

Data-Driven and Privacy-Preserving Frameworks for Enhanced Interoperability in Modern Power Systems

Zur Erlangung des akademischen Grades eines

**DOKTORS DER INGENIEURWISSENSCHAFTEN
(Dr.-Ing.)**

von der KIT-Fakultät für
Elektrotechnik und Informationstechnik
des Karlsruher Instituts für Technologie (KIT)

angenommene

DISSERTATION

von

M.Sc. Burak Dindar
geb. in Istanbul, Türkei

Tag der mündlichen Prüfung:
Hauptreferent:
Korreferent:

12.10.2026
Prof. Dr. Veit Hagenmeyer
Prof. Dr.-Ing. Marc Hiller
Prof. Dr. Antonello Monti

Acknowledgment

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Dr. Veit Hagenmeyer, for his invaluable mentorship, continuous support, and thoughtful guidance throughout my Ph.D. journey.

I would also like to extend my sincere thanks to Dr.-Ing. Hüseyin Kemâl Çakmak, whose endless motivation, invaluable support, and countless inspiring conversations have greatly enriched my experience at KIT and made challenging tasks much easier.

I would furthermore like to thank Prof. Dr.-Ing. Marc Hiller and Prof. Dr. Antonello Monti for serving as reviewers of this dissertation and for dedicating their time and expertise to the evaluation of my work.

I am also deeply grateful to my team, ESA, for the opportunity to work together in a harmonious and enjoyable multicultural environment, and for their friendship and support.

A special thanks goes to Dr. Can Berk Saner for his numerous contributions to my Ph.D. journey, our valuable collaborations, and the many scientific discussions we shared during countless meetings, despite the significant time difference.

I am deeply thankful to my family, whose endless support and belief have made this journey possible, whether we were living under the same roof or separated by miles.

Finally, and most importantly, I would like to express my deepest gratitude to my wife, Çağla Dindar, whose contributions to this thesis go far beyond my own. Without her constant strength in lifting me up whenever I stumbled and standing by me through every difficult moment, this work would not have been possible.

Karlsruhe, Germany, October 2026

Burak Dindar

Abstract

With the ongoing transformation of modern power systems, driven by the rapid proliferation of distributed generators (DGs), electric vehicles (EVs), photovoltaic (PV) systems, heat pumps, and other flexibility-providing units (FPUs), the grid is evolving into a more complex and dynamic structure. This transformation, coupled with bidirectional power flows and the inherent uncertainties associated with distributed resources, presents significant challenges in maintaining system balance, ensuring secure operation, and coordinating among multiple actors such as market operators, transmission system operators (TSOs), and distribution system operators (DSOs). Addressing these challenges requires effective interoperability among all stakeholders to enable reliable and economically efficient operation. However, achieving such interoperability necessitates extensive data exchange between parties, which is often constrained by data privacy concerns. Therefore, developing new frameworks that enhance interoperability while preserving data confidentiality has become essential for the operation of modern power systems. In response to these challenges, this thesis proposes data-driven and privacy-preserving frameworks designed to strengthen interoperability between market operators and TSOs, as well as between TSOs and DSOs. To this end, the thesis introduces a set of complementary methodologies, each addressing a distinct interoperability challenge within different layers of the power system.

First, a multi-agent system (MAS)-based AC optimal power flow (AC-OPF) model empowered by machine learning (ML) is proposed to enhance interoperability between market operators and TSOs. In this approach, the TSO develops neural network (NN) models and valid inequalities to approximate the feasible space of the AC-OPF problem using only active power. Importantly, these models mask sensitive system information before being transmitted to the market operator. The market operator agent then integrates the received NN models and valid inequalities into a mixed-integer linear programming (MILP) problem. This MILP problem can be solved under various market-based objective functions and constraints while still respecting power system limits. As a result, private market-related data remain confidential and are not shared with the TSO. Hence, the proposed framework enhances interoperability by ensuring mutual data privacy between both entities.

Second, the interaction between the TSOs and DSOs is addressed using the feasible operating region (FOR) approach, which serves as a privacy-preserving coordination framework. The FOR

captures all feasible combinations of active and reactive power at the point of common coupling (PCC) while respecting DS-level constraints such as voltage limits, thermal limits, and FPU operating ranges. By sharing this compact PQ representation instead of sensitive information, such as network topology, line parameters, or customer data, TSOs and DSOs can coordinate effectively without compromising privacy. Although the FOR concept has been widely explored in the literature, its practical implementation within TSO operations remains challenging due to three key factors: (i) accurately determining the FOR in large-scale, meshed DS networks; (ii) obtaining a tractable analytical representation; and (iii) quantifying its economic value. To address these challenges, this thesis proposes a novel AC-OPF-based method for constructing a three-dimensional PQV-FOR that explicitly accounts for voltage variability and diverse FPU characteristics. An implicit polynomial fitting technique is then employed to obtain an analytical representation of the FOR. Furthermore, a quadratic cost function defined over the PQV domain is introduced to monetize the FOR. Collectively, these developments enable single-round TSO-DSO coordination that ensures data privacy while enhancing interoperability.

Lastly, the interaction between TSOs and DSOs is addressed from a different perspective. While the FOR approach provides an effective foundation for privacy-preserving coordination, it is often constrained by its reliance on a single PCC, as the dispatch of DGs is typically determined through the PCC interface. To overcome this limitation, this thesis introduces an ML-based framework in which the technical constraints of DSs are represented through ML models trained exclusively on non-sensitive data. Using these models, the TSO can solve the OPF problem and directly determine the dispatch of FPUs in a single round of communication, even in systems with multiple PCCs. Within this framework, a novel NN architecture is developed to accurately and efficiently represent the feasible operating region of DSs, ensuring both computational efficiency and reliability. By encapsulating DS technical constraints within ML models, the exchange of sensitive information between TSOs and DSOs is effectively eliminated. Consequently, the proposed approach enables the exploitation of DS flexibility for system operation without compromising data privacy, thereby further enhancing interoperability among stakeholders.

Overall, the frameworks developed in this thesis collectively contribute to advancing the secure, efficient, and privacy-aware operation of modern power systems. By integrating data-driven and optimization techniques within privacy-preserving architectures, the proposed approaches enable seamless interoperability among market operators, TSOs, and DSOs without compromising data confidentiality.

Kurzfassung

Mit der fortschreitenden Transformation moderner Energiesysteme-vorangetrieben durch die rasche Verbreitung von dezentralen Erzeugungsanlagen (DGs), und weiteren Flexibilitätsbereitstellungseinheiten (FPU), entwickelt sich das Netz zu einer zunehmend komplexen und dynamischen Struktur. Diese Transformation stellt erhebliche Herausforderungen in Bezug auf die Sicherstellung des Leistungsgleichgewichts im System, den sicheren Netzbetrieb sowie die Koordination zwischen verschiedenen Akteuren, wie Marktbetreibern, Übertragungs- (TSOs) und Verteilnetzbetreibern (DSOs), dar. Zur Bewältigung dieser Herausforderungen ist eine effektive Interoperabilität zwischen allen Stakeholdern erforderlich, um einen zuverlässigen und wirtschaftlich effizienten Betrieb zu ermöglichen. Die Realisierung einer solchen Interoperabilität setzt jedoch einen umfangreichen Datenaustausch zwischen den Parteien voraus, der häufig durch Datenschutzbedenken eingeschränkt wird. Daher ist die Entwicklung neuer Konzepte, die die Interoperabilität erhöhen und gleichzeitig die Vertraulichkeit der Daten gewährleisten, für den Betrieb moderner Energiesysteme von entscheidender Bedeutung geworden. Als Antwort auf diese Herausforderungen schlägt die vorliegende Dissertation datengetriebene und datenschutzwahrende Frameworks vor, die darauf ausgelegt sind, die Interoperabilität sowohl zwischen Marktbetreibern und TSOs als auch zwischen TSOs und DSOs zu stärken. Zu diesem Zweck führt die Dissertation eine Reihe komplementärer Methoden ein, die jeweils eine spezifische Interoperabilitäts herausforderung in unterschiedlichen Ebenen des Energiesystems adressieren.

Zunächst wird ein Multi-Agenten-System (MAS) basierendes AC-Optimal-Power-Flow-(AC-OPF)-Modell, das durch maschinelles Lernen (ML) erweitert wird, vorgeschlagen, um die Interoperabilität zwischen Marktbetreibern und TSOs zu verbessern. Hierbei entwickelt der TSO neuronale Netzwerke (NNs) sowie gültige Ungleichungen, um den zulässigen Lösungsraum des AC-OPF-Problems ausschließlich basierend auf Wirkleistung zu approximieren. Entscheidend ist, dass diese Modelle sensible Systeminformationen verschleiern, bevor sie an den Marktbetreiber übermittelt werden. Der Marktbetreiber integriert anschließend die empfangenen NN-Modelle und gültigen Ungleichungen in eine gemischt-ganzzahlige lineare Programmierung (MILP). Dieses MILP kann unter verschiedenen marktbasierenden Zielfunktionen und Nebenbedingungen gelöst werden, wobei die Netzgrenzen weiterhin berücksichtigt werden. Somit bleiben private marktbezogene Daten vertraulich und werden nicht mit dem TSO geteilt. Das vorgeschlagene Framework verbessert die Interoperabilität, indem es den gegenseitigen Datenschutz sicherstellt.

Im zweiten Schritt wird die Interaktion zwischen TSOs und DSOs mithilfe des sogenannten Feasible Operating Region (FOR)-Ansatzes adressiert, der als ein datenschutzwahrendes Koordinationsverfahren dient. Die FOR umfasst sämtliche zulässigen Kombinationen von Wirk- und Blindleistung am Netzverknüpfungspunkt (PCC) unter Einhaltung verteilenebene-spezifischer Randbedingungen. Durch die Bereitstellung dieser kompakten PQ-Darstellung statt sensibler Informationen wie Netzstruktur, Leitungsparametern oder Kundendaten können TSOs und DSOs effektiv kooperieren, ohne den Datenschutz zu gefährden. Obwohl der FOR-Ansatz in der Literatur umfassend untersucht wurde, bleibt seine praktische Umsetzung im TSO-Betrieb aufgrund dreier wesentlicher Faktoren herausfordernd: (i) die genaue Bestimmung der FOR in großen, vermaschten Verteilnetzen, (ii) die Gewinnung einer handhabbaren analytischen Darstellung sowie (iii) die Quantifizierung ihres wirtschaftlichen Kosten. Zur Bewältigung dieser Herausforderungen wird in dieser Dissertation eine neue AC-OPF-basierte Methode zur Konstruktion einer dreidimensionalen PQV-FOR vorgeschlagen, welche Spannungsvariabilitäten sowie unterschiedliche FPU-Eigenschaften berücksichtigt. Eine implizite Polynomapproximationstechnik wird anschließend eingesetzt, um eine analytische Darstellung der FOR zu erhalten. Außerdem wird eine quadratische Kostenfunktion über dem PQV-Raum eingeführt, um die FOR zu monetarisieren. Zusammengefasst ermöglichen diese Entwicklungen eine einrundige TSO-DSO-Koordination, die den Datenschutz gewährleistet und zugleich die Interoperabilität verbessert.

Abschließend wird die Interaktion zwischen TSOs und DSOs aus einer weiteren Perspektive betrachtet. Obwohl der FOR-Ansatz eine effektive Grundlage für eine datenschutzwahrende Koordination darstellt, ist er häufig durch seine Beschränkung auf einen einzigen PCC limitiert, da die Einsatzplanung dezentraler Erzeugungsanlagen über diesen PCC erfolgt. Um diese Einschränkung zu überwinden, wird ein ML-basiertes Framework vorgestellt, in dem die technischen Restriktionen von Verteilnetzen mittels ML-Modellen abgebildet werden, die ausschließlich auf nicht-sensiblen Daten trainiert werden. Mithilfe dieser Modelle kann der TSO das OPF-Problem lösen und den Einsatz von FPUs direkt in einer einzigen Kommunikationsrunde bestimmen, selbst in Systemen mit mehreren PCCs. Innerhalb dieses Frameworks wird eine neuartige NN-Architektur entwickelt, die die zulässige Betriebsregion von Verteilnetzen präzise und effizient abbildet und dabei sowohl Rechenaufwand als auch Zuverlässigkeit berücksichtigt. Durch ML-Modellen entfällt vollständig der Austausch sensibler Informationen zwischen TSOs und DSOs. Das vorgestellte Verfahren ermöglicht somit die Nutzung der Flexibilität von Verteilnetzen für den Systembetrieb, ohne den Datenschutz zu gefährden, und verbessert die Interoperabilität.

Insgesamt leisten die entwickelten Frameworks einen wesentlichen Beitrag zur sicheren, effizienten und datenschutzbewussten Betriebsführung moderner Energiesysteme. Durch die Integration datengetriebener und optimierungsbasierter Verfahren in datenschutzwahrende Architekturen ermöglichen die vorgeschlagenen Ansätze eine Interoperabilität zwischen Marktbetreibern, TSOs und DSOs, ohne die Vertraulichkeit der Daten zu beeinträchtigen.

Contents

Acknowledgment	i
Abstract	iii
Kurzfassung	v
Acronyms	xi
1 Introduction	1
1.1 Motivation	1
1.2 Challenges and Research Questions	2
1.3 Contributions and Outline	4
2 Background and Preliminaries	9
2.1 Design and Operation of the Modern Power Systems	9
2.2 Increasing Need for Interoperability	13
2.3 Data Privacy Aspects	19
2.4 Data-Driven Approaches in Power Systems	24
2.5 Conventions and Notations	27
3 Privacy-Preserving Cross-Entity AC-OPF Between Market Operators and TSOs	29
3.1 State of the Art	30
3.2 Structure of the Proposed Multi-Agent System	33
3.3 Overview of the Proposed Methodology	34
3.3.1 Formulation of the Standard AC-OPF	35
3.3.2 Formulation of the Machine Learning-Driven Multi-Agent-Based AC-OPF	35
3.4 Valid Inequalities and Dataset Creation	37
3.4.1 Generating Valid Inequalities	38
3.4.2 N-ball Approach	40
3.4.3 Multivariate Gaussian Distribution	42
3.4.4 Perturbation Approach	43

3.5	Neural Network Models and Their Transformation to MILP	44
3.5.1	Training of the Classifier Model	45
3.5.2	Training of the Regressor Model	45
3.5.3	Neural Network MILP Representation	46
3.6	Case Studies and Discussion	47
3.6.1	Dataset Creation Process	48
3.6.2	ML Training and Approximation Quality	49
3.6.3	Benchmark Results with Different Cost Coefficients	50
3.6.4	Benchmark Results with Different Cost Coefficients and Constraints	51
3.7	Summary	53
4	Privacy-Preserving Analytical Representation and Monetization of DS Flexibility Using the Feasible Operating Region Approach for TSO-DSO Interoperability	55
4.1	State of the Art	57
4.2	Overview of the Proposed Methodology	59
4.2.1	Standard AC-OPF for Integrated Transmission-Distribution System	60
4.2.2	FOR-Based AC-OPF for Privacy-Preserving DS Flexibility Utilization	61
4.3	Sampling Strategies for Flexibility Modeling	63
4.3.1	Boundary Sampling for the PQV-FOR	64
4.3.2	Interior Sampling for the Cost of the FOR	67
4.4	Construction of the Analytical Functions	68
4.4.1	Implicit Polynomial Representation of the PQV-FOR	68
4.4.2	Quadratic Representation of the Cost of the FOR	70
4.4.3	Theoretical Justification for the Analytical Representations	71
4.4.4	Remarks	72
4.5	Case Studies and Discussion	73
4.5.1	Distribution Systems Specifications	73
4.5.2	Derivation of Analytical FOR and Cost Functions for DSs	74
4.5.3	Computation Time Analysis of the Proposed Sampling Strategy	77
4.5.4	Analysis of the Impact of Polynomial Degree	79
4.5.5	Incorporation of Analytical DS Representations into TS and Benchmarking Against AC-OPF	80
4.5.6	Analysis of the Impact of the Varying Voltage at the PCCs	82
4.5.7	Benchmark Against an Iterative Communication Scheme	83
4.6	Summary	85
5	A Novel TSO-DSO Framework for Multi-PCC Operation and Direct Utilization of DS Flexibility	87
5.1	State of the Art	88

5.2 Overview of the Proposed Methodology	91
5.2.1 Formulation of the Standard AC-OPF	91
5.2.2 Formulation of the ML-Based Privacy-Preserving AC-OPF	93
5.3 Dataset Creation	95
5.4 Representation Technical Constraints of the Distribution Systems with Machine Learning Models	98
5.4.1 NN-Guided Polytope Representation of Feasible Region	99
5.4.2 Quadratic Regression-Based Power Flow Approximator	102
5.5 Case Studies and Discussion	102
5.5.1 Power System Creation	103
5.5.2 Dataset Generation, Training, and Approximation Quality in ML Models	103
5.5.3 Analysis of the Impact of the Conservativeness Parameter	107
5.5.4 Analysis of the Impact of the Varying Voltage at the PCCs	110
5.5.5 Computational Time Performance of the Proposed Method	111
5.5.6 Scalability Analysis of the Proposed Method	112
5.6 Summary	114
6 Conclusion and Future Work	115
List of Figures	119
List of Tables	121
References	123
Publications	139
Journal Articles	139
Conference Proceedings	140

Acronyms

AC	Alternating Current
BBPS	Bounding Box Projection Sampling
DC	Direct Current
DER	Distributed Energy Resource
DG	Distributed Generator
DP	Differential Privacy
DSO	Distribution System Operator
EV	Electric Vehicle
FDS	Fibonacci Direction Sampling
FOR	Feasible Operating Region
FPU	Flexibility-Providing Unit
LHS	Latin Hypercube Sampling
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MAS	Multi-Agent System
MGD	Multivariate Gaussian Distribution
MILP	Mixed-Integer Linear Programming
ML	Machine Learning
MOA	Market Operator Agent
NN	Neural Network
OPF	Optimal Power Flow
PCC	Point of Common Coupling
PV	Photovoltaic
RMSE	Root Mean Square Error
RS	Random Sampling
SOC	Second-Order Cone Programming
TOA	Technical Operator Agent
TSO	Transmission System Operator

1 Introduction

The electric power system is undergoing a fundamental transformation. The traditional power system, characterized by unidirectional energy flows from large synchronous generators through transmission and distribution networks to passive loads, is rapidly being replaced by a highly distributed, actively managed system. New components, predominantly connected to the distribution network, such as distributed generators (DGs), photovoltaics (PV), electric vehicles (EVs), heat pumps and other flexibility-providing units (FPU), introduce bidirectional power flows, increased uncertainty, and greater operational complexity. Simultaneously, electricity markets are becoming more granular, demanding tighter coordination among institutional actors, including market operators, transmission system operators (TSOs), and distribution system operators (DSOs).

This transformation introduces new challenges to ensure reliable and efficient system operation. Among these, interoperability has emerged as a key requirement. Enhancing interoperability goes beyond establishing standardized data exchange formats or regulatory frameworks and it also necessitates mechanisms that preserve the confidentiality of system participants. The present thesis is situated within this evolving context, and aims to develop frameworks that enhance interoperability among power system stakeholders while safeguarding data privacy.

Further in this chapter, Section 1.1 presents the motivation behind the study, Section 1.2 discusses the key challenges and formulates the research questions, and Section 1.3 details the main contributions and outline the thesis.

1.1 Motivation

Considering the accelerating pace of global warming and the commitment of many countries to carbon-neutrality targets, the power system emerges as a key element in achieving these goals. The ongoing large-scale transformation of power systems must therefore be completed efficiently, and the challenges arising in this process must be addressed promptly. One of the main motivations behind this thesis is to contribute to this global transformation of the power system.

A central aspect of this transformation is the increasing variability caused by DGs and the growing number of controllable FPU. These elements introduce rapid fluctuations in the grid, making

it essential to accurately model, manage, and coordinate their effects. Consequently, system stakeholders such as market operators, TSOs and DSOs, must collaborate far more closely than in traditional, centrally operated power systems. Numerous mechanisms and initiatives are being developed worldwide to facilitate such coordination, and another key motivation of this research is to support and enhance these processes through more effective technical frameworks.

An equally important consideration in this context is data privacy. As power systems become increasingly digitalized and interconnected, stakeholders are required to exchange a wide range of data, including grid topology, customer load profiles, and operational constraints. However, the unrestricted sharing of such data across institutional boundaries may be legally prohibited, commercially sensitive, or vulnerable to cybersecurity threats. Thus, each stakeholder must legitimately protect their data for reasons ranging from commercial confidentiality to consumer privacy. Effective interoperability solutions must therefore be designed to enable reliable coordination without compromising data privacy. This challenge constitutes one of the core motivations of this thesis.

At the same time, ensuring privacy should not come at the cost of system scalability or computational efficiency. Power systems are becoming increasingly complex, with a rapidly rising number of active components. As a result, standard problems such as AC optimal power flow (AC-OPF) become computationally expensive and challenging to scale. Developing methods that balance privacy preservation, scalability, and computational efficiency is thus another central motivation of this thesis.

Overall, the fundamental motivation of this thesis is to enhance interoperability among power system stakeholders while preserving data confidentiality and computational tractability. By doing so, this research aims to contribute to the development of a future-ready power system that is sustainable, efficient, and securely coordinated across all operational layers.

1.2 Challenges and Research Questions

Enhancing interoperability in power systems while maintaining data privacy requires the development of innovative coordination frameworks. However, designing such frameworks involves overcoming several fundamental challenges related to information exchange, model fidelity, and computational tractability [1, 2, 3].

For instance, consider the interoperability between market operators and TSOs. In practice, market operators are often required to solve optimization problems, such as those based on the AC-OPF or other formulations that incorporate specific market constraints and objectives, to determine efficient market clearing and dispatch decisions. With the increasing volatility and dynamism

introduced by DGs and FPU, such optimization tasks have become even more frequent and complex [4, 5].

However, enabling cross-entity optimization typically necessitates sharing detailed models, parameters, or measurements that can reveal sensitive information, such as network topology, line parameters, or customer load profiles [6, 7]. To mitigate these risks, TSOs often share only simplified network equivalents consisting of a few representative nodes, rather than the full transmission grid model, with market operators [8]. Conversely, market operators may possess private contractual or pricing information, for example, bilateral agreements or market bids that cannot be disclosed to the TSOs for confidentiality or regulatory reasons. This mutual restriction on data exchange highlights a fundamental challenge in modern power system coordination: ensuring efficient joint operation and decision-making without compromising the privacy of any stakeholder [9]. Accordingly, this leads to the first research question: ***How can cross-entity optimization between market operators and TSOs be achieved effectively without requiring the exchange of sensitive data?***

Similarly, when considering interoperability between TSOs and DSOs, the ongoing transformation of the power system necessitates the development of new coordination frameworks that account for the evolving role of the DS [10, 11]. As the penetration of DGs and other FPUs continues to grow, the flexibility offered by DSs is becoming increasingly valuable [12, 13]. This flexibility has significant potential to enhance network management efficiency at both transmission and distribution levels [14].

A key challenge, however, lies in determining how to utilize DS flexibility effectively while preserving data privacy among stakeholders. In this context, one of the most prominent approaches proposed in the literature is the feasible operating region (FOR) method [15, 16, 17]. In the FOR-based approach, the flexibility of each DS is aggregated at the interface between the TSO and DSO, known as the point of common coupling (PCC). Through this aggregation, sensitive DS information, such as detailed network topology or load data, remains local to the DSO and is not disclosed to the TSO. Instead, only the aggregated flexibility representation is shared, thereby ensuring privacy while enabling coordination.

Despite numerous studies proposing various techniques for computing the FOR [18, 19], a fundamental gap remains in understanding how the FOR can be effectively utilized by TSOs for system-level network management. To achieve this, several key issues must be addressed in an integrated manner: (i) the efficient computation of FORs for large-scale and meshed distribution networks, (ii) their analytical representation in a tractable mathematical form, and (iii) their economic valuation (monetization) to enable market-based coordination between TSOs and DSOs. Accordingly, this leads to the following second research question: ***How can the FOR***

of large-scale, meshed DSs be efficiently computed, analytically represented, and economically valued to enable the utilization of DS flexibility while preserving data privacy?

Although the FOR approach offers a promising pathway toward privacy-preserving coordination between TSOs and DSOs, it also presents certain limitations that need to be addressed to ensure its applicability to modern, large-scale power systems. Most existing FOR-based frameworks are designed under the assumption of a single PCC between the TS and DS [20, 21]. However, in today's increasingly complex and interconnected grid, DSs are typically connected to the TS through multiple PCCs [22, 23]. Neglecting this multi-PCC structure can lead to suboptimal coordination, as the interactions among these coupling points significantly influence system behavior and the exchange of flexibility.

Moreover, in conventional FOR-based implementations, the dispatch decisions for DGs and other FPU's are determined indirectly through the operating point selected at the PCC [24]. This hierarchical process, while effective for maintaining privacy, can introduce unnecessary computational complexity and coordination delays. Enabling direct dispatch of DGs and FPU's, without relying solely on PCC-level decision-making, could greatly simplify the coordination mechanism.

Therefore, to overcome these critical limitations, it is essential to develop a new framework that allows for the utilization of DS flexibility while explicitly accounting for multiple PCCs and enabling direct dispatch of distributed resources. This motivation gives rise to the third research question: *How can the utilization of DS flexibility be realized by considering multiple PCCs and enabling the direct dispatch of DGs and FPU's by TSOs?*

Finally, while addressing these three research questions, a fundamental design principle throughout this thesis is to ensure computational efficiency and scalability. Any proposed method or framework must not only preserve data privacy and increase interoperability but also remain tractable and practical for large-scale system implementations. Accordingly, this thesis systematically investigates and provides solutions to these three research questions.

1.3 Contributions and Outline

This thesis addresses, in detail, the three research questions introduced in the previous chapter. For each question, novel and tailored frameworks are developed with the overarching goal of enhancing interoperability in modern power systems through privacy-preserving and computationally efficient mechanisms. Throughout this research, the principle of *data privacy by design* is adopted, ensuring that sensitive information remains protected at all stages of coordination and optimization among system stakeholders.

To address the first research question, a new privacy-preserving cross-entity optimization framework is developed to facilitate coordination between market operators and TSOs [25]. Unlike conventional approaches, the proposed framework eliminates the need for sensitive data exchange by ensuring that each stakeholder retains its private information locally. Only the active power outputs of generators that are already known to both entities are shared. This minimal data exchange enables the market operator to perform economically optimal decisions while the TSO maintains network feasibility. Consequently, the framework achieves privacy preservation by design and significantly enhances interoperability between the two entities.

For the second research question, this thesis develops an innovative FOR-based framework that enables DSs to communicate their flexibility to the TSO without disclosing sensitive data [26]. The proposed method constructs the FOR using a novel AC-OPF-based approach capable of accurately capturing the behavior of complex and meshed DSs. Moreover, the FOR and its associated cost are represented analytically, ensuring both precision and tractability. This analytical representation allows for effective modeling of the DS feasible region using only non-sensitive data. Interoperability between the TSO and DSO is established through the PCC, where both entities already share access to measurable quantities such as active and reactive power and voltage magnitude. Consequently, the framework enables efficient coordination and reliable system operation while inherently preserving data privacy.

Lastly, to address the third research question, this thesis develops a multi-PCC direct dispatch framework [27, 28, 29]. Classical FOR-based coordination methods are typically constrained to single-PCC systems and rely on indirect coordination through aggregated flexibility representations. In contrast, the proposed framework models the internal behavior of DSs using ML models trained exclusively on non-sensitive data. These models accurately capture the DS operational responses and are securely shared with the TSO, enabling direct dispatch of DGs and FPU's without revealing any confidential network or customer information. This approach allows multi-PCC coordination, simplifies the decision-making process, and improves system-wide efficiency while maintaining strict privacy guarantees.

In summary, this thesis introduces a suite of integrated, privacy-preserving, and computationally efficient frameworks that collectively enhance interoperability across different layers of the modern power system. Following this overview, the scope and contributions of the subsequent chapters of this thesis are detailed as follows:

Chapter 2 provides the foundational background required to support the privacy-preserving and interoperability-oriented frameworks developed in the remainder of the thesis. The chapter first outlines the design and operation of modern power systems, emphasizing the transition from traditional, centrally managed, unidirectional grids toward decentralized, renewable-integrated architectures characterized by increased variability, bidirectional power flows. Building on this,

the chapter highlights the escalating need for interoperability among TSOs, DSOs, market operators, aggregators, and prosumers, showing how insufficient coordination contributes to rising balancing costs, renewable curtailments, and operational inefficiencies. The chapter then examines the data privacy challenges that arise from intensified data exchange, identifying the sensitive information held by each stakeholder and explains how confidentiality concerns create a fundamental privacy-coordination dilemma. Subsequently, the chapter surveys data-driven approaches widely used in forecasting, optimization, and control, illustrating their relevance and potential for enabling privacy-preserving coordination in modern grids. Finally, the chapter establishes the conventions and notations adopted throughout the thesis.

Chapter 3 introduces a privacy-preserving AC-OPF framework that leverages machine learning (ML) and a multi-agent system (MAS) architecture to enhance interoperability between market operators and TSOs. The chapter formulates a two-level coordination scheme in which both agents preserve data privacy while jointly achieving system-wide optimality. Within this framework, the TSO agent develops neural network (NN) models and valid inequalities to approximate the feasible region of the AC-OPF problem using only active power, thereby concealing sensitive network parameters. To ensure accurate model training, an effective dataset is generated using novel data preparation techniques. The trained NN models are then transmitted to the market operator agent, who integrates them into a mixed-integer linear programming (MILP) formulation to perform market clearing under various operational objectives. This approach ensures that neither TSO-specific network data nor market-sensitive information is disclosed, thereby enabling secure and interoperable coordination between the two entities.

Chapter 4 focuses on the privacy-preserving analytical representation and economic valuation of DS flexibility to enable efficient coordination between TSOs and DSOs. In this chapter, the FOR concept is advanced into a three-dimensional PQV-FOR framework that explicitly accounts for voltage variability and heterogeneous FPU characteristics. A novel AC-OPF-based sampling strategy, combining bounding box projection sampling (BBPS) and Fibonacci direction sampling (FDS) methods, is developed to efficiently capture the boundary of the FOR in large-scale, meshed DSs. Subsequently, an implicit polynomial fitting method is introduced to construct a tractable analytical representation of the FOR. Furthermore, a quadratic cost function defined in the PQV domain is proposed to quantify and monetize DS flexibility. These analytical representations are then integrated into a FOR-based AC-OPF framework at the transmission level, enabling single-round TSO-DSO coordination that ensures privacy preservation and computational scalability.

Chapter 5 extends the research by introducing a novel ML-based TSO-DSO coordination framework for multi-PCC operation and direct utilization of DS flexibility. While the FOR approach serves as a powerful privacy-preserving tool, it is often limited by its single-PCC formulation and indirect flexibility dispatch. To overcome these limitations, this chapter models the technical

constraints of DSs using machine learning models trained exclusively on non-sensitive coupling variables. These models encapsulate DS behavior and are shared with the TSO, allowing it to determine the dispatch of DGs and other FPU's directly. The proposed NN architecture effectively captures the complex nonlinear characteristics of the DS, ensuring high accuracy and tractability. As a result, this framework supports direct, scalable, and privacy-compliant coordination across multiple PCCs, thereby further enhancing interoperability among system stakeholders.

Finally, Chapter 6 concludes the thesis by summarizing the key findings, highlighting the main contributions, and outlining potential directions for future research and development.

2 Background and Preliminaries

The primary objective of this dissertation is to enhance interoperability among power system stakeholders, TSOs, DSOs, market operators, and aggregators through the development of novel privacy-preserving frameworks. Achieving effective coordination among these entities is a cornerstone of modern power system operation. Yet, it remains a significant challenge due to the increasing complexity, decentralization, and digitalization of the grid. As discussed in the previous chapter, ongoing transformations driven by renewable integration, DGs, and active market participation have fundamentally altered traditional operational paradigms. To address these emerging challenges, innovative methods that simultaneously enable seamless data exchange and preserve the confidentiality of stakeholder information are required. Accordingly, this chapter provides the necessary theoretical and contextual background to support the subsequent developments presented in the thesis.

2.1 Design and Operation of the Modern Power Systems

With the intensification of global warming, the resulting environmental challenges have significantly increased the concerns of many countries and compelled them to take concrete actions toward decarbonization. According to the Paris Agreement, numerous nations have committed to substantially reducing their carbon emissions by 2030 and achieving net-zero emissions by 2050 [30]. Consequently, various sectors (industry, transportation, energy sector etc.) worldwide are undergoing extensive structural and technological transformations to meet these goals.

Among all these sectors, the electric power system occupies a central role in global decarbonization efforts. The power system remains one of the most significant sources of carbon emissions in many countries. For instance, in China, one of the highest carbon-emitting countries, electricity generation is the primary contributor to national emissions, accounting for about 45% of the country's direct carbon emissions [31]. Similar patterns are observed in other industrialized and developing economies, where electricity production represents a significant share of national emission inventories.

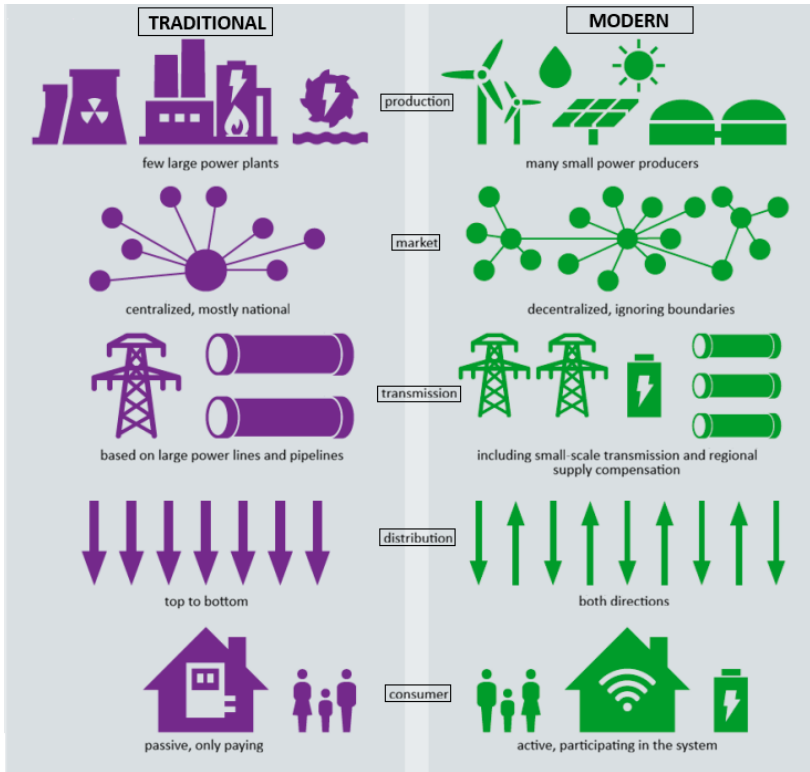


Figure 2.1: Comparison of traditional and modern power systems [32].

Consequently, the pressure to decarbonize the electric power system has become both urgent and unavoidable. This pressure is driving a fundamental transformation of power system design and operation. The transition involves not merely the replacement of carbon-intensive generation technologies with renewable energy sources, but also a comprehensive architectural and operational transformation of the grid itself, as illustrated in Fig. 2.1.

In a traditional power system, electricity is typically generated by large, centralized fossil-fuel-based synchronous machines located in conventional power plants. The generated electricity is then transmitted via high-voltage transmission lines and distributed to passive end-users. Within this framework, power flow is strictly unidirectional, from generation through transmission and distribution to consumers. The overall control paradigm follows a top-down structure, in which system operators centrally dispatch generation based on demand forecasts. At the same time, end-users remain non-participatory, exerting little to no influence on real-time system operation.

In this centralized structure, electricity markets are also highly regulated and often vertically integrated, leaving limited room for competition or flexibility. Market participation is typically

restricted to large generators and utilities, with minimal interaction between transmission and distribution levels.

Although this conventional architecture provides high reliability, operational stability, and predictability, it is increasingly challenged by global decarbonization goals and carbon neutrality commitments. As a result, the traditional power system structure is undergoing a profound transformation toward more decentralized, flexible, and market-driven architectures.

A central aspect of this transformation is the paradigm shift from large, centralized fossil-fuel-based power plants to a highly distributed generation architecture. Rather than depending predominantly on a limited number of high-capacity conventional generators, modern power systems increasingly incorporate a vast array of small-scale DGs. Unlike conventional synchronous generators, these DGs are largely dependent on variable renewable energy sources. Consequently, their power output fluctuates according to natural conditions such as solar irradiance and wind speeds, resulting in increased uncertainty and variability in power production [33].

Moreover, most DGs are connected to the distribution network, which was historically designed for unidirectional power flow. Their widespread deployment transforms the distribution grid into an active layer of the power system, where electricity may flow bidirectionally depending on local generation and demand conditions. This multi-directional power flow introduces operational challenges [34]. These developments also introduce substantial changes within transmission networks. In particular, the increased penetration of distributed and renewable energy resources has accelerated the deployment of large-scale battery installations and other advanced flexibility solutions across transmission systems.

At the consumer level, the role of end-users is also fundamentally changing. Historically, consumers are treated as passive loads with no direct influence on network operations or energy management. However, the widespread adoption of rooftop PVs, EVs, battery storage systems, and other distributed energy technologies has shifted this paradigm. Consumers are increasingly evolving into active stakeholders (commonly referred to as prosumers) who can not only consume energy but also generate, store, and strategically manage their electricity demand [35].

Parallel to these physical transformations, the architecture of electricity markets is also evolving. In contrast to traditional centralized market structures, modern electricity markets are becoming increasingly decentralized and multi-layered, facilitating participation from a diverse set of stakeholders, including aggregators, flexibility service providers, and distributed energy resource operators. Furthermore, market boundaries are expanding beyond conventional national, transmission-level frameworks. Collectively, these transformations are fundamentally redefining the roles and responsibilities of stakeholders across the power system.

When examining the evolving roles of system stakeholders, TSOs continue to bear primary responsibility for ensuring the secure and reliable operation of the transmission network. Their core duties include maintaining frequency stability, managing system congestion, and overseeing power exchanges in accordance with both technical requirements and market regulations. However, with increasing DGs integration at the distribution level, TSOs are now required to collaborate more closely with DSOs. This heightened coordination is essential for accessing distributed flexibility resources and ensuring effective system-wide balancing under increased decentralization [36].

DSOs are likewise transitioning from passive asset managers to active system operators. Historically, their primary role was to deliver electricity from the transmission network to end-users. In modern power systems, however, DSOs oversee increasingly complex distribution networks populated with DGs, EVs, and flexible demand units connected at low- and medium-voltage levels. As a result, they are now responsible for managing local congestion, maintaining voltage quality, and enabling the participation of distributed resources in energy and ancillary-service markets. This evolution necessitates the development of advanced operational tools and interoperable data-exchange frameworks to facilitate effective coordination with both TSOs and market participants.

Market operators play a pivotal role in ensuring the economic efficiency of power-system operation. They are responsible for organizing and clearing day-ahead, intraday, and balancing markets, where generation and consumption bids are matched according to established market principles. With the growing participation of flexible and distributed energy resources, market operators must now interact more dynamically with both TSOs and DSOs to ensure that market outcomes remain technically feasible and compliant with network constraints.

Another emerging entity in modern power systems is the aggregator. Aggregators coordinate small-scale flexible resources and represent them collectively in electricity markets. By aggregating these distributed resources, they enable competitive participation and serve as intermediaries that bridge operational and commercial interactions among consumers, DSOs, and market operators.

Finally, consumers are transforming into prosumers. Through smart devices, storage, and rooftop PV, etc., prosumers can respond to price signals, provide demand-side flexibility, and even offer ancillary services. Their participation not only increases overall system flexibility but also necessitates secure and efficient information exchange with higher-level operators.

All of these evolving roles, together with the changing design of the power system, are transforming its operational practices. The operation of modern power systems relies on a combination of physical control mechanisms and optimization-based decision processes aiming to ensure security, reliability, and economic efficiency. These operations take place across multiple temporal

and hierarchical layers, ranging from millisecond-scale frequency control to day-ahead market scheduling.

Day-ahead market scheduling determines which generation units will be dispatched to meet electricity demand the following day. Under this framework, dispatchable generators are scheduled to produce electricity based on forecasted system conditions. In addition to meeting expected demand, a portfolio of reserves is maintained to address potential contingencies, including transmission outages, generator failures, and unexpected fluctuations in supply and demand. These scheduling processes must also account for unit-specific operational constraints, including ramping limits, minimum up- and down-time requirements, and fuel or environmental restrictions.

As real time approaches, intraday and hours-ahead redispatch mechanisms progressively refine these schedules to accommodate updated forecasts and operational conditions. In the final operational layer, minutes-ahead balancing actions and automatic frequency-restoration mechanisms are activated to maintain real-time system equilibrium. Following any disturbance, a coordinated sequence of control actions, consisting of primary control (seconds), secondary control (tens of seconds to minutes), and tertiary control (minutes to hours), is deployed to restore system balance and ensure operational security.

Although this hierarchical control structure remains fundamental in modern power systems, its implementation is becoming increasingly complex due to high penetration of renewable energy, widespread distributed energy resources, and active demand-side participation.

2.2 Increasing Need for Interoperability

The primary objective of any power system is to maintain instantaneous balance between generation and load while accounting for network limits and operational constraints. In a traditional power system this goal is achieved through centralized planning and dispatch: day-ahead market clearing and unit commitment establish schedules that are relatively stable and reliable because most generation is large-scale, controllable, and predictable. However, the rapid transformation of power systems driven by the proliferation of DGs (e.g., wind and solar), along with the emergence of flexible loads and energy-storage technologies, has significantly increased the uncertainty and variability that system operators must manage. Renewable generation fluctuates across time scales ranging from seconds to hours and is closely linked to weather patterns, which are inherently imperfect to forecast. Consequently, achieving accurate predictions has become considerably more challenging than in the past, leading to substantial deviations between forecasted and actual system conditions [37].

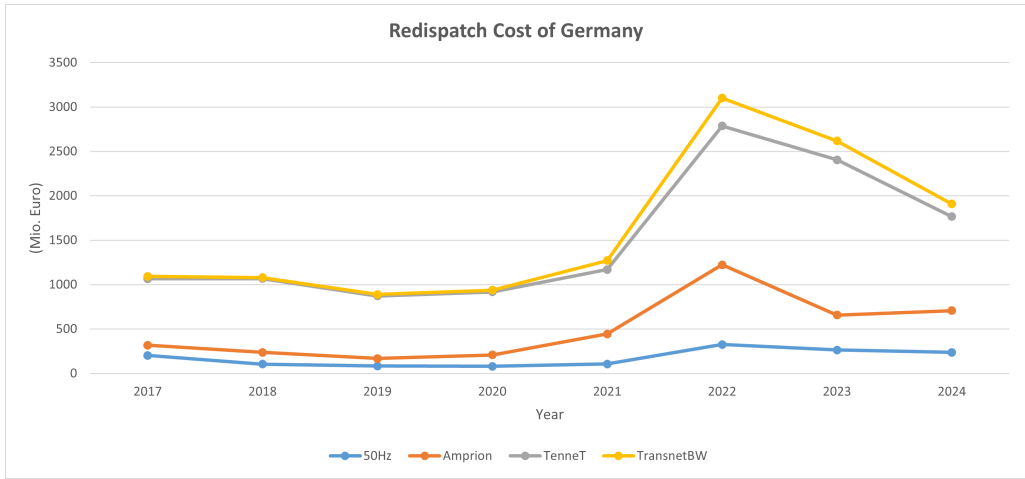


Figure 2.2: Redispatch costs in Germany [38].

As a direct consequence of increased uncertainty and variability in generation and demand, system operators are required to perform more frequent intraday adjustments and redispatch actions, and to procure additional reserves to ensure secure and balanced system operation. These corrective measures, while essential for maintaining system reliability, introduce substantial operational costs. For instance, Fig. 2.2 illustrates the evolution of redispatch expenditures in Germany across all four TSOs.

As seen in the figure, redispatch costs in Germany have experienced a significant upward trend in the last years. Particularly noticeable is the sharp increase following the pandemic and the geopolitical energy crisis in Europe, both of which amplified price volatility and stressed grid operations. Although a partial decline has been observed in more recent years, current redispatch expenditures remain well above pre-pandemic levels. Moreover, the aggregated annual redispatch costs across Germany consistently amount to several billion euros, reflecting the substantial economic burden associated with balancing actions in modern power systems.

This phenomenon is not unique to Germany. A similar trend can be observed in other regions undergoing rapid integration of renewable energy. For instance, in California, the significant growth in PV generation has led to substantial curtailment of renewables in recent years. Fig. 2.3 illustrates the curtailment of PV and wind generation over time. In 2024 alone, approximately 3.4 million MWh of utility-scale wind and solar energy were curtailed, representing a considerable loss of zero-marginal-cost clean energy and highlighting the operational challenges associated with variable generation.

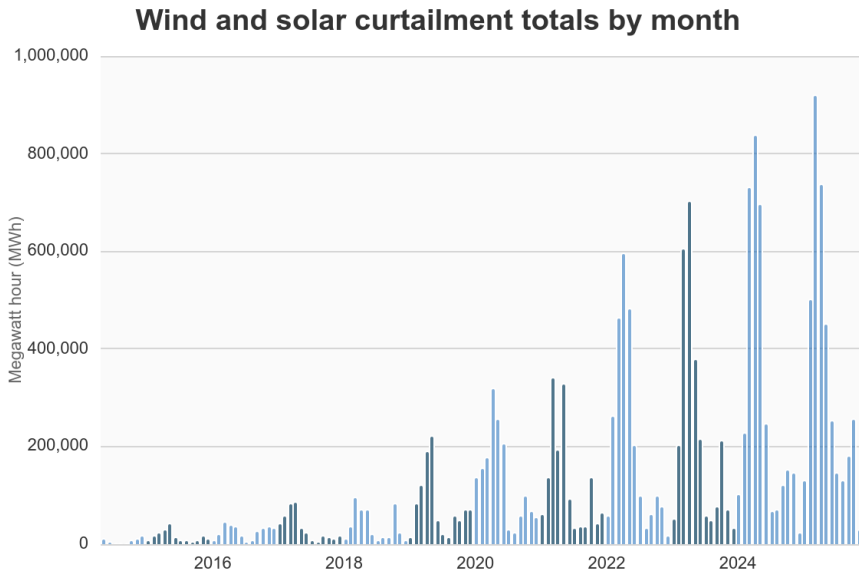


Figure 2.3: Renewable energy curtailment in California [39].

A similar situation can be observed in Great Britain. As shown in Fig. 2.4, the National Energy System Operator (NEMO), the system operator for Great Britain, has experienced a sharp increase in balancing service costs. In 2024/2025, the total cost of balancing services reached £2.7 billion, approximately 10% higher than the £2.5 billion recorded in the previous period. This increase is largely driven by the growing need for inertia and fast-frequency-response resources, resulting from the displacement of conventional synchronous generators by renewable-based and inverter-interfaced resources.

These examples collectively illustrate that the transition toward carbon neutrality, while essential, brings significant operational and economic challenges. As renewable penetration increases, maintaining real-time balance between supply and demand becomes more complex and costly.

To preserve the supply-demand balance and ensure secure operation under increasing system uncertainty, power system stakeholders are deploying a wide spectrum of technological, operational, and market-based solutions. One major approach involves grid reinforcement and expansion. Transmission and distribution system operators continue to invest in new substations, transformers, overhead lines, and underground cables to alleviate congestion and enhance transfer capacity. For example, German TSOs have identified extensive grid expansion requirements to support the energy transition, with total investment projected to reach approximately €732 billion by 2045 [41, 42].

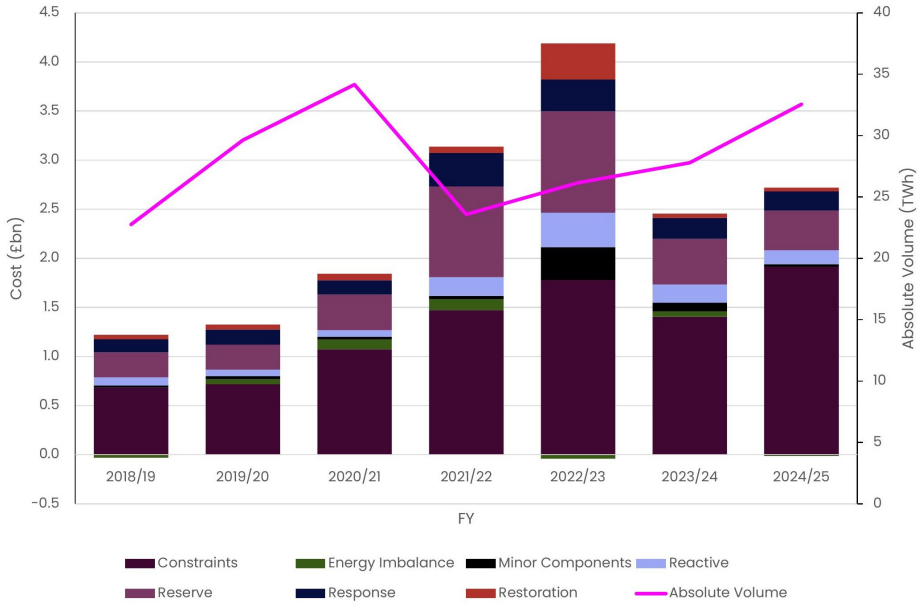


Figure 2.4: Balancing service cost in Great Britain [40].

In parallel, operators are accelerating digitalization initiatives, deploying advanced metering infrastructure, wide-area monitoring systems, and real-time data platforms to enhance situational awareness and operational coordination. Through these developments, system operators aim to transition from traditional reactive control strategies towards predictive and data-driven operational paradigms.

Another critical solution focuses on flexibility enhancement. One strategy in this context is to deploy energy storage technologies, including utility-scale batteries and pumped-hydro storage systems. These resources are capable of delivering fast frequency response, alleviating network congestion, and providing various other ancillary services essential for system stability. Although the cost of battery energy storage has significantly declined over the past decade, capital expenditures remain substantial, and large-scale deployment continues to represent a major financial undertaking for system operators and investors [43]. Nevertheless, energy storage systems are increasingly recognized as indispensable assets for facilitating the integration of renewable energy and maintaining secure and efficient power-system operation.

A cost-effective and increasingly prominent strategy for enhancing power system flexibility is the utilization of existing DGs and other FPU within the system, thereby avoiding substantial new

infrastructure investments. Since these resources are predominantly connected to the distribution network, distribution-level flexibility has emerged as a critical component of modern grid operation [44].

DS flexibility can be activated by dynamically adjusting consumption or generation patterns of these distributed assets. For example, photovoltaic output may be curtailed or optimally scheduled, battery systems can be charged or discharged in response to system requirements, electric-vehicle charging can be shifted to off-peak periods, and industrial or residential loads can be modulated through demand-side management programs. Such coordinated flexibility actions enable distributed assets to support grid stability, improve operational efficiency, and mitigate the variability associated with renewable-energy production.

The benefits of this approach are multi-dimensional. First, by utilizing existing distributed resources, it reduces the need for network expansion and infrastructure upgrades, thereby lowering capital expenditures for both system operators and consumers. Second, flexibility at the distribution level facilitates greater integration of renewable energy by reducing curtailment and thus enabling higher penetration of variable resources. Third, local balancing of supply and demand helps alleviate stress on transmission and distribution networks, contributing to congestion management, voltage regulation, and loss reduction, thereby improving operational efficiency. The final benefit of DS flexibility is the participation of the consumer-owned assets in flexibility services. Thus, traditional passive consumers can be active prosumers, creating new revenue streams.

Active participation of diverse stakeholders in power system operation has become increasingly essential for cost-effective grid management. With the transformation, traditional centralized operational practices are no longer sufficient; instead, system operators are progressively integrating a wider range of actors into the operations. In this context, several regulatory and operational advancements have been introduced worldwide to enable more inclusive and coordinated system operation.

A prominent example is Germany's implementation of *Redispatch 2.0*, which fundamentally restructured redispatch procedures. Historically, redispatch was primarily conducted using conventional power plants with capacities greater than 10 MW. However, under *Redispatch 2.0*, participation has been extended to all generation units with a capacity of 100 kW or more, including DGs [45]. Moreover, remotely controllable units below 100 kW are also eligible for redispatch participation, significantly broadening the resource pool. As a direct consequence, DSOs now play an active role in managing network constraints and facilitating the use of flexibility resources at lower voltage levels.

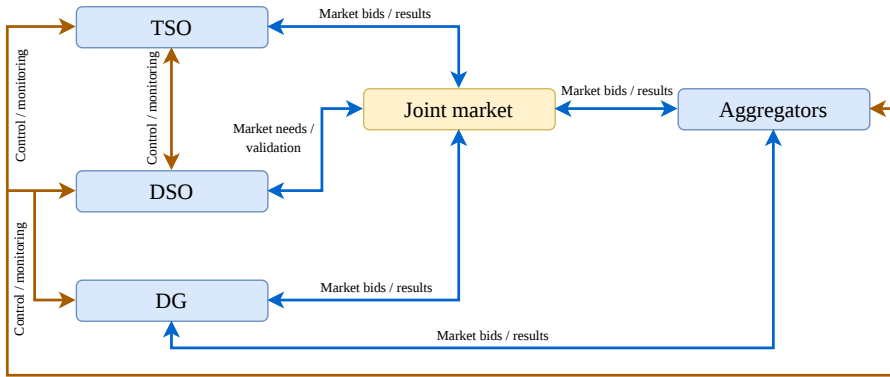


Figure 2.5: Example coordination scheme of power system stakeholders [48].

Building on this framework, the planned *Redispatch 3.0* initiative seeks to further expand flexibility provision. The aim is to incorporate controllable assets below 100 kW installed in low- and medium-voltage networks, including battery storage systems, electric vehicles, heat pumps, and small-scale photovoltaic-plus-storage systems [46]. In addition, *Redispatch 3.0* emphasizes enhanced communication infrastructures, increased automation, and deeper digitalization to enable real-time coordination and monitoring of distributed assets.

Through these initiatives, Germany is moving toward a multi-stakeholder redispatch ecosystem in which numerous distributed units contribute to grid stability. This transition not only expands the available flexibility resources but also enables balancing tasks to be performed more cost-effectively.

All these transformations in the design and operation of modern power systems inevitably require enhanced coordination and information exchange among the various stakeholders involved [47]. The transition from centralized to decentralized architectures has introduced a wide range of actors, including TSOs, DSOs, market operators, aggregators, prosumers, and flexibility service providers, each with distinct roles, objectives, and data ownership responsibilities. As a result, effective coordination mechanisms between these entities have become essential.

Multiple coordination schemes exist among power-system stakeholders, each reflecting varying degrees of information exchange, control interaction, and decision-making authority. An illustrative example of such coordination is presented in Fig. 2.5.

In this scheme, market operators determine energy schedules and dispatch decisions primarily based on economic optimization, whereas TSO and DSO validate the technical feasibility of these schedules with respect to network constraints. Similarly, coordination between TSOs and DSOs involves continuous information exchange to enable joint optimization of system operation and

flexibility management. In parallel, aggregators play an increasingly pivotal role by pooling DGs, interfacing them with both market and grid operators.

However, the effectiveness of these coordination frameworks depends heavily on the interoperability between different entities, i.e., the ability to exchange, interpret, and utilize information across organizational and technical boundaries. Consequently, enhancing interoperability among stakeholders has become a top priority for achieving secure and flexible operation in modern power systems.

At the same time, as the power system continues to transform, data exchange has become increasingly intensive and frequent. In this context, one of the most crucial challenges arises from data privacy concerns among stakeholders. Within these coordination frameworks, sensitive information, such as network parameters, market bids, or customer consumption profiles, must be frequently exchanged among different entities. Hence, this data must be protected not only from third parties but also, in some cases, from other system stakeholders themselves.

Therefore, there is a growing need for the development of privacy-preserving frameworks and approaches that enable seamless coordination and interoperability in power systems without compromising data confidentiality.

2.3 Data Privacy Aspects

As discussed above, the ongoing transformation of modern power systems, driven by digitalization, decentralization, and increasing interconnectivity, has made data privacy one of the most critical and complex aspects of secure and efficient system operation. The shift from centralized architectures to distributed energy systems has led to a rapid escalation in the amount and diversity of data exchanged among stakeholders. At the same time, in modern grids, maintaining the delicate balance between generation and demand has become considerably more challenging due to the high penetration of DGs. Consequently, system operators and other entities increasingly depend on real-time monitoring, forecasting, and control mechanisms, all of which require the continuous exchange of detailed operational data.

This transformation is further reinforced by the widespread deployment of advanced digital and communication technologies, such as smart meters, sensors, and intelligent electronic devices, which collectively generate vast amounts of granular data at different levels of the power system. While this data provides valuable insights for enhancing system flexibility, efficiency, and resilience, it also introduces new vulnerabilities and risks [49].

In traditional power systems, most operational processes are centrally managed, and the exchange of information among entities is relatively limited, structured, and well-defined. However, as the system transforms into a more distributed and dynamic ecosystem, the volume, frequency, and sensitivity of data exchanged have increased dramatically. Each stakeholder now possesses unique datasets that are essential for maintaining secure, efficient, and coordinated operation of the grid. At the same time, many of these datasets contain commercially sensitive, operationally critical, or personally identifiable information, which must be protected against unauthorized disclosure or misuse.

This situation creates an inherent privacy-coordination dilemma: on one hand, effective coordination and system optimization require the sharing of detailed operational data among stakeholders; on the other hand, data privacy and regulatory constraints limit the extent to which such information can be openly exchanged. Balancing these conflicting requirements, enabling data-driven coordination while preserving stakeholder confidentiality, has thus become a central challenge in the modern power system landscape.

For instance, TSOs possess highly detailed information regarding transmission network topology, line parameters, contingency models, and real-time system states. DSOs, in turn, manage extensive datasets describing distribution network configurations and consumer-level load and distributed generation profiles. Market operators handle commercially sensitive information, including bidding strategies, price forecasts, and contractual commitments. Meanwhile, aggregators and prosumers generate behavioral and locational data that can reveal individual consumption patterns, operational schedules, or even financial characteristics. Although these datasets are indispensable for ensuring secure, efficient, and coordinated system operation, they are also inherently confidential and often subject to strict regulatory protection. Consequently, unrestricted data sharing among entities is neither feasible nor desirable, as it could expose stakeholders to competitive disadvantages, privacy breaches, or cyber risks.

The unrestricted sharing of such data across institutional boundaries can lead to multiple forms of risk. From a technical perspective, it increases exposure to cyberattacks, data interception, and malicious model inversion, which could compromise operational security. Such vulnerabilities can trigger cascading failures or widespread blackouts. A striking example occurred during the 2015 cyberattack on Ukraine's power grid, where attackers gained unauthorized access to substation SCADA systems, resulting in the disconnection of multiple substations and leaving approximately 230,000 customers without electricity for several hours [50]. Similarly, in 2022, a major cyber incident in Germany disrupted the communication infrastructure used to control and monitor nearly 2,000 wind turbines [51].

From a commercial perspective, unrestricted or poorly protected data exchange can lead to the leakage of sensitive market information, such as bidding strategies, price forecasts, and contractual

positions. Such exposures may enable market manipulation, collusion, or unfair competition among market participants, thereby undermining both system efficiency and market integrity. Several studies have demonstrated that malicious actors can exploit vulnerabilities in electricity markets, for instance, by applying cyberattacks on day-ahead or real-time markets to maximize their own profits [52].

In addition to these risks, cyberattacks can also compromise the privacy of end-users, prosumers, and DG owners. Through unauthorized access to advanced metering infrastructure, communication networks, or cloud-based energy management platforms, attackers may reconstruct individual load or generation profiles, revealing sensitive behavioral, locational, or financial information [53]. Such data can reveal patterns of occupancy, appliance usage, or even industrial production schedules, posing significant privacy and security concerns for both residential and commercial stakeholders.

With the increasing reliance on data exchange for coordination and optimization in modern power systems, ensuring data privacy while maintaining operational efficiency has become a central research objective [54]. Several methodologies have been proposed in the literature to achieve this balance, each grounded in distinct theoretical principles and offering different trade-offs between privacy, computational complexity, and coordination efficiency. Among the most prominent are distributed optimization algorithms, differential privacy techniques, and federated learning frameworks.

One of the most used approaches to privacy-preserving coordination in power systems is distributed optimization. In these frameworks, the global optimization problem, such as OPF or other market-clearing optimizations, is decomposed into smaller subproblems, each handled by a local agent (e.g., a TSO, DSO, or aggregator) [55]. In this architecture, local agents solve their respective problems using only their own data and exchange only limited information (such as voltage-current values at the interface). Through iterative updates, agents coordinate their decisions until global feasibility and optimality are achieved [56]. Through this architecture, the global optimization problem can be solved collaboratively without requiring agents to disclose their sensitive internal data, such as network topology, line parameters, or cost coefficients, to other entities. This ensures data privacy, since each participant retains full control over its proprietary information. Moreover, because the global problem is decomposed into smaller local subproblems, the approach exhibits excellent scalability: each agent solves only a limited portion of the overall task, thereby substantially reducing the computational requirements and facilitating parallel processing. An additional advantage of this structure is its modularity. Each stakeholder can independently maintain, modify, or update its local model without affecting the others. This feature makes distributed optimization particularly well-suited for TSO-DSO coordination problems, where both parties operate under different objectives and regulatory frameworks but

must nonetheless achieve joint system feasibility and optimality [57]. However, despite these advantages, distributed optimization algorithms also face several important limitations. A primary drawback is their high communication burden. Since convergence to the global optimum typically requires multiple iterations of data exchange among agents, the resulting communication overhead can be substantial, leading to latency. Moreover, ensuring secure and reliable communication between agents necessitates robust technical infrastructure, including encrypted data channels, authentication mechanisms, and synchronization protocols. These additional requirements not only increase system complexity but also impose non-negligible implementation and maintenance costs, particularly in large-scale power system environments. Furthermore, although sensitive data are not explicitly shared, the iterative exchange of coordination variables can unintentionally reveal confidential information through inference or correlation analysis.

Another prominent privacy-preserving approach widely discussed in the literature is the differential privacy mechanism. The core principle of differential privacy is to introduce carefully calibrated random noise into datasets or query results, thereby ensuring that the inclusion or exclusion of any single data record, such as an individual consumer's load or a distributed generator's production profile, cannot be inferred with high confidence. The amount of noise added is governed by a tunable parameter, which quantifies the trade-off between data accuracy and privacy protection. By properly adjusting this parameter, system operators can ensure that the released data remains useful for analysis and decision-making while maintaining a mathematically provable level of privacy. In power systems, differential privacy has been applied to a variety of use cases, including smart meter data aggregation, load forecasting, market data publication, and grid operation [58]. For example, an OPF framework can be designed using differential privacy methods for TSO and DSO to prevent the disclosure of individual customer behavior [59]. Through this mechanism, even if an attacker intercepts the shared dataset, it becomes statistically impossible to infer sensitive information about any specific customer or DG operator. Moreover, unlike distributed optimization algorithms, which require multiple iterative rounds of data exchange, differential privacy can be implemented as a single-round operation, thereby reducing communication overhead and the risk of information leakage during coordination. However, despite these advantages, the differential privacy mechanism also entails significant challenges. The most fundamental limitation arises from the accuracy-privacy trade-off: while increasing the amount of noise enhances privacy protection, it simultaneously degrades the precision of shared data and the reliability of derived operational results. This can be particularly problematic in applications requiring high temporal resolution or real-time control, such as frequency regulation, state estimation, or market clearing, where even small inaccuracies may propagate into unstable or suboptimal system behavior. Furthermore, selecting an appropriate noise parameter is inherently complex, as it depends on the sensitivity of the underlying data, the scale of aggregation, and the operational context. Another important concern is that excessive noise addition may distort the

physical interpretability of the data, potentially decoupling it from the true behavior of the underlying electrical system. This limits the feasibility of differential privacy for direct operational coordination between system actors. Consequently, although the method has demonstrated strong potential for statistical data publication, anonymized market reporting, and privacy-aware analytics, its direct application in real-time or optimization-based decision-making remains constrained by the inherent balance between privacy protection and data utility.

Another type of privacy protection method that has been widely explored in recent years is Federated Learning. Unlike conventional centralized approaches, where all raw data must be collected, stored, and processed at a central location, federated learning enables multiple entities to collaboratively train a shared global model without exchanging their private datasets. Instead, each stakeholder performs training locally using its own data and transmits only model parameters, gradients, or compressed updates to a coordinating server. The central server aggregates these updates, typically through weighted averaging, to produce an improved global model, which is then redistributed back to participants for the next training round. In this way, learning occurs cooperatively while sensitive operational data never leaves the owner's local environment. In the domain of power system applications, federated learning has gained growing attention due to the increasing availability of high-resolution operational data and the need for privacy-aware analytics. Recent studies have demonstrated the successful application of federated learning to load and renewable forecasting, fault detection, and market optimization [60, 61]. Federated learning also enhances cybersecurity, as raw data never passes through communication networks, significantly reducing the attack surface for potential adversaries. Moreover, the framework inherently supports real-time or near real-time model updates, which is valuable for power systems characterized by fast-changing load patterns, volatile renewable generation, and dynamic operational conditions. The approach encourages active participation of decentralized stakeholders, strengthens data sovereignty, and aligns well with the increasing digitalization and decentralization of power system operation.

Despite these advantages, several challenges restrict the direct adoption of federated learning in operational decision-making. A major limitation is the communication overhead associated with frequent exchange of model updates, especially when models are large, training rounds are numerous, or network conditions are constrained. Compared with single-round data-sharing methods, federated learning may require dozens or even hundreds of communication iterations to reach convergence. Ensuring secure and authenticated communication channels further increases system complexity.

Another difficulty arises from the fact that data distributions across power system stakeholders are often highly non-uniform, unbalanced, and non-independent. Such characteristics can slow convergence, degrade model accuracy, or even lead to biased global models.

As highlighted, each privacy-preserving technique proposed in the literature offers valuable benefits but also exhibits important limitations when applied to real-world power system coordination. Given these constraints, there is a growing need for new privacy-preserving frameworks that ensure both operational feasibility and practical implementability, particularly in large-scale, nonconvex, and data-intensive grid operations. This motivates the adoption of data-driven approaches as a powerful alternative. Instead of exchanging sensitive system models, network parameters, or real-time measurements, stakeholders can share learned representations, surrogate models, or behavioral approximations that encapsulate only the essential input-output relationships required for coordination.

The core idea is to achieve privacy by design: rather than protecting sensitive data after it has been exchanged, the framework is constructed such that sensitive information never leaves its owner in the first place. In a data-driven setting, the detailed internal characteristics of a system (e.g., topology, line limits, voltage constraints, bidding strategies) are replaced with locally trained models, such as neural networks, analytical parameterizations, or feasible region representations, that map shared variables to the necessary operational outputs. Additionally, when these models are created with non-sensitive information already shared among stakeholders, the protection of private data can be fully ensured.

Furthermore, data-driven models provide several additional advantages. They are computationally lightweight compared to cryptographic or iterative distributed schemes, enabling their integration in time-critical operations. Unlike differential privacy, they do not rely on artificially added noise that might compromise system security or accuracy. They also naturally handle nonlinearities and nonconvexities that central characteristics of AC-OPF and flexibility modeling.

2.4 Data-Driven Approaches in Power Systems

The growing complexity, decentralization, and uncertainty of modern power systems have accelerated a shift from purely model-based methodologies toward data-driven approaches. Traditionally, power-system operation and optimization have relied on deterministic, physics-based models supported by detailed system parameters such as network topology, generator characteristics, demand profiles, and other system attributes. Although these models remain essential, their accuracy and practical applicability are increasingly constrained by the rapid integration of DGs, EVs, flexible prosumers, and advanced control technologies, among other emerging system components. These emerging components exhibit highly dynamic, uncertain, and location-dependent behaviors that are difficult to represent accurately using conventional analytical formulations alone.

In this context, data-driven techniques have emerged as powerful tools for enhancing forecasting, control, and optimization in modern grids. By learning patterns directly from historical or simulated data, these methods can approximate complex nonlinear relationships that are otherwise difficult or computationally intensive to model explicitly.

Data-driven approaches are used for a range of purposes in the literature. For instance, a major research direction focuses on forecasting applications, where machine learning (ML) and statistical learning methods have demonstrated significant improvements over classical methods. In particular, data-driven methods have been widely used for short-term and long-term load forecasting, PV and wind generation forecasting, electricity price prediction, and demand response estimation [62]. Approaches such as neural networks, support vector regression, and more recently deep learning architectures, including long short-term memory (LSTM), convolutional neural network (CNN), have become state-of-the-art due to their ability to capture nonlinear temporal dependencies and exogenous influences [63]. These forecasting tools play a central role in unit commitment, economic dispatch, market clearing, and renewable integration [64].

Another important area is data-driven control and optimization, in which surrogate models or learned approximations are employed to replace computationally intensive analytical formulations. Many studies have explored learning-based approximations of AC power flow or optimal power flow [65]. These methods enable real-time decision-making by significantly reducing computation time while maintaining reasonable accuracy. Also, Physics-informed neural networks further integrate physical laws into the learning process, thereby improving generalization and reliability [66]. These methods are particularly valuable in systems where detailed network models are unavailable or where system conditions change too rapidly for traditional optimization to be executed reliably.

Data-driven methods also play a prominent role in system monitoring, anomaly detection, and cyberattack identification. By leveraging diverse data sources, such as smart meter data, sensor information, and communication logs, ML-based anomaly detection schemes have been proposed to identify false data injection attacks, abnormal consumption behaviors, and various forms of communication intrusion [67]. These tools enhance situational awareness and improve the resilience of cyber-physical power systems against emerging threats.

Another active domain in which data-driven methods are increasingly employed is stability and security assessment. Traditional transient-stability evaluation depends on computationally intensive time-domain simulations. In contrast, ML-based classifiers trained on large simulation datasets can rapidly estimate transient stability margins or classify post-contingency operating points [68]. Similarly, data-driven models have been used to evaluate voltage stability limits,

approximate dynamic responses, and assist in preventive control [69]. These approaches substantially accelerate security assessment and enable near-real-time screening of large sets of system contingencies.

Overall, the literature demonstrates that data-driven approaches have become indispensable in forecasting, optimization, control, stability assessment, and anomaly detection within modern power systems. Overall, the existing literature demonstrates that data-driven approaches have become indispensable across a wide range of power system applications, including forecasting, optimization, control, stability assessment, and anomaly detection. Their ability to model complex nonlinear relationships and operate effectively under uncertainty has fundamentally reshaped operational and planning paradigms in modern grids. Importantly, the shift toward data-centric methodologies has also revealed their significant potential in addressing one of the most challenging aspects of contemporary power system interoperability: data privacy.

As data-driven models rely primarily on learned statistical representations rather than on explicit sharing of raw measurements, topological information, or device-level parameters, they offer inherent structural advantages for privacy-preserving collaboration. For example, model-based abstractions, feature embeddings, surrogate mappings, or parameterized representations can serve as substitutes for sensitive datasets, enabling stakeholders to contribute valuable operational knowledge without disclosing confidential information. In this sense, data-driven approaches not only enhance the accuracy, scalability, and adaptability of modern power system applications but also form a foundational building block for developing next-generation privacy-preserving interoperability frameworks.

While data-driven approaches offer significant potential for improving operational efficiency and enabling privacy-preserving coordination, their successful deployment in real power system environments requires meeting several methodological and infrastructural requirements. A fundamental requirement is the availability of high-quality and sufficiently diverse datasets. Data-driven models must capture a wide range of operating conditions, including typical system behavior, rare events, contingencies, extreme weather impacts, and variability. In situations where real-world data are insufficient, incomplete, or unavailable, synthetic datasets must be generated using appropriate techniques. These synthetic frameworks must faithfully reflect physical constraints and operational uncertainties to avoid introducing systematic errors into learning-based models. The generation of realistic synthetic datasets has therefore become a crucial prerequisite for the practical use of data-driven tools.

Furthermore, the adoption of data-driven approaches necessitates access to adequate computational infrastructure and scalable algorithmic frameworks. Many modern ML models require substantial computational resources for both training and inference. Therefore, to ensure their effective and reliable use in power-system applications, appropriate solutions must be developed

not only for model development but also for their real-time or near-real-time implementation in operational environments.

Meeting these requirements is essential not only for achieving high model performance but also for enabling the integration of data-driven techniques into privacy-preserving interoperability frameworks that support the evolving needs of modern power system operations. In this thesis, all proposed methodologies and frameworks are developed with these considerations at the forefront. Particular emphasis has been placed on ensuring that the models rely exclusively on non-sensitive data. This design choice is intentional: because the resulting models may be shared among different stakeholders or potentially accessed by third parties in the event of data leakage, it is imperative that no confidential operational parameters, commercially sensitive information, or personally identifiable data can be reconstructed or inferred from them.

By embedding these principles directly into the methodological design, often referred to as *privacy by design*, the developed frameworks inherently safeguard stakeholder confidentiality, even under adversarial inference scenarios. At the same time, they maintain compatibility with the coordination requirements of modern grids, thus contributing to enhanced interoperability. Through this approach, the thesis demonstrates that effective and privacy-preserving coordination can be achieved without compromising model performance or operational practicality.

2.5 Conventions and Notations

Throughout this thesis, the following conventions are adopted. Parameters are denoted by standard letters (a , A), while variables are represented using boldface letters ($\mathbf{a}$, $\mathbf{A}$). Sets are expressed using calligraphic letters ($\mathcal{A}$). Matrices appear in uppercase letters (A), whereas scalars and (column) vectors are denoted in lowercase (a). Functions are written in the form $A(\cdot)$.

For a vector a , the n -th element is written as $a^{(n)}$, whereas $a^{(-n)}$ represents all elements of a vector except the n -th element. For a matrix A , $A^{(n,\cdot)}$ refers to the n -th row, and $A^{(i,j)}$ denotes the element in the i -th row and j -th column. Furthermore, $\leq$ and $\geq$ are used for element-wise comparisons, $\leqslant$ and $\geqslant$ denote standard scalar comparisons. Finally, $\odot$ is the Hadamard product, and $\oslash$ is the Hadamard division.

It should also be noted that each of the subsequent chapters (Chapters 3, 4, and 5) focuses on different aspects and perspectives of interoperability and, consequently, employs distinct mathematical formulations. For instance, Chapter 4 addresses TSO-DSO coordination through the point of common coupling (PCC), whereas Chapter 5 considers direct dispatch of DGs and FPU. Therefore, each chapter introduces its own notation tailored to the specific methods

developed therein. To ensure clarity and readability, all notations are defined locally and remain consistent within the respective chapters.

3 Privacy-Preserving Cross-Entity AC-OPF Between Market Operators and TSOs

In this chapter, the first research question of the dissertation is addressed, focusing on how effective cross-entity optimization between market operators and TSOs can be achieved without requiring the exchange of sensitive data, thereby enhancing interoperability between these entities. As modern power systems increasingly rely on distributed, data-intensive operational processes, the need for secure and efficient exchange of information across institutional boundaries has become more critical than ever. However, traditional AC-OPF coordination frameworks typically require the disclosure of sensitive network parameters, cost structures, or market strategies that stakeholders are unwilling or unable to share. This chapter proposes a novel solution that reconciles the operational necessity of coordination with the preservation of data confidentiality.

To this end, the chapter develops a privacy-preserving AC-OPF framework based on a multi-agent system (MAS) architecture supported by ML. The framework enables a market operator and a TSO to jointly achieve system-wide optimality while ensuring that neither party exposes private data. The central idea is to approximate the feasible operating region of the TSO's AC-OPF problem using neural network (NN) models and valid inequalities trained exclusively on non-sensitive active power data. These models, generated by the TSO, capture the complex AC behavior of the transmission system without revealing network topology, line parameters, or other operationally sensitive attributes. Furthermore, these models are trained using an effective dataset generated through n-ball, multivariate Gaussian distribution (MGD), and perturbation techniques. These techniques are employed to ensure the creation of a balanced dataset containing high-quality data points around the boundary. Once trained and validated, the NN models are transmitted to the market operator, who incorporates them into a mixed-integer linear programming (MILP) formulation to perform market clearing under different operational objectives. This two-layer coordination mechanism ensures mathematically consistent operation while preserving complete data confidentiality.

Beyond the overarching contribution of interoperability, the chapter also provides several methodological contributions as follows:

- The use of an NN representation in combination with a set of valid inequalities to represent the feasible set of AC-OPF in the MILP formulation prevents the sharing of sensitive data.
- An effective dataset generation methodology, which uses only active power, enables the training of an NN with accurate representation capability of the feasible set of an AC-OPF, even with a relatively small dataset and shallow NN architecture.
- The incorporation of the trained NN into a MILP formulation, which can be efficiently solved using commercial solvers. The resulting MILP is highly flexible, allowing for modifications to the objective function of the AC-OPF and the integration of additional constraints, as needed.

The rest of this chapter is organized as follows: The state of the art is reviewed in Section 3.1. In Section 3.2, the structure of the proposed MAS is described. In Section 3.3, the proposed methodology is presented. In Section 3.4, an effective dataset creation technique is introduced. Subsequently, in Section 3.5, the structure of the NN models and their transformation to the MILP are detailed. Then, the proposed method is benchmarked with the AC-OPF, which does not consider data privacy, to evaluate its effectiveness through different case studies in Section 3.6. Finally, this chapter is summarized in Section 3.7. This chapter is primarily based on the findings presented in [25].

3.1 State of the Art

To manage the electricity market effectively and prevent energy bottlenecks, it is crucial to resolve AC-OPF and make dispatch decisions [70]. The use of the AC-OPF and dispatch mechanisms in the electricity market has the potential to generate significant economic gains. For example, in the United States, increasing market efficiency by 5% through AC-OPF can result in annual savings of up to 6 billion dollars [71]. Given the expected increase in uncertainty in the near future, the use of effective AC-OPF and dispatch mechanisms can lead to even greater economic gains.

In the transformed electricity grid, TSOs, DSOs, DERs companies, and aggregators must work together. For instance, in Germany's Redispatch 2.0, all conventional plants and DERs with an installed capacity of 100 kW or more, as well as DSOs, are required to participate in redispatch [45]. Additionally, in the future Redispatch 3.0, private customers will be allowed to participate in the market via aggregators [72]. Therefore, AC-OPF and redispatch should now be realized with the participation of different stakeholders, not only by system operators, as in the past. This

will increase the need for cooperation among the institutions. To this end, various institutions, including universities, TSOs, DSOs, and consulting firms, are working together on research and development projects, such as Digiplat and Enera, to develop and implement effective cooperation mechanisms for flexibility and redispatch decisions [73, 74].

Data privacy is one of the most important factors that complicates cooperation between industry partners. For instance, in a project that plans to establish a common platform for the redispatch mechanism, a TSO needs to share sensitive data, such as grid topology, voltage, and current magnitudes, with other partners that perform market research. Therefore, there is a need for a mechanism that protects data privacy between the model owner and other parties. To protect data privacy, institutions have developed various methods such as sharing simplified "light models" of the grid [56]. Nevertheless, these light models may still contain sensitive data, and it remains unclear to what extent simplified models can accurately represent the underlying power system. In the present study, a framework that prevents the sharing of sensitive data between institutions, who own grid models and do not own grid models, is the main objective to increase cooperation by providing data privacy.

In the literature, to preserve data privacy, a distributed OPF approach is used. In this approach, the problem is divided into sub-problems to prevent complete grid model sharing. However, the complex voltage and/or active and reactive power of tie-lines are shared between neighboring regions [75, 76]. Despite being less sensitive, this data could still pose privacy risks due to potential re-identification issues [7]: a privacy-preserving distributed OPF algorithm is developed and various encryption techniques are used to prevent violations. In [77] an encryption step is added to differentially private projected subgradient algorithm to preserve data privacy. With the added encryption step, OPF's accuracy remains the same, but it affects computation negatively by increasing the number of iterations. A privacy-preserving quadratic optimization algorithm is introduced in [78] to address security-constrained OPF in the smart grid. Although the number of iterations remains relatively unchanged, this method is 2^4 times slower compared to the non-privacy-preserving approach. In [7] a differentially private distributed algorithm is proposed to preserve re-identification of shared data. This method ensures the feasibility of the solution while adhering to constraints. However, its computation time is notably long; for instance, the 118-bus and 189-bus systems require 82.160 and 86.908 minutes, respectively. A partially homomorphic encryption technique is proposed in [79]. With this method, the OPF can be solved with several iterations, and the optimality gap is very small. While the paper presents an innovative approach to data privacy through encryption, its implementation may pose significant challenges, as the method cannot be readily solved through commercial solvers and lacks flexibility in accommodating additional constraints beyond standard optimal power flow constraints. As can be seen, there are trade-offs between data privacy and efficiency of the AC-OPF in the literature. However, for an effective privacy-preserving OPF, the computational burden should be low, the

model and solution performance, and the flexibility of the method should be high. Therefore, innovative methods are needed. The method proposed in the present study ensures that all these requirements are satisfied.

Instead of using distributed optimization methods, an alternative approach for achieving privacy-preserving OPF involves employing machine learning models. Specifically, the party with access to the system model can train a machine learning model to obtain an equivalent representation of the system model. This trained model can then be shared with other parties instead of the original model, which can be used within the OPF framework to seek optimal power dispatch without requiring sensitive data. Establishing such a framework is the primary objective of the present chapter. In recent years, numerous studies have explored the use of machine learning-based approaches in the context of OPF problems. In general, these models are employed for "end-to-end" prediction, wherein they are trained to output the optimal generation dispatch for a given input of load distribution [80, 81]. These methods are generally used to reduce the computational burden of OPF. In [82], an integration of deep NN and Lagrangian duality is proposed, enabling OPF to be solved in milliseconds for systems with up to 3400 buses while accounting for system constraints. A physics-informed, data-driven OPF approach utilizing a stacked extreme learning machine is introduced in [83], which improves the algorithm efficiency by simplifying the time-consuming parameter tuning process. Additionally, [84] develops a deep NN for high-dimensional load-to-solution mapping to solve the OPF, resulting in a significant acceleration of computation time. However, because these end-to-end approaches do not account for data privacy and their ML models are challenging to adapt to varying constraints and objective functions, they are not directly applicable to privacy-preserving OPF.

In order to achieve privacy-preserving OPF using an ML-based representation of the power system model, two essential components are required. Firstly, the ML model should accurately represent the underlying system model to ensure that the OPF results obtained via the ML-based representation do not differ significantly from the results obtained using the actual system model. Secondly, an efficient integration of the trained ML-model within an optimization formulation is necessary to ensure that the problem can be solved to (near-) optimality without undue computational expense.

The accuracy of the ML-based representation is largely contingent upon the representational power of the employed ML model, as well as the quality of the training data [85, 86]. To ensure effectiveness, datasets should feature a balanced ratio of feasible/infeasible data points, including high-quality samples around decision boundary. Consequently, there is considerable research emphasis on effective dataset generation techniques. Recent advancements in this domain are summarized in [87]. Despite considerable efforts, many existing methods only represent a small portion of the feasible space, impacting overall model accuracy. Addressing these challenges,

a new split-based method proposed in [87] aims to generate diverse data samples covering the entire feasible space, however it does not adequately address class imbalance. Another approach, outlined in [88], offers an efficient dataset generation method leveraging hyperplanes to eliminate large portions of the input space and create a balanced dataset. However, the required data points to produce a balanced dataset are too many and ML model accuracies are poor. Moreover, these methodologies do not guarantee the acquisition of high-quality data, particularly around the decision boundary, thereby hindering the accurate representation of the power system model using ML models.

3.2 Structure of the Proposed Multi-Agent System

In the present chapter, a multi-agent system (MAS) is introduced to ensure data privacy among institutions. The primary agents within this framework include the technical operator agent (TOA) and market operator agent (MOA).

The TOA is equipped with a comprehensive grid model and is responsible for the secure operation of the grid. Consequently, the TOA must adhere to various technical constraints, including voltage thresholds, line and transformer loading limits, as well as the active and reactive power constraints of the generators. The TOA can accurately estimate the active and reactive power requirements of the demand load at each bus by using its access to the complete grid model and historical grid data.

Based on this information, the TOA can define the feasible operational space of the power system. However, due to sensitive data such as the power system's topology and customer load information, sharing this data raises concerns about data privacy. Therefore, to safeguard data privacy, the feasible space of the power system is defined using NN models. Additionally, the TOA generates valid inequalities to reduce the search space and enables the MOA to solve the optimization problem more accurately and efficiently. These valid inequalities and NN models can be shared with the MOA conveniently, preserving the sensitive data.

The MOA lacks a grid model, but has bid/cost data and market constraints. This agent leverages this information to minimize dispatch or redispatch costs and maintain a balance between demand and generation through optimization techniques.

The MOA must consider the limits of the power system when conducting market based optimization. In this regard, the NN models developed by the TOA are used. Leveraging the structure of NNs, the MOA can integrate these models as constraints into its own optimization problem. By integrating these constraints alongside its own objective functions and constraints, the MOA can

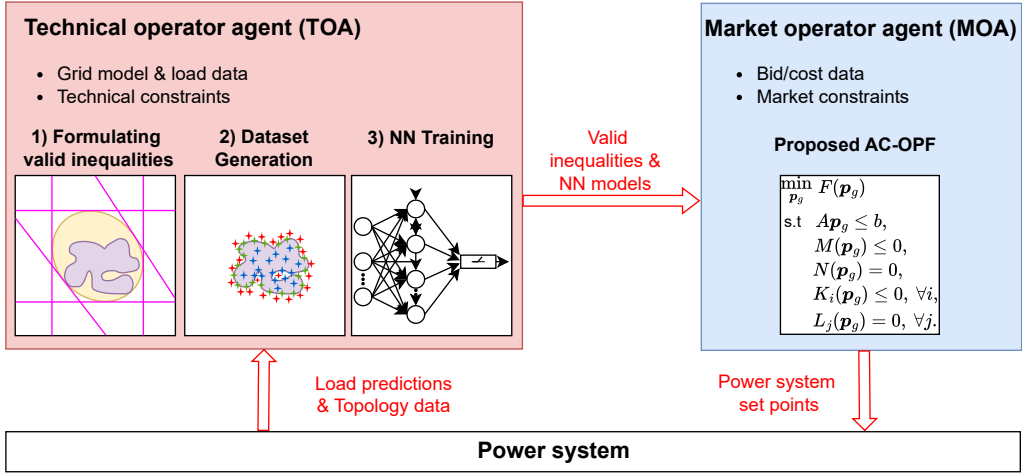


Figure 3.1: The schematic representation of the proposed method.

effectively solve the optimization problem while respecting the limits of the power system. The schematic representation of the proposed method is shown in Fig. 3.1.

Fig. 3.1 illustrates the process flow: The TOA begins by constructing a dataset to train the NN models and formulating valid inequalities. It is important to note that generic cost functions are employed during the training of these models, and an efficient dataset generation technique is designed. Following this, the trained NN models and valid inequalities are transferred to the MOA. Subsequently, the MOA integrates these NN models and valid inequalities as constraints within its own optimization problem.

As a result of this approach, the AC-OPF problem can be effectively addressed by considering both the power system and market constraints. Notably, by transmitting power system data to the MOA in the form of NN models, it guarantees the privacy of sensitive data. Furthermore, there is no requirement to share special constraints and cost data, if any, belonging to the MOA with the TOA, thereby further supporting data privacy. In this proposed method, the roles of the TOA and MOA can be fulfilled by various institutions, including universities, TSOs, DSOs, and consulting firms. Detailed information regarding the proposed method will be provided in the subsequent sections.

3.3 Overview of the Proposed Methodology

In this section, the mathematical formulation of the standard AC-OPF is presented, followed by a detailed examination of the proposed method.

3.3.1 Formulation of the Standard AC-OPF

$$\min_{\mathbf{v}, \boldsymbol{\theta}, \mathbf{p}_g, \mathbf{q}_g} F(\cdot) \quad (3.1a)$$

$$\text{s.t.} \quad G_P(\mathbf{v}, \boldsymbol{\theta}; Y_{bus}) + p_d - C_g \mathbf{p}_g = 0, \quad (3.1b)$$

$$G_Q(\mathbf{v}, \boldsymbol{\theta}; Y_{bus}) + q_d - C_g \mathbf{q}_g = 0, \quad (3.1c)$$

$$G_S(\mathbf{v}, \boldsymbol{\theta}; Y_{bus}) \leq \bar{s}_f, \quad (3.1d)$$

$$\underline{v} \leq \mathbf{v} \leq \bar{v}, \quad \underline{\theta} \leq \boldsymbol{\theta} \leq \bar{\theta}, \quad (3.1e)$$

$$\underline{p}_g \leq \mathbf{p}_g \leq \bar{p}_g, \quad \underline{q}_g \leq \mathbf{q}_g \leq \bar{q}_g, \quad (3.1f)$$

where $\mathbf{v}, \boldsymbol{\theta} \in \mathbb{R}^{n_b}$ are the vectors of bus voltage magnitude, voltage angle, $\mathbf{p}_g, \mathbf{q}_g \in \mathbb{R}^{n_g}$ are the vectors of active and reactive power generations, C_g is the $n_b \times n_g$ *generation connection matrix* such that the element (i, j) is one if generator j is located at bus i , and zero otherwise. $p_d, q_d \in \mathbb{R}^{n_b}$ are the bus active and reactive power demand vectors, respectively, and Y_{bus} is the bus admittance matrix. In (3.1a), $F(\cdot)$ is the objective function. (3.1b) and (3.1c) represent the active and reactive balance equations, in which $G_P(\cdot)$ and $G_Q(\cdot)$ are the relevant functions. In (3.1d), $G_S(\cdot)$ denotes the line apparent power flows, which is bounded by the line flow limit vector $\bar{s}_f$. (3.1e) and (3.1f) imposes the upper and lower bounds on the variables.

3.3.2 Formulation of the Machine Learning-Driven Multi-Agent-Based AC-OPF

In the present section, the proposed privacy-preserving AC-OPF approach are formulated through collaboration between the TOA and MOA. Standard AC-OPF constraints (3.1b) - (3.1f) define a non-convex feasible set in $\mathbf{v}, \boldsymbol{\theta}, \mathbf{p}_g, \mathbf{q}_g$ -space $\mathcal{S}$. However, this formulation encompasses sensitive data, such as network topology. The goal is to find an alternative but equivalent representation of $\mathcal{S}$ without revealing sensitive data, thereby enabling the sharing of this model with the MOA.

Process begins by noting that the objective function $F(\cdot)$ in (3.1) typically depends only on $\mathbf{p}_g$, while $\mathbf{v}, \boldsymbol{\theta}, \mathbf{q}_g$ are involved only in constraints. Given this, the AC-OPF constraints (3.1b) - (3.1f) can be represented as a feasible set $\mathcal{P} \subseteq \mathbb{R}^{n_g}$. The set $\mathcal{P}$ includes all $\mathbf{p}_g$ values for which there exists at least one pair of $(\mathbf{v}, \boldsymbol{\theta}, \mathbf{q}_g)$ that satisfies (3.1b) - (3.1f) [89, 90]. Notably, $\mathcal{P}$ can be thought of as the projection of $\mathcal{S}$ onto the $\mathbf{p}_g$ -dimension. Then, (3.1) can be equivalently represented as:

$$\min_{\mathbf{p}_g} F(\mathbf{p}_g) \quad (3.2a)$$

$$\text{s.t. } \mathbf{p}_g \in \mathcal{P}. \quad (3.2b)$$

For a general network, explicitly representing $\mathcal{P}$ mathematically is a complex task. In this work, the aim is to find an approximation of $\mathcal{P}$, denoted as $\hat{\mathcal{P}}$, which can be represented by a set of constraints that are functions of $\mathbf{p}_g$. These functions are designed to avoid explicit parameterization by sensitive data exclusive to the TOA, allowing them to be shared with the MOA without data privacy concerns. Specifically, it can be hypothesized that $\hat{\mathcal{P}}$ takes the following form:

$$\hat{\mathcal{P}} = \{\mathbf{p}_g \mid A\mathbf{p}_g \leq b, M(\mathbf{p}_g) \leq 0, N(\mathbf{p}_g) = 0\}, \quad (3.3)$$

where, $A\mathbf{p}_g \leq b$ is a system of linear inequalities, which defines a convex polytope in $\mathbb{R}^{n_g}$ encompassing $\mathcal{P}$. The values of the matrix A and vector b are determined using *valid inequalities* derived from the convex relaxation of the AC-OPF, which is detailed in Section 3.4.1.

The function $M(\cdot)$ is designed to represent the feasibility condition of a generation dispatch $\mathbf{p}_g$. In this regard, $M(\cdot)$ is constructed such that $M(\mathbf{p}_g) \leq 0$ holds only if $\mathbf{p}_g \in \mathcal{P}$. To obtain $M(\cdot)$, an ML-based binary classifier model is trained that classifies a given $\mathbf{p}_g$ as *feasible* (Class 0) if $\mathbf{p}_g \in \mathcal{P}$, and *infeasible* (Class 1) otherwise.

Lastly, the function $N(\cdot)$ is aimed at capturing the power balance in the system. Specifically, $N(\mathbf{p}_g) = 0$ holds only if the generation dispatch $\mathbf{p}_g$ meets the total load demand and system losses. To obtain $N(\cdot)$, an ML-based regressor model is trained. This model takes $\mathbf{p}_g^{(-s_g)}$ as the input, where s_g is the slack generator of the system, and outputs $\mathbf{p}_g^{(s_g)}$. In particular, the regressor model can be represented as $N'(\mathbf{p}_g^{(-s_g)}) = \mathbf{p}_g^{(s_g)}$, therefore $N(\mathbf{p}_g) = N'(\mathbf{p}_g^{(-s_g)}) - \mathbf{p}_g^{(s_g)}$.

To represent the functions $M(\cdot)$ and $N(\cdot)$, NN models for classification and regression are proposed, respectively. The reason of this choice is two-fold: on one hand, NNs have high generalization capabilities; on the other hand, the NN models can be exactly represented by a set of mixed-integer linear constraints [91], which enables us to integrate them into a mixed-integer linear programming (MILP) problems.

Indeed, the accuracy of approximation of $\hat{\mathcal{P}}$ for $\mathcal{P}$, particularly in terms of the NN-based functions $M(\cdot)$ and $N(\cdot)$, is contingent upon the quality of the training dataset. In Section 3.4, a rigorous workflow is introduced for creating the dataset consisting of highly informative samples. Using

this workflow, within the proposed methodology, the TOA creates the dataset and trains the models for obtaining $M(\cdot)$ and $N(\cdot)$. As these functions are not explicitly parametrized by sensitive data, but rather by the weights and biases of the NN, they, as well as A and b , can be shared with the MOA without data privacy concerns. Additionally, even if an attempt is made to reconstruct the feasible space by reverse-engineering the NN models, it is important to note that these models are solely based on the active power outputs of the generators. As such, sensitive data, such as load demand, voltage profiles, and system topology, are not directly accessible through these models. Then, the MOA can solve the *machine learning-driven multi-agent-based AC-OPF problem* for the dispatch decisions, which is formulated as:

$$\min_{\mathbf{p}_g} F(\mathbf{p}_g) \quad (3.4a)$$

$$\text{s.t. } A\mathbf{p}_g \leq b, \quad (3.4b)$$

$$M(\mathbf{p}_g) \leq 0, \quad (3.4c)$$

$$N(\mathbf{p}_g) = 0, \quad (3.4d)$$

$$K_i(\mathbf{p}_g) \leq 0, \quad \forall i, \quad (3.4e)$$

$$L_j(\mathbf{p}_g) = 0, \quad \forall j, \quad (3.4f)$$

where, (3.4b)-(3.4d) represent $\hat{\mathcal{P}}$ as given in (3.3). As indicated earlier, the NN-based functions in (3.4b) and (3.4c) can be represented as mixed-integer linear constraints, as detailed in Section 3.5. In (3.4e) and (3.4f), the functions $K_i(\cdot)$ and $L_j(\cdot)$ represent the additional inequality and equality constraints, respectively, defined by the MOA, if any. These constraints, for example, may stem from such as bilateral agreements, contracts, and market rules that the MOA wishes to keep confidential. Thanks to the proposed privacy-preserving formulation, the MOA does not need to disclose these constraints to the TOA and can incorporate them into the dispatch decisions. Furthermore, the MOA can freely define the objective function $F(\mathbf{p}_g)$ without revealing it to the TOA, allowing for flexible and privacy-preserving adjustments based on bid or cost data. Moreover, the MOA can still reach the optimal solution of the freely defined objective function by leveraging the valid inequalities and NN models generated by the TOA, ensuring that the technical constraints are respected while maintaining privacy.

3.4 Valid Inequalities and Dataset Creation

In the present section, the objective is to construct an effective dataset that enhances the performance of the NN models and elevates the accuracy of the proposed methodology. In the AC-OPF

problem, the feasible space has a high-multidimensional non-convex structure and it is therefore quite complex to be analysed [92, 93]. In some networks, disconnected feasible spaces may even emerge [94]. Additionally, besides the non-convex structure of the AC-OPF, non-convexities also arise on the market side [95]. The inherent challenge is that data-driven approaches face significant barriers without a well-suited dataset. Methods developed without a suitable dataset generally map only a small portion of the feasible region, which causes the models to produce accurate results only in these regions [96]. However, for a good ML training process, it's essential to thoroughly map the feasible space and maintain a balanced dataset (feasible/infeasible) [87, 97]. In particular, the errors made by the NN tend to increase near the boundaries, as expected [88]. Therefore, creating a dataset requires high-quality samples to map these boundaries appropriately [98].

Considering the aforementioned challenges, in the present chapter a novel approach is introduced to dataset creation. As detailed in the previous section, the dataset is constructed using only the active power outputs of the generators p_g . The first step involves employing second-order cone programming (SOCP) relaxations to formulate valid inequalities, thereby reducing the search space (see Section 3.4.1). It is important to note that the SOCP relaxation is employed exclusively for generating valid inequalities, while all subsequent related calculations in this study are performed using the non-linear AC-OPF formulation. Subsequently, the n-ball approach is presented to generate feasible samples on the boundary of the $\mathcal{P}$ (see Section 3.4.2). Then, multivariate Gaussian distribution (MGD) is applied to the data points obtained in the previous section to generate samples within the $\mathcal{P}$, including undefined regions of the boundary (see Section 3.4.3). Finally, a perturbation approach is introduced to generate infeasible samples around the boundary by using feasible samples, generated in the previous steps (see Section 3.4.4). The projection of the multi-step dataset creation methodology is outlined in two dimensions in Fig. 3.2.

3.4.1 Generating Valid Inequalities

In the present section, the methodology is illustrated for formulating valid inequalities crucial for the optimization problem (3.4) solved by the MOA. It's worth noting that these valid inequalities are devised by the TOA in the form $Ap_g \leq b$. Indeed, the definition of space by valid inequalities plays a pivotal role in reducing the search space. In [88], the authors show that a significant portion of the search space in the AC-OPF problem comprises infeasible data points, which do not provide significant value. Instead, high-quality data points from the boundary between the feasible and infeasible regions should be prioritized. Therefore, narrowing down the search space using valid inequalities and focusing on the region bounded by these valid inequalities is crucial. This approach facilitates the attainment of quicker and more precise results in the optimization

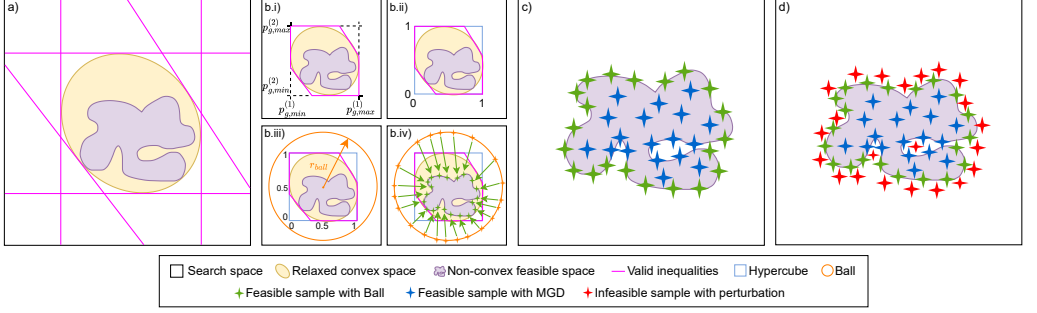


Figure 3.2: The projection of the multi-step dataset creation methodology in two dimensions. a) Firstly, valid inequalities are established using SOCP relaxation, effectively delineating the area within which feasible data points reside. The region outside these valid inequalities is guaranteed to contain infeasible data points, effectively narrowing down the search space. Additionally, $p_{g,min}$ and $p_{g,max}$ are derived from these valid inequalities (see Algorithm 1). b.i) The relaxed convex space is restricted using $p_{g,min}$ and $p_{g,max}$. b.ii) All p_g values are normalized to fall within the range of 0 to 1, effectively confined within a hypercube. b.iii) A ball encompassing the hypercube is constructed, with a center located at 0.5 and a radius of r_{ball} . b.iv) Points are sampled from this ball, and through an optimization, the nearest feasible data points on the boundary of non-convex feasible space are determined (see Algorithm 2). c) Using MGD, feasible samples are acquired from the boundary, which cannot be sampled with the ball approach due to the complex non-convex nature of the feasible space. Additionally, data points within the feasible space are obtained through MGD (see Algorithm 3). d) Feasible data points are perturbed to generate infeasible data points around the boundary (see Algorithm 4).

problem (3.4) addressed by the MOA. To achieve this, the SOCP relaxation technique is employed [99]. It is essential to note that the relaxed feasible space is used exclusively in the formulation of valid inequalities. The detailed process for generating valid inequalities is explained in Algorithm 1.

For this, $A \in \mathbb{R}^{(4n_g+2) \times n_g}$ represents the matrix of coefficients, while $b \in \mathbb{R}^{4n_g+2}$ is the vector of constants. To construct the A matrix and b vector an optimization problem is formulated as shown in (3.5) where $c \in \mathbb{R}^{n_g}$ represents the coefficient vector. This optimization yields the optimal active power output of the generators, denoted as $p_g^* \in \mathbb{R}^{n_g}$ vector, based on the given c . Various linearly independent c vectors are considered, namely e_d , o_d , and u . Here, $e_d \in \mathbb{R}^{n_g}$ denotes the standard unit vector, having a value of one at the $e_d^{(d)}$ element and zeros elsewhere; $o_d \in \mathbb{R}^{n_g}$ is a vector, with a value of zero at the $o_d^{(d)}$ element and ones elsewhere; and $u \in \mathbb{R}^{n_g}$ is the vector of ones.

The A matrix is constructed using $-c^\top$, and the b vector is created using $-c^\top p_g^*$ by leveraging the relaxation technique. When (3.5) is solved according to the SOCP relaxation, the optimal solution p_g^* must satisfy $c^\top p_g^* \leq c^\top p_g$ for all feasible p_g due to the optimality condition of this convex problem. Therefore $-c^\top p_g \leq -c^\top p_g^*$ is a valid inequality for the AC-OPF. It's important to emphasize that there cannot be even a single feasible sample outside the region defined by valid inequalities. In addition, any sample that is infeasible according to relaxation remains infeasible

Algorithm 1 Valid inequalities generation**Input:** Power system data**Output:** $A, b, p_{g,min}, p_{g,max}$

1: Define an optimization problem according to SOCP relaxation

$$\begin{aligned} \min_{\mathbf{p}_g} \quad & c^\top \mathbf{p}_g \\ \text{s.t.} \quad & \text{SOCP relaxation.} \end{aligned} \tag{3.5}$$

2: $A \leftarrow []; b \leftarrow [];$ 3: **for** c in $\{e_d, -e_d, o_d, -o_d\}$ **do**4: **for** $d \leftarrow 1$ to n_g **do**5: Solve (3.5) and obtain p_g^* 6: $A \leftarrow [A; -c^\top]; b \leftarrow [b; -c^\top p_g^*];$ 7: **for** c in $\{u, -u\}$ **do**8: Solve (3.5) and obtain p_g^* 9: $A \leftarrow [A; -c^\top]; b \leftarrow [b; -c^\top p_g^*];$ 10: $p_{g,min} \leftarrow -[b^{(1)} : b^{(n_g)}];$ 11: $p_{g,max} \leftarrow [b^{(n_g+1)} : b^{(2n_g)}];$

for AC-OPF. Thereby, by employing valid inequalities, the search space is significantly reduced. Moreover, the first and second n_g elements of the b vector are used, corresponding to the e_d and $-e_d$ vectors, to derive $p_{g,min} \in \mathbb{R}^{n_g}$ and $p_{g,max} \in \mathbb{R}^{n_g}$ vectors, respectively, according to SOCP relaxation. These $p_{g,min}$ and $p_{g,max}$ vectors are then utilized to tighten the boundaries of $\mathbf{p}_g$. Subsequently, these refined boundaries are employed in the generation of the datasets.

3.4.2 N-ball Approach

After establishing valid inequalities and reducing the search space, the next step is to obtain high-quality samples at the feasible/infeasible boundary of the AC-OPF. This is crucial for effectively defining the boundary and increasing the success of ML models. To achieve this objective, in this section, the focus is on generating feasible data points at the boundary of the $\mathcal{P}$. The n-ball approach is employed for this purpose.

The initial step involves normalizing each $\mathbf{p}_g$ using $p_{g,min}$ and $p_{g,max}$, such that $(\mathbf{p}_g - p_{g,min}) \oslash (p_{g,max} - p_{g,min})$. This normalization ensures that the $\mathcal{P}$ remains confined within a unit hypercube. Subsequently, a ball of dimension n_g is imagined that entirely encloses this unit hypercube. It's essential to note that for the ball to encompass the unit hypercube, its radius r_{ball}

must satisfy $r_{ball} \geq \sqrt{2}/2$, and its center must be at 0.5. This configuration ensures that the ball covers the unit hypercube, thereby ensuring that the $\mathcal{P}$ remains inside the ball. Next, a sample is projected taken from the surface of this n_g -dimensional ball onto the $\mathcal{P}$. This approach enables to sample points from the boundary of the $\mathcal{P}$. To accomplish this, an optimization problem is formulated designed for identifying the closest feasible point on the AC-OPF to the sample taken from the ball. By iterative solving of this problem a specified number of times, a dataset is compiled comprising feasible data points from the boundary. Algorithm 2 describes the creation of this dataset using the n-ball approach.

Algorithm 2 Feasible dataset generation with n-ball

Input: Power system data, $p_{g,min}$, $p_{g,max}$, n_{ball} , r_{ball}

Output: D_{ball}

1: Define an optimization problem according to standard AC-OPF

$$\begin{aligned} \min_{\mathbf{p}_g} \quad & \|p_{g,ball} - \mathbf{p}_g\|_2^2 \\ \text{s.t.} \quad & (3.1b) - (3.1f). \end{aligned} \tag{3.6}$$

2: $D_{ball} \leftarrow []$;

3: **for** $idx \leftarrow 1$ to n_{ball} **do**

4: $p_{g,ball} \leftarrow$ sample a vector of size n_g from $\mathcal{N}(0, 1)$;

5: $p_{g,ball} \leftarrow r_{ball} \times (p_{g,ball} / \|p_{g,ball}\|_2) + 0.5$;

6: $p_{g,ball} \leftarrow p_{g,ball} \odot (p_{g,max} - p_{g,min}) + p_{g,min}$;

7: Solve (3.6) and obtain p_g^*

8: $D_{ball} \leftarrow [D_{ball}; p_g^{*\top}]$;

For this, the process starts by generating a random vector $p_{g,ball} \in \mathbb{R}^{n_g}$ using a standard normal distribution. Given that all $\mathbf{p}_g$ are normalized, it is necessary to normalize $p_{g,ball}$ by dividing its L_2 -norm. Afterwards, $p_{g,ball}$ needs to be scaled to represent the ball surface. For this purpose, this vector is scaled by multiplying it by r_{ball} . Additionally, to ensure proper alignment with the center, this vector is shifted by adding 0.5. This process yields a random point on the surface of the ball. Next, this point is denormalized using $p_{g,min}$ and $p_{g,max}$ to utilize it in the optimization problem. Finally, employing (3.6), the feasible sample $p_g^* \in \mathbb{R}^{n_g}$ at the boundary of the AC-OPF is computed that is closest to $p_{g,ball}$ with the assistance of the L_2 -norm. By iteratively repeating this process n_{ball} times, the dataset $D_{ball} \in \mathbb{R}^{n_{ball} \times n_g}$ is generated, which includes feasible samples from the boundary.

3.4.3 Multivariate Gaussian Distribution

Following the sampling from the boundaries of the AC-OPF using the n-ball approach, the focus in this section is to sample from the boundaries that cannot be captured using the n-ball approach due to the intricate structure of the AC-OPF. Furthermore, the aim is obtain feasible samples from within the $\mathcal{P}$. To achieve this, the multivariate Gaussian distribution (MGD) is employed. A dataset is created consisting of feasible samples using the MGD, leveraging the dataset generated with the n-ball approach and solving an optimization problem similar to the previous section. This process is explained in Algorithm 3.

Algorithm 3 Feasible dataset generation with MGD

Input: Power system data, D_{ball} , n_{mgd}

Output: D_f

1: Define an optimization problem according to standard AC-OPF

$$\begin{aligned} \min_{\mathbf{p}_g} \quad & \|p_{g,mgd} - \mathbf{p}_g\|_2^2 \\ \text{s.t.} \quad & (3.1b) - (3.1f). \end{aligned} \tag{3.7}$$

2: $\mu \leftarrow \text{mean}(D_{ball})$;

3: $\Sigma \leftarrow \text{cov}(D_{ball})$;

4: $D_{mgd} \leftarrow []$;

5: **for** $idx \leftarrow 1$ to n_{mgd} **do**

6: $p_{g,mgd} \leftarrow \text{sample a vector of size } n_g \text{ from } \mathcal{N}(\mu, \Sigma)$;

7: Solve (3.7) and obtain p_g^*

8: $D_{mgd} \leftarrow [D_{mgd}; p_g^{*\top}]$;

9: $D_f \leftarrow [D_{ball}; D_{mgd}]$;

Utilizing the dataset D_{ball} generated in the previous section, the row mean vector $\mu \in \mathbb{R}^{n_g}$ is computed and the covariance matrix $\Sigma \in \mathbb{R}^{n_g \times n_g}$. Subsequently, a random vector $p_{g,mgd} \in \mathbb{R}^{n_g}$ is generated based on μ and Σ . By solving (3.7) with $p_{g,mgd}$ a feasible sample $p_g^* \in \mathbb{R}^{n_g}$ is determined. It's important to note that since this optimization is solved in accordance with the AC-OPF, the results are always feasible. This process is iterated n_{mgd} times to produce the dataset $D_{mgd} \in \mathbb{R}^{n_{mgd} \times n_g}$. Finally, by combining both datasets D_{ball} and D_{mgd} , a dataset $D_f \in \mathbb{R}^{n_f \times n_g}$ is obtained that effectively describes $\mathcal{P}$, comprising high-quality feasible data points. For this, n_f is defined as $n_f = n_{ball} + n_{mgd}$.

3.4.4 Perturbation Approach

As the final step of dataset generation, infeasible data points are produced based on the feasible data points generated in the previous sections. In general, small perturbations are introduced to the feasible data points in the D_f dataset to derive infeasible data points. Given that the D_f primarily comprises feasible points from the boundaries, and these points undergo only minor perturbations, the resulting infeasible points also lie predominantly along the boundary. Consequently, both high-quality feasible and infeasible data points are obtained in the vicinity of the boundary, ensuring its accurate delineation. Additionally, it's worth mentioning that the NN regression model $N(\cdot)$ is employed to enhance the quality of the infeasible data points during the perturbation process. Further details regarding $N(\cdot)$ are provided in Section 3.5.2. This perturbation process and the production of infeasible dataset are summarized in Algorithm 4.

Algorithm 4 Infeasible dataset generation with perturbation

Input: Power system data, n_{ball} , n_{mgd} , D_f , $N(\cdot)$, lim , $step$
Output: D_{inf} , D

```

1:  $D_{inf} \leftarrow []$ ;
2: for  $idx \leftarrow 1$  to  $(n_{ball} + n_{mgd})$  do
3:    $p_{g,f} \leftarrow D_f^{(idx,:)\top}$ ;
4:    $suc \leftarrow 1$ ;  $ctr \leftarrow 0$ ;
5:   while  $suc == 1$  and  $ctr < lim$  do
6:      $p_{g,rand} \leftarrow \text{sample a vector of size } n_g \text{ from } \mathcal{N}(0, 1)$ ;
7:      $p_{g,rand} \leftarrow step \times (p_{g,rand} / \|p_{g,rand}\|_2)$ ;
8:      $p_{g,pert} \leftarrow p_{g,f} + p_{g,rand}$ ;
9:      $p_{g,pert}^{(sg)} \leftarrow N'(p_{g,pert}^{(-sg)})$ ;
10:    if (3.1) with  $p_g = p_{g,pert}$  is infeasible then
11:       $suc \leftarrow 0$ 
12:     $ctr \leftarrow ctr + 1$ ;
13:  if  $suc == 0$  then
14:     $D_{inf} \leftarrow [D_{inf}; p_{g,pert}^\top]$ ;
15:  $D \leftarrow [D_f; D_{inf}]$ ;

```

To generate an infeasible data point, the process starts by creating a random vector $p_{g,rand} \in \mathbb{R}^{n_g}$ through a standard normal distribution. Subsequently, $p_{g,rand}$ is normalized by dividing its L_2 -norm, and determine its magnitude using $step \in \mathbb{R}$. Employing $p_{g,rand}$, the $p_{g,f} \in \mathbb{R}^{n_g}$ vector is perturbed from the D_f , resulting the perturbed vector $p_{g,pert} \in \mathbb{R}^{n_g}$. However, rather than utilizing a completely random vector, the aim is to use a vector that ensures the relationship between the slack generator and non-slack generators is maintained to increase the quality of

the resulting perturbed vector. To achieve this, the NN regression model $N(\cdot)$ is leveraged. By predicting the slack generator $p_{g,pert}^{(sg)}$ using the non-slack generators $p_{g,pert}^{(-sg)}$, as determined by $N(\cdot)$, $p_{g,pert}$ is updated. Subsequently, by solving Equations (3.1) with the modified $p_{g,pert}$, its feasibility status is investigated. If the optimization yields an infeasible result, $p_{g,pert}$ is added to the infeasible dataset $D_{inf} \in \mathbb{R}^{n_{inf} \times n_g}$.

It's worth noting that $p_{g,f}$ is typically perturbed with a small magnitude represented by *step* during the perturbation process. Consequently, $p_{g,pert}$ may always remain feasible. Thus, it's essential to limit the number of iterations to a specific value denoted as $lim \in \mathbb{Z}_+$. Following this iterative process, a balanced dataset $D \in \mathbb{R}^{(n_f+n_{inf}) \times n_g}$ is generated, comprising high-quality data points with a feasible-to-infeasible ratio of approximately one.

3.5 Neural Network Models and Their Transformation to MILP

Now the training of the NN models are discussed for representing the functions $M(\cdot)$ and $N(\cdot)$, as well as the MILP representation of the trained NNs. A feed-forward NN architecture is considered with ReLU activation in the hidden layers. Without loss of generality, single hidden layer networks are used, but the concepts extend naturally to networks with multiple hidden layers.

The NN with an input $\mathbf{z} \in \mathbb{R}^{n_z}$ and an output $\mathbf{y} \in \mathbb{R}$ is represented as follows:

$$\mathbf{h} = \text{ReLU}(W_h \mathbf{z} + \beta_h), \quad (3.8a)$$

$$\mathbf{o} = W_o \mathbf{h} + \beta_o, \quad (3.8b)$$

$$\mathbf{y} = \sigma(\mathbf{o}), \quad (3.8c)$$

where, $W_h \in \mathbb{R}^{n_h \times n_z}$ and $W_o \in \mathbb{R}^{1 \times n_h}$ are the weight matrices for the hidden and output layers, respectively, while n_h is the number of hidden nodes. The bias vectors are $\beta_h \in \mathbb{R}^{n_h}$ and $\beta_o \in \mathbb{R}$ for the hidden and output layers, respectively. The ReLU function is defined as $\text{ReLU}(\cdot) = \max(0, \cdot)$. The activation function σ for the output layer is taken as the sigmoid function for the binary classification task, which is used to obtain $M(\cdot)$, whereas it is taken as a linear function for the regression problem, which is used to obtain $N'(\cdot)$, and therefore $N(\cdot)$.

3.5.1 Training of the Classifier Model

The NN binary classifier is trained to take a generation dispatch vector, $\mathbf{z} = \mathbf{p}_g$, as input and output zero if $\mathbf{p}_g$ is feasible, i.e., $\mathbf{p}_g \in \mathcal{P}$, and one otherwise. To train this network, a training dataset consisting of feasible samples from D_f and infeasible samples from D_{inf} is compiled. Let the index set of the training dataset be denoted as $\mathcal{T}_M$. The loss function L_M , minimized during training, is a weighted binary cross-entropy loss defined as follows:

$$L_M = - \sum_{i \in \mathcal{T}_M} \omega_{10} y_i \log(\hat{y}_i) + \omega_{01} (1 - y_i) \log(1 - \hat{y}_i), \quad (3.9)$$

where $y_i \in \{0, 1\}$ is the true label (i.e., 0 for feasible, and 1 for infeasible instances) and $\hat{y}_i$ is the predicted label for instance $i \in \mathcal{T}_M$. The parameter $\omega_{10} > 0$ represents the weight assigned to incorrectly classifying an infeasible instance as feasible, while $\omega_{01} > 0$ represents the weight for the opposite error.

In this problem context, distinguishing the significance of incorrect predictions by the NN is crucial. Predicting a feasible point as infeasible incurs an economic loss, whereas predicting an infeasible point as feasible can lead to undesirable outcomes in the power grid. To mitigate the risk of identifying an infeasible point as feasible, the weights are set such that $\omega_{10} > \omega_{01}$. Thus, after all these steps, the function $M(\cdot)$ can be constructed.

3.5.2 Training of the Regressor Model

The NN regressor is trained to take a generation dispatch vector with the entry associated with the slack generator removed, $\mathbf{z} = \mathbf{p}_g^{(-s_g)}$, as input and output the power output of the slack generator, $\mathbf{p}_g^{(s_g)}$. To compile the training dataset for the regressor model, only the feasible samples D_f are considered. From these, the rows are extracted corresponding to non-slack generators as the predictors and the slack generator as the target variable. Let the index set of the training dataset be denoted as $\mathcal{T}_N$. The mean squared error loss function L_N , minimized during training, is defined as follows:

$$L_N = \frac{1}{|\mathcal{T}_N|} \sum_{i \in \mathcal{T}_N} (\hat{y}_i - y_i)^2, \quad (3.10)$$

where y_i is the true value of the slack generator's power output and $\hat{y}_i$ is the predicted value for instance $i \in \mathcal{T}_N$. Finally, after the training process the function $N(\cdot)$ is obtained.

3.5.3 Neural Network MILP Representation

Once the NNs are trained and their weights and biases are identified, they are integrated into the optimization problem as constraints (3.4c) and (3.4d). To achieve this, first the approach is employed given in [91], which exploits the piecewise-linear nature of the ReLU function to obtain a mixed-integer linear representation. The set of constraints to represent (3.8a)-(3.8b) is given as follows:

$$W_h \mathbf{z} + \beta_h = \mathbf{h} - \mathbf{h}_-, \quad (3.11a)$$

$$0 \leq \mathbf{h} \leq \bar{\mathbf{h}} \odot \mathbf{u}, \quad (3.11b)$$

$$0 \leq \mathbf{h}_- \leq \bar{\mathbf{h}}_- \odot (1 - \mathbf{u}), \quad (3.11c)$$

$$\mathbf{o} = W_o \mathbf{h} + \beta_o, \quad (3.11d)$$

$$\underline{\mathbf{z}} \leq \mathbf{z} \leq \bar{\mathbf{z}}, \quad \underline{\mathbf{o}} \leq \mathbf{o} \leq \bar{\mathbf{o}}, \quad (3.11e)$$

$$\mathbf{h}, \mathbf{h}_- \in \mathbb{R}^{n_h}, \quad \mathbf{u} \in \mathbb{B}^{n_h}. \quad (3.11f)$$

The constraints (3.11a) to (3.11c) model the ReLU activation. Constraint (3.11a) decomposes the pre-activation output into the positive part $\mathbf{h}$ and the non-activated part $\mathbf{h}_-$. Constraints (3.11b) and (3.11c) ensure that $\mathbf{h}$ and $\mathbf{h}_-$ follow the ReLU behavior based on the binary variables $\mathbf{u}$. Constraint (3.11d) defines the linear transformation of the output layer: Constraints (3.11e) ensure the variables are within specified bounds, and (3.11f) specifies the domains of the variables.

Then, to represent (3.4c), the following constraint is used together with (3.11):

$$\mathbf{o} \leq -\rho, \quad (3.12)$$

where, $\rho \geq 0$ represents a *conservativeness* parameter. In the context of the NN classifier, this parameter is used to ensure that the output $\mathbf{o}$ remains below a certain threshold, thereby providing a safety margin. This margin helps to account for inaccuracies in the NN predictions, ensuring that the generation dispatch remains within safe and feasible limits.

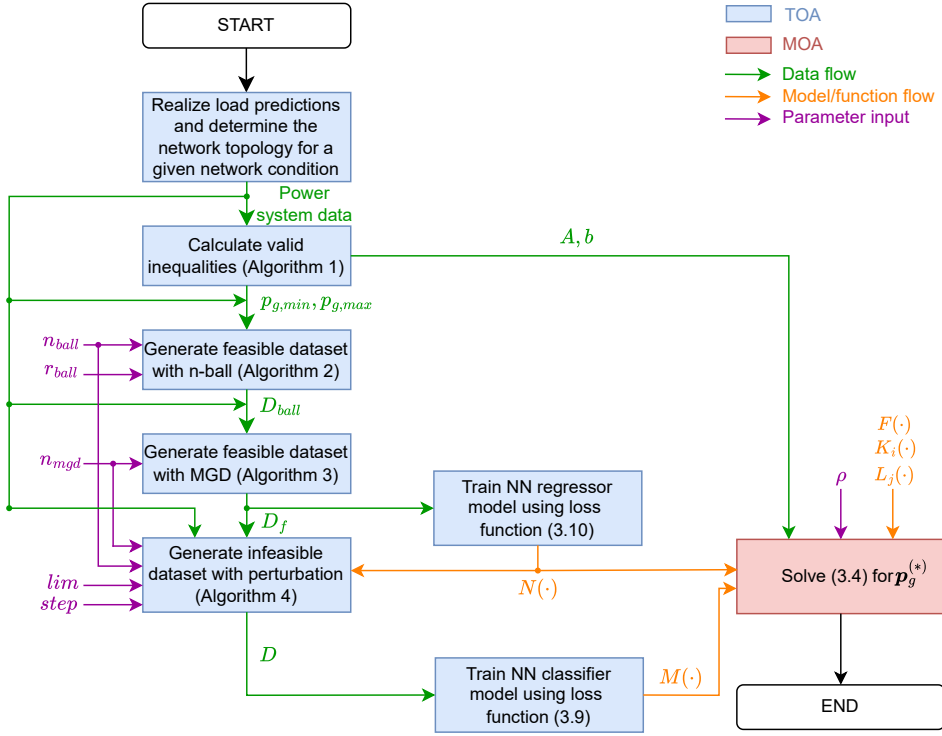


Figure 3.3: Flowchart for the proposed methodology.

In order to represent (3.4d), the following constraint is used together with (3.11)*:

$$o - p_g^{(s_g)} = 0. \quad (3.13)$$

Finally, the flowchart of the entire process, as described in the preceding sections, is presented in Fig. 3.3.

3.6 Case Studies and Discussion

In this study, the performance of the proposed method are evaluate by comparing it to the AC-OPF, which does not consider data privacy, using the 30-, 57-, and 162-bus benchmark models from

*For brevity, the same set of variables is used for the NN models in this explanation. However, in practice, two distinct sets of variables are used: one for the classifier and another for the regressor, with the corresponding weights and biases.

Table 3.1: Summary of the Dataset Creation

Power system	D_{ball}	D_{mgd}	D_f	D_{inf}	D	Feasibility ratio
Case 30	7,500	7,500	15,000	10,478	25,478	58.87%
Case 57	15,000	15,000	30,000	24,160	54,160	55.39%
Case 162	50,000	50,000	100,000	99,834	199,834	50.04%

PGLib-OPF [100]. The power system data is used exactly as it is. The evaluation process begins with the generation of an effective dataset, employing the techniques outlined in the previous sections to acquire high-quality data points on the boundary of the $\mathcal{P}$. Subsequently, NN models are trained to map the $\mathcal{P}$, assessing the quality of their approximations. Finally, the efficacy of the proposed method are tested by conducting comparative analyses against the AC-OPF, using various objective functions and constraints.

MATLAB is selected as the primary simulation environment for the case studies. MATPOWER [101] is utilized with KNITRO solver [102] for the AC-OPF, while GUROBI [103] and YALMIP [104] are employed for the MILP. Furthermore, NN models are trained and tested using TensorFlow/Keras [105, 106].

3.6.1 Dataset Creation Process

Datasets are created by following the steps explained in detail in Section 3.4. Firstly, valid inequalities are generated according to Algorithm 1. When examining the Algorithm 1, it can be observed that the dimension of matrix A is $4n_g + 2 \times n_g$ and vector b is $4n_g + 2$. Therefore, for 30-, 57-, and 162-bus power systems, A has dimensions 26×6 , 30×7 , and 50×12 , while b has dimensions 26, 30, and 50, respectively.

After generating valid inequalities, feasible samples are produced at the boundary of the $\mathcal{P}$ using the n-ball approach, following Algorithm 2. Specifically, r_{ball} is set to 3 and adjust n_{ball} for each case, as detailed in Table 3.1. Next, the multivariate Gaussian distribution is applied to D_{ball} as described in Algorithm 3, generating feasible data points from both the unsampled boundary and within the $\mathcal{P}$. It's worth noting that $n_{ball} = n_{mgd}$, so the size of the D_{mgd} and D_{ball} is equal. Subsequently, D_f is obtained by combining these two datasets containing feasible samples. Finally, infeasible samples are generated by perturbing the feasible samples according to Algorithm 4, creating D_{inf} . lim and $step$ are selected as 5 for this perturbation process. Consequently, D is obtained by combining D_f and D_{inf} . After all these steps, the generated number of data points and the feasibility ratio for each case are summarized in Table 3.1.

Table 3.2: Metrics for the NN Classification

Power system	Accuracy	Recall	Specificity
Case 30	98.64%	98.45%	98.90%
Case 57	90.32%	84.40%	97.76%
Case 162	97.64%	96.17%	99.11%

When examining Table 3.1, it can be observed that a balanced dataset of approximately 50% feasibility ratio is created for each case. The slight deviation from exactly 50% can be attributed to the fact that infeasible data points may not always be reachable during the perturbation process (see Algorithm 4). Therefore, the size of D_{inf} may be smaller than D_f . Additionally, the proposed data generation techniques facilitate the acquisition of high-quality samples around the boundary. Consequently, suitable datasets are generated for effective utilization in ML models with a relatively small number of data points.

3.6.2 ML Training and Approximation Quality

After acquiring the datasets, the ML models are trained to construct the functions $M(\cdot)$ and $N(\cdot)$. Subsequently, the approximation quality of these models are assessed using test sets. Consistent with the typical procedure, the dataset are partitioned into 80% train set and 20% test set for both classification and regression models. As detailed in Section 3.5.1, an NN classification model is employed to classify feasible and infeasible samples. For each case, a single hidden layer and 150 hidden nodes are utilized. Additionally, w_{10} is set to 2.5, 5, and 10 for the respective power systems, while $w_{01} = 1$ for all systems.

Once the model is trained, the approximation quality of the NN classification model is evaluated using accuracy, recall and specificity metrics [107] as shown in Table 3.2. As outlined in Section 3.5.1, a higher weight value is assigned to predicting an infeasible point as feasible, aiming to prevent the model from misclassifying infeasible points. Consequently, the specificity metric, which indicates the prediction performance of infeasible points, is quite high. However, due to these weights, the accuracy and recall metrics are slightly lower. Although this situation may lead to some economic loss, it enhances the feasibility of the proposed method as is designed it for.

The training process for the NN regression model follows a similar approach as explained in Section 3.5.2. a single hidden layer is choose with 500 hidden nodes for all power systems. It should be noted that only D_f is used for training the NN regression model. Subsequently, the model's performance is evaluated using numerical metrics such as root mean square error

Table 3.3: Metrics for the NN Regression

Power system	RMSE	MAE	MAPE
Case 30	0.529	0.373	0.176%
Case 57	0.686	0.334	0.424%
Case 162	0.227	0.119	0.025%

Table 3.4: Comparison of Feasibility Ratio and Average Cost Difference for Different Cost Coefficients

Power system	Feasibility ratio	Average cost difference
Case 30	100%	0.154%
Case 57	100%	0.006%
Case 162	96.40%	0.786%

(RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), as presented in Table 3.3. As demonstrated in the results, the NN regression models achieve performance that are close to 100% accuracy for each power system. Overall, the high performance achieved by both the NN classification and regression models indicates that the functions $M(\cdot)$ and $N(\cdot)$ collectively can effectively capture the behavior of the power systems using only the active power outputs of the generators p_g .

3.6.3 Benchmark Results with Different Cost Coefficients

In this section, the efficiency of the proposed method is assessed by conducting a comparative analysis with the AC-OPF, which does not consider data privacy, in terms of feasibility ratio and total cost. To achieve this, 1,000 different sets of cost coefficients are randomly generated for each power system. This approach allows for a detailed evaluation of the proposed method and represents various special objective functions that the market operator agent may have. Only linear cost coefficients are considered as in PGLib-OPF library [100] without loss of generality. Here, conservativeness is used only for the most complex model, Case 162, as $\rho = 10$.

For each set of randomly generated cost coefficients, first a solution is obtained using the proposed method and subsequently validate its feasibility in the AC-OPF formulation. During the validation process, a tolerance of 1×10^{-2} is applied to avoid numerical errors. Then, the total costs obtained from both methods are compared for the same set of cost coefficients. The comparison results for feasibility ratio and average cost difference are presented in Table 3.4.

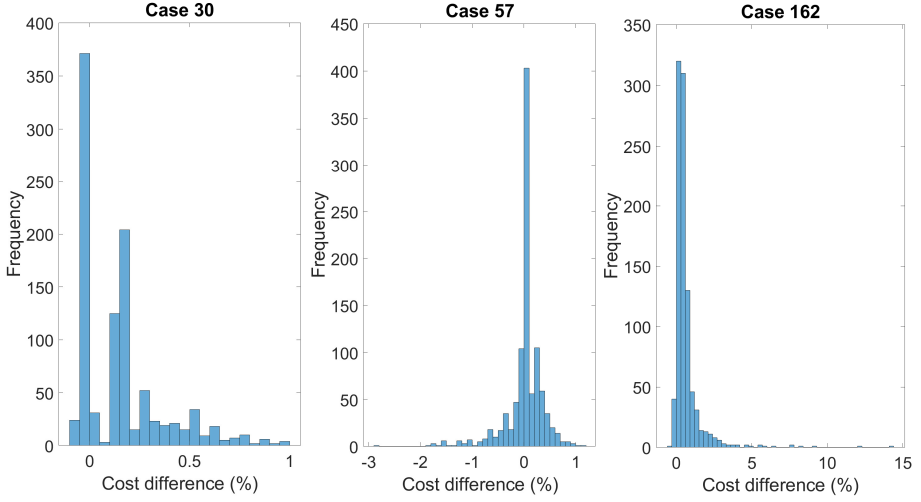


Figure 3.4: The histogram of the total cost differences for different cost coefficients.

As evident from the results, the proposed method achieves a 100% feasibility ratio for Case 30 and 57, while achieving a slightly lower ratio of 96.40% for Case 162. Similarly, the average cost difference between the proposed method and the AC-OPF remains negligible, closely approaching 0% for Cases 30 and 57, and only reaching 0.786% for Case 162. For a more comprehensive analysis of the numerical results, the histogram of the total cost differences is depicted in Fig. 3.4.

Negative values in the histogram indicate instances where the proposed method achieves a lower total cost compared to the AC-OPF. For Case 30 and 57, the maximum cost difference is approximately 1%. Although Case 162 shows some high values, these instances are relatively rare, with only 10 out of 1,000 samples exceeding a 5% cost difference. This situation can be associated with the use of high weights and conservativeness to ensure a high feasibility ratio for this power system. Overall, the proposed method demonstrates performance that is comparable to the AC-OPF, while effectively preserving data privacy.

3.6.4 Benchmark Results with Different Cost Coefficients and Constraints

In this section, similar to the previous one, 1,000 different sets of cost coefficients are randomly generated for each case study. However, in addition to this, different combinations of constraints are introduced. Consequently, the proposed method is compared with the AC-OPF, which does not consider data privacy, across various cost coefficients and constraints. This approach allows

Table 3.5: Comparison of Feasibility Ratio and Average Cost Difference for Different Cost Coefficients and Constraints

Constraint combinations	Feasibility ratio	Average cost difference
(3.14a)	95.20%	0.790%
(3.14a) + (3.14b)	94.70%	0.739%
(3.14a) + (3.14b) + (3.14c)	95.70%	0.793%

to represent special constraints along with the special objective functions that the market operator agent may have, providing a comprehensive evaluation of the method's performance. Rather than focusing on Case 30 and 57, the attention is directed to the more intricate Case 162 to provide a more comprehensive assessment of the proposed method's efficiency. To facilitate this investigation, the following three constraints are set:

$$p_g^{(11)} + p_g^{(12)} - 1000 \geq 0, \quad (3.14a)$$

$$p_g^{(7)} - p_g^{(4)} \geq 0, \quad (3.14b)$$

$$2 \times (p_g^{(2)} + p_g^{(8)}) - p_g^{(1)} \geq 0. \quad (3.14c)$$

In the standard case, the active power lower limits of the 11th and 12th generators, $\underline{p}_g^{(11)}$ and $\underline{p}_g^{(12)}$ are 0 MW. In (3.14a), this is modified by imposing a constraint that their total active power output must be greater than 1,000 MW. Additionally, although the upper limits of the active power for the 7th and 4th generators are $\bar{p}_g^{(7)} = 308$ MW and $\bar{p}_g^{(4)} = 366$ MW, a constraint is added that the $p_g^{(7)}$ must always be greater than $p_g^{(4)}$ in (3.14b). As a final constraint, twice $p_g^{(2)}$ plus $p_g^{(8)}$ is required to exceed $p_g^{(1)}$ in (3.14c), while $\bar{p}_g^{(1)}$, $\bar{p}_g^{(2)}$, $\bar{p}_g^{(8)}$ are 1,147 MW, 451 MW, and 455 MW, respectively. The analyses are performed by adding these constraints incrementally, and the feasibility ratio and average cost difference, in comparison to the AC-OPF, are summarized in Table 3.5.

When the results are examined, it can be observed that, similar to the previous section, a feasibility ratio of around 95% is achieved in each case, and the average cost difference remains approximately 0.8%. This indicates that the proposed method maintains its effectiveness even with the introduction of various additional constraints. The histogram of the total cost differences is presented in Fig. 3.5.

Furthermore, when examining the histogram, it is evident that the distributions are similar across cases and even consistent with the previous section. This suggests that the inclusion of additional constraints does not impact the performance of the proposed method. Consequently, it can

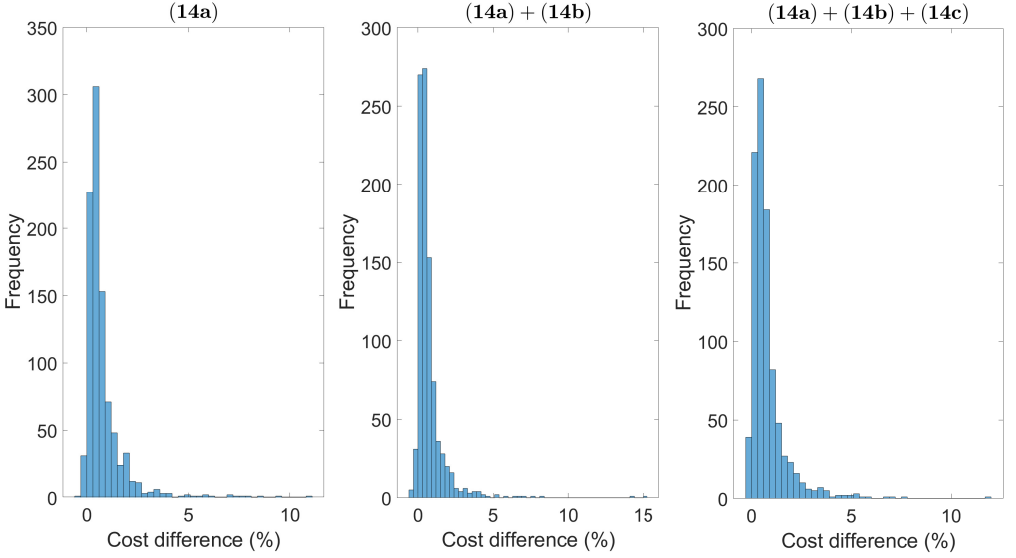


Figure 3.5: The histogram of the total cost differences for different cost coefficients and constraints.

be inferred that various other constraints, beyond those specified, can be incorporated without any degradation in performance. In summary, the proposed method demonstrates acceptable performance compared to the AC-OPF while maintaining data privacy. Moreover, the integration of the trained NN model into the MILP formulation enables flexible utilization of different cost coefficients and additional constraints.

3.7 Summary

The need for interoperability in power system management is becoming increasingly critical to improve operational efficiency and reduce costs. For the effective use of mechanisms such as AC-OPF and redispatch, coordination among institutions is key. However, such coordination is often hindered by strict confidentiality requirements and the reluctance of institutions to share sensitive network or market data. Addressing the first research question of this thesis, this chapter introduced a privacy-preserving AC-OPF framework that enables secure and interoperable decision-making by combining multi-agent system principles with data-driven modeling.

A key contribution of this chapter lies in the development of a data-efficient methodology for generating an effective training dataset. Instead of relying on conventional sampling schemes

that require large-scale data generation to capture the feasible-infeasible boundary of the AC-OPF, the proposed technique makes use of n-ball sampling, multivariate Gaussian distributions, and perturbation-based exploration to create a balanced, boundary-rich dataset with significantly fewer samples. Leveraging this effective dataset, NN models that map feasible space using only active power, achieve high performance. By employing these NN models, sensitive data such as grid topology, current, voltage values, etc., are concealed, ensuring data privacy.

Subsequently, the NN models are embedded into a MILP formulation, enabling the market operator to solve AC-OPF, without access to confidential TSO data. Thanks to the NN structure, this integration enables the optimization problem to be addressed with varying market-based objective functions and constraints. Moreover, the incorporation of valid inequalities further enhances the efficiency of the optimization process by reducing the search space, leading to more accurate and expedited results. As a result, the proposed method allows the AC-OPF to be solved in a manner that respects both market requirements and system security constraints, all without direct data sharing between institutions. Furthermore, it is important to emphasize that the trained NN models can be reused as long as the underlying network model, operating constraints, and relevant operating conditions remain sufficiently consistent with those represented in the training dataset. However, substantial changes in network topology, component parameters, or operating conditions may affect the feasible operating region and therefore require the training dataset and the corresponding NN models to be updated or retrained.

The proposed method is evaluated using 30-bus, 57-bus, and 162-bus benchmark models from the PGLib-OPF. The test results reveal high performance in terms of NN accuracy and overall results, underscoring the efficacy of the proposed method. Furthermore, the results demonstrate that the proposed method is applicable to systems with varying dimensions and characteristics. Most importantly, the chapter demonstrates that accurate system-wide optimization can be achieved without compromising sensitive information, providing a concrete, data-driven answer to the first research question.

4 Privacy-Preserving Analytical Representation and Monetization of DS Flexibility Using the Feasible Operating Region Approach for TSO-DSO Interoperability

The rapid proliferation of DGs and FPU, combined with increasing uncertainties at the distribution level, has significantly increased the operational and economic importance of distribution system flexibility. TSOs increasingly depend on DS flexibility to ensure secure and cost-efficient system operation. However, the effective utilization of this flexibility is severely hindered by data privacy concerns. DSOs are generally unwilling to disclose sensitive internal information, such as network topology, line parameters, and detailed load or generation profiles, thereby preventing TSOs from directly accessing the necessary data to model DS behavior. For this reason, indirect characterization techniques, such as feasible operating region (FOR)-based approaches, have become essential. Yet, deriving and sharing the FOR is especially difficult in large-scale, meshed DSs due to the nonlinear nature of AC power flows and the complexity of the feasible boundary. Moreover, once the FOR is computed, it must be analytically represented and economically valued in a manner that remains tractable and maintains privacy. Addressing these computational, operational, and privacy-related challenges constitutes the core of the second research question examined in this thesis.

To answer these challenges, this chapter introduces a novel, data-driven framework that enables the efficient computation, analytical representation, and economic valuation of DS flexibility while strictly preserving data privacy. The framework builds on a comprehensive PQV-FOR formulation, which extends the conventional two-dimensional PQ capability representation to a three-dimensional PQV domain. This enriched representation explicitly captures boundary voltage variations and the heterogeneous characteristics of FPU, thereby providing a more realistic and detailed description of the DS operational capabilities. To compute the FOR efficiently, the chapter proposes a dedicated AC-OPF-based sampling methodology that is capable of exploring

the high-dimensional feasibility boundary of large, meshed distribution networks while substantially reducing the number of required OPF evaluations. Using the sampled boundary points, an implicit polynomial fitting technique is developed to construct an analytical representation of the FOR. This analytical model can be seamlessly integrated into TSO-level optimization problems without revealing any sensitive DS data. By sharing only this fitted representation, the DSO retains full confidentiality over its topology, operational constraints, and internal states.

In addition to computing and representing the PQV-FOR, the chapter introduces a quadratic flexibility cost function defined directly in the PQV domain. This function enables the monetization of DS flexibility in a manner consistent with the physical system behavior. The analytical FOR and associated cost model are then incorporated into a FOR-based AC-OPF formulation at the transmission level, enabling single-round TSO-DSO coordination. This eliminates the need for iterative communication, reduces coordination overhead, enhances computational scalability, and preserves data privacy by design.

Beyond the overarching contribution of interoperability, the chapter also provides several methodological contributions as follows:

- A realistic characterization of DS flexibility is proposed by moving beyond conventional PQ-only models and explicitly defining flexibility in the three-dimensional PQV domain, capturing the impact of voltage variability.
- An analytical representation of DS flexibility is formulated to enable its direct integration into the TSO's AC-OPF, comprising two complementary components: (i) an implicit polynomial to represent the non-convex FOR boundary, and (ii) an analytical quadratic cost function to monetize operating points, completing the proposed analytical interface.
- A data-driven methodology is developed to construct the analytical models through AC-OPF-based effective sampling strategies. A complementary approach, combining bounding box projection and Fibonacci directions, is used to efficiently capture the FOR boundary, while Latin hypercube sampling generates a homogeneous dataset for the cost function.
- The proposed analytical representations facilitate the realization of a single-round TSO-DSO communication scheme, without iterative coordination or computationally intensive disaggregation. The resulting interface is inherently privacy-preserving, as only analytical functions defined over non-sensitive coupling variables are exchanged, without requiring disclosure of sensitive data such as network topology and load profiles.

The remainder of this chapter is structured as follows. Section 4.1 provides a review of the state of the art. Section 4.2 introduces the overall methodological framework. Section 4.3 describes the sampling strategies used to generate representative boundary points of the FOR. Section 4.4

details the development of the analytical FOR representation and the associated cost functions. Section 4.5 presents the numerical case studies conducted to evaluate the proposed framework. Finally, Section 4.6 summarizes the chapter. This chapter is primarily based on the findings presented in [26].

4.1 State of the Art

Enhanced coordination between TSOs and DSOs is essential for managing the overall power system effectively [108]. However, this coordination is often hindered by stakeholders' concerns regarding data privacy [109]. Therefore, privacy-preserving methodologies are crucial to enable secure and effective TSO-DSO interactions. In this context, the FOR approach, based on the PQ chart, has been proposed to represent the aggregated flexibility of the DS. By sharing this compact PQ representation instead of sensitive data (e.g., topology or customer information), TSOs and DSOs can coordinate effectively [110]. Once the FOR is defined, it can be directly utilized by the TSO.

In the literature, the FOR is computed using various techniques, including geometric methods (e.g., Minkowski sum), random sampling (RS), and optimization-based approaches, particularly those relying on OPF. In geometric methods, the flexibilities of FPU are geometrically aggregated to form the FOR [111, 112]. However, a major limitation is that it does not account for the underlying grid constraints [113]. Alternatively, the RS-based approach involves generating a large number of random operating points, followed by power flow analyses to determine whether each point satisfies grid constraints. The collection of feasible operating points is then used to construct the FOR [114]. While this method accounts for grid limitations, it is computationally intensive, as a substantial number of samples are required [115]. Due to these limitations, optimization-based methods have gained prominence as a more efficient solution [116].

Among OPF-based methods, one commonly used approach involves linearizing the power system model [117, 118]. While this method offers computational efficiency, the linearization of nonlinear power flow equations leads to inaccuracies in the determination of the FOR [116]. Another category is sampling strategies [19], which are typically classified into angle-based and set-point-based approaches. In these methods, the OPF is solved for varying objective function coefficients in the PQ domain [119, 120]. Compared to other techniques, these strategies can more effectively identify the boundary of the FOR [113]. However, most existing studies apply these methods to radial test systems with a limited number of buses and often assume a fixed voltage at the PCC [110, 121]. In practice, the operation of DSs as meshed networks is becoming increasingly prevalent, necessitating that these characteristics be accounted for in analysis. Moreover, to fully exploit the flexibility potential of DSs, the PCC voltage should not be assumed constant;

instead, analyses should be conducted in the three-dimensional PQV domain [122]. Given these requirements, there is a need for innovative approaches to accurately determine the FOR.

After the FOR is generated, a critical question arises: *how can the FORs computed by DSOs be effectively utilized by TSOs in their operational planning?* One common solution involves iterative coordination methods that establish a hierarchical interaction between TSOs and DSOs. In these frameworks, the DSO first determines the FOR, and the TSO independently solves its own OPF problem, subsequently proposing a power exchange operating point at the PCC. The DSO then verifies whether the proposed point lies within its FOR [123, 124, 125]. However, this process often requires a significant number of iterations, leading to high communication overhead. To mitigate this issue, it is desirable for the FOR to be represented analytically, enabling its direct integration into the TSO's OPF formulation. Yet, due to the potentially non-convex nature of the FOR, achieving an accurate and tractable analytical representation remains challenging. As a result, several simplifying assumptions are typically employed in the literature.

For example, [126] approximates the FOR as a polygon and integrates it into a linear OPF through a set of linear inequalities. Similarly, in [127] and [128], the FOR is modeled using a simple rectangular PQ box, which facilitates its incorporation but neglects the underlying complexity of the actual FOR. [129] employs the LinDistFlow model to compute a conservative approximation of the FOR, ensuring that the resulting bounds remain within the actual FOR, thus avoiding the risk of selecting infeasible setpoints, a common issue in convex hull-based approaches. Despite their practical utility, they often lead to over-conservative or inaccurate representations. Therefore, there is a pressing need for novel analytical representations that strike a balance between computational tractability and accurate depiction of the FOR's complex geometry.

Another important consideration is that the FOR solely characterizes the feasibility of operating points. However, for the FOR to be effectively integrated into the TSO's OPF problem, it is also necessary to monetize the aggregated flexibility within the PQV (or PQ) domain. This requires a representation of the cost of flexibility in an analytical form, analogous to the analytical FOR representation. Achieving this is particularly challenging due to the nonlinear nature of power system. One of the earliest efforts in this direction is presented in [130], where monetization is approached through an aggregation and disaggregation process. However, this method relies on a brute-force evaluation of the cost at each point on the FOR, resulting in significant computational burden. Similarly, [131] uses a brute-force method to assign costs across the FOR but does not provide an analytical cost representation.

A notable advancement is found in [132], where the cost of aggregated flexibility is represented analytically. In this study, piecewise linear cost functions are independently derived for active and reactive power. Despite this innovation, the approach is only tested on the 34-bus radial test system and is not validated through integration into an OPF formulation. In [133], the cost is

Table 4.1: Overview of the Related Work.

Work	Uses Full AC-Power Flow Model	Meshed Distribution System	Considered Variables at PCC	Analytical FOR	Analytical Cost	TSO-DSO Coordination Scheme
[123]	✗	✗	P, Q	✓ (Polygon)	✗	Iterative
[124]	✓	✓	P, Q	✗	✗	Iterative
[125]	✗	✗	P, Q	✓ (Polygon)	✗	Iterative
[126]	✗	✗	P, Q	✓ (Polygon)	✗	Iterative
[127]	✓	Not Specified	P, Q	✓ (PQ Box)	✗	Single Round
[128]	✗	Not Specified	P, Q	✓ (PQ Box)	✗	Iterative
[129]	✗	✗	P, Q, V	✓ (Polyhedron)	✓	Not Specified
[130]	✓	✗	P, Q	✗	✗	Not Specified
[131]	✓	✓	P, Q	✗	✗	Not Specified
[132]	✓	✗	P or Q	✗	✓	Not Specified
[133]	✓	Not Specified	Q	✓ (1D Q-Range)	✓	Single Round
This Work	✓	✓	P, Q, V	✓ (Implicit Polynomial)	✓	Single Round

approximated using a continuous quadratic function, but only for the reactive power, with the active power and voltage assumed constant. This significantly simplifies the problem but limits applicability to broader flexibility modeling. Meanwhile, [129] emphasizes that the variables at the PCC, including P, Q, and V, must be considered holistically when defining the cost function. This study fits the cost using a quadratic function, but relies on the simplified LinDistFlow model and is validated only on a small 15-bus radial test system. These studies collectively highlight that, to fully leverage DS flexibility within TSO operations, the cost of the FOR must be represented in a form that is both analytically tractable and compatible with OPF formulations. Furthermore, to ensure practical applicability, these approaches must be extended to more realistic conditions, including meshed network topologies, large-scale systems, and AC-OPF. An overview of the related work is given in Table 4.1.

4.2 Overview of the Proposed Methodology

In this section, the mathematical formulation of the standard AC-OPF for integrated transmission-distribution systems is presented, followed by the proposed FOR-based AC-OPF framework for privacy-preserving DS flexibility utilization. Although the proposed framework is applicable to any type of FPU, these units are referred to simply as DGs throughout the remainder of the chapter for ease of exposition.

4.2.1 Standard AC-OPF for Integrated Transmission-Distribution System

For an integrated transmission-distribution system with n_b total buses, including n_g conventional generator and $n_{b,ts}$ buses in TS, and comprising n_{ds} DSs, where the j -th DS contains $n_{dg,j}$ DGs, the standard AC-OPF can be formulated as follows:

$$\min_{\substack{\hat{v}, \hat{\theta}, \\ \check{p}_g, \check{q}_g, \\ p_{dg,j}, \\ q_{dg,j}}} \sum_{i=1}^{n_g} C_i(\check{p}_g^{(i)}) + \sum_{j=1}^{n_{ds}} \sum_{k=1}^{n_{dg,j}} C_{jk}(p_{dg,j}^{(k)}) \quad (4.1a)$$

$$\text{s.t. } G_P(\hat{v}, \hat{\theta}; \hat{Y}) + \hat{p}_d - K\check{p}_g - \sum_{j=1}^{n_{ds}} H_j p_{dg,j} = 0, \quad (4.1b)$$

$$G_Q(\hat{v}, \hat{\theta}; \hat{Y}) + \hat{q}_d - K\check{q}_g - \sum_{j=1}^{n_{ds}} H_j q_{dg,j} = 0, \quad (4.1c)$$

$$G_{\text{line}}(\hat{v}, \hat{\theta}; \hat{Y}) \leq \hat{l}_{\text{line}, \max}, \quad (4.1d)$$

$$G_{dg,j}(p_{dg,j}, q_{dg,j}) \leq 0, \quad \forall j \in \{1, \dots, n_{ds}\}, \quad (4.1e)$$

$$\hat{v}_{\min} \leq \hat{v} \leq \hat{v}_{\max}, \quad \hat{\theta}_{\min} \leq \hat{\theta} \leq \hat{\theta}_{\max}, \quad (4.1f)$$

$$\check{p}_{g,\min} \leq \check{p}_g \leq \check{p}_{g,\max}, \quad \check{q}_{g,\min} \leq \check{q}_g \leq \check{q}_{g,\max}, \quad (4.1g)$$

$$p_{dg,j,\min} \leq p_{dg,j} \leq p_{dg,j,\max}, \quad \forall j \in \{1, \dots, n_{ds}\}, \quad (4.1h)$$

$$q_{dg,j,\min} \leq q_{dg,j} \leq q_{dg,j,\max}, \quad \forall j \in \{1, \dots, n_{ds}\}, \quad (4.1i)$$

where $\hat{v}, \hat{\theta}, \hat{p}_d, \hat{q}_d \in \mathbb{R}^{n_b}$ denote the bus voltage magnitudes, voltage angles, and active and reactive power demand vectors, respectively. The bus admittance matrix is represented by $\hat{Y} \in \mathbb{R}^{n_b \times n_b}$. Active and reactive power generation vectors for the TS are $\check{p}_g, \check{q}_g \in \mathbb{R}^{n_{b,ts}}$, with the corresponding connection matrix $K \in \mathbb{R}^{n_b \times n_{b,ts}}$, where $K^{(t,v)} = 1$ if this element is in the TS, and zero otherwise. For the j -th DS, $p_{dg,j}, q_{dg,j} \in \mathbb{R}^{n_{dg,j}}$ denote the active and reactive power generation vectors of DGs. Their connection to the network is captured by matrix $H_j \in \mathbb{R}^{n_b \times n_{dg,j}}$ with $H_j^{(m,n)} = 1$ if the n -th DG of DS j connects at bus m , and zero otherwise.

The OPF objective in (4.1a) minimizes the total generation cost, including that of DGs. Let $C_i(\cdot)$ and $C_{jk}(\cdot)$ be the generation cost functions for TS generators at bus i , and the k -th DG in DS j , respectively. Both functions are modeled as standard quadratic cost functions as $C_l(p) = a_l p^2 + b_l p + c_l$, without loss of generality. For notational simplicity, the first n_g

buses are correspond to conventional generators in the integrated system. Equations (4.1b)-(4.1c) enforce active and reactive power balance through functions $G_P(\cdot)$ and $G_Q(\cdot)$. The line flow limits are enforced in (4.1d) via $G_{line}(\cdot)$, bounded by line flow limit vector $\hat{l}_{line,max}$. Additionally, the function $G_{dg,j}(\cdot)$ in (4.1e) characterizes the operating limits of DGs that do not exhibit conventional rectangular PQ capability curves. Constraints on voltages and generations are imposed in (4.1f)-(4.1i).

4.2.2 FOR-Based AC-OPF for Privacy-Preserving DS Flexibility Utilization

A key observation from (4.1) is that utilizing DS flexibility within an OPF framework typically necessitates access to sensitive system information. For instance, the admittance matrix $\hat{Y}$ encodes detailed grid topology, while the vectors $\hat{p}_d$ and $\hat{q}_d$ contain detailed load profiles. As the OPF problem is typically coordinated by the TSO, DSOs are often reluctant to disclose such data due to privacy concerns. To overcome this limitation, a novel FOR-based AC-OPF formulation is proposed that enables the integration of DS flexibility without requiring the exchange of sensitive information between TSOs and DSOs.

In the proposed framework, each DS is represented via two analytical functions: one representing its FOR and the other modeling the associated cost of FOR. These functions are defined over non-sensitive coupling variables that are already exchanged between TSOs and DSOs. Specifically, n_{ds} DSs are considered, where the j -th DS comprises $n_{dg,j}$ DGs, and is connected to a designated transmission buses tb_j , which serve as points of common coupling (PCCs). These PCCs are assumed to be *empty buses*, meaning they do not host any directly connected generation or load units. Each DS j is therefore characterized by functions defined over the coupling variables in the PQV domain, $\mathbf{p}_j, \mathbf{q}_j, \mathbf{v}_j \in \mathbb{R}$, namely active and reactive power flow at the PCC-directed from DS towards TS-and voltage magnitude at the corresponding PCC. For notational convenience, a concatenated vector of coupling variables is defined as $\mathbf{x}_j = [\mathbf{p}_j \quad \mathbf{q}_j \quad \mathbf{v}_j]^\top \in \mathbb{R}^3$. With this, the FOR-based AC-OPF problem can be expressed as follows:

$$\min_{\substack{\check{\mathbf{v}}, \check{\boldsymbol{\theta}}, \\ \check{\mathbf{p}}_g, \check{\mathbf{q}}_g, \\ \mathbf{x}_j}}, \quad \sum_{i=1}^{n_g} C_i(\check{\mathbf{p}}_g^{(i)}) + \sum_{j=1}^{n_{ds}} C_j(\mathbf{x}_j) \quad (4.2a)$$

$$\text{s.t.} \quad G_P(\check{\mathbf{v}}, \check{\boldsymbol{\theta}}; \check{\mathbf{Y}}) + \check{p}_d - \check{\mathbf{p}}_g = 0, \quad (4.2b)$$

$$G_Q(\check{\mathbf{v}}, \check{\boldsymbol{\theta}}; \check{\mathbf{Y}}) + \check{q}_d - \check{\mathbf{q}}_g = 0, \quad (4.2c)$$

$$G_{\text{line}}(\check{\mathbf{v}}, \check{\boldsymbol{\theta}}; \check{\mathbf{Y}}) \leq \check{l}_{\text{line}, \max}, \quad (4.2d)$$

$$\check{v}_{\min} \leq \check{\mathbf{v}} \leq \check{v}_{\max}, \quad \check{\theta}_{\min} \leq \check{\boldsymbol{\theta}} \leq \check{\theta}_{\max}, \quad (4.2e)$$

$$\check{p}_{g, \min} \leq \check{\mathbf{p}}_g \leq \check{p}_{g, \max}, \quad \check{q}_{g, \min} \leq \check{\mathbf{q}}_g \leq \check{q}_{g, \max}, \quad (4.2f)$$

$$FOR_j(\mathbf{x}_j) \leq 0, \quad \forall j \in \{1, \dots, n_{ds}\}, \quad (4.2g)$$

$$\check{\mathbf{p}}_g^{(tb_j)} = \mathbf{p}_j, \quad \check{\mathbf{q}}_g^{(tb_j)} = \mathbf{q}_j, \quad \check{\mathbf{v}}^{(tb_j)} = \mathbf{v}_j, \quad (4.2h)$$

$$x_{j, \min} \leq \mathbf{x}_j \leq x_{j, \max}, \quad \forall j \in \{1, \dots, n_{ds}\}, \quad (4.2i)$$

$$\mathbf{x}_j = \begin{bmatrix} \mathbf{p}_j & \mathbf{q}_j & \mathbf{v}_j \end{bmatrix}^\top, \quad \forall j \in \{1, \dots, n_{ds}\}. \quad (4.2j)$$

Examining (4.2b) - (4.2f), it is evident that only TS-related variables are explicitly included, while DS-related variables are encapsulated within the functions $FOR_j(\mathbf{x}_j)$, as defined in (4.2g). $FOR_j(\mathbf{x}_j)$ characterize the FOR of the DSs, ensuring compliance with internal technical constraints such as voltage and line flow limits. Specifically, $FOR_j(\mathbf{x}_j) \leq 0$ holds if and only if $\mathbf{x}_j$ lies within the FOR. Otherwise, it indicates a violation of DS constraints. Additionally, the function $C_j(\mathbf{x}_j)$ represent the cost associated with the FOR. Furthermore, (4.2h) enforces the physical coupling between TS and DS by ensuring that the coupling variables match at the PCCs. Lastly, (4.2i) defines the bounds on the coupling variables.

Both $FOR_j(\cdot)$ and $C_j(\cdot)$ must be represented in analytical form to ensure their tractability within the AC-OPF framework. Crucially, these functions are constructed solely using non-sensitive coupling variables, thereby preserving data privacy between TSOs and DSOs. A schematic overview of the proposed framework is illustrated in Fig. 4.1. In this architecture, each DSO computes its respective $FOR_j(\cdot)$ and $C_j(\cdot)$ and communicates them to the TSO. The TSO subsequently incorporates these functions into the FOR-based AC-OPF in (4.2). Once an optimal solution within the feasible region is obtained, each DSO independently solves its local AC-OPF to determine the internal dispatch of its DGs (or more broadly, FPU). This process eliminates the need for computationally intensive disaggregation or iterative coordination. As a result, the proposed method enables cost-effective integration of DS flexibility into system-level decision-making, while simultaneously maintaining data confidentiality and complying with the technical constraints of both TSs and DSs.

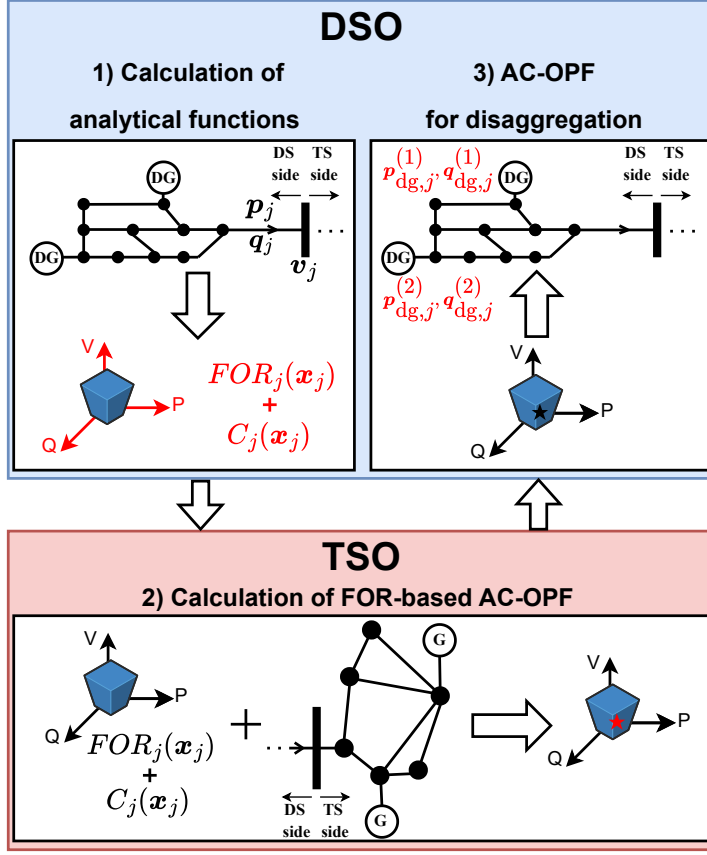


Figure 4.1: Schematic representation of the proposed method. For clarity, only a single DS is illustrated; however, the framework supports multiple DSs. The results obtained at each step are highlighted in red.

4.3 Sampling Strategies for Flexibility Modeling

Constructing the analytical functions $FOR_j(\cdot)$ and $C_j(\cdot)$ requires two distinct datasets because they serve fundamentally different modeling purposes. The FOR representation aims to accurately capture the boundary of the feasible region, which requires dense sampling on the surface of the region. In contrast, the cost functions $C_j(\cdot)$ aims to approximate the cost distribution within the interior of the feasible region, which requires a homogeneous sampling of interior points. Therefore, efficient and computationally tractable strategies are employed to generate both datasets with minimal overhead, as detailed in the following subsections.

It is important to note that the sampling is not performed by independently prescribing active power, reactive power, and voltage at the PCC. Instead, all PQV samples are obtained as feasible

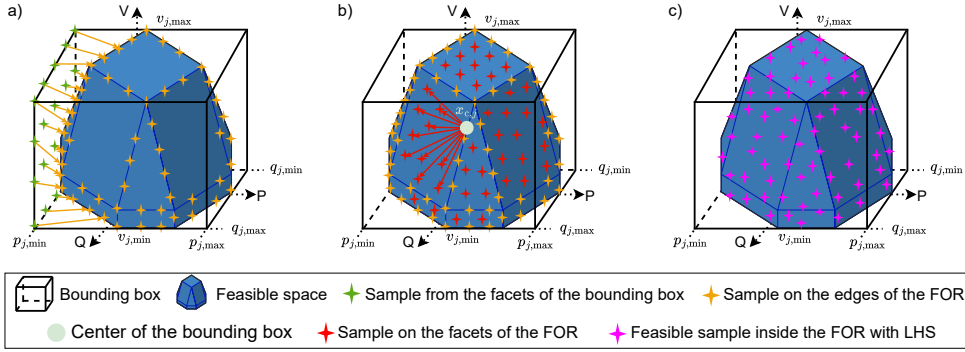


Figure 4.2: The sampling procedure with a) BBPS b) FDS c) LHS. For clarity in the illustration, arrows are indicated only one side.

solutions of AC-OPF problems, in which the physical interdependence among these variables is fully enforced by the AC power flow equations and operational constraints. Accordingly, the constructed PQV datasets represent physically achievable PCC operating points rather than arbitrary combinations of P , Q , and V .

4.3.1 Boundary Sampling for the PQV-FOR

As highlighted in the literature (see Section 4.1), AC-OPF-based methods are particularly effective for identifying the FOR boundary. However, capturing the FOR's complete geometry, which includes both the edges and facets, requires a comprehensive strategy. To this end, an approach is developed for generating FOR boundary data that combines two complementary methods: *Bounding Box Projection Sampling (BBPS)* to effectively capture the edges, and *Fibonacci Direction Sampling (FDS)* to ensure full coverage of the facets. The combination of these methods, which are detailed in the following subsections, yields a well-distributed and representative dataset that accurately characterizes the FOR's complex shape.

4.3.1.1 Bounding Box Projection Sampling (BBPS)

The BBPS method, detailed in Algorithm 5, is the first of two complementary strategies. The process is executed in two phases.

In the first phase, a tight bounding box is established that encloses the FOR. This is achieved by solving six AC-OPF problems to find the minimum and maximum feasible values for each of the three coupling variables (p_j, q_j, v_j) at the PCC, resulting in the limit vectors $x_{j,\min}, x_{j,\max} \in \mathbb{R}^3$.

This targeted approach avoids inefficient sampling of the inherently large infeasible space and focuses the data generation effort on the most relevant region.

Algorithm 5 The BBPS-based Algorithm for Generating FOR Boundary Data

- Input:** Power system data for DS j
Output: $x_{j,\min}, x_{j,\max}, D_{\text{bbps},j}$
- 1: *Phase 1: Determine Bounding Box*
 - 2: **for** $idx \leftarrow 1$ to 3 **do**
 - 3: $x_{j,\min}^{(idx)} \leftarrow \text{Solve } \min \mathbf{x}_j^{(idx)} \text{ s.t. AC-OPF constraints}$
 - 4: $x_{j,\max}^{(idx)} \leftarrow \text{Solve } \max \mathbf{x}_j^{(idx)} \text{ s.t. AC-OPF constraints}$
 - 5: *Phase 2: Project Samples onto FOR Boundary*
 - 6: $D_{\text{bbps},j} \leftarrow [];$
 - 7: **for** $k \leftarrow 1$ to $n_{\text{bbps},j}$ **do**
 - 8: $x_{\text{lhsf},j} \leftarrow \text{sample a vector from the facets of the bounding box defined by } x_{j,\min} \text{ and } x_{j,\max} \text{ using LHS}$
 - 9: Define optimization problem:

$$\min_{\mathbf{x}_j} \|\mathbf{x}_{\text{lhsf},j} - \mathbf{x}_j\|_2^2 \quad (4.3a)$$

s.t. Standard AC-OPF constraints for DS j
 - 10: Solve (4.3) and obtain optimal point x_j^*
 - 11: $D_{\text{bbps},j} \leftarrow [D_{\text{bbps},j}; x_j^{*\top}];$
-

In the second phase, the boundary dataset is generated itself. A set of $n_{\text{bbps},j}$ points, denoted individually as $x_{\text{lhsf},j} \in \mathbb{R}^3$, is sampled on the facets of the previously determined bounding box using Latin hypercube sampling (LHS) [134]. Each external point $x_{\text{lhsf},j}$ is then projected onto the FOR boundary by solving the optimization problem defined in (4.3). This problem finds the closest feasible point $x_j^* \in \mathbb{R}^3$ on the FOR by minimizing the Euclidean (L_2) distance, an approach similar to the previous chapter.

The resulting optimal points x_j^* are collected to form the dataset $D_{\text{bbps},j} \in \mathbb{R}^{n_{\text{bbps},j} \times 3}$. The points generated via BBPS tend to concentrate along the sharp edges of the FOR (see Fig. 4.2a), effectively capturing these critical features.

4.3.1.2 Fibonacci Direction Sampling (FDS)

While the BBPS method effectively captures the edges of the FOR, a complementary approach is needed to sample its facets. To this end, the FDS method is introduced, detailed in Algorithm 6, which ensures comprehensive coverage of the FOR's entire surface (see Fig. 4.2b).

Algorithm 6 The FDS-based Algorithm for Generating FOR Boundary Data

Input: Power system data for DS j , $x_{j,\min}$, $x_{j,\max}$

Output: $D_{\text{fds},j}$

1: Define an optimization problem for a given direction d_k :

$$\max_{t_k} t_k \quad (4.4a)$$

$$\text{s.t. } \mathbf{x}_j = \mathbf{x}_{c,j} + d_k t_k \quad (4.4b)$$

Standard AC-OPF constraints for DS j

2: $D_{\text{fds},j} \leftarrow []$;

3: $\mathbf{x}_{c,j} \leftarrow 0.5 \times (\mathbf{x}_{j,\max} + \mathbf{x}_{j,\min})$

4: $\phi \leftarrow \pi(3 - \sqrt{5})$ ▷ Golden angle for Fibonacci lattice

5: **for** $k \leftarrow 0$ to $n_{\text{fds},j} - 1$ **do**

6: $d_k \leftarrow \begin{bmatrix} r \cos(\theta) \\ r \sin(\theta) \\ z \end{bmatrix}^\top$,
 where $z = 1 - \frac{2k+1}{n_{\text{fds},j}}$, $\theta = \phi \cdot k$, $r = \sqrt{1 - z^2}$

7: Solve (4.4) to obtain the optimal step size t_k^*

8: Calculate the boundary point: $\mathbf{x}_j^* \leftarrow \mathbf{x}_{c,j} + d_k t_k^*$

9: $D_{\text{fds},j} \leftarrow [D_{\text{fds},j}; \mathbf{x}_j^{*\top}]$;

The core idea is to cast *virtual arrows* from the FOR's interior to its boundary along systematically chosen directions. The process begins by defining an origin point, $\mathbf{x}_{c,j} \in \mathbb{R}^3$, at the center of the bounding box. From this origin, a set of $n_{\text{fds},j}$ direction vectors, $d_k \in \mathbb{R}^3$, is generated using the Fibonacci lattice method, which ensures uniform angular coverage in three-dimensional space. For each direction d_k , the optimization problem in (4.4) is solved to find the maximum feasible step size, $t_k \in \mathbb{R}$, along that direction. The resulting intersection point with the FOR boundary, $\mathbf{x}_j^* \in \mathbb{R}^3$, is then recorded. This procedure is repeated for all directions, yielding the dataset $D_{\text{fds},j} \in \mathbb{R}^{n_{\text{fds},j} \times 3}$.

Finally, to create a complete representation of the FOR's geometry, the datasets from both sampling methods are combined into a single dataset: $D_{\text{FOR},j} = [D_{\text{bbps},j}; D_{\text{fds},j}]$, where,

$D_{\text{FOR},j} \in \mathbb{R}^{n_{\text{FOR},j} \times 3}$. For this, $n_{\text{FOR},j}$ is defined as $n_{\text{FOR},j} = n_{\text{bbps},j} + n_{\text{fds},j}$. This final dataset captures both the sharp edges and the smooth facets of the FOR with high fidelity, serving as the foundation for the data-driven construction of the analytical model $\text{FOR}(\mathbf{x})$.

4.3.2 Interior Sampling for the Cost of the FOR

To create the cost function, a dataset of feasible operating points from the interior of the FOR is required. Unlike the boundary-focused methods, this step necessitates a homogeneous sampling of the entire FOR. This is achieved using the LHS method, as illustrated in Fig. 4.2c and detailed in Alg. 7.

Algorithm 7 The LHS-based Algorithm for Generating Cost Data

Input: Power system data for DS j , $x_{j,\min}$, $x_{j,\max}$

Output: $D_{\text{cost},j}$, $y_{\text{cost},j}$

- 1: Define an optimization problem for a candidate point $x_{\text{lhsc},j}$:

$$\min C_j = \sum_{k=1}^{n_{\text{dg},j}} C_{jk}(\mathbf{p}_{\text{dg},j}^{(k)}) \quad (4.5a)$$

$$\text{s.t. } \mathbf{x}_j = x_{\text{lhsc},j} \quad (4.5b)$$

Standard AC-OPF constraints for DS j

- 2: $D_{\text{cost},j} \leftarrow []$;
 - 3: $y_{\text{cost},j} \leftarrow []$;
 - 4: $idx \leftarrow 1$;
 - 5: **while** $idx \leq n_{\text{cost},j}$ **do**
 - 6: $x_{\text{lhsc},j} \leftarrow$ sample a vector from the bounding box using LHS
 - 7: **if** (4.5) with $x_{\text{lhsc},j}$ is feasible **then**
 - 8: Obtain the optimal total cost C_j^*
 - 9: $D_{\text{cost},j} \leftarrow [D_{\text{cost},j}; x_{\text{lhsc},j}^\top]$;
 - 10: $y_{\text{cost},j} \leftarrow [y_{\text{cost},j}; C_j^*]$;
 - 11: $idx \leftarrow idx + 1$;
-

The process begins by generating a candidate operating point, $x_{\text{lhsc},j} \in \mathbb{R}^3$, from within the bounding box using LHS. This point is then tested for feasibility by solving the AC-OPF problem defined in (4.5), where the PCC variables are fixed to the candidate's values. If feasible, the candidate point is valid, and the point itself is stored in the feature dataset $D_{\text{cost},j} \in \mathbb{R}^{n_{\text{cost},j} \times 3}$,

and the corresponding optimal total cost, C_j^* , is stored in the target vector $y_{\text{cost},j} \in \mathbb{R}$. This procedure is repeated until the desired number of samples, $n_{\text{cost},j}$, is collected, resulting in a dataset that homogeneously covers the cost characteristics of the FOR interior.

4.4 Construction of the Analytical Functions

In the following, the procedure for constructing the analytical functions $FOR_j(\cdot)$ and $C_j(\cdot)$ is detailed based on the datasets $D_{\text{FOR},j}$ and $D_{\text{cost},j}$, respectively.

Polynomial representations are utilized for several practical and theoretical reasons. First, fitting polynomial functions requires short training times and minimal hyperparameter tuning, making it computationally attractive for large-scale power system applications. Second, once fitted, the function is fully characterized by a set of numerical coefficients. This enables a privacy-preserving exchange of flexibility information while ensuring communication is trivial in size and highly data-efficient. Third, polynomials are smooth and differentiable, which makes them well-suited for gradient-based optimization solvers used in AC-OPF. This ensures numerical stability and seamless integration into the OPF formulation while maintaining computational efficiency. Finally, as detailed in the theoretical insights provided in Section 4.4.3, the use of polynomials is mathematically substantiated by their strong approximation capabilities over bounded domains. This provides a theoretical foundation for the proposed analytical representations, which are capable of capturing complex, non-convex flexibility characteristics across different network configurations.

4.4.1 Implicit Polynomial Representation of the PQV-FOR

For notational brevity, the DS subscript j is omitted in the remainder of this section.

The objective of this subsection is to approximate the boundary of the three-dimensional FOR using a single implicit polynomial function. For this, an implicit polynomial of degree d_{FOR} is used. A point $x = (p, q, v)$ is considered to be on the boundary of the FOR if it satisfies $FOR(x) = 0$. The polynomial is of the form:

$$FOR(\mathbf{x}) = \sum_{\alpha+\beta+\theta \leq d_{\text{FOR}}} a_{\alpha\beta\theta} \mathbf{p}^\alpha \mathbf{q}^\beta \mathbf{v}^\theta. \quad (4.6)$$

The coefficients $a_{\alpha\beta\theta}$ are the unknown parameters. To facilitate a linear algebraic solution, an explicit ordering is established for the monomial terms. A bijective mapping is defined

$\sigma : \{(\alpha, \beta, \theta) \in \mathbb{N}_0^3 \mid \alpha + \beta + \theta \leq d_{\text{FOR}}\} \rightarrow \{1, \dots, K\}$ that uniquely maps each exponent triplet (α, β, θ) to an index s , where, $K = \binom{d_{\text{FOR}}+3}{3}$ is the total number of monomial terms. This mapping allows the coefficients to be arranged into a single column vector $a \in \mathbb{R}^K$, where the s -th element, $a^{(s)}$, corresponds to the coefficient $a_{\alpha\beta\theta}$ such that $\sigma(\alpha, \beta, \theta) = s$.

4.4.1.1 Volumetric Constraint Generation

To ensure that the polynomial defines a true volume, yielding negative values inside the FOR, positive values outside, and values (near) zero on the boundary, the original boundary data matrix is augmented with additional interior and exterior points. This volumetric constraint strategy, inspired by the 3L algorithm for fitting implicit surfaces [135], enforces a clear separation between the inside and outside of the region, which cannot be achieved using boundary points alone.

Let the original data matrix be denoted $D_{\text{bnd}} \equiv D_{\text{FOR}}$. Additional constraint sets are generated by scaling the boundary points relative to their centroid, c , defined as:

$$c = \frac{1}{n_{\text{FOR}}} \sum_{l=1}^{n_{\text{FOR}}} x_l \quad (4.7)$$

where x_l is the l -th row (sample point) of D_{bnd} . From each point x_{bnd} in the rows of D_{bnd} , new points are generated by applying a scaling factor γ . An inner point x_{in} is generated using a shrink factor $\gamma_{\text{in}} < 1$, and one or more sets of outer points $\{x_{\text{out},m}\}$ are generated using growth factors $\gamma_{\text{out},m} > 1$. The transformation is given by:

$$x_{\text{new}} = c + \gamma(x_{\text{bnd}} - c). \quad (4.8)$$

This procedure yields an inner data matrix D_{in} and a set of S outer data matrices $\{D_{\text{out},m}\}_{m=1}^S$, each of size $n \times 3$.

4.4.1.2 Linear System Formulation and Solution

The problem of finding the coefficient vector a is framed as solving an overdetermined system of linear equations. For each constraint data matrix, a corresponding monomial matrix is constructed. Let D' be a generic data matrix. The monomial matrix $M(D') \in \mathbb{R}^{|D'| \times K}$ is constructed such that each row corresponds to a point x_l (the l -th row of D') and each column s corresponds to a unique monomial basis term, ordered according to the mapping σ . The entry for the l -th row and the s -th column is given by the evaluation of the s -th monomial at point x_l as $M(D')^{(l,s)} = p_l^\alpha q_l^\beta v_l^\theta$, where $s = \sigma(\alpha, \beta, \theta)$.

Using this definition, a monomial matrix is constructed for each of our constraint sets: $M_{\text{in}} = M(D_{\text{in}})$, $M_{\text{bnd}} = M(D_{\text{bnd}})$, and $M_{\text{out},m} = M(D_{\text{out},m})$ for $m = 1, \dots, S$. These matrices are vertically concatenated to form a single composite system matrix M_{FOR} . Correspondingly, a constraint vector b_{FOR} is defined by concatenating vectors of target values for each region. These target vectors consist of small constants chosen to enforce the desired sign of the polynomial: $c_{\text{in}} < 0$, $c_{\text{bnd}} \approx 0$, and $c_{\text{out},m} > 0$. The complete linear system is thus formulated as $M_{\text{FOR}}\mathbf{a} = b_{\text{FOR}}$, where

$$M_{\text{FOR}} = \begin{bmatrix} M_{\text{in}} \\ M_{\text{bnd}} \\ M_{\text{out},1} \\ \vdots \\ M_{\text{out},S} \end{bmatrix}, \quad b_{\text{FOR}} = \begin{bmatrix} c_{\text{in}} \cdot \mathbf{1} \\ c_{\text{bnd}} \cdot \mathbf{1} \\ c_{\text{out},1} \cdot \mathbf{1} \\ \vdots \\ c_{\text{out},S} \cdot \mathbf{1} \end{bmatrix}. \quad (4.9)$$

Thereby, $\mathbf{1}$ is a column vector of ones of appropriate dimension. Since this system is overdetermined, an exact solution generally does not exist. The objective is to determine the coefficient vector $\mathbf{a}^*$ that minimizes the squared Euclidean norm of the residual, $\|M\mathbf{a} - b\|^2$. This least-squares solution is found using the Moore-Penrose pseudoinverse, denoted by $M_{\text{FOR}}^\dagger$:

$$\mathbf{a}^* = M_{\text{FOR}}^\dagger b_{\text{FOR}}. \quad (4.10)$$

The resulting vector $\mathbf{a}^*$ contains the coefficients of the implicit polynomial $\text{FOR}(x)$ in (4.6), providing an analytical representation of the FOR.

4.4.2 Quadratic Representation of the Cost of the FOR

To monetize the FOR, the cost function $C(x)$ is modeled as a trivariate quadratic polynomial. This is a special case of the polynomial fitting described in the preceding section, with the degree fixed to $d_{\text{cost}} = 2$. The cost function takes the general form:

$$C(\mathbf{x}) = \sum_{\alpha+\beta+\theta \leq 2} w_{\alpha\beta\theta} \mathbf{p}^\alpha \mathbf{q}^\beta \mathbf{v}^\theta. \quad (4.11)$$

The fitting procedure seeks the coefficients $w_{\alpha\beta\theta}$ that best match the data generated by Algorithm 7. It follows the same linear least-squares formulation previously described. The monomial matrix

$M_{\text{cost}} \in \mathbb{R}^{n_{\text{cost}} \times K}$ is constructed from the feature dataset D_{cost} , where $K = 10$ for a second-degree trivariate polynomial. The vector y_{cost} serves directly as the target for the regression. The coefficient vector $w^* \in \mathbb{R}^{10}$, which contains the ordered coefficients $w_{\alpha\beta\theta}$, is then found by solving the overdetermined system $M_{\text{cost}}w = y_{\text{cost}}$ using the Moore-Penrose pseudoinverse:

$$w^* = M_{\text{cost}}^\dagger y_{\text{cost}}. \quad (4.12)$$

The resulting vector w^* provides the coefficients for the analytical cost function $C(x)$ in (4.11).

4.4.3 Theoretical Justification for the Analytical Representations

Recall that the FOR is defined over the coupling variables x , which are bounded as in (4.2i), and therefore lie in a compact domain. Let $FOR^*(x)$ denote an exact, but generally unknown, scalar function such that

$$FOR^*(x) \leq 0 \iff x \text{ is feasible.} \quad (4.13)$$

Assume that $FOR^*(\cdot)$ is continuous over this domain. Then, by the Stone-Weierstrass theorem [136], for any $\varepsilon > 0$, there exists a multivariate polynomial $FOR(x)$ such that

$$\sup_x |FOR^*(x) - FOR(x)| < \varepsilon. \quad (4.14)$$

This implies that the feasible region can be approximated arbitrarily well, up to an arbitrarily small boundary error, by polynomial sublevel sets of the form $FOR(x) \leq 0$. Therefore, implicit polynomial representations, as adopted in (4.6), provide a theoretically justified and expressive framework for approximating the FOR.

A similar argument holds for the analytical representation of the flexibility cost. Given that the underlying generation costs are continuous and the operating domain is compact, the Stone-Weierstrass theorem ensures that the aggregated cost function can be approximated arbitrarily well by a polynomial proxy $C(x)$. This provides a theoretical justification for using a polynomial cost model, while the choice of a trivariate quadratic form reflects a trade-off between approximation accuracy and computational tractability.

4.4.4 Remarks

In the following, several remarks are provided to clarify key modeling assumptions and discuss practical considerations related to the construction of the analytical functions.

4.4.4.1 Dependence of PQV Variables

The FOR is represented in the PQV domain using the coupling variables $\mathbf{x} = (\mathbf{p}, \mathbf{q}, \mathbf{v})$. These variables are not treated as independent. As described in Section 4.3, all PQV samples are obtained as feasible solutions of AC-OPF problems, ensuring that the physical interdependence among active power, reactive power, and voltage is fully preserved. Consequently, the constructed function $FOR(\mathbf{x})$ represents a projection of the underlying AC-OPF feasible manifold onto the PQV space, rather than an approximation over arbitrary combinations of $(\mathbf{p}, \mathbf{q}, \mathbf{v})$.

4.4.4.2 Interpretation of Voltage in the PQV Domain

Within the proposed formulation, voltage is treated as a controllable variable subject to operational limits, as defined in (4.2i). This is consistent with its role in AC-OPF, where voltage is regulated within a safe operating range (e.g., 0.95-1.05 p.u.). The focus of this work is on feasibility and flexibility. Therefore, potential long-term effects associated with operating at extreme voltage levels are not explicitly modeled. If desired, such preferences or penalties can be incorporated into the analytical cost function $C(\mathbf{x})$. While the case studies in this work primarily evaluate the accuracy of the FOR and its cost representation, the proposed analytical structure allows such quality-of-service considerations to be incorporated in a straightforward manner.

4.4.4.3 Choice of Polynomial Order

As discussed in Section 4.4.3, polynomial functions possess strong approximation capabilities over bounded domains. The accuracy of the implicit polynomial representation in (4.6) depends on the selected degree d_{FOR} . In this sense, the polynomial degree can be interpreted as a hyperparameter. Standard hyperparameter tuning procedures, such as cross-validation or validation-based selection, can be employed to identify an appropriate degree for a given system. The impact of the polynomial degree is further examined empirically in Section 4.5.4.

4.4.4.4 Re-computation of the Analytical functions under System Variations

The FOR depends on system operating conditions, including load levels, renewable generation, network topology, and operational constraints. Changes in these factors may alter the shape of the FOR. Typical examples include significant load variations (e.g., peak vs off-peak demand),

fluctuations in renewable generation (e.g., PV intermittency or wind variability), network reconfiguration due to switching actions, and topology changes/equipment outages caused by faults or maintenance activities. Such events directly affect the feasible operating region defined by the underlying AC-OPF constraints. These changes necessitate recomputation of the FOR, including both the boundary dataset D_{FOR} and the corresponding model fitting $FOR(\mathbf{x})$.

In addition, changes in the feasible region also impact the associated flexibility cost. Since the cost dataset is generated from feasible operating points within the FOR, as described in Section 4.3.2, any modification to the FOR requires recomputation of the interior dataset D_{cost} , followed by re-fitting of $C(\mathbf{x})$. Variations in DG cost coefficients may further require generation of a new dataset and recomputation of the cost function.

Despite this, the proposed sampling and fitting procedures are computationally efficient. As demonstrated in Section 4.5.3, the total data generation time, including both boundary and interior sampling, remains within typical operational time frames and can be further accelerated via parallelization. This allows the analytical representations to be computed offline (e.g., day-ahead or intra-day) and updated periodically as system conditions evolve.

4.5 Case Studies and Discussion

The performance of the proposed method is assessed through a comprehensive set of case studies. First, DSs with diverse PQ characteristics are considered, and their FORs and associated costs are approximated using analytical functions. The accuracy, computation time, and sensitivity of these approximations are then evaluated, including the impact of sampling strategies, polynomial degree, and PCC voltage variability. Finally, the analytically represented DSs are integrated into specified TSs, where the proposed FOR-based AC-OPF framework is applied and benchmarked against both a centralized AC-OPF and an iterative privacy-preserving coordination scheme. The overall performance is evaluated in terms of feasibility, optimality, computational efficiency, and communication requirements.

All simulations are conducted in the MATLAB environment, utilizing the MATPOWER toolbox [101] with the KNITRO solver [102] for the AC-OPF problems.

4.5.1 Distribution Systems Specifications

To rigorously evaluate the performance and scalability of the proposed method, case studies are conducted on 33-, 136-, and 533-bus DSs (i.e., $n_{\text{ds}} = 3$). These DSs are augmented with 5, 10, and 20 DGs, respectively. Notably, the voltage at all PCCs is allowed to vary within a range

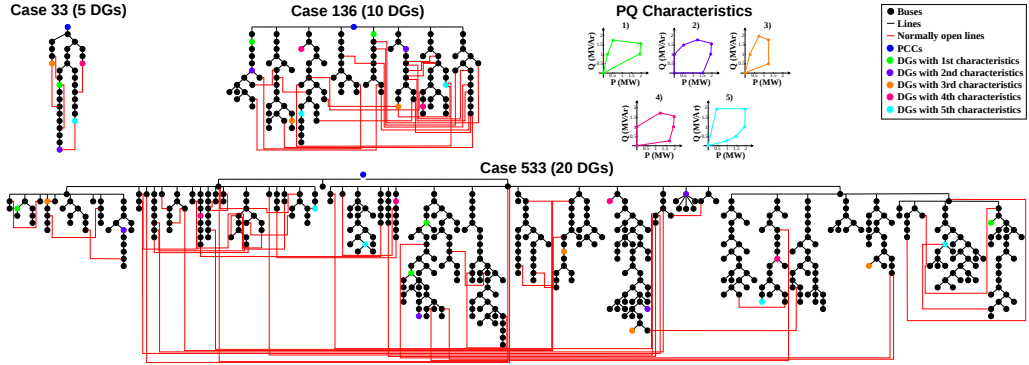


Figure 4.3: Single line diagrams of the DSs and characteristics of the DGs.

of 0.95 to 1.05 p.u., rather than being fixed to a nominal value as commonly assumed in the literature. The configurations and key specifications of these systems are illustrated in Fig. 4.3.

As depicted in Fig. 4.3, all normally open lines in the selected DSs are closed, enabling meshed network topologies. This allows the proposed method to be tested under more complex meshed configurations, in contrast to many existing studies that focus solely on radial networks. Moreover, instead of assuming idealized rectangular PQ characteristics, each DS is equipped with diverse non-ideal convex PQ characteristics, in line with the modeling approach presented in [27]. Specifically, five distinct types of DG characteristics are incorporated to reflect the heterogeneity of FPU. This results in complex, three-dimensional FORs varying with voltage, allowing a comprehensive assessment of the proposed methodology.

4.5.2 Derivation of Analytical FOR and Cost Functions for DSs

In this subsection, analytical approximations of the FOR and the corresponding cost functions for each DS are constructed. These approximations are derived from the datasets $D_{bbps,j}$, $D_{fds,j}$, i.e., $D_{FOR,j}$ and $D_{cost,j}$, which are obtained through the tailored sampling strategies described earlier in Algorithms 5, 6, 7. The number of generated data points for each DS is summarized in Table 4.2. Note that, due to the inclusion of voltage as an additional variable, constructing the three-dimensional FOR requires relatively more data points compared to its two-dimensional counterpart. Furthermore, the datasets are visualized in the three-dimensional PQV domain in Fig. 4.4a and 4.4b. As observed, the FOR and its associated cost characteristics can be effectively captured. As expected, the BBPS method tends to populate points along the edges of the FOR, whereas the FDS method generates samples that lie predominantly on its facets. Merging these

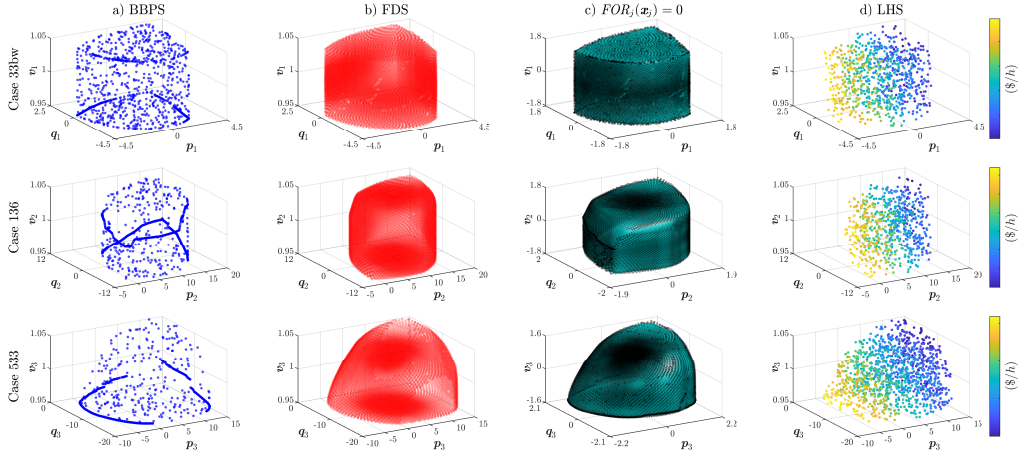


Figure 4.4: a) Dataset generated using BBPS b) Dataset generated using FDS c) Constructed FOR functions (with z-score normalization) d) Cost distribution obtained via LHS. Active and reactive power are expressed in MW and MVAR, respectively, while voltage is expressed in p.u.

Table 4.2: Parameters for the Design of the Functions

DS	n_{bbps}	n_{fds}	n_{cost}	d_{FOR}	d_{cost}	γ_{in}	$\gamma_{\text{out},1}$	$\gamma_{\text{out},2}$	c_{in}	c_{bnd}	$c_{\text{out},1}$	$c_{\text{out},2}$
Case 33bw	10^3	10^4	10^3	8	2	0.999	1.005	1.07	-0.15	0	0.1	0.2
Case 136	10^3	10^4	10^3	8	2	0.999	1.02	1.03	-0.05	0	0.1	0.2
Case 533	10^3	10^4	10^3	8	2	0.999	1.005	1.07	-0.08	0	0.15	0.6

datasets yields an effective representation of the FOR. The visualizations also reveal the emergence of complex FOR geometries, due to meshed topologies and the presence of a large number of DGs.

For the cost modeling, samples are drawn homogeneously from the interior of the FOR using the LHS method (see Fig. 4.4d). The cost function reflects only the generation costs of the DGs; the cost of power imported through the PCC is excluded, as it is already incorporated into the generation costs when the DS model is integrated to the TS. Here, operating points corresponding to higher power export from the DS to the TS, represented by more negative active power values under the adopted sign convention, require increased DG production and therefore incur higher costs.

After generating the datasets, the analytical functions are constructed by representing them in polynomial form. The hyperparameters used for this fitting process are summarized in Table 4.2. Specifically, one inner data matrix D_{in} and two outer data matrices D_{out} (i.e., $S = 2$) are

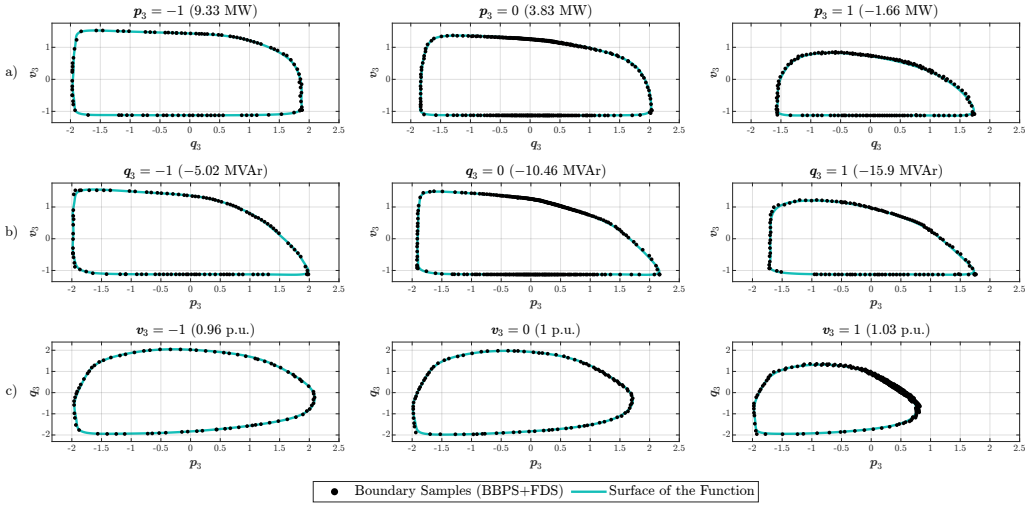


Figure 4.5: Two-dimensional slice plots of the FOR for Case 533, illustrating the sampled boundary points and the surface of the fitted function. Slices are taken at fixed (a) active power, (b) reactive power, and (c) voltage, respectively. All axis values correspond to z-score normalized quantities. Figure titles correspond to the slice value in normalized and actual units.

generated. These hyperparameters are deliberately chosen to ensure that the resulting polynomial defines a conservative feasible region. Such conservatism is essential in power system operations: while classifying a feasible operating point as infeasible results in only a small cost penalty, misclassifying an infeasible operating point as feasible can lead to severe operational and reliability issues. The fitted polynomials, shown in Fig. 4.4c together with the boundary data points, demonstrate that the proposed approach achieves a highly accurate approximation of the FOR with minimal conservativeness, ensuring that the approximation remains entirely within the true FOR. Additionally, to further improve interpretability, two-dimensional slice plots for the Case 533 corresponding to Fig. 4.4c are provided in Fig. 4.5, which illustrate both the sampled boundary points and the fitted implicit surface more clearly. Due to space limitations, these plots are shown only for Case 533.

To evaluate the performance of the functions $FOR_j(x_j)$, random 10^5 samples are generated within the bounding box. These samples naturally contain a specific amount of feasible and infeasible samples. Using these samples, the volume of the FOR is estimated that is captured by the polynomial approximation. For this purpose, the classical confusion matrix framework is adopted. The quality of the approximation is then quantified using standard performance metrics derived from the confusion matrix, and the corresponding results are reported in Table 4.3.

The specificity metric evaluates the ability of the function to correctly identify infeasible points. Since all functions achieve 100% specificity, no infeasible points are misclassified as feasible,

Table 4.3: Performance Metrics of the FOR Functions

Function	Fitting time (s)	Accuracy	Recall	Specificity
$FOR_1(\mathbf{x}_1)$	0.36	99.67%	99.16%	100%
$FOR_2(\mathbf{x}_2)$	0.37	99.68%	98.97%	100%
$FOR_3(\mathbf{x}_3)$	0.39	99.50%	98.14%	100%

Table 4.4: Performance Metrics of the Cost Functions

Function	Fitting time (s)	RMSE	MAE
$C_1(\mathbf{x}_1)$	0.041	0.0021	0.0016
$C_2(\mathbf{x}_2)$	0.045	0.0055	0.0042
$C_3(\mathbf{x}_3)$	0.057	0.0341	0.0266

indicating that the approximation does not allow any infeasible operating points. In contrast, the recall metric measures performance on feasible samples. For example, for $FOR_1(\mathbf{x}_1)$, the recall is 99.16%, meaning that 99.16% of feasible points are correctly identified, while 0.84% of the FOR volume is not captured. This illustrates the trade-off between ensuring complete feasibility and fully representing the volume of the FOR. Overall, when the results are examined, the proposed approximation performs as intended: a small portion of the FOR is sacrificed to reach 100% feasibility. Note that, the impact of this slight volume loss on the overall operational cost is analyzed in the following section.

Furthermore, the cost functions are constructed using the hyperparameters specified in Table 4.2, with their performance metrics summarized in Table 4.4. As shown, the cost functions exhibit high accuracy, indicating that they can effectively approximate the cost of the FOR. The corresponding cost distribution is illustrated in Fig. 4.4d. Also, the fitting times of both $FOR_j(\mathbf{x}_j)$ and $C_j(\mathbf{x}_j)$ are quite low, demonstrating the computational efficiency of the proposed approach. Overall, given that the DSs considered involve a large number of buses and DGs with complex meshed structures, the analytical approximations are obtained with consistently high performance.

4.5.3 Computation Time Analysis of the Proposed Sampling Strategy

As shown in Table 4.3 and 4.4, once the sampling datasets are generated, the fitting of the analytical FOR and cost functions is completed in less than one second. Consequently, the computation

Table 4.5: Sample Generation Time (s) for Different Distribution Systems

DS	Single-Core				56-Core			
	$D_{\text{bbps},j}$	$D_{\text{fds},j}$	$D_{\text{cost},j}$	Total	$D_{\text{bbps},j}$	$D_{\text{fds},j}$	$D_{\text{cost},j}$	Total
Case 33bw	95.91	1044.31	93.60	1233.82	2.53	20.46	2.47	25.46
Case 136	203.55	1997.67	196.39	2397.61	4.28	40.41	4.61	49.30
Case 533	789.01	7577.64	653.87	9020.52	18.81	155.54	25.85	200.20

time of the proposed method is dominated by the sampling stage, while the construction of the analytical functions and their subsequent integration into the TSO's AC-OPF introduce negligible overhead (The computational impact of integrating the analytical functions into the TSO's OPF is separately analyzed in Section 4.5.5).

As discussed earlier, the proposed sampling strategy is specifically designed to reduce computational complexity by focusing on boundary-relevant points and by limiting the number of required AC-OPF solves. Moreover, the three datasets $D_{\text{bbps},j}$, $D_{\text{fds},j}$, and $D_{\text{cost},j}$ are independent and can be generated in parallel. In addition, each dataset generation process is fully parallelizable internally. To evaluate the computation time of the sampling process, all simulations are performed on a server equipped with two Intel Xeon Platinum 8176M CPUs (2.10 GHz, 28 cores each).

Table 4.5 shows the sample generation times under single-core and 56-core execution. Among the datasets, for the largest test system (Case 533), the total generation time is 200.2 s under 56-core execution. This generation time fits within typical operational time frames; for instance, market-based operations are commonly conducted in 15-minute intervals. Furthermore, since the FOR construction can be performed offline (e.g., day-ahead or ahead of market clearing), the computational burden does not pose a practical limitation for operational deployment in such cases.

The results also demonstrate near-linear scalability with respect to parallelization. Specifically, speedups of 48.46, 48.63, and 45.06 are achieved for Cases 33, 136, and 533, respectively, when comparing single-core and 56-core execution. Despite an approximately 16-fold increase in system size from Case 33 to Case 533, the observed speedup scales proportionally, indicating that the proposed method remains computationally tractable for large-scale DSs.

Finally, it should be noted that recalculation of the analytical functions is required when significant changes occur in network topology, operational limits, or forecasted operating conditions. As demonstrated by the computational results, such recalculations can be completed within a relatively short time and well within typical operational time windows. Therefore, the proposed framework is suitable for practical deployment in current power system operations.

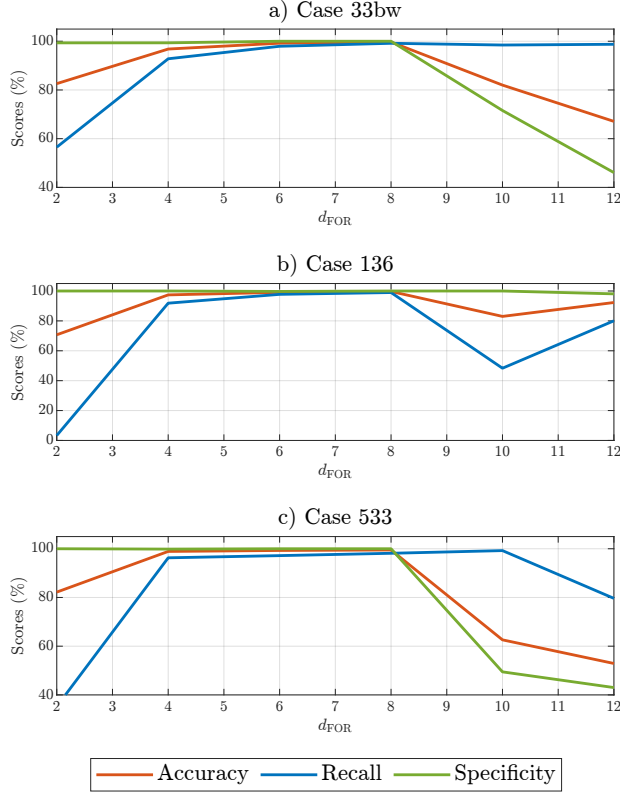


Figure 4.6: Change of performance metrics according to polynomial degree.

4.5.4 Analysis of the Impact of Polynomial Degree

The most influential parameter affecting the performance of the analytical FOR representation is the polynomial degree, denoted by d_{FOR} . This parameter directly affects the expressive capability of the implicit polynomial and, consequently, the accuracy and reliability of the resulting FOR. To systematically assess its impact, a sensitivity analysis is conducted by varying d_{FOR} while keeping all other parameters fixed.

Fig. 4.6 illustrates the evolution of key performance metrics for different values of d_{FOR} across all DSs. As discussed earlier, the specificity metric is of critical importance, since values below 100% indicate that infeasible operating points may be incorrectly classified as feasible. Such misclassification would compromise system security when the FOR is embedded into the TSO-level AC-OPF. Therefore, only polynomial degrees achieving 100% specificity are considered acceptable.

From Fig. 4.6, it can be observed that lower polynomial degrees are insufficient to fully capture the non-convex geometry of the PQV-FOR, resulting in reduced recall and accuracy. Furthermore, excessively high polynomial degrees lead to a degradation in all metrics. Across all DSs, a polynomial degree of $d_{\text{FOR}} = 8$ consistently achieves perfect specificity while simultaneously maximizing accuracy. Consequently, $d_{\text{FOR}} = 8$ is selected as the nominal value for constructing the analytical FOR and is used throughout the subsequent integration of the FOR into the TSO-level AC-OPF formulations.

4.5.5 Incorporation of Analytical DS Representations into TS and Benchmarking Against AC-OPF

In this chapter, the approximated DSs are integrated into the TS as formulated in Equation (4.2). To this end, Case 33bw, Case 136, and Case 533 DSs are integrated with the corresponding TS benchmark models from the PGLib-OPF library, namely Case 30, Case 57, and Case 162, respectively [100]. In the Case 30, buses 11, 16, and 20 are designated as the PCCs; in the Case 57, buses 7, 34, and 48 serve as the PCCs; and in the Case 162, buses 75, 109, and 129 are selected as the PCCs. In this way, three distinct TS-DS test systems are constructed, each consisting of one TS interconnected with three DSs.

After that, simulations are conducted using 1,000 randomly generated sets of cost coefficients. The results are compared against the standard AC-OPF in terms of both total cost and computational time. Fig. 4.7 presents histograms of the cost and time differences, using the standard AC-OPF as the reference. The feasibility ratio is consistently 100%, indicating that the proposed method never produces infeasible solutions. The cost difference remains negligible across all cases, with average deviations of 0.058%, 0.031%, and 0.002%, respectively, where these values indicate that the proposed method results in higher cost compared to the reference AC-OPF. Although the FOR of the DSs is constructed in a relatively conservative manner, its impact on the total cost is observed to be minimal. As the size of the TS increases, the influence of the DSs diminishes, leading to smaller cost deviations. In terms of computational time, the proposed method outperforms the standard AC-OPF in the majority of tests, demonstrating high computational efficiency. The average time difference are -44.22%, -58.11%, -47.62% (-0.23, -0.24, and -0.33 seconds), respectively. Note that, negative values indicate that the proposed method is faster than the standard AC-OPF. These efficiency gains are achieved by representing the complex DSs with compact polynomial approximations.

Note that, within the proposed method, the TSO determines an operating point located on the FOR by considering the overall system cost. The corresponding cost value at this operating point is already available to the DSO through the previously constructed cost function. Consequently,

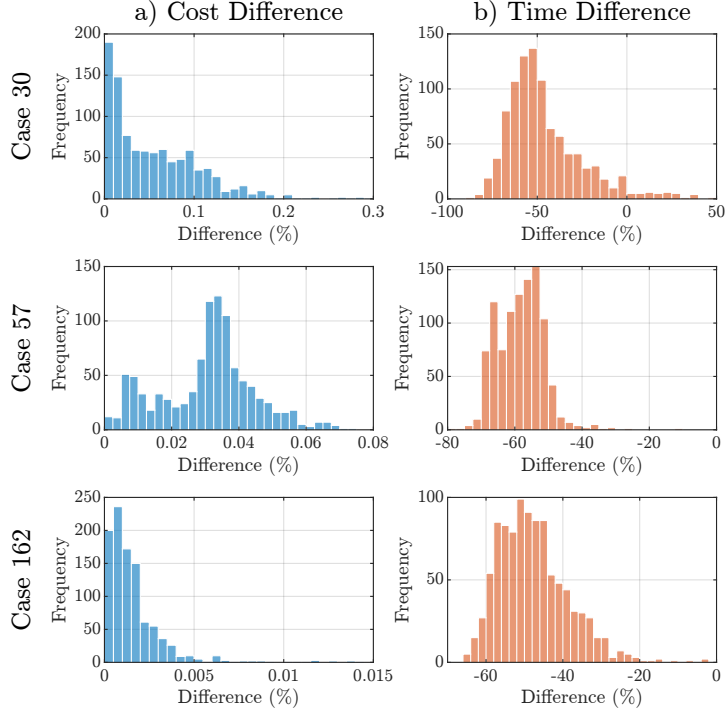


Figure 4.7: Histogram of total cost and computational time differences, with AC-OPF as the reference, for integrated TS-DS. Each TS is combined with three analytically represented DSs (Case 33bw, Case 136, and Case 533).

disaggregation at the DS level can be seamlessly carried out by solving a standard AC-OPF that incorporates this operating point as an input.

The results demonstrate that the proposed analytical representations enable the tractable integration of DS flexibility into the TSO-level OPF without significant loss of optimality. Although the FOR is constructed conservatively, its impact on the total cost remains negligible, even as system size increases. By replacing complex, high-dimensional DS models with compact polynomials, the method achieves a consistent computational speedup compared to a centralized AC-OPF, while maintaining 100% feasibility and preserving DS-level data privacy. While this section focuses on the efficiency of the analytical modeling approach; the impact on coordination and communication requirements is examined in Section 4.5.7.

4.5.6 Analysis of the Impact of the Varying Voltage at the PCCs

Most existing FOR formulations characterize DS flexibility solely in terms of active and reactive power, i.e., in the PQ space, assuming a fixed voltage at the PCC. Under this assumption, voltage-dependent network phenomena are neglected. In practice, however, voltage deviations at the PCC directly influence network losses, reactive power flows, and operational constraints, thereby affecting both the shape and size of the FOR (see Fig. 4.4c and Fig. 4.5c).

Allowing voltage to vary expands the FOR in the PQV space, resulting in a larger flexibility volume. This expanded region enables the TS to exploit DS flexibility more effectively, which in turn can improve overall system operation and reduce operational costs. Consequently, incorporating variable voltage is essential for accurately capturing DS flexibility.

To quantify the impact of PCC voltage variability, a comparative analysis is conducted between the proposed PQV-based FOR and a conventional PQ-based FOR. The analysis is performed on the same TS-DS test systems used in the previous section. For the PQ-based FOR, the PCC voltage is fixed at 1.0 p.u., while all other parameters and modeling assumptions are kept identical to isolate the effect of voltage variability. For each test system, 1,000 randomly generated sets of cost coefficients are used, following the same procedure as in the previous analysis. The cost differences are evaluated by taking the PQV-based FOR as the reference and are summarized in Fig. 4.8.

The results indicate that the mean cost differences are 2.01%, 1.29%, and 2.65% for Cases 30, 57, and 162, respectively, while the maximum observed differences reach 21.53%, 30.16%, and 11.11%. Here, positive values indicate that the PQ-FOR with fixed PCC voltage results in higher costs compared to the PQV-FOR reference. In all cases, feasibility is 100%, indicating that both FOR representations remain operationally valid. However, fixing the PCC voltage consistently leads to higher system costs, demonstrating that restricting voltage results in suboptimal utilization of DS flexibility.

These results show that incorporating voltage variability at the PCCs is crucial not only for accurately representing the FOR geometry but also for improving the economic efficiency of TSO-DSO coordination. The proposed PQV-FOR therefore provides a more realistic and effective flexibility representation than conventional PQ-based approaches.

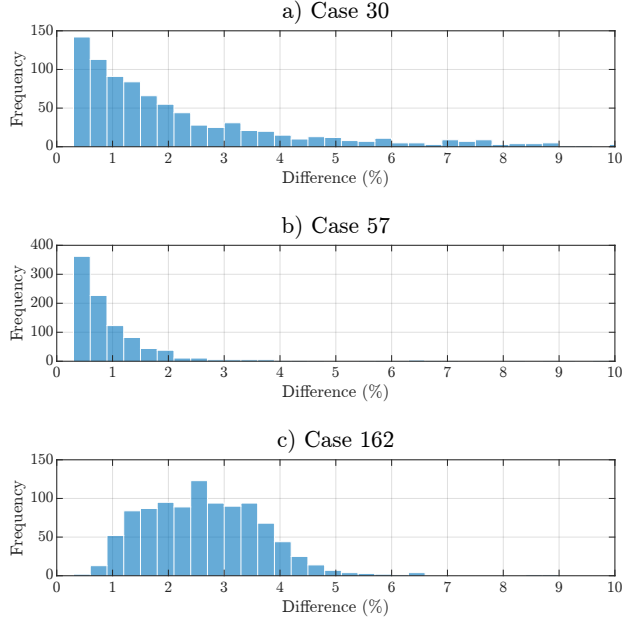


Figure 4.8: Distribution of cost differences between proposed PQV-FOR and PQ-FOR with fixed PCC voltage (PQV-FOR is used as the reference). Outliers are omitted from the plots.

4.5.7 Benchmark Against an Iterative Communication Scheme

The proposed analytical representation framework facilitates a single-round TSO-DSO communication scheme by shifting the bulk of the execution time to offline and highly parallelizable operations, such as sample generation and model fitting by DSOs. The only online execution involves a single operation where the TSO solves the FOR-based AC-OPF. This approach offers structural advantages over iterative communication schemes that require online local computations and continuous data exchange at every iteration, which imposes strict requirements on communication reliability and latency. To quantify these benefits, the proposed framework is benchmarked against an iterative communication scheme based on the alternating direction method of multipliers (ADMM). The ADMM-based scheme is developed using DPLib, an open-source MATLAB-based library for distributed power system optimization [137].

In the ADMM-based scheme, the TSO and DSOs are modeled as multiple regions coupled through tie lines. For each interface, every incident region maintains local copies of the boundary bus voltage magnitude and phase angle, together with auxiliary variables representing neighboring regions. At each iteration, all regions solve local AC-OPF problems in parallel using the most

Table 4.6: Comparison of Standard (Centralized), ADMM-based, and Proposed TSO-DSO Coordination Methods

Method	Offline phase	Online phase				Time	Cost
	Total offline time (s)	Online iterations	Avg. local comput. time (s/iter)	Avg. coord. time (s/iter)	Total online time (s)	Total exec. time (s)	Diff. vs standard AC-OPF
Standard AC-OPF	–	1	–	–	1.06	1.06	–
ADMM	–	130	1.48	0.88	319.95	319.95	0.36%
Proposed	200.59	1	–	–	0.73	201.32	0.07%

recent interface variables and dual multipliers. Afterward, interface voltage magnitudes and phase angles are exchanged and aggregated, and the dual variables are updated to enforce interregional consistency. This iterative exchange continues until interface consistency convergence criteria are satisfied. To reflect a *best case* situation for ADMM, an *oracle* approach is adopted, where dispatch results from 1,000 iterations are evaluated using a centralized AC-OPF to identify the iteration that yields the lowest cost while remaining feasible. The penalty parameter (ρ) is also optimized by a trial-and-error approach. This oracle-based evaluation avoids penalizing ADMM for suboptimal stopping criteria and separates coordination and communication overhead from convergence tuning effects. The ADMM benchmark employs a standard implementation, while more advanced variants with improved convergence properties exist in the literature. The purpose of this comparison is not to optimize ADMM performance, but to highlight differences in online coordination and communication requirements.

The results on the TS-DS test system with Case 162 as the TS are summarized in Table 4.6 and include offline computation time, online computation time, iteration counts, and total execution time under an assumed one-way communication delay of 50 ms. Compared to the ADMM-based scheme, which requires 130 online iterations and incurs a total execution time of 319.95 s, the proposed framework completes online coordination in a single iteration and reduces the online execution time from 319.95 s to 0.73 s, which corresponds to a reduction of more than 99.7%. Despite an offline model construction cost of 200.59 s, the proposed method reduces the total execution time by 118.63 s, which corresponds to a 37.1% reduction, while achieving a smaller cost deviation of 0.07% compared to 0.36% for ADMM.

It is observed that the communication payload per exchange is negligible for both approaches. In ADMM, each online iteration exchanges two interface variables, namely the PCC voltage magnitude and angle, per DS in each direction, which results in $4n_{\text{DS}}$ double-precision scalars per iteration. In contrast, the proposed framework exchanges $\left(\binom{d_{\text{FOR}}+3}{3} + 10\right)n_{\text{DS}}$ scalars once in the offline phase to communicate the analytical FOR and cost models, with no iterative online data exchange. For an 8th-degree FOR polynomial, this equals $175n_{\text{DS}}$ scalars. These payloads are on the order of 10 to 1,000 double-precision scalars and do not pose a bandwidth limitation. The key difference lies in the frequency of data exchange required during online coordination.

These results confirm that the proposed framework substantially lowers the online computation and communication burden while preserving solution quality.

4.6 Summary

This chapter developed a comprehensive framework for characterizing, representing, and utilizing DS flexibility in a manner that is computationally scalable and inherently privacy preserving. The proposed methods address emerging operational challenges in modern power systems, in which the increasing penetration of DGs and FPU requires TSOs to rely on flexibility from distribution networks without direct access to their internal data or structure.

A key contribution of this chapter is the introduction of an extended three-dimensional PQV-FOR. By incorporating voltage variability at the PCC and capturing the heterogeneous limits of embedded FPUs, the PQV-FOR offers a significantly richer and more accurate representation of DS flexibility than conventional two-dimensional PQ capability curves. This enhanced formulation reflects the true physical flexibility of distribution networks and enables more reliable system-level decision-making.

To efficiently characterize this high-dimensional region, the chapter proposed an AC-OPF-based sampling strategy tailored to the properties of large-scale, meshed DSs. The combination of bounding box projection sampling and Fibonacci direction sampling enables effective exploration of the feasible boundary with a substantially reduced number of OPF evaluations. This targeted sampling approach addresses long-standing scalability challenges and provides a computationally tractable path toward high-quality FOR estimation.

Based on sampled boundary points, an implicit polynomial fitting technique is developed to obtain a compact analytical representation of the PQV-FOR. The resulting formulation provides a conservative yet sufficiently accurate approximation of the true feasible region. Importantly, this analytical model is fully compatible with optimization-based decision-making, allowing DS flexibility to be incorporated into TSO-level operational problems without exposing network topology, parameters, or internal load patterns. Through this design, the DSO maintains strict privacy, sharing only a fitted analytical function rather than raw system data.

In addition to feasibility characterization, the chapter introduced a quadratic flexibility cost function defined directly in the PQV domain. This analytical cost model captures the marginal effort associated with active and reactive power adjustments across varying PCC voltage levels, enabling an economic valuation of DS flexibility. When combined with the analytical FOR, it forms a complete, optimization-ready flexibility product that can be seamlessly integrated into TSO operations.

To facilitate this integration, a FOR-based AC-OPF coordination mechanism is developed. Under this scheme, the TSO determines its optimal operating point at the PCC using the analytical FOR and cost function provided by the DS. Each DSO then dispatches its internal FPU via its own AC-OPF problem, adhering to the TSO's decision. This single-round communication approach eliminates the need for iterative exchanges, avoids disaggregation complexity, and significantly reduces computational overhead, demonstrating clear advantages in scalability and practicality.

The effectiveness of the proposed framework is validated through extensive case studies involving large-scale, meshed DSs integrated into TSs. The analyses assess the performance of the analytical FOR and cost representations, the impact of PCC voltage variability on flexibility and economic performance, the computation time of the sampling and fitting procedures, the coordination efficiency relative to iterative benchmark and the overall efficiency of the proposed framework against the standard AC-OPF. Across all test cases, the proposed method achieves average cost deviations below 0.06% while delivering computational speedups of up to 58% relative to standard AC-OPF. These results demonstrate that DS flexibility can be effectively integrated into power system operation under realistic modeling and privacy constraints.

Moreover, the proposed method is developed under fixed network topology and operating conditions. Uncertainties in renewable generation and load, as well as topology changes due to faults or maintenance, may alter the shape of the FOR and necessitate its recomputation. However, computational time analyses demonstrate that these recalculations can be executed efficiently, with execution times comfortably fitting within standard operational time frames.

Overall, this chapter presents a unified methodology that enables reliable characterization, economic valuation, and efficient incorporation of DS flexibility into transmission-level decision processes without compromising data privacy. The proposed techniques offer a viable pathway toward enhanced TSO-DSO coordination and represent a significant advancement in the operationalization of DS flexibility in modern power systems.

5 A Novel TSO-DSO Framework for Multi-PCC Operation and Direct Utilization of DS Flexibility

The increasing penetration of DGs, FPU and active consumers at the distribution level is prompting a paradigm shift in power system operation. As flexibility requirements increase at the transmission level, the ability to utilize DS resources in a coordinated and scalable manner has become essential. Existing privacy-preserving coordination approaches, such as FOR-based methods, provide valuable tools, yet they typically operate at a single PCC and require indirect flexibility allocation. These limitations restrict their applicability to emerging multi-PCC distribution networks, where heterogeneous FPU, diverse PQ characteristics, and complex internal constraints must be considered jointly. Enabling the coordinated and direct dispatch of DS flexibility across multiple PCCs, while maintaining strict stakeholder privacy, is a key challenge addressed in this chapter.

This chapter introduces an advanced ML-based coordination framework that overcomes the structural limitations of earlier FOR-driven approaches. The central idea is to shift from aggregate PQ-boundary representations at a single PCC toward ML models that encapsulate the internal technical constraints of DSs using only non-sensitive coupling variables. By training tailored neural networks (NN) on datasets generated via Latin hypercube sampling (LHS), the proposed framework enables accurate modeling of DS feasibility across multiple PCCs without exposing topology, line parameters, or internal load profiles. The NN architecture is specifically designed to capture nonlinear interactions among DGs, FPU, and multi-PCC power flows, providing a tractable surrogate that can be seamlessly embedded into transmission-level optimization problems.

A further contribution lies in the direct utilization of DS flexibility at the transmission level. Rather than exchanging PCC-based FOR descriptions, the proposed approach enables the TSO to determine the dispatch of individual DGs and FPU through a centralized optimization problem that incorporates ML-derived DS models. To reflect the diversity of real-world flexibility resources, the framework accounts for FPU with non-rectangular PQ characteristics and adapts

the LHS-based dataset generation process accordingly. This delivers a scalable, adaptable, and operationally realistic representation of DS flexibility, enabling enhanced coordination across multiple DSs and multiple PCCs.

Taken together, the developments in this chapter directly address the third research question of the thesis by demonstrating how DS flexibility can be leveraged across multiple PCCs through direct dispatch of DGs and FPU, while preserving data privacy and ensuring interoperability.

Beyond the overarching contribution of interoperability, the chapter also provides several methodological contributions as follows:

- Direct integration of FPU into the OPF problem, enabling DG dispatch within a single communication round, thus eliminating the need for an additional disaggregation.
- Effective incorporation of DS flexibility in complex meshed systems, accommodating scenarios with multiple DSs and multiple PCCs, including treating PCC voltage as a variable to enhance DS flexibility.
- Broad adaptability, allowing integration of FPU with diverse PQ characteristics.
- The novel NN architecture enables efficient and accurate representation of the DSs' feasible operating region, improving computational performance.

The rest of this chapter is structured as follows. Section 5.1 reviews the state of the art. Section 5.2 introduces the proposed coordination methodology. Section 5.3 presents the dataset generation strategy. Section 5.4 describes the construction of the ML models that capture the technical constraints of DSs. Section 5.5 evaluates the performance of the proposed approach through a series of benchmark studies. Finally, Section 5.6 summarizes the main findings and discusses their implications. This chapter is primarily based on the findings presented in [27].

5.1 State of the Art

The inherent fluctuations associated with DGs complicate the management not only of the DS but also of the TS [138]. On the other hand, the flexibility provided by DSs can be effectively leveraged for the provision of ancillary services, contributing to the stability and reliability of the entire power system [10]. As the number of FPU continues to rise, the necessity for effectively managing these flexibilities is becoming increasingly critical. However, the effective utilization of these flexibilities necessitates a high level of coordination between TSOs and DSOs [139].

In recent years, increasing the coordination between TSOs and DSOs and the utilization of DSs flexibility in ancillary services have garnered significant attention from researchers, leading to numerous studies focused on developing innovative coordination schemes [108, 140, 141]. These schemes typically require specific data exchanges between TSOs and DSOs in predefined formats. However, despite existing agreements governing such data transfers, the implementation of these coordination schemes in real-world projects faces numerous challenges and barriers [142]. One major issue is the unwillingness of stakeholders, such as TSOs and DSOs, to share essential data [2]. Current bilateral agreements often fail to address key concerns, such as data leakage, which can lead to the unintended disclosure of sensitive information. For instance, coordination schemes may expose DS system topology data (e.g., line parameters) or customer-specific load data, jeopardizing both commercially sensitive information and the privacy of individual customers. These concerns hinder interoperability and pose a significant challenge to the efficient operation of the power system [143].

In this context, differential privacy (DP) has been investigated as a method for protecting sensitive data in power systems [144]. For example, DP has been applied to obscure transmission line and transformer parameters during data exchange in power grids [6]. Similarly, customer load data in distributed OPF has been protected using DP techniques [7]. In this method, noise is added to the data to prevent the exposure of sensitive information. While this approach enhances data protection, the introduction of noise can pose significant challenges in complex optimization algorithms such as OPF, potentially leading to infeasible solutions [145]. This, in turn, limits the effective utilization of FPU potential. As highlighted in [6], additional mechanisms are necessary to maintain high accuracy while preserving data privacy. However, implementing such mechanisms introduces extra computational overhead. Moreover, existing studies primarily focus on protecting specific types of data, without offering comprehensive solutions to safeguard all sensitive data simultaneously.

Another commonly used approach to ensure comprehensive data privacy in TSO-DSO interactions is the distributed OPF method [146]. In this approach, the OPF problem is decomposed into sub-problems to prevent the need for sharing complete grid models. However, it still requires the exchange of sensitive information such as complex voltages and/or active and reactive power flows at tie-lines between neighboring regions. While this method allows for the effective integration of FPUs from DSs into the TSO's OPF, it suffers from several limitations [147]. Firstly, the approach relies on iterative information exchanges between regions to achieve convergence, which significantly increases communication complexity. Secondly, as the number of DSOs in the system grows, the number of iterations and the time required for convergence rise considerably, posing scalability challenges [57]. Additionally, these methods often model FPUs using idealized rectangular PQ characteristics, failing to capture the diversity of real-world PQ capabilities.

Additionally, wide range of approaches focuses on the concept of PQ capability charts to ensure data privacy in TSO-DSO coordination [119, 110]. In this approach, the DSO calculates the aggregated flexibility at the TSO-DSO interface within the PQ domain [148, 149]. This PQ region, often represented as a polygon, defines the feasible operating region (FOR) of the DS [150]. The TSO can then leverage these aggregated flexibilities for power system operations without the need to exchange sensitive data, such as the grid model [151, 118].

In addition to the advantages related to data privacy, the PQ capability chart approach has a key limitation [152]: Specifically, the cost associated with any point on the PQ chart reflects the aggregate costs of various DGs, making it difficult to directly incorporate the cost implications of a TSO's selected point in the analysis [130]. Consequently, an additional disaggregation problem must be addressed to account for the individual costs of DGs [153]. For instance, in [154], a two-level hierarchical optimization scheme is proposed, where DGs are first aggregated, a multi-step OPF is performed, and then an optimization-based disaggregation problem is solved. Similarly, in [155], a top-down disaggregation process across voltage levels, based on a linear OPF model, is introduced and tested on a real distribution system. As illustrated, the PQ capability chart approach necessitates solving the disaggregation problem to effectively utilize aggregated flexibility in ancillary services, which introduces additional workload and requires iterative communication between TSOs and DSOs.

Another important aspect to consider regarding the PQ capability chart approach is the simplifications often employed in the method. In many studies, it is assumed that the DS is connected to the TS through a single point of common coupling (PCC), and radial test systems are utilized [117, 156, 157]. However, in reality, many DSs are operated in a meshed configuration, with multiple PCCs between TSOs and DSOs. Germany is a prominent example of this complexity; its 110 kV grid is meshed, connected to the TS via multiple PCCs, and managed by DSOs [158]. Moreover, a common assumption in the literature is that the voltage at the TSO-DSO interface, i.e., the PCC, remains constant [159]. However, in practical scenarios, the voltage at the PCC fluctuates depending on dispatch decisions. Assuming a constant voltage at the PCC can lead to an inaccurate assessment of DS flexibility potential, ultimately limiting its effective utilization. These simplifications and assumptions hinder the practical application of the PQ capability chart approach in real-world scenarios.

Considering the aforementioned challenges, our previous works [29, 28] introduced an ML-based methodology to integrate DGs located within the DS into the OPF problem, which is solved by the TSO, while maintaining data privacy. Although various coordination schemes exist, ENTSO-E asserts that TSOs hold the primary responsibility for overall system security, while DSOs are tasked with ensuring the secure operation of their respective DSs [160]. In alignment with these responsibilities, the approach involves the DSO developing ML models that encapsulate the

technical constraints of the DS based solely on the active and reactive power outputs of the DGs and the voltage magnitude at the PCC. By training ML models with this limited dataset, which comprising only information already known and shared between the TSO and DSO, commercially (e.g., system topology) and individual (e.g., customer load profiles) sensitive data is inherently protected, as these details are excluded from the dataset used for training. It is important to note that while ML models are generally susceptible to model inversion attacks, the proposed method ensures that even if reverse engineering is applied, no sensitive information is exposed, as the ML models are trained exclusively with non-sensitive data. Once trained, these ML models are transferred to the TSO, which subsequently utilizes them to solve the OPF problem, including the direct determination of DG dispatch within a single communication round. This approach not only guarantees data privacy by enabling the DSO to share only ML models trained on non-sensitive data, but also ensures that the overall system is managed by the TSO in compliance with ENTSO-E's operational framework. Simultaneously, the method considers the technical constraints of both the DS (through ML models) and the TS, facilitating a secure and coordinated operation.

5.2 Overview of the Proposed Methodology

In this section, the mathematical formulation of the standard AC-OPF for integrated transmission-distribution systems with multiple distribution systems is presented. Subsequently, the formulation of the proposed ML-based privacy-preserving AC-OPF is introduced.

5.2.1 Formulation of the Standard AC-OPF

In the present chapter, an integrated power system is considered with a total of n_b buses, comprising a TS with n_g conventional generator, and $n_{b,ts}$ buses, as well as n_{ds} DSs, where the j -th DS contains $n_{dg,j}$ distributed generators (DGs). Note that some DSs have multiple PCCs with the TS. Following this consideration the standard AC-OPF can be defined as follows:

$$\min_{\substack{\hat{\mathbf{v}}, \hat{\boldsymbol{\theta}}, \\ \check{\mathbf{p}}_{\mathbf{g}}, \check{\mathbf{q}}_{\mathbf{g}}, \\ \mathbf{p}_{\text{dg},j}, \\ \mathbf{q}_{\text{dg},j}}} \sum_{i=1}^{n_{\mathbf{g}}} C_i(\check{\mathbf{p}}_{\mathbf{g}}^{(i)}) + \sum_{j=1}^{n_{\text{ds}}} \sum_{k=1}^{n_{\text{dg},j}} C_{jk}(\mathbf{p}_{\text{dg},j}^{(k)}) \quad (5.1a)$$

$$\text{s.t. } G_{\text{P}}(\hat{\mathbf{v}}, \hat{\boldsymbol{\theta}}; \hat{\mathbf{Y}}) + \hat{p}_{\text{d}} - K\check{\mathbf{p}}_{\mathbf{g}} - \sum_{j=1}^{n_{\text{ds}}} H_j \mathbf{p}_{\text{dg},j} = 0, \quad (5.1b)$$

$$G_{\text{Q}}(\hat{\mathbf{v}}, \hat{\boldsymbol{\theta}}; \hat{\mathbf{Y}}) + \hat{q}_{\text{d}} - K\check{\mathbf{q}}_{\mathbf{g}} - \sum_{j=1}^{n_{\text{ds}}} H_j \mathbf{q}_{\text{dg},j} = 0, \quad (5.1c)$$

$$G_{\text{line}}(\hat{\mathbf{v}}, \hat{\boldsymbol{\theta}}; \hat{\mathbf{Y}}) \leq \hat{l}_{\text{line,max}}, \quad (5.1d)$$

$$\hat{v}_{\min} \leq \hat{\mathbf{v}} \leq \hat{v}_{\max}, \quad \hat{\theta}_{\min} \leq \hat{\boldsymbol{\theta}} \leq \hat{\theta}_{\max}, \quad (5.1e)$$

$$\check{p}_{\mathbf{g},\min} \leq \check{\mathbf{p}}_{\mathbf{g}} \leq \check{p}_{\mathbf{g},\max}, \quad \check{q}_{\mathbf{g},\min} \leq \check{\mathbf{q}}_{\mathbf{g}} \leq \check{q}_{\mathbf{g},\max}, \quad (5.1f)$$

$$p_{\text{dg},j,\min} \leq \mathbf{p}_{\text{dg},j} \leq p_{\text{dg},j,\max}, \quad \forall j \in \{1, \dots, n_{\text{ds}}\}, \quad (5.1g)$$

$$q_{\text{dg},j,\min} \leq \mathbf{q}_{\text{dg},j} \leq q_{\text{dg},j,\max}, \quad \forall j \in \{1, \dots, n_{\text{ds}}\}. \quad (5.1h)$$

For clarity and ease of reference, the following notation is used in the present chapter: variables associated with the integrated system (including both TS and DS) are denoted with a hat ($\hat{\mathbf{a}}$), variables associated solely with the TS are denoted with an inverted hat ($\check{\mathbf{a}}$), and variables related exclusively to the DS are presented without a hat ($\mathbf{a}$). For example, $\hat{\mathbf{v}}$, represents the voltage magnitudes of all buses in the integrated system, while $\check{\mathbf{v}}$ refers only to the TS buses.

Following this convention, $\hat{\mathbf{v}}, \hat{\boldsymbol{\theta}}, \hat{p}_{\text{d}}$, and $\hat{q}_{\text{d}} \in \mathbb{R}^{n_{\text{b}}}$ represent the vectors of bus voltage magnitude, voltage angle, active and reactive power demand vectors respectively, for the integrated system, which include both TS and DSs. $\hat{\mathbf{Y}} \in \mathbb{R}^{n_{\text{b}} \times n_{\text{b}}}$ denotes the bus admittance matrix. $\check{\mathbf{p}}_{\mathbf{g}}, \check{\mathbf{q}}_{\mathbf{g}} \in \mathbb{R}^{n_{\text{b},\text{ts}}}$ are the vectors of active and reactive power generation for the TS buses. K is the $n_{\text{b}} \times n_{\text{b},\text{ts}}$ *transmission generation connection matrix* such that the element (t, v) is one if this element is located inside the TS, and zero otherwise. The vectors $\mathbf{p}_{\text{dg},j}, \mathbf{q}_{\text{dg},j} \in \mathbb{R}^{n_{\text{dg},j}}$ correspond to the active and reactive power generation of the DGs in the j -th DS. H_j is the $n_{\text{b}} \times n_{\text{dg},j}$ *distributed generation connection matrix* such that the element (m, n) is one if n -th DG of the j -th DS is located at bus m , and zero otherwise. It is important to note that the size of the vectors $\check{\mathbf{p}}_{\mathbf{g}}$ and $\check{\mathbf{q}}_{\mathbf{g}}$ corresponds to the number of TS buses, $n_{\text{b},\text{ts}}$, while the size of the vectors $\mathbf{p}_{\text{dg},j}$ and $\mathbf{q}_{\text{dg},j}$ corresponds to the number of DGs, $n_{\text{dg},j}$.

Moreover, in (5.1a), the objective function minimizes the total cost of generation dispatch, including DGs. Here, $C_i(\cdot)$ represents the cost of active power generation at bus i , similarly, $C_{jk}(\cdot)$ represents the cost of active power generation for the k -th DG in the j -th DS. Without

loss of generality, a standard quadratic cost function is considered for both functions, expressed as $C_l(\mathbf{p}) = a_l \mathbf{p}^2 + b_l \mathbf{p} + c_l$. Note that, in this integrated system, it is assumed that the first n_g buses are associated with conventional generators for notational convenience. Equations (5.1b) and (5.1c) represent the active and reactive balance equations, where $G_P(\cdot)$ and $G_Q(\cdot)$ are the corresponding functions. In (5.1d), $G_{line}(\cdot)$ denotes the line apparent power flows, which is bounded by the line flow limit vector $\hat{l}_{line,max}$. Finally, (5.1e) - (5.1h) establish the upper and lower bounds for the respective variables.

Examining Equation (5.1), it becomes evident that the utilization of flexibility from DSs in network management requires access to sensitive data for the entire system. For instance, the admittance matrix $\hat{Y}$ encapsulates the topology of the system, while the demand vectors $\hat{p}_d$ and $\hat{q}_d$ contain load data. Typically, since the OPF problem is solved by the TSO, DSOs are reluctant to share such sensitive data with TSOs. To address this issue, in the present chapter, a novel AC-OPF formulation is introduced designed to eliminate the need for sensitive data exchange between TSOs and DSOs. This new formulation allows for the effective use of DS flexibility in power system management while maintaining data privacy.

5.2.2 Formulation of the ML-Based Privacy-Preserving AC-OPF

In the novel AC-OPF formulation, the primary goal is to prevent the exchange of sensitive data between TSOs and DSOs. To achieve this, the DS-related variables and parameters are separated from the integrated system. As previously described, it is assumed that there are n_{ds} distribution systems, and the j -th DS contains $n_{dg,j}$ distributed generators, where $j \in \{1, 2, \dots, n_{ds}\}$. This setup is extended by assuming that each distribution system j is connected to specific buses $\{s_{j,1}, s_{j,2}, \dots, s_{j,r_j}\}$ (i.e., the points of common coupling (PCCs)) of the TS, where r_j denotes the number of PCCs for the j -th DS. These PCCs in the TS are treated as *empty* buses, meaning these buses do not have any directly connected generators or loads.

Accordingly, each DS is modeled at the corresponding PCCs as dependent active and reactive power injections. These injections represent the power flows at the PCCs. For instance, a DS with a single PCC is modeled at that PCC, while a DS with multiple PCCs is represented by separate active and reactive power flows at each respective PCC. This representation depends on the vector $\mathbf{v}_j \in \mathbb{R}^{r_j}$, which consists of the voltage magnitudes at the PCCs $(s_{j,1}, s_{j,2}, \dots, s_{j,r_j})$ of the j -th DS. It also depends on active and reactive power generation vectors of DGs, $\mathbf{p}_{dg,j}$ and $\mathbf{q}_{dg,j}$ for the j -th DS. For convenience, these variables are concatenated into a single vector $\mathbf{x}_j = \begin{bmatrix} \mathbf{v}_j^\top & \mathbf{p}_{dg,j}^\top & \mathbf{q}_{dg,j}^\top \end{bmatrix}^\top \in \mathbb{R}^{n_j}$, where $n_j = r_j + 2n_{dg,j}$. With this setup, the proposed privacy-preserving AC-OPF can be defined as follows:

$$\min_{\substack{\check{\mathbf{v}}, \check{\boldsymbol{\theta}}, \\ \check{\mathbf{p}}_g, \check{\mathbf{q}}_g, \\ \check{\mathbf{p}}_{dg,j}, \\ \check{\mathbf{q}}_{dg,j}}} \sum_{i=1}^{n_g} C_i(\check{\mathbf{p}}_g^{(i)}) + \sum_{j=1}^{n_{ds}} \sum_{k=1}^{n_{dg,j}} C_{jk}(\mathbf{p}_{dg,j}^{(k)}) \quad (5.2a)$$

$$\text{s.t. } G_P(\check{\mathbf{v}}, \check{\boldsymbol{\theta}}; \check{\mathbf{Y}}) + \check{p}_d - \check{\mathbf{p}}_g = 0, \quad (5.2b)$$

$$G_Q(\check{\mathbf{v}}, \check{\boldsymbol{\theta}}; \check{\mathbf{Y}}) + \check{q}_d - \check{\mathbf{q}}_g = 0, \quad (5.2c)$$

$$G_{\text{line}}(\check{\mathbf{v}}, \check{\boldsymbol{\theta}}; \check{\mathbf{Y}}) \leq \check{l}_{\text{line,max}}, \quad (5.2d)$$

$$\check{v}_{\min} \leq \check{\mathbf{v}} \leq \check{v}_{\max}, \quad \check{\theta}_{\min} \leq \check{\boldsymbol{\theta}} \leq \check{\theta}_{\max}, \quad (5.2e)$$

$$\check{p}_{g,\min} \leq \check{\mathbf{p}}_g \leq \check{p}_{g,\max}, \quad \check{q}_{g,\min} \leq \check{\mathbf{q}}_g \leq \check{q}_{g,\max}, \quad (5.2f)$$

$$P_{j,u}(\mathbf{x}_j) + \check{\mathbf{p}}_g^{(s_{j,u})} = 0, \quad \forall j \in \{1, \dots, n_{ds}\}, \\ \forall u \in \{1, \dots, r_j\}, \quad (5.2g)$$

$$Q_{j,u}(\mathbf{x}_j) + \check{\mathbf{q}}_g^{(s_{j,u})} = 0, \quad \forall j \in \{1, \dots, n_{ds}\}, \\ \forall u \in \{1, \dots, r_j\}, \quad (5.2h)$$

$$FR_j(\mathbf{x}_j) \leq 0, \quad \forall j \in \{1, \dots, n_{ds}\}, \quad (5.2i)$$

$$x_{j,\min} \leq \mathbf{x}_j \leq x_{j,\max}, \quad \forall j \in \{1, \dots, n_{ds}\}, \quad (5.2j)$$

$$\mathbf{x}_j = \left[\mathbf{v}_j^\top \mathbf{p}_{dg,j}^\top \mathbf{q}_{dg,j}^\top \right]^\top, \quad \forall j \in \{1, \dots, n_{ds}\}. \quad (5.2k)$$

Examining Equations (5.2b) - (5.2f), it can be seen that these equations contain only TS-related variables. The DS-related variables are expressed through the functions defined in Equations (5.2g) - (5.2i). The functions $P_{j,u}(\mathbf{x}_j)$ and $Q_{j,u}(\mathbf{x}_j)$ are designed to represent DS-related variables to the active and reactive power flow at the PCCs. Thanks to these functions, DSs are modeled as active and reactive power sources at the PCCs from the perspective of the TS. Note that the variables $\check{\mathbf{p}}_g^{(s_{j,u})}$ and $\check{\mathbf{q}}_g^{(s_{j,u})}$ represent the active and reactive power flow at the PCC, respectively, directed from DS towards TS.

The functions $FR_j(\mathbf{x}_j)$ are designed to represent the feasible region of the DSs. These functions ensure that technical constraints, such as line flow and voltage magnitude limits within the DS, are satisfied. Specifically, if $\mathbf{x}_j$ represents a feasible operating point that complies with all DS constraints, the condition $FR_j(\mathbf{x}_j) \leq 0$ is satisfied. If this condition is not met, it indicates that $\mathbf{x}_j$ lies outside the feasible region. Moreover, (5.2j) defines the bounds for the DS-related variables. It should be noted that for each DS, only a single $FR_j(\mathbf{x}_j)$ function is created, regardless of the number of PCCs. However, for each DS, separate $P_{j,u}(\mathbf{x}_j)$ and $Q_{j,u}(\mathbf{x}_j)$ functions must be defined for each PCC.

In summary, non-sensitive DS-related variables are encapsulated within a specific set of functions to represent the technical constraints of the DS while preserving data privacy. These functions are constructed using ML models trained exclusively on non-sensitive DS-related variables, ensuring that sensitive data remains protected throughout the process. It should be emphasized that the ML-based functions do not arbitrarily replace the physical constraints of the DS. Instead, the datasets used to train the ML models are generated by evaluating candidate operating points using the physical DS model, where power flow feasibility and operational limits such as voltage magnitude and line flow constraints are explicitly verified. Based on these samples, the ML models inherently learn an implicit representation of the feasible operating region of the DS. Although misclassification cannot be entirely ruled out in general, a conservativeness parameter will be introduced in the following sections to reduce the risk of misclassification. In the case studies, this parameter is selected to achieve 100% feasibility among the evaluated operating points. Consequently, the learned function $FR_j(\mathbf{x}_j)$ acts as a surrogate constraint that approximates the feasible region. When embedded into the TSO-level OPF problem, this surrogate constraint ensures that candidate operating points remain consistent with the underlying DS operational limits without requiring the explicit DS network model.

The procedure for generating the datasets required to train the ML models is described in detail in Section 5.3. Subsequently, the NN-based construction of the $FR_j(\mathbf{x}_j)$ functions is introduced in Section 5.4.1, while the quadratic regression models used to obtain the $P_{j,u}(\mathbf{x}_j)$ and $Q_{j,u}(\mathbf{x}_j)$ functions are detailed in Section 5.4.2.

Fig. 5.1 illustrates the schematic representation of the proposed method. As outlined in previous sections, the OPF should be solved by TSOs. To facilitate this, the ML models and the cost functions of the DGs are shared with the TSO. By employing these ML models, the TSO can effectively solve the proposed ML-based privacy-preserving OPF as determined in (5.2). This approach allows the OPF to be solved and the dispatch decisions for the DGs to be determined in a single round of communication, without requiring any additional disaggregation processes. Consequently, the flexibility obtained from DSs can be utilized for various network management purposes in a cost-effective manner, while ensuring the protection of sensitive data and adhering to the technical constraints of both TSOs and DSOs.

5.3 Dataset Creation

In the proposed method, the technical constraints of the DSs are represented using a set of functions developed through ML models. The effective training of these models necessitates a comprehensive dataset. To generate this dataset, the Latin hypercube sampling (LHS) method

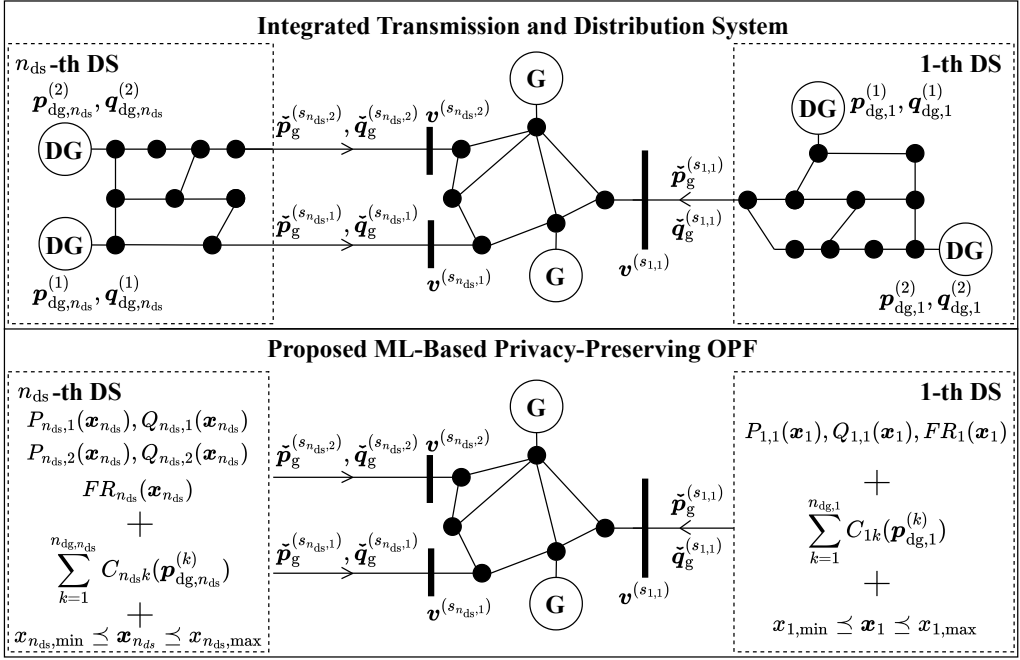


Figure 5.1: Schematic representation of the proposed method.

is employed [134]. LHS allows for sampling from a multidimensional distribution while maintaining the marginal probability distributions for each variable. This technique ensures efficient exploration of the entire range of each variable, even when the number of samples is relatively small.

To create the dataset, the DSO generates various operating points, represented by different values of $\mathbf{x}_j$, within the specified limits of these variables, as outlined in (5.2j). Particular attention must be given to the variables $\mathbf{p}_{dg,j}$ and $\mathbf{q}_{dg,j}$, as they define the PQ chart of the FPU. It is important to note that the DGs used in present study can be also considered as FPUs.

In most studies, the PQ characteristics of FPUs are typically considered as rectangular (ideal or generic) [161] (see Fig. 5.2a). However, FPUs exhibit varying PQ characteristics, which are often modeled as convex polygons [162]. In [113], rather than focusing on specific FPU shapes, such as triangular or square configurations, the methodology is demonstrated using arbitrary convex polygons. This approach illustrated the applicability of the method across diverse characteristics. Following this direction, the present chapter also adopts arbitrary convex polygon PQ characteristics, thereby demonstrating the effectiveness of the proposed method for different FPU characteristics (see Fig. 5.2b).

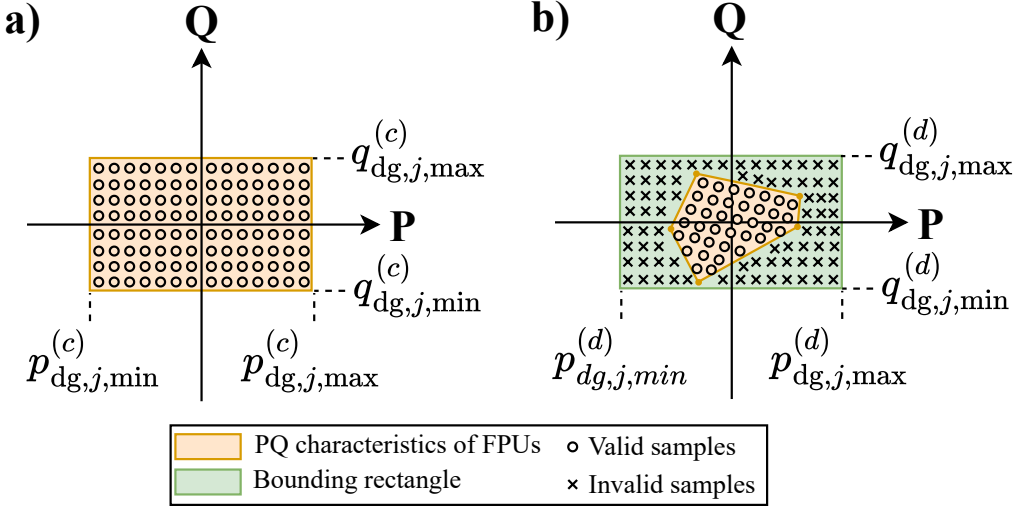


Figure 5.2: Data sampling approach using LHS.

In generating the dataset, a novel approach is introduced for sampling with LHS in scenarios involving arbitrary convex polygons. Here, the bounding rectangles formed by the $p_{dg,j,min}$, $p_{dg,j,max}$, $q_{dg,j,min}$, and $q_{dg,j,max}$ are used, and these bounding rectangles fully encapsulate the arbitrary convex polygons. Subsequently, LHS is applied within these bounding rectangles, enabling sampling from the entire arbitrary convex polygon that lies within the bounding rectangle.

It is important to note that, as a natural consequence of this approach, some samples are taken from the area between the arbitrary polygon and the bounding rectangle. However, these samples do not represent feasible operating points. Fig. 5.2 illustrates the data sampling approach using LHS. Specifically, Fig. 5.2a illustrates a rectangular PQ characteristic, while Fig. 5.2b shows a PQ characteristic of a convex polygon along with its bounding rectangle. It also distinguishes between valid samples that fall within the convex polygon and invalid samples that are located in the region between the polygon and the bounding rectangle.

To ensure that only valid samples within the polygon are considered, an additional filtering approach is required. As is well known, convex polygons are characterized by a set of linear inequalities. Accordingly, these inequalities are expressed mathematically as follows:

$$A_{PQ,jk} \begin{bmatrix} \mathbf{p}_{dg,j}^{(k)} & \mathbf{q}_{dg,j}^{(k)} \end{bmatrix}^\top \leq b_{PQ,jk} \forall j \in \{1, \dots, n_{ds}\} \text{ and } \forall k \in \{1, \dots, n_{dg,j}\}, \quad (5.3)$$

where $A_{PQ,jk} \in \mathbb{R}^{n_{v,jk} \times 2}$ represents the matrix of coefficients, while $b_{PQ,jk} \in \mathbb{R}^{n_{v,jk}}$ is the vector of constants. Also, $n_{v,jk}$ indicates the number of vertices that define the convex polygon for a given DG. Note that, for DGs with rectangular characteristics, the vertices are determined by the $p_{dg,j,\min}$, $p_{dg,j,\max}$, $q_{dg,j,\min}$ and $q_{dg,j,\max}$ values. After defining these linear inequalities, they can be incorporated into the OPF problem defined in (5.2) to ensure that invalid samples, which lie between the convex polygon and the bounding rectangle, are identified as infeasible. To achieve this, the OPF problem can be extended as follows:

$$\begin{aligned} \min \quad & (5.2a) \\ \text{s.t.} \quad & (5.2b) - (5.2k), \\ & (5.3). \end{aligned} \tag{5.4}$$

This ensures that invalid samples are appropriately classified as infeasible, allowing only valid samples to be evaluated. This approach facilitates dataset generation using standard LHS without the need for additional sampling techniques. Consequently, the proposed method can effectively handle FPU's with arbitrary convex polygon characteristics beyond rectangular ones. Furthermore, since the FPU's characteristics are represented by linear inequalities, they can be integrated into the OPF problem in a computationally efficient manner.

Overall, each operating point $\mathbf{x}_j$ generated by LHS is assessed based on the security limits of the DSs using power flow analysis. Based on this evaluation, the operating points are classified as either feasible or infeasible. Subsequently, datasets are compiled consisting of feasible instances $\mathcal{F}$ and infeasible instances $\mathcal{I}$.

5.4 Representation Technical Constraints of the Distribution Systems with Machine Learning Models

After generating the dataset, ML models are trained to construct the previously defined functions. Specifically, a novel, tailored NN classification model is introduced to construct the $FR_j(\mathbf{x}_j)$ functions. In addition, quadratic regression models are employed to construct $P_{j,u}(\mathbf{x}_j)$ and $Q_{j,u}(\mathbf{x}_j)$ functions.

It is important to recall that the $FR_j(\mathbf{x}_j)$ functions determine the feasible region of the DSs by incorporating their technical constraints. In this way, they ensure that all operational limits of the DSs are satisfied in any dispatch of the DGs. Meanwhile, the functions $P_{j,u}(\mathbf{x}_j)$ and $Q_{j,u}(\mathbf{x}_j)$

establish the coupling between DS-related variables and the active/reactive power at the PCCs. Consequently, from the perspective of the TS, each DS can be effectively modeled as an equivalent active and reactive power source at its corresponding PCC.

5.4.1 NN-Guided Polytope Representation of Feasible Region

In this section, the functions $FR_j(\mathbf{x}_j)$ are modeled in the form of a convex polytope, ensuring that the condition $FR_j(\mathbf{x}_j) \leq 0$ is satisfied. To construct this model, a previously generated dataset is utilized that consists of a finite number of feasible and infeasible instances i.e., $\mathcal{F}$ and $\mathcal{I}$. Following this, the function can be described as follows:

$$FR_j(\mathbf{x}_j) = A_{FR,j}\mathbf{x}_j - b_{FR,j}. \quad (5.5)$$

To construct a convex polytope that encompasses all feasible instances, it is needed to determine a matrix $A_{FR,j} \in \mathbb{R}^{n_{f,j} \times n_j}$ and a vector $b_{FR,j} \in \mathbb{R}^{n_{f,j}}$. This formulation ensures that all feasible instances $\mathbf{x}_j \in \mathcal{F}$ satisfy the inequality $A_{FR,j}\mathbf{x}_j \leq b_{FR,j}$, while all infeasible instances $\mathbf{x}_j \in \mathcal{I}$ do not satisfy this inequality, i.e., $A_{FR,j}\mathbf{x}_j \not\leq b_{FR,j}$. Note that, $n_{f,j}$ represents the number of facets of the polytope, assuming that there is no redundancy in $A_{FR,j}\mathbf{x}_j \leq b_{FR,j}$.

The next step in constructing the polytope involves determining under what circumstances an operating point $\mathbf{x}_j$ satisfies the defined inequality. It is considered $\mathbf{x}_j$ to satisfy the inequality if and only if every element of $\mathbf{z}$ is less than or equal to zero, where $\mathbf{z} = A_{FR,j}\mathbf{x}_j - b_{FR,j}$. As a result, $\max(\mathbf{z}) \leq 0$ indicates that $\mathbf{x}_j$ lies inside the polytope, making it a feasible point. On the contrary, if at least one element of $\mathbf{x}_j$ is strictly greater than zero, this implies $\max(\mathbf{z}) > 0$, meaning that $\mathbf{x}_j$ is outside the polytope and therefore an infeasible instance. To better understand this polytope, Fig. 5.3 provides a schematic representation with $n_{f,j} = 6$, where feasible instances are depicted by $-$ and infeasible instances by $+$.

After determining the approach to assess whether a given operating point $\mathbf{x}_j$ lies inside or outside the polytope, the next crucial step is to define the appropriate parameters $A_{FR,j}$ and $b_{FR,j}$. To achieve this, a novel tailored NN architecture is leveraged specifically designed for this purpose. Upon training, the weights and biases of this NN model are directly mapped to the parameters $A_{FR,j}$ and $b_{FR,j}$. In this framework, feasible instances are labeled as Class 0, and infeasible instances as Class 1. The proposed NN architecture can be mathematically represented as follows:

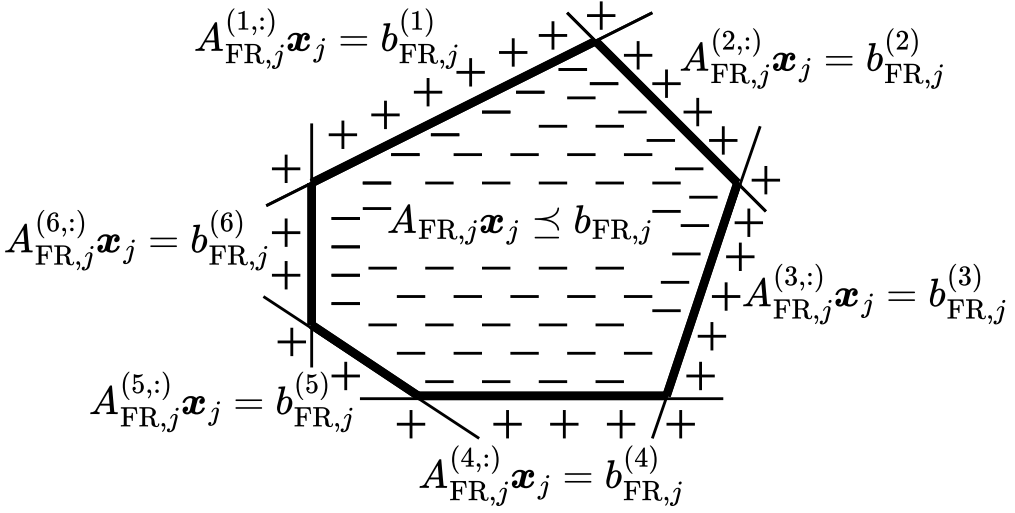


Figure 5.3: Schematic representation of the polytope.

$$\mathbf{o}_j = W_j \mathbf{x}_j + b_j, \quad (5.6a)$$

$$\mathbf{f}_j = \max(\mathbf{o}_j), \quad (5.6b)$$

$$\mathbf{y}_j = \text{sigmoid}(\mathbf{f}_j). \quad (5.6c)$$

Equation (5.6) describes a feed-forward architecture. Firstly, (5.6a) represents a hidden layer with $n_{h,j}$ nodes, where $W_j \in \mathbb{R}^{n_{h,j} \times n_j}$ denotes the weight matrix and $b_j \in \mathbb{R}^{n_{h,j}}$ denotes the bias vector. In (5.6b), instead of using a standard activation function, the output of the hidden layer $\mathbf{o}_j \in \mathbb{R}^{n_{h,j}}$ is processed by a max aggregator function, resulting in $\mathbf{f}_j \in \mathbb{R}$. Then, $\mathbf{f}_j$ is passed through the sigmoid activation in (5.6c), producing the final output $\mathbf{y}_j \in [0, 1]$, which can be interpreted as the probability of infeasibility of a given input $\mathbf{x}_j$. Fig. 5.4 shows the architecture of the proposed NN. Finally, to train the NN, the standard binary cross-entropy loss function is employed.

As the final step, the relationship between the weights W_j and biases b_j of the NN and matrix $A_{FR,j}$ and vector $b_{FR,j}$ needs to be established. According to (5.6), a sample $\mathbf{x}_j$ is classified as feasible if $\max(W_j \mathbf{x}_j + b_j) \leq 0$ and as infeasible if $\max(W_j \mathbf{x}_j + b_j) > 0$. This implies that the decision region for feasible samples is defined by $W_j \mathbf{x}_j \leq -b_j$. Thus, the desired polytope can be defined by setting $A_{FR,j} = W_j$ and $b_{FR,j} = -b_j$. Note that, the number of hidden nodes $n_{h,j}$

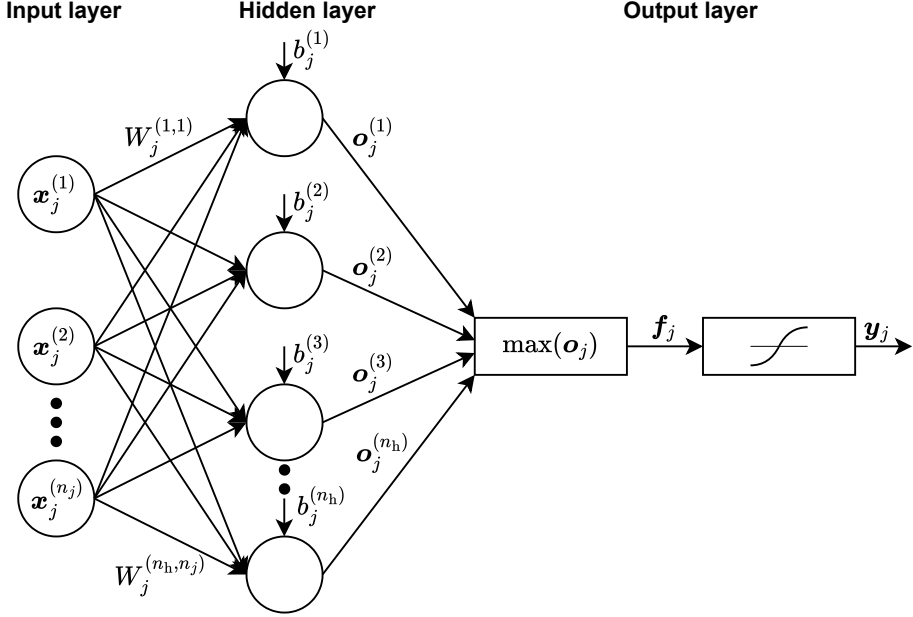


Figure 5.4: The architecture of the novel tailored NN.

provides an upper bound on the number of facets of the polytope (though this is only an upper bound, as some rows of $W_j x_j \leq -b_j$ may be redundant).

After obtaining $A_{FR,j}$ and $b_{FR,j}$ from the NN, one critical aspect in accurately mapping the feasible region by using the polytope approach is the possibility of misclassification. Incorrectly classifying a feasible point as infeasible may only lead to economic losses, while misclassifying an infeasible point as feasible can result in significant issues within the power system. To mitigate this risk, a *conservativeness parameter* $c_{FR,j}$ is introduced, which is directly embedded into the inequality constraints of the polytope, modifying its formulation as:

$$A_{FR,j} x_j \leq b_{FR,j} - c_{FR,j}. \quad (5.7)$$

Assigning a positive value to $c_{FR,j}$ shifts the polytope's boundaries inward, resulting in a smaller but more reliable feasible region. This adjustment reduces the risk of misclassifying infeasible points as feasible. Naturally, the choice of $c_{FR,j}$ entails a trade-off: larger values increase conservativeness, but at the expense of excluding some feasible points, thereby slightly increasing the total cost. Selecting this parameter appropriately is therefore crucial. By employing this

strategy, represent the non-convex feasible area can be represented as a conservative convex polytope, ensuring a more reliable representation of the feasible region. Finally, the constraint $A_{\text{FR},j} \mathbf{x}_j \leq b_{\text{FR},j} - c_{\text{FR},j}$ can be incorporated into the OPF problem as specified (5.2i), the feasible region of the DS can be effectively approximated. Utilizing such a polytope allows the OPF to be implemented in a computationally efficient and privacy-preserving manner.

5.4.2 Quadratic Regression-Based Power Flow Approximator

After defining the feasible region of the DSs, the focus is on defining the functions $P_{j,u}(\mathbf{x}_j)$ and $Q_{j,u}(\mathbf{x}_j)$. These functions are designed to map the DS-related variables $\mathbf{x}_j$ to the active and reactive power flows at the PCCs, respectively. Considering the inherent quadratic relationship between power injections and system losses [163], a quadratic regression model is selected to define these mappings. Accordingly, these functions can be described as follows:

$$P_{j,u}(\mathbf{x}_j) = \mathbf{x}_j^\top A_{\text{P},j,u} \mathbf{x}_j + b_{\text{P},j,u}^\top \mathbf{x}_j + c_{\text{P},j,u}, \quad (5.8a)$$

$$Q_{j,u}(\mathbf{x}_j) = \mathbf{x}_j^\top A_{\text{Q},j,u} \mathbf{x}_j + b_{\text{Q},j,u}^\top \mathbf{x}_j + c_{\text{Q},j,u}, \quad (5.8b)$$

where $A_{\text{P},j,u}, A_{\text{Q},j,u} \in \mathbb{R}^{n_j \times n_j}$, and $b_{\text{P},j,u}, b_{\text{Q},j,u} \in \mathbb{R}^{n_j}$, and $c_{\text{P},j,u}, c_{\text{Q},j,u} \in \mathbb{R}$ represent the model parameters, which are determined through standard ML training procedures.

To train the quadratic regression models, $\mathbf{x}_j$ is used as input, active and reactive power at the PCCs, i.e., $\check{p}_g^{(s_{j,u})}$ and $\check{q}_g^{(s_{j,u})}$ as output of the models. The values of $\check{p}_g^{(s_{j,u})}$ and $\check{q}_g^{(s_{j,u})}$ are derived from power flow calculations. Note that, while both feasible instances $\mathcal{F}$ and infeasible instances $\mathcal{I}$ are utilized in the NN models, only feasible instances $\mathcal{F}$ are employed for training the quadratic regression models. This ensures that the regression models are trained exclusively on physically valid operating points, thereby avoiding the risk of learning invalid relationships introduced by infeasible samples. Upon completion of the training process, these functions accurately represent the relationship between DG-related variables and the corresponding power flows at the PCCs.

5.5 Case Studies and Discussion

In the present chapter, the performance of the proposed method is evaluated by comparing it with the traditional AC-OPF, which does not consider data privacy. The evaluation process involves several steps: first, a suitable power system is established, followed by the generation of a comprehensive dataset. ML models are then trained using this dataset, and their approximation

accuracy is assessed. Finally, the effectiveness of the proposed method are examined through extensive case studies.

The primary simulation environment for this study is MATLAB, where MATPOWER [101] is utilized with KNITRO solver [102] for performing AC-OPF calculations. ML models are developed and trained using TensorFlow/Keras [105, 106].

5.5.1 Power System Creation

To create an appropriate integrated transmission and distribution system, IEEE 30-bus test system is employed as TS, and three IEEE 33-bus test systems are employed as the DSs. Fig. 5.5 shows the integrated power system configuration. For all DSs, the normally open lines are closed, converting the systems into meshed grids to assess the effectiveness of the proposed method in meshed grids. The first DS is connected to the 11th bus of the TS (i.e., $s_{1,1} = 11$) through a single PCC, and includes only one DG. This setup is designed such that the first DS is represented in a three-dimensional space (i.e., $x_1 \in \mathbb{R}^3$), enabling the proposed method to be visualized within three-dimension.

To demonstrate the effectiveness of the proposed method in a more complex system, five DGs are integrated into the second DS. Additionally, as is commonly observed in real-world scenarios, this DS is connected to the TS through two PCCs located at the 16th and 17th buses of the TS (i.e., $s_{2,1} = 16$ and $s_{2,2} = 17$), effectively showcasing the scenario involving multiple PCCs. The third DS demonstrates the effectiveness of the proposed method on FPU with varying PQ characteristics. To achieve this, the same topology as the second DS is employed, but instead of DGs with only rectangular PQ charts as in the first two DSs, this system incorporates DGs with varying convex polygon PQ characteristics, as illustrated in Fig. 5.6. This setup demonstrates that the proposed method maintains high performance regardless of the specific PQ characteristics. Additionally, the third DS is connected to the TS through two PCCs at the 19th and 20th buses of the TS (i.e., $s_{3,1} = 19$ and $s_{3,2} = 20$). Furthermore, all DGs are designed to have active power outputs ranging between 0 and 2 MW, and reactive power outputs between 0 and 2 MVAR, respectively. Note that these values also define the bounding rectangle for the DGs within the third DS.

5.5.2 Dataset Generation, Training, and Approximation Quality in ML Models

To develop the ML models, the first step involves generating the dataset, for which LHS is employed, as detailed earlier. For the first DS, a total of 20,000 data points are generated, while

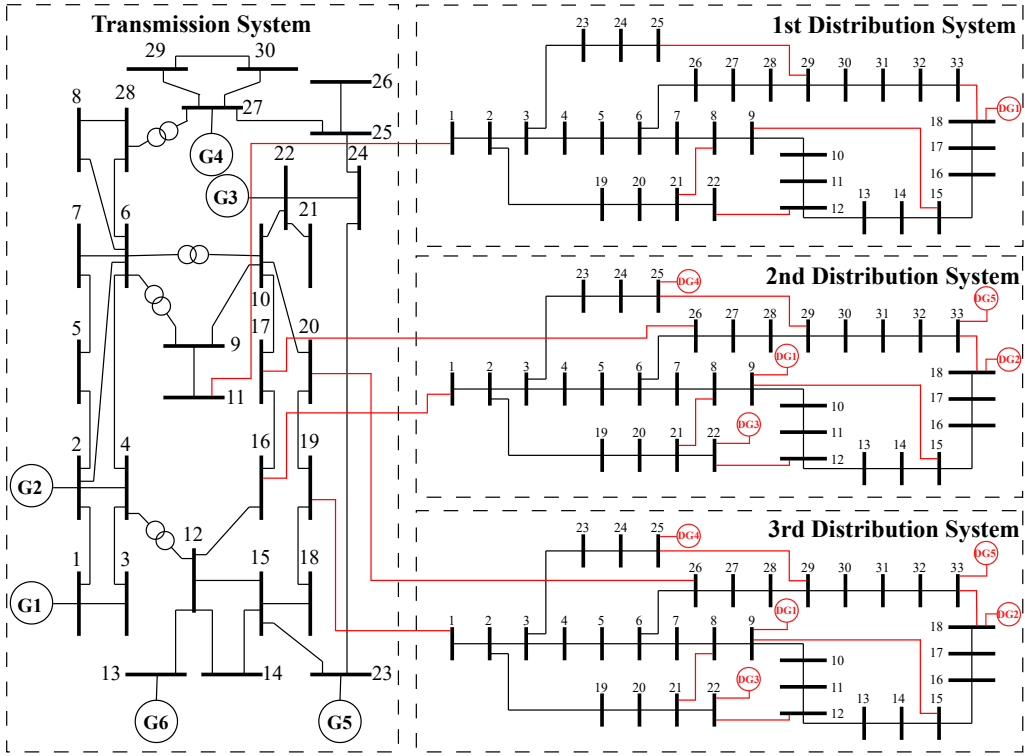


Figure 5.5: Single-line diagram of the integrated power system, where the modifications introduced to the network are highlighted in red.

100,000 data points are generated for both the second and third DSs. It is important to note that, as previously discussed, the same dataset and ML models are used for both the second and third DSs to demonstrate that the proposed method maintains high performance, irrespective of the specific PQ characteristics. After generating the datasets, ML models are trained by following standard procedures for both classification and regression tasks. The dataset is divided into a training set (80%) and a test set (20%) to validate the model's performance. In addition, the hyperparameters of the NN models are optimized using k-fold cross-validation applied to the training set.

NN classification models are employed to distinguish between feasible and infeasible operating points based on the generated x_j values. For constructing these NN models, random search hyperparameter tuning is employed, selecting $n_{h,1} = 20$ hidden nodes for the first DS, and $n_{h,2} = 1,000$ and $n_{h,3} = 1,000$ hidden nodes for the second and third DSs, respectively. As the complexity of the power system increases, the number of facets required to accurately represent the feasible space of the DSs also rises, hence the increased number of hidden nodes. It is important to recall that the number of hidden nodes, $n_{h,j}$, provides an upper bound on the number

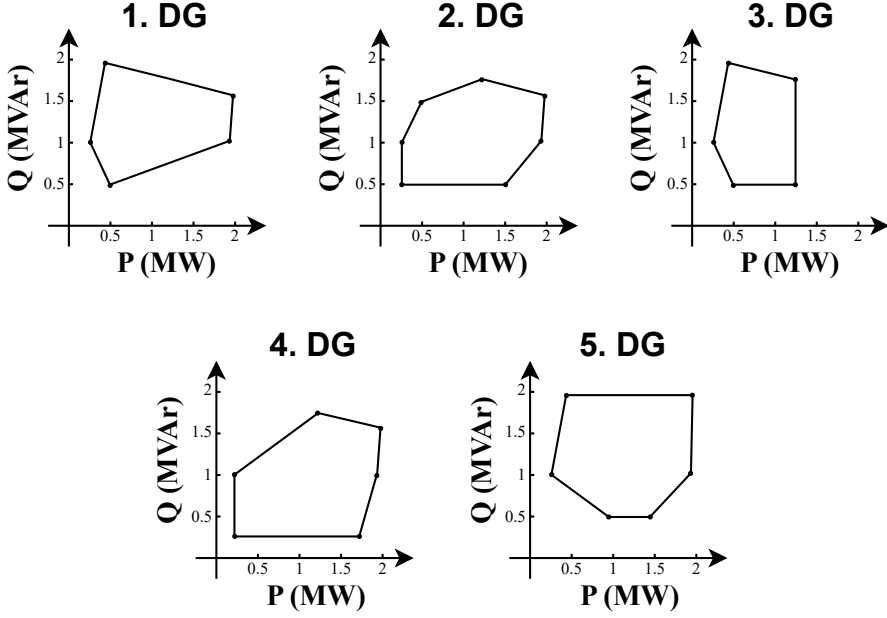


Figure 5.6: Convex polygon PQ characteristics for DGs in the third DS.

of facets that define the polytope. Consequently, some rows of the matrix W_j and vector b_j may be redundant. Therefore, even when a large number of hidden nodes are defined, the NN only generates as many facets as necessary to describe the feasible region effectively.

After training the models, their approximation quality is assessed by evaluating them on the test sets. Let $y_i \in \{0, 1\}$ denote the true feasibility status of instance i , where $y_i = 0$ indicates a feasible instance and $y_i = 1$ indicates an infeasible instance. Let $\tilde{y}_i$ denote the feasibility status determined by the model $FR_j(\cdot)$. For a dataset containing N instances, the performance of $FR_j(\cdot)$ is evaluated using the accuracy, recall, and specificity metrics [107].

The accuracy is defined as the fraction of instances for which the determined feasibility status matches the true status:

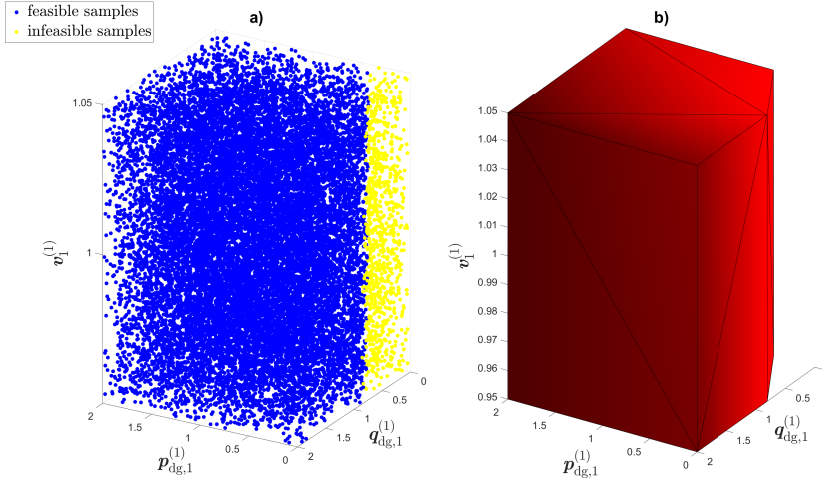
$$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(\tilde{y}_i = y_i). \quad (5.9)$$

The recall measures the fraction of feasible instances that are correctly identified:

$$\text{Recall} = \frac{\sum_{i=1}^N \mathbf{1}(\tilde{y}_i = 0 \wedge y_i = 0)}{\sum_{i=1}^N \mathbf{1}(y_i = 0)}. \quad (5.10)$$

Table 5.1: Accuracy, Recall and Specificity Metrics of the NN Models

Model	Accuracy	Recall	Specificity
$FR_1(\mathbf{x}_1)$	99.90%	99.89%	100.00%
$FR_2(\mathbf{x}_2)$ & $FR_3(\mathbf{x}_3)$	96.47%	97.23%	95.53%

**Figure 5.7:** a) Dataset indicating feasible and infeasible samples. b) Feasible region approximation of the NN indicating the facets.

The specificity measures the fraction of infeasible instances that are correctly identified:

$$\text{Specificity} = \frac{\sum_{i=1}^N \mathbf{1}(\tilde{y}_i = 1 \wedge y_i = 1)}{\sum_{i=1}^N \mathbf{1}(y_i = 1)}. \quad (5.11)$$

Here, $\mathbf{1}(\cdot)$ denotes the indicator function, which equals 1 when the condition holds and 0 otherwise. The results are summarized in Table 5.1. For the NN model of the first DS, i.e., $FR_1(\mathbf{x}_1)$, which is a relatively less complex system, the all metrics are observed to be nearly 100%. Additionally, Fig. 5.7 provides a visual representation of the generated dataset and the NN's approximation of the feasible region for the first DS. Notably, if the voltage were assumed to be constant, the feasible region would be represented as a two-dimensional area. However, the figure illustrates a larger three-dimensional region, demonstrating that incorporating voltage variations allows for better utilization of DS flexibility potential. As depicted in Fig. 5.7, the NN model accurately approximates the feasible region and successfully establishes a well-defined decision boundary.

When examining the models for the second and third DS, i.e., $FR_2(\mathbf{x}_2)$ and $FR_3(\mathbf{x}_3)$, it is observed that the accuracy and recall metrics are 96.47% and 97.23%, respectively, while the specificity metric is 95.53%. These results indicate that, although the NN achieves high performance even for complex DSs, it does not reach 100%. In particular, the specificity metric reflects the model's ability to correctly classify infeasible samples. Misclassifying infeasible points as feasible can pose serious risks in power system operation, highlighting the importance of the conservativeness parameter introduced in Equation (5.7). This parameter is therefore critical in ensuring reliable system operation, and its impact will be further analyzed in the following section.

Following the NN-based classification models, quadratic regression models are developed and evaluate their performance using numerical metrics such as root mean square error (RMSE) and mean absolute error (MAE). Let z_i denote the true value (i.e., real or reactive power flow at a PCC) for instance i , and let $\tilde{z}_i$ denote the value determined by the regression model (i.e., $P_{j,u}(\cdot)$ or $Q_{j,u}(\cdot)$). Then, the RMSE and MAE are defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\tilde{z}_i - z_i)^2}. \quad (5.12)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\tilde{z}_i - z_i|. \quad (5.13)$$

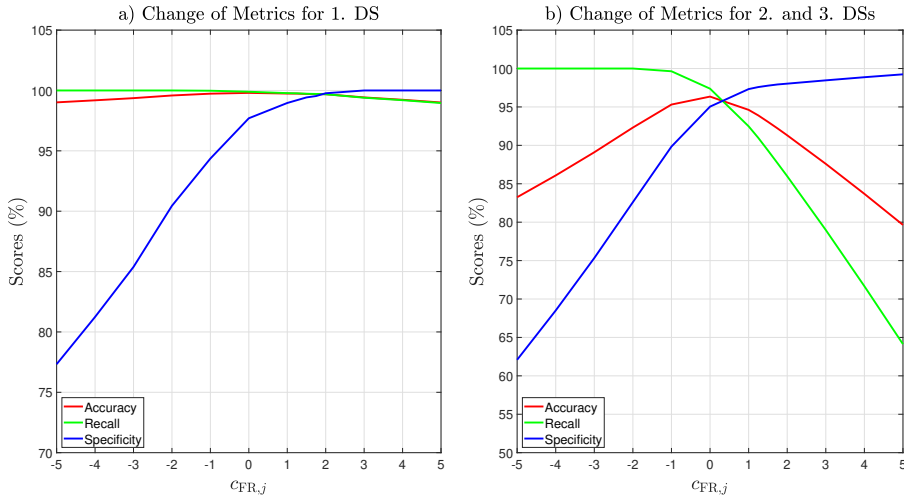
RMSE penalizes larger deviations more strongly due to the squared term, whereas MAE measures the average magnitude of the deviations. The results, displayed in Table 5.2, indicate that all regression models achieved performances close to 100%. These numerical results clearly demonstrate that the ML models are highly effective in capturing and mapping the characteristics of the DSs.

5.5.3 Analysis of the Impact of the Conservativeness Parameter

After constructing the ML models, it is essential to examine the effect of the conservativeness parameter defined in Equation (5.7), as this parameter directly mitigates the risk of misclassification. This analysis should first be conducted from the perspective of the DSOs responsible for generating the ML models. As discussed in the previous sections, the parameters $A_{FR,j}$ and $b_{FR,j}$ are derived from the NN parameters W_j and b_j . Once these values are obtained, the influence

Table 5.2: RMSE and MAE Metrics of the Quadratic Regression Models

Model	RMSE	MAE
$P_{1,1}(x_1)$	5.0×10^{-4}	3.6×10^{-4}
$P_{2,1}(x_2) \& P_{3,1}(x_3)$	2.9×10^{-4}	2.0×10^{-4}
$P_{2,2}(x_2) \& P_{3,2}(x_3)$	4.9×10^{-4}	3.4×10^{-4}
$Q_{1,1}(x_1)$	4.3×10^{-4}	3.0×10^{-4}
$Q_{2,1}(x_2) \& Q_{3,1}(x_3)$	2.7×10^{-4}	1.8×10^{-4}
$Q_{2,2}(x_2) \& Q_{3,2}(x_3)$	4.5×10^{-4}	3.1×10^{-4}

**Figure 5.8:** Change of metrics according to conservativeness parameter.

of different values of $c_{FR,j}$ can be systematically examined by substituting the corresponding x_j samples from the test set (generated in the previous section) into Equation (5.7). The outcomes are then assessed using the relevant performance metrics, with the results presented in Fig. 5.8.

As expected, increasing $c_{FR,j}$ shifts the polytope's boundary inward, leading to an improvement in the specificity metric. This indicates that the model becomes more effective at avoiding the critical misclassification of infeasible points as feasible, which is the desired outcome. Conversely, the recall metric decreases with larger values of $c_{FR,j}$, reflecting a reduction in the model's ability to correctly identify feasible points. This decline corresponds to the loss of feasible space, thereby highlighting the inherent trade-off between conservativeness and the completeness of the feasible region representation.

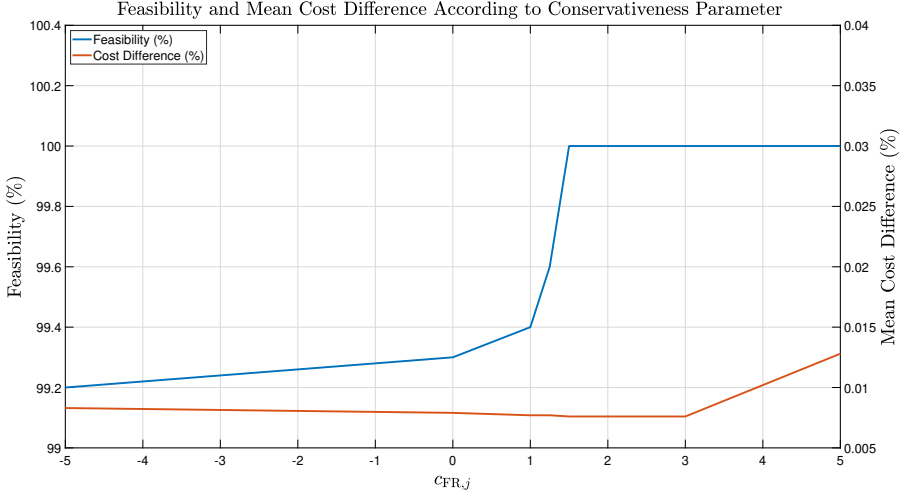


Figure 5.9: Feasibility and mean cost difference according to conservativeness parameter, taking AC-OPF as reference.

The impact of the conservativeness parameter on system cost should be assessed from the perspective of the TSO, once the ML models are transferred. To this end, both the feasibility and the total cost of the system are evaluated under varying values of $c_{FR,j}$. Specifically, the proposed method is benchmarked against the standard AC-OPF formulation across 1,000 randomly generated sets of cost coefficients using the created power system. In this comparison, the proposed method is implemented using (5.4), while the benchmark AC-OPF solution is obtained from (5.1). The results, presented in Fig. 5.9, illustrate the effect of the conservativeness parameter relative to the AC-OPF reference, highlighting the trade-offs between feasibility and total system cost.

Feasibility increases with higher values of $c_{FR,j}$, reaching 100% once this parameter exceeds approximately 1.5. In other words, the proposed method consistently identifies operating points that remain feasible within the standard AC-OPF. This outcome underscores the effectiveness of prioritizing the avoidance of infeasible operating points, albeit at the expense of a slight increase in total cost. Notably, the mean cost difference does not rise sharply as feasibility improves; rather, the increase becomes more pronounced only beyond a certain threshold of $c_{FR,j}$. From a practical perspective, this suggests that the TSO and DSO may prioritize feasibility and select $c_{FR,j}$ values that ensure 100% feasibility, while still accounting for the system's operational and economic conditions when determining the final choice.

Table 5.3: Metrics of the NN Models under Constant PCC Voltage

Model	Accuracy	Recall	Specificity
$FR_1(\mathbf{x}_1)$	100.00%	100.00%	100.00%
$FR_2(\mathbf{x}_2)$ & $FR_3(\mathbf{x}_3)$	99.58%	99.85%	99.18%

5.5.4 Analysis of the Impact of the Varying Voltage at the PCCs

In this subsection, the impact of varying PCC voltages on the proposed methodology is investigated. To isolate this effect, all parameters described previously, such as the power system configuration, dataset size, and other modeling assumptions, are kept constant, while only the PCC voltages (i.e., $\mathbf{v}_j$) are excluded from the variable set $\mathbf{x}_j$, and the entire process is repeated. Unlike the general approach, the dataset in this case is generated by fixing the PCC voltages at 1 p.u. The resulting NN performance metrics are summarized in Table 5.3.

An examination of the NN metrics reveals that the models perform better when PCC voltages are held constant compared to the case with varying voltages. This outcome is fundamentally expected, as fixing the PCC voltages reduces the dimensionality of the NN input space, thereby simplifying the learning task. Furthermore, the quadratic regression models also demonstrate consistently strong performance, with all results approaching 100%. For brevity, these numerical values are not included in a table.

Subsequently, the constructed model is transferred to the TSO, as in the previous section, and the proposed method is solved for 1,000 different sets of cost coefficients. The results are then compared with the standard AC-OPF in which the PCC voltage is not fixed, and the comparison is presented in the form of histograms in Fig. 5.10a. In addition, for a more detailed comparison, the case with varying PCC voltages (taken from the Section 5.5.3) is also benchmarked against AC-OPF, with the results shown in Fig. 5.10b. For a fair comparison between the two cases, the conservativeness parameter is set to $c_{FR,j} = 2$ in both scenarios.

Note that, feasibility reaches 100% in both cases. As shown in Fig. 5.10, the average and maximum cost differences for the varying-voltage case are 0.0076% and 0.2794%, respectively, whereas these values increase to 1.638% and 4.7924% when the PCC voltage is fixed. This finding indicates that, although NN models achieve higher predictive performance in the constant-voltage case, the resulting cost differences compared to AC-OPF are significantly larger. Consequently, allowing PCC voltages to vary enables a more cost-effective utilization of DS flexibility. This

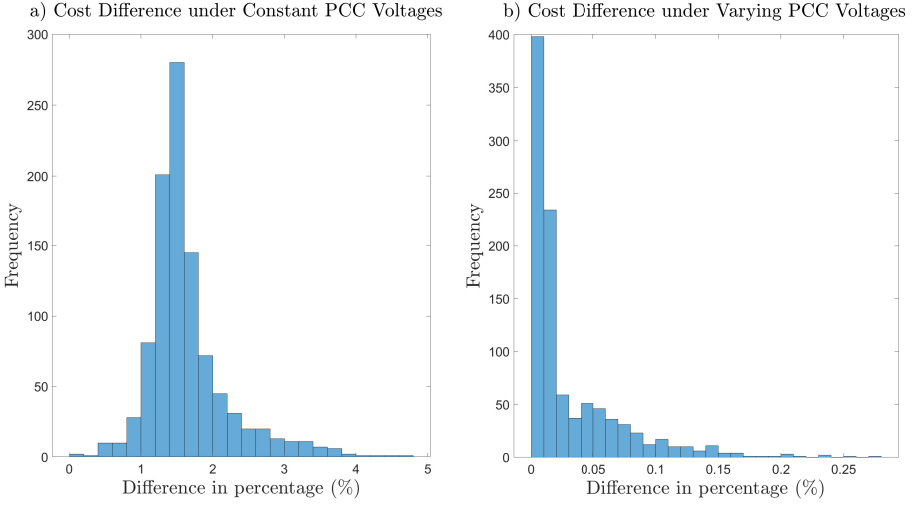


Figure 5.10: Cost difference comparison under varying and constant PCC voltages.

underscores the critical role of voltage variability at PCCs in improving the economic efficiency of the proposed method.

5.5.5 Computational Time Performance of the Proposed Method

Another important criterion for evaluating the effectiveness of the proposed method is its computational performance. In this regard, the case with $c_{FR,j} = 2$, examined in Section 5.5.3, is compared against the standard AC-OPF in terms of solution time, with the results illustrated as a histogram in Fig. 5.11.

The analysis shows that the average time difference is only 0.0647 seconds, with a maximum difference of 0.2594 seconds, highlighting the minimal computational overhead introduced by the proposed method. Negative time differences correspond to cases where the proposed method is faster, and indeed, several instances demonstrate improved speed relative to standard AC-OPF. This efficiency can be attributed to the tailored NN architecture embedded in the framework, which enhances computational speed while preserving accuracy. Overall, the proposed method provides privacy-preserving solutions with negligible time difference compared to AC-OPF.

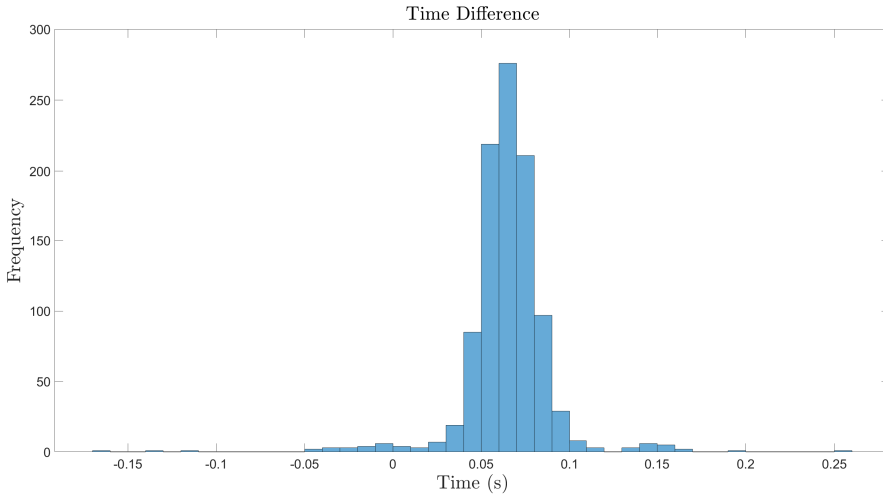


Figure 5.11: Histogram of the computational time differences taking AC-OPF as reference.

5.5.6 Scalability Analysis of the Proposed Method

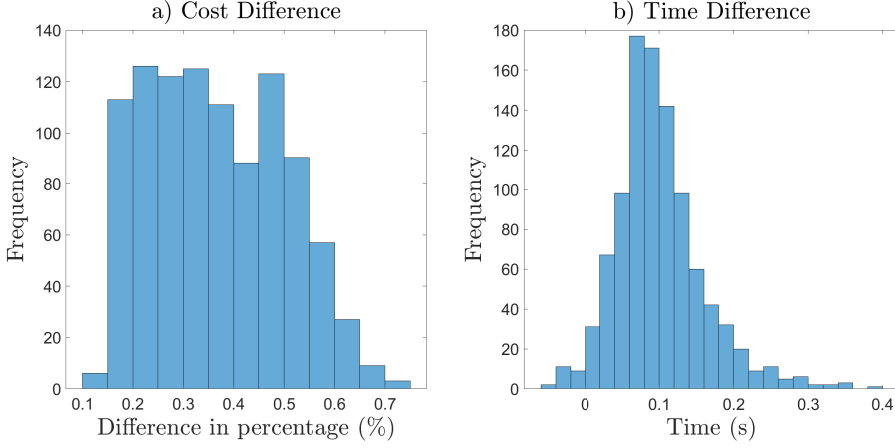
To demonstrate the scalability performance of the proposed method, a larger test system is constructed consisting of a 162-bus TS interconnected with three 136-bus DSs. Each DS in this setup includes 10 DGs. For consistency and brevity, the same parameter settings, dataset sizes, and PCC bus numbers as in the previous cases are adopted. Furthermore, to assess the method's performance under diverse FPU characteristics, the convex polygon PQ characteristics are employed illustrated in Fig. 5.6, with each characteristic type assigned to two DGs. These characteristics are mathematically represented as sets of linear inequalities, as defined in (5.3).

Following the construction of this extended power system, the datasets are generated and the ML models are trained. The performance metrics of the NN models are presented in Table 5.4. Despite the increased number of busbars and DGs, the NN models maintain high accuracy. It is important to note that these results correspond to the case where $c_{FR,j} = 0$; as discussed earlier, the specificity metric can be further improved by appropriately tuning $c_{FR,j}$. Similarly, the quadratic regression models again achieve performance levels close to 100%, and for brevity these values are not reported in a separate table.

Subsequently, the trained models are transferred to the TSO, which applies the proposed method across 1,000 randomly generated sets of cost coefficients with $c_{FR,j} = 2$. The results, shown in Fig. 5.12, compare the proposed method with the standard AC-OPF in terms of both cost difference and computational time.

Table 5.4: Metrics of the NN Models for Scalability Analysis

Model	Accuracy	Recall	Specificity
$FR_1(\mathbf{x}_1)$	98.03%	96.17%	98.91%
$FR_2(\mathbf{x}_2)$ & $FR_3(\mathbf{x}_3)$	97.04%	96.39%	97.37%

**Figure 5.12:** Histogram of total cost and computational time differences for scalability analysis, with AC-OPF taken as the reference.

In this scalability study, the proposed method again achieves 100% feasibility. When examining the cost performance, the mean cost difference is 0.36%, while the maximum cost difference is 0.72%. This demonstrates the efficiency of the proposed method in terms of cost. In terms of computational performance, the average time difference is 0.1 seconds, with a maximum of 0.39 seconds, further confirming the computational efficiency of the approach. These results highlight that the proposed framework remains both accurate and efficient even in large-scale, meshed TS-DS systems.

Moreover, the proposed method effectively accommodates diverse FPU characteristics while maintaining strong performance. Given that different FPU characteristics are treated as constraints, this approach demonstrates the capacity to incorporate other potential market-based or operational constraints between TSOs and DSOs without any loss in performance. Consequently, the proposed method achieves a balance between data privacy and operational efficacy, yielding comparable performance to the standard AC-OPF. This capability enables DS flexibility to be leveraged for network management purposes without compromising data privacy.

5.6 Summary

The increasing penetration of DGs and FPU at the distribution level has transformed DSs from passive demand repositories into active participants in system operation. Realizing the operational value of this flexibility requires a coordinated interaction between TSOs and DSOs that goes beyond conventional information exchange at a single point of common coupling. In particular, modern large-scale DSs are often operated through multiple PCCs and exhibit meshed structures that interact dynamically with the transmission network. Under these conditions, traditional approaches based on aggregated PQ capability boundaries or single-interface models become increasingly insufficient.

To address these challenges, this chapter introduced a machine-learning-based privacy-preserving AC-OPF framework that enables a TSO to determine the dispatch of DGs and other FPU associated with multiple DSs and subject to multiple PCC configurations. In this approach, the technical constraints of the DSs are represented using ML models, which can be shared with the TSO without compromising data privacy. By leveraging these ML models, the TSO can solve the OPF problem and directly determine the dispatch of FPU. This allows for dispatch decisions to be made in a single round of communication, eliminating the need for an additional disaggregation step. Furthermore, these ML models capture the nonlinear operational constraints of DSs with high fidelity, and their tailored NN enables computationally efficient integration into the TSO's OPF problem.

An additional contribution of this chapter lies in demonstrating that the proposed framework can accommodate FPU with varied PQ operating characteristics rather than assuming idealized rectangular capability charts. By employing Latin hypercube sampling for dataset generation and customizing model training accordingly, the approach captures realistic operational flexibility characteristics and their interaction with varying PCC voltage levels. As a result, the TSO is able to dispatch FPU directly at their individual terminals, rather than scheduling aggregated active and reactive power at a single interface point. This capability significantly enhances the operational utilization of DS flexibility.

The proposed method has been assessed through multiple case studies, including meshed DSs coupled at several PCCs. The results confirm that the ML models provide accurate representations of DS feasibility and allow the TSO to obtain near-optimal OPF solutions without knowledge of internal DS details. Moreover, high computational efficiency is achieved, illustrating that the approach is compatible with practical operational timeframes. In doing so, the proposed scheme demonstrates that DS flexibility can be exploited at scale while maintaining a strict data-privacy-by-design philosophy, thereby contributing directly to enhanced interoperability across system operators.

6 Conclusion and Future Work

The ongoing transformation of electric power systems, driven by the increasing penetration of distributed energy resources, the decentralization of control, the emergence of prosumers, and the growing complexity of market and operational interactions, has fundamentally reshaped how system operators coordinate across network layers. As the technical and economic interdependencies between markets, transmission networks, and distribution systems deepen, interoperability has become a prerequisite for achieving secure, reliable, and cost-efficient system operation. Yet, enabling such interoperability remains hindered by growing concerns around data privacy. As a result, although real-time collaborative operation has become technically necessary, stakeholders continue to restrict information exchange, creating a fundamental tension between cooperation and privacy.

This thesis addressed this tension and the associated three research questions, as discussed in detail in Section 1.2. The first research question examined how cross-entity optimization between market operators and TSOs can be carried out without revealing sensitive network or market information. The second research question focused on how DS flexibility can be effectively represented through FORs that are tractable, analytically expressible, and monetized, particularly for large-scale meshed DSs. The third research question extended this perspective by asking how DS flexibility can be utilized in scenarios involving multiple PCCs and direct dispatch of DGs and FPU, overcoming the inherent limitations of classical FOR-based approaches. Across these three research questions, this thesis develops a set of novel, data-driven, and privacy-preserving coordination frameworks that enable interoperability without requiring stakeholders to disclose sensitive data.

The first contribution of this thesis concerns the challenge of performing joint optimization between market operators and TSOs, where neither party is willing nor permitted to disclose sensitive data. To this end, Section 3 introduces a privacy-preserving AC-OPF framework where the TSO constructs NN representations of the feasible operating space using only non-sensitive active power data, supported by a novel and computationally efficient dataset-generation technique. In this way, the TSO avoids revealing any confidential system parameters. These NN models, complemented with valid inequalities, are transmitted to the market operator, who integrates them into a MILP-based market clearing formulation. The market operator similarly preserves its own

privacy by avoiding the disclosure of bids, contracts, or price-sensitive information. As demonstrated in numerical studies, the resulting market clearing decisions remain highly accurate while providing data privacy. This provides a concrete answer to the first research question by showing that the proposed framework can establish privacy-preserving yet operationally meaningful coordination between market operators and TSOs.

The second research question centered on the representation and utilization of DS flexibility in a manner that preserves DSO privacy while remaining tractable for TSO-level decision-making. Section 4 proposes a comprehensive PQV-FOR framework that significantly extends the classical two-dimensional PQ representation by adding voltage as a third dimension and incorporating heterogeneous FPU characteristics. Furthermore, a novel OPF-based sampling and analytical reconstruction methodology is proposed, enabling efficient construction of an implicit polynomial FOR representation in large-scale meshed DSs. By deriving analytical expressions for both feasibility boundaries and associated flexibility costs, the resulting PQV-FOR becomes directly compatible with system-level optimization and allows economic valuation of DS flexibility. Importantly, only these aggregated analytical models are exchanged with the TSO, while detailed DS information remains local to DSOs. Furthermore, the resulting FOR-based coordination framework operates in a single round of communication. These developments demonstrate that DS flexibility can be incorporated into transmission-level decision making while fully preserving data privacy.

While the FOR methodology successfully enables privacy-aware TSO-DSO coordination, its classical form is constrained by two structural limitations: it assumes a single PCC, and restricts the TSO to indirect dispatch of FPUs. The third research question, therefore, sought strategies to overcome these limitations to better align flexibility utilization. Section 5 develops an ML-based framework designed to support multi-PCC operation and enable the direct dispatch of DGs and FPUs from the TSO side. In this approach, DS behavior is represented through tailored NN architectures that map non-sensitive coupling variables. The resulting models allow the TSO to directly dispatch DGs and FPUs associated with multiple PCCs and DSs, thereby removing the need for hierarchical disaggregation. Simulation studies demonstrated that the proposed method maintains high accuracy even in meshed DSs connected through multiple PCCs, while solving AC-OPF problems. This chapter, therefore, provided a flexible, scalable, and privacy-preserving solution for direct utilization of DS flexibility, thereby addressing the third research question.

Taken together, the methods developed in Sections 3, 4, and 5 form a coherent methodological foundation for achieving interoperability in power systems under strict data privacy constraints. Each research question corresponds to a specific coordination context, market-TSO, TSO-DSO via aggregated flexibility, and multi-PCC TSO-DSO with direct dispatch, and each is answered with a tailored data-driven framework. Through these innovative methodological developments, the

research presented in this thesis contributes to enabling secure, effective, and privacy-preserving interoperability in modern power systems.

Although the proposed methods significantly advance privacy-preserving interoperability, several opportunities remain for future investigations. First, despite the high predictive accuracy of the developed ML models, approximation errors cannot be entirely eliminated. For industrial deployment, formal guarantees on feasibility, safety, and constraint satisfaction may be required. Future research could therefore focus on developing certification mechanisms or integrating verification techniques to ensure reliability under all operating conditions. Second, the current frameworks operate on a single-period basis, however some FPU, such as battery energy storage systems, exhibit inherently time-coupled behaviors, and operational constraints such as ramping limits play a critical role in real-world decision making. Extending the proposed methods to multi-period formulations would enable a comprehensive representation of the power system. Approaches such as physics-informed neural networks offer promising pathways to embed domain knowledge into models, potentially improving generalization and robustness while reducing the required volume of training data.

List of Figures

2.1	Comparison of traditional and modern power systems [32].	10
2.2	Redispatch costs in Germany [38].	14
2.3	Renewable energy curtailment in California [39].	15
2.4	Balancing service cost in Great Britain [40].	16
2.5	Example coordination scheme of power system stakeholders [48].	18
3.1	The schematic representation of the proposed method.	34
3.2	The projection of the multi-step dataset creation methodology in two dimensions. . .	39
3.3	Flowchart for the proposed methodology.	47
3.4	The histogram of the total cost differences for different cost coefficients.	51
3.5	The histogram of the total cost differences for different cost coefficients and constraints.	53
4.1	Schematic representation of the proposed method.	63
4.2	The sampling procedure with a) BBPS b) FDS c) LHS.	64
4.3	Single line diagrams of the DSs and characteristics of the DGs.	74
4.4	a) Dataset generated using BBPS b) Dataset generated using FDS c) Constructed FOR functions (with z-score normalization) d) Cost distribution obtained via LHS. Active and reactive power are expressed in MW and MVAR, respectively, while voltage is expressed in p.u.	75
4.5	Two-dimensional slice plots of the FOR for Case 533, illustrating the sampled boundary points and the surface of the fitted function. Slices are taken at fixed (a) active power, (b) reactive power, and (c) voltage, respectively. All axis values correspond to z-score normalized quantities. Figure titles correspond to the slice value in normalized and actual units.	76
4.6	Change of performance metrics according to polynomial degree.	79
4.7	Histogram of total cost and computational time differences, with AC- OPF as the reference, for integrated TS-DS. Each TS is combined with three analytically represented DSs (Case 33bw, Case 136, and Case 533).	81
4.8	Distribution of cost differences between proposed PQV-FOR and PQ- FOR with fixed PCC voltage (PQV-FOR is used as the reference). Outliers are omitted from the plots.	83

5.1	Schematic representation of the proposed method.	96
5.2	Data sampling approach using LHS.	97
5.3	Schematic representation of the polytope.	100
5.4	The architecture of the novel tailored NN.	101
5.5	Single-line diagram of the integrated power system.	104
5.6	Convex polygon PQ characteristics for DGs in the third DS.	105
5.7	a) Dataset indicating feasible and infeasible samples. b) Feasible region approximation of the NN indicating the facets.	106
5.8	Change of metrics according to conservativeness parameter.	108
5.9	Feasibility and mean cost difference according to conservativeness parameter, taking AC-OPF as reference.	109
5.10	Cost difference comparison under varying and constant PCC voltages.	111
5.11	Histogram of the computational time differences taking AC-OPF as reference.	112
5.12	Histogram of total cost and computational time differences for scala- bility analysis, with AC-OPF taken as the reference.	113

List of Tables

- 3.1 Summary of the Dataset Creation 48
- 3.2 Metrics for the NN Classification 49
- 3.3 Metrics for the NN Regression 50
- 3.4 Comparison of Feasibility Ratio and Average Cost Difference for
Different Cost Coefficients 50
- 3.5 Comparison of Feasibility Ratio and Average Cost Difference for
Different Cost Coefficients and Constraints 52
- 4.1 Overview of the Related Work. 59
- 4.2 Parameters for the Design of the Functions 75
- 4.3 Performance Metrics of the FOR Functions 77
- 4.4 Performance Metrics of the Cost Functions 77
- 4.5 Sample Generation Time (s) for Different Distribution Systems 78
- 4.6 Comparison of Standard (Centralized), ADMM-based, and Proposed
TSO-DSO Coordination Methods 84
- 5.1 Accuracy, Recall and Specificity Metrics of the NN Models 106
- 5.2 RMSE and MAE Metrics of the Quadratic Regression Models 108
- 5.3 Metrics of the NN Models under Constant PCC Voltage 110
- 5.4 Metrics of the NN Models for Scalability Analysis 113

References

- [1] Y. Ruwaida, J. P. Chaves-Avila, N. Etherden, I. Gomez-Arriola, G. Gürses-Tran, K. Kessels, C. Madina, A. Sanjab, M. Santos-Mugica, D. N. Trakas, *et al.*, “TSO-DSO-customer coordination for purchasing flexibility system services: Challenges and lessons learned from a demonstration in Sweden,” *IEEE Transactions on Power Systems*, vol. 38, no. 2, pp. 1883–1895, 2022.
- [2] C. Ziesemann, F. Gaumnitz, C. S. Köhnen, and I. E. Stoyanova, *Challenges and Barriers to the Implementation of TSO-DSO Coordination Concepts: Discussion Paper*. Universitätsbibliothek der RWTH Aachen, 2023.
- [3] S. Zhang, S. Maharjan, L. A. Bygrave, and S. Yu, “Data sharing, privacy and security considerations in the energy sector: A review from technical landscape to regulatory specifications,” *arXiv preprint arXiv:2503.03539*, 2025.
- [4] A. Babiker, S. S. Ahmad, I. Ahmed, M. Khalid, M. A. Abido, and F. S. Al-Ismail, “Optimal power flow: A review of state-of-the-art techniques and future perspectives,” *IEEE Access*, 2025.
- [5] Z. Wang, A. Younesi, M. V. Liu, G. C. Guo, and C. L. Anderson, “AC optimal power flow in power systems with renewable energy integration: A review of formulations and case studies,” *IEEE Access*, vol. 11, pp. 102681–102712, 2023.
- [6] F. Fioretto, T. W. Mak, and P. Van Hentenryck, “Differential privacy for power grid obfuscation,” *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1356–1366, 2019.
- [7] T. W. Mak, F. Fioretto, and P. Van Hentenryck, “Privacy-preserving obfuscation for distributed power systems,” *Electric Power Systems Research*, vol. 189, p. 106718, 2020.
- [8] A. Holzmann, L. Zwieb, K. Knaus, A. Harrucksteiner, and A. Sahin, “Increasing transparency for flow-based market coupling in european electricity trading,” *Vienna: Austrian Energy Agency GmbH*, 2020.

-
- [9] Y. Tong, J. Deyton, J. Sun, and F. Li, " S^3A : A secure data sharing mechanism for situational awareness in the power grid," *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 1751–1759, 2013.
- [10] J. Ringelstein, M. Vogt, A. M. Khavari, R. Ciavarella, M. Di Somma, and G. Graditi, "A methodology for improved TSO-DSO coordination in grid operation planning," *Electric Power Systems Research*, vol. 211, p. 108445, 2022.
- [11] G. Migliavacca, M. Rossi, H. Gerard, M. Džamarija, S. Horsmanheimo, C. Madina, I. Kockar, G. Leclecq, M. Marroquin, and H. Svendsen, "TSO-DSO coordination and market architectures for an integrated ancillary services acquisition: the view of the SmartNet project," *CIGRE*, 2018.
- [12] Y. Zahraoui, B. R. Pokhrel, I. Oleinikova, and S. N. Gros, "Assessment of TSO-DSO coordination to enhance flexible grid operation," *Current Sustainable/Renewable Energy Reports*, vol. 12, no. 1, p. 20, 2025.
- [13] A. Vicente-Pastor, J. Nieto-Martin, D. W. Bunn, and A. Laur, "Evaluation of flexibility markets for retailer-DSO-TSO coordination," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 2003–2012, 2018.
- [14] S. I. Vagropoulos, P. N. Biskas, and A. G. Bakirtzis, "Market-based TSO-DSO coordination for enhanced flexibility services provision," *Electric Power Systems Research*, vol. 208, p. 107883, 2022.
- [15] S. Mittal, P. Pareek, and A. Verma, "Convexified flexibility area identification at TSO-DSO interface," *IEEE Transactions on Industry Applications*, 2024.
- [16] G. Papazoglou and P. Biskas, "Review of methodologies for the assessment of feasible operating regions at the TSO-DSO interface," *Energies*, vol. 15, no. 14, p. 5147, 2022.
- [17] L. Ageeva, M. Majidi, and D. Pozo, "Coordination between TSOs and DSOs: Flexibility domain identification," in *The 12th Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion (MEDPOWER 2020)*, vol. 2020, pp. 429–434, IET, 2020.
- [18] S. Mittal and A. Verma, "Quantifying the flexibility at the tso-dso interface under uncertainties," in *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, pp. 1–5, IEEE, 2024.
- [19] R. Vijay and P. Mathuria, "Feasibility and flexibility regions estimation at TSO-DSO interconnection node using grid structure optimization," *Sustainable Energy, Grids and Networks*, vol. 32, p. 100952, 2022.

-
- [20] A. Patig, O. Stanojev, P. Aristidou, A. Kiprakis, and G. Hug, "Fast mapping of flexibility regions at TSO-DSO interfaces under uncertainty," in *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pp. 1–6, IEEE, 2022.
- [21] S. Mittal, P. Pareek, and A. Verma, "Assessment of flexibility region at TSO-DSO interface with dispatchable resources," in *2022 22nd National Power Systems Conference (NPSC)*, pp. 500–505, IEEE, 2022.
- [22] J. Silva, J. Sumaili, B. Silva, L. Carvalho, F. Retorta, M. Staudt, and V. Miranda, "A data-driven approach to estimate the flexibility maps in multiple TSO-DSO connections," *IEEE Transactions on Power Systems*, vol. 38, no. 2, pp. 1908–1919, 2022.
- [23] C. Gu, J. Wang, and L. Wu, "Distributed energy resource and energy storage investment for enhancing flexibility under a TSO-DSO coordination framework," *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 3, pp. 2961–2973, 2023.
- [24] H. Bakhtiari, M. R. Hesamzadeh, and D. W. Bunn, "TSO-DSO operational coordination using a look-ahead multi-interval framework," *IEEE Transactions on Power Systems*, vol. 38, no. 5, pp. 4221–4239, 2022.
- [25] **B. Dindar**, C. B. Saner, H. K. Çakmak, and V. Hagenmeyer, "Machine learning-driven multi-agent-based AC optimal power flow with effective dataset creation for data privacy and interoperability," *Sustainable Energy, Grids and Networks*, vol. 42, p. 101672, 2025.
- [26] **B. Dindar**, C. B. Saner, H. K. Çakmak, and V. Hagenmeyer, "Analytical representation and monetization of distribution system flexibility for TSO-DSO coordination via advanced sampling techniques," *IEEE Transactions on Smart Grid*, 2026. Accepted, in press.
- [27] **B. Dindar**, C. B. Saner, H. K. Çakmak, and V. Hagenmeyer, "Privacy-preserving utilization of distribution system flexibility for enhanced TSO-DSO interoperability: A novel machine learning-based optimal power flow approach," *Applied Energy*, vol. 414, p. 127848, 2026.
- [28] **B. Dindar**, C. B. Saner, D. Y. Polat, H. K. Çakmak, and V. Hagenmeyer, "A machine learning-based privacy-preserving approach to incorporate distributed generators in AC optimal power flow," in *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, pp. 1–6, IEEE, 2024.
- [29] **B. Dindar**, C. B. Saner, H. K. Çakmak, and V. Hagenmeyer, "TSO-DSO interaction: Privacy-preserving optimal power flow with distributed generators using a machine learning-based approach," in *2023 IEEE PES 15th Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pp. 1–6, IEEE, 2023.

-
- [30] Z. Jahangiri, R. Hendriks, and M. McPherson, “A machine learning approach to analysis of Canadian provincial power system decarbonization,” *Energy Reports*, vol. 11, pp. 4849–4861, 2024.
- [31] N. H. Commission *et al.*, “China health statistics year book 2020,” 2020.
- [32] R. Štefko, Z. Čonka, and M. Kolcun, “Case study of power plants in the Slovak republic and construction of microgrid and smart grid,” *Applied Sciences*, vol. 11, no. 11, p. 5252, 2021.
- [33] M. A. Abdullah, K. M. Muttaqi, and A. P. Agalgaonkar, “Sustainable energy system design with distributed renewable resources considering economic, environmental and uncertainty aspects,” *Renewable Energy*, vol. 78, pp. 165–172, 2015.
- [34] Y. Zhang, J. Wang, and Z. Li, “Uncertainty modeling of distributed energy resources: Techniques and challenges,” *Current Sustainable/Renewable Energy Reports*, vol. 6, no. 2, pp. 42–51, 2019.
- [35] A. Gautier, J. Jacqmin, and J.-C. Poudou, “The prosumers and the grid,” *Journal of Regulatory Economics*, vol. 53, no. 1, pp. 100–126, 2018.
- [36] Z. Yuan and M. R. Hesamzadeh, “Hierarchical coordination of TSO-DSO economic dispatch considering large-scale integration of distributed energy resources,” *Applied energy*, vol. 195, pp. 600–615, 2017.
- [37] E. Ela, V. Diakov, E. Ibanez, and M. Heaney, “Impacts of variability and uncertainty in solar photovoltaic generation at multiple timescales,” tech. rep., National Renewable Energy Laboratory (NREL), Golden, CO (United States), 2013.
- [38] ENTSO-E, “Costs of congestion management.” <https://transparency.entsoe.eu/congestion-management/r2/costs/show>, 2025.
- [39] California ISO, “Managing the evolving grid.” <https://www.caiso.com/about/our-business/managing-the-evolving-grid>, 2025.
- [40] National Energy System Operator (NESO), “2025 annual balancing costs report.” <https://www.neso.energy/document/362561/download>, 2025.
- [41] C. Büttner and U. Müller, “The impact of redispatch on grid and storage expansion planning in the German energy system,” *ACM SIGENERGY Energy Informatics Review*, vol. 4, no. 4, pp. 129–141, 2025.

-
- [42] F. Probst, C. Wagner, M. Greve, A. Nouvirdourst, P. Willers, A. Reinecke, and P. Kien-scherf, “Abschätzung der netzausbaukosten und die resultierenden netzentgelte für baden-württemberg und deutschland zum jahr 2045. studie der ef.-ruhr gmbh und ewi universität köln,” 2024.
- [43] W. Cole, A. W. Frazier, and C. Augustine, “Cost projections for utility-scale battery storage: 2021 update,” tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2021.
- [44] S. Dalhues, Y. Zhou, O. Pohl, F. Rewald, F. Erlemeyer, D. Schmid, J. Zwartscholten, Z. Hagemann, C. Wagner, D. M. Gonzalez, *et al.*, “Research and practice of flexibility in distribution systems: A review,” *CSEE Journal of Power and Energy Systems*, vol. 5, no. 3, pp. 285–294, 2019.
- [45] Bundesverband der Energie- und Wasserwirtschaft, “Redispatch 2.0.” <https://www.bdeew.de/energie/redispatch-20/>, 2023.
- [46] G. Blumberg, C. Schneller, H. Schuster, F. Ocker, S. Ried, and J. Stenglein, “Redispatch 3.0:—regulatorischer rahmen, markt-und produktdesign,” *Zielmodell für eine ergänzende markt-basierte Einbindung kleinteiliger dezentraler Flexibilitäten in den Redispatch-Prozess. Aachen, Stuttgart, Bayreuth.*, vol. 2, p. 2024, 2022.
- [47] A. Kocher, J. Kovačević, A.-C. Süß, C. B. Saner, **B. Dindar**, M. A. Ehsan, H. K. Çakmak, and V. Hagenmeyer, “Privacy protection in collaborative geographically distributed co-simulation of multimodal energy systems,” in *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*, pp. 639–647, 2024.
- [48] H. G. Morais Valente, E. Lambert, and J. Cantenot, “Chapter 7 - transmission system operator and distribution system operator interaction,” in *Local Electricity Markets* (T. Pinto, Z. Vale, and S. Widergren, eds.), pp. 107–125, Academic Press, 2021.
- [49] S. S. Ottenburger, D. Trybushnyi, H. K. Çakmak, **B. Dindar**, B. Chowdhury, R. Mikut, and V. Hagenmeyer, “Smart adaptive power distribution via hierarchical communication and action cascades under contingency conditions,” *IEEE Transactions on Smart Grid*, 2026, submitted.
- [50] V. S. Rajkumar, A. Ştefanov, A. Presekál, P. Palensky, and J. L. R. Torres, “Cyber attacks on power grids: Causes and propagation of cascading failures,” *IEEE Access*, vol. 11, pp. 103154–103176, 2023.

-
- [51] V. S. Rajkumar, A. Ştefanov, J. L. R. Torres, and P. Palensky, “Dynamical analysis of power system cascading failures caused by cyber attacks,” *IEEE Transactions on Industrial Informatics*, vol. 20, no. 6, pp. 8807–8817, 2024.
- [52] S. Ahmadian, X. Tang, H. A. Malki, and Z. Han, “Modelling cyber attacks on electricity market using mathematical programming with equilibrium constraints,” *IEEE Access*, vol. 7, pp. 27376–27388, 2019.
- [53] P. O. Ajiboye, K. O.-B. O. Agyekum, and E. A. Frimpong, “Privacy and security of advanced metering infrastructure (AMI) data and network: a comprehensive review,” *Journal of Engineering and Applied Science*, vol. 71, no. 1, p. 91, 2024.
- [54] **B. Dindar**, D. Müller, M. McCulloch, H. K. Çakmak, V. Hagenmeyer, and D. Graeber, “Privacy-preserving flexibility provisioning with an optimization approach for enhanced TSO-DSO interaction in modern power systems,” in *2025 21st International Conference on the European Energy Market (EEM)*, pp. 1–6, IEEE, 2025.
- [55] D. K. Molzahn, F. Dörfler, H. Sandberg, S. H. Low, S. Chakrabarti, R. Baldick, and J. Lavaei, “A survey of distributed optimization and control algorithms for electric power systems,” *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2941–2962, 2017.
- [56] T. Mühlfordt, X. Dai, A. Engelmann, and V. Hagenmeyer, “Distributed power flow and distributed optimization—formulation, solution, and open source implementation,” *Sustainable Energy, Grids and Networks*, vol. 26, p. 100471, 2021.
- [57] X. Dai, A. Kocher, J. Kovačević, **B. Dindar**, Y. Jiang, C. Jones, H. K. Çakmak, and V. Hagenmeyer, “Ensuring data privacy in AC optimal power flow with a distributed co-simulation framework,” *Electric Power Systems Research*, vol. 235, p. 110710, 2024.
- [58] Z. Erkin, J. R. Troncoso-Pastoriza, R. L. Lagendijk, and F. Pérez-González, “Privacy-preserving data aggregation in smart metering systems: An overview,” *IEEE Signal Processing Magazine*, vol. 30, no. 2, pp. 75–86, 2013.
- [59] V. Dvorkin, F. Fioretto, P. Van Hentenryck, P. Pinson, and J. Kazempour, “Differentially private optimal power flow for distribution grids,” *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2186–2196, 2020.
- [60] T. Miller, I. Durlik, E. Kostecka, P. Kozłowska, and A. Nowak, “Federated learning for decentralized electricity market optimization: A review and research agenda,” *Energies*, vol. 18, no. 17, p. 4682, 2025.

-
- [61] R. Zheng, A. Sumper, M. Aragüés-Peñalba, and S. Galceran-Arellano, "Advancing power system services with privacy-preserving federated learning techniques: A review," *IEEE access*, vol. 12, pp. 76753–76780, 2024.
- [62] O. Bertozzi, H. R. Chamorro, E. O. Gomez-Diaz, M. S. Chong, and S. Ahmed, "Application of data-driven methods in power systems analysis and control," *IET Energy Systems Integration*, vol. 6, no. 3, pp. 197–212, 2024.
- [63] B. Biswal, S. Deb, S. Datta, T. S. Ustun, and U. Cali, "Review on smart grid load forecasting for smart energy management using machine learning and deep learning techniques," *Energy Reports*, vol. 12, pp. 3654–3670, 2024.
- [64] A. Botterud, Z. Zhou, J. Wang, J. Sumaili, H. Keko, J. Mendes, R. J. Bessa, and V. Miranda, "Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: A case study of Illinois," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 1, pp. 250–261, 2012.
- [65] A. S. Zamzam and K. Baker, "Learning optimal solutions for extremely fast AC optimal power flow," in *2020 IEEE international conference on communications, control, and computing technologies for smart grids (SmartGridComm)*, pp. 1–6, IEEE, 2020.
- [66] R. Nellikkath and S. Chatzivasileiadis, "Physics-informed neural networks for AC optimal power flow," *Electric Power Systems Research*, vol. 212, p. 108412, 2022.
- [67] A. Aghazadeh Ardebili, O. Hasidi, A. Bendaouia, A. Khalil, S. Khalil, D. Luceri, A. Longo, E. H. Abdelwahed, S. Qassimi, and A. Ficarella, "Enhancing resilience in complex energy systems through real-time anomaly detection: a systematic literature review," *Energy Informatics*, vol. 7, no. 1, p. 96, 2024.
- [68] C. B. Saner, Y. Yaslan, and I. Genc, "An ensemble model for wide-area measurement-based transient stability assessment in power systems," *Electrical Engineering*, vol. 103, no. 6, pp. 2855–2869, 2021.
- [69] J. Cao, M. Zhang, and Y. Li, "A review of data-driven short-term voltage stability assessment of power systems: Concept, principle, and challenges," *Mathematical Problems in Engineering*, vol. 2021, no. 1, p. 5920244, 2021.
- [70] J. Rahman, C. Feng, and J. Zhang, "Machine learning-aided security constrained optimal power flow," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, IEEE, 2020.
- [71] E. I. A. (US), *Annual Energy Outlook 2012: With Projections to 2035*. Government Printing Office, 2012.

-
- [72] TransnetBW, “Studie zu redispatch 3.0 vorgestellt.” <https://www.transnetbw.de/de/newsroom/pressemitteilungen/studie-zu-redispatch-3-0-vorgestellt>, 2022.
- [73] M. McCulloch, J. Dierenbach, M. Baatar, D. Graeber, K. Tolstrup, S. Fanta, and V. Zobernig, “How interoperability of flexibility platforms enables market design opportunities,” in *2023 19th International Conference on the European Energy Market (EEM)*, pp. 1–4, IEEE, 2023.
- [74] Bundesverband der Energie- und Wasserwirtschaft, “enera.” <https://projekt-enera.de/>, 2020.
- [75] K. Christakou, D.-C. Tomozei, J.-Y. Le Boudec, and M. Paolone, “AC OPF in radial distribution networks—Part II: An augmented Lagrangian-based OPF algorithm, distributable via primal decomposition,” *Electric Power Systems Research*, vol. 150, pp. 24–35, 2017.
- [76] X. Dai, Y. Lian, Y. Jiang, C. N. Jones, and V. Hagenmeyer, “Hypergraph-based fast distributed AC power flow optimization,” in *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 4572–4579, IEEE, 2023.
- [77] M. Ryu and K. Kim, “A privacy-preserving distributed control of optimal power flow,” *IEEE Transactions on Power Systems*, vol. 37, no. 3, pp. 2042–2051, 2021.
- [78] X. Niu, H. K. Nguyen, J. Sun, and Z. Han, “Privacy-preserving computation for large-scale security-constrained optimal power flow problem in smart grid,” *IEEE Access*, vol. 9, pp. 148144–148155, 2021.
- [79] T. Wu, C. Zhao, and Y.-J. A. Zhang, “Privacy-preserving distributed optimal power flow with partially homomorphic encryption,” *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4506–4521, 2021.
- [80] N. Guha, Z. Wang, M. Wytock, and A. Majumdar, “Machine learning for AC optimal power flow,” *arXiv preprint arXiv:1910.08842*, 2019.
- [81] F. Hasan, A. Kargarian, and A. Mohammadi, “A survey on applications of machine learning for optimal power flow,” in *2020 IEEE Texas Power and Energy Conference (TPEC)*, pp. 1–6, IEEE, 2020.
- [82] M. Chatzos, F. Fioretto, T. W. Mak, and P. Van Hentenryck, “High-fidelity machine learning approximations of large-scale optimal power flow,” *arXiv preprint arXiv:2006.16356*, 2020.

-
- [83] X. Lei, Z. Yang, J. Yu, J. Zhao, Q. Gao, and H. Yu, "Data-driven optimal power flow: A physics-informed machine learning approach," *IEEE Transactions on Power Systems*, vol. 36, no. 1, pp. 346–354, 2020.
- [84] X. Pan, "DeepOPF: deep neural networks for optimal power flow," in *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pp. 250–251, 2021.
- [85] T. W. Mak, M. Chatzos, M. Tanneau, and P. Van Hentenryck, "Learning regionally decentralized ac optimal power flows with admittance," *IEEE Transactions on Smart Grid*, 2023.
- [86] I. Mezghani, S. Misra, and D. Deka, "Stochastic AC optimal power flow: A data-driven approach," *Electric Power Systems Research*, vol. 189, p. 106567, 2020.
- [87] A.-A. B. Bugaje, J. L. Cremer, and G. Strbac, "Split-based sequential sampling for real-time security assessment," *International Journal of Electrical Power & Energy Systems*, vol. 146, p. 108790, 2023.
- [88] A. Venzke, D. K. Molzahn, and S. Chatzivasileiadis, "Efficient creation of datasets for data-driven power system applications," *Electric Power Systems Research*, vol. 190, p. 106614, 2021.
- [89] B. Zhang and D. Tse, "Geometry of feasible injection region of power networks," in *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1508–1515, IEEE, 2011.
- [90] B. C. Lesieutre and I. A. Hiskens, "Convexity of the set of feasible injections and revenue adequacy in FTR markets," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1790–1798, 2005.
- [91] M. Fischetti and J. Jo, "Deep neural networks and mixed integer linear optimization," *Constraints*, vol. 23, no. 3, pp. 296–309, 2018.
- [92] F. Rossi, E. P. Araujo, M. C. Mañe, and O. G. Bellmunt, "Data generation methodology for machine learning-based power system stability studies," in *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pp. 1–5, IEEE, 2022.
- [93] J. Jalving, M. Eydenberg, L. Blakely, A. Castillo, Z. Kilwein, J. K. Skolfield, F. Boukouvala, and C. Laird, "Physics-informed machine learning with optimization-based guarantees: Applications to AC power flow," *International Journal of Electrical Power & Energy Systems*, vol. 157, p. 109741, 2024.

-
- [94] D. K. Molzahn, "Computing the feasible spaces of optimal power flow problems," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4752–4763, 2017.
- [95] M. Bichler and J. Knörr, "Getting prices right on electricity spot markets: On the economic impact of advanced power flow models," *Energy Economics*, vol. 126, p. 106968, 2023.
- [96] T. Joswig-Jones, K. Baker, and A. S. Zamzam, "OPF-learn: An open-source framework for creating representative AC optimal power flow datasets," in *2022 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, IEEE, 2022.
- [97] A.-A. B. Bugaje, J. L. Cremer, and G. Strbac, "Generating quality datasets for real-time security assessment: Balancing historically relevant and rare feasible operating conditions," *International Journal of Electrical Power & Energy Systems*, vol. 154, p. 109427, 2023.
- [98] I. V. Nadal and S. Chevalier, "Scalable bilevel optimization for generating maximally representative OPF datasets," in *2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, pp. 1–6, IEEE, 2023.
- [99] C. Bingane, M. F. Anjos, and S. Le Digabel, "Tight-and-cheap conic relaxation for the AC optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7181–7188, 2018.
- [100] S. Babaeinejadsarookolae, A. Birchfield, R. D. Christie, C. Coffrin, C. DeMarco, R. Diao, M. Ferris, S. Fliscounakis, S. Greene, R. Huang, *et al.*, "The power grid library for benchmarking ac optimal power flow algorithms," *arXiv preprint arXiv:1908.02788*, 2019.
- [101] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.
- [102] R. H. Byrd, J. Nocedal, and R. A. Waltz, "Knitro: An integrated package for nonlinear optimization," *Large-scale nonlinear optimization*, pp. 35–59, 2006.
- [103] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2023.
- [104] J. Löfberg, "YALMIP : A toolbox for modeling and optimization in MATLAB," in *In Proceedings of the CACSD Conference*, (Taipei, Taiwan), 2004.
- [105] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.
- [106] F. Chollet *et al.*, "Keras." <https://keras.io>, 2015.

-
- [107] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International journal of data mining & knowledge management process*, vol. 5, no. 2, p. 1, 2015.
- [108] A. G. Givisiez, K. Petrou, and L. F. Ochoa, "A review on TSO-DSO coordination models and solution techniques," *Electric Power Systems Research*, vol. 189, p. 106659, 2020.
- [109] A. Mohammadi, M. Mehrtash, and A. Kargarian, "Diagonal quadratic approximation for decentralized collaborative TSO+ DSO optimal power flow," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2358–2370, 2018.
- [110] F. Capitanescu, "TSO–DSO interaction: Active distribution network power chart for TSO ancillary services provision," *Electric Power Systems Research*, vol. 163, pp. 226–230, 2018.
- [111] S. Kundu, K. Kalsi, and S. Backhaus, "Approximating flexibility in distributed energy resources: A geometric approach," in *Power Systems Computation Conference (PSCC)*, pp. 1–7, IEEE, 2018.
- [112] E. Öztürk, K. Rheinberger, T. Faulwasser, K. Worthmann, and M. Preißinger, "Aggregation of demand-side flexibilities: A comparative study of approximation algorithms," *Energies*, vol. 15, no. 7, p. 2501, 2022.
- [113] D. A. Contreras and K. Rudion, "Computing the feasible operating region of active distribution networks: Comparison and validation of random sampling and optimal power flow based methods," *IET Generation, Transmission & Distribution*, vol. 15, no. 10, pp. 1600–1612, 2021.
- [114] S. Alsharif, M. Sarstedt, and E. M. Veith, "Comparison of heuristic optimization-based methods for determining the flexibility potential at vertical system interconnections," in *IEEE PES Innov. Smart Grid Technol. Conf. Europe (ISGT EUROPE)*, pp. 1–6, IEEE, 2023.
- [115] M. Heleno, R. Soares, J. Sumaili, R. J. Bessa, L. Seca, and M. A. Matos, "Estimation of the flexibility range in the transmission-distribution boundary," in *2015 IEEE Eindhoven PowerTech*, pp. 1–6, IEEE, 2015.
- [116] M. Sarstedt, L. Kluß, J. Gerster, T. Meldau, and L. Hofmann, "Survey and comparison of optimization-based aggregation methods for the determination of the flexibility potentials at vertical system interconnections," *Energies*, vol. 14, no. 3, p. 687, 2021.

-
- [117] M. Kalantar-Neyestanaki, F. Sossan, M. Bozorg, and R. Cherkaoui, "Characterizing the reserve provision capability area of active distribution networks: A linear robust optimization method," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2464–2475, 2019.
- [118] P. Fortenbacher and T. Demiray, "Reduced and aggregated distribution grid representations approximated by polyhedral sets," *International Journal of Electrical Power & Energy Systems*, vol. 117, p. 105668, 2020.
- [119] J. Silva, J. Sumaili, R. J. Bessa, L. Seca, M. A. Matos, V. Miranda, M. Caujolle, B. Goncer, and M. Sebastian-Viana, "Estimating the active and reactive power flexibility area at the TSO-DSO interface," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4741–4750, 2018.
- [120] L. Lopez, A. Gonzalez-Castellanos, D. Pozo, M. Roozbehani, and M. Dahleh, "Quickflex: a fast algorithm for flexible region construction for the tso-dso coordination," in *2021 International Conference on Smart Energy Systems and Technologies (SEST)*, pp. 1–6, IEEE, 2021.
- [121] A. Rabiee, R. J. Bessa, J. Sumaili, A. Keane, and A. Soroudi, "Exploiting the determinant factors on the available flexibility area of ADNs at TSO-DSO interface," *IET Renewable Power Generation*, vol. 18, no. 14, 2024.
- [122] L. Stark and L. Hofmann, "Novel approach for flexibility aggregation at multiple vertical TSO-DSO interconnections," in *Transformation der Stromversorgung–Netzregelung und Systemführung; 15. ETG/GMA-Fachtagung, Netzregelung und Systemführung*, VDE, 2024.
- [123] K. Yoon, J. Leem, S. H. Lee, and J.-W. Park, "Cooperative control of tso and dso: Management of line congestion and frequency response," *IEEE Access*, 2024.
- [124] L. Stark, M. Sarstedt, and L. Hofmann, "Determination of interdependent feasible operation regions at multiple TSO-DSO interconnections," in *2023 IEEE PES Innov. Smart Grid Technol. Conf. Europe (ISGT EUROPE)*, pp. 1–6, IEEE, 2023.
- [125] J. Liu, Z. Tang, Y. Liu, Y. Zhou, P. P. Zeng, and Q. Wu, "Region-inspired distributed optimal dispatch of flexibility providers in coordinated transmission-distribution framework," *Energy*, p. 134985, 2025.
- [126] D. A. Contreras, S. Müller, and K. Rudion, "Congestion management using aggregated flexibility at the TSO-DSO interface," in *2021 IEEE Madrid PowerTech*, pp. 1–6, IEEE, 2021.

-
- [127] F. Capitanescu, "OPF integrating distribution systems flexibility for TSO real-time active power balance management," in *Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion (MEDPOWER 2018)*, pp. 1–5, IET, 2018.
- [128] M. Kalantar-Neyestanaki and R. Cherkaoui, "Grid-cognizant TSO and DSO coordination framework for active and reactive power flexibility exchange: The Swiss case study," *Electric Power Systems Research*, vol. 235, p. 110747, 2024.
- [129] M. B. Bandeira, T. Faulwasser, and A. Engelmann, "An ADP framework for flexibility and cost aggregation: Guarantees and open problems," *Electric Power Systems Research*, vol. 234, p. 110818, 2024.
- [130] M. Sarstedt and L. Hofmann, "Monetarization of the feasible operation region of active distribution grids based on a cost-optimal flexibility disaggregation," *IEEE Access*, vol. 10, pp. 5402–5415, 2022.
- [131] A. Churkin, M. Sanchez-Lopez, M. I. Alizadeh, F. Capitanescu, E. A. M. Ceseña, and P. Mancarella, "Impacts of distribution network reconfiguration on aggregated DER flexibility," in *2023 IEEE Belgrade PowerTech*, pp. 1–7, IEEE, 2023.
- [132] F. Capitanescu, "Computing cost curves of active distribution grids aggregated flexibility for TSO-DSO coordination," *IEEE Transactions on Power Systems*, vol. 39, no. 1, pp. 2381–2384, 2023.
- [133] J. Bozzonek, T. Wolgast, and A. Nieße, "Design and evaluation of a multi-level reactive power market," *Energy Informatics*, vol. 5, no. 1, p. 6, 2022.
- [134] D. Huntington and C. Lyrantzis, "Improvements to and limitations of Latin hypercube sampling," *Probabilistic engineering mechanics*, vol. 13, no. 4, pp. 245–253, 1998.
- [135] M. M. Blane, Z. Lei, H. Civi, and D. B. Cooper, "The 3L algorithm for fitting implicit polynomial curves and surfaces to data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 3, pp. 298–313, 2002.
- [136] M. H. Stone, "The generalized Weierstrass approximation theorem," *Mathematics Magazine*, vol. 21, no. 5, pp. 237–254, 1948.
- [137] M. Hasanzadeh and A. Kargarian, "DPLib: A standard benchmark library for distributed power system analysis and optimization," *arXiv preprint arXiv:2506.20819*, 2025.

-
- [138] H. Jia, W. Qi, Z. Liu, B. Wang, Y. Zeng, and T. Xu, "Hierarchical risk assessment of transmission system considering the influence of active distribution network," *IEEE Transactions on Power Systems*, vol. 30, no. 2, pp. 1084–1093, 2014.
- [139] H. Gerard, E. Rivero, and D. Six, "Basic schemes for TSO-DSO coordination and ancillary services provision," *SmartNet Deliv. D*, vol. 1, p. 12, 2016.
- [140] X. Dai, Y. Guo, Y. Jiang, C. N. Jones, G. Hug, and V. Hagenmeyer, "Real-time coordination of integrated transmission and distribution systems: Flexibility modeling and distributed NMPC scheduling," *Electric Power Systems Research*, vol. 234, p. 110627, 2024.
- [141] L. Lind, R. Cossent, and P. Frías, "Evaluation of TSO–DSO coordination schemes for meshed-to-meshed configurations: Lessons learned from a realistic Swedish case study," *Sustainable Energy, Grids and Networks*, vol. 35, p. 101125, 2023.
- [142] G. Migliavacca, *TSO-DSO Interactions and Ancillary Services in Electricity Transmission and Distribution Networks: Modeling, Analysis and Case-Studies*. Springer, 2019.
- [143] M. Habibi, V. Vahidinasab, and M. S. Sepasian, "A privacy-preserving approach to day-ahead TSO-DSO coordinated stochastic scheduling for energy and reserve," *IET Generation, Transmission & Distribution*, vol. 16, no. 1, pp. 163–180, 2022.
- [144] T. W. Mak, F. Fioretto, L. Shi, and P. Van Hentenryck, "Privacy-preserving power system obfuscation: A bilevel optimization approach," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1627–1637, 2019.
- [145] F. Fioretto and P. Van Hentenryck, "Constrained-based differential privacy: Releasing optimal power flow benchmarks privately: Releasing optimal power flow benchmarks privately," in *Integration of Constraint Programming, Artificial Intelligence, and Operations Research: 15th International Conference, CPAIOR 2018, Delft, The Netherlands, June 26–29, 2018, Proceedings 15*, pp. 215–231, Springer, 2018.
- [146] X. Dai, J. Zhai, Y. Jiang, Y. Guo, C. N. Jones, and V. Hagenmeyer, "Advancing distributed AC optimal power flow for integrated transmission-distribution systems," *IEEE Transactions on Network Science and Engineering*, 2025.
- [147] T. Jiang, C. Wu, R. Zhang, X. Li, and F. Li, "Risk-averse TSO-DSOs coordinated distributed dispatching considering renewable energy and demand response uncertainties," *Applied Energy*, vol. 327, p. 120024, 2022.
- [148] A. Churkin, W. Kong, J. N. M. Gutierrez, E. A. M. Ceseña, and P. Mancarella, "Tracing, ranking and valuation of aggregated DER flexibility in active distribution networks," *IEEE Transactions on Smart Grid*, vol. 15, no. 2, pp. 1694–1711, 2023.

-
- [149] S. Wang and W. Wu, "Aggregate flexibility of virtual power plants with temporal coupling constraints," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5043–5051, 2021.
- [150] D. A. Contreras and K. Rudion, "Time-based aggregation of flexibility at the TSO-DSO interconnection point," in *IEEE Power & Energy Society General Meeting*, pp. 1–5, IEEE, 2019.
- [151] R. Vijay and P. Mathuria, "Complex power flexibility evaluation using energy arbitrage between transmission and distribution," *Electric Power Systems Research*, vol. 203, p. 107641, 2022.
- [152] M. Usman, M. I. Alizadeh, F. Capitanescu, I.-I. Avramidis, and A. G. Madureira, "A novel two-stage TSO–DSO coordination approach for managing congestion and voltages," *International Journal of Electrical Power & Energy Systems*, vol. 147, p. 108887, 2023.
- [153] X. Chen and N. Li, "Leveraging two-stage adaptive robust optimization for power flexibility aggregation," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 3954–3965, 2021.
- [154] E. Polymeneas and S. Meliopoulos, "Aggregate modeling of distribution systems for multi-period OPF," in *Power Systems Computation Conference (PSCC)*, pp. 1–8, IEEE, 2016.
- [155] H. Früh, S. Müller, D. Contreras, K. Rudion, A. von Haken, and B. Surmann, "Coordinated vertical provision of flexibility from distribution systems," *IEEE Transactions on Power Systems*, vol. 38, no. 2, pp. 1834–1844, 2022.
- [156] S. Riaz and P. Mancarella, "Modelling and characterisation of flexibility from distributed energy resources," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 38–50, 2021.
- [157] Z. Tan, H. Zhong, Q. Xia, C. Kang, X. S. Wang, and H. Tang, "Estimating the robust PQ capability of a technical virtual power plant under uncertainties," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4285–4296, 2020.
- [158] D. S. Stock, F. Sala, A. Berizzi, and L. Hofmann, "Optimal control of wind farms for coordinated TSO-DSO reactive power management," *Energies*, vol. 11, no. 1, p. 173, 2018.
- [159] Y. Xu, L. Yao, T. Pu, S. Liao, F. Cheng, Y. Li, and X. Wang, "Voltage-dependent PQ reserve capacity evaluation at TSO-DSO interface considering uncertainties of DGs and FLs," *CSEE Journal of Power and Energy Systems*, 2022.

-
- [160] ENTSO-E, “Towards smarter grids: developing TSO and DSO roles and interactions for the benefit of consumers,” Available: https://eepublicdownloads.entsoe.eu/clean-documents/Publications/Position%20papers%20and%20reports/150303_ENTSOE_Position_Paper_TSO-DSO_interaction.pdf, 2015.
- [161] D. A. C. Schneider, *Estimation of flexibility potentials in active distribution networks*, vol. 34. BoD–Books on Demand, 2021.
- [162] S. Riaz and P. Mancarella, “On feasibility and flexibility operating regions of virtual power plants and TSO/DSO interfaces,” in *2019 IEEE Milan PowerTech*, pp. 1–6, IEEE, 2019.
- [163] B. Liu, F. Liu, W. Wei, and J. Wang, “Estimating B-coefficients of power loss formula considering volatile power injections: an enhanced least square approach,” *IET Generation, Transmission & Distribution*, vol. 12, no. 12, pp. 2854–2860, 2018.

Relevant Publications

Journal Articles

Peer-Reviewed Articles:

- [25] **B. Dindar**, C. B. Saner, H. K. Çakmak, and V. Hagenmeyer, “Machine learning-driven multi-agent-based AC optimal power flow with effective dataset creation for data privacy and interoperability,” *Sustainable Energy, Grids and Networks*, vol. 42, p. 101672, 2025
- [26] **B. Dindar**, C. B. Saner, H. K. Çakmak, and V. Hagenmeyer, “Analytical representation and monetization of distribution system flexibility for TSO-DSO coordination via advanced sampling techniques,” *IEEE Transactions on Smart Grid*, 2026. Accepted, in press
- [27] **B. Dindar**, C. B. Saner, H. K. Çakmak, and V. Hagenmeyer, “Privacy-preserving utilization of distribution system flexibility for enhanced TSO-DSO interoperability: A novel machine learning-based optimal power flow approach,” *Applied Energy*, vol. 414, p. 127848, 2026
- [57] X. Dai, A. Kocher, J. Kovačević, **B. Dindar**, Y. Jiang, C. Jones, H. K. Çakmak, and V. Hagenmeyer, “Ensuring data privacy in AC optimal power flow with a distributed co-simulation framework,” *Electric Power Systems Research*, vol. 235, p. 110710, 2024

Articles Under Review:

- [49] S. S. Ottenburger, D. Trybushnyi, H. K. Çakmak, **B. Dindar**, B. Chowdhury, R. Mikut, and V. Hagenmeyer, “Smart adaptive power distribution via hierarchical communication and action cascades under contingency conditions,” *IEEE Transactions on Smart Grid*, 2026, submitted

Conference Proceedings

Peer-Reviewed Articles:

- [28] **B. Dindar**, C. B. Saner, D. Y. Polat, H. K. Çakmak, and V. Hagenmeyer, “A machine learning-based privacy-preserving approach to incorporate distributed generators in AC optimal power flow,” in *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, pp. 1–6, IEEE, 2024
- [29] **B. Dindar**, C. B. Saner, H. K. Çakmak, and V. Hagenmeyer, “TSO-DSO interaction: Privacy-preserving optimal power flow with distributed generators using a machine learning-based approach,” in *2023 IEEE PES 15th Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pp. 1–6, IEEE, 2023, (recipient of the Best Paper Award)
- [47] A. Kocher, J. Kovačević, A.-C. Süß, C. B. Saner, **B. Dindar**, M. A. Ehsan, H. K. Çakmak, and V. Hagenmeyer, “Privacy protection in collaborative geographically distributed co-simulation of multimodal energy systems,” in *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*, pp. 639–647, 2024
- [54] **B. Dindar**, D. Müller, M. McCulloch, H. K. Çakmak, V. Hagenmeyer, and D. Graeber, “Privacy-preserving flexibility provisioning with an optimization approach for enhanced TSO-DSO interaction in modern power systems,” in *2025 21st International Conference on the European Energy Market (EEM)*, pp. 1–6, IEEE, 2025