2 algorithms since both algorithms have the same mean function (3.23). In contrast, the EM mean function depends on input covariance; in our case, the RMSE is better. However, the CPU time of optimization is greater for EM, with a sharper increase for the larger system.

4.4.4   Chance-constraint feasibility and uncertainty margins

We plot the spread of centered CC-OPF output variables (voltages at PQ buses, reactive powers at controllable generators, and apparent power flows on the lines) for the IEEE 9-bus system in Fig. 4-2. The three standard deviations (3 STD) intervals for uncertainty margins are analytically derived using the estimated GP-variance, see Eq. (4.6e). The different approximations considered for uncertainty propagation are marked as blue (TA1), red (TA2), and black (EM). The empirical spread for the same output variables is derived using both GP-function and AC-PF equation

System	Parameters	TA1	TA2	EM
IEEE 9	RMSE [p.u.]	$6.35 \cdot 10^{-3}$	$6.35 \cdot 10^{-3}$	$5.95 \cdot 10^{-3}$
	No. iteration	16	14	16
	CPU time [s]	0.35	4.97	6.60
IEEE 39	RMSE [p.u.]	$1.37 \cdot 10^{-2}$	$1.37 \cdot 10^{-2}$	$1.35 \cdot 10^{-2}$
	No. iteration	18	18	20
	CPU time [s]	19.38	717.6	1221.2
IEEE 118	RMSE [p.u.]	$8.37 \cdot 10^{-1}$	$8.37 \cdot 10^{-1}$	$8.21 \cdot 10^{-1}$
	No. iteration	20	21	21
	CPU time [s]	212.1	2987.1	4102.3

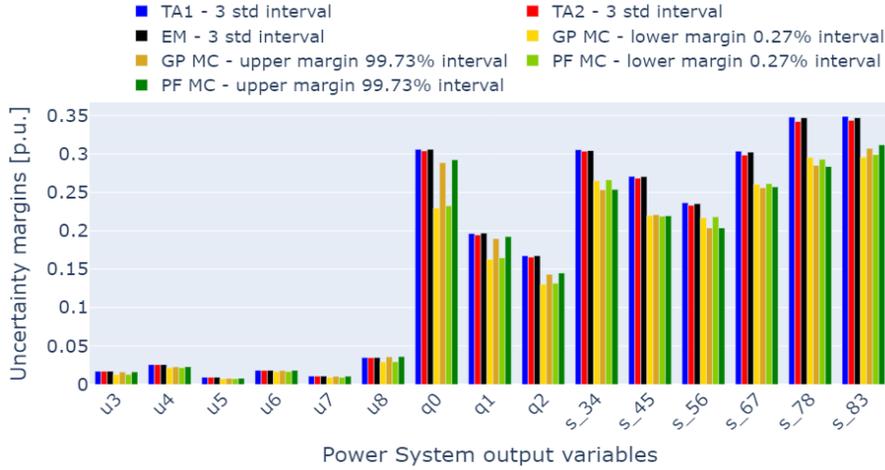Table 4.1: GP CC-OPF method results for IEEE 9, IEEE 39 and IEEE 118 bus systems.
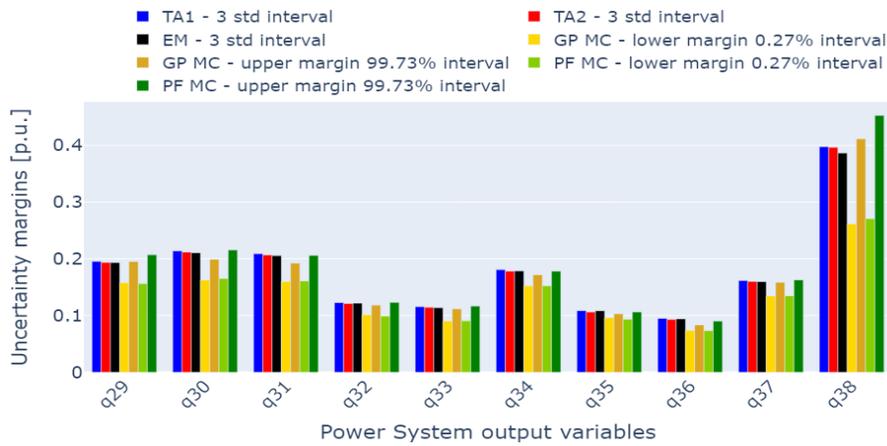

Figure 4-2: Output uncertainty spread of IEEE 9.

with 5000 Monte Carlo (MC) samples of the input uncertainty ω. Brown and yellow bars represent the upper and lower margins of the MC-based GPR model, while dark and light green bars are for the upper and lower margins of the MC-based AC-PF equation. Due to the asymmetrical MC-based distributions owing to nonlinear AC-PF, the upper and lower uncertainty margins computed as 99.73% and 0.27% probability, respectively (3 STD), are depicted separately.

Similar plots for nodal voltage, reactive power injections, and apparent power flows on lines of the 39 and 118 bus systems are presented in Fig. 4-3 and Fig. 4-4. Note that part of the variables is shown in Fig. 4-4 due to the impossibility of presenting all of them in the figure. Comparing the analytic results with the Monte Carlo based uncertainty margins, it is clear that they are close for most outputs, with a few exceptions where the analytic spread *overestimate* the MC ones.
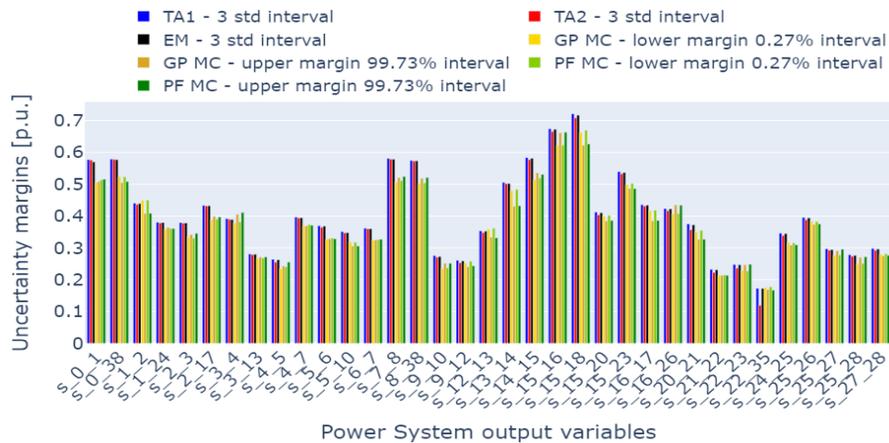
Note that as analytic margins satisfy the chance constraints, see Eq. (4.6f) and Eq. (4.6g), the smaller MC empirical violations will also be within the chance-
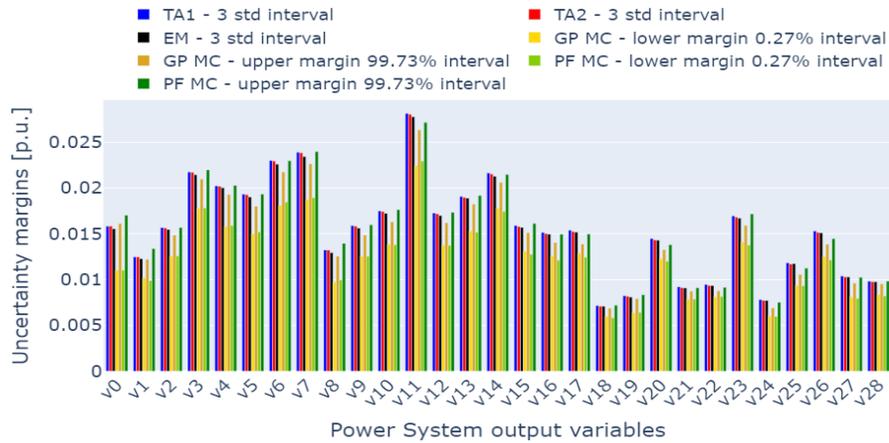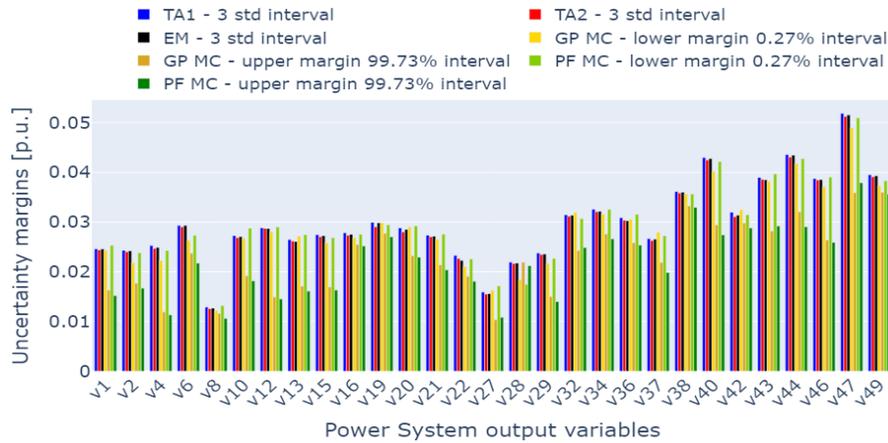
(a) Voltage output distributions

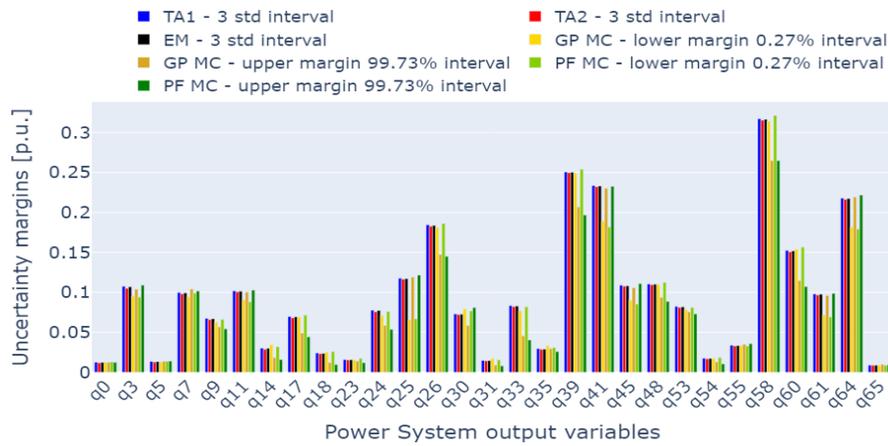

(b) Reactive power output distributions

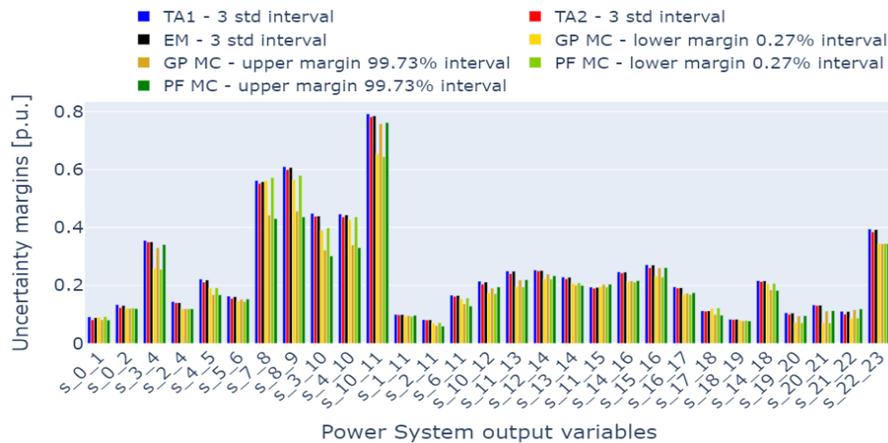(c) Apparent power output distributions

Figure 4-3: Output uncertainty spread of IEEE 39.

(a) Voltage output distributions

(b) Reactive power output distributions

(c) Apparent power output distributions

Figure 4-4: Output uncertainty spread of IEEE 118.

constrained limits, and hence feasible. Moreover, from Fig. 4-2, Fig. 4-3 and Fig. 4-4, MC margins for the GPR model and AC-PF equation are comparable, which implies that the GPR model can learn the nonlinear AC-PF accurately. For the IEEE 39 and 118 bus systems, we do observe small deviations between upper and/or lower margins for some variables, but this can be overcome by increasing the number of training data samples.

4.4.5 Comparison with Model-based Scenario CC-OPF

Here, we compare the cost, computational time, and CC-feasibility of the GP CC-OPF method (with TA1 approximation), with model-based scenario CC-OPF [Vrakopoulou et al., 2013a, Mezghani et al., 2020], with a varying number of scenarios. The empirical constraint violation for determined set-points is computed using 1000 test MC samples. We also compare our method with two model-based setups:

(A) AC-OPF will full recourse, i.e., separate AC-OPF for each MC sample,

(B) and AC-OPF for the base case, i.e., AC-OPF for the mean input.

Note that though unrealistic, (A) considers generalized nonlinear feedback, while (B) doesn't optimize the participation factor α and uses equal generator participation for test MC samples. Hence they represent, respectively, model-based solutions with the best and worst feasible solutions.

In Table 4.2, we can observe that GP CC-OPF solution produces the desired reliability level of $\epsilon_y = 2.5\%$, with a cost between that of (A) and (B) for both test cases. Note that (A) solves separate AC-OPF for each MC sample and hence has no constraint violation but a higher cost. On the other hand, (B) 's set point is optimized by OPF for the mean loads and RES. Thus, it has a lower cost but large violations for test uncertainty. Their computational time is that of a single AC-OPF. For scenario CC-OPF, increasing the number of scenarios increases cost and reduces constraint violations for either test case. Although the number of scenarios necessary to get desired CC-feasibility is greater than 100 for IEEE 39 and greater than what leads to a significantly higher computational time for scenario CC-OPF over GP CC-OPF where CC is implemented analytically.

To study the empirical spread of output variables, we illustrate the voltage magnitude under uncertainty at bus 8 of IEEE9 in Fig. 4-5, the voltage magnitude at bus 25 of IEEE39 in Fig. 4-6, and the same variable at bus 47 of IEEE118 in Fig. 4-7. As one can observe, voltage values for GP CC-OPF for the IEEE9 (Fig. 4-5a) are centered further from the minimal voltage than the corresponding voltage values for 100 scenarios CC-OPF (Fig. 4-5b). The latter leads to reduced violation of GP CC-OPF compared to scenario CC-OPF for the important lower voltage limit. Moreover, the latter's spread is more skewed to the constraint border. Note that (A) with full recourse has no violation while (B) for base-case has a higher violation. Similarly, for voltage at bus 25 of IEEE39 and bus 47 of IEEE118, the GP CC-OPF voltage spread (Fig. 4-6a and 4-7a) is further from the important upper limit compared to the solution of scenario CC-OPF (Fig. 4-6b and 4-7b).

System	IEEE 9		
Parameters	Cost [\$]	Infeas. Prob. [%]	Time (s)
A (full recourse)	$4.056 \cdot 10^3$	-	0.86
B (base-case)	$3.467 \cdot 10^3$	9.76	0.86
GP CC-OPF	$4.039 \cdot 10^3$	0.44	0.35
20 CC-OPF	$3.481 \cdot 10^3$	7.52	4.4
50 CC-OPF	$3.836 \cdot 10^3$	4.92	12.1
100 CC-OPF	$3.985 \cdot 10^3$	1.24	22.8

System	IEEE 39		
Parameters	Cost [\$]	Infeas. Prob. [%]	Time (s)
A (full recourse)	$7.869 \cdot 10^6$	-	0.93
B (base case)	$7.533 \cdot 10^6$	35.36	0.93
GP CC-OPF	$7.752 \cdot 10^6$	2.36	19.38
50 CC-OPF	$7.642 \cdot 10^6$	14.20	96.4
100 CC-OPF	$7.696 \cdot 10^6$	8.24	184.7
200 CC-OPF	$7.813 \cdot 10^6$	0.16	505.0

System	IEEE 118		
Parameters	Cost [\$]	Infeas. Prob. [%]	Time (s)
A (full recourse)	$26.91 \cdot 10^6$	-	1.59
B (base case)	$26.39 \cdot 10^6$	49.97	1.59
GP CC-OPF	$26.91 \cdot 10^6$	2.47	212.1
100 CC-OPF	$26.52 \cdot 10^6$	28.12	683.3
200 CC-OPF	$26.71 \cdot 10^6$	19.62	1856.5
500 CC-OPF	$26.90 \cdot 10^6$	4.78	39537.1

Table 4.2: Cost function values and probability of violation of a constraint at a solution
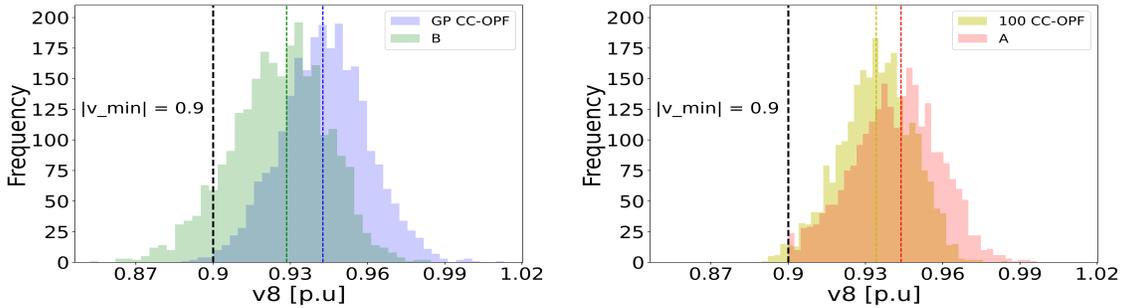
(a) Spread of voltages for GP CC-OPF and AC-OPF for base case.

(b) Spread of voltages for scenario CC-OPF and (A) AC-OPF with full recourse.

Figure 4-5: Case IEEE-9. Empirical Spread of voltages at Bus 8 under uncertainty.

(a) Spread of voltages for GP CC-OPF and (B) AC-OPF for the base case.

(b) The spread of voltages for scenario CC-OPF and (A) AC-OPF with full recourse.

Figure 4-6: Case IEEE-39. Empirical Spread of voltages at Bus 25 under uncertainty.

(a) Spread of voltages for GP CC-OPF and (B) AC-OPF for the base case.

(b) The spread of voltages for scenario CC-OPF and (A) AC-OPF with full recourse.

Figure 4-7: Case IEEE-118. Empirical Spread of voltages at Bus 47 under uncertainty.

4.4.6 Discussion
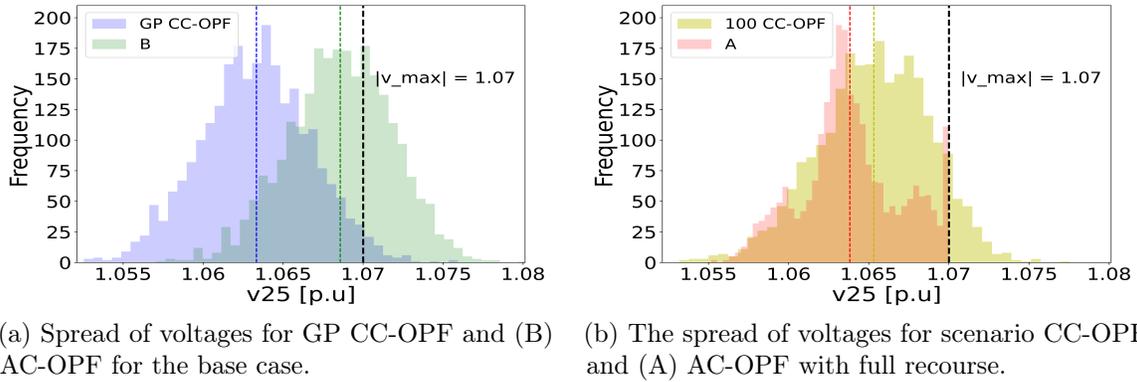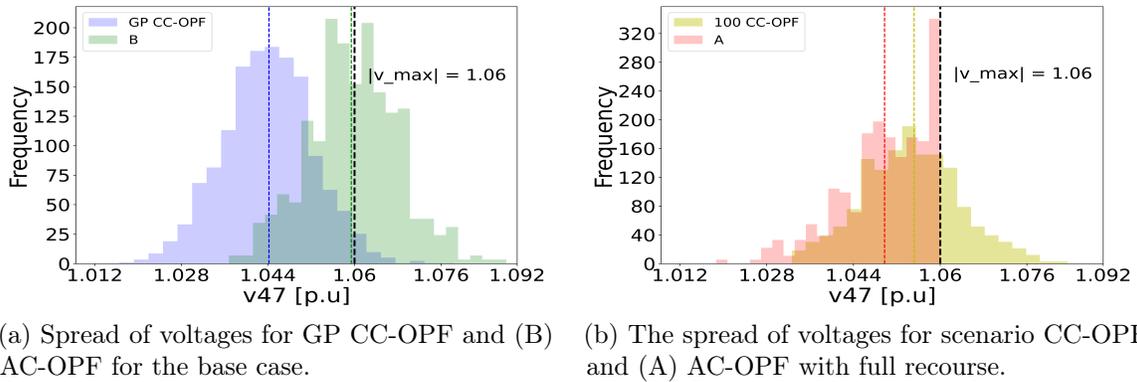
Comparing the results of different approximation methods within GP CC-OPF, we can conclude that all three methods (TA1, TA2, and EM) approximate uncertainty margins very well with small probability violations for MC test samples (Figs. 4-2, 4-3). Table 4.1 shows that with the system scaling, the accuracy (RMSE) of the mean performance decreases. It can be explained by the fact that for the IEEE 39 and IEEE 118 bus systems, we considered 200 and 400 training data samples randomly sampled in space. More samples and active learning techniques can help improving the accuracy of large systems. The presented new approach for solving the CC-OPF problem could be equally efficiently applicable for power networks of essentially larger size than the test networks used for the algorithm testing. In general, the problem of GPR is scalability, where the computational complexity of training and prediction increases with the number of samples and input/output dimensions. Due to the inclusion of the second derivative term, TA2 and EM are more complicated than TA1 as confirmed by the CPU time results in Table 4.1. Computational complexity problems can be overcome by using sparse GPR and dimensionality reduction techniques [Hewing et al., 2019], which we plan to explore in the future. Generally, from the obtained results

81

that the TA1 method gives similar performance accuracy as TA2 and EM, but with less complexity.

4.5   Conclusion

In this chapter, we have introduced a novel stochastic OPF method that is model-free and data-driven. The key idea is to replace the standard AC power flow equation and security constraints with a Gaussian process regression model. We have shown that even with a relatively small number of samples, the GP function can accurately model the AC-PF equations and efficiently propagate uncertainty from input to output.

In our experiments, we have demonstrated that applying the first Taylor Approximation method in GP CC-OPF outperforms other proposed approximation methods and two widely used sample-based approaches. Despite the notable accuracy of the data-driven CC-OPF, it is limited by the propagation space of the training data. To address this issue, we propose a new hybrid GP CC-OPF method in the next **Chapter** 5 that combines linear and non-linear approximation methods. Additionally, we propose improving the scalability of the proposed approach using sparse GPR formulations.

Chapter 5

Data Driven CC-OPF using Hybrid Sparse Gaussian Process

5.1 Introduction

In this chapter, we introduce an improved version of the data-driven stochastic OPF method proposed in **Chapter 4**. This improved approach is a hybrid method that combines the linear DC-PF balance equation with the GPR-based estimation of the residuals between DC and AC power flow equations. By incorporating the DC power flow equation, our method is more robust and less sensitive to the propagation space of the training data. To further improve the scalability of the method, we use sparse techniques for the GPR kernel matrix. For the input uncertainty propagation, the first-order Taylor Approximation (TA1) is applied. As in **Chapter 4**, the proposed approach has been evaluated on IEEE9, 39 and 118 bus systems. The approach and results of this chapter have been presented and published at the *IEEE Belgrade PowerTech 2023* conference [1].

The remaining parts of this chapter are organized as follows: In Section 5.2, we explain the reasons that motivated us to develop a hybrid approach for GP CC-OPF. Section 5.3 presents the mathematical formulation of the proposed hybrid approach. The experimental results that demonstrate the effectiveness of the proposed method are presented in Section 5.4. Finally, in Section 5.6, we summarize the key findings of this study.

5.2 Motivation

In **Chapter 4**, we introduce a novel data-driven GP CC-OPF approach that is based on the full AC-PF equations. However, this approach may not always be robust and its effectiveness heavily depends on the distribution and spread of training data in the space. If the training data do not cover the space properly, the GPR model may tend towards the prior mean, which is zero, resulting in an inadequate approximation of the AC-PF function. To illustrate this drawback, we provide a toy example in Fig. 5-1

The disadvantage of the full approach is that GP is learned from historical data on
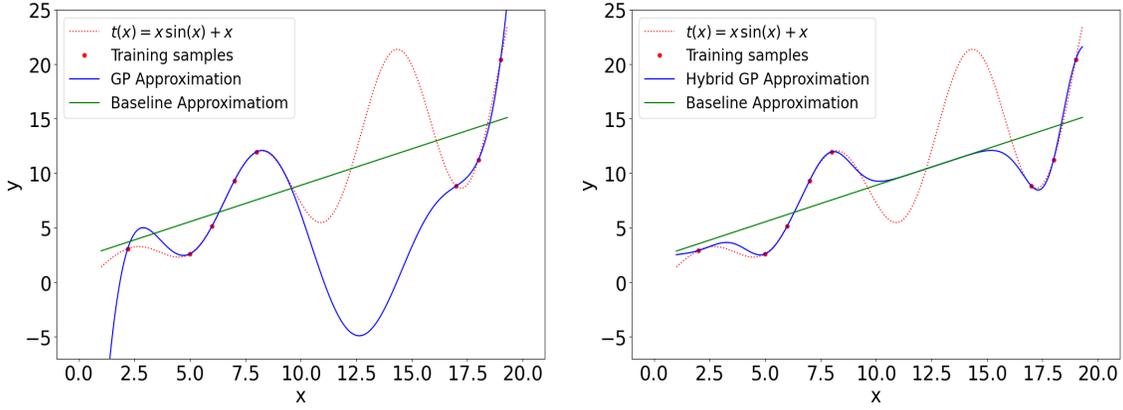
(a) Full approach: GP learned on historical data; (b) Hybrid approach: GP learned on residuals between the linear model and data.

Figure 5-1: Non-linear function $x\sin(x) + x$ (in red) is approximated with GP (blue; left) based on several observations (red dots). The hybrid GP approximation (blue; right) uses a linear proxy (green) to improve the approximation.

a full non-linear AC model using the entire synthetic data space. This means that this model will make a significant prediction error if the historical data is poorly distributed in space due to the technical nature of the system. A simple one-dimensional example can be observed in Fig. 5-1a. One can see the difference in regression results between $x = 10$ and $x = 17.5$ for this method. A simple GPR will tend to predict zero prior in this region since the absence of training data. To overcome this problem and build a more robust approach, we propose a hybrid chance constraints GP-based structure. Whereas hybrid GPR has non-zero prior in this region – linear approximation – and regression result is more adequate.

The advantage of the hybrid approach can be demonstrated with a simple one-dimensional example in Fig. 5-1b. Consider a non-linear function $t(x) = x sin(x) + x$ (red dashed line) and a linearized function $h(x) = ax + b$ (green solid line - baseline approximation), where coefficients (a, b) are fitted according to the training data corresponding to the non-linear toy function t. Fig. 5-1a presents the dependence learned by GPR (blue solid line) from the full non-linear function, while in Fig. 5-1b) the hybrid approach of the summation of the linearized function and the additive GP function learned on the residuals between $t(x)$ and $h(x)$ is shown. It can be seen from Fig. 5-1a that when training data does not cover the function domain well enough, the GP function will tend toward the prior mean equal to zero. In the case of a hybrid approach, the function will tend to the values of the linearized function. Therefore, the worst-case output results will not be equal to zero, but to the linearized model outputs. Thus, a more robust model is proposed, in which the efficiency will not be so strongly violated in the case that the training set is not fully presented. In the case of the hybrid GP chance-constrained AC-OPF, the worst-case output results will be equal to those obtained with the approximated DC CC-OPF.

The hybrid framework combines a linear DC approximated physical model and

84

an additive GP-part learned on the residuals between a non-linear AC and a linear DC power flow model. Thus, the hybrid approach to the AC power flow equation is modeled as follows:

$$y^{out} = z^{out}(x_i^{in}) + (g^{out}(x_i^{in}) + \omega_i), \qquad i = 1, ..., m_s \tag{5.1}$$

where z^{out} is a known linear part and g^{out} is an additive GP-term that describes unknown non-linearities of the system that need to be learned from historical data. According to 5.1, a residual output r_i^{out} is equal to:

$$r_i^{out} = y^{out} - z^{out}(x_i^{in}) = g^{out}(x_i^{in}) + \omega_i, \qquad i = 1, ..., m_s \tag{5.2}$$

Even though the hybrid approach is a more robust method, it may still face computational complexity issues due to the GP kernel matrix. To address this challenge, we introduce a sparse GP approximation to enhance computational efficiency while maintaining the robustness and accuracy of the method.

5.3 Hybrid approach

As previously discussed in Section 4.3, utilizing Gaussian processes to model power flow equations with uncertainty offers a more tractable solution. Unlike the full approach, our proposed hybrid approach replaces the power flow balance equation with a combination of a linear approximation and an additive GPR model learned on the residuals:

$$\mu = z^{out}(\mu_{x_*^{in}}) + \mu(x_*^{in}) \tag{5.3a}$$
$$\sigma^2 = diag(\Sigma_{z^{out}}(\Sigma_{x_*^{in}})) + diag(\Sigma(x_*^{in})) \tag{5.3b}$$

where $z^{out}(\mu_{x_*^{in}}) = [v_{dc}, A\mu_{x_*^{in}}^T + b]$; $v_{dc} = 1$ $p.u.$ is constant DC output voltage, A is linear coefficient matrix and b is intercept vector; $diag(\Sigma_{z^{out}}(\Sigma_{x_*^{in}}))$ and $diag(\Sigma(x_*^{in}))$ are column vectors of diagonal matrix elements where

$$\Sigma_{z^{out}} = \begin{bmatrix} 0 & 0 \\ 0 & A\Sigma_{x_*^{in}}A^T \end{bmatrix}$$

Applying the GP-balance equation 5.3, an analytical reformulation of chance-constraints with uncertainty margins Eq. 4.1-4.5 derived in Section 4.3, the hybrid

GP CC-OPF problem is:

$$\min_{p_g, \alpha, \mu, \sigma} \sum_{i \in \mathcal{G}} \{ c_{2,i}(p_{g,i}^2 + tr(\Sigma_\omega)\alpha_i^2) + c_{1,i}p_{g,i} + c_{0,i} \} \tag{5.4a}$$

$$\text{s.t.} \sum_{i \in \mathcal{G}} \alpha_i = 1, \ \alpha_i \geq 0 \tag{5.4b}$$

$$\sum_{i \in \mathcal{G}} p_{g,i} = \sum_{i \in \mathcal{L}} p_{l_i} - \sum_{j \in \mathcal{L}} p_{rs_j} \tag{5.4c}$$

$$\mu_i = z^{out}(\mu_{x_*^{in}}) + \mu(x_*^{in}) \tag{5.4d}$$

$$\sigma_i^2 = diag(\Sigma_{z^{out}}(\Sigma_{x_*^{in}})) + diag(\Sigma(x_*^{in})) \tag{5.4e}$$

$$y_i^{out_{min}} + \lambda_i \leq \mu_i \leq y_z^{out_{max}} - \lambda_i \tag{5.4f}$$

$$p_{g,i}^{min} + \lambda_{p_{g,i}} \leq p_{g,i} \leq p_{g,i}^{max} - \lambda_{p_{g,i}} \tag{5.4g}$$

where $\lambda_i = r_i \sqrt{\sigma_i^2}$ for $i \in n_y$ and $\lambda_{p_{g,i}} = r_{p_{g,i}}\alpha_i \sqrt{tr(\Sigma_\omega)}$ are uncertainty margins.

Similar to the full approach, the non-linear optimization framework CasADi is used to solve the hybrid GP CC-OPF formulation in Eq. 5.4 with the available IPOPT solver.

5.3.1 Input Approximation as Normal Distributions

In the GPR model, for the given deterministic inputs, the predicted outputs have a normal (Gaussian) distribution described with mean and variance (Eq. 3.20). Since we consider a stochastic problem, the inputs are also distributed. Evaluating the outputs of a GPR from an input distribution is generally intractable and the resulting outputs' distribution is not normal. For certain assumptions of the input distribution, some strict over-approximations of the resulting distributions exist [Koller et al., 2018], but they are computationally demanding and very conservative. Accordingly, we consider computationally cheap and a practical approximation where the control and uncertain inputs are approximated as jointly Gaussian distributed:

$$x_*^{in} \sim \mathcal{N}(\mu_{x_*^{in}}, \Sigma_{x_*^{in}}) = \mathcal{N}\left(\begin{bmatrix} p_g \\ p_\omega \end{bmatrix}, \begin{bmatrix} \Sigma_g & \Sigma_{g\omega} \\ \Sigma_{g\omega}^T & \Sigma_\omega \end{bmatrix} \right) \tag{5.5}$$

where p_g, $p_\omega = [p_l^T, p_{rs}^T]^T$ are vectors of optimized control variables and forecasted operating points; $\Sigma_g \in \mathbb{R}^{n_\mathcal{G} \times n_\mathcal{G}}$, $\Sigma_\omega \in \mathbb{R}^{n_\mathcal{L} \times n_\mathcal{L}}$ and $\Sigma_{g\omega} \in \mathbb{R}^{n_\mathcal{G} \times n_\mathcal{L}}$ are input covariance sub-matrices.

The covariance matrices Σ_ω, Σ_g and $\Sigma_{g\omega}$ are equal to the equations 4.8, 4.9, and 4.10, respectively.

5.4   Experimental Setup and Results

The simulations in this chapter are performed on an Intel Core i7-5500U CPU @ 2.40GHz and 8GB of RAM. The python *hybrid-gp*[1] framework was developed to simulate the investigation and *pandapower* [Thurner et al., 2018] package was used to validate the results.

5.4.1   Case Study

The performance and scalability of the proposed hybrid GP CC-OPF approach are evaluated on 9, 39 and 118 bus IEEE test systems. All systems are provided with `pandapower` [Thurner et al., 2018].

IEEE 9-bus Test Systems consists of 3 generators and 3 loads on the high-voltage level of $V_n = 345kV$. We introduce two renewable source generators at buses 3 and 5 with a total forecast power output of 80 MW, which is approximately 25.4% of the total active power demand. We assume that the reactive to real power ratio is 0.3. All loads and renewable sources are uncertain.

IEEE 39-bus Test Systems consists of 10 generators and 21 loads with $V_n = 345kV$. Renewable source generators are placed at buses 1, 11, 14, 21, 23, and 28 with a total forecast power output of 1260 MW, which is approximately 20.2% of the total active power demand. The assumed power ratio is 0.3. Also, we consider that all loads and renewable sources are assumed to be uncertain.

IEEE 118-bus Test Systems consists of 54 generators and 99 loads with $V_n = 345kV$. The six renewable source generators are placed at buses 10, 27, 47, 51, 78, and 92 with a total forecast power output of 1260 MW, which is approximately 29.7% of the total active power demand. All loads and RES are considered as uncertain elements with an assumed reactive to real power ratio of 0.3.

The forecast errors are modeled as zero mean, multivariate Gaussian random variables with a standard deviation corresponding to 15% of the forecasted loads ($\sigma_l = 0.15\,p_l$) and 30% of the forecasted renewable sources ($\sigma_{rs} = 0.3\,p_{rs}$). The acceptable violation probabilities ϵ are set to $\epsilon_{p_g} = 0.1\%$ and $\epsilon = 2.5\%$. We assume a very small percentage of violations (ϵ_{p_g}) according to [Abbaspourtorbati and Zima, 2015], since generation constraints are physically impossible to violate. However, output constraints are soft constraints where some violations can be tolerated (ϵ_y) if the magnitude and duration are not too large, or removed through additional control actions such as generation re-dispatch [Roald and Andersson, 2017].

5.4.2   Model Performance

The hybrid GPR model is trained on residuals between datasets collected from full AC-OPF and DC-OPF approximation. The synthetic data of the IEEE 9-bus system consists of 8 inputs features and 15 outputs. The IEEE39 model has 37 inputs and 74 outputs, while the GPR model of the IEEE 118-bus system has 159 inputs and

---

[1] https://github.com/mile888/hybrid_gp

291 outputs. We generated 75, 200 and 400 random training samples for IEEE 9, 39 and 118-bus systems and considered a different number of inducing points of 10, 30, and 50 for both full and hybrid GPR models. We use IPOPT solver to solve both hybrid and full GP CC-OPF and find a feasible solution with a convergence tolerance of $\epsilon_{tol} = 10^{-5}$.

The results in Table 5.1 indicate that the hybrid GPR model with TA1 outperforms both RMSE and CPU time with respect to the full GPR model with TA1 in **Chapter** 4. Similarly, sparse approximation allows to significant reduce the time complexity for both hybrid and full approaches. The power of the sparse approximation is presented for all cases. Decreasing the number of inducing points reduces the computation time in all test cases (IEEE9, IEEE39 and IEEE118) and approaches (full and hybrid), but increases RMSE. Accordingly, we can find a trade-off between the computation time and the accuracy of the faster and more robust hybrid GP approach. To determine the level of error of the mean output values in prediction, the RMSE of the output data from the analytical DC approximated CC-OPF was considered as the worst-case error. For the given same inputs, the RMSE of the CC DC-OPF is RMSE $= 6.65e^{-2}$ p.u. of IEEE9, RMSE $= 3.04e^{-2}$ p.u. of IEEE39, and RMSE $= 5.84e^{-1}$ p.u. of IEEE118. In other words, applying the GPR model with a few samples leads to at least 6 times more accurate approximation than the DC baseline.

Therefore, our proposed hybrid GP approach in both cases gives a smaller error than the well-known DC approximation. The comparison of the results are summarized in Table 5.1.

Test Case	IEEE 9		IEEE 39		IEEE 118	
Parameters	RMSE 10^{-2} p.u.	Time sec.	RMSE 10^{-2} p.u.	Time sec.	RMSE 10^{-2} p.u.	Time sec.
hybrid TA1	**0.30**	0.24	**0.53**	10.28	**21.4**	162.4
hybrid-sparse 50 TA1	0.56	0.10	1.12	2.93	23.2	29.3
hybrid-sparse 30 TA1	0.79	0.06	1.22	1.23	31.4	14.5
hybrid-sparse 10 TA1	1.22	**0.02**	1.41	**0.51**	39.9	**7.4**
full TA1	0.63	0.35	1.37	19.4	83.7	212.1
full-sparse 50 TA1	0.97	0.17	2.69	4.12	87.8	41.1
full-sparse 30 TA1	1.67	0.13	2.80	1.42	93.1	24.1
full-sparse 10 TA1	2.43	0.03	3.25	**0.51**	99.7	10.2

Table 5.1: GP CC-OPF approximation method results for IEEE 9, IEEE 39 and IEEE 118 bus systems.

To support readers in better understanding the results presented in Table 5.1, we provide graphical illustrations for all systems in Fig. 5-2, 5-3 and 5-4 .

5.4.3 Comparison of GP CC-OPF to Sample-Based Methods

In this section, we compare optimal objective function value, computational time, and reliability of the proposed hybrid and full GP CC-OPF approaches with Scenario
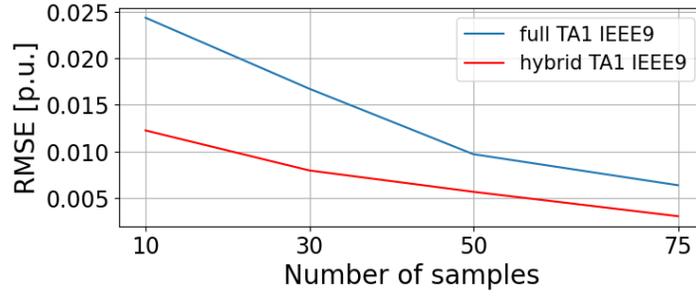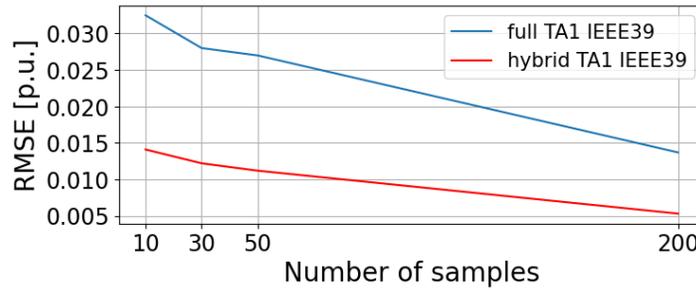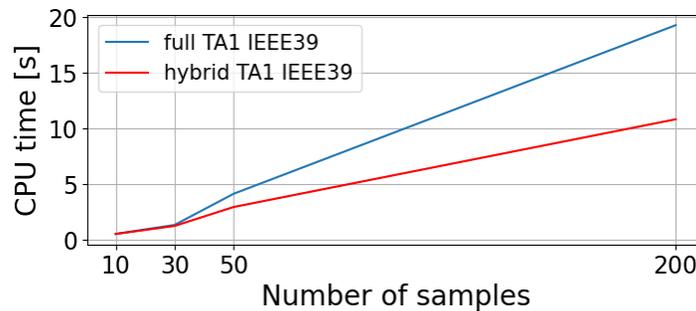
(a) RMSE IEEE9
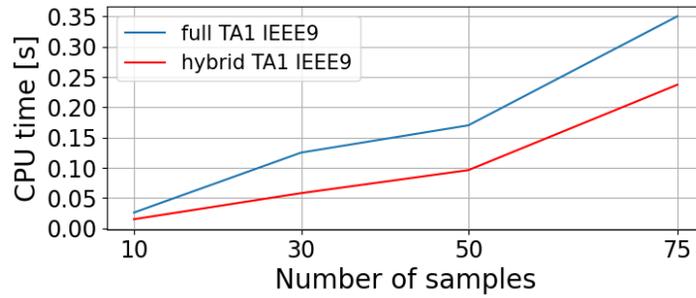
(b) CPU time IEEE9

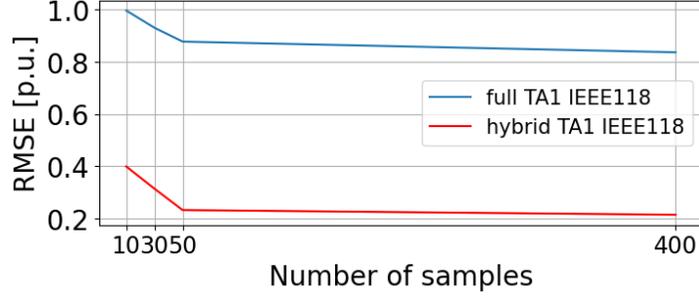Figure 5-2: Ilustration of the RMSE and CPU time results for the IEEE9 bus system.
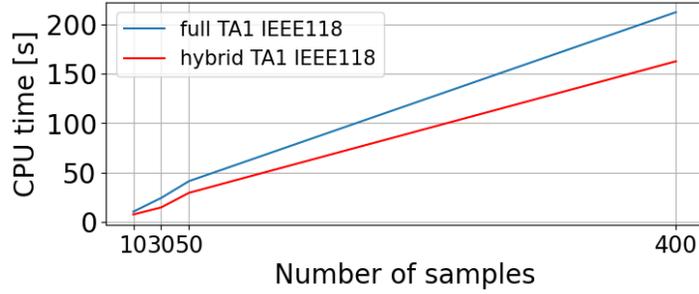

(a) RMSE IEEE39

(b) CPU time IEEE39

Figure 5-3: Ilustration of the RMSE and CPU time results for the IEEE39 bus system.

(a) RMSE IEEE118

(b) CPU time IEEE118

Figure 5-4: Ilustration of the RMSE and CPU time results for the IEEE118 bus system.

Approximation (SA) CC-OPF [Mezghani et al., 2020]. In all cases, we use the first-order Taylor approximation (TA1) as a function of linearization. The empirical constraint violation is computed using 1000 Monte-Carlo samples that follow (Gaussian) uncertainty distribution.

We compare our approach to two baselines. (A) Deterministic AC-OPF problem for each uncertainty realization and taking a corresponding quantile in the output distribution. (B) A deterministic AC-OPF for the mean load and RES, with fixed and equal participation factor α^*. Notice that (A) is unrealistic as it solves AC-OPF after uncertainty realization (full-resource), while (B) does not optimize the participation factor α. Thus (A) gives the upper (though unrealistic) bound and (B) gives the lower bound on the feasible solution value. Numerical evaluation of the bounds, given in Table 5.2, shows that the hybrid-sparse GP modeling provides a flexible scheme that can improve computational effort without sacrificing accuracy.

Results in Table 5.2 show that the proposed hybrid GP CC-OPF model outperforms sample-based chance constraints reformulation in the context of computational complexity. In all system cases, the proposed GP method needs an enormously lower CPU time to find a solution. Moreover, the hybrid approach is faster than GP CC-OPF based on full AC-PF from **Chapter 4**.

System	IEEE 9		
Parameters	Cost [$]	Infeas. Prob. [%]	Time (s)
A (full recourse)	$4.056 \cdot 10^3$	-	0.86
B (base-case)	$3.467 \cdot 10^3$	9.76	0.86
full GP CC-OPF	$4.039 \cdot 10^3$	0.44	0.35
hybrid GP CC-OPF	$4.024 \cdot 10^3$	0.80	0.12
20 CC-OPF	$3.481 \cdot 10^3$	7.52	4.4
50 CC-OPF	$3.836 \cdot 10^3$	4.92	12.1
100 CC-OPF	$3.985 \cdot 10^3$	1.24	22.8

System	IEEE 39		
Parameters	Cost [$]	Infeas. Prob. [%]	Time (s)
A (full recourse)	$7.869 \cdot 10^6$	-	0.93
B (base case)	$7.533 \cdot 10^6$	35.36	0.93
full GP CC-OPF	$7.752 \cdot 10^6$	2.36	19.38
hybrid GP CC-OPF	$7.752 \cdot 10^6$	2.48	10.82
50 CC-OPF	$7.642 \cdot 10^6$	14.20	96.4
100 CC-OPF	$7.696 \cdot 10^6$	8.24	184.7
200 CC-OPF	$7.813 \cdot 10^6$	0.16	505.0

| System	IEEE 118		
Parameters	Cost [$]	Infeas. Prob. [%]	Time (s)
A (full recourse)	$26.91 \cdot 10^6$	-	1.59
B (base case)	$26.39 \cdot 10^6$	49.97	1.59
full GP CC-OPF	$26.91 \cdot 10^6$	2.47	212.1
hybrid GP CC-OPF	$26.91 \cdot 10^6$	2.49	162.4
100 CC-OPF	$26.52 \cdot 10^6$	28.12	683.3
200 CC-OPF	$26.71 \cdot 10^6$	19.62	1856.5
500 CC-OPF	$26.90 \cdot 10^6$	4.78	39537.1

Table 5.2: Cost function values and probability of violation of a constraint at a solution

### 5.4.4 Interpretation of results

In this section, we will provide a visual interpretation of the results obtained from the previous section 5.4.3. We will consider all three parameters: cost, security (infeasible probability), and time, as shown in Table 5.2. These parameters will be graded on a scale from 0 to 5, where 0 represents the worst case and 5 represents the best.

Fig. 5-5 presents a visual comparison of all approaches in terms of the highlighted parameters. In each case (IEEE9, IEEE39, and IEEE118), the red horizontal lines represent the trade-off between security and cost. It is evident that both the hybrid and full approaches are capable of finding a suitable trade-off between security and cost. However, the hybrid approach exhibits faster performance.

Furthermore, as the system size increases and the approaches become more scalable, the time required for both the hybrid and full approaches also increases compared to
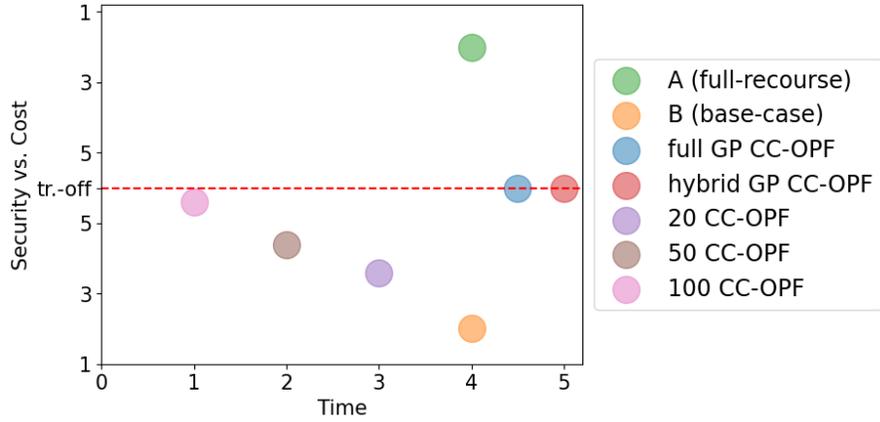
(a) IEEE9
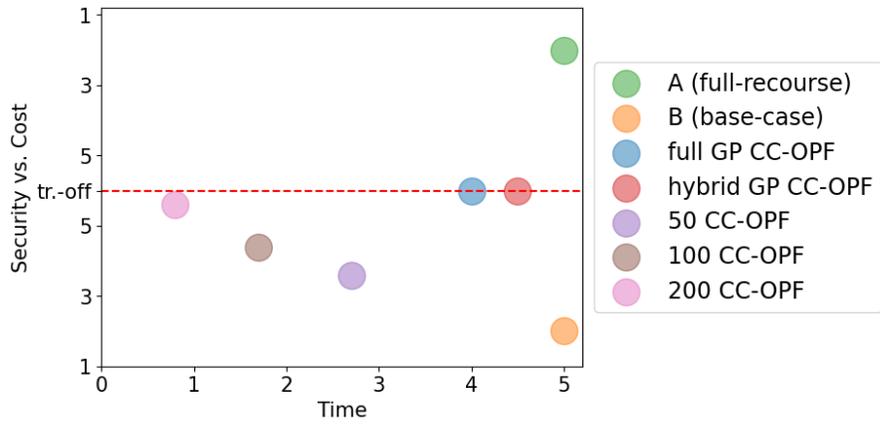
(b) IEEE39
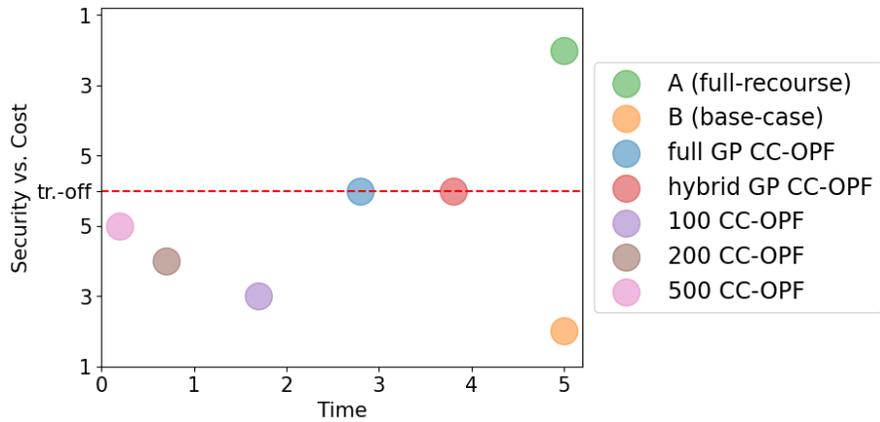
(c) IEEE118

Figure 5-5: Illustration of cost, security, and time comparison approaches for solving stochastic AC-OPF.

approaches A (full-resources) and B (base-case). Despite this increase in time, the trade-off between security and cost is maintained.

5.4.5 Comparison of Uncertainty Margins

Fig. 5-6 compares the uncertainty margins [Schmidli et al., 2016] calculated with the hybrid and full GP CC-OPF (red and blue), the Monte Carlo simulation of AC-OPF for base-case (B) (yellow and green) and the Monte Carlo simulation for CC-OPF solution (brown and dark salmon) with 100 (IEEE9), 200 (IEEE39) and 500 (IEEE118) scenarios. The bar plots represent an example of voltage, reactive power and apparent power flow uncertainty margins for each system.

The three standard deviations (3 std) intervals for uncertainty margins are analytically derived using the estimated GP-variance (5.4e) around the estimated GP-mean (5.4d). The uncertainty margins from Monte Carlo are asymmetrical due to MC-based distributions owing to non-linear AC-PF. Thus, Monte Carlo uncertainty margins determine upper $(1 - \epsilon) = 99.73\%$ and lower $(\epsilon) = 0.27\%$ quantiles (equal to 3 std) of the output distribution around the AC-PF solution $y(x)$, denoted by $y_{1-\epsilon}^{out}$ and y_{ϵ}^{out}. Thus, the constraint tightenings are calculated by:

$$\lambda^{upper} = y_{1-\epsilon}^{out} - y^{out}(x^{in}) \tag{5.6a}$$

$$\lambda^{lower} = y^{out}(x^{in}) - y_{\epsilon}^{out} \tag{5.6b}$$

Compared to the voltage margins, the power margins are much larger in absolute terms. The GP CC-OPF margins of both approaches are symmetric, leading to larger or smaller margins compared with the upper and lower empirical quantiles obtained from the Monte Carlo AC-OPF and CC-OPF simulations in (5.6). This indicates that the GP uncertainty margins are well approximated.

### 5.4.6 Analysis in the Case of Distribution Gaps

In this section, we analyze the presence of gaps in distributions in the IEEE9 case. The purpose of the experiment is to test the robustness of the ML algorithm in case of corrupted data. The gaps in distributions may appear in case of an information attack, or poor data collection. A simple example of that was presented in Fig. 5-1 – the interval from, approximately, 7.5 to 17 has a lack of input data coverage which leads to poor performance of full GP. However, the hybrid GP, presented in this thesis handles such intervals better. We expect to observe a similar result in the case of power grids as well.

First, let us specify the experiment setup. The experiment consists of three major steps: corrupting the data, training on the corrupted data and testing the predictions of models. We compare two models: full GP from **Chapter 4** and the proposed hybrid GR model. The corruption step is conducted as follows: after generating the training dataset, we throw out a part of data, i.e., samples of pairs (x_i^{in}, y_i^{out}), such that it would result in crossing out an interval of some buses' input data. Examples of such corruption are provided in Fig. 5-7. Next, both full and hybrid GP are trained on the corrupted dataset. Finally, the RMSE metric is being evaluated on both regression models on the crossed-out regions.

We conduct the outlined experiment on case IEEE9, with the dataset generated
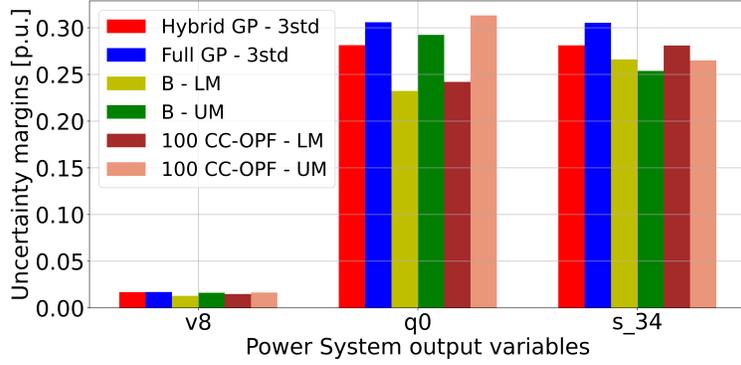
(a) IEEE9
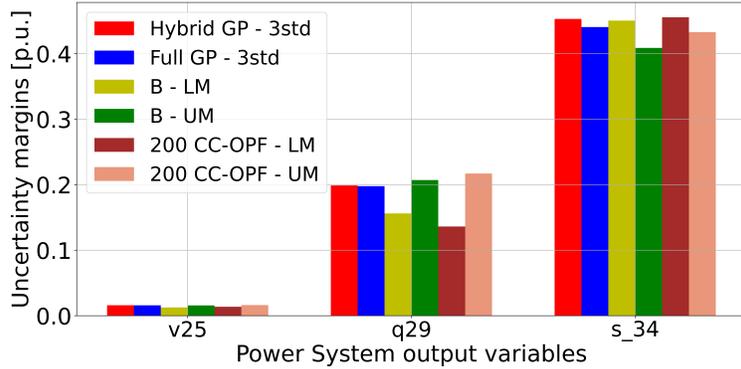
(b) IEEE39
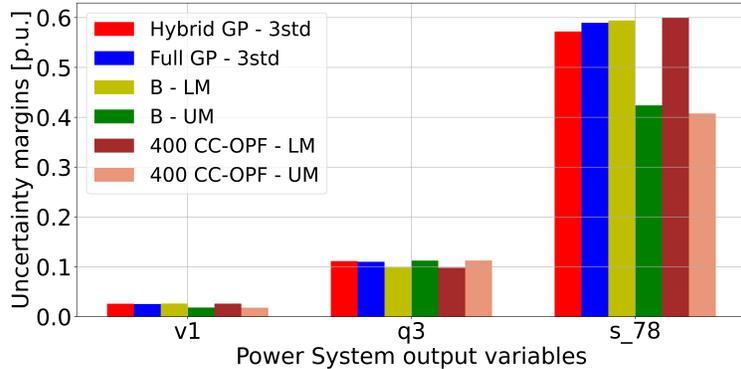
(c) IEEE118

Figure 5-6: Output uncertainty margins λ, for example, voltage, reactive and apparent power constraints. (LM - lower margin 0.27% interval; UM - upper margin 99.73% interval; B - deterministic AC-OPF for the mean load and RES, with fixed and equal participation factor α^*).

as in Section 2.4. The corruption of data is conducted on three buses separately, i.e., in each of the three experiments only one bus is a target for corruption. The load fluctuation data is crossed out at buses 4, 6 and 8. In each case, 15 MW interval has been thrown out. The histograms of load data are presented in Figure 5-7. The others' distributions have not experienced any significant change. Next, we trained ordinary

and hybrid GPR on these data. The evaluated RMSE metrics are presented in Table 5.3.
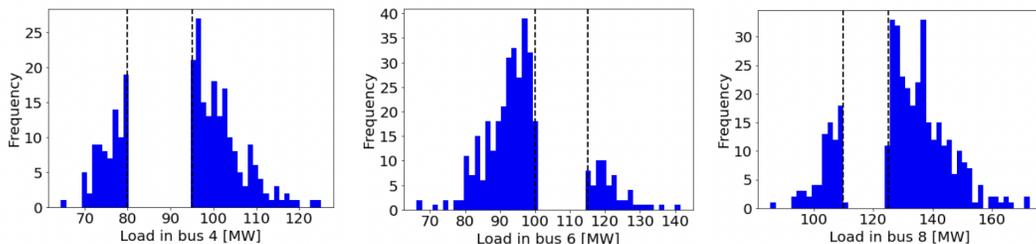

Figure 5-7: An example of misspecified input, where a 15 MW interval around the mean bus load value is inaccessible. The latter may happen as a result of cyberattacks or meter failures. The proposed algorithm demonstrates stability over significant adversarial changes in input data.

In each case, the hybrid GP approach shows from one to two orders of magnitude lower RMSE than the full GP approach. On a closer look, one can observe that the difference in RMSE magnitude is larger when the number of thrown-out samples is larger, in other words, the larger probability mass is being crossed out. Specifically, the vicinity of expectation, the interval with the highest density (55%), has been excluded in the experiment with bus 4. This led to the difference of 2 magnitude order in RMSE of ordinary and hybrid GPR. Conversely, for experiments with loads 6 and 8, the difference in magnitude order is 1, since the distribution mass is lower – 38 % and 33.5 %, respectively. The results are summarized in Table 5.3.

Experiment	RMSE GPR	RMSE Hybrid GPR	% dropped
Load 4	$3.65 \cdot 10^{-1}$	$\mathbf{9.11 \cdot 10^{-3}}$	55.0%
Load 6	$3.89 \cdot 10^{-1}$	$\mathbf{1.16 \cdot 10^{-2}}$	38.0%
Load 8	$9.25 \cdot 10^{-2}$	$\mathbf{8.34 \cdot 10^{-3}}$	33.5%

Table 5.3: Summary of RMSE comparison in the case of missing data. The last column specifies the portion of data distribution of the corresponding load that has been dropped.

In summary, we studied the robustness of the proposed novel hybrid GPR method for CC AC-OPF against previously proposed full GPR for CC AC-OPF on an example of a missing part of the data domain. One can conclude that the proposed hybrid is more robust since it provides smaller RMSE even in the regions of the domain that have not been covered in the training data.

Another approach that may improve the results stands for introducing ranking regularization when learning the Gaussian processes [Sidana et al., 2017, Sidana et al., 2021]. The latter allows to prioritize the error minimization across certain most critical lines and/or most uncertain outputs.

5.5   Discussion

To evaluate the optimization performance and uncertainty margins of the proposed hybrid approach, different stochastic approaches are considered. These approaches include the well-known sample-based approaches and the analytical state-of-the-art GP-based approach proposed in **Chapter** 4. This Chapter represents an improvement of the full GP approach that is faster and more robust.

The advantage of the GP approach is that it is based on Gaussian processes and thus enables to avoid of local sub-minima during optimization and realization of chance constraints in an easy and cheap way. The hybrid approach is faster since initial optimizing points are better defined with the help of a nominal linear model. Robustness is reflected in the fact that in the case where the historical data is poorly distributed in the space due to the system's technical nature, the proposed model will be equal to the linearly approximated DC model. Thus, the possible prediction error will be significantly reduced.

Although the proposed approach reduces the computational time, still it will be increased with system scale. This problem can be easily overcome by applying sparse GPR and dimensionality reduction techniques while maintaining a similar solution. Also, this framework can be easily extended to multi-period extensions of CC-OPF, where forecasts and errors may be correlated.

5.6   Conclusion

This Chapter discusses the use of Gaussian process regression (GPR) to learn non-linearities in stochastic OPF for improved performance. By combining the additive GPR with a linear power system model, we can improve solution accuracy and reduce computational complexity. This approach also helps to avoid spurious local minimums, making the AC-OPF problem more tractable. However, since GPR is not very scalable, a sparsity trick is employed to significantly reduce complexity. Accordingly, we propose a hybrid GPR approach that shows higher robustness to missing and corrupted data than the classical full GPR. Through simulation examples, we demonstrate how the proposed formulations provide secure control with improved performance for numerous IEEE test cases, with low optimization time. In particular, our approach provides a trade-off between the system's security, optimal cost, and time complexity.

Chapter 6

GP CC-OPF Software

6.1 Introduction

In this chapter, a brief overview is presented on the software that has been created from scratch to facilitate the realization of the contributions made in this thesis. The code for this software is accessible through GitHub https://github.com/mile888/full, https://github.com/mile888/hybrid and Codeocean platforms https://codeocean.com/. In addition to this, a paper has been published that describes this code in *Software Impact* journal [3].

6.2 User guide

The software implements the GP CC-OPF approach, which is depicted in Fig. 6-1. This approach has been developed within a Python environment and relies on various libraries such as NumPy, SciPy, Pandas, and CasADi. NumPy is utilized for efficient matrix and vector operations, which are used for constructing the proposed algorithm. SciPy is employed for optimizing the GP model, while Pandas played a crucial role in generating and preparing synthetic datasets. CasADi, on the other hand, serves as a solver for non-linear optimization problems and is used in our case to model and realize the CC-OPF approach. Code details are presented below.

6.2.1 Data preparation

We investigate two software approaches: the full GP CC-OPF and the hybrid GP CC-OPF. The key distinction between these approaches lies in their training methodologies.

The full approach is trained on the complete AC-PF balance equations, while the hybrid approach consists of a linear DC part and a residual part. The linear part can be implemented in two ways: using physical linear DC-PF equations or adopting a data-driven way. In this study, we employ the data-driven way, where data from the DC-PF model is generated to learn the linear function. The linear model, illustrated in the Fig. 6-1, utilizes ordinary least squares Linear Regression provided by the scikit-learn library [Pedregosa et al., 2011]. Additionally, the data for the residual
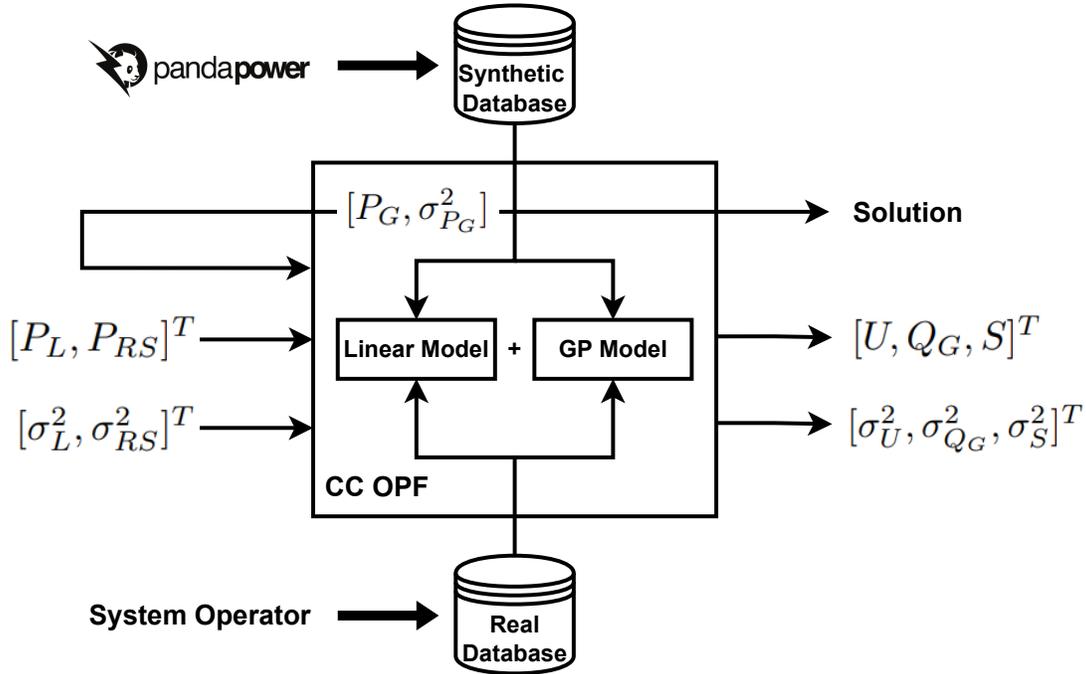
Figure 6-1: General scheme of GP CC-OPF approach realized in the software.

part is generated by calculating the residuals between the AC-PF and DC-PF sample results on which the GP is trained.

Since all simulations are performed on synthetic datasets, we have developed a program for generating these datasets. We utilize the pandapower framework [Thurner et al., 2018] to simulate power systems and generate the corresponding synthetic datasets. While the ultimate goal is for this software to be applicable to real-world data, we currently evaluate its performance using synthetic data. In both cases and approaches, the GP model is trained on standardized data.

The folder named *<IEEE{x} - pandapower>* for full and *<IEEE{x}_hybrid - pandapower>* for hybrid approach provide the implementation details on how to simulate static power system flows and generate synthetic datasets using the pandapower framework.

6.2.2   GP CC-OPF setup

The GP CC-OPF approach combines conventional chance-constrained optimization with an integrated data-driven machine learning model, as illustrated in Fig. 6-1. The software implementation comprises several Python modules that interact with each other:

- **gp_functions.py**: This module creates the machine learning model using the SEard (Squared Exponential with Automatic Relevance Determination) kernel and various approximations such as TA1, TA2, and EM.

- **gp_optimization.py**: The gp_optimization.py module defines the optimization structure for the GP model.

- **gp_regression.py**: This module integrates both gp_functions.py and gp_optimization.py to train the GP model.

- **gp_ccopf.py** (for the full approach): The gp_ccopf.py module incorporates the trained GP model for the full GP CC-OPF approach.

- **gp_ccopf_hybrid.py** (for the full approach): The gp_ccopf_hybrid.py module incorporates the trained GP model for the hybrid GP CC-OPF approach.

- **linear_model.py** (for the hybrid approach): In this module, a Linear Regression model is created to generate the linear DC part within the hybrid approach.

These modules work together to implement the full GP CC-OPF approach and its hybrid variant. These modules work together to implement the full GP CC-OPF approach and its hybrid variant. A schematic representation of the module interaction is shown in Fig. 6-2.
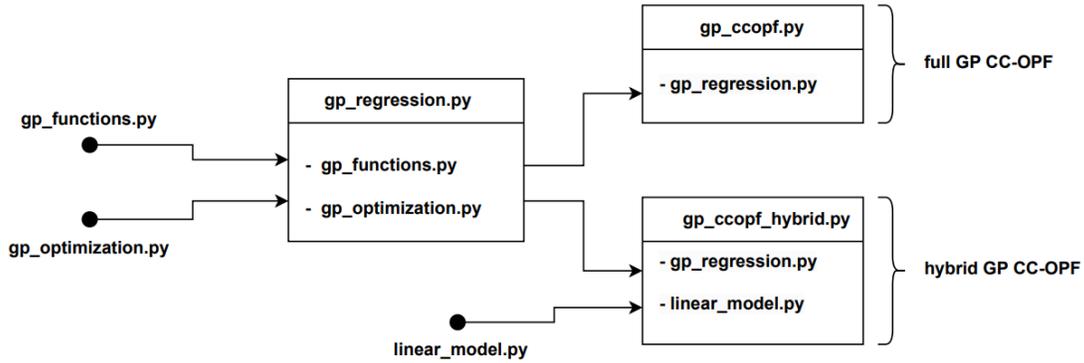

Figure 6-2: General scheme of GP CC-OPF approach realized in the software.

The gp_functions.py module creates the GP machine learning model while gp_optimization.py provides the optimization structure of the GP model. In both approaches gp_regression.py combines them for training the GP model. The resulting model is then integrated into gp_ccopf.py for the full approach as well as gp_ccopf_hybrid.py for the hybrid approach. Additionally, the linear_model.py module is included in gp_ccopf_hybrid.py module and handles the creation of the Linear Regression model for the linear DC part.

6.2.3 Outputs of the tool

The GP CC-OPF approach, implemented in the developed tool, enables users to obtain optimal solutions for controllable generators while minimizing the cost. Additionally, the tool provides valuable information such as output variables and their associated propagated standard deviations. These variables include voltage magnitudes at non-controllable buses, reactive power at controllable buses, and apparent power flow

in the grid lines. By examining these outputs, users can assess the feasibility and accuracy of the obtained solution.

In summary, the developed tool utilizing the GP CC-OPF approach not only delivers optimal solutions for controllable generators but also provides crucial information regarding the system's variables and their propagated standard deviations. This allows users to monitor the feasibility and accuracy of the obtained solution effectively.

6.3 Conclusion

Currently, the industry lacks developed software that effectively solves stochastic OPF problems. System operators predominantly rely on deterministic OPF solutions that do not account for generation uncertainties. In the scientific community, although considerable progress has been made in stochastic OPF, the algorithms proposed in the literature are often inaccessible to the public. This presents challenges for researchers in terms of comparing their methods with existing work and necessitates significant effort to re-implement existing algorithms. Consequently, the availability of open-source code can significantly impact both industry and academia.

In future versions, our plan is to enhance the tool by introducing a user-friendly interface that is intuitive to use. We firmly believe that the proposed data-driven approach, presented in the form of public code, will have a positive impact on future research in stochastic OPF. Furthermore, we anticipate that it will serve as a stimulus for engineers to adopt these methods in the industry.

Chapter 7

Conclusion and Future Perspectives

7.1 Concluding remarks

This thesis addresses the challenge of solving stochastic OPF using a data-driven chance-constrained approach. To achieve this, we use a Gaussian process regression (GPR) model that enables us to propagate uncertainties and satisfy the analytical reformulation of CC-OPF. The first significant contribution of the thesis is to replace the full AC-PF balance equations with GPR and incorporate them into the CC-OPF problem. This approach leads to significant improvements in both computational complexity and the accuracy of the solution. Furthermore, this approach facilitates the calculation of uncertainty margins in stochastic AC-OPF, achieving less than a 2.5% deviation in chance constraint optimization. A key advantage of this approach is the use of the Squared Exponential covariance function within the Gaussian Process model. This allows us to approximate smoother functions, consequently avoiding local and sub-local minima, leading to faster and more robust solutions.

However, this approach can be affected by the quality of the dataset, which can impact the solution's robustness. To address this issue, we propose a more robust solution that is at worst-case equal to the linearly approximated DC model. We achieve this by proposing a hybrid approach that combines a linear DC model with a GP-based model learned on the residuals between AC-PF and DC-PF. This approach overcomes the drawbacks of the full GP approach and provides a more robust and efficient solution. To further improve the computational complexity and find a trade-off between complexity and accuracy, we use a sparse GP technique. This technique enables us to significantly reduce computational complexity while maintaining a high level of accuracy. Through the implementation of this hybrid and sparse approach, we can enhance efficiency, achieving a minimum 2x speedup and up to 6x more accurate solutions when compared to state-of-the-art approaches.

Overall, our approach offers a promising solution for addressing stochastic OPF problems by combining data-driven strategies with chance-constrained methodologies and leveraging machine learning techniques. Additionally, it is worth noting that neither of the proposed GP approaches necessitates system parameters or topology information. Consequently, the implementation of these methods can be seamlessly

executed in real-time, with the overall efficiency contingent upon the quality of the dataset.

In summary, this thesis has yielded significant research outcomes such as:

1. developed a data-driven approach that analytically calculates uncertainty margins in stochastic AC-OPF with less than 2.5% deviation in chance constraints optimization;

2. local and sub-local minima could be avoided by using the Squared Exponential covariance function of the Gaussian Process model;

3. implementing a hybrid and sparse approach can improve efficiency with a minimum speedup of 2 times and up to 6 times more accurate solutions compared to state-of-the-art approaches;

4. information about system parameters and topologies is not required.

It is worth noting that the first contribution to this dissertation, assuming a full GP CC-OF approach, is published in the *International Journal of Electrical Power & Energy Systems (IJEPES)* [2]. The proposed hybrid GP CC-OF approach is presented and published at the *IEEE Belgrade PowerTech 2023* conference [1]. The software developed to implement the approaches presented in this thesis is explained and published in the *Software Impact* journal [3].

7.2   Future perspectives

The current work is based on synthetic data. Although many ML solutions and applications have been proposed, they have not yet been widely implemented in real-world settings and are often only tested on open-source data. To evaluate the effectiveness of proposed software solutions, it would be valuable to test them on real power system data generated by Transmission System Operators (TSOs).

If the proposed software solution works well on real-world data, it has a high probability of becoming a primary data-driven solution for solving stochastic Optimal Power Flow (OPF) problems globally.

Bibliography

[Abbaspourtorbati and Zima, 2015] Abbaspourtorbati, F. and Zima, M. (2015). The swiss reserve market: Stochastic programming in practice. *IEEE Transactions on Power Systems*, 31(2):1188–1194.

[Amini et al., 2022] Amini, M.-R., Feofanov, V., Pauletto, L., Devijver, E., and Maximov, Y. (2022). Self-training: A survey. *arXiv preprint arXiv:2202.12040*.

[Andersson et al., 2019] Andersson, J. A., Gillis, J., Horn, G., Rawlings, J. B., and Diehl, M. (2019). Casadi: a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11(1):1–36.

[Anikin et al., 2022] Anikin, A., Gasnikov, A., Gornov, A., Kamzolov, D., Maximov, Y., and Nesterov, Y. (2022). Efficient numerical methods to solve sparse linear equations with application to pagerank. *Optimization Methods and Software*, 37(3):907–935.

[Baker and Toomey, 2017] Baker, K. and Toomey, B. (2017). Efficient relaxations for joint chance constrained ac optimal power flow. *Electric Power Systems Research*, 148:230–236.

[Ben-Tal et al., 2009] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust optimization*, volume 28. Princeton university press.

[Bienstock et al., 2014] 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.

[Boyd and Vandenberghe, 2004] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. New York: Cambridge University Press.

[Burashnikova, 2022] Burashnikova, A. (2022). Large-scale sequential learning for recommender and engineering systems. *arXiv preprint arXiv:2205.06893*.

[Burashnikova et al., 2022] Burashnikova, A., Li, W., Amini, M., Deka, D., and Maximov, Y. (2022). Ranking-based physics-informed line failure detection in power grids. *arXiv preprint arXiv:2209.01021*.

[Burashnikova et al., 2021] Burashnikova, A., Maximov, Y., Clausel, M., Laclau, C., Iutzeler, F., and Amini, M.-R. (2021). Learning over no-preferred and preferred sequence of items for robust recommendation. *Journal of Artificial Intelligence Research*, 71:121–142.

[Cain et al., 2012] Cain, M. B., O'neill, R. P., Castillo, A., et al. (2012). History of optimal power flow and formulations. *Federal Energy Regulatory Commission*, 1:1–36.

[Calafiore and Campi, 2013] Calafiore, G. and Campi, M. c. (2013). Probabilistic security-constrained ac optimal power flow. *IEEE Trans. Autom. Control*, 51(5):742–753.

[Calafiore and Campi, 2006] Calafiore, G. C. and Campi, M. C. (2006). The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753.

[Capitanescu, 2016] Capitanescu, F. (2016). Critical review of recent advances and further developments needed in ac optimal power flow. *Electric Power Systems Research*, 136:57–68.

[Capitanescu et al., 2012] Capitanescu, F., Fliscounakis, S., Panciatici, P., and Wehenkel, L. (2012). Cautious operation planning under uncertainties. *IEEE Transactions on Power Systems*, 27(4):1859–1869.

[Capitanescu et al., 2011] Capitanescu, F., Ramos, J. M., Panciatici, P., Kirschen, D., Marcolini, A. M., Platbrood, L., and Wehenkel, L. (2011). State-of-the-art, challenges, and future trends in security constrained optimal power flow. *Electric power systems research*, 81(8):1731–1741.

[Carpentier, 1962] Carpentier, J. (1962). Contribution a l'etude du dispatching economique. *Bulletin de la Société Francaise des électriciens*, 3(1):431–447.

[Castillo et al., 2015] Castillo, A., Lipka, P., Watson, J.-P., Oren, S. S., and O'Neill, R. P. (2015). A successive linear programming approach to solving the iv-acopf. *IEEE Transactions on Power Systems*, 31(4):2752–2763.

[CIGRE, 2009] CIGRE (2009). Technical brochure on grid integration of wind generation. *International Conference on Large High Voltage Electric Systems*.

[Cunningham et al., 2008] Cunningham, P., Cord, M., and Delany, S. J. (2008). Supervised learning. *Machine learning techniques for multimedia: case studies on organization and retrieval*, pages 21–49.

[Deisenroth, 2010] Deisenroth, M. P. (2010). *Efficient reinforcement learning using Gaussian processes*, volume 9. KIT Scientific Publishing.

[DENA, 2005] DENA (2005). Energy management planning for the integration of wind energy into the grid in germany, onshore and offshore by 2020.

[Donnot, 2019] Donnot, B. (2019). *Deep learning methods for predicting flows in power grids: novel architectures and algorithms*. PhD thesis, Université Paris Saclay (COmUE).

[Du et al., 2021] Du, X., Lin, X., Peng, Z., Peng, S., Tang, J., and Li, W. (2021). Chance-constrained optimal power flow based on a linearized network model. *International Journal of Electrical Power & Energy Systems*, 130:106890.

[Dudley, 2010] Dudley, R. M. (2010). Sample functions of the gaussian process. In *Selected Works of RM Dudley*, pages 187–224. Springer.

[EERE, 2008] EERE (2008). 20% of wind energy by 2030: Increasing wind energy's contribution to us electricity supply. *International Conference on Large High Voltage Electric Systems*.

[EIA,] EIA. Energy information administration, international energy statistics. Retrieved from http://www.eia.gov/cfapps/ipdbproject/IEDIndex3.cfm?tid=2&pid=2&aid=2.

[Gelman et al., 1995] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

[Ghahramani, 2004] Ghahramani, Z. (2004). Unsupervised learning. *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003, Tübingen, Germany, August 4-16, 2003, Revised Lectures*, pages 72–112.

[Girard et al., 2002] Girard, A., Rasmussen, C., Candela, J. Q., and Murray-Smith, R. (2002). Gaussian process priors with uncertain inputs application to multiple-step ahead time series forecasting. *Advances in neural information processing systems*, 15.

[Girard et al., 2003] Girard, A., Rasmussen, C. E., Quinonero-Candela, J., and Murray-Smith, R. (2003). Gaussian process priors with uncertain inputs? application to multiple-step ahead time series forecasting. In *Advances in neural information processing systems*, volume 15.

[Gonzalez et al., 2006] Gonzalez, G., ALVIRA, D., LÓPEZ, S., LANCHA, D., SANTOS, D., CASTAÑS, E., and FERNÁNDEZ, E. (2006). Experience integrating and operating wind power in the peninsular spanish power system. point of view of the transmission system operator and a wind power producer.

[Gordon and Tibshirani, 2012] Gordon, G. and Tibshirani, R. (2012). Karush-kuhn-tucker conditions. *Optimization*, 10(725/36):725.

[Grinis, 2016] Grinis, R. (2016). *On the soliton resolution conjecture for wave maps*. PhD thesis, University of Oxford.

[Grinis, 2022] Grinis, R. (2022). Differentiable programming for particle physics simulations. *Journal of Experimental and Theoretical Physics*, 134(2):150–156.

[Gupta et al., 2021] Gupta, S., Misra, S., Deka, D., and Kekatos, V. (2021). Dnn-based policies for stochastic ac opf. *arXiv preprint arXiv:2112.02441*.

[Halilbašić et al., 2018] Halilbašić, L., Pinson, P., and Chatzivasileiadis, S. (2018). Convex relaxations and approximations of chance-constrained ac-opf problems. *IEEE Transactions on Power Systems*, 34(2):1459–1470.

[Hastie et al., 2009] Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Tibshirani, R., and Friedman, J. (2009). Overview of supervised learning. *The elements of statistical learning: Data mining, inference, and prediction*, pages 9–41.

[Hewing et al., 2019] Hewing, L., Kabzan, J., and Zeilinger, M. N. (2019). Cautious model predictive control using gaussian process regression. *IEEE Transactions on Control Systems Technology*, 28(6):2736–2743.

[Hou and Roald, 2020] Hou, A. M. and Roald, L. A. (2020). Chance constraint tuning for optimal power flow. In *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 1–6. IEEE.

[Kadilenko and Grinis, 2023] Kadilenko, E. M. and Grinis, R. (2023). Adjoint sensitivity method for kohn-sham equations.

[Kaelbling et al., 1996] Kaelbling, L. P., Littman, M. L., and Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285.

[Kocijan et al., 2004] Kocijan, J., Murray-Smith, R., Rasmussen, C. E., and Girard, A. (2004). Gaussian process model based predictive control. In *Proceedings of the 2004 American control conference*, volume 3, pages 2214–2219. IEEE.

[Koller et al., 2018] Koller, T., Berkenkamp, F., Turchetta, M., and Krause, A. (2018). Learning-based model predictive control for safe exploration. In *2018 IEEE conference on decision and control (CDC)*, pages 6059–6066. IEEE.

[Kraft, 1988] Kraft, D. (1988). A software package for sequential quadratic programming. *Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt*.

[Krechetov et al., 2018] Krechetov, M., Marecek, J., Maximov, Y., and Takac, M. (2018). Entropy penalized semidefinite programming. *arXiv preprint arXiv:1802.04332*.

[Kundur et al., 1994] Kundur, P., Balu, N. J., , and Lauby, M. G. (1994). *Power system stability and control*, volume 7. McGraw-hill New York.

[Kundur, 2012] Kundur, P. S. (2012). *Power system stability.* Power System Stability and Control, Third Edition.

[LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.

[Li, 2017] Li, Y. (2017). Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274.*

[Liu et al., 2020] Liu, H., Ong, Y.-S., Shen, X., and Cai, J. (2020). When gaussian process meets big data: A review of scalable gps. *IEEE transactions on neural networks and learning systems*, 31(11):4405–4423.

[Liu et al., 2012] Liu, Z., van der Sluis, L., Winter, W., Paeschke, H., Becker, R., Weber, C., Eickmann, J., Schroeders, C., Oldewurtel, F., Roald, L., et al. (2012). Challenges, experiences and possible solutions in transmission system operation with large wind integration. In *11th International Workshop on Large-Scale Integration of Wind Power into Power Systems*. 11th International Workshop on Large-Scale Integration of Wind Power into . . . .

[Low, 2014a] Low, S. H. (2014a). Convex relaxation of optimal power flow—part i: Formulations and equivalence. *IEEE Transactions on Control of Network Systems*, pp.15.

[Low, 2014b] Low, S. H. (2014b). Convex relaxation of optimal power flow—part ii: Exactness. *IEEE Transactions on Control of Network Systems*, pp.177.

[Low, 2019] Low, S. H. (2019). Solutions of dc opf are never ac feasible. *arXiv:1912.00319.*

[Lowd and Meek, 2005] Lowd, D. and Meek, C. (2005). Adversarial learning. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 641–647.

[Lubin et al., 2015] Lubin, M., Dvorkin, Y., and Backhaus, S. (2015). A robust approach to chance constrained optimal power flow with renewable generation. *IEEE Transactions on Power Systems*, 31(5):3840–3849.

[Lubin et al., 2019] Lubin, M., Dvorkin, Y., and Roald, L. (2019). Chance constraints for improving the security of ac optimal power flow. *IEEE Transactions on Power Systems*, 34(3):1908–1917.

[Lukashevich et al., 2021] Lukashevich, A., Gorchakov, V., Vorobev, P., Deka, D., and Maximov, Y. (2021). Importance sampling approach to chance-constrained dc optimal power flow. *arXiv preprint arXiv:2111.11729.*

[Lukashevich and Maximov, 2021] Lukashevich, A. and Maximov, Y. (2021). Power grid reliability estimation via adaptive importance sampling. *IEEE Control Systems Letters*, 6:1010–1015.

[Margellos et al., 2014] Margellos, K., Goulart, P., and Lygeros, J. (2014). On the road between robust optimization and the scenario approach for chance constrained optimization problems. *IEEE Trans. Autom. Control*, 59(8):2258–2263.

[Mezghani et al., 2020] Mezghani, I., Misra, S., and Deka, D. (2020). Stochastic ac optimal power flow: A data-driven approach. *Electric Power Systems Research*, 189:106567.

[Mikhalev et al., 2020] Mikhalev, A., Emchinov, A., Chevalier, S., Maximov, Y., and Vorobev, P. (2020). A bayesian framework for power system components identification. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE.

[Morillo et al., 2022] Morillo, J. L., Zephyr, L., Pérez, J. F., Cadena, A., and Anderson, C. L. (2022). Distribution-free chance-constrained load balance model for the operation planning of hydrothermal power systems coupled with multiple renewable energy sources. *International Journal of Electrical Power & Energy Systems*, 142:108319.

[Mühlpfordt et al., 2016] Mühlpfordt, T., Faulwasser, T., and Hagenmeyer, V. (2016). Solving stochastic ac power flow via polynomial chaos expansion. In *2016 IEEE Conference on Control Applications (CCA)*, pages 70–76. IEEE.

[Mühlpfordt et al., 2019] Mühlpfordt, T., Roald, L., Hagenmeyer, V., Faulwasser, T., and Misra, S. (2019). Chance-constrained ac optimal power flow: A polynomial chaos approach. *IEEE Transactions on Power Systems*, 34(6):4806–4816.

[Murray-Smith et al., 2003] Murray-Smith, R., Sbarbaro, D., Rasmussen, C. E., and Girard, A. (2003). Adaptive, cautious, predictive control with gaussian process priors. *IFAC Proceedings Volumes*, 36(16):1155–1160.

[Ng et al., 2018] Ng, Y., Misra, S., Roald, L. A., and Backhaus, S. (2018). Statistical learning for dc optimal power flow. In *2018 Power Systems Computation Conference (PSCC)*, pages 1–7. IEEE.

[Nocedal and Wright, 2006] Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. New York: Springer.

[ONeill et al., 2011] ONeill, R., Dautel, T., and Krall, E. (2011). Recent iso software enhancements and future software and modeling plans. federal energy regulatory commission, washington. Technical report, DC, Tech. Rep.

[Owen et al., 2019] Owen, A. B., Maximov, Y., and Chertkov, M. (2019). Importance sampling the union of rare events with an application to power systems analysis. *Electronic Journal of Statistics*, 13(1):231–254.

[Pan et al., 2020] Pan, X., Zhao, T., Chen, M., and Zhang, S. (2020). Deepopf: A deep neural network approach for security-constrained dc optimal power flow. *IEEE Transactions on Power Systems*, 36(3):1725–1735.

[Pareek and Nguyen, 2020] Pareek, P. and Nguyen, H. D. (2020). Gaussian process learning-based probabilistic optimal power flow. *IEEE Transactions on Power Systems*, 36(1):541–544.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

[Pena-Ordieres et al., 2020] Pena-Ordieres, A., Molzahn, D. K., Roald, L. A., and Wächter, A. (2020). Dc optimal power flow with joint chance constraints. *IEEE Transactions on Power Systems*, 36(1):147–158.

[Quinonero-Candela and Rasmussen, 2005] Quinonero-Candela, J. and Rasmussen, C. E. (2005). A unifying view of sparse approximate gaussian process regression. *The Journal of Machine Learning Research*, 6:1939–1959.

[Rasmussen, 2003] Rasmussen, C. E. (2003). Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer.

[Rasmussen et al., 2006] Rasmussen, C. E., Williams, C. K., et al. (2006). *Gaussian processes for machine learning*, volume 1. Springer.

[Roald and Andersson, 2017] Roald, L. and Andersson, G. (2017). Chance-constrained ac optimal power flow: Reformulations and efficient algorithms. *IEEE Transactions on Power Systems*, 33(3):2906–2918.

[Roald et al., 2016] Roald, L., Misra, S., Krause, T., and Andersson, G. (2016). Corrective control to handle forecast uncertainty: A chance constrained optimal power flow. *IEEE Transactions on Power Systems*, 32(2):1626–1637.

[Roald et al., 2015] Roald, L., Oldewurtel, F., Van Parys, B., and Andersson, G. (2015). Security constrained optimal power flow with distributionally robust chance constraints. *arXiv preprint arXiv:1508.06061*.

[Roald et al., 2017] Roald, L. A., Molzahn, D. K., and Tobler, A. F. (2017). Power system optimization with uncertainty and ac power flow: Analysis of an iterative algorithm. In *10th IREP Symp. Bulk Power Syst. Dynamics Control*.

[Schellenberg et al., 2005] Schellenberg, A., Rosehart, W., and Aguado, J. (2005). Cumulant-based probabilistic optimal power flow (p-opf) with gaussian and gamma distributions. *IEEE Transactions on Power Systems*, 20(2):773–781.

[Schmidli et al., 2016] Schmidli, J., Roald, L., Chatzivasileiadis, S., and Andersson, G. (2016). Stochastic ac optimal power flow with approximate chance-constraints. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pages 1–5. IEEE.

[Schulz et al., 2018] Schulz, E., Speekenbrink, M., and Krause, A. (2018). A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, 85:1–16.

[Seeger et al., 2003] Seeger, M. W., Williams, C. K., and Lawrence, N. D. (2003). Fast forward selection to speed up sparse gaussian process regression. In *International Workshop on Artificial Intelligence and Statistics*, pages 254–261. PMLR.

[Sidana et al., 2017] Sidana, S., Trofimov, M., Horodnitskii, O., Laclau, C., Maximov, Y., and Amini, M.-R. (2017). Representation learning and pairwise ranking for implicit feedback in recommendation systems. *arXiv preprint arXiv:1705.00105*.

[Sidana et al., 2021] Sidana, S., Trofimov, M., Horodnytskyi, O., Laclau, C., Maximov, Y., and Amini, M.-R. (2021). User preference and embedding learning with implicit feedback for recommender systems. *Data Mining and Knowledge Discovery*, 35:568–592.

[Singh et al., 2021] Singh, M. K., Kekatos, V., and Giannakis, G. B. (2021). Learning to solve the AC-OPF using sensitivity-informed deep neural networks. *IEEE Transactions on Power Systems*.

[Sjödin et al., 2012] Sjödin, E., Gayme, D. F., and Topcu, U. (2012). Risk-mitigated optimal power flow for wind powered grids. In *2012 American Control Conference (ACC)*, pages 4431–4437. IEEE.

[Smola and Bartlett, 2001] Smola, A. and Bartlett, P. (2001). Sparse greedy gaussian process approximation. *Advances in Neural Information Processing Systems*, 13.

[Stulov et al., 2019] Stulov, N., Sobajic, D. J., Maximov, Y., Deka, D., and Chertkov, M. (2019). Learning a generator model from terminal bus data. *arXiv preprint arXiv:1901.00781*.

[Stulov et al., 2020] Stulov, N., Sobajic, D. J., Maximov, Y., Deka, D., and Chertkov, M. (2020). Learning model of generator from terminal data. *Electric Power Systems Research*, 189:106742.

[Thurner et al., 2018] Thurner, L., Scheidler, A., Schäfer, F., Menke, J.-H., Dollichon, J., Meier, F., Meinecke, S., and Braun, M. (2018). pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems. *IEEE Transactions on Power Systems*, 33(6):6510–6521.

[Titsias, 2009] Titsias, M. (2009). Variational learning of inducing variables in sparse gaussian processes. In *Artificial intelligence and statistics*, pages 567–574. PMLR.

[Ullah et al., 2022] Ullah, Z., Wang, S., Wu, G., Hasanien, H. M., Jabbar, M. W., Qazi, H. S., Tostado-Véliz, M., Turky, R. A., and Elkadeem, M. R. (2022). Advanced studies for probabilistic optimal power flow in active distribution networks: A scientometric review. *IET Generation, Transmission & Distribution*, 16(18):3579–3604.

[Usama et al., 2019] Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K.-L. A., Elkhatib, Y., Hussain, A., and Al-Fuqaha, A. (2019). Unsupervised machine learning for networking: Techniques, applications and research challenges. *IEEE access*, 7:65579–65615.

[Van Engelen and Hoos, 2020] Van Engelen, J. E. and Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine learning*, 109(2):373–440.

[Vapnik, 1999] Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.

[Venzke et al., 2020] Venzke, A., Halilbašić, L., Barré, A., Roald, L., and Chatzivasileiadis, S. (2020). Chance-constrained ac optimal power flow integrating hvdc lines and controllability. *International Journal of Electrical Power & Energy Systems*, 116:105522.

[Venzke et al., 2017] Venzke, A., Halilbasic, L., Markovic, U., Hug, G., and Chatzivasileiadis, S. (2017). Convex relaxations of chance constrained ac optimal power flow. *IEEE Transactions on power systems*, 33(3):2829–2841.

[Vinoth and Datta, 2022] Vinoth, A. and Datta, S. (2022). Fundamentals of machine learning. *Machine Learning in Industry*, pages 1–27.

[Vrakopoulou et al., 2013a] Vrakopoulou, M., Katsampani, M., Margellos, K., Lygeros, J., and Andersson, G. (2013a). Probabilistic security-constrained ac optimal power flow. In *2013 IEEE Grenoble Conference*, pages 1–6. IEEE.

[Vrakopoulou et al., 2013b] Vrakopoulou, M., Katsampani, M., Margellos, K., Lygeros, J., and Andersson, G. (2013b). Probabilistic security-constrained ac optimal power flow. *PowerTech*, 28(4):2013.

[Vrakopoulou et al., 2013c] Vrakopoulou, M., Margellos, K., Lygeros, J., and Andersson, G. (2013c). Probabilistic framework for reserve scheduling and n-1 security assessment of systems with high wind power penetration. *IEEE Transactions on Power Systems*, 28(4):3885–3896.

[Vrakopoulou et al., 2013d] Vrakopoulou, M., Margellos, K., Lygeros, J., and Andersson, G. (2013d). Probabilistic guarantees for the n-1 security of systems with wind power generation. *Reliability and risk evaluation of wind integrated power systems*, pages 59–73.

[Wächter and Biegler, 2006] Wächter, A. and Biegler, L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1):25–57.

[Warrington et al., 2013] Warrington, J., Goulart, P., Mariéthoz, S., and Morari, M. (2013). Policy-based reserves for power systems. *IEEE Transactions on Power Systems*, 28(4):4427–4437.

[Weiss et al., 2016] Weiss, K., Khoshgoftaar, T. M., and Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, 3(1):1–40.

[Wu et al., 2019] Wu, J., Zhang, B., Jiang, Y., Bie, P., and Li, H. (2019). Chance-constrained stochastic congestion management of power systems considering uncertainty of wind power and demand side response. *International Journal of Electrical Power & Energy Systems*, 107:703–714.

[Wächter and Biegler, 2006] Wächter, A. and Biegler, L. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program. 106*, pp.25.

[Xiao et al., 2001] Xiao, Y., Song, Y., and Sun, Y. (2001). A hybrid stochastic approach to available transfer capability evaluation. *IEE Proceedings-Generation, Transmission and Distribution*, 148(5):420–426.

[Xie and Ahmed, 2017] Xie, W. and Ahmed, S. (2017). Distributionally robust chance constrained optimal power flow with renewables: A conic reformulation. *IEEE Transactions on Power Systems*, 33(2):1860–1867.

[Xu et al., 2020] Xu, Y., Hu, Z., Mili, L., Korkali, M., and Chen, X. (2020). Probabilistic power flow based on a gaussian process emulator. *IEEE Transactions on Power Systems*, 35(4):3278–3281.

[Zhang and Li, 2010] Zhang, H. and Li, P. (2010). Probabilistic analysis for optimal power flow under uncertainty. *IET Generation, Transmission & Distribution*, 4(5):553–561.

[Zhang and Li, 2011] Zhang, H. and Li, P. (2011). Chance constrained programming for optimal power flow under uncertainty. *IEEE Transactions on Power Systems*, 26(4):2417–2424.

[Zhou, 2021] Zhou, Z.-H. (2021). *Machine learning*. Springer Nature.