Advanced Uncertainty- and Physics-Aware Optimization for Energy Systems Operation

Ph.D. Dissertation submitted on the 15th July 2022 in partial fulfilment of the requirements to attain the degree of

Doctor of Philosophy in Engineering Science

presented by

Adriano Arrigo

Supervisors:

Prof. François Vallée	(UMONS, supervisor)
Prof. Jalal Kazempour	(DTU, co-supervisor)

Members of the examination jury:

Prof. Christophe Caucheteur	(UMONS, Chairman)
Prof. Yury Dvorkin	(JHU, External member)
Prof. Bertrand Cornélusse	(ULiège, External member)
Prof. Kenneth Bruninx	(TUDelft, External member)
Prof. Daniel Tuyttens	(UMONS, Internal member)
Dr Jean-François Toubeau	(UMONS, Internal member)
Prof. Zacharie De Grève	(UMONS, Secretary)

July 2022

Acknowledgements

This dissertation summarizes the work realized during the four years of my Ph.D. thesis that I pursued within the Power Systems & Markets Research group at the University of Mons. This period of time represents a very enriching period of my life, during which I met wonderful people.

First and foremost, I would like to thank my supervisor Prof. François Vallée, for giving me the opportunity to enter the academic world. His clear guidance and advices along the thesis have had a great positive impact on my personal and technical development. François has a constant joy of life, integrity and devotion to research which makes him the person who lightens your day and, as a researcher, your research work (and I needed these lights at multiple occasions). Beyond the impact that François had on my day-to-day work, today we may count on each other as friends. I will always be grateful for everything he brought me which will, without a doubt, shape the future of my life.

My day-to-day work has also been co-supervised by Prof. Jalal Kazempour. We first met during my master's thesis that I realized under his supervision during an Erasmus Exchange in Denmark. The student that I was at that time certainly did not realize the awesome impulse Jalal gave me towards the beginning of the Ph.D. thesis. I learnt a lot from Jalal and enhanced myself thank to him. The work realized during the four last years capitalizes on his expertise and knowledge in the fields of power systems, optimization and energy markets. Our discussions and meetings have always been a great pleasure and I hope that our ways may cross again in the future of our careers.

My next and special thanks go to Dr Jean-François Toubeau for his importance all along the pursuit of my thesis. He provided our discussions and meetings with plenty of original ideas that have shaped my thesis. He has triggered in me the taste for the good side of research. His devotion for equity, rigorous research and innovation, will remain an inspiring example to me. I wish you the best in the pursuit of his academic career. Thank you for everything !

I would like to thank Prof. Olivier Deblecker and Prof. Zacharie De Grève for the support they provided me. I would also like to thank Dr Jérémie Bottieau and Dr Martin Hupez for the friendly atmosphere they set up around me in the lab. We were "in the same boat" during our Ph.D. theses. Our correlated research topics have made us naturally collaborate on several projects, and I can only recommend their expertise in their respective fields of research. Thanks also to Manu, Thomas, Véronique, Charline, Alexis, François, Bashir, Behzad, Ahmad, Hooman, Jamal, Arnaud, Martin, Idriss, Thuy-Hai, Clélia, Louise and Aurélia for all the funny moments we have shared at the lab. Last but not least, I would like to thank all my partners during our unbelievably high-level table tennis sessions, they are memorable.

I would also like to thank the people I encountered from outside the lab. A special thank goes to the members of the examination committee for properly reviewing my manuscript, and especially the external members, Prof. Yury Dvorkin, Prof. Bertrand Cornélusse and Prof. Kenneth Bruninx. It has also been a great pleasure to exchange ideas and collaborate with Christos, Anna, Niklas, and Anubhav. I will be happy to meet you again.

This dissertation will focus on the ways of incorporating the uncertainties and complex operating constraints in decision-making problems related to energy systems operation. I will close my acknowledgements quoting Bertrand Russell on uncertainties :

When one admits that nothing is certain one must, I think, also admit that some things are much more nearly certain than others.

Adriano Arrigo

Abstract

In the current energy transition towards decarbonized energy systems, the mission for system operators is to continue maintaining a cost-efficient and reliable supply of energy. However, this is hindered by the increasing uncertainty and variability stemming from weather-dependent renewable energy sources which call system operators to adapt the short-term operational procedures. In that context, the main objective of this thesis is to develop novel data-driven optimization methods for decision-making in energy systems that embed an accurate representation of uncertainties, and non-convex network constraints to improve the cost-efficiency and reliability.

A first part of this thesis investigates uncertainty-aware optimal power flow problems where the system operator schedules the available generating units while being aware of uncertainties pertaining to, e.g., renewable power generation. In this direction, traditional stochastic optimization approaches have been extensively studied in the literature and applied to energy systems but they all rely on a unique probability distribution to describe the uncertainty, though a *true* probability distribution is not accessible in general. To address this issue, Distributionally Robust Optimization (DRO) has gained recognition in handling the potential misrepresentations of uncertainties in energy systems. To do so, DRO considers a family of probability distributions, i.e., the *ambiguity set*. Hence, the definition of ambiguity set is a prerequisite for the underlying performances of the obtained solution. Following this rationale, in this thesis, we explore additional features of uncertainty, such as physical bounds and dependence structure, to refine the definition of the ambiguity set. On the one hand, physical bounds allow to ensure that the uncertainty only takes realistic values, avoiding unrealistic ones such as, e.g., negative wind power generation. On the other hand, the dependence structure guarantees that the distributions within the ambiguity set follow a dependence structure which is consistent with the empirically observed one.

A second part of this thesis focuses on improving the representation of non-convex network constraints within the integrated operation of power and natural gas sys-

tems. Regarding electrical transmission, we consider the DC power flow equations that are usually deemed reasonable for high voltage transmission systems. The gas flow dynamics, however, are governed by the so-called Weymouth equation which is highly non-convex and challenging to include within optimization programs. To address this issue, the approaches based on second-order cone relaxation or linear interpolation usually allow to derive a tractable solution but they are not sufficiently tight for practical purpose and still need to address research questions. To bridge this gap, we introduce a new neural-network-constrained optimization framework for improving the modeling of gas flow dynamics within the integrated operation of energy systems. The method consists of capturing the non-convex Weymouth relation between pressures and gas flows via a neural network. Then, the neural network is encoded via a tractable mixed-integer program within the set of constraints.

Through the developments and experiments in this thesis, we show that distributionally robust and neural-network-constrained methodologies allow to make informed decisions which improve cost-efficiency and reliability of the underlying operational procedures. Overall, the developed methodologies show promising results which pave the way towards the data-driven operation of energy systems.

Contents

1 Introduction **1**
- 1.1 Background and Motivation 1
- 1.2 Literature Review and Research Questions 2
- 1.3 Research Contributions 7
- 1.4 Thesis Organization 8
- 1.5 List of Relevant Publications 10

2 The Energy Context **13**
- 2.1 The Energy Transition 14
- 2.2 Basic Operation of Energy Systems 16
- 2.3 Impacts of the Transition on the Energy Systems 24
- 2.4 Ensuring a Reliable Operation in an Uncertain Environment . . . 27
- 2.5 Chapter Summary and Conclusions 30

3 Distributionally Robust Optimal Power Flow **31**
- 3.1 Hands on Distributionally Robust Optimization 32
- 3.2 The Distributionally Robust OPF Problem 39
- 3.3 A State-of-The-Art Setting 44
- 3.4 Chapter Summary and Conclusions 49

4 An Exact and Physically-Bounded Distributionally Robust OPF **51**
- 4.1 Motivation 52
- 4.2 The Exact and Physically-Bounded DR-OPF Model 54
- 4.3 Numerical Experiments 63
- 4.4 Chapter Summary and Conclusions 76

5 Embedding Dependencies within Distributionally Robust OPF **77**
- 5.1 Motivation 78
- 5.2 The Moment-Metric-Based DR-OPF Model 81
- 5.3 The Copula-Based DR-OPF Model 86
- 5.4 Numerical Experiments 97
- 5.5 Chapter Summary and Conclusions 105

6 Distributionally Robust Optimization for the Coordinated Dispatch of Power and Gas **107**
6.1 Motivation . . . 108
6.2 The Distributionally Robust Coordinated Power and Gas Dispatch 110
6.3 Numerical Experiments . . . 117
6.4 Chapter Summary and Conclusions . . . 121

7 Machine Learning for Improved Gas Network Models **123**
7.1 Motivation . . . 124
7.2 The Neural-Network-Constrained Power and Gas Dispatch . . . 127
7.3 Benchmark Convexification Techniques . . . 135
7.4 Numerical Experiments . . . 139
7.5 Chapter Summary and Conclusions . . . 145

8 Conclusion **147**
8.1 Summary . . . 147
8.2 Conclusions . . . 148
8.3 Prospects . . . 150

Appendices **153**
A The IEEE 24-node Reliability Test System . . . 153
B The Belgian Power and Gas Systems . . . 155
C Procedure of the Out-of-Sample Analysis . . . 157
D Proof of the First Assertion in Theorem 5.1 . . . 160

List of Relevant Publications **165**

References **166**

Acronyms

ACE	:	Area Control Error.
AC-OPF	:	AC Optimal Power Flow.
aFRR	:	Automatic Frequency Restoration Reserve.
BRP	:	Balance Responsible Party.
BSP	:	Balancing Service Provider.
CVaR	:	Conditional Value-at-Risk.
DA	:	Day-Ahead.
DC-OPF	:	DC Optimal Power Flow.
DR-OPF	:	Distributionally Robust Optimal Power Flow.
DRCC	:	Distributionally Robust Chance Constraint.
DRO	:	Distributionally Robust Optimization.
DSO	:	Distribution System Operator.
EU	:	European Union.
FCR	:	Frequency Containment Reserve.
MILP	:	Mixed-Integer Linear Program.
ML	:	Machine Learning
mFRR	:	Manual Frequency Restoration Reserve.
NN	:	Neural Network.
NRV	:	Net Regulation Volume.
OPF	:	Optimal Power Flow.
PDE	:	Partial Differential Equation.
ReLU	:	Rectifier Linear Unit.
RES	:	Renewable Energy Source.
RT	:	Real-Time.
SDP	:	Semi-Definite Program.
SI	:	System Imbalance.
SOCP	:	Second-Order Cone Program.
TSO	:	Transmission System Operator.
VaR	:	Value-at-Risk.

Nomenclature

The main notations for sets, decision variables and parameters, are defined in the following.

Sets

$t \in \mathcal{T}$	:	Set of time periods.
$e \in \mathcal{E}$	:	Set of conventional generating units.
$\ell \in \mathcal{L}$	:	Set of electricity transmission lines.
$d \in \mathcal{D}$	:	Set of electricity demands.
$k \in \mathcal{K}$	:	Set of renewable generating units.
$g \in \mathcal{G}$	:	Set of gas suppliers.
$b \in \mathcal{B}$	:	Set of gas demands.
$m \in \mathcal{M}$	:	Set of gas network nodes.
$(m,n) \in \mathcal{Z}$	:	Set of gas pipelines.
$(m,n) \in \mathcal{C}$	:	Set of compressors.
$i \in \{1,...,N\}$	:	Set of scenarios for in-sample simulation.

Decision variables

p_e	:	Power generation of conventional generating unit e.
$\overline{r}_e$	:	Upward balancing capacity of conventional generating unit e.
$\underline{r}_e$	:	Downward balancing capacity of conventional generating unit e.
p_g	:	Gas supply from gas supplier g.
$q_{m,n,t}$	:	Gas flow in pipeline $(m,n) \in \mathcal{Z}$.
pr_m	:	Gas pressure at node m.
$Y_{e,w}$	:	Participation factor of generating unit e in system balancing.
$V_{g,w}$	:	Participation factor of gas supplier g in system balancing.

Uncertainty-related parameters

$\widetilde{\xi}$	:	Uncertain parameter.
$\widehat{\xi}_i$	:	Historical observation of uncertain parameter $\widetilde{\xi}$.
$\mathbb{Q}$	:	Probability distribution of uncertain parameter $\widetilde{\xi}$.

$\widehat{\mathbb{Q}}_N$	:	Empirical probability distribution of uncertain parameter $\widetilde{\xi}$, based on historical observations $\widehat{\xi}_i$.
μ	:	Day-ahead forecast of renewable power generation uncertainty.
Σ	:	Covariance matrix of uncertain parameter.
$\mathcal{M}(\Xi)$	:	Set of all probability distributions on the support Ξ.
$\mathcal{A}$	:	Ambiguity set.
θ	:	Wasserstein radius.
ϵ	:	Violation probability of operating constraints.
$\mathbb{C}$	:	Copula of uncertain parameter $\widetilde{\xi}$.
$\widehat{\mathbb{C}}_N$	:	Empirical copula of uncertain parameter $\widetilde{\xi}$, based on historical observations $\widehat{\xi}_i$.
Q	:	Matrix parameter for the definition of support.
h	:	Vector parameter for the definition of support.

Parameters

C_e	:	Marginal generation cost of conventional generating unit e.
$\overline{C}_e$	:	Procurement cost for upward balancing capacity of conventional generating unit e.
$\underline{C}_e$	:	Procurement cost for downward balancing capacity of conventional generating unit e.
C_g	:	Marginal gas supply cost of gas supplier g.
p_d	:	Consumption level of demand d.
$F_\ell^{\max}$	:	Capacity of transmission line ℓ.
$R_e^{\max}$	:	Maximum balancing capacity procurement of conventional generating unit e.
$\overline{P}_e$	:	Maximum generating limit of conventional generating unit e.
$\underline{P}_e$	:	Minimum generating limit of conventional generating unit e.
$M^{\mathcal{E}}$	:	Power transfer distribution factor for conventional generating units.
$M^{\mathcal{W}}$	:	Power transfer distribution factor for renewable generating units.
$M^{\mathcal{D}}$	:	Power transfer distribution factor for demands.
$\overline{\mathrm{PR}}$	:	Maximum pressure bound in gas system.
$\underline{\mathrm{PR}}$	:	Minimum pressure bound in gas system.
$K_{m,n}$	:	Pipeline Weymouth constant.
$S_{m,n}$	:	Pipeline linepack constant.
η_e	:	Conversion efficiency of gas-fired power plants.
$H_{m,n,0}$	:	Initial linepack in gas pipeline $(m,n) \in \mathcal{Z}$.

CHAPTER 1

Introduction

1.1 Background and Motivation

The need to address climate change has triggered a major transition in the energy sector which paves the way towards the growing integration of renewables in energy systems [1]. In this transitioning framework, the objective for system operators is to continue ensuring a cost-efficient and reliable supply of energy, which is being challenged in two different dimensions.

On the one hand, the *limited predictability* of renewable power generation induces uncertainty in the supply-side, which forces the system operators to rethink the existing short-term operational procedures (e.g., day-ahead scheduling, procurement and activation of operating reserves, network reconfiguration, ...). These procedures usually result in complex decision-making problems, whose solution may increase the operational risk if it does not properly accommodate the uncertain renewable power in-feeds [2]. On the other hand, natural-gas-fired power plants are becoming key flexible resources for mitigating the real-time *fluctuations* from renewable power generation, therefore increasing the need to improve the coordination between power and natural gas systems [3]-[4]. However, this is being hindered by the complexity arising from modeling network limitations in a coordinated framework, which are governed by non-convex equations. In the current practice, these equations are relaxed or approximated, potentially yielding non-trivial violations of network restrictions.

In this context, it becomes essential to adapt the operational procedures in energy systems to cope with the uncertainty and variability induced by weather-dependent renewable energy sources. In that regard, the recent advances in mathematical optimization methods foster the incorporation of probabilistic information and

improved network models in decision-making problems. The main rationale is that making more informed decisions (i.e., uncertainty and physics-aware) can help improve the cost-efficiency while maintaining the security of supply.

1.2 Literature Review and Research Questions

The objective of this thesis is to make a contribution towards addressing the aforementioned challenges by developing novel data-driven optimization methods for decision-making in energy systems[1] that embed an accurate representation of *i)* uncertainties, and *ii)* non-convex network constraints. In that direction, the focus is set on the Optimal Power Flow (OPF) problem, which represents an important part of short-term operational procedures. The OPF problem, aims at cost-effectively scheduling the available generating units, while ensuring the balance between generation and demand, as well as supply- and demand-side constraints and network limitations. Along this thesis, we adopt the system operator viewpoint, who solves an OPF problem in the day-ahead stage, and is exposed to uncertain deviations stemming from renewable power in-feeds between the day-ahead and real-time stages (i.e., the actual delivery stage).

In what follows, we review the literature regarding the uncertainty-aware optimization methods for OPF problem in Section 1.2.1, and the modeling of network limitations in Section 1.2.2. Both reviews lead to unexplored research questions that are the focus of the subsequent research.

1.2.1 Uncertainty-Aware Optimization for OPF Problem

In contrast with the current practice to solve OPF problems (i.e., based on single-point forecast of uncertainty), two main research strands exist in the literature to mitigate the uncertainty. The first one focuses on deriving *probabilistic forecasts*, whose accuracy has continuously increased in the last decades due to the advances in machine learning [5]-[6]. With access to probabilistic forecasts of renewable power generation, the system operator can make improved decisions, e.g., based on dedicated closed-form formulas. The second research strand relates to the incorporation of probabilistic information within the decision-making problems. The so-called *uncertainty-aware optimization* enables the system operator to make optimal data-driven decisions, that are informed with a probabilistic insight of uncertainty [7]. This second approach is considered in this thesis.

[1]This thesis is mainly concerned with energy systems, combining electricity and natural gas.

Traditional approaches[2] are extensively studied in the literature, among which stochastic programming [10], robust optimization [11] and chance-constrained programming [12]. *Stochastic programming* aims at minimizing the operating cost in expectation over a discrete set of scenarios of uncertain parameters [13] -[14]. The size of the resulting problem to solve depends on the number of in-sample scenarios, therefore ensuing a trade-off between the solution quality and computational time [15]. To counter this limitation, *robust optimization* [16] has gained attention as it relieves the computational burden. Robust decisions are optimal for the worst-case realization within an uncertainty set, which collects all potential realizations of the uncertain parameters. However, they are known to potentially produce over-conservative solutions. To reduce the conservativeness of robust solutions, *chance-constrained programming* can offer an alternative that usually suits well the definition of reliability criterion in energy systems [2]. This technique uses chance constraints which enforce the operating restrictions (e.g., transmission capacity limits, supply- and demand-side operational constraints) with high probability [17].

Nonetheless, *all* these techniques have their own limitations and drawbacks [18]. In particular, they all assume the availability of a distribution function that describes accurately the uncertainty. Though, in real-life applications, the *true* distribution function is not accessible, and probabilistic forecasts are never 100% accurate. Therefore, any uncertainty modeling technique relying on a specific distribution may fail in achieving the optimal dispatch of generating units. To address this issue, *Distributionally Robust Optimization (DRO)* has been re-introduced[3] recently in the operations research literature [21]. The technique has then gained attention in the power system literature due to its ability to hedge against the misrepresentations of uncertainty (e.g., in the form of biased probabilistic forecasts). To do so, DRO considers an *ambiguity set*, i.e., a collection of candidate distributions for describing the uncertainty. The optimal decisions are then minimized in expectation for the worst-case distribution within the ambiguity set. It has been shown that this setting may outperform the traditional approaches, in particular when the probabilistic forecasts are proven inaccurate (e.g., when few scenarios are available) [22]. In this perspective, the definition of ambiguity set is a prerequisite for the underlying performance of the distributionally robust solutions. There exist different definitions of ambiguity sets in the literature, two of which have been mainly introduced in the power system literature [23], namely

[2]These traditional approaches are currently being implemented in real-life applications. For instance, the Swiss balancing capacity market uses stochastic programming to determine the optimal amount of operating reserves to be booked in day-ahead [8]. The same applies in Sweden where chance constraints are leveraged to optimally schedule operating reserves [9].

[3]The introduction of DRO in the literature can be traced back to 1958 [19]. Recent developments allowed to derive tractable formulations that can be solved efficiently by computers [20].

the moment-based and the Wasserstein metric-based ambiguity sets.

The *moment-based* ambiguity set collects the distributions with the same moments, e.g., mean and covariance, achieved from the historical data. We refer to [23–26] for various applications of moment-based DRO to different problems in power systems. The shortcoming of this technique is that the whole empirical data is reduced to the value of moments only, therefore overlooking additional empirical information. In this work, the emphasis is set on the *Wasserstein metric-based* ambiguity set, which collects the distributions whose distance from a central empirical distribution (which is assessed via the Wasserstein probability metric) must be lower than or equal to a predefined value, called *radius*. The latter can be seen as a tuning parameter to adjust the decision maker's risk attitude. For instance, a metric-based ambiguity set with a larger radius collects more potential distributions for renewable power generation uncertainty, resulting in a more conservative dispatch solution. References [27–29] make use of the metric-based DRO for various applications in power systems, including OPF problems.

Despite the appealing properties of Wasserstein metric-based DRO, there is no guarantee that the endogenously selected worst-case distribution exhibits the empirically observed characteristics of uncertainty. This can negatively impact the distributionally robust solution which may be optimized for an inexact representation of uncertainty. Therefore, in the recent literature, we observe a willingness to enhance the modeling of ambiguity sets by removing the unrealistic distributions [30]. Among the potential enrichment of Wasserstein ambiguity sets, leveraging the *support* information (i.e., physical bounds of uncertainty) enables excluding distributions with unrealistic realizations of uncertainty [20], such as, e.g., distributions with a negative renewable power generation. Differently, references [30] and [31] include *modality* information (i.e., the number of spikes in a probability distribution) within the ambiguity set to get rid of unrealistic distributions, e.g., those with two or more spikes. Another potential enhancement of ambiguity set consists in embedding information on *dependencies*, such that the distributions within the set follows a dependence structure which is similar to the empirically observed one [32, 33]. Following this rationale, the first research question in this thesis is envisioned from the perspective of the most recent advances in DRO [34], where exact formulations of metric-based Distributionally Robust counterparts are derived.

Research Question 1

How additional meaningful features of ambiguity set, such as physical bounds and dependence structure, may help improving the cost efficiency and reliability of distributionally robust decision-making in energy systems ?

1.2.2 Modeling Network Limitations

In parallel to the accommodation of uncertainties, system operators may improve the operational procedures based on OPF problems by having recourse to more accurate network models. In this research work, we consider the coordinated operation of electricity and natural gas systems as an extension of the OPF problem, in which the power and natural gas flows are governed by non-convex equations [35]. There exist two main approaches to deal with network constraints. The first consists in simply neglecting the network limitations, assuming that the producers and consumers may completely freely exchange energy as if they were directly connected to each others. This results in the so-called *economic dispatch*, which may provide infeasible solutions with respect to the physics of the power and natural gas networks. To resolve this issue and prevent congestions, the system operator must check the operating constraints and have recourse to redispatching actions, if necessary, to make the day-ahead schedule comply with the physical constraints [36]. This approach is the common practice in the EU, where the cost of redispatching actions increases the total system cost [37]. The second approach relates to the integration of a representation of the network topology and its limitations within the OPF problem. This second approach is commonly used in the US [38, 39] and resort to the alliance of engineering and economical concepts to perform cost-efficient dispatch solutions with regards to physics constraints. Aiming at improved operational procedures, we focus on the second approach and discuss the modeling aspects regarding the power and natural gas flow equations, in the following paragraphs.

Regarding electricity, the alternating current power flows are governed by non-convex equations involving quadratic terms and sine-cosine functions [40]. The version of OPF problem embedding this complete set of equations is called AC-OPF, and results in a non-convex optimization problem with non-linear equality constraints, which is hard to solve in practice. To deal with such non-convex equations, different (quadratic, conic, SDP) relaxation techniques are proposed in the literature [41] but finding a zero duality gap solution remains an open research question [42]-[43]. Alternatively, the so-called DC-OPF consists of a linear approximation of the non-convex power flow equations, which is commonly implemented in real-life OPF problems (e.g., in the US PJM market clearing, or in the EU flow-based market clearing for cross-zonal exchanges). We refer to [44–46] for works in the literature using the DC power flow model. Following the common practice, we assume a DC approximation of power flows, which provides reasonable results in general for high voltage high power transmission systems [47].

The gas flow dynamics in the gas transmission system are governed by highly non-convex equations, such as the Weymouth relation, which links the pressures

at inlet and outlet nodes to the flow within the pipeline via a quadratic equality. Different approaches exist in the literature to cope with the non-convexity of the Weymouth equation, which mainly rely on convexification techniques[4]. The most commonly used convexification technique is the so-called Second-Order Cone (SOC) relaxation [49] and its enhanced versions based on McCormick envelopes [50] or bound tightening algorithms [51]. However, these SOC-based relaxations remain loose and strongly jeopardizes the feasibility of the obtained solution. Differently, authors in [**?**] and [52] consider an interpolation of the Weymouth equation, via an incremental piecewise linear formulation, but the size of the resulting program to solve increases with the number of intervals, therefore ensuing a trade-off between the approximation error and the computational burden [53].

Overall, the previously explored convexification techniques still need to address fundamental research questions. In particular, they all rely on specific modeling assumptions and it is not straightforward to enhance these models for considering gas flow bidirectionality, see [54]. In order to reduce the inherent modeling errors arising from these convexification techniques, machine-learning-based optimization methods have been explored in the power system literature. A first approach consists in deriving a regression-based proxy of the solution to OPF problems [55–57]. These approaches reveal a substantial computational speed-up, however, their application in the scope of the power and gas dispatch problem is hindered by the number of operating points required in the training phase to yield an accurate model. To address this issue, several additional works consider the incorporation of machine learning models into the set of constraints. For instance, authors in [58] propose a neural-network-constrained AC-OPF encoding electricity dynamic security restrictions. Following the same rationale, machine learning can help improving the modeling of gas flow dynamics but limited research contributions have strived in that direction. To fill this gap in the current state of the art, it is important to address the following research question.

Research Question 2

How the recent breakthrough in neural network-constrained optimization may help modeling non-convex gas flow dynamics within integrated energy systems operation ?

[4]Convexification techniques are of two main types [48], namely, convex *relaxations* (which enlarge the original feasible set), and convex *approximations* (which may disregard some feasible solutions).

1.3 Research Contributions

The scientific publications related to the subsequent research contributions are reported under alphanumeric citations, and gathered within Section 1.5.

On the path towards addressing the aforementioned research questions, a first contribution in **Paper [G]** introduces a survey of the traditional uncertainty-aware decision-making approaches (i.e., stochastic programming, robust optimization and chance-constrained programming) and their limitations, based on an electricity-only OPF problem. Pursuing an improved modeling of uncertainty in OPF problems, we propose enhanced DRO methodologies (as listed in Section 1.3.1) which incorporate additional features of uncertainty (such as physical bounds and dependence structure) and bring useful insights for addressing Question 1. Next, we propose a neural-network-constrained optimization method in Section 1.3.2 to cope with the non-convexity arising from modeling gas flow dynamics, as a framework that sheds light on the answer to Question 2.

1.3.1 Enhanced DRO Methodologies

As a first step towards addressing Question 1, our contribution in **Paper [A]** adds a constraint binding the physical restrictions pertaining to the renewable power generation uncertainty (i.e., the support) to the exact formulation of Distributionally Robust Chance Constraints (DRCCs). Therefore, the framework avoids the unrealistic realizations of uncertainty (e.g., negative power generation) from the ambiguity set, while preserving exactness. We apply the resulting formulation to an OPF problem and derive a comparative discussion against the traditional scenario-based programming and the state-of-the-art distributionally robust counterparts.

As a second step, we explore the benefits of embedding the dependence information within the definition of ambiguity set. Two specific scientific contributions aim at addressing this challenge.

In **Paper [C]**, we combine support information with dependence structure, within the definition of ambiguity set. The proposed approach relies on a moment-metric-based DRO methodology, which includes a constraint on the second-order moment of uncertainty, i.e., the covariance matrix. The framework is capable of considering the space-time dependencies pertaining to the renewable power generation uncertainty.

Moving beyond the second-order moment of uncertainty which only allows to capture linear dependence structure, our second specific contribution in **Paper**

[B] considers the non-linear dependencies. In particular, we propose a new copula-based ambiguity set which is tailored to capture any type of dependencies among renewable power generating units.

Finally, **Paper [D]** explores the coordination of power and gas systems accounting for uncertainties in the multivariate and correlated energy demand. We propose a distributionally robust framework which leverages the previous developments for considering the dependencies among uncertain power and gas demands. We consider a case study inspired by the Belgian power and gas networks, showing the scalability of the proposed DRO methodologies.

1.3.2 Neural Networks for Modeling Gas Network

On the path towards addressing Question 2, **Papers [D] and [F]**[5] study the coordination of integrated energy systems. In these papers, the gas flow dynamics are approximated via a second-order cone relaxation which is sufficiently tight in the scope of these works, but may remain too loose (regarding ex-post feasibility of the obtained solutions) for real-life implementations.

Therefore, in **Paper [E]**, we propose a neural-network-constrained optimization method to improve the modeling of non-convex natural gas flow dynamics within the coordinated power and gas dispatch. The proposed framework includes a regression model of the Weymouth equation, based on supervised machine learning, which is encoded into the set of constraints as a mixed-integer program. We further enhance our model by introducing a reformulation of the leaky rectifier linear unit activation function, which improves the computational efficiency.

1.4 Thesis Organization

This thesis is structured as follows.

Chapter 2 brings additional insights contextualizing the research in the global energy sector framework, before diving into the technical matter. Essentially, this chapter includes basic concepts and key numbers regarding the operation of energy systems for the non-initiated reader.

Chapter 3 aims to introduce the mathematical background required for distributionally robust optimization and shows the numerical benefits pertaining to

[5] **Paper [F]** considers the coordination of power, heat, and natural gas systems and focuses on revealing the maximum "network flexibility," corresponding to the ability of natural gas and district heating pipelines to store energy.

DRO compared to the traditional approaches. As an introductory case study, we develop an electricity-only OPF problem and its DRO counterpart. Based on a simple illustrative example, we show the limitations of the state-of-the-art DRO methodologies.

Chapter 4 partially answers the first research question, by introducing an additional constraint on the support of uncertainty within the definition of exact DRCCs. The resulting model is bilinear and is solved efficiently via an alternating algorithm. The improvements in terms of total operating costs and ex-post operational violation probability are shown through an extensive numerical study.

Chapter 5 brings additional insights regarding the first research question, by embedding the dependence information into the ambiguity set definition. First, we investigate the moment-metric-based ambiguity set, which is an enriched version of the traditional ambiguity set with a constraint on covariance matrix. Next, we aim to study non-linear dependencies and propose a copula-based ambiguity set that is capable of capturing any type of dependence structure.

Chapter 6 gathers the previously acquired knowledge in DRO and proposes a distributionally robust coordinated power and gas dispatch model. The chapter sheds light on the benefits of the proposed DRO methodologies applied to a more complex problem and considering a case study inspired by the Belgian power and gas networks. The outcomes also allow to reveal limitations about the modeling of gas flow dynamics.

Chapter 7 reviews the traditional convexification approaches (i.e., relaxation and approximation) for handling the non-convex Weymouth equation, and introduces basic concepts of machine learning. Next, the second question is addressed by proposing a neural network-constrained power and gas dispatch problem, which embeds an accurate regression model of gas flow dynamics.

Chapter 8 closes the manuscript with summary, concluding remarks and prospects.

The developments underpinning the research contributions in this Ph.D. thesis are related to mathematical optimization. Overall, open-source programming languages are leveraged. In particular, we use Julia Programming Language [59], together with the modeling language JuMP [60]. Regarding the solvers that are required to optimize the resulting models, we use both Mosek [61] for conic and semi-definite programming problems and Gurobi for other types of mathematical programs. It is worth mentioning that all source codes are made publicly available in the different online appendices [62], [63], [64], [65].

1.5 List of Relevant Publications

The following publications report the research contributions that are embodied in this dissertation.

Journal Papers

[A] **A. Arrigo**, C. Ordoudis, J. Kazempour, Z. De Grève, J.-F. Toubeau, and F. Vallée, "Wasserstein distributionally robust chance-constrained optimization for energy and reserve dispatch: An exact and physically-bounded formulation," European Journal of Operations Research, vol. 269, no. 1, pp. 304-322, January 2022.

[B] **A. Arrigo**, J. Kazempour, Z. De Grève, J.-F. Toubeau, and F. Vallée, "Embedding Dependencies Between Wind Farms in Uncertainty-Aware Optimal Power Flow," 2022, submitted to IEEE Transactions on Power Systems, under the second round of review.

Conference Papers

[C] **A. Arrigo**, J. Kazempour, Z. De Grève, J.-F. Toubeau, and F. Vallée, "Enhanced Wasserstein Distributionally Robust OPF With Dependence Structure and Support Information," IEEE PowerTech Conference, Madrid, June 2021.

[D] **A. Arrigo**, J.-F. Toubeau, I. Fattahi, Z. De Grève and François Vallée, "Distributionally Robust Power and Gas Dispatch with Multivariate and Correlated Uncertainty," IEEE PMAPS Conference, Manchester, June 2022.

Other Reports

[E] **A. Arrigo**, M. Dolányi, K. Bruninx, J.-F. Toubeau, "Machine Learning for Improved Gas Network Modeling in Coordinated Energy Systems," Working paper[6].

The following works have been realized during the course of the PhD thesis. Their content is relevant but is not directly included in this dissertation.

Journal Papers

[F] A. Schwele, **A. Arrigo**, C. Vervaeren, J. Kazempour, and F. Vallée, "Coordination of electricity, heat, and natural gas systems accounting for network flexibility," Electric Power Systems Research, 189:106776, 2020.

[6] This work is foreseen to be submitted as a journal paper by October 2022.

Conference Papers

[G] **A. Arrigo**, C. Ordoudis, J. Kazempour, Z. de Grève, J.-F. Toubeau, and F. Vallée, "Optimal power flow under uncertainty: An extensive out-of-sample analysis," in 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGTEurope), pages 1–5, 2019.

[H] J. Bottieau, **A. Arrigo**, Z. D. Grève, F. Vallée, and J.-F. Toubeau, "A distributionally robust framework for providing passive balancing services," IEEE PowerTech Conference, Madrid, June 2021.

CHAPTER 2

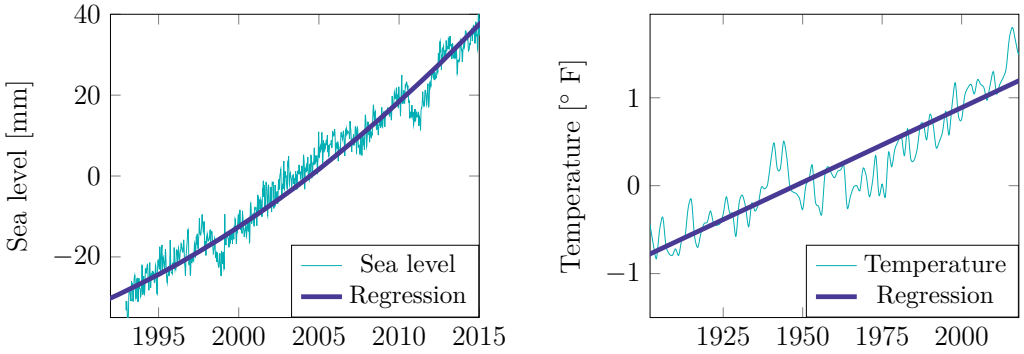
The Energy Context

Energy, in its many existing forms, facilitates the development and sustainability of the human civilization. Our intricate relation with energy and its diverse utilities (e.g., heating, cooking, farming, ...) has shaped our well-being and living throughout history. However, in spite of its physical abundance, useful energy is difficult to harvest and remains a scarce good. Therefore, the selection of the appropriate form to produce, store, transport and use energy still represents a strategic issue for countries, nowadays. Among the available forms of energy, electricity and natural gas elect as the most notable ones, due to their wide application range, high conversion efficiency and comparatively easy transportation through wires and pipelines. Nowadays, electricity and natural gas are used as the energy carriers of about 50 % of the main end utilities [66].

Following intergovernmental climate change agreements pursuing the decarbonization of our society, our utilization of energy is currently evolving. This *transition* impacts the overall energy sector, at all levels, and especially the electricity and natural gas industries. In this framework, the following introductory chapter aims at contextualizing the goals of the underlying research. First, the main drivers and key numbers of the energy transition are showcased in Section 2.1. Then, Section 2.2 presents the basic operation of power and natural gas systems. The impact of the energy transition on the system operation is the focus of Section 2.3. In particular, ensuring a reliable and cost-efficient system operation in an increasingly uncertain environment (e.g., stemming from weather-dependent renewable energy sources) is becoming challenging. In that regard, Section 2.4 identifies the key challenges that need to addressed within the current energy transition.

2.1 The Energy Transition

The rise in global temperature and the rise in sea level represent the most widely known symptoms witnessing the ongoing climate crisis. The evolution of these key indicators is illustrated in the graphs of Figure 2.1, with the regression curves showing their global increasing trend. The brutal variations (in such a comparatively short period of time) induced by this climate change have dramatic direct and indirect consequences for the ecosystems worldwide with, e.g., the increasing number of extreme weather events such as storms, hurricanes or floodings. Human-beings are taking part in these ecosystems, not only as victims of these consequences, but also as their causal agent. The partial responsibility of humanity in the climate change already took birth during the first industrial revolution (19^{th} century) when coal-fueled engines were introduced and allowed the mechanization of large-scale industrial processes. Following this, the discovery of oil and natural gas (and their intensive use in industry) have bolstered the emissions of greenhouse gases (which stem from the combustion of fossil fuels such as coal, oil or gas) and have caused the rise of the temperature via the well-known greenhouse effect [67].



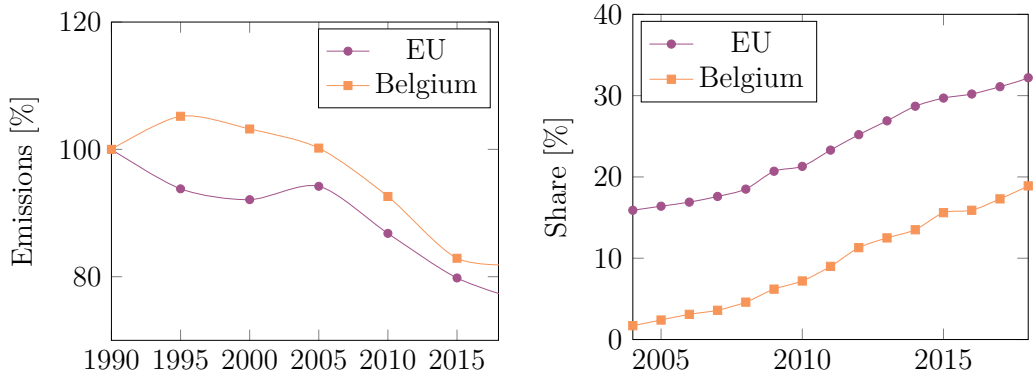
(a) Global average sea level [68].

(b) Global average temperature [69].

Figure 2.1: Evolution of key climate change indicators.

While becoming aware of its unprecedented influence on the climate system, humanity also endorses a key role in finding the solution to fight climate change. In that direction, the diverse intergovernmental agreements on climate change endeavour to tackle this enormous challenge. These agreements usually translate into European regulations and directives, which are eventually adopted into national laws. In particular, the EU envisions *carbon-neutrality*[1] by 2050 for the European Member States, which has been translated into the EU Green deal since

[1]Carbon neutrality refers to a system state where residual greenhouse gas emissions are completely balanced by their capture in the atmosphere.



(a) Greenhouse gas emissions compared to 1990 levels.

(b) Renewable share in electricity generation.

Figure 2.2: Achievements in terms of climate change goals in EU and Belgium [66].

2019 [70]. To pave the way towards this ambitious goal, the EU has primarily set the "20-20-20" targets, which correspond, respectively, to a 20% decrease in greenhouse gas emissions compared to the 1990 levels, a 20% increase in energy efficiency and an objective of 20% share of renewables in the energy mix, by the year 2020. These goals were achieved at the EU level in 2020. Following this, the EU leaders have set new goals, replacing the "20-20-20" targets at multiple occasions. Essentially, two legislative packages updated the European goals, namely, the "2030 Climate and Energy targets" (2014, [71]), and the "Clean Energy for all Europeans" (2019, [72])[2]. The current European goals for climate at the horizon 2030 are listed below:

The current European goals for climate at the horizon 2030:

- 40% decrease in greenhouse gas emissions compared to 1990 levels,
- 32% share of renewable energy sources in the EU's energy mix,
- 32.5% improvement in energy efficiency.

It is worth mentioning that the "20-20-20" targets were not achieved by all individual Member States (i.e., only 11 Member States achieved the combined realization of these three goals [74]). For instance, Belgium did not manage to achieve its second and third goals. The graphs in Figure 2.2 illustrate the efforts made by Belgium and the EU. In particular, the left-hand side plot refers to the greenhouse gas emissions compared to the emission levels in 1990 (i.e., the index equals 100% for the year 1990). The right-hand side plot refers to the

[2]The "Fit for 55" package (2021, [73]) is under debate to align the European objectives with the European climate ambitions at the horizon 2050, specifically aiming at a 55% decrease in greenhouse gas emissions compared to 1990 levels.

share of renewable energy sources in electricity generation. As it can be observed, these indicators are currently evolving, and the end of the transition period is yet unknown.

2.2 Basic Operation of Energy Systems

This thesis focuses on the operational procedures for the energy systems, e.g., the short-term scheduling of injections and offtakes. At the considered time horizon, the main degrees of freedom for the system operators include generation dispatch and reserve allocation[3]. However, this operation is currently being challenged by the energy transition (see Section 2.3 and 2.4 for the main impacts and challenges it may induce). In the following, we introduce the basic operation of power systems in Section 2.2.1 and that of natural gas systems in Section 2.2.2.

Historically, the electricity and natural gas systems were operated by a single entity. In the last decades, however, the energy sector has been liberalized by means of four legislative packages delivered by the European Commission (and a fifth one which is still under debate at the time of writing). The first energy package in 1996 (corresponding to directives 96/92/EC and 98/30/EC) resulted in the unbundling of the vertically-integrated monopolies and the aperture of liberalized electricity and natural gas markets. These measures primarily aimed at reduced electricity bills for the end-consumers and increased investments, notably in renewable energy installations. In this direction, new roles have been settled, redefining the framework of electricity and natural gas sectors.

The roles of *Suppliers* and *Regulators* have been introduced by means of the second energy package in 2003 (two directives 2003/54/EC and 2003/55/EC and a regulation No. 1228/2003). The regulators aim at protecting the final consumers by monitoring the energy prices and the illegal market behaviours, but also by ensuring that the European regulations are correctly integrated into the market operation. Suppliers are intermediate actors who make the link between the wholesale and retail markets, by selling energy to the end-users.

The third energy package was delivered in 2009 and introduced the adoption of two supplementary directives (2009/72/EC and 2009/73/EC) and three regulations (No. 713/2009, No. 714/2009 and No. 715/2009) which triggered the rules governing the electricity markets and strengthened the independence of regulation agencies. The ACER, ENTSO-E and ENTSO-G agencies were created, which correspond to the aggregation of, respectively, European regulators, electricity

[3]In contrast with the operation, the long-term planning of the system refers to investment decisions in, e.g., system or generating assets.

transmission system operators and gas transmission system operators.

The fourth energy package is also referred to as the "Clean Energy for all Europeans" package and it contains four directives and four regulations (Directives 2018/844, 2018/2001, 2018/2002, 2019/944 and regulations No. 2018/1999, No. 2019/941, No. 2019/942, No. 2019/943) which intend to deliver the EU's engagement on the Paris Agreement. In particular, this package sets new energy targets, but also new rules for electricity markets, consumer empowerment and risk preparedness.

The fifth energy package, "Delivering the European Green Deal", was released on 14th July 2021 with the aim at aligning the EU's energy targets with the new European climate ambitions for 2030 and 2050 [73]; the debate on its energy aspects is still ongoing at the time of writing.

2.2.1 Basic Operation of Power Systems

Electricity is a good that is exchanged between, e.g., electricity generating units and demand centers, via the electricity grid. The latter is composed of a transmission system (meshed network of high voltage nodes, transformers and transmission lines which run over long distances) and a distribution system (mainly radial network composed of medium to low voltage nodes and distribution lines). Therefore, the flows of electricity throughout the grid are not only governed by the laws of economics (i.e., offer should match demand) but also by the laws of physics (e.g., Kirchhoff's law, thermal operation limits, energy balance). Figure 2.3 illustrates the current organization of the power systems and the main actors involved.

The European commission regulation No. 2017/1485 establishing the power system operation guidelines [75] provides a legal framework for delivering the EU-wide power system operation security. These guidelines mainly infer to system operators, but also to significant grid users.

The *system operators* are the owner of the fixed infrastructure. Due to the high investment cost, system operators represent natural monopolies which are subject to regulation. At the transmission level, the Transmission System Operator (TSO) operates the high voltage high power electricity grid. At the distribution level, the Distribution System Operator (DSO) operates the medium-to-low voltage electricity grid. These entities must ensure a reliable and safe operation for the load-frequency control area they are responsible for, thereby, playing the role of facilitators for the electricity exchanges.

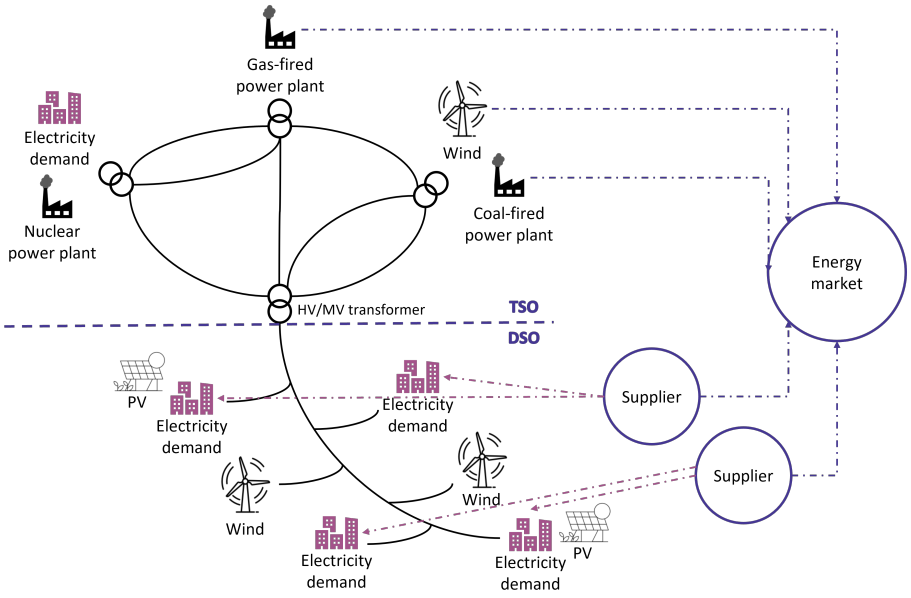


Figure 2.3: Basic operation of the power systems. Plain lines represent the flow of electricity, dashed lines represent the monetary flows. TSO: Transmission System Operator, DSO: Distribution System Operator, PV: Photovoltaic.

To ensure the safe and reliable operation, injections and off-takes must be continuously balanced over the electricity grid. However, system operators are prohibited to generate and sell electricity[4]. Therefore, TSOs outsource their balancing responsibility to *Balance Responsible Parties* (BRPs). These private legal entities are responsible for the balance at one or multiple access points of the transmission system. To safeguard the balance at these access points, BRPs establish a portfolio combining injections, off-takes, exchanges with other BRPs and cross-border exchanges. BRPs have to submit their nominations in day-ahead, which correspond to the scheduling of injections and off-takes pertaining to the assets within their portfolio.

In this direction, the *Producers* and *Consumers*, who are willing to exchange electricity throughout the grid, must have a contract with a BRP[5]. The producers generate electrical energy and inject it within the electricity grid. The off-takes by consumers are related to their utilities.

Albeit this settlement encouraging as much as possible the BRPs to maintain the balance within their portfolio, deviations may still occur with respect to the

[4]System operators are *only* allowed to own generating units that are intended to compensate the power losses within the lines.
[5]Large producers or consumers may act as their own BRP.

operational scheduling. In practice, this difference between the actual injections/offtakes and the scheduled ones is measured via the Area Control Error (ACE). In order to restore the balance between generation and consumption and to come back as close as possible to the generation schedule, the system operator activates balancing reserves (namely, frequency containment reserves, and frequency restoration reserves) in real-time. Since transmission and generation missions have been separated following the sector liberalization, TSOs must have recourse to *Balancing Service Providers* (BSPs) to keep the balance between generation and consumption. The BSPs procure generation capacity to TSOs that remains available to be activated in real-time, if required. There exist different types of balancing services that are offered via market-based auctions operated by the TSO, see [76]. The Net Regulation Volume (NRV) corresponds to the sum of activated balancing capacity. The residual System Imbalance (SI), which is not resolved, corresponds to the difference between the ACE and NRV and is an important economic signal

$$\mathrm{SI} = \mathrm{ACE} - \mathrm{NRV}. \tag{2.1}$$

Short-Term Electricity markets

Beyond operations, the physical exchanges throughout the grid involve monetary exchanges, that are settled in different short-term electricity markets. The electricity products are of two main types, i.e., *energy* and *capacity*. Energy is expressed in MWh and refers to the compound that allows to operate the electrical utilities. Capacity is expressed in MW and refers to the electrical power (i.e., the extent to which the generator may deliver a quantity of energy in a given period of time). These products are exchanged in different trading floors (power exchange), or Over-The-Counter (OTC), at different time horizons before real-time delivery. Figure 2.4 illustrates the way the markets are arranged in the liberalized framework.

Day-ahead market is cleared once a day at 12 a.m. and sets electricity exchanges for the 24 hours of the next day. This market is solved by the EPEX SPOT centralized platform, which performs a "pay-as-cleared"[6] auction defining a clearing price and a clearing quantity, by matching the merit-order demand and offer curves. The merit-order curves are obtained by sorting the offer (demand) bids in ascending (descending) order of price. The day-ahead market is the last market to be cleared before the nominations (i.e., the operational schedule of injections and off-takes) should be submitted by the BRPs to the TSO. It should be noted

[6]The pay-as-cleared auction refers to a market place where a single price is determined for all the exchanges, unlike the "pay-as-bid" auction which considers different prices for each agent participating in the market, corresponding to their own bid.

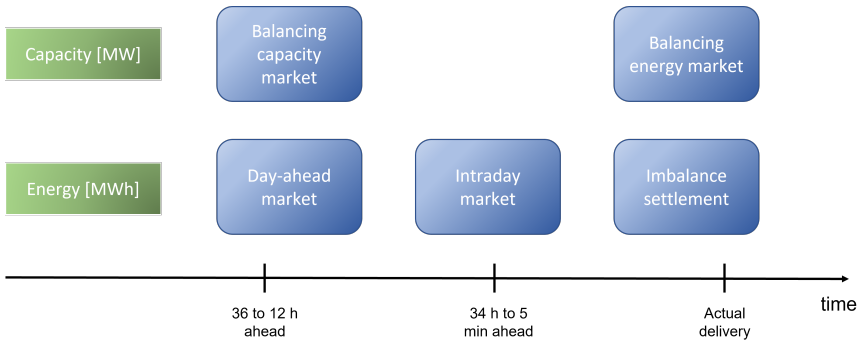


Figure 2.4: The short-term electricity trading floors.

that it is up to the TSO to undergo the necessary verification and to reject a nomination if it is expected to violate transmission constraints, since day-ahead electricity markets disregard these operating restrictions in EU.

The *intraday market* opens at 4 p.m. the day before delivery (after the nomination process) and ends up to 5 minutes before delivery [77]. This market allows the BRPs to buy or sell energy that is missing or in excess to respect the day-ahead schedule, for instance, due to an outage or a deviation in generation profile (e.g., deviation from forecast for renewable energies). This type of market is designed as a continuous session where buyers and sellers are paired as soon as their (demand or offer) bids correspond.

The remaining imbalances occurring in real-time (after the BRPs have participated in the intraday market) result in an Area Control Error (ACE) that must be resolved by the TSO. Since he may not possess generating capacity, the TSO books balancing capacity via the *balancing capacity market*. These markets are auctioned by the TSO itself, which requires a minimum amount of balancing capacity to be booked (in order to safely ensure the operation of the system). Beginning from 2020 in Belgium, these markets are cleared on a day-to-day basis before the day-ahead energy market. Different balancing capacity products are available, and required by the TSO, which range based on their reaction time.

The Frequency Containment Reserve (FCR) must react in less than 5 minutes. This type of reserve allows the TSO to maintain the frequency in a certain range around the constant frequency of 50 Hz. The activation process in real-time is automatic and decentralized in the interconnected continental European power system. The automatic Frequency Restoration Reserve (aFRR) corresponds to a balancing capacity that must react in maximum 7.5 minutes in order to restore the frequency to the value of 50 Hz. The manual Frequency Restoration Reserve (mFRR) must be activated by the TSO in case the default persists (i.e., the

frequency can not reach back the value of 50 Hz). This type of reserves must attain its maximum output power in 15 minutes.

After the activation of balancing reserves to resolve the ACE, a residual system imbalance may still occur. If the system imbalance lasts too long, the remaining recourse actions for the TSO comprises load shedding (i.e., highly expensive disconnection of some part of the grid) and renewable energy spillage. The *imbalance settlement* is the real-time market for energy, in which the BRPs are penalized for their individual imbalance that has triggered the overall system imbalance. With the aim at further harmonizing the European electricity markets, the European commission has delivered guideline on electricity balancing [78] with the aim at establishing a single-price imbalance settlement. This settlement fosters the BRPs to reduce their own imbalance.

2.2.2 Basic Operation of Natural Gas Systems

Natural gas is a commodity that is mainly exchanged throughout the natural gas grid[7]. Similarly to the power system, the natural gas system is composed of a high pressure transmission system (meshed network that transports the natural gas over long distances) and a low pressure distribution system that serves the end-users. A schematic natural gas system is represented in Figure 2.5.

The transmission system operator must ensure the reliable operation of the transmission system, mainly composed of gas pipelines and compressors (that create a difference of pressure along a pipeline). Similarly to the power system, due to the high investment cost in the natural gas facilities, transmission system operators represent natural monopolies, which are subject to regulation. Natural gas TSOs are also facilitators for the physical exchanges between the gas suppliers and the consumers. On the one hand, gas suppliers buy gas from domestic or foreign producers and are willing to inject it at one or multiple entry points. On the other hand, the consumers are willing to withdraw gas for different utilities in retail or industry (e.g., heating, cooking, industrial processes, ...).

The inherent characteristics of natural gas results in multiple differences in the natural gas system operation, compared to the power system operation. First, the gas is flowing at a velocity, which is much lower than the velocity of electricity (i.e., the speed of light) and has a compressible nature which enables a cost-efficient storage. Therefore, the instantaneous balance between supply and consumption is not crucial, but the injections and off-takes must rather be

[7]Note that the share of natural gas exchanged in the liquid state (i.e., liquefied natural gas) is growing, but the transmission system remains the main carrier for natural gas, comprising of a major part of the international trades.

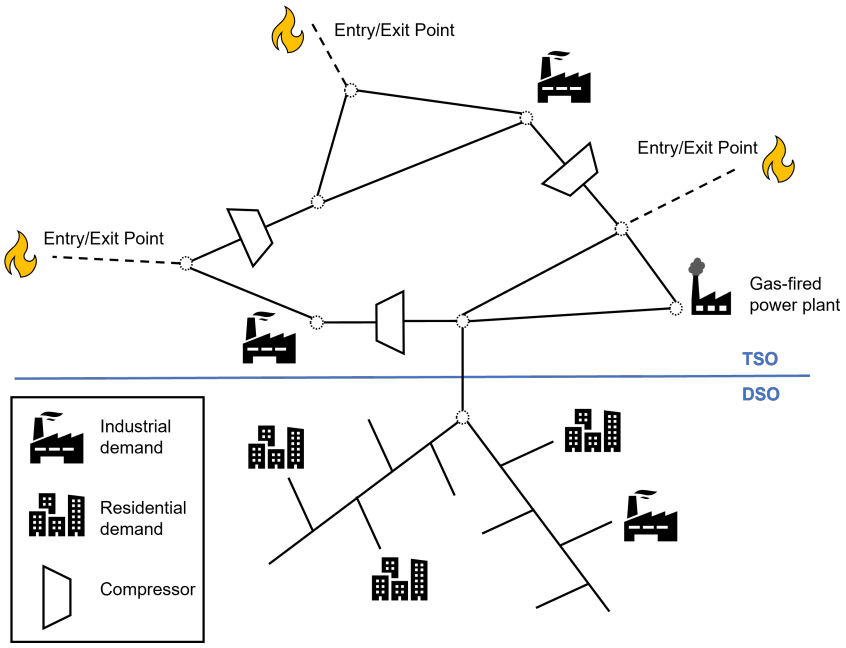


Figure 2.5: Basic operation of the natural gas system. TSO: Transmission System Operator, DSO: Distribution System Operator.

balanced in the long run (e.g., over the day). The natural gas TSO is responsible for the overall balance of the system. Analogously to the electricity system, the balancing responsibility is transferred as much as possible on third parties, the so-called *shippers*. Shippers have contracts with gas producers and consumers, and also with the system operators, in view of transporting the natural gas via the network. Second, natural gas, unlike electricity can easily be stored either in gas tanks or within the pipelines. The inherent storage capability within the gas pipelines, is referred to as *linepack*. It allows, to have different inlet and outlet flows at the interfaces of each pipeline. This results in an additional flexibility that can be provided by the difference between injections and off-takes in the short-run.

Short-Term Natural Gas Markets

The natural gas has been traditionally traded via long-term contracts. However, recently, the liquidity of short-term natural gas markets has increased.

In Europe, the day-ahead natural gas market is cleared without considering the gas network limitations, under the so-called entry-exit model. In this model, the shippers acquire the right to transport gas throughout the network, i.e., injections and/or withdraws at one or multiple entry/exit points. The right to

transport is booked as a firm or interruptible capacity. The firm capacity must be ensured by the TSO (under normal operating conditions), whereas the interruptible capacity can be reduced or shed by the TSO, if required for technical purposes.

The shippers who are willing to exchange natural gas via the transmission grid, must balance their portfolio on a daily-basis (e.g., in Belgium) with within-day tolerances. These tolerances aim at mitigating the aggregate imbalance in the gas network. When the aggregate imbalance reaches the balancing threshold, the TSO must take balancing actions, whose cost will be repercuted to the shippers that have caused the overall imbalance[8]. Finally, at the end of the day, the residual imbalance is settled in money via an imbalance settlement.

2.2.3 Interplay Between Power and Natural Gas Systems

The natural gas-fired power plants may play a key role in the future of energy systems, as these flexible and efficient power plants represent a potential candidate for mitigating the uncertainties stemming from the renewable-based power generation[9]. In turn, their location at the interface of both power and natural gas systems may induce the propagation of uncertainties from power to natural gas systems. In this context, the cost-efficient operation of NGFPPs (and the underlying exchanges between power and natural gas) calls for the convergence of operation between the power and natural gas systems.

Another incentive towards the coordination of their operation lies within the scope of Power-to-Gas (P2G) facilities [79]. The P2G facilities allow to produce gas from electricity surplus and water or carbon dioxide. The two most common types of P2G technologies are the electrolyzer and the methanizer, which respectively allow to produce hydrogen or methane that can directly be injected within the natural gas system. It is worth mentioning that the emphasis of our work is not on P2G technologies. The links between the power and natural gas systems are schematically summarized in Fig. 2.6.

In that context, it becomes important to coordinate the operation of power and natural gas systems, as the bidirectional interplay between both systems is expected to be tightened in the future. This becomes more and more important as the magnitude of uncertainties stemming from renewable power generation

[8]Note that these actions are not often used in the operation of natural gas system. Consequently, it is often said that the natural gas system does not need to be balanced. The flexibility provided by shippers via interruptible capacities, however, is used.

[9]It is worth mentioning that, following the recent evolution of energy markets in Europe, and depending on the evolution of gas prices, NGFPPs may not represent a cost-efficient solution for providing the flexibility required by the power network.

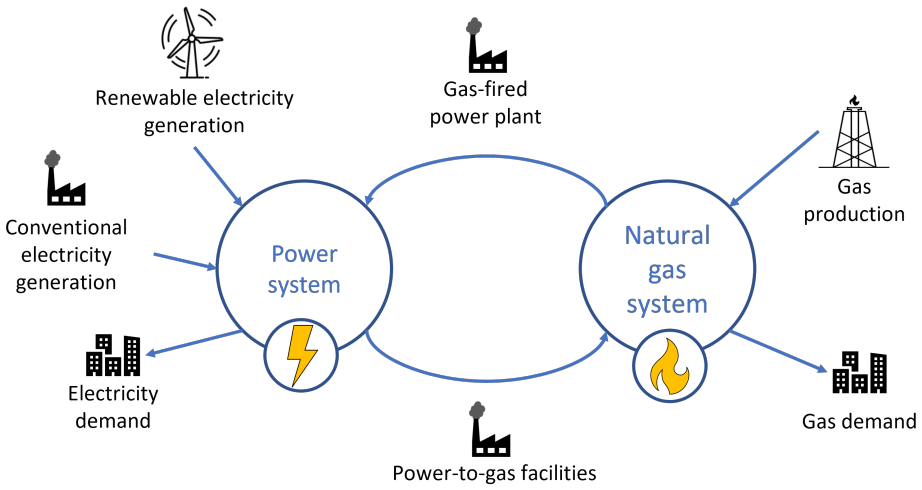


Figure 2.6: Schematic representation of the power and gas systems showing the interplay between both systems via the Natural Gas-Fired Power Plants (NGFPPs) and Power-to-Gas (P2G) facilities.

may propagate from one to another system, which calls for additional flexibility that can be found in the inherent storage capability of natural gas. Different coordination schemes are investigated in the literature from a sequential decoupled operation to a simultaneous and joint operation of both systems [80].

2.3 Impacts of the Transition on the Energy Systems

The ongoing energy transition has three main objectives:

The main goals of the energy transition:

- **Affordability**: granting the access to energy for everyone,
- **Sustainability**: ensuring the supply of energy for the next generations,
- **Reliability**: ensuring a safe and reliable supply of energy.

In the following, we identify three major evolutions pursuing these objectives, which mainly impact the current operation of power systems.

2.3.1 From Fossil-Fueled to Renewable-Based Generation

The European objectives in terms of integration of renewables in the energy generation mix, notably translates into the massive connection of renewable-based

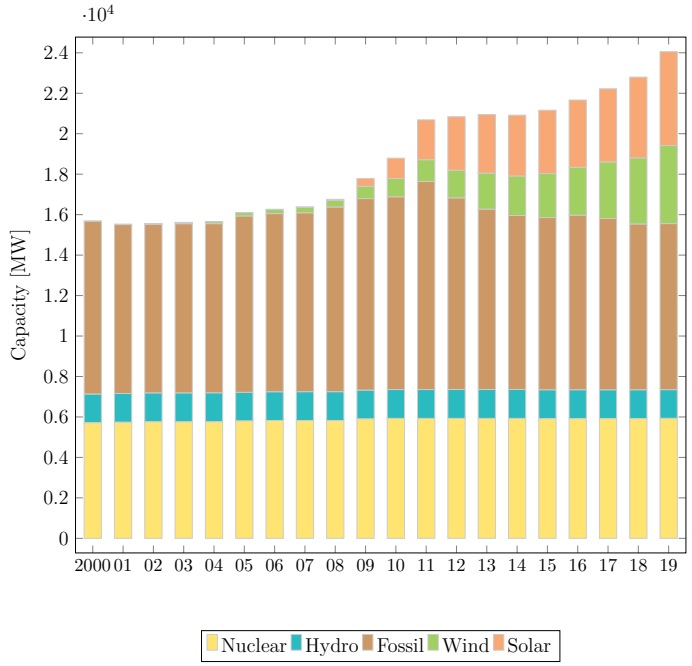


Figure 2.7: Mix of electricity generation capacity installed in Belgium.

electricity generating units, e.g., wind turbines and solar panels, into the power systems. The evolution of generation capacity in Belgium broken down by type of primary energy source is represented in Figure 2.7. As one can see from this graph, the shift towards renewable energy is already taking place with the installed wind and solar power capacity connected to the electricity grid, which is expected to continue growing in the near future.

The main characteristics of the weather-dependent renewable energy sources pertain to their *i)* non-controllability (i.e., they can not be set to a constant generation setpoint over a fixed period of time), and *ii)* limited predictability (i.e., the accuracy of scheduling these wheather-dependent resources depends on the period ahead of the delivery time). This hinders the large-scale integration of renewables into the usual operation of power systems.

Renewables play also an essential role in markets where they are involved. With the governmental subsidies, renewable energy sources may offer bids at very low marginal prices. This results in the shift of the offer curve towards the right and a potentially reduced electricity price. This imperfection of markets is one of the sources for the losses incurred by medium-cost assets (e.g., flexible gas-fired units) and the potentially reduced investments in other types of generation capacity.

In Belgium, the proliferation of renewable energy sources is taking place in parallel with the willingness to phase-out nuclear power plants. At the time of writing, negotiations are still ongoing to decide whether a reduced number of nuclear power plants may still be operated after 2025, in Belgium. Nonetheless, nuclear electricity generation represents a dense (i.e., requires a comparatively few space to deliver high amount of energy) and controllable generation mean. Besides, it increases the rotating inertia of the system, which is paramount to safely ensure a constant frequency at 50 Hz [81]. In that direction, it seems that an energy mix relying only on renewable assets is not prone to replace nuclear energy in the short-run due to its variability and uncertainty.

2.3.2 From Centralized to Distributed Generation

Renewable energy sources are also characterized by their lower investment cost comparatively to large-scale generation centers. This creates the opportunity for end-users to make their own investments in electricity generating units, e.g., solar panels on the roof of a house or wind turbine supplying an industrial park. This results in the distribution of the production means, i.e., from large-scale generation (e.g., nuclear) connected to the high voltage grid, to small generating units spread around the region and connected to the medium-to-low voltage grid.

This key transition challenges the usual operation of the power system in different manners. First, the injections of distributed renewable energy sources is often based on electronic power devices (e.g., power electronic inverters which converts the DC current of PV panels to AC current of the grid). This creates harmonics (voltage signals with smaller amplitude but with a frequency multiple of 50 Hz) which may trigger technical issues in distribution systems [82]. Second, the flow of electricity at the interface between TSO and DSO, which was previously most of the time unidirectional (from TSO to DSO), becomes more challenging to predict, as renewable power may be injected to the transmission system in period of favorable weather conditions. This may create over-voltages in the distribution grid, and new voltage control methodologies must be developed to counter this effect [83]. Overall, this calls for a proper coordination between DSO and TSO, but also coordination at local levels between all agents (consumers, prosumers, DSO, ...) which are expected to take an active role in the operation of the grid.

2.3.3 From Generation to Consumer-Centric Operation

By owning their generation assets, the consumers are expected to gradually become *prosumers* (from the contraction of “producers” and “consumers”). In addition to generation assets, the prosumer may also own, electricity storage (via on-site batteries), electric vehicles or heat pumps. Thereby, the prosumer is

capable of providing flexibility that may help the overall system in balancing the weather-dependent and intermittent generation. This is encouraged by ongoing projects which are focusing on market designs and local coordination that would incentivize a socially-compatible behaviour from prosumers. It is worth noting that the active role of prosumers will require real-time communication tools for accessing, e.g., system state or price information.

2.4 Ensuring a Reliable Operation in an Uncertain Environment

The increase in renewable energy resources, drives a growing dependency on weather conditions, therefore increasing the operational risk in power system (by being exposed to more extent to variability and uncertainty). Though, the system operators must ensure safety and security of supply. In what follows, we define the different sources and features of uncertainty in Section 2.4.1 and the main action levers for ensuring reliability within the short-term operational procedures in Section 2.4.2.

2.4.1 Thriving in Uncertainty

Energy systems have always been exposed to operational risk since the beginning of their creation [84]. *Risk* usually defines as the exposure to *uncertainty*. Different sources and features characterize the uncertainty affecting the system operations, which are introduced in the following.

Sources of Uncertainty

Energy system operation has always been influenced by the **weather conditions**. An example of such dependency between the weather and the power system operation stems from the Texas 2021 winter storm and the subsequent energy catastrophy [85]. During this event, temperatures dropped to -16°C, disrupting the operation of power system in a unprecedented manner which resulted in the disconnection of 4.4 million inhabitants. An non-exhaustive list of weather-related impacts on the main components involved in this event is as follows. Natural gas being frozen in pipelines hindered the gas supply of gas-fired power plants, frozen coal piles resulted in the interruption of several coal plants, wind turbines and one nuclear power plant also stopped operating.

Nowadays, an increasing share of renewable energy sources are connected to the electricity grid. As previously introduced in Section 2.3.1, renewable energy is variable (i.e., not controllable) and uncertain (i.e., not accurately predictable),

ensuing a growing uncertainty in the supply side, driven by the weather conditions.

The **electricity load** is also uncertain. In particular, the Belgian transmission system operator ELIA has established different adequacy scenarios [86] characterizing the future of the Belgian electricity demand. Beyond the long-term and medium-term uncertainty (e.g., stemming from different load patterns depending on the season of the year), the demand is also uncertain from a short-term perspective. In particular, the electrification of the economy (with e.g., the massive incentives towards electric vehicles and heat pumps) is ongoing at an uncertain velocity and towards an uncertain global level.

Uncertainties may also stem from the **operation** itself. For instance, outages, current faults, transmission line defaults or transformer explosions, are uncertain events that affect the supply of electricity and are limitedly predictable.

Finally, in the liberalized **market**-based framework, agents are incentivized by price signals. These prices have a high impact on the economic viability of the asset owner (recent discussions set the debate on whether the real-time price are suitable signals for long-term investments [87]). In this perspective, prices also represent a source of uncertainty for these stakeholders, as well as for system operators.

Features of Uncertainty

The uncertainty pertaining to either, economic signal, system operation, renewable-based generation, electrical load or weather, may come in different forms, which requires different (re-)actions by system operators [78].

A first type of random event affecting the power system is usually classified as **contingencies**. These random events are defined as discrete disturbances such as generation or transmission line outages. A famous real-life example of an impactful contingency resides in the 2003 outage of a transmission line between Italy and France. The cascading outage effect resulted in a large blackout affecting about 56 million people [88]. Usually, system operators may have recourse to contingency reserves (such as replacement reserves) that are manually activated, to respond to such random events.

Besides, the second type of uncertainties affecting the energy system operation are **fluctuations**, e.g., continuous disturbances in supply, demand, or price signals. The recourse actions available for system operators ranges among the use of spinning operating reserves (such as primary and secondary reserves) or the recourse to load shedding (i.e., the disconnection of electrical demands when

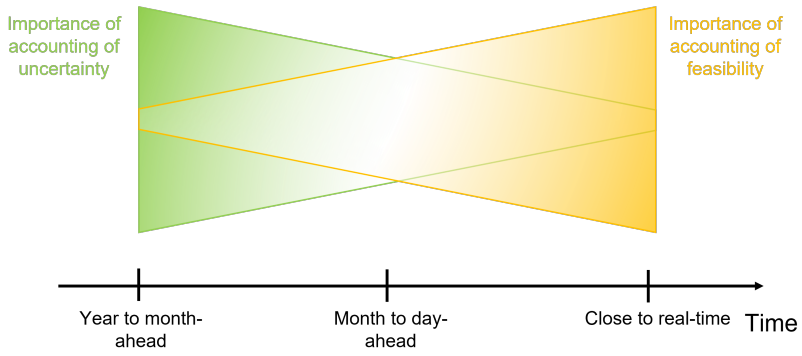


Figure 2.8: Evolution of importance of feasibility and uncertainty within decision-making problems.

demand exceeds supply). With the increase in renewable energy sources connected to the electrical grids, the intensity of such fluctuations in the supply side is also increasing. This, in turn, calls for an increasing need in flexible assets. The event happening in Texas (28^{th} February 2008) is an example of such large fluctuations in wind power generation [89]. During this event, the wind power generation unexpectedly dropped from 2000 MW to 360 MW in almost 3.5 hours. The lessons learned from this event, call for an improvement in generation and load forecast and scheduling.

2.4.2 Short-Term Operational Reliability

To mitigate the operational risks stemming from unpredicted events, system operators schedule the operation of the energy system at different time frames ahead of actual delivery to ensure safety and security of supply. To do so, TSOs have the ability to run algorithms that simulate the system state under different conditions. Depending on the time of scheduling, the magnitude of uncertainty affecting the operating scheduling decisions varies (for instance, long-term demand is not straightforward to predict, whereas day-ahead renewable energy generation may be more accurately predicted). Accounting for those probabilistic forecasts within the operational procedures is taking importance, as the magnitude of uncertainty gradually increases. However, another important feature of such operating scheduling is to ensure feasibility in the system operation (for instance, ensuring that the system state complies with the transmission line thermal limits). This requires intensive computation to finely simulate the overall system. Due to restrictions in computational times (e.g., especially close to real-time), a trade-off between those two antagonist objectives is often required. This trade-off and the evolution between the importance of accounting to either uncertainty or feasibility is illustrated in Figure 2.8.

With the increasing computing power and recent developments of network model representation, there is the potential to improve the short-term operational procedures by embedding a more accurate network model into the scheduling process. In other words, modeling the network constraints in the short-term operational scheduling is key enabler for a cost-efficient system operation.

2.5 Chapter Summary and Conclusions

In this chapter, we introduce key information regarding the ongoing energy transition. The climate change, and the subsequent intergovernmental climate agreements call for the massive integration of renewable energy sources within the energy generation mix. In turn, the traditional operation of both the power and natural gas systems is challenged by the evolutions in the energy framework. In particular, the uncertainty and variability induced by renewable energy sources as well as the decentralization of electricity generation capabilities increase the operational risk of energy systems. In that context, to ensure a cost-efficient yet reliable supply of energy, it becomes crucial for system operators to appropriately accommodate the uncertainty and improve the modeling of network limitations within operational procedures.

In what follows, we introduce DRO in Chapter 3, and propose enhanced distributionally robust methodologies in Chapters 4 and 5, as potential frameworks for improving the modeling of uncertainty within decision-making problems in energy systems. In Chapters 6 and 7, we consider the coordination of power and gas systems and study the existing convexification techniques for modeling gas flow dynamics. Then, we propose a new neural-network-constrained framework that is capable of accurately modeling the non-convex gas flow equations within the coordinated power and gas dispatch.

CHAPTER 3

Distributionally Robust Optimal Power Flow

With the increasing share of renewables in the electricity generation mix, the magnitude of uncertainty in the electrical power in-feeds is also increasing. In the recent years, research efforts have turned towards including those uncertainties within the short-term operational decision-making problems, such as the Optimal Power Flow (OPF) problem. To do so, the traditional uncertainty-aware optimization techniques assume the availability of a unique and representative distribution function. However, such distribution is not straightforwardly accessible in practice, which may lead to suboptimal solutions based on a naive representation of uncertainty. To raise this issue, Distributionally Robust Optimization (DRO) has emerged as a new modeling approach, which considers a family of potential distributions characterizing the uncertainty, the so-called, *ambiguity set*. The distributionally robust solution is then selected as the one that is optimal in expectation for the worst-case distribution within the ambiguity set. This setting is known to provide interesting out-of-sample performances compared to traditional approaches, but requires tractable reformulations.

In the following Section 3.1, we introduce the mathematical definitions regarding the distributionally robust optimization, with a special focus on the Wasserstein metric-based distributionally robust approach. In Section 3.2, we expose the distributionally robust optimal power flow problem and its state-of-the-art reformulation. Finally, in Section 3.3, we assess the performance of the distributionally robust paradigm on a small case study, and provide a discussion regarding the main limitations of the state-of-the-art approach.

3.1 Hands on Distributionally Robust Optimization

In this section, we introduce the concept of distributionally robust optimization in Section 3.1.1. Next, the main types of ambiguity sets which have been introduced in the power systems research literature, are defined in Section 3.1.2. Finally, Section 3.1.3 introduces one potential methodology to cope with two-stage decision-making problems under uncertainty.

3.1.1 Conceptual Statement

We first consider a generic optimization program, that will help us to conceptually introduce the distributionally robust optimization, such as

$$\min_{x \in \mathcal{X}} \ \mathcal{R}^{\mathrm{O}} \left(f \left(x, \tilde{\xi} \right) \right) \tag{3.1a}$$

$$\text{s.t. } \mathcal{R}^{\mathrm{C}} \left(g \left(x, \tilde{\xi} \right) \right) \leq 0. \tag{3.1b}$$

In (3.1), the decision-maker is willing to minimize the objective function (3.1a), by selecting appropriate values for decision variables $x \in \mathcal{X}$, while ensuring the feasibility constraints (3.1b). However, the cost function $f \left(x, \tilde{\xi} \right) : \mathcal{X} \times \Xi \rightarrow \mathbb{R}$ and the constraints[1] $g \left(x, \tilde{\xi} \right) : \mathcal{X} \times \Xi \rightarrow \mathbb{R}^M$ are affected by the uncertain parameter $\tilde{\xi} \in \Xi$. We denote $\mathbb{Q}$ the probability distribution, which relates the potential realizations of uncertain parameter $\tilde{\xi}$ to their occurrence probability. To correctly handle those uncertain parameters, the decision-maker has recourse to discrimination operators $\mathcal{R}^{\mathrm{O}}(.)$ and $\mathcal{R}^{\mathrm{C}}(.)$. The rationale behind $\mathcal{R}^{\mathrm{O}}(.)$ is to find the *best* random variable $f \left(x, \tilde{\xi} \right)$, from the decision-maker viewpoint. Besides, the operator $\mathcal{R}^{\mathrm{C}}(.)$ quantifies the risk of violating the feasibility constraints, given the probability distribution $\mathbb{Q}$. The setting (3.1) represents a broad range of stochastic optimization approaches. For instance, by selecting the operator $\mathcal{R}^{\mathrm{O}}(.)$ as the expected value $\mathbb{E}^{\mathbb{Q}}[.]$ over distribution $\mathbb{Q}$, problem (3.1) casts into the traditional stochastic programming approach[2]. Besides, by selecting $\mathcal{R}^{\mathrm{C}}(.)$ as the Value-at-Risk (VaR) [90], problem (3.1) amounts to a chance-constrained

[1] The vector constraint function contains M elements, i.e., $g \left(x, \tilde{\xi} \right) = \left[g_1 \left(x, \tilde{\xi} \right), ..., g_M \left(x, \tilde{\xi} \right) \right]$.

[2] Differently, robust optimization selects the worst-case realization within an uncertainty set, i.e., $\mathcal{R}^{\mathrm{O}}(.) = \max_{\tilde{\xi} \in \Xi} (.)$.

program[3]. This results in

$$\min_{x \in \mathcal{X}} \ \mathbb{E}^{\mathbb{Q}} \left[f\left(x, \tilde{\xi}\right) \right] \tag{3.2a}$$

$$\text{s.t. } \mathbb{Q}\left(g\left(x, \tilde{\xi}\right) \leq 0 \right) \geq 1 - \epsilon. \tag{3.2b}$$

Interestingly, the formulation (3.2) amounts to the traditional chance-constrained stochastic program [18]. The objective function (3.2a) is to minimize the pay-off in expectation over the probability distribution $\mathbb{Q}$ of random variable $\tilde{\xi}$. The chance constraint (3.2b) ensures feasibility with high probability level (i.e., the violation probability ϵ is usually selected as a small number between 0 and 1). Note that when the violation probability $\epsilon = 0$, the program becomes a "robustly-constrained" stochastic program, i.e., the operating constraints are enforced for all potential realizations of uncertainty. Depending on whether the probability operator collects all the constraints or is applied to each individual constraint, the chance-constraint (3.2b) represents, either a *joint* or *individual* chance constraint, respectively. This mathematically amounts to the following expressions

$$\text{Joint} : \mathbb{Q}\left(g_m\left(x, \tilde{\xi}\right) \leq 0 \ \ \forall m \in \{1, ..., M\} \right) \geq 1 - \epsilon, \tag{3.3a}$$

$$\text{Individual} : \mathbb{Q}\left(g_m\left(x, \tilde{\xi}\right) \leq 0 \right) \geq 1 - \epsilon_m \ \ \forall m \in \{1, ..., M\}\,. \tag{3.3b}$$

Individual chance constraint (3.3b) allows the decision maker to adjust his risk attitude for each constraint separately, whereas joint chance constraint (3.3a) ensures the safe operation of the overall system. In the following, we discuss the conceptual motivation for utilizing chance constraints in energy systems.

Conceptual motivation for chance-constrained programming

Chance constraints introduce a framework that suits well the definition of reliability criterion in energy systems. However, this technique usually disregards the modeling of recourse actions [91], which are typically expensive and may impact the overall operating cost of the underlying optimal decisions. In general, the decision-maker is willing to minimize the *total* operating cost (comprising the day-ahead scheduling cost and recourse action costs). Theoretically, it is expected that these costs follow the tendency shown in Figure 3.1.

[3]The equivalence between the VaR and a chance constraint is shown in [90].

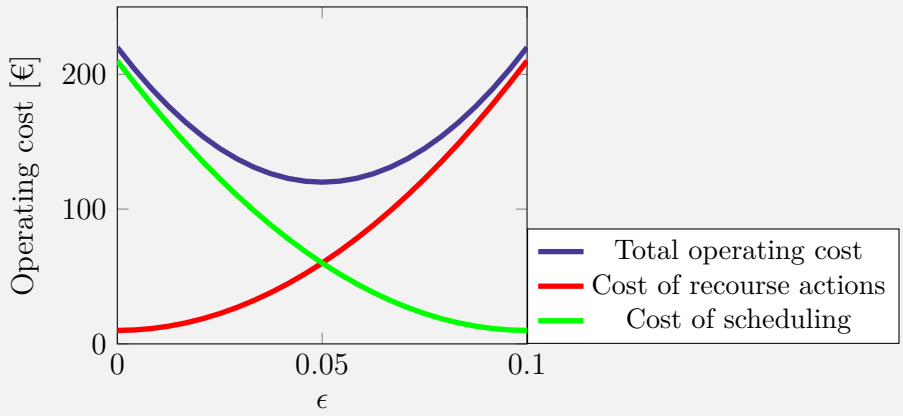


Figure 3.1: Schematic evolution of operating cost for different values of violation probability ϵ.

The cost of scheduling the system in day-ahead decreases with the allowance to violate the operating constraints. Indeed, by allowing more constraint violations, the set of feasible (i.e., acceptable) solutions grows, from a mathematical programming perspective. In this direction, the program may explore solutions that achieve lower expected scheduling cost. In contrast, by taking riskier operational positions, the operational constraints are expected to be more often violated and recourse actions (such as, e.g., load shedding or activation of expensive operating balancing capacity) will be necessary. In other words, the cost of recourse actions increases with the operational constraint violations, i.e., ϵ. As conceptually represented in Figure 3.1, summing these two contributions may result in an optimal trade-off between the cost of scheduling and the cost of recourse actions. This optimum corresponds to an optimal value for the violation probability ϵ^*. The decision-maker may therefore select the most appropriate value between 0 and ϵ^*, owing to its own risk appetite. On the contrary, beyond ϵ^*, the total operating cost increases with the violation of operating constraints, which is a situation that should be avoided. It is worth mentioning that the value ϵ^* is hard to find *a priori*, in practice.

The setting (3.2) assumes that the input distribution is unique and perfectly representative of the uncertainty. However, such distribution is not straightforwardly accessible in practice, which may lead to suboptimal solutions based on a naive representation of uncertainty. To raise this issue, DRO enables the decision-maker to feed the decision-making problem with a family of potential distributions, i.e., the so-called ambiguity set. The optimal solution is derived with respect to the expectation over the worst-case distribution within the set. By doing so, the

optimal decisions hedge against any misrepresentation of uncertainty embedded within the ambiguity set. In general, DRO may outperform traditional techniques such as stochastic programming and robust optimization [20]. The distributionally robust counterpart of problem (3.2), writes as

$$\min_{x \in \mathcal{X}} \max_{\mathbb{Q} \in \mathcal{A}} \ \mathbb{E}^{\mathbb{Q}} \left[f\left(x, \tilde{\xi}\right) \right] \tag{3.4a}$$

$$\text{s.t.} \min_{\mathbb{Q} \in \mathcal{A}} \ \mathbb{Q}\left(g\left(x, \tilde{\xi}\right) \leq 0 \right) \geq 1 - \epsilon. \tag{3.4b}$$

The objective function (3.4a) minimizes the cost in expectation for the worst-case distribution $\mathbb{Q}$, which is selected within the ambiguity set $\mathcal{A}$ by the inner maximization operator. The feasibility is enforced via (3.4b), that corresponds to a Distributionally Robust Chance Constraint (DRCC). This type of constraint enforces the inequalities within parenthesis with high probability for any distribution embedded within the ambiguity set $\mathcal{A}$. Note that the constraint (3.4b) can also be enforced *individually* or *jointly*, as discussed above for the traditional chance-constrained approach.

Intuitively speaking, the design and accuracy of the ambiguity set is linked to the underlying performance of optimal decisions. In the operations research literature, there exists plenty of definitions of ambiguity sets [21], two of which have been mainly introduced in the power system literature (i.e., the moment-based and the metric-based ones). The definitions of these types of ambiguity sets are further introduced in Section 3.1.2.

3.1.2 Selection of an Appropriate Ambiguity Set

Accounting for risk into decision-making processes often requires having recourse to a distribution function that describes the uncertainty. When the decision-maker is less confident about the accuracy of distribution functions, he is exposed to ambiguity. Ambiguity refers to distributional uncertainty, i.e., uncertainty in the distribution function itself. To incorporate this second level of uncertainty into decision-making problems, the decision-maker can collect several distribution functions within an ambiguity set. There exist a variety of ambiguity sets defined in the related literature, e.g., based on φ-divergences [92] (a special case of φ-divergence is the Kullback-Leibler divergence [93]), based on moments, see [19] and [94], or on probability metrics [95]. We refer the reader to [21] for an in-depth review of existing approaches, while we focus on the definition of moment-based and metric-based ambiguity sets in the following.

Moment-Based Ambiguity Set

Under the moment-based paradigm [96], the ambiguity set is composed of the family of distributions that share the same moments (e.g., mean and covariance matrix). Those moments may be calculated from empirical data and are usually imposed as exact values. Therefore, the ambiguity set $\mathcal{A}$ is composed of distributions with given moments but different shapes, such that

$$\mathcal{A} = \left\{ \mathbb{Q} \in \mathcal{M}(\Xi) \;\middle|\; \begin{array}{l} \mathbb{E}^{\mathbb{Q}}\left[\tilde{\xi}\right] = \mu \\ \mathbb{E}^{\mathbb{Q}}\left[(\tilde{\xi}-\mu)(\tilde{\xi}-\mu)^t\right] = \Sigma \end{array} \right\}. \tag{3.5}$$

In (3.5), the set $\mathcal{M}(\Xi)$ represents the collection of all distributions on support Ξ. The mean $\mu \in \mathbb{R}^{|\mathcal{K}|}$ and the covariance matrix $\Sigma \in \mathbb{R}^{|\mathcal{K}|\times|\mathcal{K}|}$ are assigned by the decision-maker, based on, e.g., empirically observed data. Therefore, the constraints in the definition of the ambiguity set $\mathcal{A}$ impose that any distribution $\mathbb{Q} \in \mathcal{A}$ has a mean equal to μ and a covariance matrix equal to Σ. In this formulation, it is assumed that the moment estimations are exact, i.e., imposed as fixed values. It is worth mentioning that moment-based ambiguity sets considering variations around the moment estimations also exist, see, e.g., [97].

Wasserstein Metric-Based Ambiguity Set

Recent mathematical advances [20] have allowed to take advantage of probability metrics, such as the Wasserstein[4] distance [98], in order to model the ambiguity set. The rationale behind metric-based ambiguity sets is to collect the distributions that are *close* to a central reference distribution, in the sense of a probability metric that calculates the distance between two distribution functions. In general, the reference distribution corresponds to the empirically observed one (e.g., composed of N equiprobable observed scenarios) or a probabilistic forecast derived from machine learning-based predictions. In both cases, we denote the reference distribution by $\widehat{\mathbb{Q}}_N$ hereafter. Mathematically speaking, the Wasserstein metric-based ambiguity set reads as

$$\mathcal{A} = \left\{ \mathbb{Q} \in \mathcal{M}(\Xi) \;\middle|\; d_{\mathrm{W}}(\mathbb{Q}, \widehat{\mathbb{Q}}_N) \leq \theta \right\}. \tag{3.6}$$

In (3.6), the operator $d_{\mathrm{W}}(.,.) : \mathcal{M}(\Xi) \times \mathcal{M}(\Xi) \to \mathbb{R}^+$ represents the Wasserstein distance between the reference distribution $\widehat{\mathbb{Q}}_N$ and the distribution $\mathbb{Q}$ inside $\mathcal{M}(\Xi)$.

Definition 3.1. (Wasserstein metric [98]) The Wasserstein metric $d_{\mathrm{W}}(.,.)$:

[4]Note that the spelling "Wasserstein" is currently being used by the research community for referring to Leonid N. Vaserstein [98].

$\mathcal{M}(\Xi) \times \mathcal{M}(\Xi) \to \mathbb{R}^+$ is defined via

$$d_W(\mathbb{Q}, \widehat{\mathbb{Q}}_N) = \left\{ \begin{array}{l} \min\limits_{\Pi} \int_{\Xi^2} ||\tilde{\xi} - \hat{\xi}_N|| \Pi(d\tilde{\xi}, d\hat{\xi}_N) \\ \text{s.t.} \quad \begin{array}{l} \Pi \text{ is a joint distribution of } \tilde{\xi} \text{ and } \hat{\xi}_N \\ \text{with marginals } \mathbb{Q} \text{ and } \hat{\mathbb{Q}}_N, \text{ respectively} \end{array} \end{array} \right\}. \tag{3.7}$$

This distance can be seen as the optimal cost of transporting the probability mass from $\widehat{\mathbb{Q}}_N$ to $\mathbb{Q}$. The radius θ therefore represents the maximum available budget for the transportation and binds the distance to the reference distribution, to be lower or equal to $\theta \in \mathbb{R}^+$. In other words, the ambiguity set $\mathcal{A}$ collects the distributions in the neighbourhood[5] of $\widehat{\mathbb{Q}}_N$. Note that when radius $\theta = 0$, ambiguity set reduces to a singleton which represents the reference distribution. Therefore, in this case, the distributionally robust model provides similar results than those provided by the traditional stochastic programming approach.

Discussion

The Wasserstein metric-based ambiguity set is the focus of the subsequent research work for different reasons. First, the aim of the decision-maker is to embed as much information as possible within the model, in order to make informed decisions. However, the moment-based ambiguity set relies on moments only. Therefore any additional sample would only allow to refine the moment estimates but does not provide any additional information to the model. Second, the metric-based ambiguity set offers appealing properties for decision-making under uncertainty. Notably, it provides probabilistic guarantees on the out-of-sample results as well as asymptotic consistency (meaning that, when the number of input samples N tends to ∞, the optimal solution tends towards the true stochastic one, given that the radius of the Wasserstein ball is adapted), which allow to reduce the decision maker disappointment, i.e., the difference between the expected and actual results [20]. Finally, recent developments in DRO have enabled to derive tractable reformulations which provide computational performances comparable to the traditional stochastic programming approach. We discuss these reformulations in Section 3.2.

[5] In a number of papers, the metric-based ambiguity set is referred to as the Wasserstein *ball*, due to the resemblance between definition (3.6) and the definition of a ball in the Euclidean space. In the space of distribution functions, the center of the ball is analogously the reference distribution $\widehat{\mathbb{Q}}_N$ and the radius is θ.

3.1.3 Two-stage distributionally robust optimization

In the case where the decision-maker can adapt its decisions after the uncertainty has been revealed, the optimization problem becomes two-stage, or "adaptive". The first stage corresponds to *here-and-now* decisions, while the second stage refers to *wait-and-see* decisions. This setting corresponds to a broad range of power system decision-making problems under uncertainty, where the here-and-now decisions are scheduling decisions and the wait-and-see decisions are recourse actions that can be taken after the uncertainty is revealed [99]. The corresponding cost function is written

$$f\left(x, \widetilde{\xi}\right) = \min_{y \in \mathcal{Y}} h\left(x, \widetilde{\xi}, y\right) \tag{3.8a}$$

$$\text{s.t. } q\left(x, \widetilde{\xi}, y\right) \leq d, \tag{3.8b}$$

where $y \in \mathcal{Y}$ are the recourse actions, made after the uncertainty $\widetilde{\xi}$ is revealed. Therefore, one can see the recourse action as a function of uncertain parameter, i.e., $y\left(\widetilde{\xi}\right)$. The functions $h\left(x, \widetilde{\xi}, y\right)$ and $q\left(x, \widetilde{\xi}, y\right)$ respectively, represent the second-stage cost function and the operating constraints. The parameter d is the right-hand-side of equations defining the operating conditions in the second stage. Embedding the objective function (3.8) within the distributionally robust paradigm (3.4) results in a min-max-min structure which is not straightforward to solve in practice [100]. In general, these problems are known to be computationally hard to solve.

One way to address this issue is to use linear decision rules [101], which approximate the modeling of recourse actions, with affine functions of the uncertainty, i.e., $y = Y\widetilde{\xi}$. The matrix $Y \in \mathbb{R}^{|\mathcal{Y}| \times |\mathcal{K}|}$ contains decision variables that reduce to the first-stage decision. This setting has direct applications in power systems, and is further used in this thesis, as explained in Section 3.2.

It is worth mentioning that the use of decision rules is common in multi-stage stochastic optimization. Research efforts have been realized in the direction of improving the approximation made by linear decision linear rules. Interestingly, generalized decision rules [102] represent a straightforward way for improving the accuracy of the modeling of recourse actions but at the cost of increased complexity. This extension is therefore not considered in this thesis but is left for future research.

3.2 The Distributionally Robust OPF Problem

We now introduce the Distributionally Robust Optimal Power Flow (DR-OPF) problem, which is the focus of this thesis in Chapters 4 and 5, where we consider both energy and reserve markets[6]. The Optimal Power Flow (OPF) problem is a fundamental tool for system operators, which is typically used within the unit commitment [103], market clearing [104], or security assessment [105] problems. OPF usually aims at cost-effectively dispatching the available generating units[7], while ensuring the balance between generation and consumption and the operational constraints (such as transmission line capacity limits, and generator operating bounds). However, with the recent increase in weather-dependent renewable energy sources connected to the power system, the OPF problems become increasingly exposed to supply-side uncertainty.

To cope with such uncertainty, the power system operators *reserve*[8] in practice a fraction of the capacity of *flexible resources*, such as hydro and fast-start gas-fired generators, in advance, e.g., in day ahead. This reserved capacity, if needed, will be exploited later in the real-time operation when the uncertainty is realized.

In the current real-world electricity markets, two distinct paradigms for energy and reserve dispatch exist [106, 107]. The first one, which is consistent with the current regulatory policy of European electricity markets, is to design two separate day-ahead markets for energy and reserve [108]. In this paradigm, the reserve market is cleared independently, before or after the energy market [8, 109]. The second approach is to design a single day-ahead market for energy and reserve, which is aligned with the current regulatory policy of the U.S. electricity markets [110, 111]. This market determines the energy and reserve dispatch jointly in a co-optimization manner.

The common practice in both aforementioned approaches is to define a *minimum reserve requirement* for the whole system or for each area, while using a *deterministic* dispatch model with a single-point forecast of renewable power generation. However, the growing penetration of renewables challenges this practice, requiring a proper incorporation of uncertainty into the energy and reserve

[6] Chapters 6 and 7 consider a different application, i.e., the coordinated power and gas dispatch problem, which can be seen as an extension of the OPF problem to integrated power and gas systems.

[7] There also exist formulations of the OPF problem that aim at minimizing the operational losses or the recourse action cost in a security assessment analysis.

[8] In the following, we refer indifferently to *reserve* and *balancing capacity*, for naming automatic and manual Frequency Restoration Reserves (resp., aFRR and mFRR), that are traded in day-ahead.

dispatch problem [112,113]. Following the U.S. practice, this thesis considers a co-optimization approach for energy and reserve dispatch, improved by a *probabilistic* characterization of the uncertainty.

The probabilistic energy and reserve dispatch problem has been extensively addressed in the literature, where the uncertainty is modeled using scenario-based stochastic programming [108, 114], robust optimization [115, 116], or chance-constrained programming [91, 117]. As discussed in Section 3.1, these techniques all have their limitations and drawbacks, pertaining to their naive reliance on an unique probability distribution. Therefore, in this chapter, our objective is to develop an efficient and computationally tractable formulation of DR-OPF problem.

The uncertainty affecting the DR-OPF problem may arise from, e.g., renewable in-feed fluctuations, when the real-time realization deviates from its day-ahead forecast. We model the uncertainty arising from renewable in-feeds as follows

$$\omega = \mu + \tilde{\xi}, \tag{3.9}$$

where $\omega \in \mathbb{R}^{|\mathcal{K}|}$ describes the real-time actual renewable power delivery, $\mu \in \mathbb{R}^{|\mathcal{K}|}$ defines the day-ahead renewable power generation forecast and $\tilde{\xi} \in \mathbb{R}^{|\mathcal{K}|}$ is the forecast error. In the following, we consider $\tilde{\xi}$ as the uncertain parameter affecting the OPF problem, and model it as a random variable, described by the probability distribution $\mathbb{Q}$. In this setting, the Wasserstein distributionally robust optimal power flow problem writes as

$$\min_{p,\overline{r},\underline{r},Y} \ \sum_{e\in\mathcal{E}} C_e p_e + \overline{C}_e \overline{r}_e + \underline{C}_e \underline{r}_e + \max_{\mathbb{Q}\in\mathcal{A}} \ \mathbb{E}^{\mathbb{Q}}\left[\sum_{e\in\mathcal{E}} C_e Y_{e,.}\tilde{\xi}\right] \tag{3.10a}$$

s.t.

$$p_e + \overline{r}_e \le \overline{P}_e, \ \ \forall e \in \mathcal{E} \tag{3.10b}$$

$$p_e - \underline{r}_e \ge \underline{P}_e, \ \ \forall e \in \mathcal{E} \tag{3.10c}$$

$$0 \le \underline{r}_e \le R_e^{\max}, \ \ \forall e \in \mathcal{E} \tag{3.10d}$$

$$0 \le \overline{r}_e \le R_e^{\max}, \ \ \forall e \in \mathcal{E} \tag{3.10e}$$

$$\sum_{e\in\mathcal{E}} p_e + \sum_{k\in\mathcal{K}} \mu_k - \sum_{d\in\mathcal{D}} p_d = 0, \tag{3.10f}$$

$$\sum_{e\in\mathcal{E}} Y_{e,k} + 1 = 0, \ \ \forall k \in \mathcal{K}, \tag{3.10g}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \ \mathbb{Q}\left(-\underline{r}_e \le Y_{e,.}\tilde{\xi}\right) \ge 1 - \underline{\epsilon}_e, \ \ \forall e \in \mathcal{E}, \tag{3.10h}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \ \mathbb{Q}\left(Y_{e,.}\tilde{\xi} \le \overline{r}_e\right) \ge 1 - \overline{\epsilon}_e, \ \ \forall e \in \mathcal{E}, \tag{3.10i}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \mathbb{Q}\left(\mathrm{M}_\ell^{\mathcal{E}}\left(p+Y\widetilde{\xi}\right)+\mathrm{M}_\ell^{\mathcal{K}}\left(\mu+\widetilde{\xi}\right)-\mathrm{M}_\ell^{\mathcal{D}}p_d \le F_\ell^{\max}\right)\ge 1-\epsilon_\ell,\ \ \forall \ell\in\mathcal{L}. \tag{3.10j}$$

The problem (3.10) optimizes the day-ahead dispatch $p \in \mathbb{R}^{|\mathcal{E}|}$ of conventional generating units as well as their upward and downward reserve capacity $\overline{r} \in \mathbb{R}^{|\mathcal{E}|}$ and $\underline{r} \in \mathbb{R}^{|\mathcal{E}|}$ to be booked in the day-ahead stage. Objective function (3.10a) minimizes the total operational cost of the system. The first three terms in (3.10a) are linear and contains the energy production cost C_e and the upward/-downward reserve procurement cost $\overline{C}_e$ and $\underline{C}_e$. The fourth term in (3.10a), i.e., $\max\limits_{\mathbb{Q}\in\mathcal{A}} \mathbb{E}^{\mathbb{Q}}\left[\sum\limits_{e\in\mathcal{E}} C_e Y_{e,\cdot}\widetilde{\xi}\right]$, refers to the operational cost of conventional units, incurred by the activation of their reserve capacity in the real-time operation. This cost is calculated in expectation with respect to the worst-case distribution $\mathbb{Q}$, which is endogenously selected within the ambiguity set $\mathcal{A}$. It is approximated by linear decision rules [101] using matrix $Y \in \mathbb{R}^{|\mathcal{E}|\times|\mathcal{K}|}$, whose elements are decision variables. More details about linear decision rules are provided in the following.

Constraints (3.10b) and (3.10c) ensure that the generation level of conventional units does not exceed their maximum and minimum limits, i.e., $\overline{P}_e$ and $\underline{P}_e$. Constraints (3.10d) and (3.10e) enforce the maximum amount of downward/upward reserve capacity $R_e^{\max}$ that can be provided by conventional units. Constraints (3.10f) and (3.10g) ensure the day-ahead and real-time power balance, respectively. In particular, the day-ahead constraint (3.10f) enforces the sum of total production of conventional generating units $\sum\limits_{e\in\mathcal{E}} p_e$ and total forecasted production of renewable units $\sum\limits_{k\in\mathcal{K}} \mu_k$ to be equal to total inelastic demand $\sum\limits_{d\in\mathcal{D}} p_d$. The real-time balance in (3.10g) is ensured via the elements of matrix Y, the so-called participation factors, which can be interpreted as follows. The conventional units respond to any deviation in renewable power generation $\widetilde{\xi}_k$ (with respect to the day-ahead forecast μ_k) based on their corresponding participation factors $Y_{e,k}$. Therefore, for each renewable power deviation $\widetilde{\xi}_k$, the recourse action of conventional units is $Y_{e,k}\widetilde{\xi}_k$, such that the total recourse action $\sum\limits_{e\in\mathcal{E}} Y_{e,k}\widetilde{\xi}_k$ compensates the total deviation $\widetilde{\xi}_k$. Note that (3.10g) is an equality constraint, so $\widetilde{\xi}_k$ can be dropped from both sides. In the real-time operation, the recourse action $Y_e\widetilde{\xi} = \sum\limits_{k\in\mathcal{K}} Y_{e,k}\widetilde{\xi}_k$ of conventional units should be limited to the reserve capacities procured in the day-ahead stage, i.e., $\overline{r}_e$ and $\underline{r}_e$. Similarly, the real-time flow within each transmission line $\ell \in \mathcal{L}$ should respect the capacity $F_\ell^{\max}$. The real-time flows are expressed using the power transfer distribution factor matrices $\mathrm{M}_\ell^{\mathcal{E}} \in \mathbb{R}^{|\mathcal{L}|\times|\mathcal{E}|}$, $\mathrm{M}_\ell^{\mathcal{K}} \in \mathbb{R}^{|\mathcal{L}|\times|\mathcal{K}|}$ and $\mathrm{M}_\ell^{\mathcal{D}} \in \mathbb{R}^{|\mathcal{L}|\times|\mathcal{D}|}$ for conventional generating units, renewable energy sources, and demands, respectively. These restrictions are enforced via

probabilistic constraints, namely, DRCCs (3.10h) and (3.10i) for each conventional generating unit e, and (3.10j) for each transmission line ℓ. These DRCCs state that the probabilistic constraints within parentheses should be ensured under the worst-case distribution $\mathbb{Q}$ within the ambiguity set $\mathcal{A}$ with a probability not lower than $1-\epsilon$. Note that the value of parameter $\epsilon \in \mathbb{R}$ lies between zero and one, fixed by the power system operator.

The problem (3.10) has a min-max structure with probabilistic constraints involving the ambiguity set $\mathcal{A}$. Therefore, it requires mathematical reformulations which depend on the definition of the underlying ambiguity set. In the following, we showcase the state-of-the-art reformulations regarding the Wasserstein metric-based ambiguity set [20]. In particular, Section 3.2.1 introduces the reformulation of objective function and Section 3.2.2 the one of DRCCs.

3.2.1 The Worst-Case Expectation Problem

Let us focus on the fourth term in objective function (3.10a) of the distributionally robust OPF problem, which we cast into vector formulation for the ease of notation:

$$\max_{\mathbb{Q}\in\mathcal{A}} \ \mathbb{E}^{\mathbb{Q}}\left[\sum_{e\in\mathcal{E}} C_e Y_{e,.}\widetilde{\xi}\right] = \max_{\mathbb{Q}\in\mathcal{A}} \ \mathbb{E}^{\mathbb{Q}}\left[C^\top Y\widetilde{\xi}\right] \tag{3.11}$$

Problem (3.11) refers to a worst-case expectation problem, in the sense defined in [21]. The maximization operator seeks the worst-case probability distribution of $\widetilde{\xi}$, i.e., $\mathbb{Q}\in\mathcal{A}$, with respect to the expected value of the objective function. Under mild assumptions and, given a piecewise linear objective function, authors in [20] provide a tractable conic reformulation of problem (3.11) such that:

$$\max_{\mathbb{Q}\in\mathcal{A}} \ \mathbb{E}^{\mathbb{Q}}\left[C^\top Y\widetilde{\xi}\right] = \begin{cases} \min\limits_{\lambda,\sigma_i,\gamma_i} \ \lambda\theta + \dfrac{1}{N}\sum\limits_{i=1}^{N}\sigma_i & \\ \text{s.t. } C^\top Y\hat{\xi}_i + \gamma_i^\top\left(h - Q\hat{\xi}_i\right) \le \sigma_i & \forall i\in\{1,...,N\} \\ \quad ||Q^\top\gamma_i - C^\top Y||_* \le \lambda & \forall i\in\{1,...,N\} \\ \quad \gamma_i \ge 0 & \forall i\in\{1,...,N\}, \end{cases} \tag{3.12}$$

where $\lambda\in\mathbb{R}, \sigma\in\mathbb{R}^N$ and $\gamma\in\mathbb{R}^{2N|\mathcal{K}|}$ are auxiliary variables. The reference distribution is composed of N historical observations $\widehat{\xi}_i \ \forall i\in\{1,...,N\}$. The operator $||.||_*$ computes the dual norm of a vector, i.e., $||.||_* = \max_{||v||\le 1} v^\top x$. Note that the matrix $Q\in\mathbb{R}^{2|\mathcal{K}|\times|\mathcal{K}|}$ and vector $h\in\mathbb{R}^{2|\mathcal{K}|}$ define the linear equations of the linear support of uncertain parameter $\widetilde{\xi}$.

3.2.2 Distributionally Robust Chance Constraints

The operational constraints (3.10h) to (3.10j) are enforced via *individual* DRCCs in the form of (3.4b). In order to derive a tractable model for (3.10), we investigate the case where operating restrictions are enforced as linear inequalities. Therefore, we are interested into the following constraint

$$\min_{\mathbb{Q}\in\mathcal{A}} \ \mathbb{Q}\left(a\left(x\right)^{\top}\widetilde{\xi}+b\left(x\right)\leq 0\right)\geq 1-\epsilon, \tag{3.13}$$

where $a\left(x\right)\in\mathbb{R}^{|\mathcal{K}|}$ and $b\left(x\right)\in\mathbb{R}$ are decision-dependent parameters[9]. For the sake of clarity, we omit the dependency with regards to decisions $x\in\mathcal{X}$, i.e., we use notations a and b, for $a\left(x\right)$ and $b\left(x\right)$, respectively.

The emphasis is put on *individual* distributionally robust chance constraints, in contrast with *joint* distributionally robust chance constraints, for three main reasons, explained hereafter. First, they allow the decision-maker to adjust the violation probability for each operating constraint, separately. This enables the system operator to adjust the reliability of the system differently in different regions, e.g., an extremely high reliability in dense urban area while comparatively less reliability in rural area. Second, joint chance constraints are known to be conservative and computationally intensive, see [12], and third, joint chance constraints are usually reduced to a set of individual chance constraints using a Bonferroni approximation [118] or other existing approximations.

State-of-the-Art Conditional-Value-at-Risk Approximation

We firstly observe that the DRCC formulation (3.13) is equivalent to

$$\min_{\mathbb{Q}\in\mathcal{A}} \ \mathbb{Q}\left(a^{\top}\widetilde{\xi}+b\leq 0\right)\geq 1-\epsilon, \tag{3.14}$$

In the following, we are interested into introducing the state-of-the-art CVaR approximation for DRCCs. As thoroughly discussed in [94], a DRCC can be conservatively approximated by a Conditional-Value-at-Risk constraint[10]. Mathematically speaking,

$$\max_{\mathbb{Q}\in\mathcal{A}} \ \mathbb{Q}\text{-CVaR}_{\epsilon}(a^{\top}\widetilde{\xi}-b)\leq 0 \Rightarrow \min_{\mathbb{Q}\in\mathcal{A}} \ \mathbb{Q}\left(a^{\top}\widetilde{\xi}\leq b\right)\geq 1-\epsilon. \tag{3.15}$$

[9] By appropriately defining $a\left(x\right)$ and $b\left(x\right)$, one can end up in the specific formulations in (3.10h) to (3.10j).

[10] The Value-at-Risk and Conditional-Value-at-Risk are thoroughly presented in [90].

Implication (3.15) states that the $\mathbb{Q}$-CVaR$_\epsilon$ represents a sufficient condition for ensuring a DRCC given the violation probability ϵ. CVaR represents a conservative approximation (i.e., implicitly implying a lower violation probability than the one defined by the decision maker), because it inherently accounts for the magnitude of constraint violation. Using the definition of the CVaR [90], the left-hand side of (3.15) can be cast into

$$\max_{\mathbb{Q}\in\mathcal{A}} \min_{\tau\in\mathbb{R}} \ \tau + \frac{1}{\epsilon} \ \mathbb{E}^{\mathbb{Q}}\left[\lceil a^\top \widetilde{\xi} - b - \tau \rceil^+\right] \leq 0, \tag{3.16}$$

where $\tau \in \mathbb{R}$ is an auxiliary variable and $\lceil . \rceil^+ = \max(0, .)$. After rearranging the order of the optimization operators, this formulation is equivalent to

$$\min_{\tau\in\mathbb{R}} \ \tau + \frac{1}{\epsilon} \max_{\mathbb{Q}\in\mathcal{A}} \ \mathbb{E}^{\mathbb{Q}}\left[\lceil a^\top \widetilde{\xi} - b - \tau \rceil^+\right] \leq 0. \tag{3.17}$$

In (3.17), a worst-case expectation problem appears which can reformulated following a similar procedure as the one described in Section 3.2.1. This approach results in a min-min structure, in which the minimization operators can be merged. Finally, the following set of conic constraints equivalently states the distributionally robust CVaR constraint (3.15),

$$\tau + \frac{1}{\epsilon}\left(\lambda^{\text{CVaR}}\rho + \frac{1}{N}\sum_{i=1}^{N}\sigma_i^{\text{CVaR}}\right) \leq 0, \tag{3.18a}$$

$$a^\top\hat{\xi}_i - b - \tau + \gamma_{i,1}^\top\left(h - Q\hat{\xi}_i\right) \leq \sigma_i^{\text{CVaR}}, \qquad \forall i \in \{1, ..., N\}, \tag{3.18b}$$

$$\gamma_{i,2}^\top\left(h - Q\hat{\xi}_i\right) \leq \sigma_i^{\text{CVaR}}, \qquad \forall i \in \{1, ..., N\}, \tag{3.18c}$$

$$||Q^\top\gamma_{i,1} - a||_* \leq \lambda^{\text{CVaR}}, \qquad \forall i \in \{1, ..., N\}, \tag{3.18d}$$

$$||Q^\top\gamma_{i,2}||_* \leq \lambda^{\text{CVaR}}, \qquad \forall i \in \{1, ..., N\}, \tag{3.18e}$$

$$\gamma_i \geq 0, \qquad \forall i \in \{1, ..., N\}, \tag{3.18f}$$

where $\lambda^{\text{CVaR}} \in \mathbb{R}$, $\sigma^{\text{CVaR}} \in \mathbb{R}^N$, $\gamma_{i,1} \in \mathbb{R}^{2|\mathcal{W}|}$, $\gamma_{i,2} \in \mathbb{R}^{2|\mathcal{W}|}$ and $\tau \in \mathbb{R}$ are auxiliary variables. Note that support $Q\widetilde{\xi} \leq h$ is incorporated into formulation (3.18).

3.3 A State-of-The-Art Setting

Following the state-of-the-art developments presented in the previous section, we derive the reformulation of the distributionally robust optimal power flow problem to be solved. The resulting program is conic (i.e., including conic constraints)

and its detailed mathematical model is given in (3.19).

$$\min_{\Theta} \sum_{e\in\mathcal{E}} C_e p_e + \overline{C}_e \overline{r}_e + \underline{C}_e \underline{r}_e + \lambda\theta + \frac{1}{N}\sum_{i=1}^{N}\sigma_i \tag{3.19a}$$

$$\text{s.t. (3.10b) - (3.10g)}, \tag{3.19b}$$

$$\begin{cases} C^\top Y\hat{\xi}_i + \gamma_i^\top\left(h - Q\hat{\xi}_i\right) \le \sigma_i & \forall i\in\{1,...,N\},\\ ||Q^\top\gamma_i - C^\top Y||_* \le \lambda & \forall i\in\{1,...,N\},\\ \gamma_i \ge 0 & \forall i\in\{1,...,N\}, \end{cases} \tag{3.19c}$$

$$\left.\begin{cases} \overline{\tau}_e + \frac{1}{\overline{\epsilon}_e}\left(\overline{\lambda}_e\theta + \frac{1}{N}\sum_{i=1}^{N}\overline{\sigma}_{e,i}\right) \le 0 & \\ Y_e\hat{\xi}_i - \overline{r}_e - \overline{\tau}_e + \overline{\gamma}_{e,i,1}^\top\left(h - Q\hat{\xi}_i\right) \le \overline{\sigma}_{e,i} & \forall i\in\{1,...,N\}\\ \overline{\gamma}_{e,i,2}^\top\left(h - Q\hat{\xi}_i\right) \le \overline{\sigma}_{e,i} & \forall i\in\{1,...,N\}\\ ||Q^\top\overline{\gamma}_{e,i,1} - Y_e||_* \le \overline{\lambda}_e & \forall i\in\{1,...,N\}\\ ||Q^\top\overline{\gamma}_{e,i,2}||_* \le \overline{\lambda}_e & \forall i\in\{1,...,N\}\\ \overline{\gamma}_{e,i,1}\ge 0; \overline{\gamma}_{e,i,2}\ge 0 & \forall i\in\{1,...,N\} \end{cases}\right\} \forall e\in\mathcal{E}, \tag{3.19d}$$

$$\left.\begin{cases} \underline{\tau}_e + \frac{1}{\underline{\epsilon}_e}\left(\underline{\lambda}_e\theta + \frac{1}{N}\sum_{i=1}^{N}\underline{\sigma}_{e,i}\right) \le 0 & \\ -Y_e\hat{\xi}_i - \underline{r}_e - \underline{\tau}_e + \underline{\gamma}_{e,i,1}^\top\left(h - Q\hat{\xi}_i\right) \le \underline{\sigma}_{e,i} & \forall i\in\{1,...,N\}\\ \underline{\gamma}_{e,i,2}^\top\left(h - Q\hat{\xi}_i\right) \le \underline{\sigma}_{e,i} & \forall i\in\{1,...,N\}\\ ||Q^\top\underline{\gamma}_{e,i,1} + Y_e||_* \le \underline{\lambda}_e & \forall i\in\{1,...,N\}\\ ||Q^\top\underline{\gamma}_{e,i,2}||_* \le \underline{\lambda}_e & \forall i\in\{1,...,N\}\\ \underline{\gamma}_{e,i,1}\ge 0; \underline{\gamma}_{e,i,2}\ge 0 & \forall i\in\{1,...,N\} \end{cases}\right\} \forall e\in\mathcal{E}, \tag{3.19e}$$

$$\left.\begin{cases} \tau_\ell + \frac{1}{\epsilon_\ell}\left(\lambda_\ell\theta + \frac{1}{N}\sum_{i=1}^{N}\sigma_{\ell,i}\right) \le 0 & \\ M_\ell^{\mathcal{E}}\left(p + Y\hat{\xi}_i\right) + M_\ell^{\mathcal{K}}\left(\mu + \hat{\xi}_i\right) - M_\ell^{\mathcal{D}} p_d & \\ \qquad - F_\ell^{\max} - \tau_\ell + \gamma_{\ell,i,1}^\top\left(h - Q\hat{\xi}_i\right) \le \sigma_{\ell,i} & \forall i\in\{1,...,N\}\\ \gamma_{\ell,i,2}^\top\left(h - Q\hat{\xi}_i\right) \le \sigma_{\ell,i} & \forall i\in\{1,...,N\}\\ ||Q^\top\gamma_{\ell,i,1} - \left(M_\ell^{\mathcal{E}}Y + M_\ell^{\mathcal{K}}\right)||_* \le \lambda_\ell & \forall i\in\{1,...,N\}\\ ||Q^\top\gamma_{\ell,i,2}||_* \le \lambda_\ell & \forall i\in\{1,...,N\}\\ \gamma_{\ell,i,1}\ge 0; \gamma_{\ell,i,2}\ge 0 & \forall i\in\{1,...,N\} \end{cases}\right\} \forall \ell\in\mathcal{L}. \tag{3.19f}$$

In (3.19), the set of decisions variables Θ comprises $\overline{\tau}_e \in \mathbb{R}, \overline{\lambda}_e \in \mathbb{R}, \overline{\sigma}_{e,i} \in \mathbb{R}, \overline{\gamma}_{e,i,1} \in \mathbb{R}^{|\mathcal{K}|}, \overline{\gamma}_{e,i,2} \in \mathbb{R}^{|\mathcal{K}|}, \underline{\tau}_e \in \mathbb{R}, \underline{\lambda}_e \in \mathbb{R}, \underline{\sigma}_{e,i} \in \mathbb{R}, \underline{\gamma}_{e,i,1} \in \mathbb{R}^{|\mathcal{K}|}, \underline{\gamma}_{e,i,2} \in \mathbb{R}^{|\mathcal{K}|}, \tau_\ell \in \mathbb{R}, \lambda_\ell \in \mathbb{R}, \sigma_{\ell,i} \in \mathbb{R}, \gamma_{\ell,i,1} \in \mathbb{R}^{|\mathcal{K}|}, \gamma_{\ell,i,2} \in \mathbb{R}^{|\mathcal{K}|}, p \in \mathbb{R}^{|\mathcal{E}|}, \overline{r} \in \mathbb{R}^{|\mathcal{E}|}, \underline{r} \in \mathbb{R}^{|\mathcal{E}|}, Y \in \mathbb{R}^{|\mathcal{E}|\times|\mathcal{K}|}, \lambda \in \mathbb{R}, \sigma_i \in \mathbb{R}$ and $\gamma_i \in \mathbb{R}^{|\mathcal{K}|}$. The aim is to derive the day-ahead power and reserve dispatch $\{p, \overline{r}, \underline{r}\}$, the additional variables being auxiliary variables involved in the modeling of uncertainty. The decision-maker may select the exogenous parameters to describe its own risk-appetite, namely, the Wasserstein radius $\theta \in \mathbb{R}$ and the violation probabilities $\epsilon_{(.)} \in [0, 1]$. Furthermore, the data-driven setting (3.19) assumes the availability of N empirical observations of the uncertainty $\hat{\xi}_i \in \mathbb{R}^{|\mathcal{K}|} \ \forall i \in \{1, ..., N\}$. The economical and technical parameters $C_e, \underline{C}_e, \overline{C}_e, \overline{P}_e, \underline{P}_e, F_\ell^{\max}$ and $R_e^{\max}$ are case-dependent and are provided in each specific study which is considered in the following Chapters 4 and 5. Finally, the dual norm appearing in the conic constraints can be easily reformulated in a linear manner using the $\infty-$norm, see [61]. The complete final program is therefore linear and we use the programming language Julia v1.7, with the modeling language JuMP v0.22.3, and the solver Gurobi v9.5, to find the optimal power flow solution. The model (3.19) is used in this thesis for benchmarking the different proposed approaches. Despite its appealing properties, model (3.19) may still suffer from limitations that are discussed in the following two case studies.

3.3.1 Case Study 1: Physical bounds

In this section, we provide an illustrative example showing a first limitation of the state-of-the-art distributionally robust paradigm[11]. The system under study is represented in Figure 3.2. We use a two-node power system, including a conventional generator, a wind farm, an inelastic demand, and a transmission line. The technical characteristics of the conventional generator are $\overline{P} = 1200$ MW, $R^{\max} = 500$ MW, $C = 15$ €/MWh, $\overline{C} = 2$ €/MW and $\underline{C} = 3$ €/MW. The installed capacity of the wind farm is 800 MW. The demand is 1000 MW. The transmission line, whose capacity is 2000 MW, is never congested.

Consider a case where the day-ahead forecast of wind power generation μ is 320 MW. Given the day-ahead forecast and the installed capacity of the wind farm, i.e., 320 MW and 800 MW, it is trivial that the maximum required upward and downward reserves are 320 MW and 480 MW, respectively. Assuming a specific case where the sole historical observation is $\hat{\xi}_i = 0$, we solve the optimal power

[11] It is worth mentioning that the DR-OPF problem (3.19) includes information about physical bounds via the so-called support of uncertainty. However, in this section, we are concerned with showing that overlooking these bounds is limiting the performance of the obtained DRO solution.

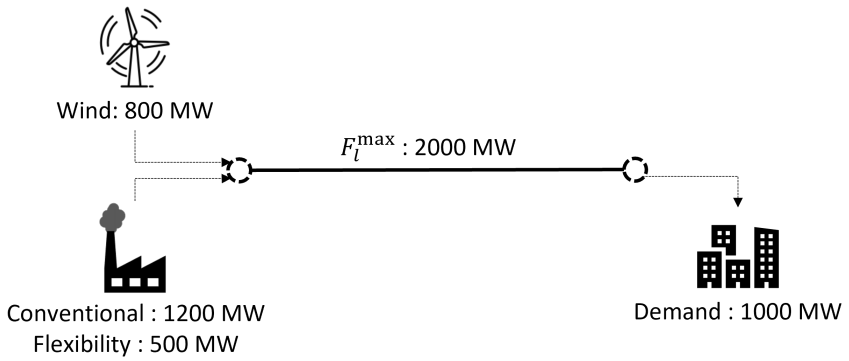


Figure 3.2: An illustrative 2-node power system connecting an inelastic demand to a conventional generator and a wind farm through a transmission line.

flow problem (3.19). The ambiguity around the predicted single-point distribution is modeled through the Wasserstein radius $\theta = \{0.03, 0.05\}$, while the violation probability of each DRCC is set to $\epsilon = 0.05$. The distributionally robust optimal dispatch of the conventional generator is given in Table 3.1.

Table 3.1: Optimal dispatch of the conventional generator

	p [MW]	$\overline{r}$ [MW]	$\underline{r}$ [MW]
$\theta = 0.03$	680	480	480
$\theta = 0.05$	Infeasible		

The interesting numerical finding is that the upward reserve dispatch of the conventional generator is $\overline{r} = 480$ MW, which is way larger than the 320-MW dispatch of the wind farm. Obviously, the upward reserve is overbooked by 160 MW of useless reserve, which has therefore increased the total operational cost of the underlying power system. Furthermore, the program may even become infeasible, e.g., for $\theta = 0.05$, owing to the need for dispatching a large amount of reserves that are not available in the system.

The reason for such naive results is that the optimization model does not impose the limits of 320 MW and 480 MW for the upward and downward reserves. These constraints can be interpreted as physical bounds, i.e., support, that should be imposed on uncertainty realizations of renewable power generation. Our work in Chapter 4 contributes to include these bounds into a more advanced and exact formulation of DRCCs [34].

3.3.2 Case Study 2: Dependencies

The second power system under study is represented in Figure 3.3. It is a slightly updated version of the one investigated in the previous section, where two wind farms are connected to the system, one at each node of the system. The maximum power capacity of each wind farm is equal to 400 MW.

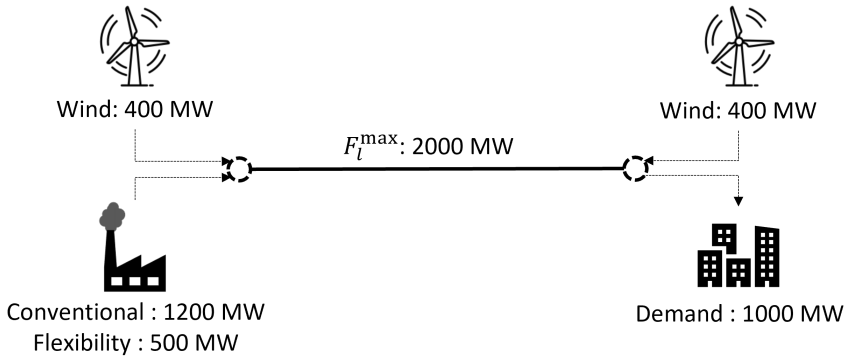


Figure 3.3: An illustrative 2-node power system connecting an inelastic demand to a conventional generator and two wind farms through a transmission line.

In this case study, the source of uncertainty comprises two dimensions, one for each wind farm. Let us assume that, for some unknown meteorological reasons, the two wind farms have a completely inversely dependent relation, i.e., when the wind farm 1 produces a lot of energy, the wind farm 2 produces less. In other words, it is (almost surely) expected from past observations, that the aggregated wind power generation always takes the same constant value, i.e., 320 MW when $\mu = 0.4$. These past observations are shown in Figure 3.4. In such a case, it is trivial that the need for reserve always equals zero (as system operator is almost sure of real-time total wind power generation).

We solve the optimal power flow problem (3.19). The ambiguity around the predicted empirical distribution distribution (shown in Figure 3.4) is modeled through the Wasserstein radius $\theta = \{0.01, 0.03, 0.05\}$, while the violation probability of each DRCC is set to $\epsilon = 0.05$. The distributionally robust optimal dispatch of the conventional generator is given in Table 3.2.

Interestingly, we observe that the upward and downward reserve are overbooked by 160 MW and 480 MW, respectively for the cases where $\theta = 0.01$ and $\theta = 0.03$. This is due to the fact that the dependence structure of past observation is not enforced in the modeling of ambiguity set, resulting in potentially misrepresentation of

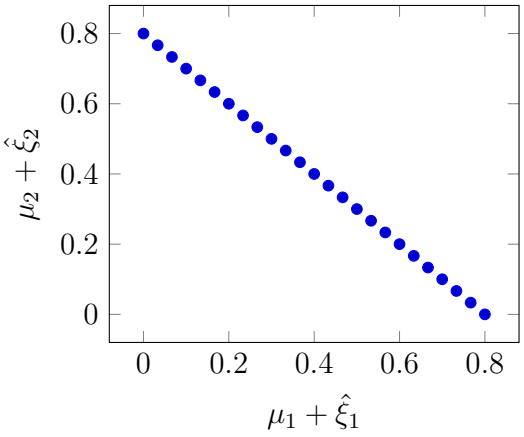


Figure 3.4: Available past observations for the second case study. In this arbitrary illustrative example, the uncertainty has two dimensions, for which a strong dependence structure exists, i.e., wind farm 1 always produces less when wind farm 2 produces more, and inversely.

Table 3.2: Optimal dispatch of the conventional generator

	p [MW]	$\overline{r}$ [MW]	$\underline{r}$ [MW]
$\theta = 0.01$	600	160	160
$\theta = 0.03$	600	480	480
$\theta = 0.05$	Infeasible		

worst-case distribution functions, i.e., which follow a different dependence structure comparatively to the empirically observed one. Furthermore, the program may even become infeasible, e.g., for $\theta = 0.05$, owing to the need for dispatching a large amount of reserves that are not available in the system. This calls for a new definition of ambiguity that incorporates a restriction on dependence structure. Our work in Chapter 5 aims at incorporating such information on dependence structure within the Wasserstein metric-based DRO.

3.4 Chapter Summary and Conclusions

This chapter conceptually introduce DRO as an advanced uncertainty modeling technique that is capable of considering a family of potential distributions describing, e.g., the wind power generation uncertainty. By doing so, the method allows to hedge against inexactness in the representation of uncertainty within decision-making problems for energy systems. In that direction, we apply DRO to the OPF problem and provide tractable reformulations, regarding the worst-case

expectation problem in the objective function and the DRCCs within the set of constraints. Interestingly, we highlight two potential limitations of the existing approaches. Firstly, overlooking the physical bounds of uncertainty into the formulation yields suboptimal and overconservative solutions which may jeopardize the cost-efficiency of the distributionally robust approach. Secondly, there is no guarantee that the distributions within the ambiguity set are consistent with the empirically observed dependence structure, which may have similar effects.

To address these two limitations in the most recent advances in DRO, we consider the incorporation of physical bounds into the exact reformulation of DRCCs in Chapter 4. Then, we enhance the definition of metric-based ambiguity sets in Chapter 5 by incorporating information on dependence structure of the uncertain parameters.

CHAPTER 4

An Exact and Physically-Bounded Distributionally Robust OPF

In the context of the energy transition, modeling the uncertainty of weather-dependent renewable power generation within the OPF problems may help improve the cost-efficiency and reliability of short-term operational procedures. In that direction, we focus on the most recent advances in DRO and develop an *exact* DRCC-OPF model with a Wasserstein metric-based ambiguity set. While preserving the exactness, we then improve the model by enforcing physical bounds on the uncertainty space. These bounds ensure that the obtained solutions are realistic and feasible. The resulting model is bilinear and we propose a solution approach based on an iterative algorithm which is computationally efficient and has convergence guarantee. A thorough out-of-sample analysis is performed to compare the proposed model against a scenario-based stochastic program. We also compare the performance of the proposed exact reformulation against a state-of-the-art approximate technique in the literature, built upon the conditional-value-at-risk measure. We eventually show that the proposed physically-bounded exact reformulation outperforms the other methods by achieving a cost-optimal yet reliable trade-off between reserve procurement and load curtailment.

The content of this chapter is mainly based on the following publication[1]

[B] **A. Arrigo**, C. Ordoudis, J. Kazempour, Z. De Grève, J.-F. Toubeau, and F. Vallée, "Wasserstein distributionally robust chance-constrained optimization for energy and reserve dispatch: An exact and physically-bounded formulation," European Journal of Operations Research, vol. 269, no. 1, pp. 304-322, January 2022.

[1]The Elsevier permission grant is available at `https://www.elsevier.com/about/policies/copyright/permissions`, Accessed on 26$^{\text{th}}$ April 2022.

4.1 Motivation

With the growing integration of renewables within the electricity generation mix, it becomes crucial to appropriately accommodate the uncertain and weather-dependent renewable power in-feeds within the OPF problems. The uncertainty-aware OPF problem has been extensively addressed in the literature, where the uncertainty is modeled using scenario-based stochastic programming [108, 114], robust optimization [115, 116], or chance-constrained programming [91, 117]. However, any inadequate representation of uncertainty in this probabilistic optimization problem may lead to a sub-optimal solution in terms of the dispatch of flexible resources, and therefore an increase in the total operational cost of the system [119]. The challenge is that the *true* probability distribution of the renewable power generation uncertainty is not necessarily known [120, 121]. Therefore, any uncertainty modeling technique relying on a specific distribution may fail in achieving the optimal dispatch of flexible resources.

In that regard, DRO represents a promising technique which allows to hedge against the misrepresentations of uncertainty. Within the DRO framework, it is desirable to incorporate the available historical data of uncertainty realization as much as possible within the model. To do so, we focus on the *metric-based* DRO[2] where the ambiguity set encompasses all distributions, whose probability distance from an empirical distribution including the historical data is lower than or equal to a predefined value, i.e., the radius. In addition, the operating restrictions within the OPF problem may be modeled via probabilistic constraints in the form of DRCCs. The rationale behind DRCCs is to allow the violation of the underlying probabilistic constraints up to a predefined extent for the worst-case distribution. This brings an extra degree of freedom for the decision-maker to impose her risk attitude by setting a violation probability for each probabilistic constraint *a priori*. This setting results in a Wasserstein metric-based distributionally robust chance-constrained OPF problem, which is the focus of this chapter.

Several works exist in the literature that use metric-based distributionally robust chance-constrained programs for power system applications, e.g., [27–29] and [122]. To the best of our knowledge, all these works *approximate* DRCCs, e.g., using a Conditional Value-at-Risk (CVaR) reformulation, as proposed by [94] and [20]. This may result in a sub-optimal or over-conservative solution. In this chapter, our main contribution is to reformulate DRCCs within the OPF problem in an *exact* manner, resulting in optimal dispatch decisions. In the following, we explain our methodological steps, and further clarify the contributions in this chapter.

[2]Unlike the moment-based DRO which only considers the moment estimation, the metric-based one takes the advantage of existing additional data, but at the potential cost of increased computational burden.

Following [34] and [123], we first provide an *exact* reformulation for the worst-case expectation in the objective function and for all DRCCs. The resulting model is a mixed-integer linear program (MILP). While keeping the exact reformulation, we then improve the proposed model by including the *physical bounds* of uncertain parameters, the so-called *support*. For this purpose, we limit the renewable power dispatch to lie within zero and the installed capacity of the corresponding renewable energy unit. The importance of including these additional limits for obtaining realistic and cost-efficient solutions is argued in Section 3.3. All analytical reformulations are introduced in the following Section 4.2, which results in a bilinear program, instead of the MILP model without support. We solve the resulting bilinear model using an iterative algorithm, which is similar to the one in [94]. Our various numerical experiments show that the algorithm is computationally efficient and converges to a solution, although a theoretical guarantee for convergence to *global optimality* does not exist. We provide an extensive numerical convergence analysis based on a case study that allows us to quantify the gap between the optimal solution obtained by the iterative algorithm and that achieved by the non-linear solver IPOPT [124], which serves as a benchmark in this chapter.

Finally, we provide a thorough numerical analysis in Section 4.3 based on an extensive out-of-sample simulation to compare different alternatives for DRCC reformulation, including (*i*) the CVaR approximation as proposed in [94], (*ii*) the exact MILP reformulation without support as proposed in [34] and [123], and (*iii*) our proposed exact physically-bounded bilinear formulation. Such a comparison is extended by taking into account the scenario-based stochastic programming as an additional alternative technique for modeling renewable power generation uncertainty.

As our main finding, we show that our proposed physically-bounded DRO model with the exact reformulation outperforms all other techniques in terms of the total expected operational cost of the system and its standard deviation. Furthermore, our numerical analysis stresses the importance of incorporating the physical bounds of uncertain parameters in order to fully leverage the benefits of the DRO approach.

Notational Remark: The symbols with a hat, e.g., $\hat{\xi}$, refer to historical observations. The operator $|.|$ returns the cardinality of the underlying set. Furthermore, $(.)^\top$ is the transpose operator, whereas $\mathbb{E}^{\mathbb{Q}}[.]$ refers to the expected value with respect to the probability distribution $\mathbb{Q}$. The operator $||a||_*$ represents the dual norm of $||a||$. This dual norm is defined as $||a||_* = \left\{\max_v \; a^\top v \text{ s.t. } ||v|| \leq 1\right\}$, where v is a vector with the same dimension than vector a.

4.2 The Exact and Physically-Bounded DR-OPF Model

In this section, we derive the necessary reformulations for the Distributionally Robust Optimal Power Flow (DR-OPF) problem (3.10) given the proposed exact and physically-bounded approach. First, we reformulate objective function (3.10a). Following the exact reformulation technique proposed by [20], we replace the inner maximization problem $\max_{\mathbb{Q}\in\mathcal{A}} \mathbb{E}^{\mathbb{Q}}\left[C^\top Y\xi\right]$ by a linear minimization problem, i.e.,

$$\begin{cases} \min\limits_{\lambda,\sigma_i,\gamma_i} \ \lambda\theta + \dfrac{1}{N}\sum\limits_{i=1}^{N}\sigma_i & \\ \text{s.t. } C^\top Y\hat{\xi}_i + \gamma_i^\top\left(h - Q\hat{\xi}_i\right) \leq \sigma_i & \forall i \in \{1,...,N\} \\ \quad ||Q^\top\gamma_i - C^\top Y||_* \leq \lambda & \forall i \in \{1,...,N\} \\ \quad \gamma_i \geq 0 & \forall i \in \{1,...,N\}, \end{cases} \tag{4.1}$$

where $\lambda \in \mathbb{R}$, $\sigma \in \mathbb{R}^N$ and $\gamma_i \in \mathbb{R}^{2|\mathcal{K}|}$ are auxiliary variables. The support, defined as $Q\xi \leq h$, restricts the worst-case distribution to take realistic values of uncertainty. The resulting min-min objective function is then collapsed to a single minimization problem.

Second, we reformulate DRCCs (3.10h)-(3.10j) using three different alternatives, namely

- The CVaR approximation[3] as proposed in [94],
- The exact MILP without support as proposed in [34] and [123],
- Our proposed physically-bounded exact bilinear reformulation.

The first alternative has been previously introduced as a state-of-the-art technique in Section 3.2.2. This alternative will be used as a benchmark in the numerical experiments of Section 4.3. The advantage of CVaR approximation is that it preserves linearity and allows for the incorporation of support information, but at the potential cost of a conservative solution [125]. The reason for the increased conservativeness is that the CVaR inherently accounts for the severity of the violation, resulting in a lower ex-post violation probability compared to the pre-defined one. Using the CVaR approximation, the OPF model (3.10) is eventually reformulated as a linear program, which is provided in 3.3.

[3] Chen et al. (2018, Corollary 1) show that the CVaR approximation is exact under a condition for which $\epsilon \leq \frac{1}{N}$.

In the remainder of this section, Subsection 4.2.1 details the second alternative, i.e., an exact MILP reformulation for DRCCs without support. We include the modeling framework of the second alternative in the main body of this chapter, since it provides the basis for our proposed model. Recall that Subsection 3.3.1 provides a stylized example, illustrating why the state-of-the-art reformulations without support may not determine the optimal energy and reserve dispatch in a desirable way. Subsection 4.2.2 presents our proposed exact reformulation with support, and highlights its advantages.

Remark 1. Hereinafter, for the sake of notational clarity, each distributionally robust chance constraint (3.10h), (3.10i) and (3.10j) is rewritten in a generic way as $\mathbb{Q}\left(\tilde{\xi} \in S(\mathcal{Y})\right) \geq 1-\epsilon$, where $S(\mathcal{Y}) = \left\{\tilde{\xi} \in \mathbb{R}^{|\mathcal{K}|} |\ a^\top \tilde{\xi} \leq b\right\}$ denoted as *safe set*. This set represents the geometrical region where the underlying constraint is realized. Note that variable vector $a = A\mathcal{Y} + \breve{a}$ and variable scalar $b = b^\top \mathcal{Y} + \breve{b}$ are both affine functions of decision variables $\mathcal{Y} = \{p, \overline{r}, \underline{r}, Y\}$ with $a, \breve{a} \in \mathbb{R}^{|\mathcal{K}|}$, $A \in \mathbb{R}^{|\mathcal{K}| \times |\mathcal{Y}|}$, $b, \breve{b} \in \mathbb{R}$ and $b \in \mathbb{R}^{|\mathcal{Y}|}$. The vector $\mathcal{Y}$ collects all day-ahead dispatch decisions including those within vectors $p, \overline{r}, \underline{r}$ and matrix Y in a vector form. We denote the *unsafe set* as $\overline{S}(\mathcal{Y}) = \left\{\tilde{\xi} \in \mathbb{R}^{|\mathcal{K}|} |\ a^\top \tilde{\xi} > b\right\}$, which is the complementary region of $S(\mathcal{Y})$.

4.2.1 Exact Reformulation of DRCCs Without Physical Bounds

We consider the availability of N samples, each corresponding to a historical observation of renewable power deviation in real time with respect to the day-ahead forecast. These samples are collected in the set $\left\{\hat{\xi}_i |\ i = 1, ..., N\right\}$. Fig. 4.1 depicts these samples as well as a half space representing the safe set $S(\mathcal{Y})$ for an individual DRCC. Note that one of the samples in Fig. 4.1 is outside the safe set, as the chance constraint allows violating the constraint to some extent. We also consider a *displacement budget* as a function of radius $\theta \in \mathbb{R}_+$, whose value is assigned by the system operator. The rationale behind DRO is to determine the worst-case distribution $\mathbb{Q}$ within the ambiguity set $\mathcal{P}$ that moves as much as possible samples $\hat{\xi}_i$ outside the safe set with the given displacement budget [126]. These moves are illustrated by red arrows in Fig. 4.1, where the worst-case distribution, for example, could fully move two samples $\hat{\xi}_2$ and $\hat{\xi}_3$ to the boundary of the safe set, while the fourth sample $\hat{\xi}_4$ is partially moved due to the limited displacement budget. In Section 4.2.2, we will propose a new framework to move the samples while accounting for the support.

In order to express the mathematical equivalence of each DRCC, we first introduce index $\pi_i(\mathcal{Y})$ that reorders samples $\hat{\xi}_i$ in increasing distance to the boundary as a

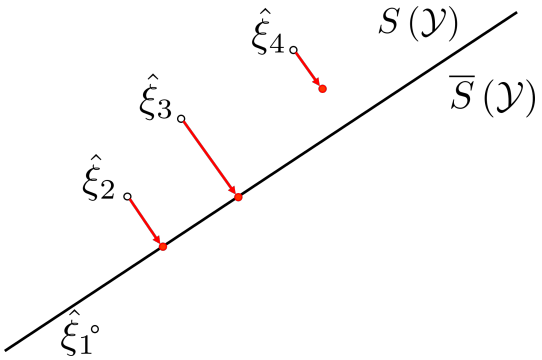


Figure 4.1: Historical observations and safe set corresponding to an arbitrarily selected distributionally robust chance constraint

function of decision variables $\mathcal{Y}$. We also define distance function $\text{dist}(\hat{\xi}_i, \overline{S}(\mathcal{Y}))$ that computes the distance between the underlying sample $\hat{\xi}_i$ within the safe set and the boundary of the unsafe set $\overline{S}(\mathcal{Y})$ in the geometrical space. The samples that are already unsafe are assigned with a distance equal to zero. According to [34], each DRCC, when $\theta > 0$, is mathematically equivalent to a regular inequality constraint of the form

$$\sum_{i=1}^{\epsilon N} \text{dist}\left(\hat{\xi}_{\pi_i(\mathcal{Y})}, \overline{S}(\mathcal{Y})\right) \geq \theta N. \tag{4.2}$$

The proof is available in [34]. The sum operator in (4.2) picks the first ϵ times N number of samples. However, this number is not necessarily an integer value. For example, assume the distance of four samples $\hat{\xi}_1$ to $\hat{\xi}_4$ close to the boundary in Fig. 4.1 are 5, 10, 12 and 15, respectively. If $\epsilon N = 2$, then the left-hand side of (4.2) is equal to 15 (i.e., 5 + 10). However, in case $\epsilon N = 2.5$, then it would be equal to 21 (i.e., 5 + 10 + 6), taking into account the half of the distance corresponding to the third sample. To get rid of the partial sum operator and the permutation index $\pi_i(\mathcal{Y})$, inequality (4.2) is replaced by the set of equations (4.3) without approximation, but at the cost of adding extra variables $t \in \mathbb{R}$ and $\beta \in \mathbb{R}_+^N$ [34, 123]:

$$\epsilon N t - e^\top \beta \geq \rho N, \tag{4.3a}$$

$$\text{dist}(\hat{\xi}_i, \overline{S}(\mathcal{Y})) \geq t - \beta_i \;\; \forall i \in \{1, ..., N\}. \tag{4.3b}$$

The last task is to reformulate the distance function $\text{dist}(\hat{\xi}_i, \overline{S}(\mathcal{Y}))$. Following [34], we replace (4.3) by (4.4) with no approximation, but at the cost of adding binary

variables $q \in \{0,1\}^N$:

$$\epsilon N t - e^\top \beta \geq \rho N ||a||_*, \tag{4.4a}$$
$$b - a^\top \hat{\xi}_i + M q_i \geq t - \beta_i \ \ \forall i \in \{1, ..., N\}, \tag{4.4b}$$
$$M\,(1 - q_i) \geq t - \beta_i \ \ \forall i \in \{1, ..., N\}. \tag{4.4c}$$

It is worth mentioning that (4.4) holds for $a \neq 0$, although Remark 2 in [34] explains how it generalizes to a case where $a = 0$. In order to maintain linearity, we pick an ∞-norm[4], i.e., $||a||_\infty$, whose dual is a 1-norm. Constraints (4.4b) and (4.4c) include a sufficiently large constant $M \in \mathbb{R}_+$. It is of importance to select a proper value for M, as a small value may affect optimality, while a very large one may result in numerical ill-conditioning. Note that if $q_i = 0$, then (4.4b) becomes binding while (4.4c) is inactive. On the contrary, $q_i = 1$ makes (4.4c) active while (4.4b) is inactive. The collection of (4.4b) and (4.4c) ensures that the value of $\text{dist}(\hat{\xi}_i, \overline{S}(\mathcal{Y}))$ is always non-negative.

The DRCCs (3.10i), (3.10h) and (3.10j) can now be replaced in an exact way by their generic equivalence (4.4). Using this exact technique, the OPF model (3.10) is eventually reformulated as a MILP, which is as follows

$$\min_{\Theta^{\text{MILP}}} \ \sum_{e \in \mathcal{E}} C_e p_e + \overline{C}_e \overline{r}_e + \underline{C}_e \underline{r}_e + \lambda\theta + \frac{1}{N}\sum_{i=1}^{N} \sigma_i \tag{4.5a}$$
$$\text{s.t. (3.10b) - (3.10g)}, \tag{4.5b}$$
$$\begin{cases} C^\top Y \hat{\xi}_i + {\gamma_i}^\top \left(h - Q\hat{\xi}_i\right) \leq \sigma_i, & \forall i \in \{1, ..., N\}, \\ ||Q^\top \gamma_i - C^\top Y||_* \leq \lambda, & \forall i \in \{1, ..., N\}, \\ \gamma_i \geq 0, & \forall i \in \{1, ..., N\}, \end{cases} \tag{4.5c}$$
$$\left.\begin{cases} \epsilon N \overline{t}_e - e^\top \overline{\beta}_e \geq \theta N ||Y_e||_*, & \\ \overline{r}_e - Y_e \hat{\xi}_i + M \overline{q}_{g,i} \geq \overline{t}_e - \overline{\beta}_{e,i}, & \forall i \in \{1, ..., N\}, \\ M\left(1 - \overline{q}_{e,i}\right) \geq \overline{t}_e - \overline{\beta}_{e,i}, & \forall i \in \{1, ..., N\}, \\ \overline{q}_e \in \{0,1\}^N, \ \ \beta \geq 0, & \end{cases}\right\} \ \forall e \in \mathcal{E}, \tag{4.5d}$$
$$\left.\begin{cases} \epsilon N \underline{t}_e - e^\top \underline{\beta}_e \geq \theta N ||-Y_e||_*, & \\ \underline{r}_e + Y_e \hat{\xi}_i + M \underline{q}_{e,i} \geq \underline{t}_e - \underline{\beta}_{e,i}, & \forall i \in \{1, ..., N\}, \\ M\left(1 - \underline{q}_{e,i}\right) \geq \underline{t}_e - \underline{\beta}_{e,i}, & \forall i \in \{1, ..., N\}, \\ \underline{q}_e \in \{0,1\}^N, \ \ \beta \geq 0, & \end{cases}\right\} \ \forall e \in \mathcal{E}, \tag{4.5e}$$

[4]This is an arbitrary choice. Alternatively, one can select the 1-norm to maintain linearity.

$$\left\{\begin{array}{ll} \epsilon N t_\ell - e^\top \beta_\ell \geq \theta N ||M_\ell^{\mathcal{E}} Y + M_\ell^{\mathcal{K}} W||_*, & \\ F_\ell^{\max} - \Big(M_\ell^{\mathcal{E}}\left(p + Y\xi\right) + M_\ell^{\mathcal{K}} W\left(\mu + \xi\right) & \\ \qquad -M_\ell^{\mathcal{D}} p_d\Big) + M q_{\ell,i} \geq t_\ell - \beta_{\ell,i}, & \forall i \in \{1, ..., N\}, \\ M\left(1 - q_{\ell,i}\right) \geq t_\ell - \beta_{\ell,i}, & \forall i \in \{1, ..., N\}, \\ q_\ell \in \{0,1\}^N, \ \beta \geq 0, & \end{array}\right\} \ \forall \ell \in \mathcal{L}, \tag{4.5f}$$

where the variable set Θ^{MILP} includes p, $\overline{r}$, $\underline{r}$, Y, $\overline{t}_e$, $\overline{\beta}_e$, $\overline{q}_e$, $\underline{t}_e$, $\underline{\beta}_e$, $\underline{q}_e$, t_ℓ, β_ℓ, q_ℓ, λ, σ_i and γ_i. Note that constraint (4.5b) introduces the operating constraints (3.10b) - (3.10g). Constraints (4.5c) are associated with the objective function reformulation. Finally, constraints (4.5d), (4.5e) and (4.5f) present the exact MILP reformulation of DRCCs (3.10h), (3.10i) and (3.10j), respectively.

4.2.2 The Proposed Exact and Physically-Bounded Reformulation of DRCCs

We add the information of physical bounds to the safe set, as schematically illustrated in Fig. 4.2 by a generic green line. This line represents one of the boundaries of a convex polyhedral support. The gray area is within the safe set but outside the support. If the worst-case distribution displaces samples $\hat{\xi}_i$ inside the gray area, the resulting dispatch solution will be overly naive as explained in the previous section. The inclusion of support modifies the sample displacement strategy to find the worst-case distribution: samples must be moved towards the unsafe set boundary without crossing the gray area. For example, samples $\hat{\xi}_3$ and $\hat{\xi}_4$ now take another path towards the unsafe set compared to the one in Fig. 4.1. This path is characterized by the direction of the shorter path towards the boundary, while being restricted by physical bounds.

The key point is to enhance the distance function in (4.3b) to be able to take into account the support information. To this purpose, we develop a mathematical framework with three consecutive steps. First, we derive an analytical expression to compute the geometrical distance between a sample and the boundary of the unsafe set while considering the support. Second, we reformulate an individual DRCC based on the updated distance function, resulting in a bilinear program. Third, we leverage an iterative algorithm to efficiently solve the resulting program.

Deriving the Expression of the Updated Distance Function

The following Lemma 4.1 incorporates the support information into the function $\text{dist}(\hat{\xi}_i, \overline{S}(\mathcal{Y}))$.

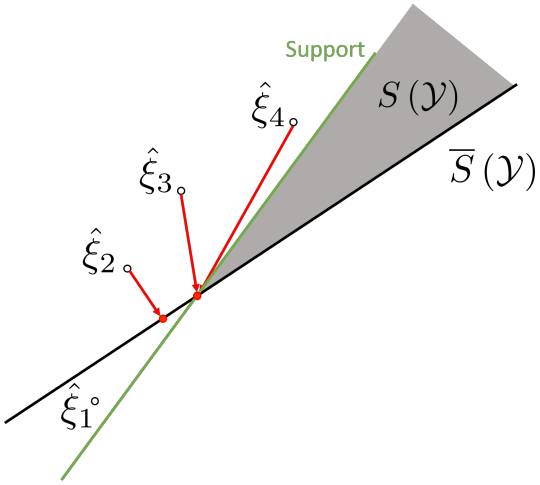


Figure 4.2: Illustration of the support (green line), enforcing the gray area to be outside the feasibility space. In comparison to Fig. 4.1, the support changes the displacement path (red arrows) of historical observations $\hat{\xi}_i$ towards the unsafe set $\overline{S}(\mathcal{Y})$.

Lemma 4.1. *Let $\{Q\xi \leq h\}$ be the set of $|\phi|$ linear constraints describing the physical bounds of uncertainty, where $Q \in \mathbb{R}^{|\phi|\times|\mathcal{K}|}$ and $h \in \mathbb{R}^{|\phi|}$. The distance between sample $\hat{\xi}_i$ and boundary of the unsafe set, including the support information, is obtained as*

$$dist(\hat{\xi}_i, \overline{S}(\mathcal{Y})) = \max_{w_i, x_i} \ (b - a^\top \hat{\xi}_i) w_i - \left(h - Q\hat{\xi}_i\right)^\top x_i \tag{4.6a}$$

$$s.t. \ \ ||a w_i - Q^\top x_i||_* \leq 1, \tag{4.6b}$$

where $w_i \in \mathbb{R}_+$ and $x_i \in \mathbb{R}_+^{|\phi|}$ are additional auxiliary variables.

Proof. The distance between sample $\hat{\xi}_i$ and boundary of the unsafe set $\overline{S}(\mathcal{Y})$, including the physical bounds information, can be formulated as

$$\text{dist}(\hat{\xi}_i, \overline{S}(\mathcal{Y})) = \min_{\zeta_i, \xi_i} \ \zeta_i \tag{4.7a}$$

$$\text{s.t.} \ \ b - a^\top \xi_i \leq 0 \qquad : w_i \geq 0, \tag{4.7b}$$

$$Q\xi_i - h \leq 0 \qquad : x_i \geq 0 \tag{4.7c}$$

$$||\xi_i - \hat{\xi}_i|| \leq \zeta_i \qquad : (v, u) \in \mathcal{K}_*, \tag{4.7d}$$

where $\zeta_i \in \mathbb{R}$ and $\xi_i \in \mathbb{R}^{|\mathcal{W}|}$ are the primal variables. The dual variables corresponding to each constraint are given alongside a colon. These dual variables are $w_i \in \mathbb{R}_+$ and $x_i \in \mathbb{R}_+^{|\phi|}$ as well as $(v, u) \in \mathbb{R}^{|\mathcal{W}|} \times \mathbb{R}$ residing in the dual cone $\mathcal{K}_*$.

The objective is to transfer the given sample $\hat{\xi}_i$ to the unsafe set imposed by (4.7b) while respecting support (4.7c) with the minimum displacement budget. This introduces the minimum distance ζ_i and variable ξ_i. The latter is the transferred $\hat{\xi}_i$, which is now located at the boundary of the unsafe set. The distance between $\hat{\xi}_i$ and ξ_i is calculated by conic constraint (4.7d). Inspired by the proof of Lemma 2 in [34], we aim at dualizing (4.7) to obtain a maximization problem. The resulting Lagrangian problem is

$$\max_{u,v,w_i,x_i} \min_{\zeta_i,\xi_i} \ \zeta_i - v^\top \left(\xi_i - \hat{\xi}_i\right) - u\zeta_i + w_i \left(b - a^\top \xi_i\right) + x_i^\top \left(Q\xi_i - h\right) \quad (4.8a)$$

$$\text{s.t. } ||v||_* \leq u, \ \ w_i \geq 0, \ \ x_i \geq 0. \quad (4.8b)$$

Deriving the Karush-Kuhn-Tucker conditions of the inner minimization problem allows us to find the dual problem. This finally renders

$$\begin{aligned} \text{dist}(\hat{\xi}_i, \overline{S}(\mathcal{Y})) = \max_{u,v,w_i,x_i} \ & \hat{\xi}_i^\top v + bw_i - h^\top x_i && (4.9a)\\ \text{s.t. } & ||v||_* \leq u && (4.9b)\\ & u = 1 && (4.9c)\\ & v = -aw_i + Q^\top x_i && (4.9d)\\ & w_i \geq 0, \ \ x_i \geq 0. && (4.9e) \end{aligned}$$

Substituting v and u as defined by (4.9c) and (4.9d) in (4.9a) and (4.9b) yields (4.6) in Lemma 4.1. ■

Reformulation

Using Lemma 4.1, Theorem 4.1 provides our proposed exact and physically-bounded reformulation of DRCCs.

Theorem 4.1. *The exact and physically-bounded reformulation of each DRCC is*

$$\epsilon N t - e^\top \beta \geq \theta N, \quad (4.10a)$$

$$(b - a^\top \hat{\xi}_i) w_i - \left(h - Q\hat{\xi}_i\right)^\top x_i \geq t - \beta_i \ \ \forall i \in \{1, ..., N\}, \quad (4.10b)$$

$$||aw_i - Q^\top x_i||_* \leq 1, \quad (4.10c)$$

$$w_i \geq 0, \ \ x_i \geq 0, \ \ \beta \geq 0. \quad (4.10d)$$

Proof. Replacing $\text{dist}(\hat{\xi}_i, \overline{S}(\mathcal{Y}))$ in (4.3b) by the maximization problem (4.6) results in a constraint in the form of $\left\{ \max_{w_i,x} \ f(w_i, x_i) \text{ s.t. (4.6b)} \right\} \geq t - \beta_i \ \ \forall i$. This implies that there exist optimal values for w and x which respect (4.6b) and the

optimal value of $f_i(w_i, x_i)$ is greater than or equal to $t - \beta_i$. Equivalently, we can drop the max operator, treat w_i and x_i as variables of min operator in (3.10a), and impose $\{f(w_i, x_i) \geq t - \beta_i \text{ and (4.6b)}\}\, \forall i$. This results in (4.10) in Theorem 4.1. ■

We replace individual DRCCs (3.10i), (3.10h) and (3.10j) by their corresponding reformulation (4.10) and write the final model as follows:

$$\min_{\Theta^{\text{Exact}}} \sum_{e\in\mathcal{E}} C_e p_e + \overline{C}_e \overline{r}_e + \underline{C}_e \underline{r}_e + \lambda\theta + \frac{1}{N}\sum_{i=1}^{N}\sigma_i \tag{4.11a}$$

$$\text{s.t. (3.10b) - (3.10g)}, \tag{4.11b}$$

$$\begin{cases} C^\top Y\hat{\xi}_i + {\gamma_i}^\top\left(h - Q\hat{\xi}_i\right) \leq \sigma_i & \forall i \in \{1,...,N\} \\ ||Q^\top\gamma_i - C^\top Y||_* \leq \lambda & \forall i \in \{1,...,N\} \\ \gamma_i \geq 0 & \forall i \in \{1,...,N\} \end{cases} \tag{4.11c}$$

$$\left.\begin{cases} \epsilon N\overline{t}_e - e^\top\overline{\beta}_e \geq \theta N, & \\ (\overline{r}_e - Y_e\hat{\xi}_i)\overline{w}_{e,i} - \left(h - Q\hat{\xi}_i\right)^\top\overline{x}_{e,i} \geq \overline{t}_e - \overline{\beta}_{e,i}, & \forall i \in \{1,...,N\}, \\ ||Y_e\overline{w}_{e,i} - Q^\top\overline{x}_{e,i}||_* \leq 1, & \forall i \in \{1,...,N\}, \\ \overline{w}_{e,i} \geq 0,\ \overline{x}_{e,i} \geq 0,\ \overline{\beta}_e \geq 0, & \end{cases}\right\} \forall e \in \mathcal{E}, \tag{4.11d}$$

$$\left.\begin{cases} \epsilon N\underline{t}_e - e^\top\underline{\beta}_e \geq \theta N, & \\ (\underline{r}_e + Y_e\hat{\xi}_i)\underline{w}_{e,i} - \left(h - Q\hat{\xi}_i\right)^\top\underline{x}_{e,i} \geq \underline{t}_e - \underline{\beta}_{e,i}, & \forall i \in \{1,...,N\}, \\ ||-Y_e\underline{w}_{e,i} - Q^\top\underline{x}_{e,i}||_* \leq 1, & \forall i \in \{1,...,N\}, \\ \underline{w}_{e,i} \geq 0,\ \underline{x}_{e,i} \geq 0,\ \underline{\beta}_e \geq 0, & \end{cases}\right\} \forall e \in \mathcal{E}, \tag{4.11e}$$

$$\left.\begin{cases} \epsilon N t_\ell - e^\top\beta_\ell \geq \theta N, & \\ \left(F_\ell^{\max} - \left(M_\ell^{\mathcal{E}}\left(p + Y\hat{\xi}_i\right) + M_\ell^{\mathcal{K}}W\left(\mu + \hat{\xi}_i\right) - M_\ell^{\mathcal{D}}p_d\right)\right)w_{\ell,i} & \\ \qquad - \left(h - Q\hat{\xi}_i\right)^\top x_{\ell,i} \geq t_\ell - \beta_{\ell,i},\ \ \forall i \in \{1,...,N\}, & \\ ||\left(M_\ell^{\mathcal{E}}Y + M_\ell^{\mathcal{K}}W\right)w_{\ell,i} - Q^\top x_{\ell,i}||_* \leq 1,\ \ \forall i \in \{1,...,N\}, & \\ w_{\ell,i} \geq 0,\ \ x_{\ell,i} \geq 0,\ \ \beta_l \geq 0, & \end{cases}\right\},\ \ \forall \ell \in \mathcal{L}, \tag{4.11f}$$

where the variable set Θ^{Exact} includes p, $\overline{r}$, $\underline{r}$, Y, $\overline{t}_e$, $\overline{\beta}_e$, $\overline{w}_{e,i}$, $\overline{x}_{e,i}$, $\underline{t}_e$, $\underline{\beta}_e$, $\underline{w}_{e,i}$, $\underline{x}_{e,i}$, t_ℓ, β_ℓ, $w_{\ell,i}$, $x_{\ell,i}$, λ, σ_i and γ_i. Constraint (4.11b) recalls the operational limits of

conventional generators and power balance conditions (3.10b) - (3.10g) within the original model (3.10) operating constraints. Constraints (4.11c) pertain to the objective function reformulation. Finally, constraints (4.11d), (4.11e) and (4.11f) present the physically-bounded exact reformulation of DRCCs (3.10h), (3.10i) and (3.10j), respectively.

The final set of decision variables is now $\{\mathcal{Y}, t, \beta, w_i, x_i\}$. It should be noted that the set of equations (4.10) contains bilinear terms bw_i and aw_i. Recall that a and b are affine functions of the dispatch decision variable $\mathcal{Y}$. However, unlike (4.4), the resulting equations (4.10) do not include any binary variable. For the sake of clarity, we rewrite the final dispatch model in a compact form as

$$\min_{\mathcal{Y},t,\beta,w_i,x_i} \quad \mathcal{J}(\mathcal{Y}) \tag{4.12a}$$

$$\text{s.t.} \quad \mathcal{H}_\alpha(\mathcal{Y}, t, \beta, w_i, x_i) \leq 0 \quad , \alpha = \{1, ..., |\alpha|\}, \tag{4.12b}$$

$$\mathcal{R}_\eta(\mathcal{Y}, t, \beta, w_i, x_i) \leq 0 \quad , \eta = \{1, ..., |\eta|\}, \tag{4.12c}$$

where $\mathcal{J}(\mathcal{Y})$ represents the objective function (3.10a) after its exact reformulation. The linear and bilinear constraints are gathered in (4.12b) and (4.12c), respectively. Note that $|\alpha|$ and $|\eta|$ represent the number of linear and bilinear constraints, respectively.

Solution Algorithm

Inspired by [94] and [122], we solve the bilinear model (4.12) using an iterative sequential algorithm, as explained in Algorithm 1. The iterative procedure is straightforward. Step 1 sets initial variables $\mathcal{Y}$ and $\mathcal{J}^0 = 10^{-3}$ which ensures that the algorithm will not terminate at the first iteration, since $\mathcal{J}^0$ appears in the denominator of the convergence criterion. In Step 2, for given decision variables $\mathcal{Y}^{\text{fixed}}$, we solve (4.13), aiming at enlarging the feasible space of bilinear constraints as a function of $\{t, \beta, w_i, x_i\}$. Note that (4.13) is linear as the decision variable set $\mathcal{Y}$ is fixed. Then, for the obtained optimal values of w_i, the linear optimization (4.14) updates the value of $\mathcal{Y}$ in Step 3. Finally, Step 4 evaluates whether the convergence criterion is fulfilled and goes back to Step 2 if it is not. We select the convergence threshold to be equal to 10^{-4}, meaning that the iterative algorithm will stop if two consecutive items of the sequence do not differ by more than 0.01%.

Algorithm 1 is similar to the iterative procedure proposed in [94]. The only difference is that instead of one bilinear constraint, we introduce $|\eta|$ number of bilinear constraints $\mathcal{R}_\eta(.)$, and then minimize the sum of their left-hand sides in Step 2, as suggested in [122]. The sequence of generated objective values $\left\{\mathcal{J}^k\right\}_{k\in\mathbb{N}}$ monotonically decreases and converges to a finite limit. The proof of convergence

Algorithm 1 Find a solution to (4.12)

Step 1: Set initial (feasible) values for dispatch decision variables $\mathcal{Y} = \mathcal{Y}^{\text{fixed}}$. Set $k = 1$ and $\mathcal{J}^0 = 10^{-3}$.
Step 2: Solve

$$\min_{t,\beta,w_i,x_i} \quad \sum_{\eta} \mathcal{R}_\eta \left(\mathcal{Y}^{\text{fixed}}, t, \beta, w_i, x_i\right) \tag{4.13a}$$

$$\text{s.t. } \mathcal{H}_\alpha \left(\mathcal{Y}^{\text{fixed}}, t, \beta, w_i, x_i\right) \leq 0 \ \ , \alpha = \{1, ..., |\alpha|\}, \tag{4.13b}$$

and fix variables w_i to their optimal values $w_i^{\text{fixed}} = w_i^*$.
Step 3: Solve

$$\min_{\mathcal{Y},t,\beta,x_i} \quad \mathcal{J}\left(\mathcal{Y}\right) \tag{4.14a}$$

$$\text{s.t. } \mathcal{H}_\alpha \left(\mathcal{Y}, t, \beta, w_i^{\text{fixed}}, x_i\right) \leq 0 \ \ , \alpha = \{1, ..., |\alpha|\}, \tag{4.14b}$$

$$\mathcal{R}_\eta \left(\mathcal{Y}, t, \beta, w_i^{\text{fixed}}, x_i\right) \leq 0 \ \ , \eta = \{1, ..., |\eta|\}, \tag{4.14c}$$

and fix dispatch variables to their optimal values $\mathcal{Y}^{\text{fixed}} = \mathcal{Y}^*$.
Step 4: Check if the optimal objective function $\mathcal{J}^k(\mathcal{Y})$ in (4.14a) verifies the convergence criterion $\frac{\mathcal{J}^k(\mathcal{Y}) - \mathcal{J}^{k-1}(\mathcal{Y})}{\mathcal{J}^{k-1}(\mathcal{Y})} \leq 10^{-4}$. If not, set $k \leftarrow k + 1$ and return to step 2.

is provided in [94], although the convergence to global optimality may not be achieved. In addition, since the convergence criterion stops the algorithm when two consecutive objective values do not differ more than a given threshold (i.e., 10^{-4} in our case), a local optimum might not be achieved, too. However, our extensive numerical study in the next section confirms the satisfactory performance of this algorithm in terms of optimality and convergence speed. This study shows that Algorithm 1 provides outcomes that are more reliable compared to a benchmark solution obtained by the non-linear solver IPOPT.

4.3 Numerical Experiments

As our case study, we consider a slightly updated version of the IEEE 24-node reliability test system [127] with 34 transmission lines, 12 conventional generators, 4 wind farms and 17 inelastic demands. The total conventional generation capacity is 2,362.5 MW, among which a maximum of 798 MW can be counted as reserve capacity. Furthermore, the total wind power capacity is 1,600 MW, and the total inelastic consumption is 2,207 MW. Further details about input data and a figure illustrating the network topology of this case study are available in Appendix A.

We use a dataset of 10,000 hourly wind power samples, which we split into two parts. The first one contains the arbitrarily selected 1,000 samples, which are used in our in-sample simulations. The second part embodies the remaining 9,000 samples, which are used a posteriori to assess the quality of decisions through an out-of-sample analysis. Using the in-sample data, we derive the mean forecast[5] μ and 1,000 wind power deviation samples $\hat{\xi}_i$, where $i = \{1, ..., 1000\}$. These samples allow constructing the ambiguity set. In addition, we compute the maximum and minimum deviations, i.e., $\xi^{\max}$ and $\xi^{\min}$, based on physical bounds 0 and 1 p.u. By doing so, we subtract the mean wind power generation μ to establish a rectangular support for the forecast error uncertainty that reflects the actual physical limits of uncertain parameters.

We solve the DR-OPF problem (3.10) using three different reformulation alternatives of DRCCs, namely (*i*) CVaR approximation, (*ii*) exact MILP reformulation without support, and (*iii*) our proposed support-based approach. As described in Section 4.2, the CVaR approximation incorporates the support information as well. These three problems are linear, mixed-integer linear, and bilinear programs, respectively. We run all aforementioned models with different numbers of historical observations, i.e., $N = \{50, 100\}$ samples selected from 1,000 in-sample data. We also solve these models with different risk-attitudes, i.e., $\epsilon = \{0.03, 0.05\}$, which are considered to be equal for all DRCCs. For each set of parameters, we solve three models[6] for Wasserstein radii θ ranging from 10^{-4} to 10^{-1}, where the exponent increases linearly with a step of 0.2.

Furthermore, we perform a computational analysis to explore the computational efficiency of Algorithm 1 in terms of both convergence speed and the gap achieved between the solution obtained by Algorithm 1 and the one obtained by the non-linear solver IPOPT. Although the bilinear reformulation is exact, the solution found by Algorithm 1 is not necessarily optimal. Our numerical analyses suggest that Algorithm 1 is more reliable than the non-linear solver. In addition, our results clearly show the importance of support information to be accounted for when one uses DRO to model uncertainties in the energy and reserve dispatch problem.

[5] We assume the availability of past observations. In real-life applications, the system operator would forecast the wind power generation based on available observations and run the optimization-based decision-making tool. In this work, the focus is on the optimization model. We assume that the forecast is calculated as the mean of the past observations.

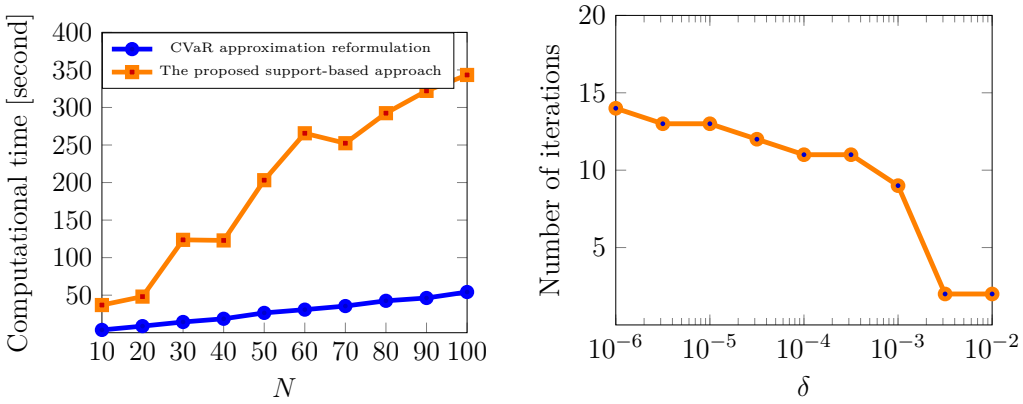
[6] It is worth mentioning that due to the high number of simulations performed in this thesis, we have omitted to report the number of variables and constraints of the underlying optimization problems to be solved. As a general rule, we have pushed the performances of our computer to their limits, where the largest admissible instances were composed of approximately 10 millions variables for linear programs, and dozens of thousands variables for MILPs.

4.3.1 Computational Performance

As the first experiment, we retrieve the evolution of computational time for each model when the number of in-sample scenarios N gradually varies from 10 to 100. Since problem (3.10) refers to a day-ahead operational problem that should typically be solved in maximum one hour in practice, e.g., in Belgian electricity market, we set a computational time limit of 1 hour. This means that we count a model computationally intractable if its computational time exceeds one hour. We realize that the exact MILP model rapidly reaches to this time limit even for a case with relatively small number of samples, e.g., $N = 20$. From now on, when a comparatively large number of scenarios is considered, the results reported for the exact MILP technique represent sub-optimal solutions obtained within the computational time limit of 1 hour. Our next simulations account for either $N = 50$ or $N = 100$ in-sample scenarios.

Given $\theta = 0.001$ and $\epsilon = 0.05$, Fig. 4.3(a) shows the computational time as a function of N for the CVaR approximation and our proposed support-based approach. We observe that the computational time increases linearly with the number of in-sample scenarios for the CVaR approximation reformulation. This appealing computational time for the CVaR approximation comes from the fact that the corresponding model is a linear program to be solved in one iteration. In the case of our proposed support-based approach, the increasing trend of computational time is non-smooth, but almost linear with a steeper slope compared to the CVaR approximation. Although our proposed reformulation has a higher computational time compared to the CVaR approximation (about 6 times higher), the results suggest that the iterative procedure of Algorithm 1 converges in a polynomial time. Fig. 4.3(b) illustrates the number of iterations required by the algorithm when the convergence criterion δ, i.e., the relative difference between the value of objective function in two consecutive iterations, increases from 10^{-6} to 10^{-2}. This figure shows that the number of iterations decreases when δ increases, as the stopping criteria becomes less strict. This decrease is less significant as the value of δ increases from 10^{-6} to 10^{-3}. However, in case of $\delta = 10^{-2}$, the algorithm converges in two iterations only.

We then study the impact of values assigned for θ and ϵ on the computational time, whose values are given in Table 4.1. We observe that the computational time increases with the number of samples N, but decreases as the value assigned for the Wasserstein radius θ increases. This is consistent with the findings of [34], showing that by increasing θ the feasible region of exactly reformulated problems could be potentially convexified.



(a) Computational time as a function of the number of in-sample scenarios N (fixed values: $\theta = 0.001$ and $\epsilon = 0.05$)

(b) Number of iterations for the iterative algorithm as a function of the stopping criteria δ (fixed values: $N = 100$, $\theta = 0.001$ and $\epsilon = 0.05$)

Figure 4.3: Computational study

In what follows, we intend to assess the gap[7] between the solution obtained by the proposed iterative algorithm and the global optimum. Recall that the underlying problem is a bilinear program which makes it non-convex, and therefore the global

[7]The optimality gap is mathematically defined as the difference between the primal best obtained solution and the dual best obtained solution. In this work, since the dual best obtained solution is not mathematically achievable, our practical alternative to assess the gap (but not necessarily the optimality gap) is to compute lower bounds on the optimal value of the objective function. We still use the term "gap" for describing the difference between our best obtained primal solution and a potential lower bound.

Table 4.1: Computational time [second]

		CVaR approximation		**The proposed support-based approach**	
		$N = 50$	$N = 100$	$N = 50$	$N = 100$
$\theta = 0.0001$	$\epsilon = 0.05$	33.3	76.8	196.4	424.3
	$\epsilon = 0.03$	29.3	74.6	352.6	464.3
$\theta = 0.001$	$\epsilon = 0.05$	26.4	54.7	203.1	343.3
	$\epsilon = 0.03$	29.7	62.6	304.4	357.4
$\theta = 0.01$	$\epsilon = 0.05$	5.2	14.3	51.6	122.4
	$\epsilon = 0.03$	5.6	15.2	50.9	189.3

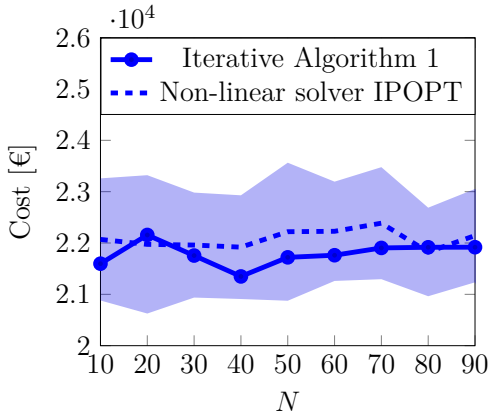
optimal point is not necessarily accessible. In general, one potential alternative to compute the optimality gap is to derive the dual problem and compare the dual best solution to our best obtained solution. However, the dual problem in our setting would not allow to find a non-trivial lower bound[8]. As the next alternative, a McCormick relaxation [130] of the bilinear terms would allow us to assess the optimality gap between the solution of the iterative algorithm and the one of the relaxed problem. However, variables w within the bilinear terms have a lower bound only, which is equal to zero, but there is no upper bound that is also required for the McCormick relaxation. We have checked the performance of McCormick relaxation by fixing the upper bound to the value given by the iterative algorithm's solution. However, by doing so, we have observed that the problem is infeasible. We hypothesize that the McCormick relaxation jeopardizes the proper displacements of samples within the physical bounds, which in turn, makes the program infeasible, as already observed in the illustrative example of Section 3.3.1.

As a pragmatic alternative, we compare the optimal value of the total expected operational cost of the system achieved from the iterative Algorithm 1 to that obtained from the non-linear solver IPOPT. This solver provides a solution which we could use as a benchmark against the solution of the iterative algorithm and assess the gap. The resulting cost profiles as a function of N are presented in Fig. 4.4(a). We run the IPOPT solver several times with a different initialization for each training sample size N in a way that the number of training samples multiplied by the number of runs stays constant[9] and equal to 200. Fig. 5(a) reports the expected value of objective function (dashed line) achieved in all these runs as well as the standard deviation (shadow area).

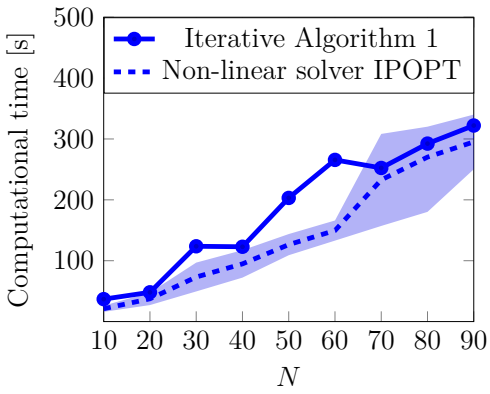
We observe that the cost profile obtained from Algorithm 1 takes lower values more frequently. In some cases, IPOPT finds a solution with a lower cost compared

[8]According to [128], the dual Lagrangian function is always concave even though the primal problem might be non-convex. Consider a set of quadratic constraints in form of $x^\top P_i x \leq d_i,\ \ \forall i \in \{1, ..., M\}$ with corresponding dual variables λ_i. In (4.12), all quadratic constraints are bilinear, meaning that P_i is not positive semidefinite. An immediate conclusion is that, $P(\lambda) = \sum_{i=1}^{M} \lambda_i P_i$ is positive semidefinite in our bilinear problem, if and only if all $\lambda_i = 0$. This refers to a very special condition under which all bilinear constraints are non-binding. In the case there is a non-zero λ_i, the dual problem gives us a trivial conclusion that the lower bound is minus infinity. In order to get a potentially non-trivial lower bound from dual problem, one has to drop *all* bilinear constraints in the primal problem. However, nearly 98% of primal constraints in (4.12) are bilinear, e.g., in the case where $N = 50$. If we drop all those constraints, the resulting problem will be highly relaxed that does not represent well the original problem (4.12). For further details on how to derive the dual problem in a bilinear program, we refer the interested reader to [129].

[9]When the theoretical number of runs, e.g., $\frac{200}{N}$, takes a non-integer value, we round it to the smallest integer greater than or equal to the theoretical number of runs.



(a) Total expected operational cost of the system as a function of the number of in-sample scenarios N, resulting from Algorithm 1 and the non-linear solver IPOPT (fixed values: $\theta = 0.001$ and $\epsilon = 0.05$)



(b) Computational time as a function of the number of in-sample scenarios N (fixed values: $\theta = 0.001$ and $\epsilon = 0.05$)

Figure 4.4: Comparison of results achieved by the iterative Algorithm 1 and the non-linear solver IPOPT. For different runs of IPOPT with different initializations, the shadow area covers the area ranging from the expected value minus the standard deviation to the expected value plus the standard deviation.

to the solution of Algorithm 1, but the gap between two solutions is relatively small. The variability around the outcome of IPOPT stems from multiple solutions achieved by IPOPT, each with a different initialization. This variability shows that this solver usually fails to provide reliable results. We retrieve the computational times and report the mean (dashed line) and the standard deviation (shadow area) in Fig. 4.4(b). The computational times in IPOPT and Algorithm 1 are comparable. Algorithm 1 slightly takes more time to find an optimal solution when $\delta = 10^{-4}$. Based on our numerical observations, we conclude that Algorithm 1 is more reliable than IPOPT, and performs well in the scope of our application.

4.3.2 In-Sample Outcomes: Dispatch Results

In this section, we present the optimal dispatch decisions $\{p_e, \overline{r}_e, \underline{r}_e, Y_e\}$ for an arbitrarily chosen conventional generator $e = 2$. For all three DRCC reformulation techniques, we compute the actual power production $p_e + Y_e\hat{\xi}_i$, which corresponds to the sum of the day-ahead power dispatch and the real-time recourse action for each historical observation $\hat{\xi}_i$ in the in-sample database. The resulting distribution is shown with blue bars in Fig. 4.5 for a case with $N = 100$, $\theta = 10^{-4}$ and $\epsilon = 0.05$.

The day-ahead energy as well as the upward and downward reserve dispatches are shown with green number and red arrows, respectively. The level of production is therefore restricted to lie within the extremity of the arrows, representing the production adjustment capability of the generator.

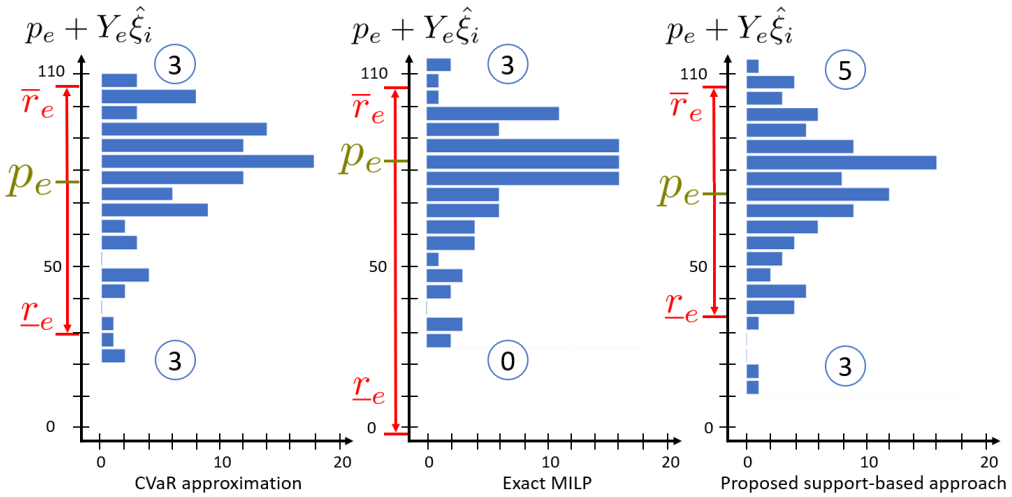


Figure 4.5: In-sample energy and reserve dispatch outcomes for conventional generator $e = 2$ in three models: Blue bars show the distribution of actual power production, i.e, the day-ahead energy dispatch p_e plus recourse action $Y_e\hat{\xi}_i$, across 100 in-sample observations $\hat{\xi}_1$, ..., $\hat{\xi}_{100}$. The energy dispatch p_e as well as the upward and downward reserve dispatches $\overline{r}_e$ and $\underline{r}_e$ are shown with green number and red arrows, respectively. The number of observations with violated downward and upward reserve constraints is given in blue circles (fixed values: $\theta = 10^{-4}$ and $\epsilon = 0.05$).

For all three reformulations, we observe that the corresponding distribution mostly takes values around the day-ahead energy dispatch p_g. However, DRCCs allow constraint violations, resulting in a total power production that lies outside the restricted zone. The number of cases with constraint violation is given in blue circles. We notice that our proposed support-based approach exhibits a higher number of cases with constraint violation, compared to the CVaR approximation. The reason for this is that the CVaR approximation is a conservative reformulation alternative and therefore the predefined level of violation probability does not necessarily occur. Finally, the exact MILP method overbooks downward reserve, which is due to the fact that the support is not included within this technique, as

explained in our illustrative example in Section 3.2. Therefore, constraints are less prone to be violated.

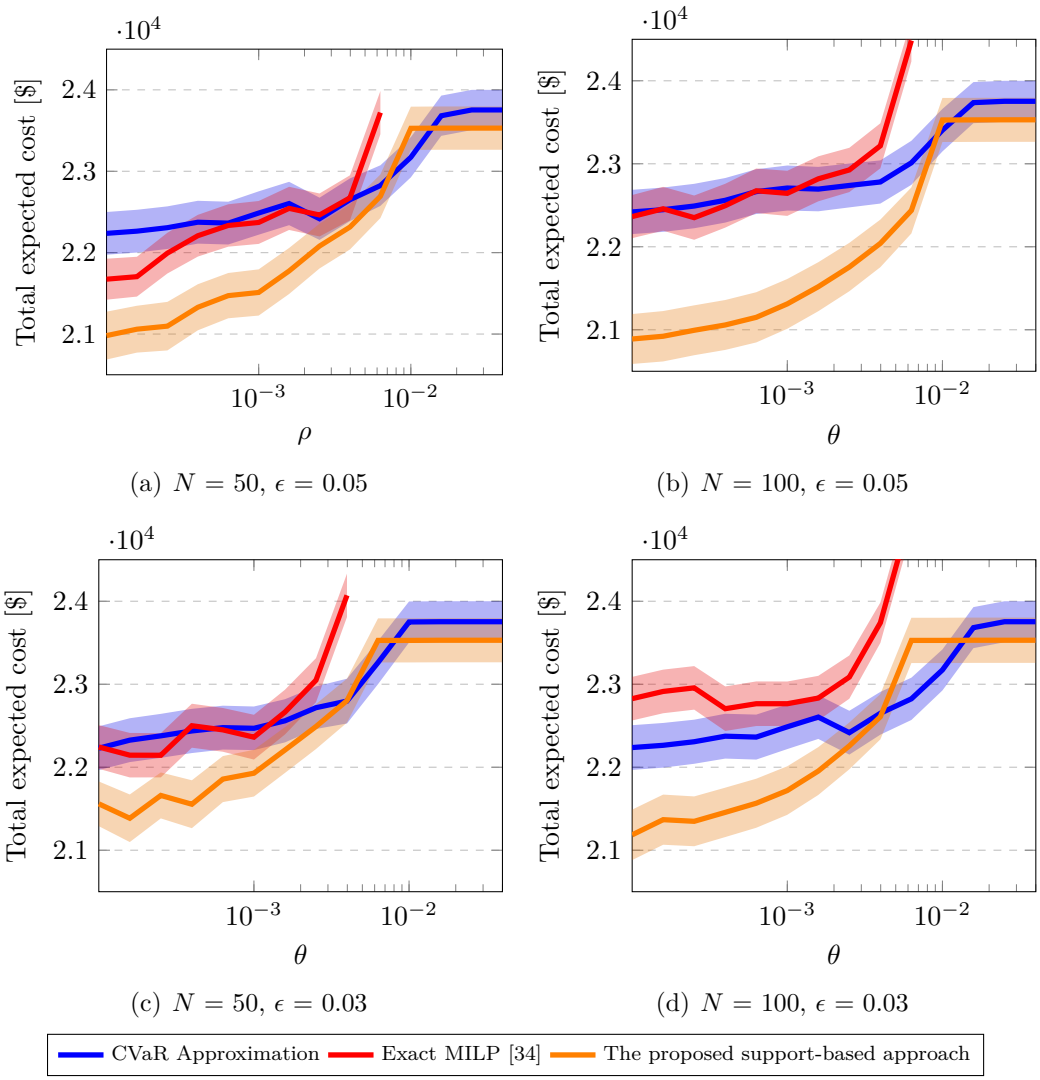
4.3.3 Out-of-Sample Outcomes: Total System Cost Without Re-Optimization

We derive the total operational cost of the system, i.e., the optimal value of objective function (3.10a), using an out-of-sample analysis as follows: we collect the optimal decision variables $\{p, \overline{r}, \underline{r}, Y\}$ obtained from each model, and calculate the optimal value of day-ahead costs $\sum_{e\in\mathcal{E}} C_e p_e + \overline{C}_e \overline{r}_e + \underline{C}_e \underline{r}_e$. For the recourse cost, we arbitrarily select first 1,000 samples from the out-of-sample data[10], say $\hat{\xi}_j$, where $j = \{1, ..., 1000\}$, and calculate the average recourse cost $10^{-3} \sum_j C^\top Y \hat{\xi}_j$. The sum of day-ahead and recourse costs gives the average of out-of-sample total system cost. We also calculate the standard deviation of this cost as a measure of variability. Here, the system operator treats the participation factors Y as informed decisions for recourse actions, and do not solve another optimization in real time under each out-of-sample realization. Consequently, the resulting cost neglects the potential cost associated with a constraint violation.

Fig. 4.6 depicts the average out-of-sample total system cost (thick curves) and the cost variability (shaded zones around the thick curves) for each model. For clarity of the figures, the shaded zones around the thick curves exhibit 10% of the standard deviation only. In most of the cases, the average cost obtained from the proposed support-based model is lower than the one achieved by the other two models. This difference is significant in some cases, e.g., in the one with $N = 100$ and $\epsilon = 0.05$. The cost achieved by the exact MILP model oscillates between the costs obtained by the other two approaches, depending on the optimality gap of the solution reached after the one-hour computational time limit. By including the support information, the average cost profile saturates for higher values of θ. Without such support information, the program either becomes infeasible, e.g., in plots 4.6(a) and 4.6(c), or optimizes for an unrealistic representation of the uncertainty.

However, it is worth noting that these results are obtained by assuming that real-time decisions are fixed in day ahead. In practice, the system operator has the opportunity to re-optimize decisions in real time to keep the balance between generation and consumption as well as to restore the feasibility by introducing

[10] In our numerical study, we have observed that the out-of-sample outcomes using either arbitrarily selected 1,000 or all 9,000 test samples are similar. For the sake of reduced computational time in our out-of-sample study, we have only picked the first 1,000 arbitrarily selected samples from the testing dataset.



(a) $N = 50$, $\epsilon = 0.05$

(b) $N = 100$, $\epsilon = 0.05$

(c) $N = 50$, $\epsilon = 0.03$

(d) $N = 100$, $\epsilon = 0.03$

Figure 4.6: Out-of-sample analysis without re-optimization: Evolution of the total average system cost and its standard deviation as a function of θ for various sets of parameters N and ϵ

extreme recourse actions such as costly load curtailment and wind spillage. The next subsection allows the system operator to re-optimize the recourse actions in real time.

4.3.4 Out-of-Sample Outcomes: Total System Cost With Re-Optimization

For a given day-ahead dispatch and under each out-of-sample realization of uncertainty $\hat{\xi}_j$, we solve a deterministic optimization problem, the so-called *re-optimization*, in real time to make decisions on the optimal recourse actions including load curtailment and wind spillage. The formulation of such an optimization problem is available in Appendix C. We take into account a cost of €500/MWh for load curtailment, while assuming a zero cost for wind spillage. There is no energy storage system in our case study to store the excess wind energy. Similar to Fig. 4.6, we report in Fig. 4.7 the average out-of-sample total system cost and its standard deviation with consideration of the re-optimization process.

We compare all three models against the Sample Average Approximation (SAA) corresponding to scenario-based stochastic programming. To apply this method, we treat the N number of in-sample observations as equiprobable scenarios and each DRCC is replaced by N number of recourse constraints, one per scenario[11]. It is worth mentioning that the SAA does not directly model the potential risk attitude of the system operator, while such an attitude is adjusted by θ and ϵ in DRO. This is the reason why the results of SAA in Fig. 4.7 are unchanged as a function of θ. Moreover, one can interpret the DRO with a large value of θ as a robust counterpart, in the sense that extreme realizations of the uncertainty are considered.

These results suggest that DRO outperforms SAA when the number of training samples N is comparatively low. This is the case where there is the lack of sufficient information about uncertainty or the number of training samples is intentionally kept below due to computational issues. To the best of our knowledge, the reserve market in Switzerland [8] and in Sweden [9] are the sole electricity markets in practice with a stochastic framework, which takes into account a very limited number of scenarios. In such a market with limited information about uncertain parameters, we hypothesize a DRO model has potential to outperform other techniques, including the scenario-based stochastic programming.

One can draw three additional important observations from Fig. 4.7. First, similar to Fig. 4.6, our proposed support-based approach provides in general

[11] Contrary to this approach, the chance-constrained SAA models the constraints as classical chance constraints. In view of Remark 1 in [34], the latter technique may not be achieved by setting $\theta = 0$. In addition, chance-constrained models are known to be computationally expensive and require analytical reformulations, which remain outside the scope of this paper.

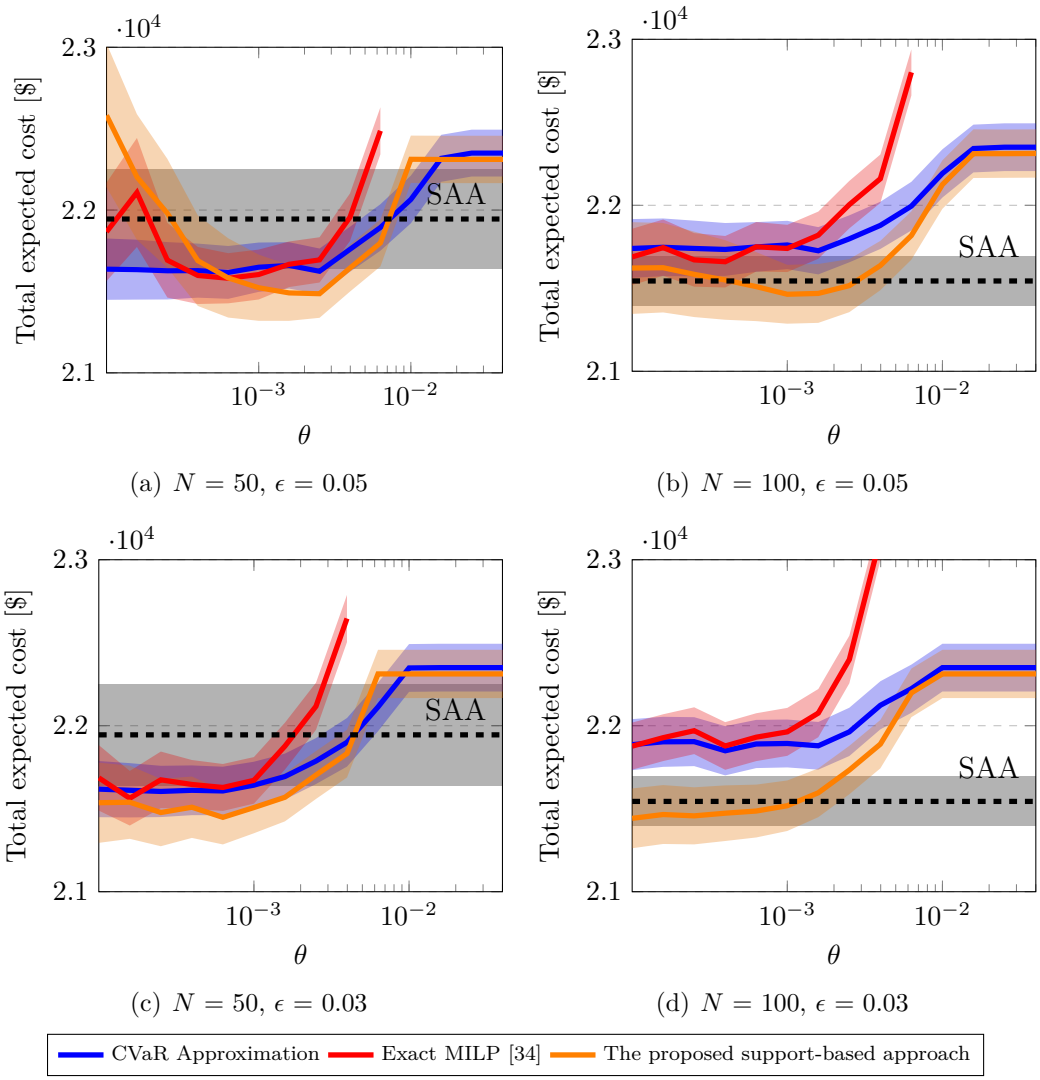
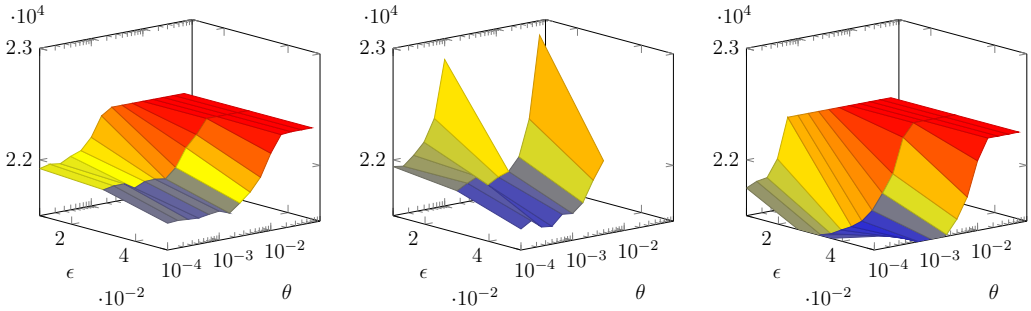


Figure 4.7: Out-of-sample analysis with re-optimization: Evolution of the total average system cost and its standard deviation as a function of θ for various set of parameters N and ϵ

better results in terms of average total operational cost of the system than the CVaR approximation and the exact MILP method without support, while the cost variability in all these three models are similar. Second, there exist several values for θ under which the proposed DRO model outperforms the SAA method. For example, in Fig. 4.7(d), the proposed DRO for a range of θ from 10^{-4} to around 10^{-3} obtains a lower average cost than the SAA. Third, the proposed DRO provides a flexibility for the decision-maker to easily adjust θ in such a way that it achieves either a solution better or similar to SAA, or a solution close to

the robust one. For example, the value of θ in Fig. 4.7(d) between 10^{-3} to 10^{-2} provides a solution that is lying between SAA and robust solutions. For values larger than 10^{-2}, the proposed DRO provides fully robust solutions.

Next, we provide three-dimensional plots of the out-of-sample average total operational cost of the system as a function of θ and ϵ, which are presented in Figures 4.14a to 4.8(c). These results suggest that the total operational cost decreases when ϵ increases. This confirms our initial intuition that the chance-constrained programming allows to reduce the total operational cost of the system by allowing constraint violation to some extent. This implies that the distributionally robust chance-constrained program finds an optimal trade-off between operational cost and reliability. These figures further show that the exact MILP model becomes infeasible when the value assigned for the constraint violation probability ϵ is comparatively low.



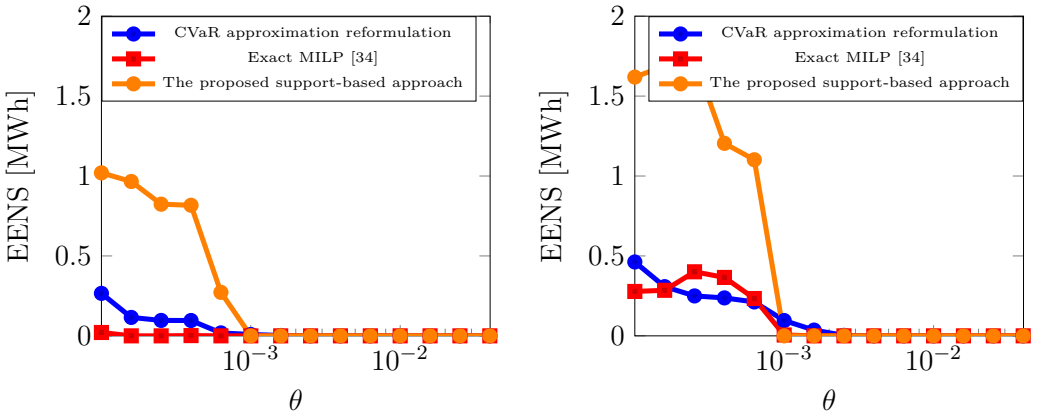
(a) CVaR Approximation (b) Exact MILP (c) The proposed support-based approach

Figure 4.8: The average total operational cost of the system with re-optimization as a function of θ and ϵ.

We emphasize that the optimal ex-ante selection of radius θ is not straightforward and may require learning from experience, i.e., past outcome of the DRO model. We have observed that there exists an optimal value for θ that leads to the minimum out-of-sample expected operational cost. Accordingly, we suggest that the decision-maker may select a value for θ ranging from the value that provides the best out-of-sample performance to infinity, leading to similar decisions to those of robust optimization. By doing so, the decision-maker is able to adapt her risk attitude. In general, choosing the value of θ that correctly mimics the risk attitude of the decision-maker is challenging and is closely dependent on her preferences in terms of cost optimality, operational reliability, computational performance, etc.

4.3.5 Expected Energy Not Served

In this subsection, we present additional numerical results in terms of the Expected Energy Not Served (EENS). This indicator shows how much load (in MWh) is expected to be curtailed. It also illustrates the violation severity of upward reserve constraints. When the upward reserve is not sufficient, the system operator loses her ability to deliver energy to the consumer and therefore load curtailment occurs. We calculate the EENS for each model when $\epsilon = \{0.03, 0.05\}$ and for Wasserstein radii ranging from 10^{-4} to 10^{-1}. The results are given in Figure 4.9.



(a) $\epsilon = 0.03$ (b) $\epsilon = 0.05$

Figure 4.9: Expected energy not served (EENS) as a function of θ (fixed value: $N = 100$)

We observe that EENS tends to decrease when θ increases as shown in Figure 4.9. These results reflect the relation that exists between θ and ϵ, which respectively set the distributional robustness of the optimal decisions and the maximum allowed violation probability of a given constraint. Even though their interpretation is different, those parameters are closely connected in the way they influence the scheduling decisions. For instance, a given violation probability ϵ will not result in the same empirical violation probabilities for two different distributions within the ambiguity set, e.g., the empirical one and the worst-case one. When the distributional robustness increases, i.e., θ takes a comparatively high value, the EENS decreases for a given violation probability ϵ. The reason for this is that the worst-case distribution in the ambiguity set (for which the decisions are optimized) results in more conservative solutions when θ increases.

Finally, it is worth mentioning that the overall system reliability violation, i.e.,

the probability of violating at least one constraint, always reaches to 100%. This naive result comes from the fact that the allowed *individual* violation probability is comparatively high, e.g., 3% or 5% in our case study, with respect to the overall system. Our numerical case study contains 58 individual chance constraints, which means that the overall violation probability would evaluate to 100% if a Bonferroni approximation would be used. Hence, even if the overall violation probability is comparatively high, the magnitude of violation remains acceptable. To support our claim, we observe that the EENS never takes a value higher than 1.8 MWh per hour. We numerically conclude that the reliability of our approach based on individual chance constraints is satisfactory from the system operator's perspective.

4.4 Chapter Summary and Conclusions

In this chapter, we develop a Wasserstein distributionally robust chance-constrained OPF model, and provide three different reformulations: (*i*) a CVaR approximation resulting in a linear program, (*ii*) an exact reformulation without support, yielding an MILP, and (*iii*) a physically-bounded exact reformulation. The last reformulation is proposed in this work for retrieving realistic and optimal solutions with respect to the uncertainty space, which results in a bilinear program. This problem is solved using an efficient iterative algorithm, and its computational performance is compared with that of a non-linear solver, IPOPT. Through an extensive out-of-sample study, we show that our proposed reformulation enables the system operator to make more informed dispatch decisions by including physical bounds within the exact reformulation of DRCCs. We observe that our proposed support-based method outperforms other techniques specially in cases wherein the value for θ is properly selected.

In the next chapter, we aim at addressing the second limitation of the state-of-the-art DRO (i.e., the accommodation of dependence structure within the definition of ambiguity set) as described in Section 3.3.2. Two different definitions of ambiguity set are proposed, namely the moment-metric-based ambiguity set and the copula-based one, which offer different degrees of accuracy regarding the modeling of dependence structure.

CHAPTER 5

Embedding Dependencies within Distributionally Robust OPF

To cope with the uncertainty stemming from weather-dependent renewable energy sources, the traditional metric-based distributionally robust approach includes the misrepresentations of distributional information (e.g., from probabilistic forecasts) within an ambiguity set, which describes the potential deviations from an empirical distribution of all uncertain parameters. However, this set typically overlooks the inherent dependencies of uncertainty, e.g., spatial dependencies of weather-dependent energy sources. In this chapter, we aim at further refining the modeling of ambiguity sets within the metric-based distributionally robust optimal power flow problems. In particular, we go beyond the state of the art by embedding such dependencies within the definition of the ambiguity set, in two different manners. We propose *i)* an enhanced moment-metric-based ambiguity set which includes a constraint on the second-order moment of uncertainty (i.e., the covariance matrix), and *ii)* a copula-based one which binds the Wasserstein distance to the empirical copula and is tailored to capture any kind of dependencies. We provide generic reformulations for both approaches such that the different outcomes could be applied to any decision-making problem under uncertainty in power systems. Finally, we illustrate the performance of the proposed models applied to an OPF problem in a power systems with high share of renewables.

The content of this chapter is mainly based on the following publications[1] :

[D] **A. Arrigo**, J. Kazempour, Z. De Grève, J.-F. Toubeau, and F. Vallée, "Embedding Dependencies Between Wind Farms in Uncertainty-Aware Optimal

[1]The IEEE permission grant is available at `https://journals.ieeeauthorcenter.ieee.org/choose-a-publishing-agreement/avoid-infringement-upon-ieee-copyright`, Accessed on 26$^{\text{th}}$ April 2022.

Power Flow," 2022, submitted to IEEE Transactions on Power Systems, under the second round of review.

[C] **A. Arrigo**, J. Kazempour, Z. De Grève, J.-F. Toubeau, and F. Vallée, "Enhanced Wasserstein Distributionally Robust OPF With Dependence Structure and Support Information," IEEE PowerTech Conference, Madrid, June 2021.

5.1 Motivation

The increasing share of weather-dependent renewable energy sources connected to the power transmission system induces uncertainty and variability in the generation-side. To address this uncertainty, it is crucial to improve the current variants of OPF problem by incorporating probabilistic forecast information (e.g., via a probability distribution function). However, the uncertainty stemming from renewable power generation exhibits multi-dimensional space-time dependencies, and it is in general difficult to represent this uncertainty by a unique distribution function. To raise this issue, we focus on metric-based DRO which considers a family of potential distributions, i.e., the so-called ambiguity set, to hedge against the inexactness of input distribution.

In the traditional Wasserstein DRO approach, all distributions in the neighborhood of a central empirical distribution, based on historically observed samples of uncertainty, are collected and fed to the decision-making problem. Although this framework accounts for potential forecast errors, there is no guarantee that the distributions within the ambiguity set remain realistic (e.g., from a correlation perspective). Therefore, in the current literature, there is a willingness to enhance the Wasserstein ambiguity set by removing the unrealistic distributions of renewable power generation uncertainty. To do so, supplementary constraints should be incorporated into the design of the ambiguity set. In the same vein as our work in Chapter 4, reference [20] includes the support information (i.e., lower and upper bounds for the renewable power generation) in the definition of Wasserstein ambiguity set, excluding distributions with unrealistic realizations of uncertainty. An example of such unrealistic distributions is those with a negative renewable power generation. References [30] and [31] include modality information, i.e., the number of spikes in a probability distribution, within the ambiguity set to get rid of unrealistic distributions, e.g., those with two or more spikes. An additional potential information to be included within the definition of ambiguity set is the correlation among uncertain parameters [33]. As our main contribution in this chapter, we go beyond the current state-of-the-art by developing two different DRO approaches, namely the *moment-metric-based* and the *copula-based*, that

are aware of the potential dependence structure among uncertain parameters, in particular the spatial dependencies of weather-dependent renewable energy sources in power systems.

The moment-metric-based approach [33] combines the benefits of the metric-based DRO with a constraint imposing the value of second-order moment. The resulting ambiguity set therefore includes all distributions in the neighborhood of a central empirical distribution with a fixed covariance matrix. By doing so, the ambiguity set shrinks to distributions with a dependence structure defined by the given covariance matrix. In this thesis, building up on [32, 33], we develop an enhanced moment-metric-based distributionally robust OPF model by incorporating both support information and dependence structure of random variables. Although the proposed moment-metric-based approach aims at embedding the dependencies within the definition of ambiguity set, it is worth mentioning that the covariance matrix may only capture linear relations and may fail in describing more complex dependence structure. Therefore, we have thrived to explore methodologies that could not only capture the linear, but also more complex dependence structure.

Several additional works in the literature have contributed towards embedding the full non-linear dependence information into the distributionally robust optimization. However, it is worth emphasizing that there are currently very few contributions in this research strand, especially focusing on power system applications. References [32, 131] envisage the use of *copula* as the mathematical object to describe the full dependence structure among random variables. A copula is a distribution function with uniformly distributed marginals that contains solely information about the dependence structure among uncertain parameters [132, 133]. To incorporate the dependence structure into the optimization problem, [32, 131] introduce a constraint binding the Wasserstein distance between the copula of the distributions inside the ambiguity set and the one of the empirical distribution.

In this thesis, inspired by [32], we propose a copula-based ambiguity set as the framework that enables the DRO problem to consider *any* type of correlation (not necessarily linear) among uncertain parameters. To do so, we combine the findings in [32] with the classical DRO approach, to produce a Wasserstein ambiguity set that uses the copula to capture the dependence structure among renewable power generators. By this, we are able to properly model the renewable power generation uncertainty that may exhibit any shape of dependencies [134].

For the reader convenience, we report in Table 5.1 the relevant features of works in the literature, which focus on improving the traditional DRO, in particular, by embedding the dependence structure, support information and unimodality

Table 5.1: Relevant features of works reported in the literature and the model proposed in this paper

Reference	Type of ambiguity set	Dependence structure	Unimodality information	Support information	Resulting model
[24]	Moment	✗	✗	✗	LP
[25]	Moment	✗	✗	✗	QCQP
[23]	Moment	✗	✓	✗	SOCP
[30], [31]	Metric	✗	✓	✗	SOCP
[20], [95], [136]	Metric	✗	✗	✓	LP
[32], [33], [137]	Moment-metric	✓ (a)	✗	✗	SDP
The enhanced moment-metric-based approach	Moment-metric	✓ (a)	✗	✓	SDP
[138], [139], [131]	Copula-based	✓ (b)	✗	✗	Conic
[32]	Copula-based	✓ (c)	✗	✓	Conic
The copula-based approach	Copula-based	✓ (c)	✗	✓	Conic

(a) The moment-metric-based approach captures linear correlations.
(b) These works can capture non-linear dependence structure, with fixed marginal distributions.
(c) These works can capture the full non-linear dependence structure.

information. To the best of our knowledge, the work presented in this chapter qualifies among the first efforts in the literature to consider the dependencies within DRO in power systems. In particular,

(i) We develop a generic SDP reformulation of the worst-case expectation problem[2] using the enhanced moment-metric-based ambiguity set, incorporating both support information and dependence structure of random variables

(ii) We develop a conic reformulation of a distributionally robust worst-case expectation problem using the proposed copula-based ambiguity set.

The derivation of these reformulations is kept generic, such that they could be applied to a broad range of decision-making problems under uncertainty in power systems. In the following, we recall the DR-OPF problem as the application considered in this chapter. Next, we introduce the moment-metric-based ambiguity set in Section 5.2, and the copula-based one in Section 5.3. In Section 5.4, we apply the proposed approaches and numerically show their benefits compared to a state-of-the-art distributionally robust model, while identifying paths for future research.

[2]The worst-case expectation problem is a well-known problem in the field of DRO [135], please see Section 3.1. Its generic reformulation usually facilitates the application of DRO to any kind of problem where it naturally appears.

5.2 The Moment-Metric-Based DR-OPF Model

In this section, we introduce the moment-metric-based distributionally robust OPF problem. In particular, we introduce the mathematical definitions regarding the moment-metric-based ambiguity set in Section 5.2.1. The reformulation of a generic worst-case expectation problem is proposed in Section 5.2.2 and the resulting distributionally robust OPF problem, with the moment-metric-based ambiguity set is given in Section 5.2.3.

5.2.1 Definition of the Moment-Metric-Based Ambiguity Set

The moment-metric-based ambiguity set, as defined in [32], is described by

$$\mathcal{A}_2 = \left\{ \mathbb{Q} \in \mathcal{M} \;\middle|\; \begin{array}{l} d_W\left(\mathbb{Q}, \widehat{\mathbb{Q}}_N\right) \leq \theta_1 \\ \mathbb{E}^{\mathbb{Q}}\left[\left(\tilde{\xi} - \mu_0\right)\left(\tilde{\xi} - \mu_0\right)^\top\right] \preceq \Sigma \end{array} \right\}, \tag{5.1}$$

which contains only those distributions within the previously defined set $\mathcal{A}_1$ that comply with the covariance matrix of the empirical distribution. The newly added constraint in (5.1) ensures that all distributions within $\mathcal{A}_2$ follow the empirical covariance matrix Σ. Note that μ_0 represents the empirical mean. Thereby, set $\mathcal{A}_2$ contains the distributions that are close to the empirical distribution $\hat{\mathbb{Q}}_N$ and satisfy the linear dependence structure expressed in Σ. This additional feature enables eliminating the unrealistic distributions from the ambiguity set $\mathcal{A}_1$, i.e., those that do not have a dependence structure which is consistent with the empirical observations of renewable power generation uncertainty.

In this thesis, we propose an enhanced moment-metric-based ambiguity set, following a similar rationale as in Chapter 4, by including the physical bounds of uncertainty within the definition of ambiguity set. The proposed ambiguity set is described by

$$\mathcal{A}_3 = \left\{ \mathbb{Q} \in \mathcal{M} \;\middle|\; \begin{array}{l} d_W\left(\mathbb{Q}, \widehat{\mathbb{Q}}_N\right) \leq \theta_1 \\ \mathbb{E}^{\mathbb{Q}}\left[\left(\tilde{\xi} - \mu_0\right)\left(\tilde{\xi} - \mu_0\right)^\top\right] \preceq \Sigma \\ \mathbb{Q}\left(\tilde{\xi} \in \mathcal{U}\right) = 1 \end{array} \right\}, \tag{5.2}$$

which contains only those distributions within the previously defined set $\mathcal{A}_2$ that satisfy an additional constraint including the support information. This constraint restricts the probability distribution $\mathbb{Q}$ to take values strictly and only within the

support $\mathcal{U}$. The support of uncertainty may correspond to physical bounds of the uncertainty, e.g., minimum and maximum renewable power generation, or to probabilistic bounds, e.g., quantile regression achieved by forecasting methods. By adding this support information, we restrict the distributions to take realistic values, thereby further eliminating unrealistic distributions from the ambiguity set. For instance, the distributions in sets $\mathcal{A}_1$ and $\mathcal{A}_2$ may take non-realistic values, e.g., negative renewable power generation, which will lead to sub-optimality or infeasibility of the resulting program for large values of θ_1. To add the support information, we consider an ellipsoidal support $\mathcal{U}$, where the random vector has to satisfy a constraint as

$$\begin{pmatrix}\widetilde{\xi} & 1\end{pmatrix} M \begin{pmatrix}\widetilde{\xi}\\ 1\end{pmatrix} \leq 0, \text{where } M = \begin{pmatrix}\Sigma_0 & -\Sigma_0\xi_0\\ -\xi_0^\top\Sigma_0 & \xi_0^\top\Sigma_0\xi_0 - 1\end{pmatrix}. \tag{5.3}$$

Note that $\xi_0 \in \mathbb{R}^{|\mathcal{K}|}$ corresponds to the center of the ellipsoid, and $\Sigma_0 \in \mathbb{R}^{|\mathcal{K}|\times|\mathcal{K}|}$ parametrizes the span of the ellipsoid.

5.2.2 Reformulation of the Worst-Case Expectation Problem

By design, DRO aims to determine the *worst-case* distribution within the given ambiguity set, and makes decisions in *expectation* with respect to such a worst-case distribution. This section introduces a generic distributionally robust worst-case expectation problem, which will be the focus of the reformulations derived in this chapter. This problem writes as

$$\min_{x\in\mathcal{X}} \overbrace{\max_{\mathbb{Q}\in\mathcal{A}_3} \mathbb{E}^{\mathbb{Q}}\left[a\left(x\right)^\top \widetilde{\xi} + b\left(x\right)\right]}^{\text{Worst-case expected cost}}, \tag{5.4}$$

where $x \in \mathcal{X}$ is the vector of decision variables. The inner maximization operator in (5.4) picks the worst-case distribution $\mathbb{Q}$ in the ambiguity set $\mathcal{A}_3$. The probability distribution of the uncertain parameter $\widetilde{\xi} \in \mathbb{R}^{|\mathcal{K}|}$ is $\mathbb{Q}$. The objective function (5.4) is linear[3], comprising of the decision-dependent vector $a(x) \in \mathbb{R}^{|\mathcal{K}|}$ and the decision-dependent scalar $b(x) \in \mathbb{R}$.

The worst-case expectation problem under the moment-metric-based ambiguity

[3]For the sake of simplicity, we assume linearity of the objective function. This assumption is aligned with the current practice of electricity markets with linear bids. An extension to a non-linear objective function is straightforward, but requires additional reformulations that are out of the scope of this work.

set $\mathcal{A}_2$, as proposed in [32], writes as

$$\max_{\mathbb{Q}\in\mathcal{A}_2} \ \mathbb{E}^{\mathbb{Q}}\left[a\left(x\right)^{\top}\xi + b\left(x\right)\right] = \tag{5.5a}$$

$$\begin{cases} \min\limits_{\lambda,\sigma_i,\Lambda\succeq 0,\zeta_i} \ \lambda\theta_1 + \langle\Lambda,\Sigma\rangle_{\mathrm{F}} + \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\sigma_i \\ \text{s.t. } \ \mathcal{F}_i \succeq 0 \ \ \forall i \in \{1,...,N\} \\ \quad\quad ||\zeta_i||_* \leq \lambda \ \ \forall i \in \{1,...,N\}\,, \end{cases} \tag{5.5b}$$

where $\lambda \in \mathbb{R}$, $\sigma \in \mathbb{R}^N$, $\zeta_i \in \mathbb{R}^{|\mathcal{W}|}$ and $\Lambda \in \mathbb{R}^{|\mathcal{K}|\times|\mathcal{K}|}$ are auxiliary variables. Operator $\langle\Lambda,\Sigma\rangle_{\mathrm{F}} = \sum\limits_{m=1}^{|\mathcal{K}|}\sum\limits_{n=1}^{|\mathcal{K}|}\Lambda_{mn}\cdot\Sigma_{mn}$ refers to the Frobenius inner product between matrices Λ and Σ. In addition, operator $||.||_*$ refers to the dual norm, which is defined as $||v||_* = \max\limits_{||b||\leq 1} v^{\top}b$. Problem (5.5) is a Semi-Definite Program (SDP), where a constraint involving the operator $\succeq$ imposes that the matrix on the left-hand side is semi-definite positive. We define matrix $\mathcal{F}_i \in \mathbb{R}^{|\mathcal{K}|+1\times|\mathcal{K}|+1}$ as

$$\mathcal{F}_i = \begin{bmatrix} \Lambda & -\frac{1}{2}a\left(x\right) + \frac{1}{2}\zeta_i - \Lambda\mu_0 \\ \left(-\frac{1}{2}a\left(x\right) + \frac{1}{2}\zeta_i - \Lambda\mu_0\right)^{\top} & \begin{array}{r}\sigma_i - \zeta_i^{\top}\hat{\xi}_i \\ - b\left(x\right) + \mu_0^{\top}\Lambda\mu_0\end{array} \end{bmatrix}. \tag{5.6}$$

Following the formulation (5.5), we now introduce the worst-case expectation problem under the enhanced moment-metric-based ambiguity set $\mathcal{A}_3$.

Theorem 5.1. *The worst-case expectation problem under ambiguity set $\mathcal{A}_3$ described by* (5.2) *with an ellipsoidal support of uncertainty $\mathcal{U}$ given by* (5.3) *is equivalent to*

$$\max_{\mathbb{Q}\in\mathcal{A}_3} \ \mathbb{E}^{\mathbb{Q}}\left[a\left(x\right)^{\top}\xi + b\left(x\right)\right] = \tag{5.7a}$$

$$\begin{cases} \min\limits_{\lambda,\sigma_i,\Lambda\succeq 0,\zeta_i,\beta_i\geq 0} \ \lambda\theta_1 + \langle\Lambda,\Sigma\rangle_{F} + \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\sigma_i \\ s.t. \ \mathcal{F}_i \succeq -\beta_i M \ \ \forall i \in \{1,...,N\} \\ \quad\quad ||\zeta_i||_* \leq \lambda \ \ \forall i \in \{1,...,N\}\,. \end{cases} \tag{5.7b}$$

where $\beta \in \mathbb{R}^N$ is an additional auxiliary variable and $M \in \mathbb{R}^{|\mathcal{K}+1|\times|\mathcal{K}+1|}$ is defined by (5.3).

Proof. We depart from Theorem 1 in [32], which is equivalently stated for discrete

empirical distributions as

$$\max_{\mathbb{Q}\in\mathcal{A}_3}\mathbb{E}^{\mathbb{Q}}\left[a\left(x\right)^{\top}\xi+b\left(x\right)\right]= \tag{5.8a}$$

$$\begin{cases}\min\limits_{\lambda,\sigma_i,\Lambda\succeq 0}\ \lambda\theta_1+\langle\Lambda,\Sigma\rangle_{\mathrm{F}}+\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\sigma_i\\ \text{s.t. }\max\limits_{\xi\in\mathcal{U}}\ a\left(x\right)^{\top}\xi+b\left(x\right)-\lambda||\xi-\hat{\xi}_i||\\ \qquad -\left(\xi-\mu_0\right)^{\top}\Lambda\left(\xi-\mu_0\right)\leq\sigma_i\ \ \forall i\in\{1,...,N\}\,.\end{cases} \tag{5.8b}$$

Constraint in (5.8b) requires further reformulations. First, we use the dual norm to recast the norm $||\xi-\hat{\xi}_i||$, yielding

$$\begin{aligned}&\max_{\xi\in\mathcal{U}}\ a\left(x\right)^{\top}\xi+b\left(x\right)-\lambda\max_{||\zeta_i||_*\leq 1}\zeta_i^{\top}\left(\xi-\hat{\xi}_i\right)\\ &\qquad -\left(\xi-\mu_0\right)^{\top}\Lambda\left(\xi-\mu_0\right)\leq\sigma_i\ \ \forall i\in\{1,...,N\}\,.\end{aligned} \tag{5.9}$$

The inner maximization operator is brought back to the left with a change in the optimization sense (i.e., $-\max=\min-$). Next, we permute the order of the optimization operators to result in a min-max structure; the permutation is allowed since the optimization sets are convex and independent. For the sake of simplicity, we realize a change of variable $\lambda\zeta_i\longmapsto\zeta_i$ and drop the min operator inside the "less or equal to" constraint, adding the variable ζ_i to the decision variable, which allows us to equivalently reformulate the constraint (5.9) as

$$\begin{aligned}&\max_{\xi\in\mathcal{U}}\ a\left(x\right)^{\top}\xi+b\left(x\right)-\zeta_i^{\top}\left(\xi-\hat{\xi}_i\right)\\ &\qquad -\left(\xi-\mu_0\right)^{\top}\Lambda\left(\xi-\mu_0\right)\leq\sigma_i\ \ \forall i\in\{1,...,N\}\end{aligned} \tag{5.10a}$$

$$||\zeta_i||_*\leq\lambda. \tag{5.10b}$$

Constraint (5.10a) corresponds to a quadratically-constrained quadratic program (QCQP) when the uncertainty set $\mathcal{U}$ is ellipsoidal and described by quadratic equations (5.3). By means of S-Lemma [140], this QCQP can be equivalently reformulated as a SDP

$$\mathcal{F}_i\succeq-\beta_i M\ \ \forall i\in\{1,...,N\} \tag{5.11a}$$

$$||\zeta_i||_*\leq\lambda\ \ \forall i\in\{1,...,N\} \tag{5.11b}$$

$$\beta_i\geq 0\ \ \forall i\in\{1,...,N\}\,, \tag{5.11c}$$

where $\beta_i\in\mathbb{R}$ is an auxiliary variable and $\mathcal{F}_i$ is defined by (5.6). Constraint (5.8b) can therefore be equivalently recast as (5.11). This completes the proof. ■

The outcome of Theorem 5.1 can be used to reformulate the worst-case expectation in the objective function (3.10a) of the moment-metric-based DR-OPF problem, as well as the worst-case expectation appearing in the probabilistic constraints (when using the CVaR approximation to deal with DRCCs as described in Section 3.2.2). Following this procedure, we end up in a final tractable SDP formulation, which is introduced in the next section.

5.2.3 Final Model

Using Theorem 5.1, we reformulate the distributionally robust OPF problem (3.10) under ambiguity set $\mathcal{A}_3$. The resulting model is a SDP and is as follows:

$$\min_{\substack{p,\overline{r},\underline{r},B \\ y_i,\zeta_i,\lambda\geq 0,\Lambda\succeq 0}} \quad \sum_{e\in\mathcal{E}} C_e p_e + \overline{C}_e \overline{r}_e + \underline{C}_e \underline{r}_e + \lambda\theta_1 + \langle \Lambda, \Sigma\rangle_{\mathrm{F}} + \frac{1}{N}\sum_{i=1}^{N} y_i \tag{5.12a}$$

$$\text{s.t. (3.10b) - (3.10g)}, \tag{5.12b}$$

$$\begin{bmatrix} \Lambda & -\frac{1}{2}Y^\top C + \frac{1}{2}\zeta_i \\ \left(-\frac{1}{2}Y^\top C + \frac{1}{2}\zeta_i\right)^\top & -\zeta_i^\top \hat{\xi}_i + y_i \end{bmatrix} \succeq -\sigma_i M \ \ \forall i \in \{1,...,N\} \tag{5.12c}$$

$$||\zeta_i||_* \leq \lambda \ \ \forall i \in \{1,...,N\} \tag{5.12d}$$

$$\underline{\tau}_e + \frac{1}{\underline{\epsilon}_e}\left(\underline{\lambda}_e\theta_1 + \langle \underline{\Lambda}_e, \Sigma\rangle_{\mathrm{F}} + \frac{1}{N}\sum_{i=1}^{N} \underline{y}_{e,i}\right) \leq 0 \quad \forall e \in \mathcal{E} \tag{5.12e}$$

$$\begin{bmatrix} \underline{\Lambda}_e & \frac{1}{2}Y_e + \frac{1}{2}\underline{\zeta}_{e,i,1} \\ \left(\frac{1}{2}Y_e + \frac{1}{2}\underline{\zeta}_{e,i,1}\right)^\top & \underline{\tau}_e + \underline{r}_e - \underline{\zeta}_{e,i,1}^\top \hat{\xi}_i + \underline{y}_{e,i} \end{bmatrix} \succeq -\underline{\sigma}_{e,i,1} M \ \ \forall i \in \{1,...,N\} \ \ \forall e \in \mathcal{E} \tag{5.12f}$$

$$\begin{bmatrix} \underline{\Lambda}_e & \frac{1}{2}\underline{\zeta}_{e,i,2} \\ \left(\frac{1}{2}\underline{\zeta}_{e,i,2}\right)^\top & -\underline{\zeta}_{e,i,2}^\top \hat{\xi}_i + \underline{y}_{e,i} \end{bmatrix} \succeq -\underline{\sigma}_{e,i,2} M \ \ \forall i \in \{1,...,N\} \ \ \forall e \in \mathcal{E} \tag{5.12g}$$

$$||\underline{\zeta}_{e,i,j}||_* \leq \underline{\lambda}_e \ \ \forall j \in \mathcal{J} \ \ \forall i \in \{1,...,N\} \ \ \forall e \in \mathcal{E} \tag{5.12h}$$

$$\overline{\tau}_e + \frac{1}{\overline{\epsilon}_e}\left(\overline{\lambda}_e\theta_1 + \left\langle \overline{\Lambda}_e, \Sigma\right\rangle_{\mathrm{F}} + \frac{1}{N}\sum_{i=1}^{N} \overline{y}_{e,i}\right) \leq 0 \quad \forall e \in \mathcal{E} \tag{5.12i}$$

$$\begin{bmatrix} \overline{\Lambda}_e & -\frac{1}{2}Y_e + \frac{1}{2}\overline{\zeta}_{e,i,1} \\ \left(-\frac{1}{2}Y_e + \frac{1}{2}\overline{\zeta}_{e,i,1}\right)^\top & \overline{\tau}_e + \overline{r}_e - \overline{\zeta}_{e,i,1}^\top \hat{\xi}_i + \overline{y}_{e,i} \end{bmatrix} \succeq -\overline{\sigma}_{e,i,1} M \ \ \forall i \in \{1,...,N\} \ \ \forall e \in \mathcal{E} \tag{5.12j}$$

$$\begin{bmatrix} \overline{\Lambda}_e & \frac{1}{2}\overline{\zeta}_{e,i,2} \\ \left(\frac{1}{2}\overline{\zeta}_{e,i,2}\right)^\top & -\overline{\zeta}_{e,i,2}^\top \hat{\xi}_i + \overline{y}_{e,i} \end{bmatrix} \succeq -\overline{\sigma}_{e,i,2} M \ \ \forall i \in \{1,...,N\} \ \ \forall e \in \mathcal{E} \tag{5.12k}$$

$$||\overline{\zeta}_{e,i,j}||_* \leq \overline{\lambda}_e \ \ \forall j \in \mathcal{J} \ \ \forall i \in \{1,...,N\} \ \ \forall e \in \mathcal{E} \tag{5.12l}$$

$$\tau_\ell + \frac{1}{\epsilon_\ell}\left(\lambda_\ell\theta_1 + \langle \Lambda_\ell, \Sigma\rangle_{\mathrm{F}} + \frac{1}{N}\sum_{i=1}^{N} y_{\ell,i}\right) \leq 0 \quad \forall \ell \in \mathcal{L} \tag{5.12m}$$

$$P_\ell^{\text{inj}} = M_\ell^{\mathcal{E}} p + M_\ell^{\mathcal{K}} W\mu - M_\ell^{\mathcal{D}} p_d \;\; \forall \ell \in \mathcal{L} \tag{5.12n}$$

$$\breve{a}_\ell = M_\ell^{\mathcal{E}} Y + M_\ell^{\mathcal{K}} W \;\; \forall \ell \in \mathcal{L} \tag{5.12o}$$

$$\begin{bmatrix} \Lambda_\ell & -\frac{1}{2}\breve{a}_\ell + \frac{1}{2}\zeta_{\ell,i,1} \\ \left(-\frac{1}{2}\breve{a}_\ell + \frac{1}{2}\zeta_{\ell,i,1}\right)^\top & \tau_\ell + f_\ell^{\max} - P_\ell^{\text{inj}} - \zeta_{\ell,i,1}^\top \hat{\xi}_i + y_{\ell,i} \end{bmatrix} \succeq -\sigma_{\ell,i,1} M \;\; \forall i \in \{1...N\} \; \forall \ell \in \mathcal{L} \tag{5.12p}$$

$$\begin{bmatrix} \Lambda_\ell & \frac{1}{2}\zeta_{\ell,i,2} \\ \left(\frac{1}{2}\zeta_{\ell,i,2}\right)^\top & -\zeta_{\ell,i,2}^\top \hat{\xi}_i + y_{\ell,i} \end{bmatrix} \succeq -\sigma_{\ell,i,2} M \;\; \forall i \in \{1,...,N\} \;\; \forall \ell \in \mathcal{L} \tag{5.12q}$$

$$||\zeta_{\ell,i,j}||_* \leq \lambda_\ell \;\; \forall j \in \mathcal{J} \;\; \forall i \in \{1,...,N\} \;\; \forall \ell \in \mathcal{L} \tag{5.12r}$$

5.3 The Copula-Based DR-OPF Model

In this section, we derive the copula-based distributionally robust OPF problem. In particular, we introduce the mathematical definition of copula-based ambiguity set in Section 5.3.1. The reformulation of a generic worst-case expectation problem is proposed in Section 5.3.2 and the final distributionally robust OPF problem, given the copula-based ambiguity set is showcased in Section 5.3.3.

5.3.1 Definition of the Copula-Based Ambiguity Set

We begin with the definition of copula $\mathbb{C}$ of the distribution $\mathbb{Q}$. Let us consider $|\mathcal{K}|$ number of renewable power units, e.g., wind farms. Let $\widetilde{\xi} \in \mathbb{R}^{|\mathcal{K}|}$ be linked to distribution $\mathbb{Q}$. Recall that symbols with a hat refer to historical observations, those with a tilde, e.g., $\widetilde{\xi}$, correspond to uncertain parameters. The symbol $\widetilde{\xi}_k$ refers to the uncertain generation of the renewable power unit k. Mathematically speaking, the copula $\mathbb{C}$ of distribution $\mathbb{Q}$ is defined as the cumulative distribution function of the uncertain parameter $\widetilde{U}$ [132, 133], i.e.,

$$\left(\widetilde{U}_1, ..., \widetilde{U}_k, ..., \widetilde{U}_{|\mathcal{K}|}\right) = \left(F_1\left(\widetilde{\xi}_1\right), ..., F_k\left(\widetilde{\xi}_k\right), ..., F_{|\mathcal{K}|}\left(\widetilde{\xi}_{|\mathcal{K}|}\right)\right), \tag{5.13}$$

where $\widetilde{U} \in \mathbb{R}^{|\mathcal{K}|}$ is linked to distribution $\mathbb{C}$. In addition, the function $F_k(.) = \mathbb{Q}_k\left(\widetilde{\xi}_k \leq .\right)$ represents the cumulative distribution function of element $\widetilde{\xi}_k$ of the $|\mathcal{K}|$-dimensional uncertain vector $\widetilde{\xi}$. Note that the probability operator $\mathbb{Q}_k(.)$ represents the marginal distribution function of the random variable $\widetilde{\xi}_k$. Therefore, $F_k\left(\widetilde{\xi}_k\right)$ defines a random variable which is uniformly distributed on the interval $[0, 1]$. By (5.13), the resulting distribution of $\widetilde{U}$ has marginals that are uniformly distributed on the interval $[0, 1]$ and has the property to embody the dependence between the components of $\widetilde{\xi}$. This observation is summarized in Fig. 5.1 which suggests that the information contained in any multivariate distribution function

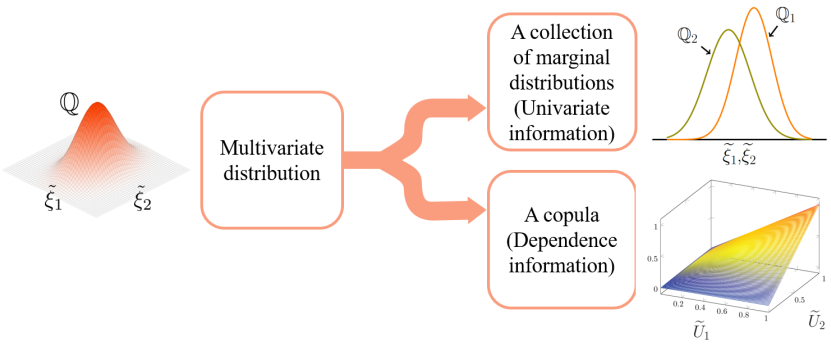


Figure 5.1: Illustration of the information contained within the copula: A stylized example with two dimensions.

can be splitted into *(i)* a collection of marginal distributions (containing the univariate information) and, *(ii)* a copula (containing the dependence information) from which the univariate information has been filtered.

We now elaborate on the need for a copula-based ambiguity set. Whereas the traditional metric-based ambiguity set $\mathcal{A}_1$ allows the decision-maker to incorporate useful information as much as possible into the distributionally robust program, it may still contain erroneous distributions, i.e., distributions that do not reflect with fidelity the potential outcome of the uncertainty. Besides, the moment-metric-based approach is only able to capture linear dependencies and may fail in representing the whole dependence structure. Hence, the decisions may unnecessarily be optimized for an over-conservative and/or non-representative insight of uncertainty, resulting in a higher expected total operational cost. In order to avoid such a situation, we aim to improve the representation of dependencies within the definition of ambiguity set, using copulas. In that direction, we introduce the copula-based ambiguity set $\mathcal{A}_4$ which generally contains more representative distribution functions, such that

$$\mathcal{A}_4 = \left\{ \mathbb{Q} \in \mathcal{M} \;\middle|\; \begin{array}{l} d_W\left(\mathbb{Q}, \widehat{\mathbb{Q}}_N\right) \leq \theta_1 \\ d_W\left(\mathbb{C}, \widehat{\mathbb{C}}_N\right) \leq \theta_2 \end{array} \right\}. \tag{5.14}$$

The first constraint in the ambiguity set $\mathcal{A}_4$ is identical to the one in $\mathcal{A}_1$, yielding the desirable properties of the classical definition of the ambiguity set. The newly added second constraint limits the distributional distance $d_W\left(\mathbb{C}, \widehat{\mathbb{C}}_N\right)$ between the endogenously selected copula $\mathbb{C}$ of the endogenously selected distribution $\mathbb{Q}$ and the empirical copula $\widehat{\mathbb{C}}_N$ of the distribution $\widehat{\mathbb{Q}}_N$. This distance should not be greater than $\theta_2 \in \mathbb{R}^+$. Again, θ_2 is a parameter to be tuned by the decision-maker.

By restricting the distributions $\mathbb{Q}$ inside the ambiguity set to have a copula $\mathbb{C}$ in the neighbourhood of the empirical one, the distributions inside the ambiguity set will follow a dependence structure which remains close to the historically observed one. Similar to θ_1, a greater value of θ_2 implies that the decision-maker is less confident about the true dependence structure of the uncertainty, and includes distributions whose dependencies are less similar to those of the empirical one within the ambiguity set.

5.3.2 Reformulation of the Worst-Case Expectation Problem

One key step in deriving the reformulation of the worst-case expectation problem (5.4), given the copula-based ambiguity set $\mathcal{A}_4$ is to establish an analytical link between the copula $\mathbb{C}$ and its distribution function $\mathbb{Q}$. This can be achieved by using (5.13), in which $\mathbb{Q}$ and $\mathbb{C}$ are linked through the marginal cumulative distribution functions $F_k(.)$.

Remark 2. Theoretically, the link between variable copula $\mathbb{C}$ and variable distribution $\mathbb{Q}$ is established via the marginal cumulative distribution functions $F_k^{\mathbb{Q}}\,\forall k$ of $\mathbb{Q}$. However, in practice, these functions are not straightforwardly accessible, due to their complex variable nature which impedes their endogenous reformulation. Consequently, considering these variable functions within the optimization framework would require research efforts that are beyond the scope of this paper. Therefore, we use the *empirical* marginal cumulative distribution functions $F_k^{\hat{\mathbb{Q}}_N}\,\forall k$ of $\hat{\mathbb{Q}}_N$. By doing so, we are able to derive a tractable reformulation of (5.4). Our claim is that the approximation made by assessing the endogenous copula $\mathbb{C}$ via the functions $F_k^{\hat{\mathbb{Q}}_N}$, is required (to avoid non-linear formulation of functions $F_k^{\mathbb{Q}}$) and is valid in practice when $F_k^{\mathbb{Q}}$ and $F_k^{\hat{\mathbb{Q}}_N}$ are close to each other (e.g., when θ_1 is small). This setting is typically suited for day-ahead probabilistic forecast embedded in optimal power flow problems, as further demonstrated in our numerical analysis in Section 5.4. In the following, functions $F_k^{\hat{\mathbb{Q}}_N}$ are denoted F_k for the ease of notation.

From now on, we denote by $\eta \in \mathbb{R}$, the value of the argument of the function, i.e., $F_k(\eta)$. Let us further clarify this function by a schematic illustration. Fig. 5.2 shows the shape of the function $F_k(\eta)$ for the renewable power unit k, given the arbitrarily selected eight equiprobable historical observations $\widehat{\xi}_{ki},\ \ i \in \{1, ..., N = 8\}$ of $\widetilde{\xi}_k$. In particular, this figure shows how the historical observations $\widehat{\xi}_{ki}$, the variable η in the x-axis and the function $F_k(\eta)$ in the y-axis are linked. Accordingly, the empirical marginal cumulative distribution function for the renewable power unit k writes as

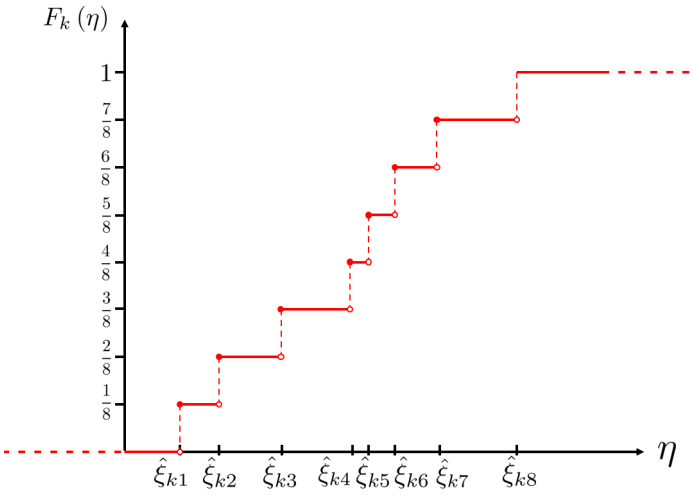


Figure 5.2: Illustration of an empirical marginal cumulative distribution function for the renewable power unit k when the number of historical observations is $N = 8$ (an arbitrarily selected number).

$$F_k(\eta) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\eta \geq \widehat{\xi}_{ki}} \text{ with } \mathbb{1}_{\eta \geq \widehat{\xi}_{ki}} = \begin{cases} 1 \text{ if } \eta \geq \widehat{\xi}_{ki} \\ 0 \text{ otherwise.} \end{cases} \tag{5.15}$$

Lemma 5.1. *The empirical marginal cumulative distribution function $F_k(\eta)$ for the renewable power unit k is equivalent to the following linear optimization program:*

$$\max_{z_{ki}} \quad \frac{1}{N}\sum_{i=1}^{N} z_{ki} \tag{5.16a}$$

$$s.t. \;\; z_{ki}\left(\eta - \widehat{\xi}_{ki}\right) \geq 0 \quad \forall i \in \{1, ..., N\}, \tag{5.16b}$$

$$0 \leq z_{ki} \leq 1 \qquad \forall i \in \{1, ..., N\}, \tag{5.16c}$$

where $\eta \in \mathbb{R}$ is the real number that corresponds to the argument of the function and $z_{ki} \in \mathbb{R}\,\forall k\,\forall i$ is the decision variable.

Proof. For a given argument value η, the program counts the number of historical observations $\widehat{\xi}_{ki}$, $i \in \{1, ..., N\}$ corresponding to the renewable power unit k under which the renewable power generation is lower than η. The constraint (5.16b) imposes that the variable z_{ki} takes a non-negative value whenever η is higher than the underlying observation $\widehat{\xi}_{ki}$, and vice versa. However, the value of z_{ki} is restricted to lie within 0 and 1 in (5.16c). Therefore, the optimal value of z_{ki} will mimic the function $\mathbb{1}_{\eta \geq \widehat{\xi}_{ki}}$. Hence, the objective function (5.16a) computes the value of the empirical marginal cumulative distribution function. ■

Equipped with this analytical link between $\mathbb{C}$ and $\mathbb{Q}$ in Lemma 5.1, it is now possible to reformulate the worst-case expectation problem (5.4) with the ambiguity set $\mathcal{A}_4$.

Theorem 5.1. *Given N historical observations $\widehat{\xi}_i,\ i \in \{1,...,N\}$ of the random variable $\widetilde{\xi}$, the worst-case expectation problem $\max\limits_{\mathbb{Q}\in\mathcal{A}_4}\ \mathbb{E}^{\mathbb{Q}}\left[a\left(x\right)^{\top}\widetilde{\xi}+b\left(x\right)\right]$ is equivalent to the following conic reformulation:*

$$\min_{\Pi}\ \ \alpha\theta_1+\beta\theta_2+\frac{1}{N}\sum_{i=1}^{N}y_i \tag{5.17a}$$

$$s.t.\ \ y_i \geq \max_{\xi\in\Xi}\ \ a\left(x\right)^{\top}\xi+b\left(x\right)-{\zeta_i^{(1)}}^{\top}\left(\widehat{\xi}_i-\xi\right) - {\zeta_i^{(2)}}^{\top}\left(F\left(\widehat{\xi}_i\right)-F\left(\xi\right)\right)\ \ \forall i \tag{5.17b}$$

$$||\zeta_i^{(1)}||_*\leq\alpha\ \ \forall i \tag{5.17c}$$

$$||\zeta_i^{(2)}||_*\leq\beta\ \ \forall i, \tag{5.17d}$$

where $i \in \{1,...,N\}$. The decision variables are collected in $\Pi = \{x,\ \alpha,\ \beta,\ y_i,\ \zeta_i^{(1)},\ \zeta_i^{(2)}\}$. In particular, $\alpha,\ \beta \in \mathbb{R}^+,\ y_i \in \mathbb{R}\ \forall i \in \{1,...,N\}$, and $\zeta_i^{(1)},\ \zeta_i^{(2)} \in \mathbb{R}^{|\mathcal{K}|}$ are auxiliary variables. Recall that parameters θ_1 and θ_2 in (5.17a) *are the Wasserstein radii defined for the ambiguity set $\mathcal{A}_4$. In addition, $F(\widehat{\xi}_i) = \left(F_1(\widehat{\xi}_{1i}),...,F_k(\widehat{\xi}_{ki}),...,F_{|\mathcal{K}|}(\widehat{\xi}_{|\mathcal{K}|i})\right)^{\top}$ and the operator $||.||_*$ computes the dual norm of a vector. The vector $F\left(\xi\right)=\left(F_1\left(\xi_1\right),...,F_{|\mathcal{K}|}\left(\xi_{|\mathcal{K}|}\right)\right)^{\top}$ is reformulated using Lemma 1, and is given by*

$$F\left(\xi\right)=\left(\begin{array}{c}\left\{\begin{array}{l}\max\limits_{z_{jki}}\ \ \dfrac{1}{N}\sum\limits_{j=1}^{N}z_{jki}\\ s.t.\ \ z_{jki}\left(\xi_k-\widehat{\xi}_{kj}\right)\geq 0\ \ \forall j\\ \qquad 0\leq z_{jki}\leq 1\ \ \forall j.\end{array}\right\},k=1\\ \vdots\\ \left\{\begin{array}{l}\max\limits_{z_{jki}}\ \ \dfrac{1}{N}\sum\limits_{j=1}^{N}z_{jki}\\ s.t.\ \ z_{jki}\left(\xi_k-\widehat{\xi}_{kj}\right)\geq 0\ \ \forall j\\ \qquad 0\leq z_{jki}\leq 1\ \ \forall j.\end{array}\right\},k=|\mathcal{K}|\end{array}\right). \tag{5.18}$$

Proof. We depart from the findings in [32, Theorem 2] which are equivalently

stated as

$$\min_{\alpha,\beta\geq 0,y_i} \alpha\theta_1 + \beta\theta_2 + \frac{1}{N}\sum_{i=1}^{N} y_i \tag{5.19a}$$

$$\text{s.t. } y_i \geq \max_{\xi\in\Xi} \; a\left(x\right)^\top \xi + b\left(x\right) - \alpha\, \mathrm{d}\left(\widehat{\xi}_i, \xi\right) - \beta\, \mathrm{d}_F\left(\widehat{\xi}_i, \xi\right) \;\; \forall i. \tag{5.19b}$$

For the sake of completeness, we also provide the proof for assertion (5.19) in the Appendix D. Complicating constraint (5.19b) requires reformulations, as it contains a maximization operator over variable ξ and the distance functions, which are defined in the following. From now on, the proof will focus on the reformulation of (5.19b). The distance functions are defined using norms, such that

$$d\left(\widehat{\xi}_i, \widetilde{\xi}\right) = ||\widehat{\xi}_i - \widetilde{\xi}||, \text{ and} \tag{5.20a}$$

$$\mathrm{d}_F\left(\widehat{\xi}_i, \widetilde{\xi}\right) = ||F\left(\widehat{\xi}_i\right) - F\left(\widetilde{\xi}\right)||, \tag{5.20b}$$

where the parameter vector $F(\widehat{\xi}_i) = \left(F_1\left(\widehat{\xi}_{1i}\right), ..., F_k\left(\widehat{\xi}_{ki}\right), ..., F_{|\mathcal{K}|}\left(\widehat{\xi}_{|\mathcal{K}|i}\right)\right)^\top$ and the function vector $F(\widetilde{\xi}) = \left(F_1\left(\widetilde{\xi}_1\right), ..., F_k\left(\widetilde{\xi}_k\right), ..., F_{|\mathcal{K}|}\left(\widetilde{\xi}_{|\mathcal{K}|}\right)\right)^\top$ are in $\mathbb{R}^{|\mathcal{K}|}$. In other words, each component of $\widehat{\xi}_i$ or $\widetilde{\xi}$ is given as argument of its corresponding marginal cumulative distribution function. Note that the resulting vector $F\left(\widehat{\xi}_i\right)$ is a sample of the copula, which can be evaluated *a priori* using (5.13). Note also that $F\left(\widetilde{\xi}\right)$ is a decision variable related to the variations within the variable copula, which requires further reformulations. Using such distance definitions, (5.19b) can be recast into

$$y_i \geq \max_{\widetilde{\xi}\in\Xi} \; a\left(x\right)^\top \widetilde{\xi} + b\left(x\right) - \alpha||\widehat{\xi}_i - \widetilde{\xi}|| - \beta||F\left(\widehat{\xi}_i\right) - F\left(\widetilde{\xi}\right)||. \tag{5.21}$$

To get rid of the norms inside the objective function of the inner maximization problem, we use dual norms ($||x||_* = \max_{||v||\leq 1} v^\top x$) as

$$\begin{aligned} y_i \geq \max_{\widetilde{\xi}\in\Xi} & \, a\left(x\right)^\top \widetilde{\xi} + b\left(x\right) - \alpha \max_{||\zeta_i^{(1)}||_*\leq 1} {\zeta_i^{(1)}}^\top \left(\widehat{\xi}_i - \widetilde{\xi}\right) \\ & - \beta \max_{||\zeta_i^{(2)}||_*\leq 1} {\zeta_i^{(2)}}^\top \left(F\left(\widehat{\xi}_i\right) - F\left(\widetilde{\xi}\right)\right) \;\; \forall i. \end{aligned} \tag{5.22}$$

Next, we eliminate the maximization operators on variables $\zeta_i^{(1)}$ and $\zeta_i^{(2)}$ by (i) switching the $-\max$ to a $\min -$, (ii) moving the resulting minimization operators to the left, (iii) merging the min operators over variables $\zeta_i^{(1)}$ and $\zeta_i^{(2)}$, (iv)

permuting with the max operator over $\widetilde{\xi}$ (which is allowed using the reformulation given by Lemma 1, because the objective function is linear and the feasible sets are convex and independent), and finally (v) dropping the min operator from the right-hand side of the $\geq$ constraint. The variables $\zeta_i^{(1)}$ and $\zeta_i^{(2)}$ are added to the overall set of decision variables and the constraints are added to the overall set of constraints, such that

$$y_i \geq \max_{\widetilde{\xi} \in \Xi} \; a(x)^\top \widetilde{\xi} + b(x) - \alpha {\zeta_i^{(1)}}^\top \left(\widehat{\xi}_i - \widetilde{\xi}\right) - \beta {\zeta_i^{(2)}}^\top \left(F\left(\widehat{\xi}_i\right) - F\left(\widetilde{\xi}\right)\right) \;\; \forall i \tag{5.23a}$$

$$||\zeta_i^{(1)}||_* \leq 1 \;\; \forall i \tag{5.23b}$$

$$||\zeta_i^{(2)}||_* \leq 1 \;\; \forall i, \tag{5.23c}$$

is equivalent to (5.19b). With the changes of variables $\alpha\zeta_i^{(1)} \rightarrow \zeta_i^{(1)}$ and $\beta\zeta_i^{(2)} \rightarrow \zeta_i^{(2)}$, the constraints become

$$y_i \geq \max_{\xi \in \Xi} \; a(x)^\top \xi + b(x) - {\zeta_i^{(1)}}^\top \left(\widehat{\xi}_i - \widetilde{\xi}\right) - {\zeta_i^{(2)}}^\top \left(F\left(\widehat{\xi}_i\right) - F\left(\widetilde{\xi}\right)\right) \;\; \forall i \tag{5.24a}$$

$$||\zeta_i^{(1)}||_* \leq \alpha \;\; \forall i \tag{5.24b}$$

$$||\zeta_i^{(2)}||_* \leq \beta \;\; \forall i. \tag{5.24c}$$

We use Lemma 5.1 to reformulate the functions $F(\xi)$. This key step allows us to make the link between the variable distribution and the variable copula, by using the empirical marginal cumulative distributions (see Remark 1). The vector $F(\xi)$ now becomes (5.18) which completes the proof. ■

The outcome of Theorem 1 is an optimization problem with constraints involving optimization operators, which is not straightforward to be solved by using off-the-shelf solvers. In the following solution approach, we aim to reformulate problem (5.17), especially constraint (5.17b), such that it can be incorporated into an optimization problem. To do so, we leverage duality theory and McCormick relaxation of bilinear terms. This will eventually enable us to reformulate the inner worst-case expectation in (5.4) in a tractable manner.

Solution Approach

From now on, we aim to reformulate the outcome of Theorem 5.1, especially constraint (5.17b), such that it becomes tractable for off-the-shelf solvers. The first step of the reformulation consists in getting rid of the maximization operators

within $F(\xi)$. In that direction, we derive the optimality conditions related to the optimization problems in (5.18), for which the strong duality theorem holds, and add them into the constraints of the outer maximization problem. Therefore, the optimality conditions for the k-th element of vector $F(\xi)$ in (5.18) are composed of *(i)* the primal constraints, *(ii)* the dual constraints, and *(iii)* the strong duality equality. This mathematically translates to the following set of constraints:

$$z_{jki}\left(\xi_k - \widehat{\xi}_{kj}\right) \geq 0 \ \ \forall j \tag{5.25a}$$

$$0 \leq z_{jki} \leq 1 \ \ \forall j \tag{5.25b}$$

$$\frac{1}{N} + \sigma_{kji}\left(\xi_k - \widehat{\xi}_{kj}\right) - \pi_{kji} \leq 0 \ \ \forall j \tag{5.25c}$$

$$\sigma_{kji}, \pi_{kji} \geq 0 \ \ \forall j \tag{5.25d}$$

$$\frac{1}{N}\sum_{j=1}^{N} z_{kji} = \sum_{j=1}^{N} \pi_{kji}. \tag{5.25e}$$

By doing so, the constraint (5.17b) becomes

$$y_i \geq \left\{ \begin{array}{l} \max\limits_{\xi, z_{kji}, \sigma_{kji}, \pi_{kji}} \ a(x)^\top \xi + b(x) - \zeta_i^{(1)\top}\left(\widehat{\xi}_i - \xi\right) \\ - \zeta_i^{(2)\top}\left(F\left(\widehat{\xi}_i\right) - \dfrac{1}{N}\begin{pmatrix} \sum_{j=1}^{N} z_{j1i} \\ \vdots \\ \sum_{j=1}^{N} z_{j|\mathcal{K}|i} \end{pmatrix}\right) \\ \text{s.t. } C\xi \leq D \\ \quad z_{jki}\left(\xi_k - \widehat{\xi}_{kj}\right) \geq 0 \ \ \forall k, j \\ \quad 0 \leq z_{jki} \leq 1 \ \ \forall k, j \\ \quad \dfrac{1}{N} + \sigma_{kji}\left(\xi_k - \widehat{\xi}_{kj}\right) - \pi_{kji} \leq 0 \ \ \forall k, j \\ \quad \sigma_{kji}, \pi_{kji} \geq 0 \ \ \forall k, j \\ \quad \dfrac{1}{N}\sum_{j=1}^{N} z_{kji} = \sum_{j=1}^{N} \pi_{kji} \ \ \forall k \end{array} \right\} \ \forall i, \tag{5.26a}$$

where the inner maximization operator has been dropped and the optimality conditions have been added to the set of constraints of the outer maximization problem. We write the support as $\left\{\widetilde{\xi} \in \mathbb{R}^{|\mathcal{K}|} \,\middle|\, C\widetilde{\xi} \leq D\right\}$, where $C \in \mathbb{R}^{2|\mathcal{K}| \times |\mathcal{K}|}$ and $D \in \mathbb{R}^{|\mathcal{K}|}$. The vectors $\widetilde{\xi}^{\max} \in \mathbb{R}^{|\mathcal{K}|}$ and $\widetilde{\xi}^{\min} \in \mathbb{R}^{|\mathcal{K}|}$ correspond to the maximum

and minimum thresholds for the uncertain parameters $\widetilde{\xi}$. In the case of renewable energy sources, $\widetilde{\xi}^{\min}$ and $\widetilde{\xi}^{\max}$ are related to zero and the installed capacity, respectively. We notice that the resulting outer maximization problem contains bilinear terms in the form of $z_{jki}\xi_k$ and $\sigma_{kji}\xi_k$. We use the McCormick relaxation of bilinear terms to restore linearity, such that the constraint can be cast into

$$y_i \geq \left\{ \begin{array}{ll} \max\limits_{\xi, z_{kji}, \sigma_{kji}, \pi_{kji}} & a\left(x\right)^\top \xi + b\left(x\right) - {\zeta_i^{(1)}}^\top \left(\widehat{\xi}_i - \xi\right) \\ & - {\zeta_i^{(2)}}^\top \left(F\left(\widehat{\xi}_i\right) - \dfrac{1}{N} \begin{pmatrix} \sum\limits_{j=1}^{N} z_{j1i} \\ \vdots \\ \sum\limits_{j=1}^{N} z_{j|\mathcal{K}|i} \end{pmatrix} \right) \\ \text{s.t.} & C\xi \leq D \\ & t_{jki} - z_{jki}\widehat{\xi}_{kj} \geq 0 \ \ \forall k, j \\ & t_{jki} \geq z_{jki}\xi_k^{\min} \ \ \forall k, j \\ & t_{jki} \geq \xi_k + z_{kji}\xi_k^{\max} - \xi_k^{\max} \ \ \forall k, j \\ & t_{jki} \leq \xi_k + z_{kji}\xi_k^{\min} - \xi_k^{\min} \ \ \forall k, j \\ & t_{jki} \geq z_{jki}\xi_k^{\max} \ \ \forall k, j \\ & 0 \leq z_{jki} \leq 1 \ \ \forall k, j \\ & \dfrac{1}{N} + v_{jki} - \sigma_{kji}\widehat{\xi}_{kj} - \pi_{kji} \leq 0 \ \ \forall k, j \\ & v_{jki} \geq \sigma_{kji}\xi_k^{\min} \ \ \forall k, j \\ & v_{jki} \geq \overline{V}\xi_k + \sigma_{kji}\xi_k^{\max} - \overline{V}\xi_k^{\max} \ \ \forall k, j \\ & v_{jki} \leq \overline{V}\xi_k + \sigma_{kji}\xi_k^{\min} - \overline{V}\xi_k^{\min} \ \ \forall k, j \\ & v_{jki} \geq \sigma_{kji}\xi_k^{\max} \ \ \forall k, j \\ & \sigma_{kji}, \pi_{kji} \geq 0 \ \ \forall k, j \\ & \dfrac{1}{N}\sum\limits_{j=1}^{N} z_{kji} = \sum\limits_{j=1}^{N} \pi_{kji} \ \ \forall k \end{array} \right\} \ \forall i. \tag{5.27a}$$

We now focus on getting rid of the outer maximization operator. The final step of the reformulation is to dualize (5.27), such that the problem becomes a minimization and the operator can be dropped from the constraint. The following problem is a tractable reformulation of (5.17):

$$\min_{\Pi} \ \ \alpha\theta_1 + \beta\theta_2 + \frac{1}{N}\sum_{i=1}^{N} y_i \tag{5.28a}$$

$$\text{s.t. } y_i \geq b(x) - {\zeta_i^{(1)}}^\top \widehat{\xi}_i - {\zeta_i^{(2)}}^\top F\left(\widehat{\xi}_i\right) + {\mu^{(0)}}^\top d$$
$$+ \sum_{k=1}^{|\mathcal{K}|} \sum_{j=1}^{N} \left(\mu_{kji}^{(3)} \widetilde{\xi}_k^{\max} - \mu_{kji}^{(4)} \widetilde{\xi}_k^{\min} + \mu_{kji}^{(6)} + \frac{1}{N} \mu_{kji}^{(7)} + \mu_{kji}^{(9)} \overline{V} \xi_k^{\max} - \mu_{kji}^{(10)} \overline{V} \xi_k^{\min} \right) \ \forall i \tag{5.28b}$$

$$a_k(x) + \zeta_{ik}^{(1)} - C_k^\top \mu^{(0)} + \sum_{j=1}^{N} \left(\mu_{kji}^{(3)} + \mu_{kji}^{(4)} + \overline{V} \mu_{kji}^{(10)} - \overline{V} \mu_{kji}^{(9)} \right) = 0 \ \forall k, i \tag{5.28c}$$

$$\frac{1}{N} \zeta_{ik}^{(2)} \leq \mu_{kji}^{(1)} \widehat{\xi}_{kj} + \mu_{kji}^{(2)} \widetilde{\xi}_k^{\min} + \mu_{kji}^{(3)} \widetilde{\xi}_k^{\max} - \mu_{kji}^{(4)} \widetilde{\xi}_k^{\min} - \mu_{kji}^{(5)} \widetilde{\xi}_k^{\max} + \mu_{kji}^{(6)} - \frac{1}{N} \lambda_k \ \forall k, j, i \tag{5.28d}$$

$$\mu_{kji}^{(1)} + \mu_{kji}^{(2)} + \mu_{kji}^{(3)} - \mu_{kji}^{(4)} - \mu_{kji}^{(5)} = 0 \ \forall k, j, i \tag{5.28e}$$

$$-\mu_{kji}^{(7)} + \mu_{kji}^{(8)} + \mu_{kji}^{(9)} - \mu_{kji}^{(10)} - \mu_{kji}^{(11)} = 0 \ \forall k, j, i \tag{5.28f}$$

$$\mu_{kji}^{(7)} - \lambda_k = 0 \ \forall k, j, i \tag{5.28g}$$

$$\mu_{kji}^{(7)} - \mu_{kji}^{(8)} \xi_k^{\min} - \mu_{kji}^{(9)} \xi_k^{\max} + \mu_{kji}^{(10)} \xi_k^{\min} + \mu_{kji}^{(11)} \xi_k^{\max} \leq 0 \tag{5.28h}$$

$$||\zeta_i^{(1)}||_* \leq \alpha \ \forall i \tag{5.28i}$$

$$||\zeta_i^{(2)}||_* \leq \beta \ \forall i. \tag{5.28j}$$

The outcome of Theorem 5.1 can be used to reformulate the worst-case expectation in the objective function (3.10a) of the copula-based DR-OPF problem, as well as the worst-case expectation appearing in the probabilistic constraints (when using the CVaR approximation to deal with DRCCs as described in Section 3.2.2). Following this procedure, we end up in a final tractable conic formulation, which is introduced in the next section.

5.3.3 Final Model

Using Theorem 5.1, we reformulate the distributionally robust OPF problem (3.10) under ambiguity set $\mathcal{A}_4$. The resulting model is a conic program and is as follows:

$$\min_{\Theta} \sum_{e \in \mathcal{E}} C_e p_e + \overline{C}_e \overline{r}_e + \underline{C}_e \underline{r}_e + \alpha\theta_1 + \beta\theta_2 + \frac{1}{N} \sum_{i=1}^{N} \sigma_i \tag{5.29a}$$

$$\text{s.t. (3.10b) - (3.10g)}, \tag{5.29b}$$

$$
\begin{cases}
-\zeta_i^{1\top}\hat{\xi}_i - \zeta_i^{2\top}\hat{F}_i + \mu^{0\top}h + \sum_{j=1}^{N}\left(\mu_{ij}^{3\top}\bar{\xi}^{\max} - \mu_{ij}^{4\top}\bar{\xi}^{\min} + \mu_{ij}^{6\top}\mathbb{1}\right) \le \sigma_i \;\; \forall i \in \{1,...,N\} \\
Y^\top C + \zeta_i^1 - Q^\top \mu_i^0 + \sum_{j=1}^{N}\left(\mu_{ij}^4 + \mu_{ij}^3\right) = 0 \;\; \forall i \in \{1,...,N\} \\
\frac{1}{N}\zeta_{wi}^2 - \mu_{wij}^{1\top}\hat{\xi}_{wi} - \mu_{wij}^{2\top}\xi_w^{\min} - \mu_{wij}^{3\top}\bar{\xi}_w^{\max} + \mu_{wij}^{4\top}\bar{\xi}_w^{\min} + \mu_{wij}^{5\top}\bar{\xi}_w^{\max} - \mu_{wij}^6 \le 0 \;\; \forall i \in \{1,...,N\} \;\; \forall j \in \{1,...,N\} \\
\mu_{wij}^1 + \mu_{wij}^2 + \mu_{wij}^3 - \mu_{wij}^4 - \mu_{wij}^5 = 0 \;\; \forall i \in \{1,...,N\} \;\; \forall j \in \{1,...,N\} \\
||\zeta_{1i}||_* \le \alpha, \;\; ||\zeta_{2i}||_* \le \beta \;\; \forall i \in \{1,...,N\}
\end{cases} \tag{5.29c}
$$

$$
\left.\begin{cases}
\overline{\tau}_e + \frac{1}{\epsilon}\left(\overline{\alpha}_e\theta_1 + \overline{\beta}_e\theta_2 + \frac{1}{N}\sum_{i=1}^{N}\overline{\sigma}_{ei}\right) \le 0 \\
-\overline{r}_e - \overline{\tau}_e - \overline{\zeta}_{1ei}^{1\top}\hat{\xi}_i - \overline{\zeta}_{1ei}^{2\top}\hat{F}_i + \overline{\mu}_e^{0\top}h + \sum_{j=1}^{N}\left(\overline{\mu}_{1eij}^{3\top}\bar{\xi}^{\max} - \overline{\mu}_{1eij}^{4\top}\bar{\xi}^{\min} + \overline{\mu}_{1eij}^{6\top}\mathbb{1}\right) \le \overline{\sigma}_{ei} \;\; \forall i \in \{1,...,N\} \\
-\overline{\zeta}_{2ei}^{1\top}\hat{\xi}_i - \overline{\zeta}_{2ei}^{2\top}\hat{F}_i + \overline{\mu}_e^{0\top}h + \sum_{j=1}^{N}\left(\overline{\mu}_{2eij}^{3\top}\bar{\xi}^{\max} - \overline{\mu}_{2eij}^{4\top}\bar{\xi}^{\min} + \overline{\mu}_{2eij}^{6\top}\mathbb{1}\right) \le \overline{\sigma}_{ei} \;\; \forall i \in \{1,...,N\} \\
Y_{ek} + \overline{\zeta}_{1eki}^1 - Q^\top\overline{\mu}_{1ei}^0 + \sum_{j=1}^{N}\left(\overline{\mu}_{1ekij}^4 + \overline{\mu}_{1ekij}^3\right) = 0 \;\; \forall i \in \{1,...,N\} \\
\overline{\zeta}_{2eki}^1 - Q^\top\overline{\mu}_{2ei}^0 + \sum_{j=1}^{N}\left(\overline{\mu}_{2ekij}^4 + \overline{\mu}_{2ekij}^3\right) = 0 \;\; \forall i \in \{1,...,N\} \\
\frac{1}{N}\overline{\zeta}_{vekij}^2 - \overline{\mu}_{vekij}^{1\top}\hat{\xi}_{ki} - \overline{\mu}_{vekij}^{2\top}\xi_k^{\min} - \overline{\mu}_{vekij}^{3\top}\bar{\xi}_k^{\max} + \overline{\mu}_{vekij}^{4\top}\bar{\xi}_k^{\min} + \overline{\mu}_{vekij}^{5\top}\bar{\xi}_k^{\max} \\
\qquad - \overline{\mu}_{vekij}^6 \le 0 \;\; \forall i \in \{1,...,N\} \;\; \forall j \in \{1,...,N\} \\
\overline{\mu}_{vekij}^1 + \overline{\mu}_{vekij}^2 + \overline{\mu}_{vekij}^3 - \overline{\mu}_{vekij}^4 - \overline{\mu}_{vekij}^5 = 0 \;\; \forall i \in \{1,...,N\} \;\; \forall j \in \{1,...,N\} \\
||\overline{\zeta}_{vei}^1||_* \le \overline{\alpha}_e, \;\; ||\overline{\zeta}_{vei}^2||_* \le \overline{\beta}_e \;\; \forall i \in \{1,...,N\}
\end{cases}\right\} \forall e \in \mathcal{E}
$$

$$
\left.\begin{cases}
\underline{\tau}_e + \frac{1}{\epsilon}\left(\underline{\alpha}_e\theta_1 + \underline{\beta}_e\theta_2 + \frac{1}{N}\sum_{i=1}^{N}\underline{\sigma}_{ei}\right) \le 0 \\
-\underline{r}_e - \underline{\tau}_e - \underline{\zeta}_{1ei}^{1\top}\hat{\xi}_i - \underline{\zeta}_{1ei}^{2\top}\hat{F}_i + \underline{\mu}_e^{0\top}h + \sum_{j=1}^{N}\left(\underline{\mu}_{1eij}^{3\top}\bar{\xi}^{\max} - \underline{\mu}_{1eij}^{4\top}\bar{\xi}^{\min} + \underline{\mu}_{1eij}^{6\top}\mathbb{1}\right) \le \underline{\sigma}_{ei} \;\; \forall i \in \{1,...,N\} \\
-\underline{\zeta}_{2ei}^{1\top}\hat{\xi}_i - \underline{\zeta}_{2ei}^{2\top}\hat{F}_i + \underline{\mu}_g^{0\top}h + \sum_{j=1}^{N}\left(\underline{\mu}_{2eij}^{3\top}\bar{\xi}^{\max} - \underline{\mu}_{2eij}^{4\top}\bar{\xi}^{\min} + \underline{\mu}_{2eij}^{6\top}\mathbb{1}\right) \le \underline{\sigma}_{ei} \;\; \forall i \in \{1,...,N\} \\
-Y_{ek} + \underline{\zeta}_{1eki}^1 - Q^\top\underline{\mu}_{1ei}^0 + \sum_{j=1}^{N}\left(\underline{\mu}_{1ekij}^4 + \underline{\mu}_{1ekij}^3\right) = 0 \;\; \forall i \in \{1,...,N\} \\
\underline{\zeta}_{2eki}^1 - Q^\top\underline{\mu}_{2ei}^0 + \sum_{j=1}^{N}\left(\underline{\mu}_{2ekij}^4 + \underline{\mu}_{2ekij}^3\right) = 0 \;\; \forall i \in \{1,...,N\} \\
\frac{1}{N}\underline{\zeta}_{vekij}^2 - \underline{\mu}_{vekij}^{1\top}\hat{\xi}_{wi} - \underline{\mu}_{vekij}^{2\top}\xi_w^{\min} - \underline{\mu}_{vekij}^{3\top}\xi_k^{\max} + \underline{\mu}_{vekij}^{4\top}\xi_k^{\min} + \underline{\mu}_{vekij}^{5\top}\xi_k^{\max} \\
\qquad - \underline{\mu}_{vekij}^6 \le 0 \;\; \forall i \in \{1,...,N\} \;\; \forall j \in \{1,...,N\} \\
\underline{\mu}_{vekij}^1 + \underline{\mu}_{vekij}^2 + \underline{\mu}_{vekij}^3 - \underline{\mu}_{vekij}^4 - \underline{\mu}_{vekij}^5 = 0 \;\; \forall i \in \{1,...,N\} \;\; \forall j \in \{1,...,N\} \\
||\underline{\zeta}_{vei}^1||_* \le \underline{\alpha}_e, \;\; ||\underline{\zeta}_{vei}^2||_* \le \underline{\beta}_e \;\; \forall i \in \{1,...,N\}
\end{cases}\right\} \forall e \in \mathcal{E} \tag{5.29d}
$$

where Θ represents the set of decision variables and includes $\underline{\alpha}_g, \underline{\beta}_g, \underline{\tau}_g, \underline{y}_{gi}, \underline{\zeta}^1_{gwi}$, $\underline{\zeta}^2_{gwi}, \underline{\mu}^0_{gwi}, \underline{\mu}^1_{gwij}, \underline{\mu}^2_{gwij}, \underline{\mu}^3_{gwij}, \underline{\mu}^4_{gwij}, \underline{\mu}^5_{gwij}, \underline{\mu}^6_{gwij}, \overline{\alpha}_g, \overline{\beta}_g, \overline{\tau}_g, \overline{y}_{gi}, \overline{\zeta}^1_{gwi}, \overline{\zeta}^2_{gwi}, \overline{\mu}^0_{gwi}, \overline{\mu}^1_{gwij}$, $\overline{\mu}^2_{gwij}, \overline{\mu}^3_{gwij}, \overline{\mu}^4_{gwij}, \overline{\mu}^5_{gwij}, \overline{\mu}^6_{gwij}, \zeta^1_{wi}, \zeta^2_{wi}, \mu^0_{wi}, \mu^1_{wij}, \mu^2_{wij}, \mu^3_{wij}, \mu^4_{wij}, \mu^5_{wij}, \mu^6_{wij}, g, \overline{r}, \underline{r}, V$, α, β, σ_i.

5.4 Numerical Experiments

In this section, we provide a numerical study to illustrate the benefits of the distributionally robust OPF models with the proposed moment-metric-based and copula-based ambiguity sets $\mathcal{A}_3$ and $\mathcal{A}_4$ over the traditional approaches in the literature. We begin with the presentation of our case study. Next, the performed numerical experiments are structured as follows. Section 5.4.1 discusses the computational times of each model. Section 5.4.2 studies the out-of-sample expected total operating cost which is achieved by the different ambiguity sets, when the Wasserstein radius θ_1 varies. The discussion regarding the performance of the copula-based ambiguity set $\mathcal{A}_4$ when the additional exogenous parameter θ_2 varies is available in Section 5.4.3, and allows to show the benefits of considering the full non-linear dependence structure via copulas. Finally, we explore the impacts of increasing the number of wind farms in Section 5.4.4

Our case study is based on a slightly updated version of the 24-node IEEE reliability test system [141], whose input data are provided in the Appendix A. This case study is composed of 12 conventional generating units with an aggregate capacity of $2,362.5$ MW, two wind farms with a total maximum installed capacity of $1,000$ MW, and eventually 17 loads with an aggregate demand of $2,207$ MW. These power suppliers and demands are connected through a network composed of 24 nodes and 34 transmission lines. The sole source of uncertainty is the deviation of renewable power generation in the real-time operation with respect to the day-ahead forecast values. To cope with such deviations, the system operator reserves a fraction of capacities of conventional units in the day-ahead stage, and activates them in the real-time operation, if necessary. The total maximum reserve capacity that conventional units can provide is 798 MW.

In this study, to assess the impacts of dependence structure on operational decisions of the system operator, we generate a dataset of $1,000$ samples representing the historical wind power observations. To do so, we use the package *DatagenCopulaBased* v1.3.0 in Julia programming language v1.4.2, which allows to generate samples with a predefined dependence structure (e.g., following the Gaussian copula). The resulting dataset mimics historical wind power observations from which we retrieve the average value μ (corresponding to the day-ahead forecast).

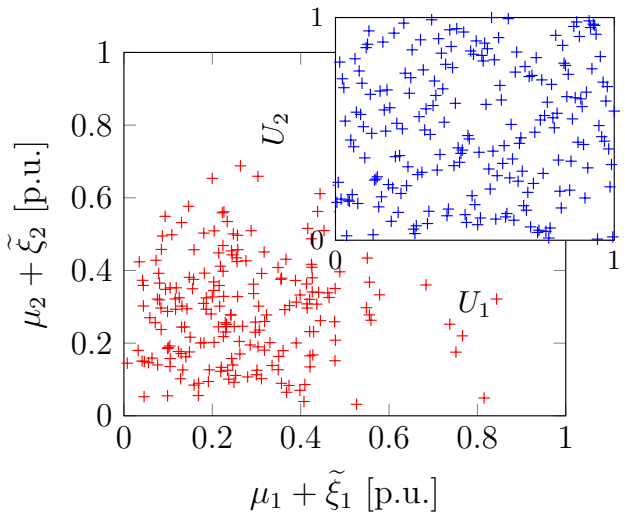


Figure 5.3: Historical wind power observations for two farms (red plot) and the corresponding empirical copula (blue plot)

The final dataset of forecast errors, and its corresponding copula are illustrated in Fig. 5.3.

The definition of additional input parameters is described in the following. We design an ellipsoidal support of uncertainty required for the enhanced moment-metric-based ambiguity set $\mathcal{A}_3$ with parameters[4]

$$\Sigma_0 = \begin{pmatrix} 2.20 & -0.25 \\ -0.25 & 2.20 \end{pmatrix} \text{ and } \xi_0 = \begin{pmatrix} 0.2046 \\ 0.2046 \end{pmatrix}. \tag{5.30}$$

Besides, the copula-based requires the definition of the empirical copula, which is derived via $F(\widehat{\xi}_i) = \left(F_1(\widehat{\xi}_{1i}), ..., F_k(\widehat{\xi}_{ki}), ..., F_{|\mathcal{K}|}(\widehat{\xi}_{|\mathcal{K}|i})\right)^\top$. The latter parameter vector is straightforwardly determined using the empirical observations of uncertainty $\widehat{\xi}_i$ and the analytical formula (5.15).

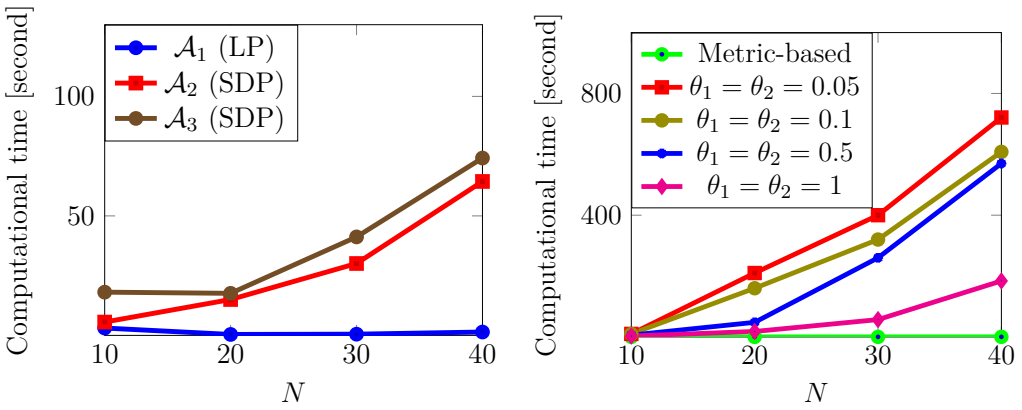
We run all aforementioned models with different numbers of historical observations, i.e., $N = \{10, 20, 30, 40\}$ samples selected from 1,000 in-sample data, for the purpose of the computational study in Section 5.4.1. Next, we select $N = 30$ fixed in-sample scenarios and solve these models with different risk-attitudes, i.e., $\epsilon = \{0.03, 0.05\}$ and different Wasserstein radii θ_1 and θ_2 ranging from 10^{-4} to 10^{-1}, where the exponent increases linearly with a step of 0.2. The linear models

[4]These parameters have been selected by trials and errors in the scope of this study, with limited impact on the final results of our proposed model. A more accurate design of the support is left for the future work.

are solved using Gurobi v8.0.1 while the SDP models are solved using Mosek 9.2, in JuMP v0.21.3 under programming language JuliaPro v1.4.2 on a computer clocking at 2.2 GHz with 16 GB of RAM. All source codes are publicly available in the online companions [64] and [63].

5.4.1 Computational Performance

We retrieve the computational time corresponding to each model with different number of in-sample scenarios N, ranging from 10 to 40. The computational times are reported in Figs. 5.4(a) and 5.4(b). As expected, we observe that the computational time for the model with the metric-based ambiguity $\mathcal{A}_1$ is the lowest due to its linearity. We now focus on the enhanced moment-metric-based ambiguity set in Fig. 5.4(a). These results show that the computational time related to the OPF model with the moment-metric-based and the enhanced moment-metric-based ambiguity sets are comparable. The reason for this is that the scale of problems (5.5) and (5.7) under ambiguity sets $\mathcal{A}_2$ and $\mathcal{A}_3$ are almost equivalent — only one supplementary scalar variable β_i appears in (5.7) compared to (5.5). These results further suggest that the computational time in both SDP models grows linearly when N increases.



(a) Moment-metric-based ambiguity set: computational time as a function of the number of in-sample scenarios N. Fixed parameters: $\epsilon = 0.05$ and $\theta_1 = 0.001$.

(b) Copula-based ambiguity set: computational time as a function of the number of historical in-sample observations N. Fixed value: $\epsilon = 0.05$.

Figure 5.4: Computational study.

We now focus on the copula-based ambiguity set $\mathcal{A}_4$ and Fig. 5.4(b) which shows the computational time as a function of the number of historical in-sample observations N under different values for θ_1 and θ_2. We observe that the computational

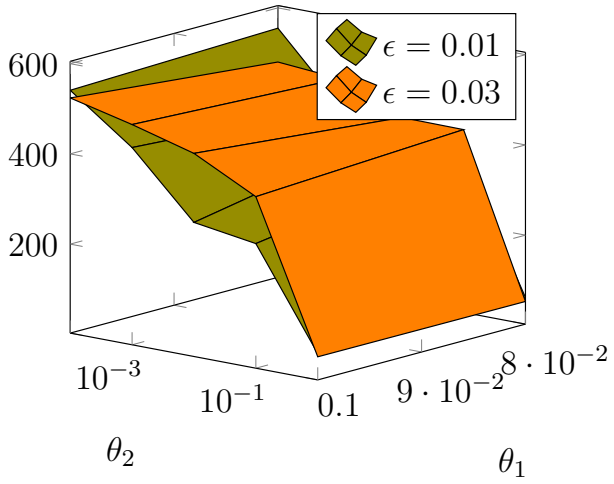


Figure 5.5: Computational time as a function of parameters θ_1 and θ_2. Fixed values: $N = 25$. The metric-based ambiguity set approach finds a solution in less than 1 second for the values of θ_1 explored in this study.

time increases with the number of historical observations in a non-linear manner. These results suggest that the computational time growth follows a quadratic trend, which is in line with the increase in the number of variables and constraints — see Theorem 5.1, where the number of certain variables and constraints increases quadratically[5] in N. It is worth mentioning that the order of magnitude of computational time is larger with the copula-based ambiguity set, compared to the moment-metric-based ambiguity set, up to a multiplying factor of 10 between both approaches.

In addition, Fig. 5.5 depicts the computational time for various values of parameters θ_1 and θ_2 for the proposed copula-based ambiguity set $\mathcal{A}_4$. The results highlight that the computational time slightly decreases when θ_2 increases. Recall that two wind farms only have been considered so far. We will investigate later in Section 5.4.4 the scalability of the proposed model with respect to the dimensionality of the uncertainty space by increasing the number of wind farms.

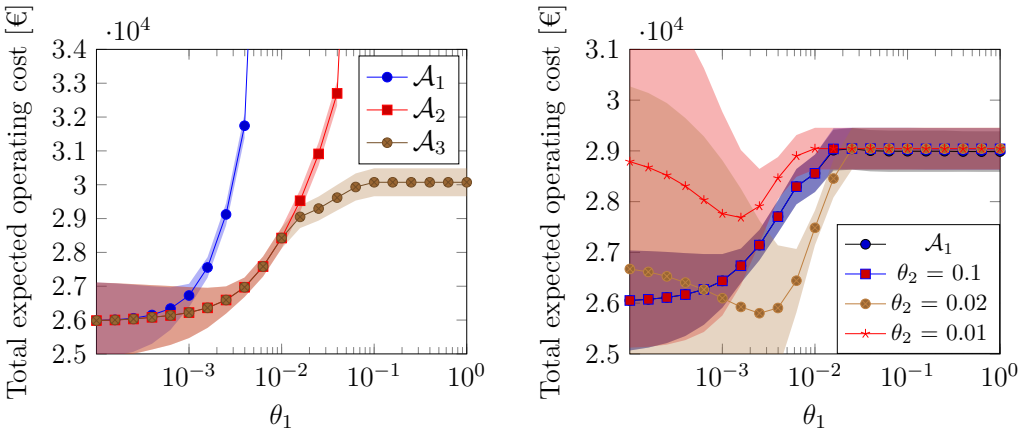
In the next two subsections, we investigate the out-of-sample performance of

[5]Compared to other uncertainty modeling techniques, e.g., scenario-based stochastic programming, the DRO approach generally provides more qualified decisions in terms of the out-of-sample performance, when the number of historical in-sample observations is relatively low. This may further motivate the use of DRO when there is limited historical data, or when the decision-maker aims to reduce the computational time by intentionally reducing the number of historical observations. As shown in Fig. 5.4(b), the computational time for the DRO approach is satisfactory when N is comparatively low.

the DRO model with the ambiguity sets $\mathcal{A}_3$ and $\mathcal{A}_4$, using the 970 remaining unseen scenarios. Recall that the ambiguity set $\mathcal{A}_3$ contains one risk-tuning parameter only[6], i.e., θ_1, that restricts the distance of distributions within the ambiguity set to the empirical one. In contrast, the ambiguity set $\mathcal{A}_4$ comprises of two risk-tuning parameters, i.e., θ_1 and θ_2, where θ_2 restricts the similarity of distributions within the ambiguity set in terms of the dependence structure to that of the empirical one. For the ease of comparison, we alternately fix one of the two risk-tuning parameters θ_1 and θ_2, while varying the other one.

5.4.2 Out-of-Sample Performance: Operational Cost of the System as a Function of θ_1

We assign three different values to parameter θ_2, namely 0.01, 0.02 and 0.1, in $\mathcal{A}_4$. We vary the value of parameter θ_1 from 10^{-4} to 10^{0}, where the exponent increases linearly with a step of 0.2, and compare the out-of-sample performance (e.g., in terms of total expected operating cost) of the DRO model for ambiguity sets $\mathcal{A}_1$ to $\mathcal{A}_4$. These results are illustrated in logarithmic Figs. 5.6(a) and 5.6(b).



(a) Moment-metric-based ambiguity set: out-of-sample total expected operational cost of the system as a function of θ_1. Fixed parameters: $\epsilon = 0.05$.

(b) Copula-based ambiguity set: out-of-sample cost as a function of θ_1 considering different values for θ_2

Figure 5.6: Numerical study. The shaded area around each curve represents the corresponding standard deviation.

[6]The violation probability $\epsilon_{(.)}$ of each DRCC can also be seen as a risk-tuning parameter. However, we keep this value unchanged (fixed to 0.05) in our numerical study, and focus on the risk-tuning parameters related to the proposed ambiguity set.

From Fig. 5.6(a), we observe that the system operator may achieve a lower expected operational cost when using the moment-metric-based ambiguity set $\mathcal{A}_2$, compared to $\mathcal{A}_1$. This interesting result stems from the smart design of $\mathcal{A}_2$ that imposes the dependence structure of probability distributions within the ambiguity set. Nevertheless, we observe that after a threshold value of radius θ_1, i.e., 0.001 for $\mathcal{A}_1$ and 0.02 for $\mathcal{A}_2$, the expected operational cost drastically increases. The program even fails to be feasible after a threshold value of 0.005 for $\mathcal{A}_1$ and 0.05 for $\mathcal{A}_2$. The reason for this is that the worst-case distribution within the ambiguity set is not bounded by the support of uncertainty. For example, it may give a non-zero probability to a negative value of wind power generation. This may lead to infeasibility when the amount of available reserve is not sufficient for the system under the worst-case distribution. Our proposed technique relying on the enhanced moment-metric-based ambiguity set $\mathcal{A}_3$ that embeds the support information allows to recover feasibility when θ_1 takes a comparatively high value. By bounding distributions within the ambiguity set to lie within the support, the values taken by the worst-case distribution remain realistic. Eventually, our proposed technique achieves a lower expected operational cost when θ_1 is greater than 0.01, compared to a moment-metric-based approach. In each of the three explored models, the standard deviation decreases when the radius θ_1 increases.

We now focus on the copula-based ambiguity set and the results reported in Fig. 5.6(b). One can observe that the results obtained from the DRO model with the metric-based ambiguity set $\mathcal{A}_1$ and those with the copula-based ambiguity set $\mathcal{A}_4$ when θ_2 is comparatively high (e.g., 0.1) are similar. This implies that the second constraint in $\mathcal{A}_4$ regarding the distance with respect to the empirical copula is not binding, and therefore the worst-case distributions for which the optimal decisions are made are identical in both DRO models. In other words, the ambiguity set $\mathcal{A}_4$ is identical to $\mathcal{A}_1$ when θ_2 takes comparatively high values. In contrast, when the value of θ_2 decreases, e.g., to 0.02, the second constraint in $\mathcal{A}_4$ becomes binding, leading to improved results in terms of the out-of-sample cost. This improvement has led to the lowest out-of-sample cost, since the constraint on dependence structure eliminates the unrealistic distributions from the ambiguity set. This numerical finding suggests that the proposed ambiguity set $\mathcal{A}_4$ outperforms $\mathcal{A}_1$, provided that appropriate values for parameters θ_1 and θ_2 are selected (i.e., $\theta_1 = 0.0025$ and $\theta_2 = 0.02$). Given these appropriate values for θ_1 and θ_2, we observe a potential cost saving of 1% in expectation with respect to the classical ambiguity set $\mathcal{A}_1$ given the value for $\theta_1 = 10^{-4}$, which achieves the minimal expected total cost for $\mathcal{A}_1$.

Overall, the copula-based ambiguity set achieves the lowest expected operating cost among all developed models. This is due to the appropriate account of the

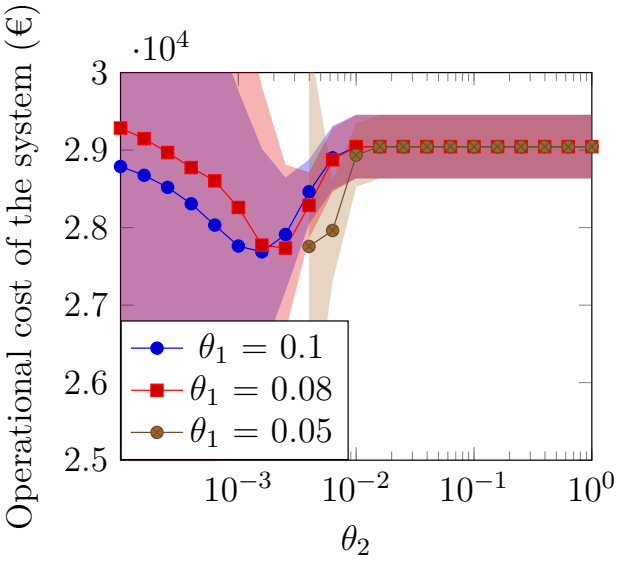


Figure 5.7: Copula-based ambiguity set: out-of-sample cost as a function of θ_2 considering different values for θ_1

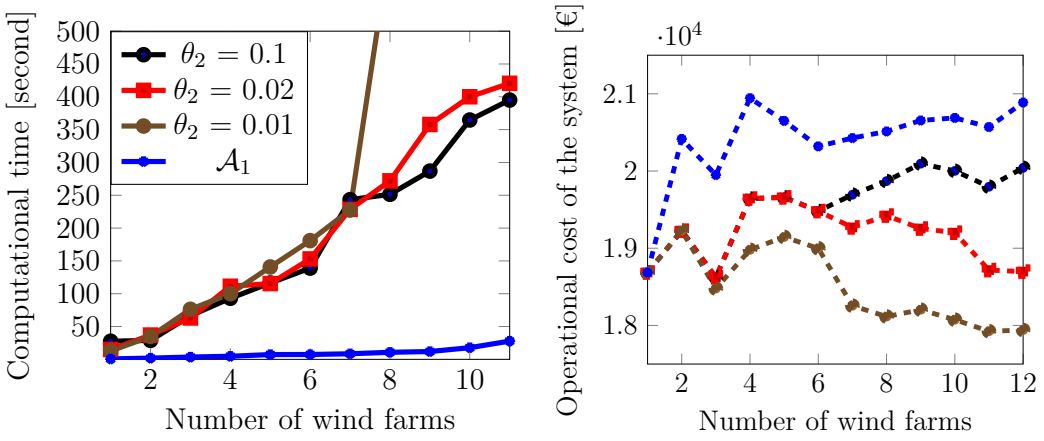
non-linear dependencies among the several wind power generation uncertainties.

It is worth mentioning that the copula-based DRO model becomes infeasible, when both parameters θ_1 and θ_2 take very low values, meaning that both constraints in $\mathcal{A}_4$ are highly restrictive. We have observed infeasibility for values of θ_1 and θ_2 lower than 0.01 in our simulations in Section V-C and for values lower than $\theta_1 = 0.05$ and $\theta_2 = 0.004$ in our simulations in Section V-D. One can intuitively interpret the observation on infeasibility as a case under which the two Wasserstein balls, defined by two constraints in $\mathcal{A}_4$, have no intersection. In other words, there is no distribution within the ambiguity set that satisfies both constraints at the same time.

5.4.3 Out-of-Sample Performance: Operational Cost of the System as a Function of θ_2

In this section, we focus on the copula-based ambiguity set to fix the value of parameter θ_1 to either 0.05, 0.08, or 0.1, and report the out-of-sample operational cost of the system for different values of θ_2. In particular, this value ranges from 10^{-4} to 10^{0}, where the exponent increases linearly with a step equal to 0.2. Fig. 5.7 shows the out-of-sample results in a logarithmic scale.

Our first observation is that for comparatively high values of θ_2, the outcomes of the DRO model with the ambiguity set $\mathcal{A}_4$ are very similar to those with



(a) Computational time as a function of the number of wind farms. Fixed values: $\theta_1 = 0.1$, $\epsilon = 0.05$, and $N = 15$.

(b) Out-of-sample operational cost of the system as a function the number of wind farms. Legends are the same at those in plot (b). Fixed values: $\theta_1 = 0.1$, $\epsilon = 0.05$, and $N = 15$.

Figure 5.8: Operational results with two (the first plot) and more (the next two plots) wind farms. The aggregate capacity of farms is always equal to $1,000$ MW.

the ambiguity set $\mathcal{A}_1$. This numerically indicates that the copula constraint is not binding. Second, we observe that the cost curves reach a minimum at intermediate values for θ_2, suggesting that an optimal value for the parameter θ_2 exists. This difference between the outcomes obtained by the DRO model with $\mathcal{A}_1$ and those with $\mathcal{A}_4$ represents the maximum cost saving that is achievable by using $\mathcal{A}_4$ instead of $\mathcal{A}_1$, for fixed value of θ_1. For instance, when $\theta_1 = 0.1$, the optimum is obtained for $\theta_2 = 0.0016$, and it corresponds to a cost saving of 4.7% compared to $\mathcal{A}_1$ with the same value of $\theta_1 = 0.1$. Third, we observe that the program becomes infeasible for comparatively low values of θ_2, when θ_1 is also low (e.g., 0.01), confirming our intuition in the previous analysis.

5.4.4 Impacts of Increasing the Number of Wind Farms

This section focuses on the copula-based ambiguity set for assessing the impact of increasing the number of wind farms, as it demonstrated to be the method that was the most suited for accounting of complex dependence structure. We increase the number of wind farms connected to the system, enlarging the dimension of the uncertainty space. We consider up to 12 wind farms, which represent the number of wind uncertainty sources which is typically considered for studies related to the real-life Belgian electricity grid (i.e., one source per province and an additional one for off-shore wind). We re-scale the wind farm capacities, such that the aggregate

capacity of wind farms is always equal to $1,000$ MW. We consider 15 historical observations.

Fig. 5.8(a) shows the computational time to solve the problem (3.10) with both ambiguity sets $\mathcal{A}_1$ (metric-based) and $\mathcal{A}_4$ (with different values of θ_2). The value of parameter θ_1 is fixed to 0.1 in all cases. In most cases, the computational time increases linearly with the number of wind farms. However, this increase is drastic when the value of θ_2 is comparatively low, indicating the case with a tight copula constraint.

Fig. 5.8(b) illustrates the out-of-sample operational cost of the system as a function of the number of wind farms. Legends are the same as those in Fig. 5.8(a). There are two important observations. First, we obtain lower values for the operational cost under the ambiguity set $\mathcal{A}_4$, in particular when the value of θ_2 is comparatively low, e.g., 0.01 or 0.02. Second, the operational cost decreases by increasing the number of wind farms, provided that the value of θ_2 is comparatively low. Both observations highlight the importance of adding dependence structure in the definition of the ambiguity set, particularly when the number of wind farms is increasing.

5.5 Chapter Summary and Conclusions

In this chapter, we explore distributionally robust optimization frameworks for the day-ahead OPF problem, that are able to capture the dependence structure among the uncertain parameters. We first propose an enhanced moment-metric-based ambiguity set that accounts for the dependencies via an additional constraint on the second-order moment of uncertainty, as well as for the support of uncertainty. Next, we introduce a copula-based ambiguity set, i.e., building up on a Wasserstein-metric-based approach with an additional constraint binding the distance with respect to the empirical copula. We provide all the mathematical reformulations related to the proposed approaches and perform thorough numerical experiments, applied to the day-ahead distributionally robust optimal power flow problem.

The outcome of this study shows *i)* the superiority of both proposed approaches in terms of the expected total operating cost of the system compared to the existing ambiguity sets, and *ii)* the improved performance of copula-based approach compared to the moment-metric-based approach. This is because the copula is able to account for the full dependence structure of the uncertain renewable energy sources. However, the copula-based approach remains the most computationally intensive approach.

Table 5.2: Summary of selection criteria for the different definitions of ambiguity sets explored in Chapters 4 and 5.

	High-dimensional uncertainty	Large-scale problem
Physical bounds	✓	✓
Moment-metric-based ambiguity set	✗	✓
Copula-based ambiguity set	✓	✗

In Table 5.2, we summarize the different criteria for selecting the appropriate definition of ambiguity set to be considered in practice. First, we recall that incorporating the physical bounds of uncertainty remains crucial in all types of application. Second, we claim that both approaches for embedding dependencies within DRO have merits, each one being more suited for different types of application. On the one hand, the moment-metric-based is suited for large-scale applications exposed to low-dimensional uncertainty, and where the time restriction to find a solution is crucial, e.g., an optimal power flow problem in a power system with limited number of wind farms. On the other hand, the copula-based approach is well suited to capture complex multi-dimensional dependencies in small-scale decision-making problems, e.g., an energy portfolio manager participating in electricity markets which is exposed to high-dimensional uncertainty but usually needs to solve a smaller scale problem (involving less variables and constraints).

In the following chapter, our aim is to show the benefits of incorporating the additional information within the ambiguity set, such as support (Chapter 4) and dependence structure (Chapter 5) on a different application. More specifically, our focus is on the coordinated power and gas dispatch problem (which can be seen as an extension of OPF problem to integrated energy systems) applied to the Belgian power and gas transmission systems.

CHAPTER 6

Distributionally Robust Optimization for the Coordinated Dispatch of Power and Gas

In the current energy transition, natural gas-fired power plants are expected to play a key role in mitigating the fluctuations from renewable power generation, therefore creating tighter links between the power and gas systems. Technical and economical risks, e.g., congestion and peak prices, may ensue from this physical and economical interplay. This calls for the coordination of operation between the power and natural gas systems as well as the anticipation of the multivariate and correlated demand uncertainty. In this chapter, leveraging the previous developments in DRO, we propose a moment-metric-based distributionally robust coordinated power and gas dispatch, which properly accommodates the uncertainty and dependencies between power and gas demands. We show the benefits of the proposed approach through numerical experiments based on a case study inspired by the Belgian power and gas systems. The results also reveal limitations on modeling gas flow dynamics within the coordinated operation of power and gas systems.

The content of this chapter is mainly based on the following publication[1]

[D] **A. Arrigo**, J.-F. Toubeau, I. Fattahi, Z. De Grève and François Vallée, "Distributionally Robust Power and Gas Dispatch with Multivariate and Correlated Uncertainty," IEEE PMAPS Conference, Manchester, June 2022.

[1]The IEEE permission grant is available at `https://journals.ieeeauthorcenter.ieee.org/choose-a-publishing-agreement/avoid-infringement-upon-ieee-copyright`, Accessed on 26^{th} April 2022.

6.1 Motivation

In the last decades, the gas and power industries have faced a substantial transition that fosters the convergence of their operation. In particular, Natural Gas-Fired Power Plants (NGFPPs) are becoming key flexible resources for cost-effectively managing the uncertainty and variability of weather-dependent renewable energy sources [3]. In addition, the electrification of energy demand results in a higher magnitude of uncertainty regarding the real-time deviation from the day-ahead demand forecast. Overall, these aspects put the system at risk by increasing the potential stresses (economical or technical) that may infer from a large deviation in energy demand from its day-ahead forecast. Recently, events happening in North America (e.g., 2008 wind power ramp down [89] or 2021 blackout happening in Texas [85]) have witnessed how this intimate interdependency between power and gas networks may lead to serious technical issues and risks of blackouts [46].

To prevent these issues, the power system research community has proposed two main potential improvements of the current operational procedures that can be seen as complementary solutions. The first solution is the *coordination* which mainly consists in the exchange of information (e.g., based on volume or price quantities) between power and gas system operators [52]. Different schemes with different degrees of operational integration have been investigated in the literature [45]. Differently, authors in [142, 143] investigate the *uncertainty-aware* day-ahead scheduling as a second solution approach, which accounts for probabilistic forecast of uncertain parameters (e.g., the multivariate and correlated demand) and for potential recourse actions when scheduling the system in day-ahead. The latter solution is the focus of this chapter. Hence, we assume a fully coordinated framework, where a unique agent, called *system operator*, operates both the power and gas systems, and has complete information about both infrastructures. In this framework, the system operator solves a coordinated power and gas dispatch problem for the overall system represented in Fig. 6.1. The aim for system operator is to minimize the total operating cost by cost-efficiently scheduling the energy suppliers, while ensuring the balance between energy supply and demand as well as grid limits. In the setting considered in this work, the demand is assumed inelastic[2] and uncertain. The uncertainty stemming from demand is multivariate (i.e., power and gas demands are either residential or industrial) and correlated. Hence, we propose modeling this uncertainty within the coordinated power and gas dispatch problem, using the moment-metric-based distributionally robust framework discussed in Chapter 5 to correctly capture the dependencies among the different power and gas demands.

[2]An inelastic demand does not vary when prices change.

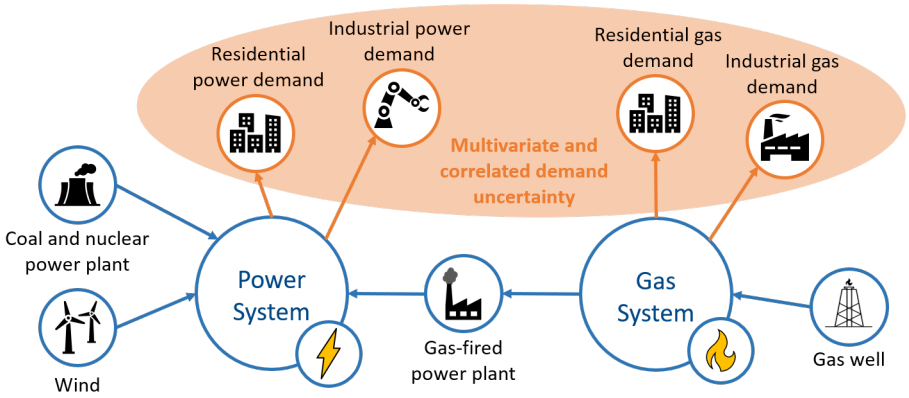


Figure 6.1: Schematic representation of power and gas systems showing involved agents and links.

Our previous work in Chapters 3, 4 and 5 analyzes the benefits of the distributionally robust optimization with respect to the traditional approaches (e.g., scenario-based programming [144], robust optimization [145, 146] and chance-constrained programming [143]). Recall that DRO may outperform these traditional approaches due to its ability to consider a family of potential distributions which describe the power and gas demand uncertainty [21]. Two main types of ambiguity set have been reported in the related power system literature and applied to the coordinated power and gas dispatch problem. First, authors in [147] use the moment-based ambiguity set which collects the distributions with given mean and covariance matrix. The shortcoming of this technique is that the information contained in such set is only based on the moment estimation and may inefficiently represent the whole ambiguity. Differently, authors in [148] use the metric-based ambiguity set [33] which collects the distributions in the neighbourhood of a central empirical one, in the sense of the Wasserstein probability metric which calculates the distance between two distributions. By doing so, the overall available historical information is fed to the program, resulting in more informed dispatch solutions. Nonetheless, with the traditional metric-based approach, there is no guarantee that the distributions within the set follow a dependence structure which is consistent with the empirically observed one. In compliance with our work in Chapter 5, we suggest supplementing the metric-based ambiguity set with an additional constraint on covariance matrix to properly accommodate the uncertainty and dependence structure in power and gas demands. This results in a moment-metric-based distributionally robust coordinated power and gas dispatch, that excludes non-representative distributions (i.e., with an inconsistent dependence structure compared to empirical data).

Following this rationale, our contributions in this work are threefold. First, we model a moment-metric-based distributionally robust coordinated power and gas

dispatch problem, enriched with an additional constraint on covariance matrix. This allows to appropriately account for the correlated uncertainty pertaining to the power and gas demand deviations from day-ahead forecast. Then, we reformulate the problem such that it becomes tractable by off-the-shelf solvers, ending up in a Semi-Definite Program (SDP). Finally, we perform numerical experiments based on a case study inspired by the real-life Belgian high voltage power and high pressure gas systems. The outcome of this analysis illustrates the benefits of the proposed approach, with respect to the traditional metric-based approach.

To the best of our knowledge, this is the first time in the literature that dependence structure is taken into account within the distributionally robust coordinated power and gas dispatch problem. In the following, Section 6.2 introduces the dispatch problem and its solution approach. Section 6.3 presents the numerical experiments performed on a case study inspired by the Belgian Power and Gas systems. It is worth mentioning that the outcome of numerical experiments also reveal limitations in the modeling of gas flow dynamics, which opens the path towards our research work in Chapter 7.

6.2 The Distributionally Robust Coordinated Power and Gas Dispatch

In this section, we derive a distributionally robust model for the coordinated power and gas dispatch problem. The formulation seeks the optimal day-ahead scheduling of generating units which minimizes the total operating cost in day-ahead, while being exposed to uncertainty in the energy demand. The operating constraints for both the electricity and natural gas systems are considered, ensuring the safe operation of the overall system, resulting in the following program.

$$\min_{\Theta} \ \max_{\mathbb{Q}\in\mathcal{A}} \ \mathbb{E}^{\mathbb{Q}}\left[\sum_{t=1}^{|\mathcal{T}|}\left(\sum_{e=1}^{|\mathcal{E}|} c_e p_{e,t} + \sum_{g=1}^{|\mathcal{G}|} c_g p_{g,t}\right)\right] \tag{6.1a}$$

s.t.

$$\sum_{e=1}^{|\mathcal{E}|} p_{e,t} + \sum_{w=1}^{|\mathcal{W}|} p_{w,t} = D_t^{\text{e,res}} + D_t^{\text{e,ind}}, \ \ \forall t \in \mathcal{T}, \tag{6.1b}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \ \mathbb{Q}\left(\underline{P}_e \leq p_{e,t}\right) \geq 1 - \underline{\epsilon}_{e,t}, \ \ \forall e \in \mathcal{E}, \ \ \forall t \in \mathcal{T}, \tag{6.1c}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \ \mathbb{Q}\left(p_{e,t} \leq \overline{P}_e\right) \geq 1 - \overline{\epsilon}_{e,t}, \ \ \forall e \in \mathcal{E}, \ \ \forall t \in \mathcal{T}, \tag{6.1d}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \mathbb{Q}\left(M_\ell^{\mathcal{E}} p_{e,t} + M_\ell^{\mathcal{W}} p_{w,t} - M_\ell^{\mathcal{D}}\left(D_t^{\text{e,res}} + D_t^{\text{e,ind}}\right) \le F_\ell^{\max}\right) \ge 1 - \epsilon_{\ell,t}, \quad \forall \ell \in \mathcal{L},\ \forall t \in \mathcal{T}, \tag{6.1e}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \mathbb{Q}\left(\underline{P}_g \le p_{g,t}\right) \ge 1 - \underline{\epsilon}_{g,t},\ \forall g \in \mathcal{G},\ \forall t \in \mathcal{T}, \tag{6.1f}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \mathbb{Q}\left(p_{g,t} \le \overline{P}_g\right) \ge 1 - \overline{\epsilon}_{g,t},\ \forall g \in \mathcal{G},\ \forall t \in \mathcal{T}, \tag{6.1g}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \mathbb{Q}\left(\underline{\text{PR}} \le pr_{m,t}\right) \ge 1 - \underline{\epsilon}_{m,t},\ \forall m \in \mathcal{M},\ \forall t \in \mathcal{T}, \tag{6.1h}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \mathbb{Q}\left(pr_{m,t} \le \overline{\text{PR}}\right) \ge 1 - \overline{\epsilon}_{m,t},\ \forall m \in \mathcal{M},\ \forall t \in \mathcal{T}, \tag{6.1i}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \mathbb{Q}\left(pr_{m,t} \le \Gamma_{m,n} pr_{n,t}\right) \ge 1 - \overline{\epsilon}_{c,t},\ \forall (m,n) \in \mathcal{C},\ \forall t \in \mathcal{T}, \tag{6.1j}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \mathbb{Q}\left(pr_{n,t} \le pr_{m,t}\right) \ge 1 - \overline{\epsilon}_{c,t},\ \forall (m,n) \in \mathcal{C},\ \forall t \in \mathcal{T}, \tag{6.1k}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \mathbb{Q}\left(q_{m,n,t} \ge 0\right) \ge 1 - \epsilon_{m,n,t},\ \forall (m,n) \in \mathcal{Z},\ \forall t \in \mathcal{T}, \tag{6.1l}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \mathbb{Q}\left(q_{m,n,t}^{\text{in}} \ge 0\right) \ge 1 - \epsilon_{m,n,t}^{\text{in}},\ \forall (m,n) \in \mathcal{Z},\ \forall t \in \mathcal{T}, \tag{6.1m}$$

$$\min_{\mathbb{Q}\in\mathcal{A}} \mathbb{Q}\left(q_{m,n,t}^{\text{out}} \ge 0\right) \ge 1 - \epsilon_{m,n,t}^{\text{out}},\ \forall (m,n) \in \mathcal{Z},\ \forall t \in \mathcal{T}, \tag{6.1n}$$

$$q_{m,n,t}^2 = K_{m,n}^2\left(pr_{m,t}^2 - pr_{n,t}^2\right)\ \forall (m,n) \in \mathcal{Z}\ \forall t \in \mathcal{T}, \tag{6.1o}$$

$$q_{m,n,t} = \frac{q_{m,n,t}^{\text{in}} + q_{m,n,t}^{\text{out}}}{2},\ \forall (m,n) \in \mathcal{Z},\ \forall t \in \mathcal{T}, \tag{6.1p}$$

$$h_{m,n,t} = S_{m,n}\frac{pr_{m,t} + pr_{n,t}}{2},\ \forall (m,n) \in \mathcal{Z},\ \forall t \in \mathcal{T}, \tag{6.1q}$$

$$h_{m,n,t} = H_{m,n}^0 + q_{m,n,t}^{\text{in}} - q_{m,n,t}^{\text{out}},\ \forall (m,n) \in \mathcal{Z},\ t = 1, \tag{6.1r}$$

$$h_{m,n,t} = h_{m,n,t-1} + q_{m,n,t}^{\text{in}} - q_{m,n,t}^{\text{out}},\ \forall (m,n) \in \mathcal{Z},\ \forall t \in \mathcal{T}_{\setminus 1}, \tag{6.1s}$$

$$\sum_{g\in\mathcal{A}_m^{\mathcal{G}}} p_{g,t} - \sum_{e\in\mathcal{A}_m^{\mathcal{E}}} \eta_e\ p_{e,t} - \sum_{u:(m,n)\in\mathcal{Z}} \left(q_{m,n,t}^{\text{in}} - q_{m,n,t}^{\text{out}}\right) = M_m^{\text{g,res}} D_t^{\text{g,res}} + M_m^{\text{g,ind}} D_t^{\text{g,ind}},\ \forall m \in \mathcal{M},\ \forall t \in \mathcal{T}, \tag{6.1t}$$

where the set of decision variables $\Theta = \left\{p_{e,t},\, p_{g,t},\, pr_{m,t},\, q_{m,n,t},\, q_{m,n,t}^{\text{in}},\, q_{m,n,t}^{\text{out}},\, h_{m,n,t}\right\}$. The sets respectively collect time periods $t \in \mathcal{T}$, electricity generating units $e \in \mathcal{E}$, wind power generators $w \in \mathcal{W}$, electricity demands $d \in \mathcal{D}$, electrical transmission lines $\ell \in \mathcal{L}$, gas suppliers $g \in \mathcal{G}$, gas network nodes $m \in \mathcal{M}$, gas network pipelines $(m,n) \in \mathcal{Z}$ (defined via their corresponding adjacent nodes) and compressors $(m,n) \in \mathcal{C}$.

The objective function (6.1a) minimizes the total cost of generating energy for the next day, in expectation for the most adversarial distribution $\mathbb{Q}$ of energy demand. Distribution $\mathbb{Q}$ is endogenously selected within the ambiguity set $\mathcal{A}$ by the inner

maximization operator. The cost of generating energy is calculated as the cost of generating electricity and supplying gas, where c_e is the marginal generation cost of generator e, c_g is the marginal cost of supplying gas for gas supplier g, $p_{e,t}$ represents the generation of electricity of generator e, and $p_{g,t}$ represents the gas supply from gas supplier g.

Equations (6.1b) to (6.1e) ensure the operational restrictions of the *power system*. In particular, equation (6.1b) balances the electricity generation of generating units $p_{e,t}$ and wind power generators $p_{w,t}$ with the residential and industrial electrical demands $D_t^{\text{e,res}}$ and $D_t^{\text{e,ind}}$, respectively. Inequalities (6.1c) and (6.1d) impose the operating limits, i.e., minimum and maximum generating capacities $\underline{P}_e$ and $\overline{P}_e$, of each electricity generating units $e \in \mathcal{E}$. Constraint (6.1e) restricts the maximum capacity $F_\ell^{\max}$ of each power transmission line $\ell \in \mathcal{L}$. The vectors $M_\ell^{\mathcal{E}} \in \mathbb{R}^{|\mathcal{E}|}$, $M_\ell^{\mathcal{W}} \in \mathbb{R}^{|\mathcal{W}|}$ and $M_\ell^{\mathcal{D}} \in \mathbb{R}^{|\mathcal{D}|}$ contain the power transmission distribution factors which map the power transmission in each line $\ell \in \mathcal{L}$ to the nodal injections, respectively for electricity generating units, wind power generators and electricity demands.

Equations (6.1f) to (6.1t) ensure the operational restrictions of the *gas system*. Inequalities (6.1f) and (6.1g) impose the minimum and maximum operating limits $\underline{P}_g$ and $\overline{P}_g$ for each gas supplier $g \in \mathcal{G}$. Constraints (6.1h) and (6.1i) enforce the minimum and maximum operating pressures $\underline{\text{PR}}$ and $\overline{\text{PR}}$ at each node $m \in \mathcal{M}$ in the gas grid. The pressure at each adjacent nodes $(m,n) \in \mathcal{C}$ of each compressor is constrained by the equations (6.1j) and (6.1k), involving the compression factor $\Gamma_{m,n}$. Equations (6.1l) to (6.1n) impose the direction of flows $q_{m,n,t}$, inflows $q_{m,n,t}^{\text{in}}$ and outflows $q_{m,n,t}^{\text{out}}$, respectively, inside each pipeline $(m,n) \in \mathcal{Z}$. Equality constraint (6.1o) corresponds to the Weymouth equation which gives the relation between the gas flow $q_{m,n,t}$ inside the pipeline and the pressures $pr_{m,t}$ and $pr_{n,t}$ at each adjacent nodes $(m,n) \in \mathcal{Z}$ given the Weymouth constant $K_{m,n}$ (dependent on the pipeline characteristics). Note that this relation relies on assumptions (i.e., isothermal and horizontal pipelines, unidirectional flows) which are usually non-limiting for the high pressure gas transmission system. Equation (6.1p) defines the gas flow as the average between inflow and outflow for each pipeline $(m,n) \in \mathcal{Z}$. The linepack $h_{m,n,t}$, i.e., the inherent storage of gas inside each pipeline, is defined via the set of equations (6.1q) to (6.1s). The linepack is linked to the pressures at input and output adjacent nodes by (6.1q), given the pipeline constant $S_{m,n}$. Equations (6.1r) and (6.1s) enforce the conservation of mass over time, where $H_{m,n}^0$ represents the initial condition of linepack. Finally, equation (6.1t) couples the power and gas systems by balancing the gas supplies $p_{g,t}$ and inflows $q_{m,n,t}^{\text{in}}$ with the gas consumption by gas-fired power plants $\eta_e \ p_{e,t}$, outflows $q_{m,n,t}^{\text{out}}$, residential gas demands $M_m^{\text{g,res}} D_t^{\text{g,res}}$, and industrial gas demands

$M_m^{\mathrm{g,ind}} D_t^{\mathrm{g,ind}}$. The mapping vectors $M^{\mathrm{g,res}} \in \mathbb{R}^{|\mathcal{M}|}$ and $M^{\mathrm{g,ind}} \in \mathbb{R}^{|\mathcal{M}|}$ allocate a fraction of the aggregate gas demand to each node. The constant η_e corresponds to the conversion factor of the gas-fired power plants. Note that the sets $\mathcal{A}_m^{\mathcal{G}}$ and $\mathcal{A}_m^{\mathcal{E}}$ collect respectively the gas suppliers and the gas-fired power plants connected to node m, while the set $u : (m, n) \in \mathcal{Z}$ collects the pipelines connected to node m.

It is worth mentioning that inequalities (6.1c) to (6.1n) are enforced as DRCCs. These probabilistic constraints enforce the operational restrictions with high probability for the worst-case distribution within the ambiguity set. We refer to Section 3.2.2 for more details about DRCCs. The modeling of demand uncertainty within DRCCs and within the objective function requires a careful attention and is further discussed in the following subsections 6.2.1 and 6.2.2.

6.2.1 Uncertainty Modeling

The day-ahead dispatch problem (6.1) is exposed to uncertainties in the demand side. We denote $\tilde{\xi} \in \mathbb{R}^{|\mathcal{K}|\times|\mathcal{T}|}$ the multivariate uncertain parameter composed of *i)* the deviation $\tilde{\xi}_{1,t}$ from residential power demand forecast $\overline{D_t^{\mathrm{e,res}}}$, *ii)* the deviation $\tilde{\xi}_{2,t}$ from industrial power demand forecast $\overline{D_t^{\mathrm{e,ind}}}$, *iii)* the deviation $\tilde{\xi}_{3,t}$ from residential gas demand forecast $\overline{D_t^{\mathrm{g,res}}}$ and, *iv)* the deviation $\tilde{\xi}_{4,t}$ from industrial gas demand forecast $\overline{D_t^{\mathrm{g,ind}}}$, as follows with $|\mathcal{K}| = 4$ sources of uncertainty $\forall t \in \mathcal{T}$:

$$D_t^{\mathrm{e,res}} = \overline{D_t^{\mathrm{e,res}}} + \tilde{\xi}_{1,t}, \qquad D_t^{\mathrm{e,ind}} = \overline{D_t^{\mathrm{e,ind}}} + \tilde{\xi}_{2,t}, \tag{6.2a}$$

$$D_t^{\mathrm{g,res}} = \overline{D_t^{\mathrm{g,res}}} + \tilde{\xi}_{3,t}, \qquad D_t^{\mathrm{g,ind}} = \overline{D_t^{\mathrm{g,ind}}} + \tilde{\xi}_{4,t}, \tag{6.2b}$$

where $D_t^{\mathrm{e,res}}$, $D_t^{\mathrm{e,ind}}$, $D_t^{\mathrm{g,res}}$, and $D_t^{\mathrm{g,ind}}$ respectively represent the real-time realizations of the aforementioned parameters and over-lined parameters $\overline{D_t^{\mathrm{e,res}}}$, $\overline{D_t^{\mathrm{e,ind}}}$, $\overline{D_t^{\mathrm{g,res}}}$, and $\overline{D_t^{\mathrm{g,ind}}}$ represent their day-ahead forecast. The multivariate uncertainty $\tilde{\xi}$ is embedded in the scheduling problem (6.1) via a moment-metric-based distributionally robust ambiguity set, which defines as

$$\mathcal{A} = \left\{ \mathbb{Q} \in \mathcal{M} \;\middle|\; \begin{array}{l} d_{\mathrm{W}}(\mathbb{Q}, \hat{\mathbb{Q}}_N) \leq \theta \\ \mathbb{E}^{\mathbb{Q}} \left[\left(\tilde{\xi} - \mu\right) \left(\tilde{\xi} - \mu\right)^{\top} \right] \preceq \Sigma \end{array} \right\}, \tag{6.3}$$

where the distribution $\mathbb{Q} \in \mathcal{M}$ describes the probability distribution function of the uncertain parameter $\tilde{\xi}$. The first constraint in (6.3) binds the Wasserstein distance[3] between the nominal distribution $\hat{\mathbb{Q}}_N$ (which consists of the collection

[3]We refer to the developments in Chapter 3 for further details about the definition of the

of N equiprobable forecasted scenarios $\widehat{\xi}_i \in \mathbb{R}^{|\mathcal{K}|\times|\mathcal{T}|}\ \forall i \in \{1,...,N\}$) and the distribution $\mathbb{Q}$ which is endogenously selected within the ambiguity set as the one that mostly affects the real-time operating cost. The parameter $\theta \in \mathbb{R}^+$ represents the Wasserstein radius. The second constraint in (6.3) implies that the distributions within the ambiguity set have a covariance matrix which is consistent with the empirically observed one $\Sigma \in \mathbb{R}^{|\mathcal{K}||\mathcal{T}|\times|\mathcal{K}||\mathcal{T}|}$. This constraint therefore enhances the accuracy of the uncertainty modeling by excluding the unrealistic distributions whose covariance matrix (i.e., dependence structure) is not representative of the observed data, see our developments in Chapter 5. This is particularly important for multivariate and correlated demand uncertainty impacting the coordinated power and gas dispatch.

To properly address this uncertainty, problem (6.1) usually embeds two decision stages, i.e., a day-ahead scheduling based on available forecasts and, a real-time balancing stage which accounts for the real-time realization of uncertainty. In real-time, when a deviation occurs, recourse actions are undertaken to ensure the safe operation of the system. To model these real-time actions, we use affine decision rules [101], as follows:

$$\begin{aligned}
&p_{e,t} = \overline{p_{e,t}} + Y_{e,t}^{\top}\tilde{\xi}_t, && p_{g,t} = \overline{p_{g,t}} + V_{g,t}^{\top}\tilde{\xi}_t,\\
&pr_{m,t} = \overline{pr_{m,t}} + \rho_{m,t}^{\top}\tilde{\xi}_t, && q_{m,n,t} = \overline{q_{m,n,t}} + \gamma_{m,n,t}^{\top}\tilde{\xi}_t,\\
&q_{m,u,t}^{\text{in}} = \overline{q_{m,n,t}^{\text{in}}} + \gamma_{m,n,t}^{\text{in}\top}\tilde{\xi}_t, && q_{m,n,t}^{\text{out}} = \overline{q_{m,n,t}^{\text{out}}} + \gamma_{m,n,t}^{\text{out}\top}\tilde{\xi}_t,
\end{aligned} \tag{6.4}$$

where $Y_{e,t}$, $V_{e,t}$, $\rho_{m,t}$, $\gamma_{m,n,t}$, $\gamma_{m,n,t}^{\text{in}}$, $\gamma_{m,n,t}^{\text{out}} \in \mathbb{R}^{|\mathcal{K}|}$ are supplementary decision variables and over-lined variables $\overline{p_{e,t}}$, $\overline{p_{g,t}}$, $\overline{pr_{m,t}}$, $\overline{q_{m,n,t}}$, $\overline{q_{m,n,t}^{\text{in}}}$ and $\overline{q_{m,n,t}^{\text{out}}}$ represent day-ahead scheduling decisions[4]. For instance, the quantity $Y_{e,t}^{\top}\tilde{\xi}_t$ corresponds to the deviation in power generation of generator e at time step t. Incorporating the affine decision rules (6.2) and (6.4) makes the problem (6.1) two-stage and uncertainty-aware. Hence, it requires reformulation to retrieve a tractable model. We begin with the power balance equation (6.1b) which may be divided into its day-ahead and real-time counterparts, as follows:

$$\sum_{e=1}^{|\mathcal{E}|}\overline{p_{e,t}} + \sum_{w=1}^{|\mathcal{W}|}\overline{\Omega_{w,t}} = \overline{D_t^{\text{e,res}}} + \overline{D_t^{\text{e,ind}}},\ \ \forall t \in \mathcal{T}, \tag{6.5a}$$

$$\sum_{e=1}^{|\mathcal{E}|} Y_{1,e,t} = 1, \sum_{e=1}^{|\mathcal{E}|} Y_{2,e,t} = 1, \sum_{e=1}^{|\mathcal{E}|} Y_{3,e,t} = 0, \sum_{e=1}^{|\mathcal{E}|} Y_{4,e,t} = 0,\ \ \forall t \in \mathcal{T}. \tag{6.5b}$$

Wasserstein probability metric.

[4]Affine decision rules are known to provide an approximation of recourse actions, which allows to improve the performance of dispatch decisions by incorporating an insight of real-time operation into the day-ahead problem, while avoiding complex multi-stage formulations.

Equation (6.5a) ensures the day-ahead balance while equations (6.5b) ensure the real-time balance, where the action of all generator compensates the deviation in electricity demand (the two first equations) and compensates the participation of other electricity generating units when a deviation in gas demand occurs (the two last equations). Besides, Weymouth equation (6.1o) may be cast into the following set of equations:

$$\overline{q_{m,n,t}}^2 = K_{m,n}^2 \left(\overline{pr_{m,t}}^2 - \overline{pr_{n,t}}^2 \right) \ \forall (m,n) \in \mathcal{Z}, \ \forall t \in \mathcal{T}, \tag{6.6a}$$

$$\gamma_{m,n,t}^2 = K_{m,n}^2 \left(\rho_{m,t}^2 - \rho_{n,t}^2 \right) \ \forall (m,n) \in \mathcal{Z}, \ \forall t \in \mathcal{T}, \tag{6.6b}$$

$$\gamma_{m,n,t} q_{m,n,t} = K_{m,n}^2 \left(\rho_{m,t} pr_{m,t} - \rho_{n,t} pr_{n,t} \right) \ \forall (m,n) \in \mathcal{Z}, \ \forall t \in \mathcal{T}, \tag{6.6c}$$

which follows after the developments in [147]. Note that equations (6.6a) and (6.6b) are non-convex quadratic equality contraints. We relax these using SOC relaxation[5], which consist of transforming the equality in a less-or-equal-to inequality constraint, ending up in a convex conic program. Note also that equation (6.6c) contains bilinear terms and is reformulated using McCormick relaxation, as proposed in [147]. Equation (6.1p) can be cast into

$$\overline{q_{m,n,t}} = \frac{\overline{q_{m,n,t}}^{\text{in}} + \overline{q_{m,n,t}}^{\text{out}}}{2}, \ \ \forall (m,n) \in \mathcal{Z} \ \ \forall t \in \mathcal{T}, \tag{6.7}$$

$$\gamma_{m,n,t} = \frac{\gamma_{m,n,t}^{\text{in}} + \gamma_{m,n,t}^{\text{out}}}{2}, \ \ \forall (m,n) \in \mathcal{Z} \ \ \forall t \in \mathcal{T}, \tag{6.8}$$

Regarding the linepack formulation (6.1q) to (6.1s), these constraints can be restated in terms of day-ahead forecast and completed with these additional constraints:

$$S_{m,n} \frac{\rho_{m,t} + \rho_{n,t}}{2} = \gamma_{m,n,t}^{\text{in}} - \gamma_{m,n,t}^{\text{out}}, \ \forall (m,n) \in \mathcal{Z} \ \ \forall t = 1, \tag{6.9}$$

$$S_{m,n} \frac{\rho_{m,t} + \rho_{n,t} - \rho_{m,t-1} - \rho_{n,t-1}}{2} = \gamma_{m,n,t}^{\text{in}} - \gamma_{m,n,t}^{\text{out}}, \ \forall (m,n) \in \mathcal{Z} \ \ \forall t \in \mathcal{T}_{\setminus 1}. \tag{6.10}$$

Finally, the coupling constraint pertaining to gas balancing is equivalently cast into:

[5]It is worth mentioning that convex relaxations of the Weymouth equation may have an impact on the feasibility of the obtained solution. We discuss this important issue in our numerical experiments and propose an innovative approach for addressing it in Chapter 7.

$$\sum_{g\in\mathcal{A}_m^{\mathcal{G}}} \overline{p}_{g,t} - \sum_{e\in\mathcal{A}_m^{\mathcal{E}}} \eta_e \overline{p}_{e,t} - \sum_{\substack{u:(m,n)\\ \in\mathcal{Z}}} \left(\overline{q}^{\text{in}}_{m,n,t} - \overline{q}^{\text{out}}_{m,n,t}\right) = M_m^{\text{g,res}}\overline{D_t^{\text{g,res}}} + M_m^{\text{g,ind}}\overline{D_t^{\text{g,ind}}}, \tag{6.11a}$$

$$\sum_{g\in\mathcal{A}_m^{\mathcal{G}}} V_{k,g,t} - \sum_{e\in\mathcal{A}_m^{\mathcal{E}}} \eta_e Y_{k,e,t} - \sum_{\substack{u:(m,n)\\ \in\mathcal{Z}}} \left(\gamma^{\text{in}}_{k,m,n} - \gamma^{\text{out}}_{k,m,n}\right) = B_k \quad \forall k\in\mathcal{K},\ \forall m\in\mathcal{M},\ \forall t\in\mathcal{T}, \tag{6.11b}$$

where $B = \begin{pmatrix} 0 & 0 & M_m^{\text{g,res}} & M_m^{\text{g,ind}} \end{pmatrix}^\top$.

6.2.2 Distributionally Robust Framework

The final distributionally robust model for the coordinated power and gas dispatch consists of (6.1) with the incorporation of additional constraints (6.4) to (6.11). The program has a min-max structure including probability operators in the constraints, and requires conic reformulations to be solved by off-the-shelf solvers. We begin with the objective function:

$$\min_{\Theta} \sum_{t=1}^{|\mathcal{T}|} \left(\sum_{e=1}^{|\mathcal{E}|} c_e \overline{p_{e,t}} + \sum_{g=1}^{|\mathcal{G}|} c_g \overline{p_{g,t}} \right) + \max_{\mathbb{Q}\in\mathcal{A}} \mathbb{E}^{\mathbb{Q}} \left[\sum_{t=1}^{|\mathcal{T}|} \left(\sum_{e=1}^{|\mathcal{E}|} c_e Y_{e,t}^\top \tilde{\xi}_t + \sum_{g=1}^{|\mathcal{G}|} c_g V_{g,t}^\top \tilde{\xi}_t \right) \right], \tag{6.12}$$

where the use of affine decision rules is made explicit. Based on the developments in Chapter 5 and assuming the availability of N empirical observations $\hat{\xi}_i$ of the uncertain parameter $\tilde{\xi}$, we derive the reformulation of the worst-case expectation problem (inner maximization), given the moment-metric-based ambiguity set $\mathcal{A}$ as follows:

$$\max_{\mathbb{Q}\in\mathcal{A}} \mathbb{E}^{\mathbb{Q}} \left[\sum_{t=1}^{|\mathcal{T}|} \left(\sum_{e=1}^{|\mathcal{E}|} c_e Y_{e,t}^\top \tilde{\xi}_t + \sum_{g=1}^{|\mathcal{G}|} c_g V_{g,t}^\top \tilde{\xi}_t \right) \right] = \tag{6.13a}$$

$$\begin{cases} \min\limits_{\lambda_t,\sigma_{i,t},\Lambda_t\succeq 0,\zeta_{i,t}} \sum\limits_{t=1}^{|\mathcal{T}|} \lambda_t\theta + \langle \Lambda_t, \Sigma_t\rangle_{\text{F}} + \dfrac{1}{N}\sum\limits_{i=1}^{N}\sigma_{i,t} \\ \text{s.t. } \mathcal{F}_{i,t} \succeq 0 \ \ \forall i\in\{1,...,N\} \ \ \forall t\in\mathcal{T} \\ \quad\ \ ||\zeta_{i,t}||_* \le \lambda_t \ \ \forall i\in\{1,...,N\} \ \ \forall t\in\mathcal{T} \end{cases} \tag{6.13b}$$

where $\lambda_t \in \mathbb{R}$, $\sigma_{i,t} \in \mathbb{R}$, $\Lambda_t \in \mathbb{R}^{|\mathcal{K}|\times|\mathcal{K}|}$, and $\zeta_{i,t} \in \mathbb{R}^{|\mathcal{K}|}$ are auxiliary variables. The operator $\langle \Lambda_t, \Sigma_t \rangle_F = \sum_{i,j}^{|\mathcal{K}|} \Lambda_{ijt}\Sigma_{ijt}$ refers to the Frobenius inner product between matrices Λ_t and Σ_t. The operator $\succeq$ requires that the matrix on the left-hand side is positive semidefinite, resulting in a SDP program. The matrices $\mathcal{F}_{i,t}$ define as follows:

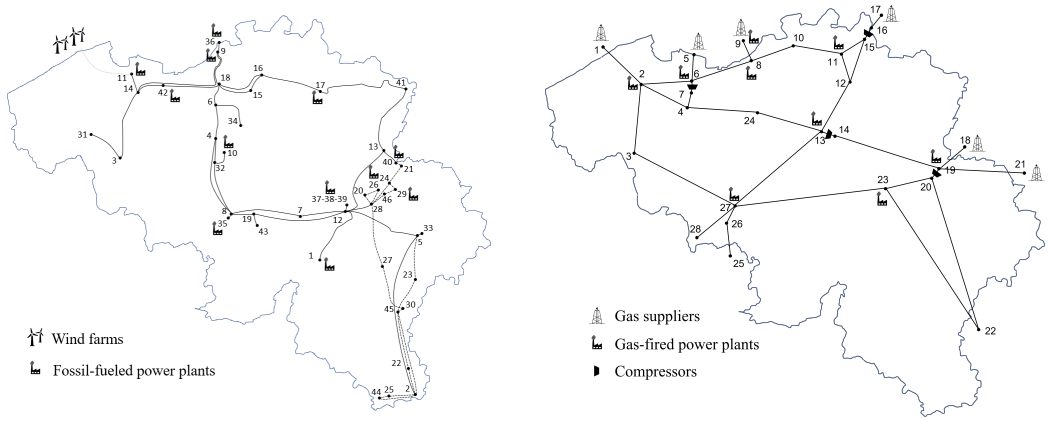
$$\mathcal{F}_{i,t} = \begin{bmatrix} \Lambda_t & \frac{1}{2}\left(\zeta_{i,t} - c_e Y_{e,t} - c_g V_{g,t}\right) \\ \frac{1}{2}\left(\zeta_{i,t} - c_e Y_{e,t} - c_g V_{g,t}\right)^\top & \sigma_{i,t} - \zeta_{i,t}^\top \hat{\xi}_{i,t} \end{bmatrix}. \tag{6.14}$$

The inner and outer minimization operators may next be merged, resulting in a global minimization problem. Second, we reformulate each DRCCs (6.1c) to (6.1n) using the CVaR approximation. The worst-case expectation problems appearing in the CVaR constraints are then reformulated given the moment-metric-based ambiguity set and model (6.13). We refer to Chapter 3 regarding the complete methodology to approximate DRCCs with CVaR constraints. For the sake of conciseness, we also refer to [65] for the final model which we exploit in the numerical analysis of Section 6.3.

6.3 Numerical Experiments

In this section, we perform numerical experiments to show the performance of the proposed distributionally robust coordinated power and gas dispatch problem (6.1). Our case study is inspired by the real-life Belgian high voltage power and high pressure gas systems. The corresponding electricity and natural gas networks are schematically represented in Figs. 6.2(a)-6.2(b) and the exhaustive list of economical and technical parameters is provided in Appendix B. On the one hand, the power system is composed of 46 electrical nodes and 69 transmission lines which connect 2 nuclear power plants for a total capacity of 6,000 MW, 10 natural-gas-fired power plants for a total capacity of 3,329 MW and 12 wind farms for a total capacity of 3,794 MW (of which 2,570 MW offshore and 1,224 MW onshore) with the electricity demand centers. On the other hand, the natural gas system is composed of 28 gas nodes and 33 gas pipelines, which connect 6 gas suppliers for a total capacity[6] of 11.2 MNm^3 with the 10 gas-fired power plants (which create the link between power and gas systems) and the gas demand centers. In addition, four compressors ensure the safe operation of the gas grid. The electricity and gas demand forecasts, as well as the wind power generation forecast are shown in Fig. 6.3(a).

[6]In this work, the gas flows are expressed in MNm3 in the standard conditions of temperature and pressure (i.e., 0° C and 1.013 bar).



(a) Belgian electrical grid topology (plain lines are 380 kV and dashed lines are 220 kV).

(b) Belgian gas grid topology.

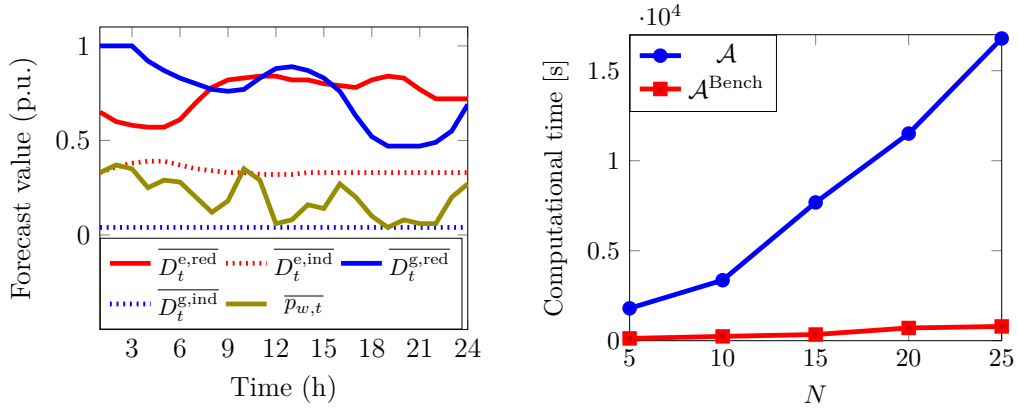
Figure 6.2: The real-life Belgian power and gas systems and related information.

The empirical distribution $\hat{\mathbb{Q}}_N$ of power and gas demand deviations is defined via historical hourly observations of power and natural gas demands retrieved from ELIA [149] and ENTSOG transparency platform [150], regarding the year 2019. We split both power and gas demands into residential and industrial compounds, resulting in a four-dimensional uncertainty for the 24 hours of the day. To feed the program with the appropriate data, we focus on 63 winter weekdays and use the probabilistic distribution composed of 15 randomly selected days. We calculate the mean over these 15 days, and use it as day-ahead forecast. Then, we retrieve the mean from the 15 scenarios and use the resulting set as in-sample scenarios for the real-time demand deviation with respect to its forecast. We calculate the covariance matrix Σ based on the in-sample scenarios to appropriately parametrize the moment-metric-based ambiguity set. Finally, the 48 remaining days are used as out-of-sample unseen scenarios to evaluate the technical and economical performances of the optimal scheduling.

Using the aforedescribed dataset as in-sample scenarios, we solve the distributionally robust coordinated day-ahead power and gas dispatch problem (6.1), given the moment-metric-based ambiguity set for different values of the Wasserstein radius $\theta = \{0.1, 0.05, 0.01, 0.005, 0.001\}$ and different values of violation probability $\epsilon = \{0.03, 0.05\}$. As a benchmark solution to compare our results, we solve the same problem (6.1), given the traditional metric-based ambiguity set

$$\mathcal{A}^{\text{Bench}} = \left\{ \mathbb{Q} \in \mathcal{M} \middle| d_W(\mathbb{Q}, \hat{\mathbb{Q}}_N) \leq \theta \right\}, \tag{6.15}$$

which does not account for the inherent dependencies among power and gas



(a) The hourly electricity and gas demands as well as the hourly wind power forecast, in per-unit.

(b) Computational time as a function of the number of in-sample scenarios N. Fixed parameters: $\epsilon = 0.05$ and $\theta = 0.001$.

Figure 6.3: Numerical study.

Table 6.1: Out-of-sample simulation results

		$\mathcal{A}$			$\mathcal{A}^{\text{Bench}}$		
ϵ	θ	Expected cost [€]	Standard deviation [€]	EENS [MW]	Expected cost [€]	Standard deviation [€]	EENS [MW]
0.03	0.001	1,050,114	120,328	0.0	1,120,326	138,207	14.24
	0.005	1,049,912	120,323	0.0	1,014,992	126,395	0.7
	0.01	1,050,171	120,326	0.0	1,143,889	138,081	21.0
	0.05	1,050,261	120,329	0.0	1,168,420	137,889	19.6
	0.1	1,049,925	120,324	0.0	1,256,392	138,144	31.8
0.05	0.001	1,032,610	120,200	0.0	1,455,122	138,251	55.5
	0.005	1,050,636	120,332	0.0	1,027,362	132,722	2.29
	0.01	1,046,249	120,300	0.0	1,265,150	138,344	35.8
	0.05	1,050,040	120,327	0.0	1,301,873	138,149	37,5
	0.1	1,050,264	120,327	0.0	1,986,344	137,997	124.6

demands. The resulting linear formulation is provided in [65]. The models are solved using Julia Programming Language v1.4.2, JuMP modeling language v0.21.3 and Mosek solver v9.2 installed on a quad-core computer clocking at 2.2 GHz. In the remainder of this section, we perform a computational time analysis in Section 6.3.1, and an out-of-sample analysis of the total operational cost and the Expected Energy Not Served (EENS) in Sections 6.3.2 and 6.3.3, respectively. Finally, we discuss the ex-post feasibility of the obtained dispatch solution in Section 6.3.4.

6.3.1 Computational Time

We run the moment-metric-based and traditional metric-based distributionally robust models for different number of in-sample scenarios, i.e., $N = \{5, 10, 15, 20, 25\}$, and for fixed values of $\theta = 0.001$ and $\epsilon = 0.05$. We report the computational time in Fig. 6.3(b). We observe a linear increasing trend for both approaches and a higher computational time for the proposed SDP formulation comparatively to the benchmark linear formulation. Note that we have not observed a significant variation of computational time when θ varies, for both the moment-metric-based and traditional metric-based models.

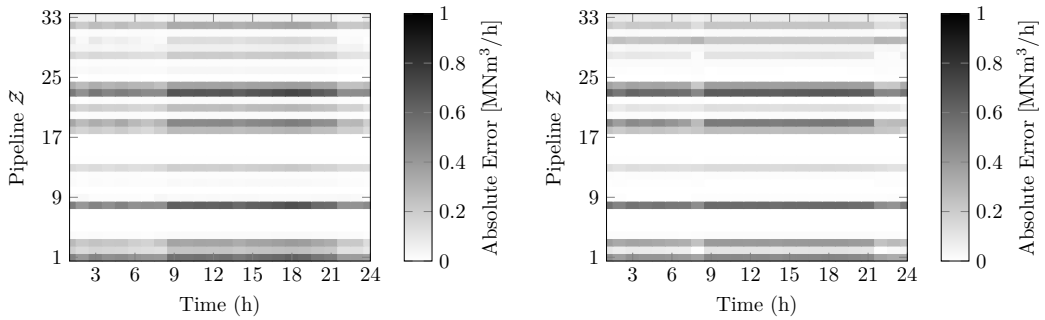
6.3.2 Out-of-sample Operational Cost

In this section, we solve the distributionally robust coodinated power and gas dispatch problem (6.1) and retrieve the optimal dispatch solutions $\overline{p_{e,t}}$ and $\overline{p_{g,t}}$, for both the moment-metric-based and traditional metric-based models, and for different values of $\theta = \{0.001, 0.005, 0.01, 0.05, 0.1\}$ and $\epsilon = \{0.03, 0.05\}$. Then, we use the remaining 48 scenarios as unseen realizations of uncertainty and solve a real-time operational re-scheduling problem[7], once for each out-of sample scenario. By doing so, we are able to calculate the total system operating cost (including the cost of load shedding) for each scenario and to report the expected value and standard deviation in Table 6.1. The lowest values in terms of cost are highlighted in blue. We observe that the proposed moment-metric-based model achieves, in general, lower total expected cost and standard deviation compared to the traditional metric-based model. For $\theta = 0.005$, the traditional approach achieves lower expected cost, but at the expense of a higher standard deviation[8].

6.3.3 Expected Energy Not Served

We calculate the Expected Energy Not Served (EENS), as the average over the out-of-sample scenarios of the load shedding actions required for balancing the energy system over the day. This quantity is indirectly related to the violation of DRCCs, e.g., in the case where power units have attained their maximum generation capacities but the deviation has not been completely compensated. Interestingly, the outcomes illustrate that, by properly considering the dependence structure among the uncertain demand deviations (using a moment-metric-based ambiguity set), the EENS is reduced to very low values compared to the traditional approach.

[7]We provide the formulation of the real-time re-scheduling problem in [65].
[8]In practice, it is not straightforward to *a priori* select the optimal value for θ.



(a) $\theta = 0.1$. (b) $\theta = 0.001$.

Figure 6.4: Matrix plot of the relative error between right-hand and left-hand sides of relaxed Weymouth equations for each time step (x-axis) and each pipeline (y-axis). Fixed parameters: $N = 10$, $\epsilon = 0.05$.

6.3.4 Ex-post feasibility

In this section, we aim to evaluate the ex-post feasibility of the solution obtained by the proposed moment-metric-based distributionally robust framework. To do so, we solve the model for two particular values of θ, i.e., 0.001 and 0.1, and retrieve the state variables $q_{m,n,t}$ and $pr_{m,t}$ at optimality. We calculate the mismatch with respect to the Weymouth equation (6.1o), as the absolute error between the right-hand side and left-hand side, i.e., $\Delta_{m,n,t} = |q^2_{m,n,t} - K^2_{m,n}\left(pr^2_{m,t} - pr^2_{n,t}\right)|$ and report the value of absolute errors in the matrix plot of Fig. 6.4. Interestingly, the Mean Absolute Error (MAE) does not vary significantly when the value of Wasserstein radius varies (MAE = 0.12 in both cases). The results suggest that the obtained distributionally robust solutions are sufficiently tight to validate the benefits of the moment-metric-based DRO over the traditional approach. Nonetheless, we highlight that a real-life implementation would require further improvement of ex-post feasibility. Intuitively speaking, these errors mainly stem from the technique which is used to relax the Weymouth equation, which opens the path towards improved modeling techniques for gas flow dynamics.

6.4 Chapter Summary and Conclusions

In this chapter, we propose a moment-metric-based distributionally robust model for the day-ahead coordinated power and gas dispatch where the uncertainty pertains to the multivariate and correlated demand. The numerical experiments show promising results in terms of cost-optimality and reliability, by appropriately considering the dependence structure among the uncertain demands. The focus on a case study inspired by the Belgian power and natural gas systems also sheds

light on the scalability of the approach.

The results also reveal that research questions still need to be addressed regarding the improvement of modeling techniques for gas networks. This challenge is the scope of the following chapter where we investigate neural-network-constrained optimization methods for modeling gas flow dynamics within the coordinated power and gas dispatch problem.

CHAPTER 7

Machine Learning for Improved Gas Network Models

The current energy transition promotes the convergence of operation between the power and natural gas systems. In that direction, it becomes paramount to improve the modeling of non-convex natural gas flow dynamics within the coordinated power and gas dispatch. In this work, we propose a neural-network-constrained optimization method which includes a regression model of the Weymouth equation, based on supervised machine learning. The Weymouth equation links gas flow to inlet and outlet pressures for each pipeline via a quadratic equality, which is captured by a neural network. The latter is encoded via a tractable mixed-integer linear program into the set of constraints. In addition, the proposed framework is capable of considering bidirectionality without having recourse to complex and potentially inaccurate convexification approaches. We further enhance our model by introducing a reformulation of the activation function, which improves the computational efficiency. An extensive numerical study based on the real-life Belgian power and gas systems shows that the proposed methodology yields promising results in terms of accuracy and tractability.

The content of this chapter is mainly based on the following publication:

[G] **A. Arrigo**, M. Dolányi, K. Bruninx, J.-F. Toubeau, "Machine Learning for Improved Gas Network Modeling in Coordinated Energy Systems," Working paper[1].

[1]This work is foreseen to be submitted as a journal paper by October 2022.

7.1 Motivation

The need to address climate change is triggering major transitions in the energy sector, one of which pertains to the increased integration of renewables in energy systems. In that context, Natural Gas-Fired Power Plants (NGFPPs) are becoming key flexible resources for cost-effectively managing the uncertainty and variability of weather-dependent renewable energy sources [3]. Moreover, the emergence of power-to-gas technologies that use electrical power to produce synthetic hydrogen or methane, paves the way towards the procurement of additional flexibility from gas storage, e.g., inside pipelines (which is referred to as linepack) or in gas tanks [151]. In light of these inherent techno-economic dependencies between power and gas systems, there is an increasing need to improve the coordination of their operation [152].

In this direction, several coordination schemes with different degrees of integration have been investigated in the literature [45]. A first solution, which is regulatory-compliant as it does not introduce a new market mechanism, consists in making the day-ahead electricity market aware of gas network restrictions [46]. A second potential solution, which increases the level of coordination between both energy systems, refers to the exchange of information (e.g., either based on volume or price quantities) between the electricity and gas market operators [142]. Alternatively, the *fully* coordinated day-ahead power and gas dispatch [153], which is the focus of this chapter, refers to a framework where a unique *system operator* jointly dispatches the power and gas generating units. The resulting decision-making problem for system operator aims at minimizing the total day-ahead scheduling cost of both systems, while ensuring the demand- and supply-side operating constraints and network limitations.

A major challenge which *all* these works must contend with relates to the modeling of natural gas flow dynamics. In particular, the relationship between pressures and flow levels alongside the gas pipelines is described via a quadratic equality, the so-called Weymouth equation, which results in highly non-linear and non-convex constraints. In this work, to address this issue, we propose a neural-network-constrained coordinated day-ahead power and gas dispatch[2], which includes a regression model of the Weymouth equation (based on supervised machine learning) resulting in a tractable Mixed-Integer Linear Program (MILP).

Different approaches exist in the literature to cope with the non-convexity of the

[2]In this work, we apply our proposed neural-network-constrained framework in a fully coordinated setting. It is worth emphasizing that this is not a limiting assumption, as our focus is on improving the modeling of gas flow dynamics and it is straightforward to incorporate our proposed framework within other coordination settings.

Weymouth equation, which mainly rely on convexification techniques[3]. Authors in [49] use the Second-Order Cone (SOC) relaxation, which transforms the equality into an inequality constraint. However, this loose relaxation requires considering the whole interior of the quadratic cone instead of its envelope, which strongly jeopardizes the feasibility of the obtained solution. Hence, the SOC relaxation is supplemented with McCormick envelopes in [50], which allow to exclude those solutions that yield a large feasibility gap with respect to the original Weymouth relation. To further improve the quality of the relaxation, authors in [51] propose tightening the bounds of the underlying McCormick envelopes, via an iterative algorithm which finds the smallest bounds around the operating point. However, this iterative procedure must be run online and results in a significant increase in computational burden. Generally speaking, all these SOC-based relaxation techniques are not sufficiently tight for practical purpose, and it is not straightforward to enhance these models for considering gas flow bidirectionality, see [54]. Differently, authors in [**?**] and [52] consider an interpolation of the Weymouth equation, via an incremental piecewise linear formulation, whose accuracy highly depends on the number of intervals selected to describe the non-linear function. The size of the resulting MILP increases with the number of intervals, therefore ensuing a trade-off between the approximation error and the computational burden [53]. Overall, the previously explored convexification techniques still need to address fundamental research questions in terms of accuracy and tractability.

In order to reduce the inherent modeling errors arising from these convexification techniques, machine-learning-based optimization methods have been explored in the power system literature, with a special focus on Optimal Power Flow (OPF) problems[4]. In particular, authors in [55] present a regression-based proxy of the solution to a real-time OPF, which is tailored to be embedded within long-term planning methodologies. Different regression tools are compared, shedding light on the appealing performances achieved by neural networks. Following the same rationale, authors in [56] and [57] leverage a deep neural network architecture which calculates the optimal solution to a security-constrained DC-OPF and AC-OPF, respectively. These approaches reveal a substantial computational speed-up, however, their application in the scope of the power and gas dispatch problem is hindered by the number of operating points required in the training phase to yield an accurate model of the overall operating feasible space. To

[3]Convexification techniques are of two main types [48], namely, convex *relaxations* (which enlarge the original feasible set), and convex *approximations* (which may disregard some feasible solutions).

[4]The OPF problems are exposed to non-convexities arising from, e.g., the modeling of AC power flows. The techniques (either based on convexification or machine learning approaches) to address these non-convexities are close to those required by the Weymouth equation, therefore making it important to review these works.

address this issue, several additional works consider the incorporation of machine learning models into the set of constraints. In particular, authors in [154] aim at predicting the set of active constraints within the DC-OPF problem and propose a solution approach based on ensemble method to reduce the size of problem. Differently, the authors in [58] propose a classification neural-network-constrained AC-OPF encoding dynamic security restrictions. To the best of our knowledge, reference [155] represents the first effort in the literature to implement a machine learning proxy within the power and gas coordinated energy systems[5]. The authors leverage decisions trees as classifiers defining operating security regions.

In this work, aiming at improving the modeling of non-convex gas flow dynamics, we propose a new neural-network-constrained optimization method. To that end, a neural network is trained offline in a fully data-driven way. In contrast to existing convex relaxation and approximation techniques, this training phase results in a *regression* model linking the gas flow to inlet and outlet pressures for each pipeline, which minimizes the mean square error from the operating data points. Unlike machine learning models dedicated at directly finding the optimal decision set, our neural network is encoded via a MILP and embedded within the set of constraints, yielding a tractable, modular and comprehensible dispatch model. The proposed framework is further enhanced with a reformulation of activation function, which improves computational efficiency. Overall, the contributions of this work are as follows:

(i) we formulate a neural-network-constrained coordinated day-ahead power and gas dispatch problem, including a regression model of the Weymouth equation, based on supervised machine learning. The resulting MILP provides a flexible framework enabling the consideration of bidirectional natural gas flow dynamics, without relying on any modeling assumption,

(ii) we improve the computational efficiency of the proposed approach by introducing a new reformulation of the activation function within the neural-network constrained power and gas dispatch problem.

Our numerical analysis based on a case study inspired by the Belgian power and gas systems stresses that the neural-network-constrained framework outperforms the existing SOC-based convexification techniques and provides comparable performances with respect to piecewise linear approximation for a similar number of mixed-integer variables involved in the optimization problem.

[5]It is worth noting that decision trees have also been applied to OPF problems [156]-[157].

7.2 The Neural-Network-Constrained Power and Gas Dispatch

In this section, we briefly recall the coordinated power and gas dispatch problem formulation in Section 7.2.1 and introduce the neural-network-constrained framework in Section 7.2.2.

7.2.1 The Coordinated Power and Gas Dispatch

In the fully coordinated framework, the system operator aims at deriving a cost-efficient day-ahead scheduling of the electrical power and natural gas in-feeds. The dispatch solution must ensure operational restrictions pertaining to *i)* the electrical network, *ii)* the gas network, and *iii)* the coupling of both, which are described in the following.

Electrical Network Constraints

The electrical network connects conventional power generating units $e \in \mathcal{E}$, wind power generating units $w \in \mathcal{W}$, and electrical power demand centers $d \in \mathcal{D}$, via high voltage transmission lines $\ell \in \mathcal{L}$. Note that the set of power generating units also contains NGFPPs. The safe operation of electrical power network is ensured by the following set of constraints:

$$\sum_{e\in\mathcal{E}} p_{e,t} + \sum_{w\in\mathcal{W}} p_{w,t} = \sum_{d\in\mathcal{D}} p_{d,t}, \ \ \forall t \in \mathcal{T}, \tag{7.1a}$$

$$\underline{P}_e \leq p_{e,t} \leq \overline{P}_e, \ \ \forall e \in \mathcal{E}, \ t \in \mathcal{T}, \tag{7.1b}$$

$$M_\ell^{\mathcal{E}} p_{e,t} + M_\ell^{\mathcal{W}} p_{w,t} - M_\ell^{\mathcal{D}} p_{d,t} \leq F_\ell^{\max}, \ \ \forall \ell \in \mathcal{L}, \ t \in \mathcal{T}. \tag{7.1c}$$

The constraint (7.1a) balances the total power generation $p_{e,t}$ and $p_{w,t}$ from the conventional and wind power generating units with the power demand $p_{d,t}$ from demand centers[6]. The minimum and maximum operating limits $\underline{P}_e$ and $\overline{P}_e$ for each conventional power generating unit is enforced via (7.1b). In (7.1c), a DC load flow model [158] is leveraged for imposing the transmission line capacity limits $F_\ell^{\max}$, using power transfer distribution factors, which map the contribution to the power flow in each line $\ell \in \mathcal{L}$ with the nodal injections. The latter are collected within $M_\ell^{\mathcal{E}}$, $M_\ell^{\mathcal{W}}$ and $M_\ell^{\mathcal{D}}$, for conventional power generating units $e \in \mathcal{E}$, wind power generating units $w \in \mathcal{W}$ and electrical power demand centers, respectively.

[6] In this work, we assume inelastic demand and deterministic wind power generation. It is worth mentioning that other works in the literature explore uncertainty-aware scheduling, such as, [144] and [147].

Gas Network Constraints

The gas network connects natural gas suppliers $g \in \mathcal{G}$ to natural gas demand centers $b \in \mathcal{B}$, via a network comprising of high pressure nodes $m \in \mathcal{M}$, high pressure gas pipelines $(m,n) \in \mathcal{Z}$ and compressors $(m,n) \in \mathcal{C}$, which are defined via their corresponding adjacent nodes m and n. The safe operation of the natural gas network is ensured by the following set of constraints:

$$\underline{P}_g \leq p_{g,t} \leq \overline{P}_g, \ \ \forall g \in \mathcal{G}, \ t \in \mathcal{T}, \tag{7.2a}$$
$$\underline{\mathrm{PR}}_m \leq pr_{m,t} \leq \overline{\mathrm{PR}}_m \ \ \forall m \in \mathcal{M}, \ t \in \mathcal{T}, \tag{7.2b}$$
$$pr_{n,t} \leq pr_{m,t} \leq \Gamma_{m,n} pr_{n,t}, \ \forall (m,n) \in \mathcal{C}, \ t \in \mathcal{T}, \tag{7.2c}$$
$$q^2_{m,n,t} = K^2_{m,n} \left(pr^2_{m,t} - pr^2_{n,t} \right), \ \ \forall (m,n) \in \mathcal{Z}, \ t \in \mathcal{T}, \tag{7.2d}$$
$$q_{m,n,t} = \frac{q^{\mathrm{in}}_{m,n,t} + q^{\mathrm{out}}_{m,n,t}}{2} \ \ \forall (m,n) \in \mathcal{Z}, \ t \in \mathcal{T}, \tag{7.2e}$$
$$h_{m,n,t} = S_{m,n} \frac{pr_{m,t} + pr_{n,t}}{2} \ \ \forall (m,n) \in \mathcal{Z}, \ t \in \mathcal{T}, \tag{7.2f}$$
$$h_{m,n,t} = H^0_{m,n} + q^{\mathrm{in}}_{m,n,t} - q^{\mathrm{out}}_{m,n,t} \ \ \forall (m,n) \in \mathcal{Z}, \ \ t = 1, \tag{7.2g}$$
$$h_{m,n,t} = h_{m,n,t-1} + q^{\mathrm{in}}_{m,n,t} - q^{\mathrm{out}}_{m,n,t} \forall (m,n) \in \mathcal{Z}, \ t \in \mathcal{T}_{\setminus \{1\}}. \tag{7.2h}$$

The constraint (7.2a) enforces the minimum and maximum operating limits $\underline{P}_g$ and $\overline{P}_g$ for gas supply $p_{g,t}$. The constraint (7.2b) enforces the nodal pressures $pr_{m,t}$ to lie within their minimum and maximum thresholds $\underline{\mathrm{PR}}_m$ and $\overline{\mathrm{PR}}_m$. In addition, the pressure at each adjacent nodes $(m,n) \in \mathcal{C}$ of each compressor is constrained by the constraint (7.2c), involving the compression factor $\Gamma_{m,n}$. Equality constraint (7.2d) corresponds to the Weymouth equation which provides the relation between the flow $q_{m,n,t}$ alongside the pipeline and the pressures $pr_{m,t}$ and $pr_{n,t}$ at each adjacent nodes $(m,n) \in \mathcal{Z}$ given the Weymouth constant $K_{m,u}$. This relation is derived from the partial differential equation underpinning the *conservation of momentum* [**?**]. Equation (7.2e) defines the physical flow as the average between inlet and outlet flows $q^{\mathrm{in}}_{m,n,t}$ and $q^{\mathrm{out}}_{m,n,t}$ for each pipeline $(m,n) \in \mathcal{Z}$. The linepack $h_{m,n,t}$, i.e., the inherent storage of gas inside each pipeline, is defined by the set of equations (7.2f) to (7.2h), which are derived from the *conservation of mass* principle. The spatial conservation of mass corresponds to equation (7.2f) which links the value of linepack to the pressures at inlet and outlet adjacent nodes, given the pipeline constant $S_{m,n}$. The mass conservation through time is enforced via equations (7.2g) and (7.2h), where $H^0_{m,n}$ represents the initial condition of linepack.

Coupling Constraint

The power and gas systems are coupled via the operation of NGFPPs, which consume natural gas for producing electrical power. This is revealed via the following relationship $\forall m \in \mathcal{M},\ t \in \mathcal{T}$

$$\sum_{g \in \mathcal{A}_m^{\mathcal{G}}} p_{g,t} - \sum_{e \in \mathcal{A}_m^{\mathcal{E}}} \eta_e p_{e,t} - \sum_{\substack{n:(m,n) \\ \in \mathcal{Z}}} \left(q^{\text{in}}_{m,n,t} - q^{\text{out}}_{m,n,t} \right) = \sum_{b \in \mathcal{B}} p_{b,t}, \tag{7.3}$$

where the sets $\mathcal{A}_m^{\mathcal{G}}$ and $\mathcal{A}_m^{\mathcal{E}}$ respectively collect the gas suppliers and the NGFPPs connected to node m, while the set $n : (m,n) \in \mathcal{Z}$ collects the pipelines connected to node m. Equation (7.3) balances the gas supplies $p_{g,t}$ and inlet flows $q^{\text{in}}_{m,n,t}$ with the gas consumption by NGFPPs $\eta_e\, p_{e,t}$, outlet flows $q^{\text{out}}_{m,u,t}$, and gas demands $p_{b,t}$. The conversion factor η_e defines the link between the energy consumed and that produced by the NGFPPs.

Resulting Model

The resulting coordinated day-ahead power and gas dispatch problem writes as

$$\min_{\Theta} \sum_{t \in \mathcal{T}} \left(\sum_{e \in \mathcal{E}} C_e p_{e,t} + \sum_{g \in \mathcal{G}} C_g p_{g,t} \right), \tag{7.4a}$$

$$\text{s.t. Electrical Network Constraints (7.1a)-(7.1c)}, \tag{7.4b}$$

$$\text{Gas Network Constraints (7.2a)-(7.2h)}, \tag{7.4c}$$

$$\text{Coupling Constraint (7.3)}, \tag{7.4d}$$

where $\Theta = \left\{ p_{e,t},\, p_{g,t},\, pr_{m,t},\, q_{m,u,t},\, q^{\text{in}}_{m,u,t},\, q^{\text{out}}_{m,u,t},\, h_{m,u,t} \right\}$ represent the set of decision variables. By solving problem (7.4a), the system operator aims at minimizing the total day-ahead cost of scheduling the electrical power and natural gas in-feeds, where C_e and C_g respectively represent the marginal generation cost of electricity and natural gas, while ensuring the operating constraints of both power and gas systems.

7.2.2 The Neural-Network-Constrained Framework

The Weymouth equation (7.2d) is a non-convex quadratic equality constraint, which highly hinders the resolution of the day-ahead coordinated power and gas dispatch problem (7.4) via off-the-shelf solvers. In the following box, we introduce the basic definitions of convexity for sets and functions, which are typically required by optimization solvers.

Basic definitions of convexity in optimization problems

Convexity is an usual requirement for optimization programs to be solved by the available solvers, which declines in two main features [128].

***i)* A convex feasible set.** The feasible set is the collection of solutions which satisfy the constraints of an optimization problem. A set $\mathcal{F} \subseteq \mathbb{R}^n$ is convex if, for any $x_1, x_2 \in \mathcal{F}$, the following condition holds for any $\alpha \in [0, 1]$:

$$\alpha x_1 + (1 - \alpha) x_2 \in \mathcal{F}. \tag{7.5}$$

Note that the expression $\alpha x_1 + (1-\alpha) x_2$ defines the line between x_1 and x_2. Intuitively speaking, the condition (7.5) requires that all points of the line between any two elements of the set remains within the set. This definition is exemplified in Figure 7.1. It is worth mentioning that, when it comes to equality constraints, only affine equality restrictions are convex.

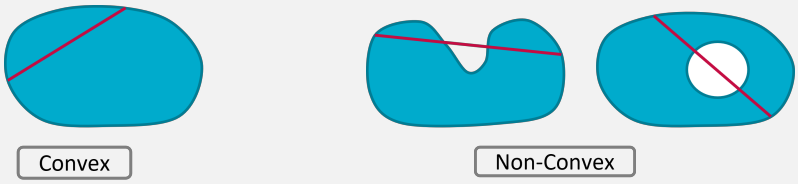


Figure 7.1: Illustrative examples of convex and non-convex feasible sets.

***ii)* A convex objective function.** A function $f(x) : \mathbb{R}^n \to \mathbb{R}$ is convex if the domain $\mathrm{dom} f$ is convex and if, for any $x_1, x_2 \in \mathbb{R}^n$, the following condition holds for any $\alpha \in [0, 1]$.

$$f(\alpha x_1 + (1 - \alpha) x_2) \leq \alpha f(x_1) + (1 - \alpha) f(x_2). \tag{7.6}$$

Intuitively speaking, the condition (7.6) requires that the line between any two points of the function should remain above the function. In other words, if one can find one line, for which at least one element is lower than the value of the function (i.e., a line that get through the function curve), the function is not convex. This basic definition is exemplified in Figure 7.2.

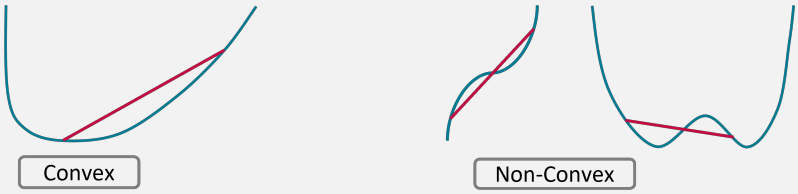


Figure 7.2: Illustrative examples of convex and non-convex functions.

In this work, we propose a neural-network-constrained optimization method which allows to recover tractability, by encoding a regression model of the non-convex Weymouth equation as a MILP into the set of constraints. In the following, we explain the methodology which requires *i)* to train the neural network offline, and *ii)* to formulate the regression model as a MILP and *iii)* to introduce the resulting formulation into the neural-network-constrained coordinated power and gas dispatch problem to solve.

Training the Neural Network

Neural networks represent a type of supervised machine learning regression tool[7] that is capable of extracting non-linear patterns from data, that would not be captured by other regression techniques (e.g., linear or polynomial regression). Their architecture is composed of neurons that are interconnected between each other via weighting branches. An example of such architecture, composed of 2 layers and 5 neurons per layer, is represented in Figure 7.3 for the specific case of Weymouth equation, which relates the inlet and outlet pressures with the gas flow within the pipeline. The fundamental building block of the neural network is the neuron, i.e., a non-linear activation function $g(.) : \mathbb{R} \rightarrow \mathbb{R}$ whose input is the weighted sum of outputs from the previous layer, and a bias term. There exist different types of activation functions that may be leveraged, such as the sigmoid, the ReLU, and the leaky ReLU which are shown in Figure 7.4.

The operation of the neural network can be written in a mathematical vectorial form, such as

$$\widetilde{y}_{k+1} = W_{k+1}\, y_k + B_{k+1} \;\; \forall k \in \{0, 1, ..., K-1\}\,, \tag{7.7a}$$
$$y_k = g\left(\widetilde{y}_k\right) \;\; \forall k \in \{1, ..., K\}\,, \tag{7.7b}$$
$$z = W_{K+1}\, y_K + B_{K+1}. \tag{7.7c}$$

In (7.7), the input variables are $y_0 = [pr_{m,t} \;\; pr_{n,t}]^\top$ and each layer $k \in \{1, ..., K\}$ is composed of N_k neurons. The weighted sum $\widetilde{y}_k \in \mathbb{R}^{N_k}$ is calculated in (7.7a) via the weight matrices $W_k \in \mathbb{R}^{N_k \times N_{k-1}}$ and the bias vector $B_k \in \mathbb{R}^{N_k}$. It is then fed to the activation function $g\left(\widetilde{y}_k\right)$ in equation (7.7b). Note that we define $g(.) : \mathbb{R}^{N_k} \rightarrow \mathbb{R}^{N_k}$ as a vector function, whose input and output are vectors of the same length, the elements of which are linked via the underlying unidimensional activation function. Finally, the output of the network $z = q_{m,n,t}$ is calculated by (7.7c), where $W_{K+1} \in \mathbb{R}^{1 \times N_k}$ and $B_{K+1} \in \mathbb{R}$, respectively represent the weight vector and the bias term of the output layer. We compactly write the resulting

[7] Neural networks may be leveraged for regression or classification tasks. In this work, we will focus on the regression capability of neural networks.

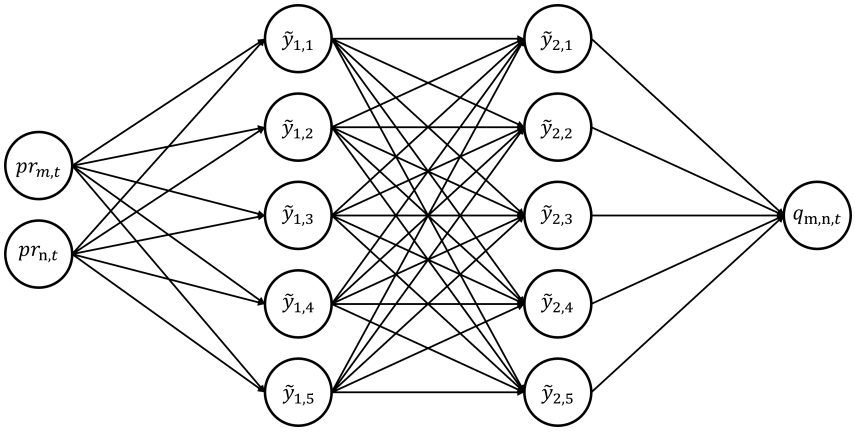


Figure 7.3: An illustrative neural network for the Weymouth equation, linking the inlet and outlet pressures to the flow. The neural network is composed of two hidden layers and five neurons per layer. Each branch is assigned with a weight which defines the weighted sum fed to the activation function.

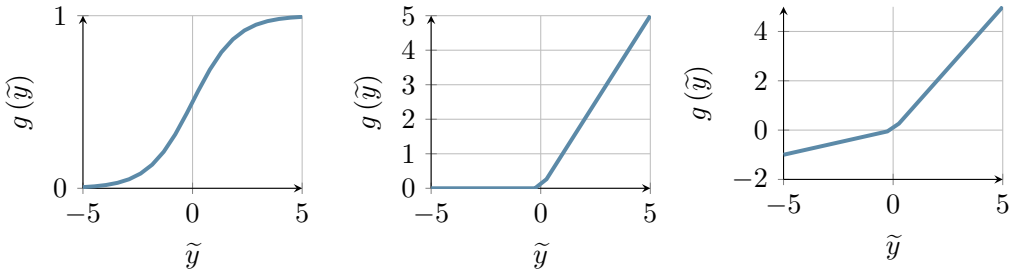


Figure 7.4: The sigmoid, ReLU and leaky ReLU activation functions.

non-linear regression function as $z = f(y_0, W, B)$, where y_0 is the input vector, and W, B respectively collect the weight matrices W_k and the bias vectors B_k, for each layer k and the output layer.

The challenge for utilizing the neural network (7.7) is to appropriately select the weights W and biases B, during the so-called *training* phase. Assuming the availability of a training dataset $\{\widehat{y}_{0,i}, \widehat{z}_i\} \quad \forall i \in \{1, ..., N\}$, this task is usually achieved by solving the following optimization problem

$$\min_{W,B} \quad \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(f\left(\widehat{y}_{0,i}, W, B\right), \widehat{z}_i\right). \tag{7.8}$$

Problem (7.8) aims at minimizing the expected value of loss function over the training dataset. The loss function $\mathcal{L}(.,.)$ gives a score for the error between

the neural network output $f\left(\widehat{y}_{0,i}, W, B\right)$ and the expected one $\widehat{z}_i$. In this work, we consider the square error as our loss function, as it is the common practice for regression models in the literature. Problem (7.8) is non-linear, due to the loss function but also due to the neural network transfer function. For solving (7.8), we use dedicated stochastic gradient descent algorithm and backpropagation techniques, that are well documented in the literature [159].

The set of equations (7.7) equipped with the optimal weight matrices obtained from (7.8), is a regression model for the Weymouth equation. In the following, our aim is to incorporate (7.7) into the set of constraints of the coordinated power and gas dispatch problem, to replace the non-convex Weymouth equation. The activation function in (7.7b) therefore requires reformulation. In that direction, we propose a tractable MILP reformulation of the activation function.

Reformulation of the Activation Function

The use of non-linear activation functions is essential for enabling neural networks to learn complex mapping functions. In this way, the adoption of the Rectifier Linear Unit (ReLU) and leaky ReLU is one of the key milestones in the advent of machine learning [160], as it has been shown that neural networks relying on these functions are easier to train[8] and often outperform traditional sigmoid and hyperbolic tangent activation functions [161]. Following these strong empirical evidences, we propose a tractable MILP reformulation of these activation functions within the proposed neural-network-constrained framework. We begin with the definition of ReLU and leaky ReLU. Mathematically speaking, these functions may be written in the form of

$$\text{ReLU}\left(\widetilde{y}\right) = \max\left(0, \widetilde{y}\right), \tag{7.9}$$

$$\ell\text{-ReLU}\left(\widetilde{y}\right) = \max\left(\alpha\widetilde{y}, \widetilde{y}\right), \tag{7.10}$$

where $\text{ReLU}\left(.\right) : \mathbb{R} \to \mathbb{R}$ and $\ell\text{-ReLU}\left(.\right) : \mathbb{R} \to \mathbb{R}$ respectively represent the ReLU and leaky ReLU activation functions. The parameter $\alpha \in [0, 1]$ defines the slope of the left-hand side non-positive part of the leaky ReLU activation function. Note that selecting $\alpha = 0$ for ℓ-ReLU amounts to the ReLU activation function. Therefore, we propose the following MILP reformulation as a generic model for

[8]The ReLU and leaky ReLU activation functions are nearly linear, and are therefore easy to optimize within gradient-based techniques.

encoding (7.7b), and for both functions:

$$\widetilde{y}_k^{\min} b_k \leq x_{1,k} \leq 0, \qquad \forall k \in \{1, ..., K\}, \tag{7.11a}$$
$$0 \leq x_{2,k} \leq \widetilde{y}_k^{\max}(1 - b_k), \qquad \forall k \in \{1, ..., K\}, \tag{7.11b}$$
$$\widetilde{y}_k = x_{1,k} + x_{2,k}, \qquad \forall k \in \{1, ..., K\}, \tag{7.11c}$$
$$y_k = \alpha x_{1,k} + x_{2,k}, \qquad \forall k \in \{1, ..., K\}, \tag{7.11d}$$
$$b_k \in \{0, 1\}^{N_k}, \qquad \forall k \in \{1, ..., K\}, \tag{7.11e}$$

where all the neurons in (7.7b) are simultaneously considered, and $x_{1,k} \in \mathbb{R}^{N_k}$, $x_{2,k} \in \mathbb{R}^{N_k}$ and $b_k \in \{0, 1\}^{N_k}$ are auxiliary continuous and binary variables. Note that we use vector notation, for the sake of clarity. The MILP (7.11) works as follows. The binary variables b_k equal to 0 or 1, respectively when the negative or positive part of the function is active. Therefore, the combination of equations (7.11a), (7.11b) and (7.11c), implies that $x_{1,k}$ and $x_{2,k}$ respectively take the negative and positive value of $\widetilde{y}_k$. The parameters $\widetilde{y}_k^{\min} \in \mathbb{R}^{N_k}$ and $\widetilde{y}_k^{\max} \in \mathbb{R}^{N_k}$, collect the minimum and maximum admissible values for the input $\widetilde{y}_k$ to each neuron. The equation (7.11d) calculates the output of the activation function, given the values of $x_{1,k}$ and $x_{2,k}$. Finally, equation (7.11e) defines b_k as a vector of binary variables. In Section 7.2.2, we incorporate the MILP (7.11) into the set of constraints of the power and gas dispatch problem to improve the modeling of gas flow dynamics.

Remark 3 (Benchmarking the proposed reformulation of activation function). To appropriately show the superiority of the proposed reformulation of activation function, we compare the approach with convexification techniques (which will be presented later in Section 7.3) as well as with an existing reformulation of ReLU proposed in [58], which is as follows:

$$\widetilde{y}_k \leq y_k \leq \widetilde{y}_k - \widetilde{y}_k^{\min}(1 - b_k), \qquad \forall k \in \{1, ..., K\}, \tag{7.12a}$$
$$0 \leq y_k \leq \widetilde{y}_k^{\max} b_k, \qquad \forall k \in \{1, ..., K\}, \tag{7.12b}$$
$$b_k \in \{0, 1\}^{N_k}, \qquad \forall k \in \{1, ..., K\}. \tag{7.12c}$$

In (7.12), the binary variables $b_k \in \{0, 1\}^{N_k}$ either activate or deactivate the right-hand side inequalities in (7.12a) and (7.12b). By doing so, $b_k = 1$ implies $y_k = \widetilde{y}_k$, and $b_k = 0$ implies $y_k = 0$. It is worth mentioning that our reformulation in (7.11) generalizes the approach in [58], since it is capable of considering both ReLU and leaky ReLU. In addition, we will show later in our numerical experiments in Section 7.4 that our reformulation outperforms (7.12) in terms of computational efficiency.

The Neural-Network-Constrained Power and Gas Dispatch

The reformulation proposed in Section 7.2.2 for incorporating the neural-network constraints in the coordinated power and gas dispatch problem results in the following MILP

$$\min_{\Theta} \sum_{t\in\mathcal{T}} \left(\sum_{e\in\mathcal{E}} C_e p_{e,t} + \sum_{g\in\mathcal{G}} C_g p_{g,t} \right), \tag{7.13a}$$

$$\text{s.t. Electrical Network Constraints (7.1a)-(7.1c)}, \tag{7.13b}$$

$$\text{Gas Network Constraints (7.2a)-(7.2c), (7.2e)-(7.2h)}, \tag{7.13c}$$

$$\text{Coupling Constraint (7.3)}, \tag{7.13d}$$

$$\text{Neural network constraints (7.7a), (7.7c), and (7.11)}, \tag{7.13e}$$

where the intractable non-convex Weymouth equation (7.2d) is replaced by the tractable MILP (7.13e) encoding the neural network, via (7.7a), (7.7c), and the reformulation of activation function in (7.11).

Remark 4 (Benchmarking the proposed reformulation of activation function). The procedure to derive the neural-network-constrained power and gas dispatch problem, given the existing reformulation of ReLU activation function [58], is similar to that explained in this section, except that (7.12) is used instead of (7.11), to reformulate the activation function.

7.3 Benchmark Convexification Techniques

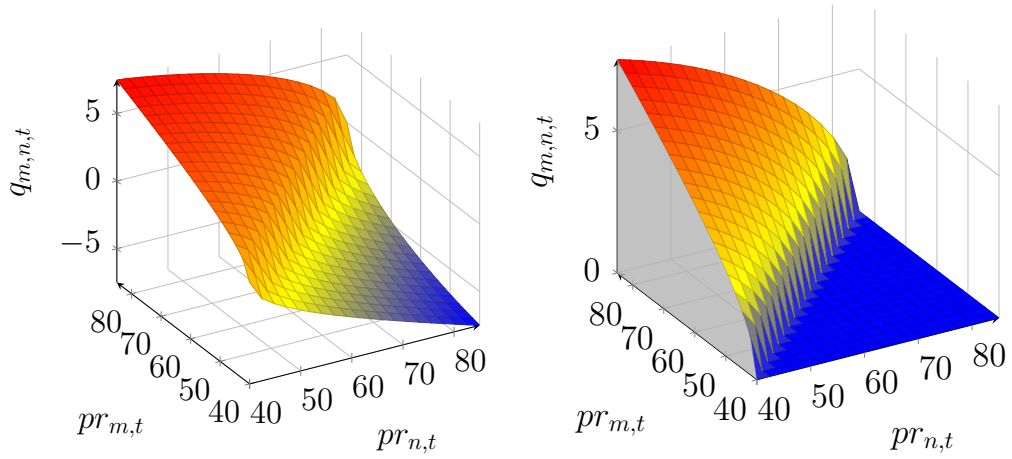
In this section, we introduce existing convexification techniques for the non-convex Weymouth equation (7.2d), that will be used later as benchmark solutions in the numerical experiments of Section 7.4. In particular, we focus on the SOC relaxation in Section 7.3.1. Next, tightened versions of the SOC relaxation are presented in Sections 7.3.2 and 7.3.3, respectively using McComick envelopes and a dedicated bound tightening algorithm. Finally, an incremental piecewise linear approximation is showcased in Section 7.3.4.

7.3.1 SOC Relaxation

The most commonly used technique to address the non-convexity is to equivalenty cast the Weymouth equation into

$$q^2_{m,n,t} \leq K^2_{m,n} \left(pr^2_{m,t} - pr^2_{n,t} \right), \ \forall (m,n) \in \mathcal{Z}, \ t \in \mathcal{T}, \tag{7.14}$$

$$q^2_{m,n,t} \geq K^2_{m,n} \left(pr^2_{m,t} - pr^2_{n,t} \right), \ \forall (m,n) \in \mathcal{Z}, \ t \in \mathcal{T}, \tag{7.15}$$



(a) The Weymouth relation between inlet, outlet pressures $pr_{m,t}$, $pr_{n,t}$, and the flows $q_{m,n,t}$ within the pipeline. Pressures are expressed in bar, flows are expressed in $\mathrm{MNm}^3/\mathrm{h}$.

(b) SOC relaxation. The grey area shows the area under the surface which is added to the feasible space by conic relaxation. The negative flows are neglected.

Figure 7.5: Three-dimensional plot of the Weymouth relation and its SOC relaxation.

and relax the non-convex constraint (7.15), thereby replacing the equality constraint by the SOC constraint in (7.14). The difference between the original Weymouth equation and the SOC constraint (7.14) is illustrated in Figure 7.5. The left-hand side Fig. 7.5(a) represents the original bidirectional link between natural gas flow and inlet and outlet pressures, captured by the non-convex conic surface. The right-hand side Fig. 7.5(b) represents the feasible space considered under the SOC relaxation (i.e., the whole grey volume under the curve). These graphs suggest that the solutions obtained from the SOC relaxation may bring non-trivial operational constraint violations. Therefore, additional tightening of the relaxation have been explored in the literature.

7.3.2 SOC Relaxation with McCormick Envelopes

The first method used to tighten the SOC relaxation consists in introducing a convex relaxation of (7.15) using the McCormick envelopes, as proposed in [50]. By doing so, the volume under the curve which contains infeasible operating points is shrunk. Mathematically speaking, this amounts to replace (7.15) with the following set of equations, from which we omit "$\forall\,(m,n) \in \mathcal{Z},\ t \in \mathcal{T}$" in each

constraint, for the sake of conciseness:

$$\kappa_{m,n,t} \geq K_{m,n}^2 \lambda_{m,n,t}, \tag{7.16a}$$
$$\kappa_{m,n,t} \geq q_{m,n,t}^2, \tag{7.16b}$$
$$\kappa_{m,n,t} \leq \left(Q_{m,n,t}^{\max} + Q_{m,n,t}^{\max}\right) q_{m,n,t} - Q_{m,n,t}^{\max} Q_{m,n,t}^{\min}, \tag{7.16c}$$
$$\lambda_{m,n,t} \geq A_{m,n,t}^{\min} b_{m,n,t} + B_{m,n,t}^{\min} a_{m,n,t} - A_{m,n,t}^{\min} B_{m,n,t}^{\min}, \tag{7.16d}$$
$$\lambda_{m,n,t} \geq A_{m,n,t}^{\max} b_{m,n,t} + B_{m,n,t}^{\max} a_{m,n,t} - A_{m,n,t}^{\max} B_{m,n,t}^{\max}, \tag{7.16e}$$
$$\lambda_{m,n,t} \leq A_{m,n,t}^{\min} b_{m,n,t} + B_{m,n,t}^{\max} a_{m,n,t} - A_{m,n,t}^{\min} B_{m,n,t}^{\max}, \tag{7.16f}$$
$$\lambda_{m,n,t} \leq A_{m,n,t}^{\max} b_{m,n,t} + B_{m,n,t}^{\min} a_{m,n,t} - A_{m,n,t}^{\max} B_{m,n,t}^{\min}. \tag{7.16g}$$

In (7.16), $\kappa_{m,n,t}$, $\lambda_{m,n,t}$, $a_{m,n,t}$ and $b_{m,n,t}$ are auxiliary variables. In addition, $a_{m,n,t} = pr_{m,t} + pr_{n,t}$ and $b_{m,n,t} = pr_{m,t} - pr_{n,t}$. In general, the convex relaxation (7.16) provides a more accurate representation of the feasible set than the one achieved by the SOC relaxation. However, it requires the introduction of the bounds $Q_{m,n,t}^{\max}$, $Q_{m,n,t}^{\min}$, $A_{m,n,t}^{\max}$, $A_{m,n,t}^{\min}$, $B_{m,n,t}^{\max}$, and $B_{m,n,t}^{\min}$, which are not straightforward to calculate. The introduction of loose bounds may still provide an insufficient representation of the Weymouth constraint.

7.3.3 Tightening the Bounds of McCormick Envelopes

To address this issue, the authors in [50]-[51] propose a bound tightening algorithm that enables to improve the accuracy of the relaxation. The iterative algorithm is able to calculate effective bounds, i.e., that reduce the available set of solutions around the operating point. Once these bounds are derived, the power and gas dispatch problem with SOC relaxation and McCormick envelopes, can be solved efficiently. However, it is worth mentioning that the procedure to derive the bounds must be run online (i.e., different bounds are required under different operating conditions), therefore resulting in a significant increase in computational burden. The flow of the bound tightening algorithm is presented in Algorithm 2.

Remark 5. In general, the SOC-based convex relaxation techniques can not straightforwardly enhance to consider bidirectionality. The previous tentatives [162] towards SOC-based relaxations of bidirectional gas flows usually rely on binary variables defining the direction of flow and big-M parameters. These modeling approaches may arise in infeasible and intractable dispatch solutions due to the impreciseness introduced by big-M parameters and the difficulty for branch-and-bound algorithms to quickly determine the direction of gas flow in all pipelines of the system.

Algorithm 2 Bound tightening

Step 1: Set $k = 1$, and initial values for the bounds $Q^{0,\max}_{m,n,t}$, $Q^{0,\min}_{m,n,t}$, $A^{0,\max}_{m,n,t}$, $A^{0,\min}_{m,n,t}$, $B^{0,\max}_{m,n,t}$, and $B^{0,\min}_{m,n,t}$. Define δ a decreasing series of positive real numbers which tends to zero, e.g., $\delta = [1, 0.5, 0.25, 0.1, 0.05, 0.01, 0, ..., 0]$. Set the convergence threshold ϵ to, e.g., 10^{-2}.
Step 2: Solve the power and gas dispatch problem, given the convex relaxation (7.16) of the Weymouth equation, and $Q^{0,\max}_{m,n,t}$, $Q^{0,\min}_{m,n,t}$, $A^{0,\max}_{m,n,t}$, $A^{0,\min}_{m,n,t}$, $B^{0,\max}_{m,n,t}$, and $B^{0,\min}_{m,n,t}$. Retrieve the optimal values $q^*_{m,n,t}$, $a^*_{m,n,t}$ and $b^*_{m,n,t}$ for variables $q_{m,n,t}$, $a_{m,n,t}$ and $b_{m,n,t}$.
Step 3: Assign the following new bounds:

$$
\begin{aligned}
Q^{\max}_{m,n,t} &\leftarrow (1+\delta_k)\max\left\{q^*_{m,n,t}\right\}\\
Q^{\min}_{m,n,t} &\leftarrow (1-\delta_k)\min\left\{q^*_{m,n,t}\right\}\\
A^{\max}_{m,n,t} &\leftarrow (1+\delta_k)\max\left\{a^*_{m,n,t}\right\}\\
A^{\min}_{m,n,t} &\leftarrow (1-\delta_k)\min\left\{a^*_{m,n,t}\right\}\\
B^{\max}_{m,n,t} &\leftarrow (1+\delta_k)\max\left\{b^*_{m,n,t}\right\}\\
B^{\min}_{m,n,t} &\leftarrow (1-\delta_k)\min\left\{b^*_{m,n,t}\right\}
\end{aligned}
$$

Step 4: Check if the convergence criterion, i.e., $|Q^{\max}_{m,n,t} - Q^{0,\max}_{m,n,t}| \leq \epsilon$, is ensured for all bounds. If not, set $k \leftarrow k+1$ and $Q^{0,\max}_{m,n,t} \leftarrow Q^{\max}_{m,n,t}$ for all bounds, and go to step 2.

7.3.4 Incremental Piecewise Linear Approximation

Piecewise linear functions represent an alternative potential approximation for the Weymouth equation. In particular, the incremental method, as described in [**?**]-[52] has shown to outperform other existing piecewise linear models for modeling gas network [53]. The main idea behind the incremental method is to linearize the non-linear terms that compose the Weymouth equation, i.e., square pressures $pr^2_{m,t}$ and square gas flows $q^2_{m,n,t}$. To do so, a collection of breakpoints are collected, i.e., $\left\{Q_j, Q_j^2\right\}$ and $\left\{PR_j^2, PR_j^2\right\}$ where $j \in \{1, ..., J\}$, and auxiliary variables $h^q_{m,n,t}$ and $h^{pr}_{m,t}$ are introduced to describe the non-linear terms $q^2_{m,n,t}$ and $pr^2_{m,t}$, respectively. For instance, the following set of equation produces a piecewise linear approximation $h^q_{m,n,t}$ of $q^2_{m,n,t}$ from which we omit "$\forall\,(m,n) \in \mathcal{Z},\ t \in \mathcal{T}$"

in each constraint for the sake of conciseness:

$$h^q_{m,n,t} = Q_1^2 + \sum_{j=1}^{J} \left(Q_{j+1}^2 - Q_j^2\right) \delta^q_{m,n,t,j}, \tag{7.17a}$$

$$q_{m,n,t} = Q_1 + \sum_{j=1}^{J-1} \left(Q_{j+1} - Q_j\right) \delta^q_{m,n,t,j}, \tag{7.17b}$$

$$0 \leq \delta^q_{m,n,t,j} \leq 1, \qquad \forall j \in \{1, ..., J\}, \tag{7.17c}$$

$$\varphi^q_{m,n,t,j} \leq \delta^q_{m,n,t,j}, \qquad \forall j \in \{1, ..., J\}, \tag{7.17d}$$

$$\delta^q_{m,n,t,j} \leq \varphi^q_{m,n,t,j-1}, \qquad \forall j \in \{2, ..., J\}, \tag{7.17e}$$

$$\varphi^q_{m,n,t,j} \in \{0, 1\}, \qquad \forall j \in \{1, ..., J\}. \tag{7.17f}$$

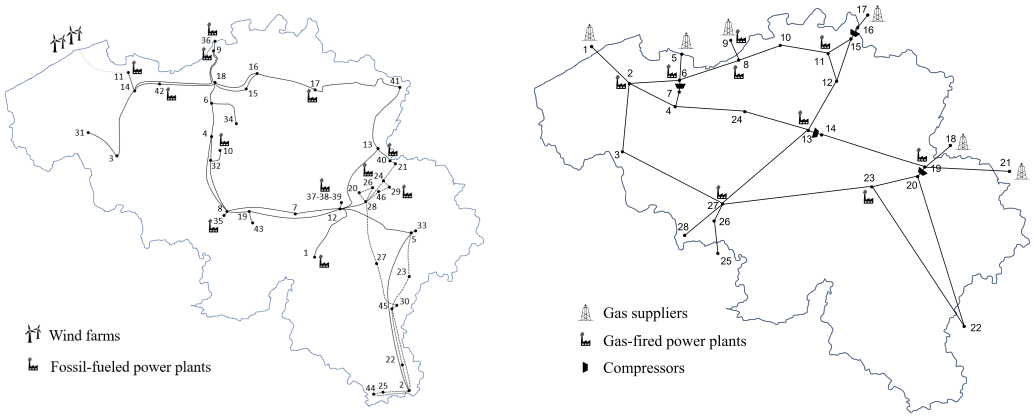
In (7.17), $\delta^q_{m,n,t,j} \in \mathbb{R}$ and $\varphi^q_{m,n,t,j} \in \{0, 1\}$ are additional auxiliary continuous and binary variables. The equations (7.17a) and (7.17b) define the linear interpolation curve between the breakpoints via the continuous variable $\delta^q_{m,n,t,j}$ which is constrained to lie between 0 and 1 in equation (7.17c). Constraints (7.17d) and (7.17e) enforce $\varphi^q_{m,n,t,j}$ equal to 1 or 0, for the intervals on the left or on the right of the current one, respectively. In addition, these equations require that $\delta^q_{m,n,t,j}$ equal to 1 for the intervals on the left, ensuing the proper incremental summation in (7.17a) and (7.17b). The piecewise linear approximation $h^{pr}_{m,t}$ of square pressures $pr^2_{m,t}$ follows immediately. Given those definitions, the Weymouth equation can be approximated via

$$h^q_{m,n,t} = K^2_{m,u} \left(h^{pr}_{m,t} - h^{pr}_{n,t}\right) \quad \forall (m, n) \in \mathcal{Z},\ t \in \mathcal{T}, \tag{7.18}$$

which is a linear expression.

7.4 Numerical Experiments

In this section, we summarize the results of numerical experiments performed on a case study inspired by the Belgian power and gas systems. In particular, we solve the proposed neural-network-constrained coordinated power and gas dispatch (7.13) with the proposed reformulation of activation function in two different settings. The first considers *unidirectional* gas flows (by adding non-negative restrictions $q_{m,n,t} \geq 0$, $q^{\text{in}}_{m,n,t} \geq 0$ and $q^{\text{out}}_{m,n,t} \geq 0$), while the second considers *bidirectional* gas flows. The main difference between the two settings stem from the relation captured by the neural network (i.e., unidirectional or bidirectional gas flows). We compare the unidirectional formulation with the existing convexication



(a) Belgian electrical grid topology (plain lines are 380 kV and dashed lines are 220 kV).

(b) Belgian gas grid topology.

Figure 7.6: The real-life Belgian power and gas systems and related information.

techniques[9]. Next, we show the superiority of the proposed neural-network-constrained framework for considering bidirectional gas flow dynamics. All models are either SOC programs or MILPs and are implemented in Julia programming language v1.7.2, using the modeling language JuMP v1.0.0, and the Gurobi solver v9.5.1, on a standard computer clocking at 2.5 GHz, with 16 GB of RAM memory. The remainder of this section is structured as follows. Section 7.4.1 introduces the Belgian power and gas systems, which are used as our case study. Section 7.4.2 presents the procedure for training the machine learning regression models. The analysis of the performance of our proposed approach compared to state-of-the-art convexification techniques is derived in Section 7.4.3, given unidirectional gas flow dynamics. The assessment of the proposed reformulation of activation function is discussed in Sections 7.4.4. Finally, we explore the bidirectional gas flow dynamics in Section 7.4.5.

7.4.1 Case Study: The Belgian Power and Gas Systems

Our case study is based on the Belgian power and gas systems. We consider the high voltage power transmission system, composed of 46 nodes that are connected via 69 transmission lines, as depicted in Fig. 7.6(a). The electrical power demand is supplied by conventional power generating units, NGFPPs, and wind farms, whose aggregate power generating capacities are respectively 6,000 MW, 3,329 MW, and 3,794 MW. The NGFPPs couple the Belgian power and natural gas

[9]Recall that, unlike our proposed neural-network-constrained framework, it is not straightforward to consider bidirectionality given the state-of-the-art convexification techniques. Therefore, the comparison with respect to those techniques is performed in the unidirectional framework.

systems. The latter is a 28-node high pressure gas transmission system, composed of 33 pipelines and 4 compressors, as depicted in Fig. 7.6(b). It connects the gas suppliers (with an aggregate capacity of 11.2 MNm^3) with the natural gas demand centers. All the technical and economical parameters pertaining to both Belgian power and natural gas transmission systems are reported in Appendix B.

7.4.2 Machine Learning Models

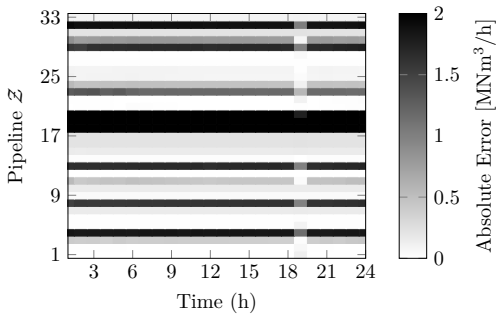
The first step for embedding neural networks within the coordinated power and gas dispatch problem is to appropriately select the optimal weight matrices W and B which minimize the value of loss function. To do so, we generate $N \geq 10^6$ operating data points from the Weymouth surface in Fig. 7.5(a). Next, we split the whole dataset into three subsets, i.e., 60% of data are used for training (to optimize the model parameters), 20% are used for validation (for stopping the iterative training procedure at the optimal time, before overfitting issues occur), and the remaining 20% of data are used as a test set (to compute unbiased performance metrics).

We repeat this procedure for two different network architectures capturing unidirectional gas flows, i.e., one composed of 2 layers and 5 neurons per layer, and one composed of one layer of 15 neurons. These two architectures are trained with both the ReLU and the leaky ReLU activation functions, resulting in four different neural networks. Note that we arbitrarily impose $\alpha = 0.3$ for the leaky ReLU activation function. Finally, we train four additional neural networks, based on the same characteristics, but capturing bidirectional gas flows. These will be used later in Section 7.4.5, to enhance the proposed neural-network-constrained framework to bidirectional gas flows dynamics.

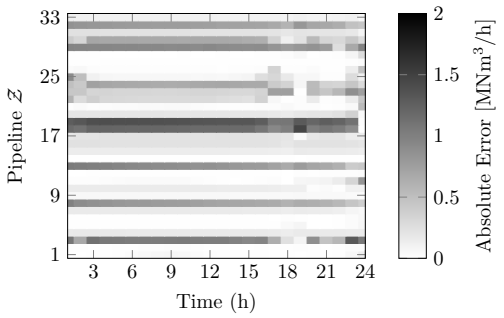
7.4.3 Ex-post Feasibility under Unidirectional Assumption

In this section, we aim to evaluate the ex-post feasibility of the obtained neural-network-constrained solutions, compared to existing SOC-based relaxation techniques and the incremental piecewise linear approximation. To do so, we solve the power and gas dispatch with *i)* our proposed neural-network-constrained framework, equipped with the optimal weight matrices W and B of the four architectures discussed in Section 7.4.2, *ii)* the three SOC-based convex relaxation techniques and *iii)* the piecewise linear approximation with two different numbers of breakpoints (i.e., 5 and 10)[10]. Next, we retrieve the optimal state variables, i.e., gas flows $q_{m,n,t}$ and pressures $pr_{m,t}$ for all models. We calculate the mismatch with respect to the Weymouth equation (7.2d), as the absolute

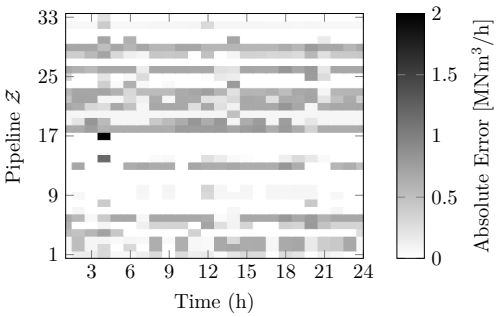
[10]We enforce an optimality gap threshold of 5% for all the MILP models to solve.



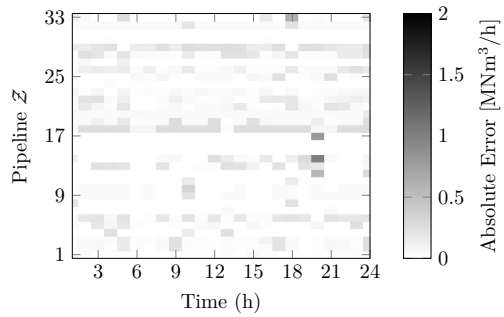
(a) SOC relaxation + McCormick envelopes.



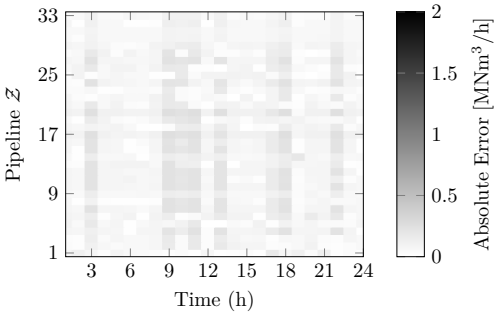
(b) SOC relaxation + bound tightening.



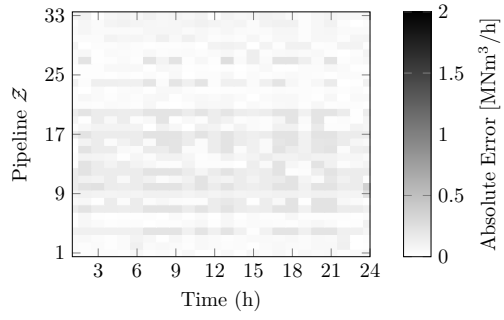
(c) Piecewise linear model (5 breakpoints).



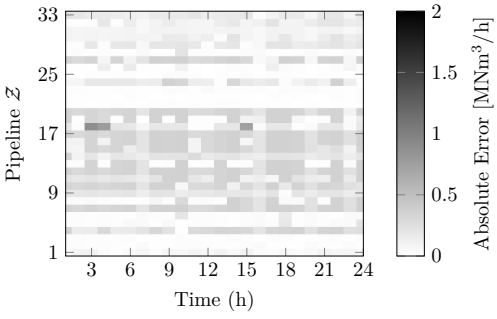
(d) Piecewise linear model (10 breakpoints).



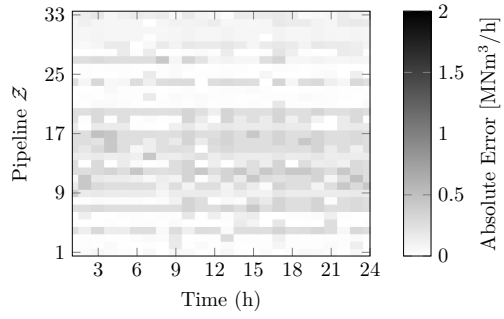
(e) NN-constrained (2x5, ReLU).



(f) NN-constrained (1x15, ReLU).



(g) NN-constrained (2x5, leaky ReLU).



(h) NN-constrained (1x15, leaky ReLU).

Figure 7.7: Matrix plot of the relative error with respect to Weymouth equation.

Table 7.1: Performance indicators for the neural-network-constrained framework and the benchmark approaches, considering unidirectional gas flows.

Model	Number of binary variables	Computational time (Gap) [s]	MAE_{Opt} (MAE_{NN}) $(\text{MNm}^3/\text{h})^2$	Expected cost [€]
SOC	N.A.	≤ 1 s	1.54	$1.41 \cdot 10^6$
SOC + McCormick	N.A.	1.5 s	0.67	$1.41 \cdot 10^6$
SOC + BT[1]	N.A.	12 s	0.30	$1.37 \cdot 10^6$
PLA[2] (5 breakpoints)	6,528	24 s (0.5 %)	0.22	$1.47 \cdot 10^6$
PLA (10 breakpoints)	14,760	60 s (0.7 %)	0.05	$1.47 \cdot 10^6$
ReLU, 2x5	7,920	1622 s (2.7 %)	0.09 (0.045)	$1.46 \cdot 10^6$
ReLU, 1x15	11,880	905 s (4,4 %)	0.08 (0.038)	$1.45 \cdot 10^6$
Leaky ReLU, 2x5	7,920	1674 s (4.4%)	0.14 (0.043)	$1.45 \cdot 10^6$
Leaky ReLU, 1x15	11,880	741 s (4.8 %)	0.13 (0.036)	$1.44 \cdot 10^6$

[1] BT: Bound tightening [2]: PLA: Piecewise Linear Approximation.

error between the right-hand side and left-hand side of the Weymouth, i.e., $\Delta_{m,n,t} = |q^2_{m,n,t} - K^2_{m,n}\left(pr^2_{m,t} - pr^2_{n,t}\right)|$ for all pipelines $(m,n) \in \mathcal{Z}$ and all time periods $t \in \mathcal{T}$, and report the value of absolute errors in the matrix plots of Fig. 7.7. In these figures, darker rectangles highlight those pipelines and time periods with a larger magnitude of error. We also report key performance indicators in Table 7.1, such as, the computational time, the Mean Absolute Error (MAE) among all pipelines and all time periods, and the expected total scheduling cost.

One can draw several conclusions from these results. First, the Figs. 7.7(a) and 7.7(b) confirm that the SOC-based relaxations are not sufficiently tight for a practical implementation (i.e., the MAE achieved by the three SOC-based techniques are the largest). Next, we observe that the proposed neural-network-constrained framework and the incremental piecewise linear approximation achieve comparable values of MAE. In particular, when comparing models involving roughly similar number of binary variables, we conclude that the 2-layer-5-neuron neural networks achieve a lower MAE compared to a piecewise linear approximation with 5 breakpoints, unlike the 1-layer-15-neuron neural networks for which opposite conclusions may be drawn when compared to the piecewise linear approximation with 10 breakpoints. It is worth emphasizing that the matrix plot in Figs. 7.7(c) and 7.7(d) suggest that the errors from the neural-network-constrained framework are smoother, compared to piecewise linear approximation. The rationale behind this is that the piecewise linear approximation is an interpolation, which may produce large errors locally. In turn, these peak errors may trigger non-trivial operating conditions for system operator. Furthermore, we want to highlight that the neural-network-constrained framework is capable of improving its inherent accuracy by considering real operating measurements in the training phase. In

Table 7.2: Performance indicators for the proposed reformulation of ReLU activation function, compared to the existing [58].

Model	Computational time (Gap) [s]	MAE_{Opt} (MAE_{NN}) $(\text{MNm}^3/\text{h})^2$	Expected cost [€]
Model (7.11), 2x5	1622 s (2.7 %)	0.086 (0.045)	$1.46 \cdot 10^6$
Model (7.11), 1x15	905 s (4,4 %)	0.081 (0.038)	$1.45 \cdot 10^6$
Model (7.12), 2x5	2602 s (4.7 %)	0.070 (0.045)	$1.48 \cdot 10^6$
Model (7.12), 1x15	1954 s (4.9 %)	0.075 (0.038)	$1.45 \cdot 10^6$

contrast with this interesting feature, the piecewise linear approximation is a model-based solution (i.e., which strongly relies on the Weymouth model) which can not incorporate real operating conditions. Finally, an additional interesting result is that the ReLU activation function achieves lower values of MAE, compared to leaky ReLU activation function, which will be confirmed later in Section 7.4.5 in the case of bidirectional gas flows.

7.4.4 Assessing the Proposed Reformulation of Activation Functions

To assess the benefits of the proposed reformulation of activation function in (7.11), we solve the coordinated power and gas dispatch problem, with the existing reformulation of ReLU defined by (7.12). We use the same weight matrices for ReLU as the ones used in Section 7.4.3, and follow the same procedure for reporting the computational time, MAE and expected total scheduling cost in Table 7.2. As our main conclusion, we observe a substantial speed-up, i.e., -37 % for the 2-layer-5-neuron neural network and -53 % for the 1-layer-15-neurons one, compared to the previously proposed reformulation. Interestingly, the benefits for utilizing our reformulation increases when the number of neurons involved in the neural network increase. In addition, recall that our proposed reformulation is capable of considering both ReLU and leaky ReLU, though we omit leaky ReLU from this section, for the sake of clarity.

7.4.5 Performance under Bidirectionality Assumption

In this section, we enhance the neural-network-constrained power and gas dispatch with bidirectional gas flow dynamics. To do so, we train the neural network based on bidirectional operating data points, and simply plug the weight matrices in a power and gas dispatch problem where the gas flow state variables (i.e., $q_{m,n,t}$, $q^{\text{out}}_{m,n,t}$ and $q^{\text{in}}_{m,n,t}$) can either be positive or negative. We solve the bidirectional power and gas dispatch with the four different architectures of neural networks

Table 7.3: Performance indicators for the neural-network-constrained framework, considering bidirectional gas flows.

Model	Number of binary variables	Computational time (Gap) [s]	MAE_{Opt} (MAE_{NN}) $(\text{MNm}^3/\text{h})^2$	Expected cost [€]
ReLU, 2x5	7,920	1257 s (3.9 %)	0.08 (0.326)	$1.13 \cdot 10^6$
ReLU, 1x15	11,880	7200 s (21.3 %)	0.08 (0.533)	$1.23 \cdot 10^6$
Leaky ReLU, 2x5	7,920	1558 s (4.0 %)	0.11 (0.491)	$1.14 \cdot 10^6$
Leaky ReLU, 1x15	11,880	4317 s (2.8 %)	0.12 (0.453)	$1.12 \cdot 10^6$

discussed in Section 7.4.2 and report the performance indicators in Table 7.3. Our main observation is that the MAE of the obtained solution is consistent with those obtained with the unidirectional model (i.e., in general, the MAE slightly decreases). However, the expected total scheduling cost is reduced by about 20 % on average for the neural networks considered in this study. The rationale behind this is that the model with bidirectional gas flow dynamics provides more flexibility to the system operator for scheduling the system, therefore revealing new opportunities to schedule the cheapest units. An additional interesting element is that the MAE of the obtained optimal solution and the one that is obtained by the neural-network transfer function *per se* are different (e.g., the 2-layer-5-neuron neural network based on ReLU achieves a MAE of 0.326 but the obtained scheduling solution attains a MAE equal to 0.08). This observation sheds light on the importance for the regression model to be accurate in the operating zones that are exploited within the optimization problem.

7.5 Chapter Summary and Conclusions

In this chapter, to cope with the non-convexity stemming from the modeling of gas flow dynamics within the coordinated power and gas dispatch, we propose a neural-network-constrained optimization method which includes a regression model of the Weymouth equation into the set of constraints. In addition, we introduce a reformulation of the activation function which improves the computational efficiency. Through several numerical experiments, we show that our framework is capable of recovering a tractable MILP, which outperforms the previously inaccurate SOC-based convexification techniques and is comparable to piecewise linear approximation methods. Interestingly, the proposed neural-network-constrained framework is capable of improving accuracy, without relying on analytical assumptions, by considering real operating measurements.

CHAPTER 8

Conclusion

8.1 Summary

This thesis introduces different optimization methods aiming at improving the short-term operational procedures in power and natural gas systems. The conducted research explores methodological improvements in mainly two different dimensions, i.e., *i)* the modeling of renewable power generation uncertainty, and *ii)* the modeling of gas flow dynamics.

Background and motivation, research questions and contributions are introduced in Chapter 1. More specifically, two different research paths are identified. On the one hand, we focus on improving the existing distributionally robust optimization methodologies for modeling uncertainty by incorporating additional features (such as physical bounds and dependence structure) into the definition of ambiguity set. On the other hand, we explore machine-learning regression tools (i.e., neural networks) combined with mathematical optimization to produce accurate models of non-convex constraints. In this work, we apply this framework for modeling gas flow dynamics within the coordinated power and gas dispatch problem.

Chapter 2 presents the motivation of the work from a broader perspective, i.e., the context of the energy transition towards carbon-neutral systems. The impacts and available action levers on system operation are also showcased in this chapter.

Chapter 3 presents the basic concepts related to the state-of-the-art distributionally robust optimization and its main limitations in view of its implementation in energy systems. To address these limitations, we firstly propose in Chapter 4 considering the physical bounds of the renewable power uncertainty and derive a distributionally robust optimal power flow problem that is exact and

physically-bounded. An efficient alternating algorithm is derived for solving the resulting bilinear program and the numerical experiments show convincing results towards the consideration of physical bounds. In a different direction, Chapter 5 proposes two models that embed the information on dependence structure within the definition of ambiguity set. The first considers an additional constraint on covariance matrix which enforces the dependence structure among renewable power generation uncertainty, thereby defining a moment-metric-based ambiguity set that is capable of representing linear dependencies. The second proposes a new copula-based formulation that is tailored for capturing any type of dependence structure. The numerical experiments show that both proposed approaches have merits and can offer a framework for improving the dispatch solutions compared to the state-of-the-art approach.

Chapter 6 leverages the previous developments in DRO and proposes a distributionally robust power and gas dispatch problem, exposed to a multivariate and correlated demand uncertainty. The numerical experiments based on a case study inspired by the Belgian power and gas systems shed light on the scalability of the proposed approach. The outcomes also reveal the limitations about the modeling of gas flow dynamics.

The non-convexity pertaining to modeling gas flow dynamics impedes the integration of a proper gas network model within the power and gas dispatch problem. In that regard, chapter 7 presents the common convexification approaches that are used in the literature, and proposes a neural-network-constrained framework, based on supervised machine learning regression tools, that accurately approximates the natural gas system behaviour. In this context, a new reformulation of the activation functions is derived, which improves the tightness of the approach. The numerical experiments show promising results in terms of accuracy, compared to the state-of-the-art convexification approaches.

8.2 Conclusions

In the following, we summarize some of the main conclusions and observations of this thesis. As a general conceptual finding, we highlight the need to properly model uncertainty and network limitations to enhance the current operational procedures in energy systems. Indeed, building up on distributionally robust and neural-network-constrained methodologies, the numerous experiments along this thesis show that accounting for these additional information improves cost-efficiency and reliability of the undertaken decisions. These improvements usually come along with a reasonable increase in computational burden.

More specifically, the proposed physically-bounded and exact distributionally robust approach shows that incorporating the physical bounds of uncertainty into the definition of ambiguity set helps retrieving feasible and robust dispatch solutions. In addition, the total operating cost can be reduced in settings where ambiguity is high, i.e., when the system operator has less confidence in the available forecast of renewable power generation. When the ambiguity is smaller, the proposed approach may still outperform the traditional techniques given an appropriate choice of exogenous parameters such as violation probabilities and Wasserstein radius. Regarding the ex-post satisfaction of operating constraints, the combination of support information with an exact distributionally robust formulation results in the enforcement of the system operator risk attitude with more fidelity.

In the same direction, the incorporation of dependence structure in the definition of ambiguity set shows improved dispatch solutions in terms of ex-post total operating cost. This is because the optimal decisions are optimized over a realistic insight of uncertainty, i.e., that is consistent with the dependence structure of empirical data. Two different approaches are explored, namely the moment-metric-based and the copula-based ones. Interestingly, the choice for the distributionally robust method to apply in practice highly depends on the type of uncertainty and the type of problem. We suggest utilizing the moment-metric-based approach for large-scale problems exposed to low-dimensional uncertainty for which the dependence structure can be easily captured by the covariance matrix. The copula-based model, however, fits well small-scale problems exposed to high-dimensional uncertainty with a complex dependence structure. As a final conclusion on these models, we want to emphasize that the incorporation of dependence structure benefits the system operator, whatsoever is the underlying dependence structure of empirical data.

Tightening the definition of ambiguity set within the distributionally robust optimal power flow problems allows to improve the modeling of uncertainty. In turn, this may result in less conservative dispatch solutions (i.e., riskier in terms of operational reliability) because the plurality of distribution functions in the ambiguity set is reduced. However, in the proposed approaches, the choice is let to the system operator to set his own desired risk attitude (e.g., via the Wasserstein radius) in a consistent manner with respect to the obtained solution, whereas state-of-the-art approaches may increase the level of conservativeness via a poor modeling of uncertainty, which is out of control from the system operator perspective.

Therefore, a particular attention must be paid to selecting the exogenous parame-

ters, such as the Wasserstein radius. In that regard, we suggest that the system operator should primarily select the radius as a large value, especially when it has less confidence about the available representation of uncertainty and dependence structure. Next, by decreasing the value of radius from a test period to another, the system operator can adapt its own risk attitude.

Finally, the limitations of traditional convexification techniques for modeling non-convex equations, such as those governing gas flow dynamics, have been experienced in this thesis. In that regard, the proposed neural-network-constrained framework for the coordinated power and gas dispatch problem improves the obtained solution in terms of ex-post operational feasibility. This is because neural networks are able to capture the non-linear equations underlying the gas flow dynamics with high accuracy. Additionally, the appropriate encoding of activation functions is crucial for embedding neural networks in the optimization process. In that direction, the reformulation of linear activation function proposed in this thesis has shown to improve computational efficiency. Overall, the setting considered in this work yields a framework enabling the incorporation of fully data-driven regression models into the set of operating constraints, i.e., trained on real operating conditions.

8.3 Prospects

The research conducted during this thesis allows to find answers for the underlying research questions, but also opened the path to future research.

One interesting future research path is to explore methodologies for ex-ante or endogenous determination of the optimal value for the exogenous parameters impacting the distributionally robust framework, e.g., the Wasserstein radius. This would allow to further improve and justify the applicability of the methodology for actual field operations. This, for example, may require a dedicated machine learning tool, leveraging information from past realizations.

Another potential research perspective relates to the modeling of recourse actions. In this thesis, we use linear decision rules for modeling the recourse actions in the real-time stage as affine functions of the uncertain parameters, thereby allowing to enhance the distributionally robust methodologies in a simplified framework. Extending the proposed approaches to a multi-stage setting with adaptive recourse may help achieving more accurate and optimal dispatch solution.

Differently, the consideration of multi-temporal uncertainties with temporal corre-

lations is a path towards advanced distributionally robust methodologies. The rationale behind this is to incorporate information on temporal correlations within the definition of ambiguity sets (e.g., leveraging conditional probability distribution functions). This enhancement should improve the decision-making process, especially in power system applications with inter-temporal restrictions.

Regarding neural-network-constrained methodologies, the contributions in this thesis opens the path for extracting the full potential of machine learning tools in view of their incorporation within optimization programs. For instance, quantile regression tools based on neural networks can be leveraged to produce probabilistic models, which may help modeling, e.g., chance constraints. Alternatively, having recourse to non-linear activation functions may further enhance the quality of the regression achieved by the neural networks. The challenge is this perspective is to accommodate the non-linear transfer function within the optimization program.

Finally, it is interesting to apply the proposed approaches to various short-term operational and long-term planning decision-making problems under uncertainty in power systems to further illustrate their potential benefits. It is interesting as well to explore the way neural-network-constrained optimization can handle other types of non-convex constraints, such as, bilinear or complementarity constraints, which affect other decision-making problems in energy systems. In that direction, combining distributionally robust with neural-network-constrained methodologies can represent a step forward in improving cost-efficiency and reliability of the operational procedures.

Appendices

A The IEEE 24-node Reliability Test System

Our case study in Chapters 4 and 5 is inspired from the IEEE reliability test system [127, 141]. All economic and technical data are provided in Table 8.1 and the transmission system topology is showcased in Fig. 8.1.

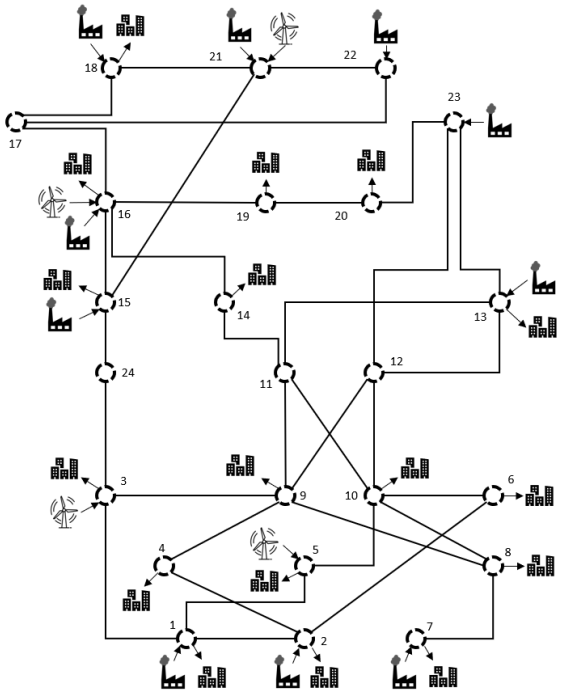


Figure 8.1: The network topology of the IEEE 24-node reliability test system [127].

In the following, we provide a description of the technical and economical parameters in Table 8.1. Technical data of conventional generating units includes their marginal generation cost C_e in €/MWh, upward balancing capacity procurement cost $\overline{C}_e$ in €/MW, downward balancing capacity procurement cost $\underline{C}_e$ in €/MW,

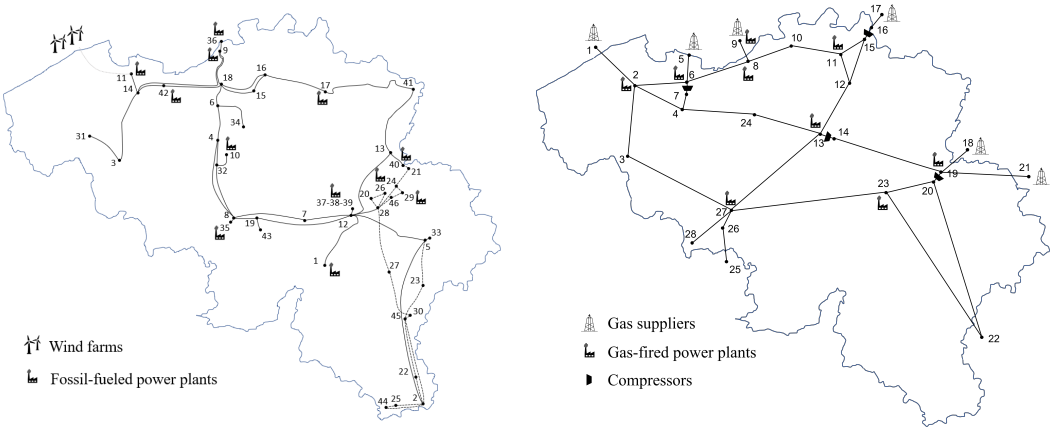
Table 8.1: Input Data

Generating units						1	2	3	4	5	6	7	8	9	10	11	12
Node						1	2	7	13	15	15	16	18	21	22	23	23
C_e [€ /MWh]						13.32	13.32	20.7	20.93	26.11	10.52	10.52	6.02	5.47	7	10.52	10.89
$\overline{C}_e$ [€ /MW]						1.68	1.68	3.30	4.07	1.89	5.48	5.48	4.98	5.53	8.00	3.45	5.11
$\underline{C}_e$ [€ /MW]						2.32	2.32	4.67	3.93	3.11	3.52	3.52	5.02	4.97	6.00	2.52	2.89
$\overline{P}_e$ [MW]						106.4	106.4	245	413.7	42	108.5	108.5	280	280	210	217	245
$\overline{R}_e$ [MW]						48	48	84	216	42	36	36	60	60	48	72	48
Wind farms														1	2	3	4
Node																16	21
$P_w^{\max}$ [MW]																500	500
Day-ahead wind power forecast [MW]																120.54	115.52
Loads	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Node	1	2	3	4	5	6	7	8	9	10	13	14	15	16	18	19	20
p_d [MWh]	84	75	139	58	55	106	97	132	135	150	205	150	245	77	258	141	100
V_{Shed} [€ /MWh]	500	500	500	500	500	500	500	500	500	500	500	500	500	500	500	500	500
Lines: From node	1	1	1	2	2	3	3	4	5	6	7	8	8	9	9	10	10
To node	2	3	5	4	6	9	24	9	10	10	8	9	10	11	12	11	12
$1/B$ [pu]	0.0146	0.2253	0.0907	0.1356	0.205	0.1271	0.084	0.111	0.094	0.0642	0.0652	0.1762	0.1762	0.084	0.084	0.084	0.084
$\overline{F}_\ell^{\max}$ [MW]	175	175	350	175	175	175	400	175	350	175	350	175	175	400	400	400	400
Lines: From node	11	11	12	12	13	14	15	15	15	16	16	17	17	18	19	20	21
To node	13	14	13	23	23	16	16	21	24	17	19	18	22	21	20	23	22
$1/B$ [pu]	0.0488	0.0426	0.0488	0.0985	0.0884	0.0594	0.0172	0.0249	0.0529	0.0263	0.0234	0.0143	0.1069	0.0132	0.0203	0.0112	0.0692
$\overline{F}_\ell^{\max}$ [MW]	500	500	500	500	250	250	500	400	500	500	500	500	500	1000	1000	1000	500

capacity $\overline{P}_e$ in MW, maximum upward and downward regulation capability $\overline{R}_e^{\max}$ in MW. In addition, Table 8.1 includes the wind farm maximum power capacity in nominal conditions $W_{(w,w)}$ in MW, as well as their day-ahead wind power forecast μ in MW. The 17 loads are characterized by their consumption p_d in MWh and the value of lost load V_{Shed} in €/MWh. Finally, the generating units, the wind farms, and the loads are connected via a 24-node network and their respective connecting node number are referred in the Table 8.1. The transmission lines are defined via their adjacent nodes, their per-unit inverse susceptance $1/B$ as well as their maximum line capacity $F_\ell^{\max}$ in MW.

B The Belgian Power and Gas Systems

In the following, we introduce our case study of Chapters 6 and 7, which is inspired from the Belgian power and gas systems [**?**]. The topology of electricity and natural networks are represented in Figs. 8.2(a) and 8.2(b), respectively. The economical and technical data are collected in Tables 8.2 and 8.3, respectively for the electricity and gas networks.



(a) Belgian electrical grid topology (plain lines are 380 kV and dashed lines are 220 kV).

(b) Belgian gas grid topology.

Fig. 8.2: The real-life Belgian power and gas systems.

We begin with the description of technical parameters related to the power system in Table 8.2. It includes information about electricity generating units such as location node, marginal generation cost C_e in €/MWh, minimum and maximum capacity $\underline{P}_e$ and $\overline{P}_e$ in MW.

Table 8.2: Technical data for electricity generating units, wind power generating units, electricity demands and transmission lines

Electricity generating units				1	2	3	4	5	6	7	8	9	10	11	12	13	14
Location [node]				29	42	24	10	11	1	36	35	17	9	37	38	39	40
Marginal generation cost c_e [€/MWh]				12.8	12.8	12.8	12.8	12.8	12.8	12.8	12.8	12.8	9.5	10	10	10	12.8
Capacity $\overline{P}_e$ [MW]				188	663	451	513	465	350	384	405	422	3000	1000	1000	1000	86
Minimum output power $\underline{P}_e$ [MW]				52.64	185.64	126.28	143.64	130.2	98.0	107.52	113.4	118.16	840.0	280.0	280.0	280.0	24.08
Wind power generating units						1	2	3	4	5	6	7	8	9	10	11	12
Location [node]						3, 11, 31	9, 14, 18, 42	8, 35, 43	15, 16, 17, 36	13, 41	6, 32	4, 10	19	5, 12, 20, 21, 24, 26, 28, 29, 33, 37, 38, 39, 40, 46	1, 7	2, 22, 23, 25, 27, 30, 44, 45	11
Installed capacity [MW]						133.3	114.9	113.4	169.4	109.4	85.4	101.7	5.3	240.3	48.3	102.7	2570.0
Electricity demands	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Location [node]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Electrical sharing [%]	2.73	3.7	5.39	4.82	2.68	0.0	2.36	0.0	0.0	2.68	2.41	2.41	0.0	0.0	4.82	2.41	5.21
Curtailment cost [€/MWh]	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000
Electricity demands	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Location [node]	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Electrical sharing [%]	5.39	2.87	0.48	0.32	2.47	0.0	0.0	0.46	5.68	0.68	0.0	5.09	0.0	0.0	5.39	8.04	5.65
Curtailment cost [€/MWh]	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000
Electricity demands						35	36	37	38	39	40	41	42	43	44	45	46
Location [node]						35	36	37	38	39	40	41	42	43	44	45	46
Electrical sharing [%]						2.97	0.0	0.0	4.82	0.0	2.41	3.38	0.0	0.94	1.34		
Curtailment cost [€/MWh]						8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000
Transmission lines: From node	1	2	3	3	3	3	4	4	4	4	5	5	5	5	6	6	6
To node	12	5	31	31	14	14	6	6	8	32	33	33	12	12	18	18	34
Reactance [per-unit]	0.121	0.313	0.075	0.075	0.130	0.130	0.051	0.053	0.155	0.046	0.007	0.007	0.145	0.145	0.051	0.050	0.071
Capacity [MW]	1184.7	1295.1	1295.1	1295.1	1184.4	1295.1	1184.4	1295.1	1295.1	1184.4	1184.4	1184.4	1184.4	1373.4	1295.1	1184.4	1295.1
Transmission lines: From node	6	7	7	8	8	8	8	9	9	9	9	9	10	11	12	12	12
To node	34	8	12	35	35	32	19	18	18	18	36	14	32	14	13	28	19
Reactance [per-unit]	0.069	0.145	0.106	0.006	0.006	0.110	0.041	0.075	0.068	0.075	0.024	0.226	0.034	0.039	0.140	0.048	0.185
Capacity [MW]	1373.4	1295.1	1295.1	1295.1	1184.4	1184.4	1184.4	1184.4	1295.1	1184.4	1184.4	1184.4	1295.1	1295.1	1184.4	1184.4	1184.4
Transmission lines: From node	12	12	12	12	13	13	14	15	15	15	16	16	17	18	19	2	2
To node	37	37	38	39	40	41	42	16	18	18	17	18	41	42	43	22	25
Reactance [per-unit]	0.010	0.010	0.009	0.011	0.034	0.139	0.046	0.046	0.069	0.069	0.107	0.115	0.189	0.112	0.025	0.044	0.062
Capacity [MW]	1184.4	1184.4	1295.1	1295.1	1295.1	1184.4	1295.1	1296.9	1295.1	1295.1	1295.1	1295.1	1295.1	1184.4	1295.1	364.5	305.1
Transmission lines: From node	2	2	20	20	21	21	5	22	23	24	24	25	26	27	27	28	29 , 30
To node	44	45	26	28	24	40	23	45	45	28	29	44	28	28	30	46	46, 45
Reactance [per-unit]	0.084	0.142	0.026	0.031	0.033	0.010	0.081	0.097	0.075	0.089	0.022	0.022	0.053	0.113	0.080	0.046	0.058, 0.021
Capacity [MW]	305.1	364.5	374.4	334.8	610.2	480.6	426.6	364.5	426.6	374.4	270.0	305.1	334.8	364.5	364.5	355.5	355.5, 364.5

The wind farms are divided into onshore and offshore regions. The aggregate wind power capacity $W_{(w,w)}$ in MW installed in the 11 provinces, is given in Table 8.2 corresponding to the wind farms numbered 1 to 11. The aggregate capacity is next equally distributed among the nodes of the corresponding province (i.e., location nodes). The offshore wind is represented by the wind farm numbered 12, which is connected to "Eeklo noord" node.

The electricity demands comprise residential and industrial compounds. Their respective location node, electrical sharing (among the nodes) in % and value of curtailed load V_{Shed} in €/MWh are referred in Table 8.2. The lines are characterized by the nodes they connect, their per-unit reactance as well as their maximum line capacity $\overline{F}_{\ell}^{\max}$ in MW.

We continue with the description of the natural gas transmission system, for which the technical and economical parameters are provided in Table 8.3. The gas system connects gas suppliers, gas demand centers as well as gas-fired power plants through a gas network composed of pipelines and compressors. In particular, the gas suppliers are characterized by their location node, marginal supplying cost C_g in €/MNm^3, and maximum debit in MNm^3/h.

The residential and industrial gas demands are distributed among regions, which is described by their location nodes. Border exchanges with France, Netherlands, and Luxembourg, are also considered assuming fixed and constant values. The nodes and transit values corresponding to theses exchanges are provided in Table 8.3.

Finally, the pipelines are characterized by their adjacent nodes, Weymouth constant $K_{m,n}$ in (MNm^3/(bar.h))^2 and linepack constant $S_{m,n}$ in MNm^3/bar. Note that the minimum and maximum pressures $\underline{\text{PR}}$ and $\overline{\text{PR}}$ are 40 and 85 bars. The gas compression factor Γ is equal for all compressors and is given by 1.5. The conversion factor $\eta_e = 5 \cdot 10^{-5}$.

C Procedure of the Out-of-Sample Analysis

In this thesis, we carry out an out-of-sample analysis to calculate unbiased performance metrics for the proposed optimization methods in Chapters 4, 5 and 6. In this section, we explain the basic principle of the out-of-sample simulation and we introduce the real-time optimization program used to run the simulation.

To perform an out-of-sample analysis, the available data is divided into two parts, i.e., *i)* the in-sample data (to run the optimization program and obtain optimal

Table 8.3: Technical data for gas suppliers, gas demands and pipelines

Gas suppliers											1	2	3	4	5	6
Location [node]											1	5	9	17	18	21
Marginal supplying cost [€/MNm^3]											2000	3000	4000	5000	6000	7000
Maximum debit [MNm^3/h]											5	1	0.5	2.2	1.5	1
Residential demands							1	2	3	4	5	6	7	8	9	
Location [node]							2	6	8	9	11	13	19	23	27	
Demand forecast[1] [MNm^3]							0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	
Industrial demands		1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Location [node]		2	3	4	8	11	12	13	14	15	19	23	26	27	28	
Demand forecast[2] [MNm^3]		0.0026	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026	
Border exchanges												1	2	3	4	
Location [node]												25	5	21	22	
Neighbouring country												France	Netherlands	Netherlands	Luxembourg	
Transit quantity[3] [MNm^3]												0.46	-0.48	-0.48	-0.002	
Pipelines: From node	1	2	2	9	8	10	11	17	15	11	12	13	18	20	20	
To node	2	6	4	8	10	11	15	16	12	12	13	27	19	23	22	
$K_{m,n}$ [(MNm^3/(bar.h))^2]	0.144	0.0827	0.0827	0.00312	0.00026	0.000936	0.000347	0.115	0.0149	0.000407	0.0140	0.00649	0.0305	0.00284	0.0000965	
$S_{m,n}$ [MNm^3/bar]	0.168	0.0349	0.0349	0.00106	0.0127	0.00353	0.00952	0.00635	0.0492	0.00811	0.0524	0.113	0.00380	0.0408	0.0342	
Pipelines: From node		23	23	26	27	27	27	2	5	7	4	24	14	6	21	
To node		22	27	25	26	28	3	3	6	4	24	13	19	8	19	
$K_{m,n}$ [(MNm^3/(bar.h))^2]		0.0000641	0.000841	0.0512	0.101	0.00601	0.0767	0.00739	0.132	0.120	0.003	0.0209	0.00794	0.0827	0.00794	
$S_{m,n}$ [MNm^3/bar]		0.0515	0.0722	0.0143	0.0309	0.0215	0.0142	0.147	0.0045	0.0198	0.0431	0.0744	0.196	0.0349	0.196	

[1] The total residential demand forecast equals 0.45 MNm^3 and is distributed uniformly among nodes with residential demands.

[2] The total industrial demand forecast equals 0.036 MNm^3 and is distributed uniformly among nodes with industrial demands.

[3] Negative quantities indicate an import of gas.

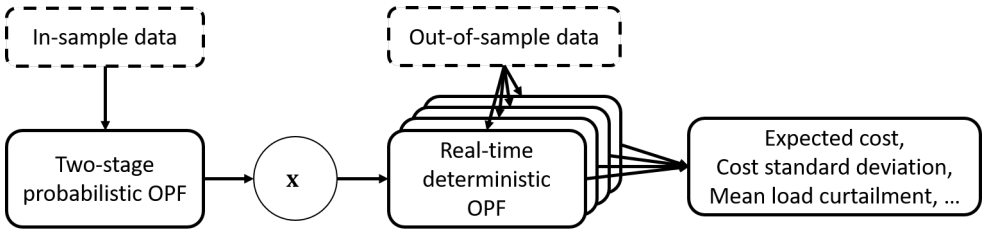


Fig. 8.3: Generic procedure for the out-of-sample analysis.

solutions), and *ii)* the out-of-sample data (to calculate unbiased performance metrics on unseen scenarios). The number of in-sample and out-of-sample scenarios is specified for each numerical experiment performed in Chapters 4, 5 or 6. The out-of-sample data is used as unseen realizations of uncertainty, e.g., of wind power generation in real-time.

The procedure of the out-of-sample analysis is schematically summarized in Fig. 8.3 and is defined as follows. We solve the day-ahead scheduling optimization programs with in-sample data and retrieved the optimal day-ahead scheduling decisions, e.g., dispatch of generating units or booked upward and downward balancing reserves (these decisions are generically named x in Figure 8.3). Next, we run a real-time optimization problem, once per each unseen realization in the out-of-sample data. By solving these problems, we compute the cost in the real-time operation for the recourse actions (including involuntarily load shedding and wind spillage) under numerous unseen scenarios. Hence, we are able to calculate the expected total operational cost of the system (which is the sum of day-ahead operational cost of the system and the average real-time operational cost calculated from the out-of-sample analysis) and its standard deviation over the out-of-sample simulations as well as the expected energy not served or the average wind spillage and report the results.

Given the real-time uncertainty realization $\widehat{\xi}_j$ that refers to, e.g., the wind power forecast error with respect to the day-ahead forecast μ, the formulation of this deterministic real-time optimization model is as follows[1]

[1]It is worth mentioning that we focus on the power system optimal power flow problem, as developed in Chapters 4 and 5. A similar formulation can straightforwardly be derived for the power and gas dispatch problem in Chapter 6, and is provided in [65].

$$\min_{V,\Delta d,\Delta w} \sum_{e\in\mathcal{E}} C_e V_{e,.}\widehat{\xi}_j + \sum_{d\in\mathcal{D}} V_d^{\text{Shed}}\Delta p_d \tag{8.1a}$$

$$\text{s.t. } 0 \le \Delta p_d \le p_d, \ \ \forall d \in \mathcal{D}, \tag{8.1b}$$

$$0 \le \Delta p_w \le W(\mu + \widehat{\xi}_j), \ \ \forall w \in \mathcal{W}, \tag{8.1c}$$

$$-\underline{r}_e \le V_{e,.}\widehat{\xi}_j \le \overline{r}_e, \ \ \forall e \in \mathcal{E}, \tag{8.1d}$$

$$\sum_{e\in\mathcal{E}} V_{e,.}\widehat{\xi}_j + \sum_{w\in\mathcal{W}} W_{w,w}\widehat{\xi}_j - \Delta p_w + \sum_{d\in\mathcal{D}} \Delta p_d = 0, \tag{8.1e}$$

$$\mathrm{M}_\ell^{\mathcal{E}}\left(p_e + V_{e,.}\widetilde{\xi}_j\right) + \mathrm{M}_\ell^{\mathcal{W}}\left(W\left(\mu + \widetilde{\xi}\right) - \Delta p_w\right) - \mathrm{M}_\ell^{\mathcal{D}}\left(d - \Delta p_d\right) \le \overline{F}_\ell \quad \forall \ell \in \mathcal{L}. \tag{8.1f}$$

The objective function (8.1a) models the total real-time operational cost of recourse actions, incurred by the recourse generation cost of conventional generating units (the first term) and the load shedding cost (the second term). The parameter V_d^{Shed} refers to the value of lost load. The wind spillage cost is assumed to be zero. Constraints (8.1b) and (8.1c) limit the wind power spillage Δp_w and the load shedding Δp_d. The recourse production $V\widehat{\xi}_j$ of conventional generating units is limited in (8.1d) by the reserve capacities $\underline{r}_e$ and $\overline{r}_e$ procured in the day-ahead stage. The real-time power balance is ensured by (8.1e). Finally, the maximum capacity limits are ensured by (8.1f) for each transmission line $\ell \in \mathcal{L}$.

D Proof of the First Assertion in Theorem 5.1

In this Section, we provide the proof for the first assertion in Theorem 5.1.

Proof. We depart from the worst-case expectation problem

$$\max_{\mathbb{Q}\in\ mathcalA_4} \ \mathbb{E}^{\mathbb{Q}}\left[a\left(x\right)^\top \widetilde{\xi} + b\left(x\right)\right], \tag{8.2}$$

where $\mathbb{Q}$ is the variable distribution function and $a\left(x\right)^\top \widetilde{\xi} + b\left(x\right) = a\left(x\right)^\top \widetilde{\xi} + b\left(x\right)$ represents the linear objective function, which depends on outer decision variables $x \in \mathcal{X}$. The problem seeks the distribution within the ambiguity set $mathcalA_4$, defined by (4), which maximizes the objective function in expectation. This problem can be developed as follows:

$$\max_{\mathbb{Q}} \ \mathbb{E}^{\mathbb{Q}}\left[a\left(x\right)^{\top}\widetilde{\xi}+b\left(x\right)\right] \tag{8.3a}$$

$$\text{s.t. } \mathrm{d}_W\left(\mathbb{Q},\widehat{\mathbb{Q}}_N\right)\leq\theta_1 \tag{8.3b}$$

$$\mathrm{d}_W\left(\mathbb{C},\widehat{\mathbb{C}}_N\right)\leq\theta_2. \tag{8.3c}$$

From now on, for the sake of clarity, we assume that distribution $\mathbb{Q}$ is discrete. Note that it is straightforward to extend the proof for continuous distribution functions. Note also that this assumption will be relaxed in the following developments, making the final result valid for continuous distributions as well. Given a set of fixed samples ξ_l (e.g., a mesh of the support of uncertainty) with $l\in\{1,...,L\}$, where L is an arbitrarily large number, the variable distribution $\mathbb{Q}$ is defined by

$$\mathbb{Q}=\sum_{l=1}^{L}\mathbb{Q}_l\,\delta_{\xi_l}, \tag{8.4}$$

where $\mathbb{Q}_l$ is the variable probability affected to sample ξ_l, and δ_{ξ_l} is the Dirac distribution centered on ξ_l. We are now able to reformulate (8.3) into

$$\max_{\mathbb{Q}_l} \ \sum_{l=1}^{L}\Psi\left(x,\xi_l\right)\mathbb{Q}_l \tag{8.5a}$$

$$\text{s.t. } \left\{\begin{array}{ll}\min\limits_{\pi_{li}^{(1)}} \ \sum\limits_{l=1}^{L}\sum\limits_{i=1}^{N} \mathrm{d}\left(\widehat{\xi}_i,\xi_l\right)\pi_{li}^{(1)} & \\ \text{s.t. } \sum\limits_{l=1}^{L}\pi_{li}^{(1)}=\left(\widehat{\mathbb{Q}}_N\right)_i & \forall i\in\{1,...,N\}\\ \quad\sum\limits_{i=1}^{N}\pi_{li}^{(1)}=\mathbb{Q}_l & \forall l\in\{1,...,L\}\end{array}\right\}\leq\theta_1 \tag{8.5b}$$

$$\left\{\begin{array}{ll}\min\limits_{\pi_{li}^{(2)}} \ \sum\limits_{l=1}^{L}\sum\limits_{i=1}^{N} \mathrm{d}_F\left(\widehat{\xi}_i,\xi_l\right)\pi_{li}^{(2)} & \\ \text{s.t. } \sum\limits_{l=1}^{L}\pi_{li}^{(2)}=\left(\widehat{\mathbb{C}}_N\right)_i & \forall i\in\{1,...,N\}\\ \quad\sum\limits_{i=1}^{N}\pi_{li}^{(2)}=\mathbb{C}_l & \forall l\in\{1,...,L\}\end{array}\right\}\leq\theta_2, \tag{8.5c}$$

where the expectation in the objective function has been developed for distribution $\mathbb{Q}$, and the Wasserstein distances $\mathrm{d}_W\left(\mathbb{Q},\widehat{\mathbb{Q}}_N\right)$ and $\mathrm{d}_W\left(\mathbb{C},\widehat{\mathbb{C}}_N\right)$ have been refor-

mulated according to their mathematical definitions. The variables $\pi_{li}^{(1)} \in \mathbb{R}^{L\times N}$ and $\pi_{li}^{(2)} \in \mathbb{R}^{L\times N}$ represent the optimal transportation plan between distributions $\widehat{\mathbb{Q}}_N$ and $\mathbb{Q}$, and $\widehat{\mathbb{C}}_N$ and $\mathbb{C}$, respectively. The distance functions $\mathrm{d}\left(\widehat{\xi}_i, \xi_l\right)$ and $\mathrm{d}_F\left(\widehat{\xi}_i, \xi_l\right)$ will be defined later on within the proof.

We reformulate (8.5) via three steps. First, we observe that the probabilities $\left(\widehat{\mathbb{P}}_N\right)_i = \left(\widehat{\mathbb{C}}_N\right)_i = \frac{1}{N}\ \forall i \in \{1, ..., N\}$ and $\mathbb{Q}_l = \mathbb{C}_l\ \forall l \in \{1, ..., L\}$. Next, we are able to drop the minimization operators since they appear in a "less-or-equal-to" constraint and to add the variables $\pi_{li}^{(1)}$ and $\pi_{li}^{(2)}$ to the set of overall variables. Finally, we drop variables $\mathbb{Q}_l$ by *(i)* substituting its expression in the third equation of (8.5b) into the objective function, and *(ii)* adding the constraint $\sum_{i=1}^{N} \pi_{li}^{(1)} = \sum_{i=1}^{N} \pi_{li}^{(2)}\ \forall l \in \{1, .., L\}$, which translates $\mathbb{Q}_l = \mathbb{C}_l\ \forall l \in \{1, ..., L\}$. We end up in

We are able to drop the min operators and add the variables $\pi_{ki}^{(1)}$ and $\pi_{ki}^{(2)}$ to the set of overall variables since they appear in a "less-or-equal-to" constraint. We observe that the probability of occurrence of each sample in the original distribution is equal to the its probability of occurrence in the copula distribution, $\mathbb{Q}_k = \mathbb{C}_k\ \forall k \in \{1, ..., K\}$.

$$\max_{\pi_{li}^{(1)}, \pi_{li}^{(2)}} \sum_{l=1}^{L}\sum_{i=1}^{N} \Psi\left(x, \xi_l\right) \pi_{li}^{(1)} \tag{8.6a}$$

$$\text{s.t.} \sum_{l=1}^{L}\sum_{i=1}^{N} \mathrm{d}\left(\widehat{\xi}_i, \xi_l\right) \pi_{li}^{(1)} \leq \theta_1 \quad : \alpha \tag{8.6b}$$

$$\sum_{l=1}^{L} \pi_{li}^{(1)} = \frac{1}{N}\ \forall i \in \{1, ..., N\} \quad : y_i \tag{8.6c}$$

$$\sum_{l=1}^{L}\sum_{i=1}^{N} \mathrm{d}_F\left(\widehat{\xi}_i, \xi_l\right) \pi_{li}^{(2)} \leq \theta_2 \quad : \beta \tag{8.6d}$$

$$\sum_{l=1}^{L} \pi_{li}^{(2)} = \frac{1}{N} \forall i \in \{1, ..., N\} \quad : t_i \tag{8.6e}$$

$$\sum_{i=1}^{N} \pi_{li}^{(1)} = \sum_{i=1}^{N} \pi_{li}^{(2)} \forall l \in \{1, .., L\} \quad : \sigma_l, \tag{8.6f}$$

where $\{\alpha, \beta, y_i, t_i, \sigma_l\}$ is the set of dual variables. Next, we dualize (8.6), yielding

$$\min_{\alpha\geq 0,\beta\geq 0,y_i,t_i,\sigma_l} \ \alpha\theta_1 + \beta\theta_2 + \frac{1}{N}\sum_{i=1}^{N} y_i + \frac{1}{N}\sum_{i=1}^{N} t_i \tag{8.7a}$$

$$\text{s.t. } y_i \geq \Psi\left(x,\xi_l\right) - \alpha\,\mathrm{d}\left(\widehat{\xi}_i,\xi_l\right) + \sigma_l \quad \forall i,\forall l \tag{8.7b}$$

$$t_i \geq -\beta\,\mathrm{d}_F\left(\widehat{\xi}_i,\xi_l\right) - \sigma_l \quad \forall i,\forall l. \tag{8.7c}$$

The constraints in (8.7) can be equivalently reformulated using maximization operators:

$$\min_{\alpha\geq 0,\beta\geq 0,y_i,t_i,\sigma_l} \ \alpha\theta_1 + \beta\theta_2 + \frac{1}{N}\sum_{i=1}^{N} y_i + \frac{1}{N}\sum_{i=1}^{N} t_i \tag{8.8a}$$

$$\text{s.t. } y_i \geq \max_{\xi_l,\sigma_l} \ \Psi\left(x,\xi_l\right) - \alpha\,\mathrm{d}\left(\widehat{\xi}_i,\xi_l\right) + \sigma_l \quad \forall i \tag{8.8b}$$

$$t_i \geq \max_{\xi_l,\sigma_l} -\beta\,\mathrm{d}_F\left(\widehat{\xi}_i,\xi_l\right) - \sigma_l \quad \forall i. \tag{8.8c}$$

The maximization operators in (8.8b) and (8.8c) can be moved to the objective function, while dropping auxiliary variables y_i and t_i. The summation and maximization operators are merged such that the variable σ_l simplifies. Furthermore, we relax the assumption on discrete set of sample ξ_l by introducing a continuous support of uncertainty $\Xi = \{\xi \in \Xi \mid C\xi \leq d\}$. Finally, for the ease of developments, we re-introduce auxiliary variable y_i, such that (8.8) is equivalent to

$$\min_{\alpha,\beta\geq 0,y_i} \ \alpha\theta_1 + \beta\theta_2 + \frac{1}{N}\sum_{i=1}^{N} y_i \tag{8.9a}$$

$$\text{s.t. } y_i \geq \max_{\xi\in\Xi} \ a\left(x\right)^{\top}\xi + b\left(x\right) - \alpha\,\mathrm{d}\left(\widehat{\xi}_i,\xi\right) - \beta\,\mathrm{d}_F\left(\widehat{\xi}_i,\xi\right) \ \ \forall i. \tag{8.9b}$$

■

List of Relevant Publications

The following publications report the research contributions that are embodied in this dissertation.

Journal Papers

[A] **A. Arrigo**, C. Ordoudis, J. Kazempour, Z. De Grève, J.-F. Toubeau, and F. Vallée, "Wasserstein distributionally robust chance-constrained optimization for energy and reserve dispatch: An exact and physically-bounded formulation," European Journal of Operations Research, vol. 269, no. 1, pp. 304-322, January 2022.

[B] **A. Arrigo**, J. Kazempour, Z. De Grève, J.-F. Toubeau, and F. Vallée, "Embedding Dependencies Between Wind Farms in Uncertainty-Aware Optimal Power Flow," 2022, submitted to IEEE Transactions on Power Systems, under the second round of review.

Conference Papers

[C] **A. Arrigo**, J. Kazempour, Z. De Grève, J.-F. Toubeau, and F. Vallée, "Enhanced Wasserstein Distributionally Robust OPF With Dependence Structure and Support Information," IEEE PowerTech Conference, Madrid, June 2021.

[D] **A. Arrigo**, J.-F. Toubeau, I. Fattahi, Z. De Grève and François Vallée, "Distributionally Robust Power and Gas Dispatch with Multivariate and Correlated Uncertainty," IEEE PMAPS Conference, Manchester, June 2022.

Other Reports

[E] **A. Arrigo**, M. Dolányi, K. Bruninx, J.-F. Toubeau, "Machine Learning for Improved Gas Network Modeling in Coordinated Energy Systems," Working paper[2].

[2]This work is foreseen to be submitted as a journal paper by October 2022.

The following works have been realized during the course of the PhD thesis. Their content is relevant but is not directly included in this dissertation.

Journal Papers

[F] A. Schwele, **A. Arrigo**, C. Vervaeren, J. Kazempour, and F. Vallée, "Coordination of electricity, heat, and natural gas systems accounting for network flexibility," Electric Power Systems Research, 189:106776, 2020.

Conference Papers

[G] **A. Arrigo**, C. Ordoudis, J. Kazempour, Z. de Grève, J.-F. Toubeau, and F. Vallée, "Optimal power flow under uncertainty: An extensive out-of-sample analysis," in 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGTEurope), pages 1–5, 2019.

[H] J. Bottieau, **A. Arrigo**, Z. D. Grève, F. Vallée, and J.-F. Toubeau, "A distributionally robust framework for providing passive balancing services," IEEE PowerTech Conference, Madrid, June 2021.

Bibliography

[1] Wind Europe, "Wind energy in Europe: 2021 statistics and the outlook for 2022-2026," 2022. [Online]. Accessed on 11$^{\text{th}}$ April 2022: windeurope.org.

[2] L. Roald, "Optimization methods to manage uncertainty and risk in power systems operation," *Ph.D. Dissertation*, 2016. ETH Zurich.

[3] P. J. Hibbard and T. Schatzki, "The interdependence of electricity and natural gas: Current factors and future prospects," *The Electricity Journal*, vol. 25, no. 4, pp. 6–17, 2012.

[4] C. Ordoudis, "Market-based approaches for the coordinated operation of electricity and natural gas systems," *Ph.D. Dissertation*, 2018. Technical University of Denmark.

[5] J. Bottieau, L. Hubert, Z. De Grève, F. Vallée, and J.-F. Toubeau, "Very-short-term probabilistic forecasting for a risk-aware participation in the single price imbalance settlement," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1218–1230, 2020.

[6] J.-F. Toubeau, J. Bottieau, F. Vallée, and Z. De Grève, "Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1203–1215, 2019.

[7] A. Shapiro and A. Philpott, "A tutorial on stochastic programming," 2007.

[8] F. Abbaspourtorbati and M. Zima, "The swiss reserve market: Stochastic programming in practice," *IEEE Transactions on power systems*, vol. 31, no. 2, pp. 1188–1194, 2016.

[9] A. Papavasiliou, A. Bouso, S. Apelfrojd, E. Wik, T. Gueuning, and Y. Langer, "Multi-area reserve dimensioning using chance-constrained optimization," *IEEE Transactions on Power Systems*, 2021. In Press.

[10] A. Papavasiliou and S. S. Oren, "Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network," *Operations research*, vol. 61, no. 3, pp. 578–592, 2013.

[11] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," *SIAM Review*, vol. 53, no. 3, pp. 464–501, 2011.

[12] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM Journal of Optimization*, vol. 17, pp. 969–996, 2007.

[13] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, "Integrating renewables in electricity markets: Operational problems," *International Series in Operations Research and Management Science*, 2014. Springer.

[14] J. R. Birge and F. Louveaux, "Introduction to stochastic programming," *Springer Series in Operations Research and Financial Engineering*, 2011. Springer.

[15] A. J. Kleywegt, A. Shapiro, and T. Homem-de Mello, "The sample average approximation method for stochastic discrete optimization," *SIAM Journal on Optimization*, vol. 12, no. 2, pp. 479–502, 2002.

[16] M. Zugno and A. J. Conejo, "A robust optimization approach to energy and reserve dispatch in electricity markets," *European Journal of Operational Research*, vol. 247, no. 2, pp. 659–671, 2015.

[17] R. Henrion, "Introduction to chance constraint programming," 2004.

[18] L. Roald, D. Pozo, A. Papavasiliou, D. K. Molzahn, J. Kazempour, and A. J. Conejo, "Power systems optimization under uncertainty: A review of methods and applications," *PSCC 2022*. forthcoming in Electric Power Systems Research.

[19] H. Scarf, "A min-max solution of an inventory problem," *Studies in The Mathematical Theory of Inventory and Production*, pp. 201–209, 1958.

[20] P. Mohajerin Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations," *Mathematical Programming*, vol. 171, no. 1-2, pp. 115–166, 2017.

[21] H. Rahimian and S. Mehrotra, "Distributionally robust optimization: A review," *Optimization Online*, August 2019. Available: www.optimization-online.org/DB_HTML/2019/08/7332.html.

[22] W. Wiesemann, D. Kuhn, and M. Sim, "Distributionally robust convex optimization," *Operations Research*, vol. 62, no. 6, pp. 1358–1376, 2014.

[23] Y. Zhang, S. Shen, and J. L. Mathieu, "Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads," *IEEE Transactions on Power Systems*, vol. 32, pp. 1378–1388, March 2017.

[24] C. Zhao and R. Jiang, "Distributionally robust contingency-constrained unit commitment," *IEEE Transactions on Power Systems*, vol. 33, pp. 94–102, January 2018.

[25] R. Mieth and Y. Dvorkin, "Data-driven distributionally robust optimal power flow for distribution systems," *IEEE Control Systems Letters*, vol. 2, pp. 363–368, July 2018.

[26] F. Pourahmadi, J. Kazempour, C. Ordoudis, P. Pinson, and S. H. Hosseini, "Distributionally robust chance-constrained generation expansion planning," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 2888–2903, 2020.

[27] C. Duan, W. Fang, L. Jiang, L. Yao, and J. Liu, "Distributionally robust chance-constrained approximate AC-OPF with Wasserstein metric," *IEEE Transactions on Power Systems*, vol. 33, pp. 4924–4936, September 2018.

[28] Y. Guo, K. Baker, E. Dall'Anese, Z. Hu, and T. H. Summers, "Data-based distributionally robust stochastic optimal power flow," *IEEE Transactions on Power Systems*, vol. 34, pp. 1483–1492, March 2019.

[29] B. K. Poolla, A. R. Hota, S. Bolognani, D. S. Callaway, and A. Cherukuri, "Wasserstein distributionally robust look-ahead economic dispatch," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2010–2022, 2021.

[30] A. Esteban-Pérez and J. M. Morales, "Partition-based distributionally robust optimization via optimal transport with order cone constraints," *Quarterly Journal of Operations Research*, 2021.

[31] B. Li, R. Jiang, and J. L. Mathieu, "The value of including unimodality information in distributionally robust optimal power flow," 2018. arXiv preprint, arXiv:1811.10217.

[32] R. Gao and A. J. Kleywegt, "Distributionally robust stochastic optimization with dependence structure." arXiv preprint, arXiv:1701.04200, 2017.

[33] C. Wang, R. Gao, F. Qiu, J. Wang, and L. Xin, "Risk-based distributionally robust optimal power flow with dynamic line rating," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6074–6086, 2018.

[34] Z. Chen, D. Kuhn, and W. Wiesemann, "Data-driven chance constrained programs over wasserstein balls," *Optimization Online*, 2018. Available: http://www.optimization-online.org/DB_FILE/2018/06/6671.pdf.

[35] D. K. Molzahn and I. A. Hiskens, "A survey of relaxations and approximations of the power flow equations," *Foundations and Trends® in Electric Energy Systems*, vol. 4, no. 1-2, pp. 1–221, 2019.

[36] European Commission, "Commission regulation 2015/1222 of 24 july 2015 establishing a guideline on capacity allocation and congestion management," 2015.

[37] K. V. den Bergh, D. Couckuyt, E. Delarue, and W. D'haeseleer, "Redispatching in an interconnected electricity system with high renewables penetration," *Electric Power Systems Research*, vol. 127, no. 10, pp. 64–72, 2015.

[38] W. Hogan, "Handbook on the economics of electricity - chapter 7: Strengths and weaknesses of the PJM market model," *Edwar Elgar Publishing*, 2019.

[39] J. Glachant, M. Hallack, and M. Vasquez, "Building competitve gas markets in the EU," *Edwar Elgar Publishing*, 2013.

[40] 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.

[41] C. Coffrin, H. L. Hijazi, and P. Van Hentenryck, "The QC relaxation: A theoretical and computational study on optimal power flow," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3008–3018, 2016.

[42] J. Lavaei and S. H. Low, "Zero duality gap in optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 92–107, 2012.

[43] D. K. Molzahn, J. T. Holzer, B. C. Lesieutre, and C. L. DeMarco, "Implementation of a large-scale optimal power flow solver based on semidefinite programming," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3987–3998, 2013.

[44] H. Wu, M. Shahidehpour, Z. Li, and W. Tian, "Chance-constrained day-ahead scheduling in stochastic power system operation," *IEEE Transactions on Power Systems*, vol. 29, no. 4, pp. 1583–1591, 2014.

[45] A. Zlotnik, L. Roald, S. Backhaus, M. Chertkov, and G. Andersson, "Coordinated scheduling for interdependent electric power and natural gas infrastructures," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 600–610, 2017.

[46] G. Byeon and P. V. Hentenryck, "Unit commitment with gas network awareness," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1327–1339, 2020.

[47] C. Coffrin, H. Hijazi, and P. Van Hentenryck, "Network flow and copper plate relaxations for ac transmission systems," in *2016 Power Systems Computation Conference (PSCC)*, pp. 1–8, 2016.

[48] C. Coffrin and L. Roald, "Convex relaxations in power system optimization: A brief introduction," 2018. arXiv preprint, arXiv:1807.07227.

[49] A. Ratha, P. Pinson, H. L. Cadre, A. Virag, and J. Kazempour, "Moving from linear to conic markets for electricity," 2021. ArXiv preprint, arXiv:2103.12122.

[50] S. Chen, A. J. Conejo, R. Sioshansi, and Z. Wei, "Unit commitment with an enhanced natural gas-flow model," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3729–3738, 2019.

[51] C. Coffrin, H. L. Hijazi, and P. Van Hentenryck, "Strengthening the sdp relaxation of ac power flows with convex envelopes, bound tightening, and valid inequalities," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3549–3558, 2017.

[52] C. M. Correa-Posada and P. Sanchez-Martin, "Integrated power and natural gas model for energy adequacy in short-term operation," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3347–3355, 2015.

[53] C. M. Correa-Posada and P. Sanchez-Martin, "Gas network optimization: A comparison of piecewise linear models," 2014.

[54] C. Shao, X. Wang, M. Shahidehpour, X. Wang, and B. Wang, "An milp-based optimal power flow in multicarrier energy systems," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 1, pp. 239–248, 2017.

[55] R. Canyasse, G. Dalal, and S. Mannor, "Supervised learning for optimal power flow as a real-time proxy," in *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, 2017.

[56] X. Pan, T. Zhao, M. Chen, and S. Zhang, "Deepopf: A deep neural network approach for security-constrained dc optimal power flow," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 1725–1735, 2021.

[57] F. Fioretto, T. W. Mak, and P. Van Hentenryck, "Predicting ac optimal power flows: Combining deep learning and lagrangian dual methods," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 630–637, Apr. 2020.

[58] I. Murzakhanov, A. Venzke, G. S. Misyris, and S. Chatzivasileiadis, "Neural networks for encoding dynamic security-constrained optimal power flow," 2021. arXiv preprint, arXiv:2003.07939.

[59] J. Bezanson and S. Karpinski, "Julia: A fast dynamic language for technical computing," 2012.

[60] I. Dunning, J. Huchette, and M. Lubin, "Jump: A modeling language for mathematical optimization," *SIAM Review*, vol. 59, no. 2, pp. 295–320, 2017.

[61] Mosek, "Mosek modeling cookbook," 2020.

[62] A. Arrigo, C. Ordoudis, J. Kazempour, Z. D. Grève, J.-F. Toubeau, and F. Vallée, "Online companion - distributionally robust chance constrained energy and reserve dispatch: An exact and physically bounded approach," 2021. Available: `https://doi.org/10.5281/zenodo.3540810`.

[63] A. Arrigo, J. Kazempour, Z. De Grève, J.-F. Toubeau, and F. Vallée, "Online companion: Enhanced Wasserstein distributionally robust OPF with dependence structure and support information," 2020. Available: `https://doi.org/10.5281/zenodo.4293635`.

[64] A. Arrigo, J. Kazempour, Z. De Grève, J.-F. Toubeau, and F. Vallée, "Online companion: Embedding dependencies between wind farms in robust optimal power flow." Available: `https://doi.org/10.5281/zenodo.5997537`, 2021.

[65] A. Arrigo, J.-F. Toubeau, I. Fattahi, Z. D. Grève, J. Kazempour, and F. Vallée, "Online companion - distributionally robust power and gas dispatch with multivariate and correlated uncertainty." Available: `https://doi.org/10.5281/zenodo.5841182`, 2022.

[66] European Commission, "EU energy in figures, statistical pocketbook 2021," *Publications Office*, 2021.

[67] Intergovernmental Panel on Climate Change, "Climate change 2021: The physical science basis - summary for policymakers," 2021. Sixth assessment report.

[68] University of Colorado: Sea Level Research Group, "Most recent global mean sea level release," 2021. [Online]. Available: www.sealevel.colorado.edu.

[69] National Oceanic and Atmospheric Administration, "Climate at a glance," 2021. [Online]. Available: www.ncdc.noaa.gov/cag.

[70] European Commission, "Commission communication COM/2019/640 of 11 december 2019: The european green deal," 2019.

[71] European Commission, "Commission communication COM/2014/015 of 22 january 2014: A policy framework for climate and energy in the period from 2020 to 2030," 2014.

[72] Florence School of Regulation, "The clean energy for all europeans package," June 2021. [Online]. Available: www.fsr.eui.eu/the-clean-energy-for-all-europeans-package.

[73] European Commission, "Commission communication COM/2021/550 of 14 July 2021: 'Fit for 55': delivering the EU's 2030 Climate Target on the way to climate neutrality," 2021.

[74] European Environment Agency, "EEA report no 13/2021: Trends and projections in europe 2021," 2021.

[75] European Commission, "Commission regulation 2017/1485 of 2 August 2017 establishing a guideline on electricity transmission system operation," 2017.

[76] J. Wang, X. Wang, and Y. Wu, "Operating reserve model in the power market," *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 223–229, 2005.

[77] G. Bertrand and A. Papavasiliou, "Adaptive trading in continuous intraday electricity markets for a storage unit," *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 2339–2350, 2020.

[78] European Commission, "Commission regulation 2017/2195 of 23 November 2017 establishing a guideline on electricity balancing," 2017.

[79] A. Belderbos, *Storage via power-to-gas in future energy systems - The need for synthetic fuel storage in systems with high shares of intermittent renewables*. 2019. Ph.D. dissertation, KULeuven.

[80] C. Ordoudis, *Market-based Approaches for the Coordinated Operation of Electricity and Natural Gas Systems*. 2018. Ph.D. dissertation, DTU.

[81] P. Denholm, T. Mai, R. W. Kenyon, B. Kroposki, and M. O'Malley, "Inertia and the power grid: A guide without the spin," *National Renewable Energy Laboratory*, 2020.

[82] J. Matevosyan, J. MacDowell, N. Miller, B. Badrzadeh, D. Ramasubramanian, A. Isaacs, R. Quint, E. Quitmann, R. Pfeiffer, H. Urdal, T. Prevost, V. Vittal, D. Woodford, S. H. Huang, and J. O'Sullivan, "A future with

inverter-based resources: Finding strength from traditional weakness," *IEEE Power and Energy Magazine*, vol. 19, no. 6, pp. 18–28, 2021.

[83] B. Bakhshideh Zad, H. Hasanvand, J. Lobry, and F. Vallée, "Optimal reactive power control of dgs for voltage regulation of mv distribution systems using sensitivity analysis method and pso algorithm," *International Journal of Electrical Power and Energy Systems*, vol. 68, pp. 52–60, 2015.

[84] D. Stenclik, A. Bloom, W. Cole, G. Stephen, A. F. Acevedo, R. Gramlich, C. Dent, N. Schlag, and M. Milligan, "Quantifying risk in an uncertain future: The evolution of resource adequacy," *IEEE Power and Energy Magazine*, vol. 19, no. 6, pp. 29–36, 2021.

[85] J. Bushnell, "To fix the power market, first fix the natural gas market," *UC Berkeley Energy Institute Blog*, 2021. [Online]. Available: https://energyathaas.wordpress.com/2021/03/01/to-fix-the-power-market-first-fix-the-natural-gas-market.

[86] ELIA, "Adequacy and flexibility study for belgium 2022-2032," 2021.

[87] M. Hogan, "Follow the missing money: Ensuring reliability at least cost to consumers in the transition to a low-carbon power system," *The Electricity Journal*, vol. 30, no. 1, pp. 55–61, 2017.

[88] UCTE, "Final report of the investigation committee on the 28 september 2003 blackout in italy," 2004.

[89] E. Ela and B. Kirby, "ERCOT Event on February 26, 2008: Lessons Learned," *National Renewable Energy Laboratory*, 2008.

[90] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of Risk*, vol. 2, no. 3, pp. 21–41, 2000.

[91] D. Bienstock, M. Chertkov, and S. Harnett, "Chance-constrained optimal power flow: Risk-aware network control under uncertainty," *SIAM review*, no. 3, pp. 461–495, 2014.

[92] A. Ben-Tal, D. den Hertog, A. D. Waegenaere, B. Melenberg, and G. Rennen, "Robust solutions of optimization problems affected by uncertain probabilities," *Management Science*, vol. 59, pp. 341–357, 2013.

[93] Z. Hu and J. Hong, "Kullback-leibler divergence constrained distributionally robust optimization," 2013.

[94] S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Mathematical Programming*, vol. 137, no. 1-2, pp. 167–198, 2013.

[95] R. Gao and A. Kleywegt, "Distributionally robust stochastic optimization with wasserstein distance." arXiv preprint, arXiv:1604.02199, 2016.

[96] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Operations Research*, vol. 58, no. 3, pp. 595–612, 2010.

[97] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Operations Research*, vol. 58, no. 3, pp. 595–612, 2010.

[98] L. V. Kantorovich and G. S. Rubinshtein, "On a space of totally additive functions," *Vestnik Leningradskogo Universiteta*, vol. 13, pp. 52–59, 1958.

[99] A. Shapiro, "Distributionally robust stochastic programming," *SIAM Journal on Optimization*, vol. 27, no. 4, pp. 2258–2275, 2017.

[100] A. Shapiro, "Tutorial on risk neutral, distributionally robust and risk averse multistage stochastic programming," *European Journal of Operational Research*, vol. 288, no. 1, pp. 1–13, 2021.

[101] D. Kuhn, W. Wiesemann, and A. Georghiou, "Primal and dual linear decision rules in stochastic and robust optimization," *Mathematical Programming*, vol. 130, no. 1, pp. 177–209, 2011.

[102] A. Georghiou, W. Wiesemann, and D. Kuhn, "Generalized decision rule approximations for stochastic programming via liftings," *Mathematical Programming*, vol. 152, no. 1, pp. 301–338, 2015.

[103] M. F. Anjos and A. J. Conejo. 2017. Now Foundations and Trends.

[104] T. V. Jensen, J. Kazempour, and P. Pinson, "Cost-optimal atcs in zonal electricity markets," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 3624–3633, 2018.

[105] M. Lubin, Y. Dvorkin, and L. Roald, "Chance constraints for improving the security of ac optimal power flow," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 1908–1917, 2019.

[106] P. González, J. Villar, C. A. Díaz, and F. A. Campos, "Joint energy and reserve markets: Current implementations and modeling trends," *Electric Power Systems Research*, vol. 109, pp. 101–111, 2014.

[107] R. Domínguez, G. Oggioni, and Y. Smeers, "Reserve procurement and flexibility services in power systems with high renewable capacity: Effects of integration on different market designs," *International Journal of Electrical Power & Energy Systems*, vol. 113, pp. 1014–1034, 2019.

[108] A. Papavasiliou, S. S. Oren, and R. P. O'Neill, "Reserve requirements for wind power integration: A scenario-based stochastic programming framework," *IEEE Transactions on Power Systems*, vol. 26, pp. 2197–2206, November 2011.

[109] J. Toubeau, Z. D. Grève, and F. Vallée, "Medium-term multimarket optimization for virtual power plants: A stochastic-based decision environment," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1399–1410, 2018.

[110] J. Morales, M. Zugno, S. Pineda, and P. Pinson, "Electricity market clearing with improved scheduling of stochastic production," *European Journal of Operational Research*, vol. 235, pp. 765–774, Jun. 2014.

[111] S. Martin, Y. Smeers, and J. A. Aguado, "A stochastic two settlement equilibrium model for electricity markets with wind generation," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 233–245, 2015.

[112] B. F. Hobbs and S. S. Oren, "Three waves of US reforms: Following the path of wholesale electricity market restructuring," *IEEE Power and Energy Magazine*, vol. 17, no. 1, pp. 73–81, 2019.

[113] E. Litvinov, F. Zhao, and T. Zheng, "Electricity markets in the United States: Power industry restructuring processes for the present and future," *IEEE Power & Energy Magazine*, vol. 17, no. 1, pp. 32–42, 2019.

[114] J. M. Morales, A. J. Conejo, and J. Perez-Ruiz, "Economic valuation of reserves in power systems with high penetration of wind power," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 900–910, 2009.

[115] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 52–63, 2013.

[116] M. Zugno and A. J. Conejo, "A robust optimization approach to energy and reserve dispatch in electricity markets," *European Journal of Operational Research*, vol. 247, pp. 659–671, 2015.

[117] M. Lubin, Y. Dvorkin, and S. Backhaus, "A robust approach to chance constrained optimal power flow with renewable generation," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3840–3849, 2016.

[118] C. E. Bonferroni, "Teoria statistica delle classi e calcolo delle probabilità," *Libreria internazionale Seeber*, 1936.

[119] T. Jonsson, P. Pinson, and H. Madsen, "On the market impact of wind energy forecasts," *Energy Economics*, vol. 32, no. 2, pp. 313–320, 2010.

[120] P. Pinson, "Wind energy: Forecasting challenges for its operational management," *Statistical Science*, vol. 28, no. 4, pp. 564–585, 2013.

[121] J. Bottieau, L. Hubert, Z. D. Grève, F. Vallée, and J.-F. Toubeau, "Very short-term probabilistic forecasting for a risk-aware participation in the single price imbalance settlement," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1218–1230, 2020.

[122] C. Ordoudis, V. A. Nguyen, D. Kuhn, and P. Pinson, "Energy and reserve dispatch with distributionally robust joint chance constraints," *Operations Research Letters*, vol. 49, no. 3, pp. 291–299, 2021.

[123] W. Xie, "On distributionally robust chance constrained program with Wasserstein distance," *Mathematical Programming*, vol. 186, pp. 115–155, 2021.

[124] A. Wachter and L. Biegler, "On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, no. 1, pp. 25–57, 2006.

[125] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM Journal of Optimization*, vol. 17, no. 4, pp. 969–996, 2007.

[126] J. Blanchet, Y. Kang, and K. Murthy, "Robust Wasserstein profile inference and applications to machine learning," *Journal of Applied Probability*, vol. 56, no. 3, pp. 830–857, 2019.

[127] C. Grigg, P. Wong, P. Albrecht, R. Allan, M. Bhavaraju, R. Billinton, Q. Chen, C. Fong, S. Haddad, S. Kuruganty, W. Li, R. Mukerji, D. Patton, N. Rau, D. Reppen, A. Schneider, M. Shahidehpour, and C. Singh, "The IEEE reliability test system 1996. a report prepared by the reliability test system task force of the application of probability methods subcommittee," *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 1010–1020, 1999.

[128] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[129] M. S. Andersen, "Bilinear constraints, hackmd note," 2021.

[130] G. P. McCormick, "Computability of global solutions to factorable nonconvex programs: Part I — Convex underestimating problems," *Mathematical Programming*, vol. 10, no. 1, pp. 147–175, 1976.

[131] G. C. Pflug and M. Pohl, "A review on ambiguity in stochastic portfolio optimization," *Set-Valued and Variational Analysis*, vol. 26, p. 733–757, 2018.

[132] R. B. Neslen, *An Introduction to Copulas*. Springer, 1999. New York, USA.

[133] A. Sklar, "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'institut de statistiques de l'université de Paris*, vol. 8, pp. 229–231, 1959.

[134] F. Vallée, J. Lobry, and O. Deblecker, "Impact of the wind geographical correlation level for reliability studies," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 2232–2239, 2007.

[135] Y. Guo, K. Baker, E. Dall'Anese, Z. Hu, and T. H. Summers, "Data-based distributionally robust stochastic optimal power flow – Part I: Methodologies," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1483–1492, 2019.

[136] T. Ding, Q. Yang, X. Liu, C. Huang, Y. Yang, M. Wang, and F. Blaabjerg, "Duality-free decomposition based data-driven stochastic security-constrained unit commitment," *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 82–93, 2019.

[137] A. Arrigo, J. Kazempour, Z. D. Z., J.-F. Toubeau, and F. Vallée, "Enhanced Wasserstein distributionally robust OPF with dependence structure and support information," in *IEEE PowerTech 2021 Conference*, 2021. Madrid, Spain.

[138] S. Agrawal, Y. Ding, A. Saberi, and Y. Ye, "Price of correlations in stochastic optimization," *Operations Research*, vol. 60, p. 150–162, 2012.

[139] R. Gao and A. Kleywegt, "Data-driven robust optimization with known marginal distributions," 2017.

[140] I. Pólik and T. Terlaky, "A survey of the S-Lemma," *SIAM Review*, vol. 49, no. 3, pp. 371–418, 2007.

[141] C. Ordoudis, P. Pinson, J. M. Morales, and M. Zugno, "An updated version of the IEEE RTS 24-bus system for electricity market and power system operation studies," 2016.

[142] C. Ordoudis, P. Pinson, and J. M. Morales, "An integrated market for electricity and natural gas systems with stochastic power producers," *European Journal of Operational Research*, vol. 272, no. 2, pp. 642–654, 2019.

[143] B. Odetayo, J. MacCormack, W. Rosehart, and H. Zareipour, "A chance-constrained programming approach to integrated planning of distributed power generation and natural gas network," *Electric Power Systems Research*, vol. 151, pp. 197–207, 2017.

[144] M. Qadrdan, J. Wu, N. Jenkins, and K. Ekanayake, "Operating strategies for a gb integrated gas and electricity network considering the uncertainty in wind power forecasts," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 1, pp. 128–138, 2014.

[145] C. Wang, W. Wei, J. Wang, and T. Bi, "Convex optimization based adjustable robust dispatch for integrated electric-gas systems considering gas delivery priority," *Applied Energy*, vol. 239, pp. 70–82, 2019.

[146] C. He, L. Wu, T. Liu, and M. Shahidehpour, "Robust co-optimization scheduling of electricity and natural gas systems via admm," *IEEE Transactions on Sustainable Energy*, vol. 8, pp. 658–670, 2016.

[147] A. Ratha, A. Schwele, J. Kazempour, P. Pinson, S. S. Torbaghan, and A. Virag, "Affine policies for flexibility provision by natural gas networks to power systems," *Electric Power Systems Research*, vol. 189, p. 106565, 2020.

[148] C. Ordoudis, V. A. Nguyen, D. Kuhn, and P. Pinson, "Energy and reserve dispatch with distributionally robust joint chance constraints," *Operations Research Letters*, vol. 49, no. 3, pp. 291–299, 2021.

[149] ELIA, "Personal correspondence," 2021.

[150] ENTSOG, "Entsog transparency platform," 2021. Online: https://transparency.entsog.eu/.

[151] C. He, T. Liu, L. Wu, and M. Shahidehpour, "Robust coordination of interdependent electricity and natural gas systems in day-ahead scheduling for facilitating volatile renewable generations via power-to-gas technology," *Journal of Modern Power Systems and Clean Energy*, vol. 5, no. 3, pp. 375–388, 2017.

[152] H. Ameli, M. Qadrdan, and G. Strbac, "Coordinated operation of gas and electricity systems for flexibility study," *Frontiers in Energy Research*, vol. 8, p. 120, 2020.

[153] Y. He, M. Shahidehpour, Z. Li, C. Guo, and B. Zhu, "Robust constrained operation of integrated electricity-natural gas system considering distributed natural gas storage," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1061–1071, 2018.

[154] Y. Ng, S. Misra, L. A. Roald, and S. Backhaus, "Statistical learning for dc optimal power flow," in *2018 Power Systems Computation Conference (PSCC)*, pp. 1–7, 2018.

[155] D. C. Costa, M. V. Nunes, J. P. Vieira, and U. H. Bezerra, "Decision tree-based security dispatch application in integrated electric power and natural-gas networks," *Electric Power Systems Research*, vol. 141, pp. 442–449, 2016.

[156] J. L. Cremer, I. Konstantelos, S. H. Tindemans, and G. Strbac, "Data-driven power system operation: Exploring the balance between cost and risk," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 791–801, 2019.

[157] I. Genc, R. Diao, V. Vittal, S. Kolluri, and S. Mandal, "Decision tree-based preventive and corrective control applications for dynamic security enhancement in power systems," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1611–1619, 2010.

[158] K. Dvijotham and D. K. Molzahn, "Error bounds on the dc power flow approximation: A convex relaxation approach," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 2411–2418, 2016.

[159] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[160] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, vol. 15, pp. 315–323, 2011.

[161] Y. LeCun, B. Yoshua, and G. Hinton, "Deep learning," *Nature*, vol. 521, p. 436–444, 2015.

[162] A. Schwele, C. Ordoudis, J. Kazempour, and P. Pinson, "Coordination of power and natural gas systems: Convexification approaches for linepack modeling," in *2019 IEEE Milan PowerTech*, pp. 1–6, 2019.