

Distributionally robust portfolio management for hybrid power plants



Anton Ruby Larsen & Mads Esben Hansen

DTU Wind-M-0617

June 2023

Author: Anton Ruby Larsen & Mads Esben Hansen

DTU Wind-M-0617
June 2023

Title:
Distributionally robust portfolio management for hybrid power plants

Project period:
January - June 2023

ECTS: 30

Education: Master of Science

Supervisor(s):

Jalal Kazempour
Andrea Kristine Gloppen Johnsen
Yannick Marcus Werner
DTU Wind and Energy Systems

Remarks:

This report is submitted as partial fulfillment of the requirements for graduation in the above education at the Technical University of Denmark.

DTU Wind and Energy Systems is a department of the Technical University of Denmark with a unique integration of research, education, innovation and public/private sector consulting in the field of wind and energy. Our activities develop new opportunities and technology for the global and Danish exploitation of wind and energy. Research focuses on key technical-scientific fields, which are central for the development, innovation and use of wind energy and provides the basis for advanced education.

Technical University of Denmark
Department of Wind and Energy Systems
Frederiksborgvej 399
DK-4000 Roskilde
www.wind.dtu.dk

Abstract

This master's thesis explores the performance and applicability of Wasserstein distributionally robust optimization (DRO) models in the context of wind/hydrogen power plants. The research aims to address four main research questions: (1) When Wasserstein DRO is needed, (2) the performance of Wasserstein DRO models, (3) the benefits of Wasserstein distributional robustness, and (4) the effect of integrating electrolysis with wind power generation.

To investigate these questions, the study formulates and compares different optimization models, including a deterministic model, a sample average approximation (SAA) model, and various DRO models. Extensive analysis and experimentation are conducted to evaluate their performance in terms of returns and volatility.

The findings reveal that Wasserstein DRO models consistently outperform both the deterministic model and the SAA model, showcasing their effectiveness in managing uncertainty. Furthermore, the study demonstrates the benefits of Wasserstein distributional robustness by allowing explicit adjustment of risk levels through the radius of the Wasserstein ball. A control algorithm is developed to dynamically adapt the risk level based on changing market conditions.

In addition, the integration of electrolysis with wind power generation is shown to reduce volatility in returns. Access to electrolysis significantly reduces volatility, irrespective of the presence of daily contracts. Moreover, the analysis highlights that electrolysis can lead to higher returns when hydrogen prices exceed specific thresholds, depending on the contractual framework.

Acknowledgements

We thank Siemens Gamesa for sharing real data as this thesis has been completed under EUDP-funded HOMEY project (64021-7010).

We want to thank our supervisors, Andrea K. G. Johnsen and Yannick M. Werner, for their guidance and expertise throughout the research process. Their input and feedback have contributed significantly to the quality of this thesis. We also acknowledge the contributions of Jalal Kazempour who also provided valuable insights and feedback during this thesis.

Contents

A Introduction	1
A.1 Research Questions	1
A.2 Literature Review	1
A.2.1 Mathematical Modeling Under Uncertainty	2
A.2.2 Summary	3
A.3 Contributions	3
A.4 Readers Guide	4
B Preliminaries	6
B.1 Power Markets	6
B.2 Wind Power Forecasting	7
B.3 A Wind/Hydrogen Model	8
C Models for Wind/Hydrogen Management Under Uncertainty	13
C.1 Optimization Under Uncertainty: A Brief Introduction	13
C.2 A Wind/Hydrogen Model with Uncertainty	15
C.3 Sample Average Approximation	16
C.3.1 A SAA Wind/Hydrogen Model	17
C.4 Distributionally Robust Optimization	18
C.4.1 Selection of Ambiguity Set	18
C.4.2 A Wasserstein DRO Wind/Hydrogen Model	19
C.4.3 A Wasserstein DRO with Correlation Wind/Hydrogen Model	22
C.5 Multi Stage Programs	25
D Evaluation of Energy Trading Models	26
D.1 Data	26
D.2 Testing	28
D.2.1 Performance Metrics	28
D.2.2 Realization	29
D.3 Parameter Fitting	30
E Results for the Wind/Hydrogen Models	31
E.1 Sample Selection	31
E.2 Results	32
E.2.1 The Deterministic Wind/Hydrogen Model	32
E.2.2 The SAA Wind/Hydrogen Model	33
E.2.3 The Wasserstein DRO Wind/Hydrogen Model	34
E.2.4 The Wasserstein DRO with Correlation Wind/Hydrogen Model	35
E.3 An Adaptive Wasserstein Radius for the Price	35

E.4	Comparison of Models	37
F	Sensitivity Analysis	41
F.1	Importance of Sample Selection	41
F.2	Importance of the Hydrogen	44
G	Discussion	49
G.1	What have we seen?	49
G.2	Future Research	50
G.2.1	Error Selection/Estimation	50
G.2.2	Control of the Risk-Awareness	50
G.2.3	The 2-Stage Model	51
G.2.4	End-to-End Optimization	52
H	Conclusion	53
List of Figures		54
List of Tables		56
A	Appendices	63
A	Relevant Extra Material	63
B	Hypograph Formulation of Concave Function	63
C	Chance Constraints, VaR and CVaR	64
D	Wasserstein Metric for Two Discrete Measures	65
E	Objective Function for the Wasserstein DRO Wind/Hydrogen Model	67
F	Dual of a Linear Program	70
G	Objective Function for the Wasserstein DRO with Correlation Wind/Hydrogen Model	71
H	K-Medoids	73
I	Pseudo Code for the Sample Selection Clustering	74
J	A PI Controller	75
K	2-Stage Models	76
K.1	2-stage Program Theory	76
K.2	Sample Selection	81
K.3	Results	82
K.4	Comparison of Models	83
K.5	Discussion	84
L	Training Results	84
L.1	Training Results for the Deterministic Wind/Hydrogen Model	85
L.2	Training Results for the SAA Wind/Hydrogen Model	85
L.3	Training Results for the Wasserstein DRO Wind/Hydrogen Model	86
L.4	Training Results for the Wasserstein DRO with Correlation Wind/Hydrogen Model	88
M	Passive Realization for Varying Hydrogen Prices	89

A | Introduction

In recent years, renewable energy has become an increasingly important source of power, with wind power, in particular, showing significant growth. However, the intermittency of wind power makes it difficult to rely on as a consistent source of energy. One solution is to use electrolyzers to convert excess electricity generated by wind turbines into hydrogen, which can then be stored and used as needed. This process, known as power-to-hydrogen, has the potential to be an important part of the transition to a more sustainable energy system.

A.1 Research Questions

In this thesis, we will focus our efforts on the problem of optimal operation of power plants with access to wind and electrolysis which we will refer to as wind/hydrogen power plants. The effect of uncertainty is important to consider in the operation of such renewable energy sources. We therefore aim at modeling a wind/hydrogen system under uncertainty. Particularly, we will try to answer the following research questions:

1. When are Wasserstein distributionally robust optimization models advantageous compared to deterministic and sample average approximation models in the context of wind/hydrogen power plants?
2. How do Wasserstein distributionally robust optimization models perform in the context of wind/hydrogen power plants?
3. What is the benefit of Wasserstein distributional robustness in the context of wind/hydrogen power plants compared to sample average approximation?
4. How does electrolysis at a wind power plant affect the daily returns?

To set the scene, we will go through recent developments in the field.

A.2 Literature Review

In the literature, we have primarily found two ways to tackle the problem of optimal operation of energy systems with access to renewable energy sources, and power-to-X. One way to solve this problem is by the use of machine learning. Zhu et al., 2022 use a deep reinforcement learning model which gradually learns how to operate a microgrid with renewable energy sources and the possibility to store energy as hydrogen. Yang et al., 2020 also use deep reinforcement learning to control a wind farm with access to energy storage. Another approach to address this problem is mathematical optimization. Within this field, two main directions can be distinguished. The first direction involves the use of

metaheuristics, which approximate the solution to the optimization problem using heuristic algorithms. Fathima and Palanisamy, 2015 outline various metaheuristics for optimizing hybrid energy systems. We will not delve deeper into metaheuristics here but instead focus on the second direction: Convex Optimization.

Convex optimization provides guarantees of optimality and offers bounds when the solution cannot be found within a specified time limit. A recent contribution by Baumhof et al., 2023 presents a convex program¹ for operating a hybrid power plant that combines wind turbines and electrolysis; however, this model does not consider uncertainty. In electricity markets, relying solely on expected values may result in suboptimal solutions, as the optimization problem may fail to account for potential asymmetric risks as discussed by Alvarez et al., 2007. Therefore, we will explore how the literature has addressed this aspect.

A.2.1 Mathematical Modeling Under Uncertainty

The process of optimizing power systems using mathematical modeling that incorporates uncertainty is described in Roald et al., 2023. The article outlines three general approaches to representing uncertainty in mathematical models:

1. Stochastic Optimization
2. Robust Optimization
3. Distributionally Robust Optimization

We will in the following subsections explore how these methods have been applied to solve the problem of optimal operation of power-to-X systems.

Stochastic Optimization

Eghbali et al., 2022 and Dong et al., 2023 use sample average approximation (SAA) to include uncertainty in a mixed integer linear program to address the optimal operation of a microgrid with access to renewable energy sources and hydrogen storage. SAA is a method within the field of stochastic optimization (SO) where the uncertainty is included as samples of the random variable(s), and is described in section C.3. When dealing with electricity markets, all uncertainty is rarely realized at the same time. We often have structures like a day-ahead market and an imbalance market where the day-ahead price is revealed before the imbalance price. To represent the time dependence, multi-stage programs have been developed, see, e.g., Birge and Louveaux, 2011. Daneshvar et al., 2020 use a two-stage SAA program to solve an optimal scheduling problem with thermal, wind and solar as energy sources, and a hydropower-pumped storage system. They model the day-ahead market as the first stage while the second stage is the imbalance market.

Robust Optimization

The second approach is robust optimization (RO) where the uncertainty is assumed to take the worst realization within a specified set. This allows the mathematical program to take unobserved worst-case realizations of the random variable(s) into account. Y. Wang et al., 2020 propose a RO model to design an off-grid solar-powered charging station that should provide fuel to both electric vehicles and hydrogen vehicles. The station has access to an electrolyzer and needs to decide how much power to convert to hydrogen. Within

¹Programs and models will be used interchangeably throughout the thesis.

the field of RO we also have the concept of multi-stage programs which is often called adaptive or adjustable robust optimization (ARO) programs. Pan et al., 2023 propose an ARO scheduling model to manage a hybrid energy generation system to maximize total profits in the day-ahead electricity and hydrogen market. In the system, they have access to solar power, hydro storage, and an electrolyzer.

Distributionally Robust Optimization

The third approach of incorporating uncertainty is distributionally robust optimization (DRO). In this approach, we consider a collection of distributions and select the one that yields the least favorable outcome for a stochastic optimization process. There are two different schools within DRO: Moment-based and metric-based. These schools differ in their definition of the distribution set, also called the ambiguity set. We will not consider the moment-based ambiguity sets but only focus on the metric-based ambiguity sets. Zheng et al., 2022 use Wasserstein DRO for optimal scheduling of power to methanol. They do not consider the day-ahead and imbalance market but only optimize the cost of scheduling the power to the methanol plant. Zheng et al., 2023 use Wasserstein DRO to find the optimal day-ahead strategy for a wind/hydrogen power plant. They do not consider uncertainty in the imbalance price nor consider the correlation between the day-ahead and the imbalance price.

Lastly, as for stochastic and robust optimization, we also have multi-stage distributionally robust programs. A study where such an approach is used is the study by Kong et al., 2022. They use a two-stage DRO model to optimize the operation of a water/hydrogen complementary system.

A.2.2 Summary

To our best knowledge Table A.1 lists the papers in the literature which are most similar to the work we have done. The table lists similarities and how our work differs.

Table A.1 – All the papers use Wasserstein distributionally robust optimization to model Wind/P2X power plants.

Literature	Trades Power	Wind/P2X	Balance price	Price dependence	Adaptive Radius
Zheng et al., 2022		x			
Zheng et al., 2023	x	x			
Our work	x	x	x	x	x

A.3 Contributions

In this thesis, we aim to advance the state-of-the-art in the field of optimal operation of wind/hydrogen power plants by addressing the aforementioned research questions. To this end, we make the following *application-oriented contributions*:

1. Our study demonstrates that distributionally robust optimization (DRO) models, specifically Wasserstein DRO with and without a correlation structure, outperform a deterministic model in the context of wind/hydrogen power plants in terms of returns and volatility.
2. Our study shows the benefit of explicitly choosing the risk level, made possible by Wasserstein DRO. Specifically, we propose an algorithm for automatically determin-

ing the risk aversion level in the DRO models, which allows for dynamically adapting the risk to the underlying market.

3. Our analysis revealed that integrating electrolysis with wind power generation reduced the volatility in daily returns. Furthermore, our investigation into returns demonstrated that electrolysis has the potential to increase returns, dependent on the hydrogen price and contractual structures.

We have also contributed with one *theoretical contribution*:

1. We present the underlying theory and principles of distributionally robust optimization. In our implementation of Wasserstein DRO with correlation, we introduce a "trust factor" that quantifies confidence in the correlation structure. This contribution enhances Wasserstein DRO when incorporating correlation information.

Overall, our contributions extend the existing literature on wind/hydrogen power plants by providing new insights into these facilities' optimal operation, trading strategies, and the potential benefits of using DRO techniques.

A.4 Readers Guide

The thesis is structured as follows:

Chapter A - Introduction, contains a general introduction to this thesis. It includes our research questions, the current state of the field, and the contributions we bring.

Chapter B - Preliminaries, gives a brief introduction to the power market, with a focus on the details needed to understand our work. There is a short discussion of how to produce wind power forecasts. Lastly, a wind/hydrogen mathematical model is presented.

Chapter C - Models for Wind/Hydrogen Management Under Uncertainty, introduces decision-making under uncertainty. The mathematical model from chapter B is presented with uncertainty. Finally, relevant theory and reformulation of the wind/hydrogen model for both SAA and DRO are presented.

Chapter D - Evaluation of Energy Trading Models, presents the data used to evaluate the performance of the uncertainty models. It is further presented how and on what metrics the models are evaluated. Finally, there is a brief remark on parameter fitting.

Chapter E - Results for the Wind/Hydrogen Models, presents the performance of the models along with how historical samples are selected. An algorithm for adaptively choosing the risk level is presented, and all models are compared.

Chapter F - Sensitivity Analysis, goes over the importance of sample selection in SAA and DRO. Further, the effect of the hydrogen price is analyzed thoroughly.

Chapter G - Discussion, presents a discussion of the thesis. Specifically, we discuss the results and what challenges remain.

Chapter H - Conclusion, contains the final conclusion of this thesis and wraps up the work done.

The code for the models used in this thesis is available at the following github. For a guide on how to use the code, we refer to the github's ReadMe file.

In the process of writing this master thesis, we have found various high-quality blog posts and YouTube videos. We thought it would be a shame not to share them, so in Appendix A, we have given a list with descriptions and links.

B | Preliminaries

B.1 Power Markets

In 2003, Denmark's energy market underwent liberalization, enabling consumers to purchase electricity on the open market. Today, electricity in Denmark is traded on an hourly basis, with each hour having its own individual electricity price. The market operates in three main segments: The day-ahead market, the intra-day market, and ancillary services. The day-ahead market is where electricity for the entire day (24 hours) is traded. Bids for supply and demand are submitted before noon on the day before (D-1). At 1:00 pm, the exchange matches these bids and reveals the prices for all hours of the following day as discussed in Danish-Energy-Agency, 2020. The intra-day market, known as XBID, facilitates trading between European countries and aims to minimize discrepancies between actual demand and supply, see, e.g., Nordpool, 2018.

To maintain balance in the electricity grid, ancillary services are utilized when significant differences arise between supply and demand. According to Energinet, 2022, these services are traded to assist the transmission system operator (TSO). The resulting imbalance price may differ from day-ahead and intra-day market prices, depending on the extent of up- or down-regulation.

Denmark's wind turbines covered 54% of electricity consumption in 2022 according to Energinet, 2023, reflecting a positive shift toward renewable energy sources. According to Soini, 2021, this poses challenges, as pricing in the imbalance market tends to be higher during periods of lower-than-expected wind power. Energy storage solutions are necessary to bridge the gap between low-cost and high-cost periods. Hydrogen production in European wind farms has been identified as a profitable method for balancing the electricity grid by Ortiz and Nørstebø, 2021.

Power-to-X solutions, including electrolysis, have gained interest in addressing the complexities of the power grid. Electrolysis is a process that uses electricity to split water into hydrogen and oxygen gas. The electrical energy required for electrolysis can come from renewable sources such as solar or wind power. Alkaline electrolyzers (AEC) are considered suitable for power-to-X systems due to their maturity, efficiency, flexibility, and cost-effectiveness. They enable the utilization of excess power during periods of low demand or low prices and can participate in the imbalance market by offering ancillary services, as shown by Danish-Energy-Agency and Energinet, 2017.

Figure B.1 shows the efficiency curve for an 52.25MW alkaline electrolyzer working at 90°C and 30 bar and its linear approximation. The figure is taken from Baumhof et al., 2023, and we base our modeling of the electrolyzer on this work, hence we will use the same curve and linear approximation technique.

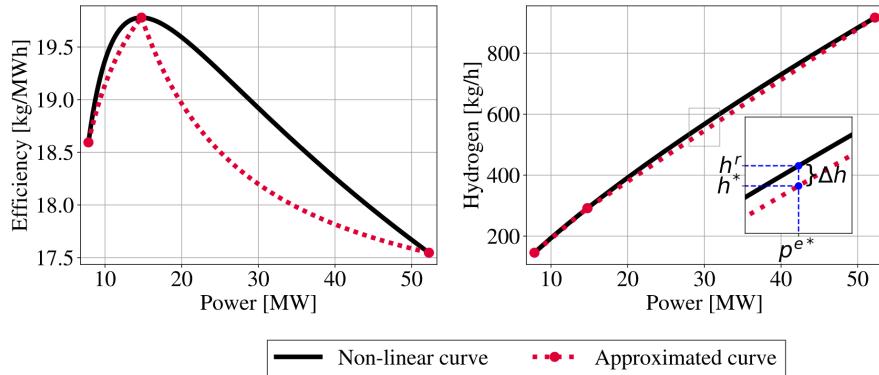


Figure B.1 – Efficiency curve for a 52.25MW alkaline electrolyzer working at 90°C and 30 bar and linear approximation thereof. The figure is taken from Baumhof et al., 2023.

Recent advancements in electrolysis technology have shown promise in improving efficiency and reducing costs, making power-to-X systems more economically viable. For example, a recent study published in Nature, Guo et al., 2023, demonstrated nearly 100% Faradic efficiency¹ in the electrolysis of seawater that has not been alkalinized nor acidified. This breakthrough could potentially revolutionize the field of hydrogen production from seawater, as it eliminates the need for costly desalination processes previously necessary to produce high-purity water for electrolysis.

B.2 Wind Power Forecasting

Wind power forecasting is essential for efficient power system operations. Giebel and Kariniotakis, 2017, Lange and Focken, 2006, and Pinson and Tastu, 2014 have discussed wind power forecasting, highlighting the importance of a physical model that predicts wind power based on weather predictions.

Pinson and Tastu, 2014 suggested predicting the output from a wind turbine based on Numerical Weather Prediction (NWP) by modeling it as a physical system. The total amount of energy that flows through a wind turbine can be expressed using physics. However, the turbine cannot extract all of the energy from the air due to several factors, including Betz's law², the turbine's maximum capacity, and cut-in and cut-out wind speeds. It is important to fit a statistical model and not just use the power curve from the manufacturer because of two things. First, the noise in the NWP has an asymmetric effect on the power output, making the power curve from the manufacturer biased, especially as the noise in NWP increases with the forecast horizon. Second, if the NWP is generally biased at the location of the wind turbine, the forecast using the power curve from the manufacturer will also be biased. Pinson and Tastu, 2014 discussed using non-parametric regression to estimate the power curve, which is often advantageous compared to a linear model with input given by the physical relation. Xu et al., 2016 further discussed using local polynomial regression (LPR) to estimate the power curve.

Based on this, we limit ourselves to using LPR, as described in Madsen and Holst, 2006 and Nielsen et al., 2001. The LPR estimates are given by

¹ "Faradaic efficiency (FE) describes the overall selectivity of an electrochemical process and is defined as the amount (moles) of collected product relative to the amount that could be produced from the total charge passed, expressed as a fraction or a percent." (Kempler and Nielander, 2023, page 1)

²See Wikipedia.

$$Y_t = P(X_t, \theta(X_t)) + \varepsilon_t, \quad (\text{B.2.1})$$

where P is a polynomial of arbitrary degree, ε_t follows a normal distribution, and θ are the locally estimated parameters that depend on the value of X_t . The parameters are found as

$$\hat{\theta}(X_t) = \arg \min_{\theta(X_t)} \frac{1}{N} \sum_{s=1}^N W(X_t, X_s)(Y_s - P(X_s, \theta(X_s)))^2, \quad (\text{B.2.2})$$

where the kernel, W , describes how the residuals are weighted, and N is the number of observations in the training data.

Figure B.2 shows a fitted power curve (Kernel regression) with the power curve from the manufacturer. Please notice the difference between the fitted curve and the manufacturer's power curve when the wind speed prediction is either high or low. As the forecast horizon increases, this difference between the fitted curve and the manufacturer's power curve is expected to increase due to an increase in the variance of the NWP. This particular power curve was fitted using a 0-degree polynomial, i.e., kernel smoothing and a Gaussian kernel.

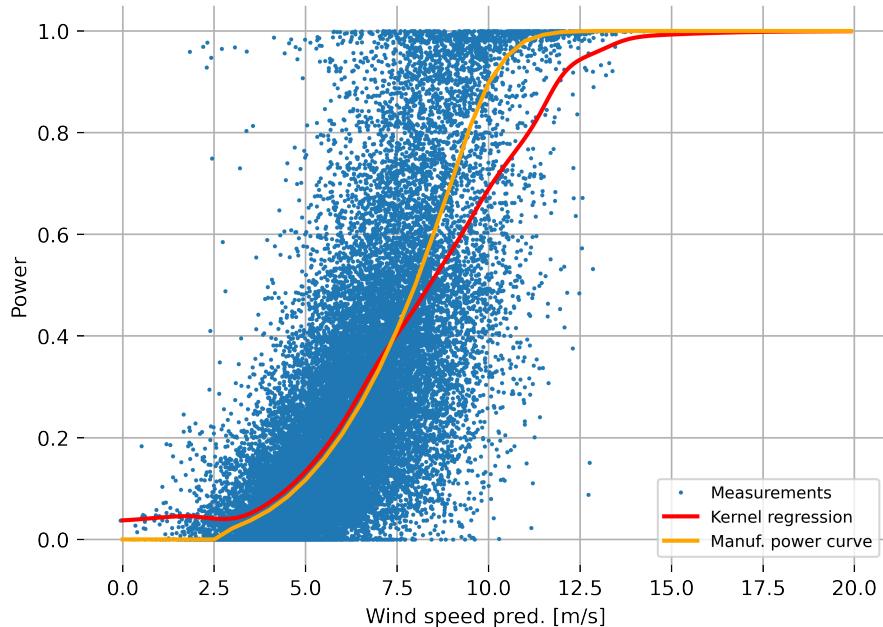


Figure B.2 – Fitted power curve and power curve from the manufacturer. Please notice that random noise has been added to measurements for confidentiality purposes.

B.3 A Wind/Hydrogen Model

This section presents the model of a wind/hydrogen power plant, which will serve as a baseline for testing and incorporating uncertainty. The model is obtained from a previous study, Baumhof et al., 2023, where they model the same wind/hydrogen power plant as the one we will focus on in this thesis. The wind/hydrogen power plant is shown in Figure B.3.

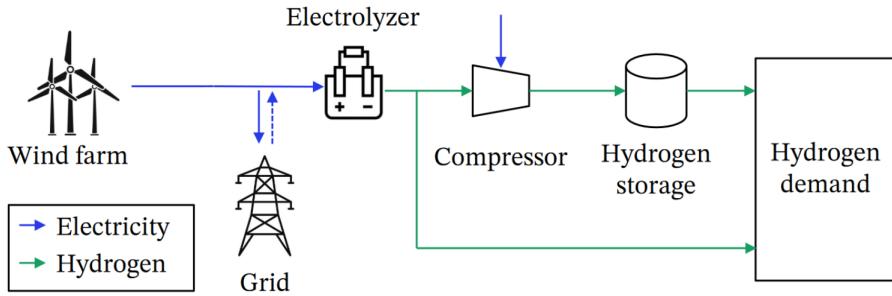


Figure B.3 – An overview of the wind/hydrogen power plant. The figure is taken from figure 2 in Baumhof et al., 2023

The model is a mixed integer linear program and is superficially described below. For a detailed explanation, we refer to the source paper.

1) *Objective function and power balance:* This part of our model differs from the original paper in that it includes the imbalance market. The objective function and power balances over the hours, $t \in \mathcal{T}$, are given as

$$\max_{\mathbf{x}} \quad \sum_{t \in \mathcal{T}} p_t^{\text{DA}} \lambda_t^{\text{DA}} + p_t^{\text{B}} \lambda_t^{\text{B}} + d_t \lambda^{\text{h}} - (p_t^{\text{DA}-} + p_t^{\text{B}-}) \lambda^{\text{TSO}} - z_t^{\text{su}} \lambda^{\text{su}} \quad (\text{B.3.1a})$$

$$\text{s.t. } p_t^w = p_t^{\text{DA}} + p_t^{\text{B}} + p_t^e + p_t^c \quad \forall t \in \mathcal{T} \quad (\text{B.3.1b})$$

$$p_t^{\text{DA}} = p_t^{\text{DA+}} - p_t^{\text{DA-}} \quad \forall t \in \mathcal{T} \quad (\text{B.3.1c})$$

$$p_t^{\text{B}} = p_t^{\text{B+}} - p_t^{\text{B-}} \quad \forall t \in \mathcal{T} \quad (\text{B.3.1d})$$

$$p_t^w \leq P_t^w \quad \forall t \in \mathcal{T} \quad (\text{B.3.1e})$$

where \mathcal{T} in our case is given by $\mathcal{T} = \{1, 2, \dots, 23, 24\}$, p_t^{DA} is power sold in day-ahead market, p_t^{B} is power sold in imbalance market, and λ_t^{DA} and λ_t^{B} are their respective prices. The placeholder for all decision variables in the model, denoted by \mathbf{x} , will be defined later in the section. The profits of the objective for each time period, t , consists of the electricity sold in the day-ahead and imbalance market, $p_t^{\text{DA}} \lambda_t^{\text{DA}} + p_t^{\text{B}} \lambda_t^{\text{B}}$, as well as the revenue generated from selling hydrogen, $d_t \lambda^{\text{h}}$. The costs are given by the electricity purchased from the grid, $(p_t^{\text{DA}-} + p_t^{\text{B}-}) \lambda^{\text{TSO}}$, and the cost of cold start-up for the electrolyzer, $z_t^{\text{su}} \lambda^{\text{su}}$. Equation B.3.1b enforces power balance, requiring that the power produced from the wind turbine(s), p_t^w , equals the grid activity, $p_t^{\text{DA}} + p_t^{\text{B}}$, plus the power consumed by the production and storage of hydrogen, $p_t^e + p_t^c$. Finally, Equation B.3.1c and Equation B.3.1d divide the day-ahead and imbalance market activities into purchase and sale components, while Equation B.3.1e permits curtailing the wind turbine(s) below the potential wind power production, P_t^w .

2) *Allowed speculation:* To maintain the primary focus on optimizing the operation of the hybrid power plant, we have implemented additional constraints in the model, specifically with regard to short selling in the day-ahead market. This part is new compared to the original paper, Baumhof et al., 2023. The reason for adding these constraints is that the original paper only considered the day-ahead market, and hence speculation or short selling was not possible within the model's framework.

The constraints are defined as follows:

$$p_t^{\text{DA}} \leq C^W \quad \forall t \in \mathcal{T} \quad (\text{B.3.2a})$$

$$p_t^{\text{DA}} \geq -p_t^w \quad \forall t \in \mathcal{T} \quad (\text{B.3.2b})$$

$$p_t^{\text{DA}} \geq -p^e - p_t^c \quad \forall t \in \mathcal{T} \quad (\text{B.3.2c})$$

Analyzing these constraints, we observe that the model permits long positions in the day-ahead market, with an upper limit defined by the maximum capacity of the wind power available at the site (denoted as C^W). This implies that the model can sell electricity up to the maximum capacity of the wind power source in the day-ahead market.

Regarding short selling in the day-ahead market, the model restricts the activity based on planned hydrogen production or the expected wind turbine output. This limitation ensures that the short positions are covered either by the expected wind power generation or by converting the purchased electricity to hydrogen. By implementing this approach, the model only engages in buying for amounts that can be offset by wind power or are projected to yield profits through electricity-to-hydrogen conversion.

3) *Electrolyzer operational states:* Next we constrain how the electrolyzer can operate.

$$1 \geq z_t^{\text{on}} + z_t^{\text{sb}} \quad \forall t \in \mathcal{T} \quad (\text{B.3.3a})$$

$$p_t^e \leq C^e z_t^{\text{on}} + P^{\text{sb}} z_t^{\text{sb}} \quad \forall t \in \mathcal{T} \quad (\text{B.3.3b})$$

$$p_t^e \geq P^{\min} z_t^{\text{on}} + P^{\text{sb}} z_t^{\text{sb}} \quad \forall t \in \mathcal{T} \quad (\text{B.3.3c})$$

$$z_t^{\text{su}} \geq z_t^{\text{on}} - z_{t-1}^{\text{on}} - z_{t-1}^{\text{sb}} \quad \forall t \in \mathcal{T} \setminus \{1\} \quad (\text{B.3.3d})$$

$$z_t^{\text{sb}} \leq z_{t-1}^{\text{on}} + z_{t-1}^{\text{sb}} \quad \forall t \in \mathcal{T} \setminus \{1\} \quad (\text{B.3.3e})$$

$$z_{t=1}^{\text{su}} = 0 \quad (\text{B.3.3f})$$

Equation B.3.3a ensures that the electrolyzer is only in one state during each time interval, t . The variable z_t^{sb} represents the standby state, z_t^{on} represents the online state where the electrolyzer produces hydrogen, and if it is neither on nor in standby, the electrolyzer is turned off. Equation B.3.3b and Equation B.3.3c guarantee that if the electrolyzer is in the on or standby state, the electricity consumption, p_t^e , remains within specific upper and lower bounds. Finally, Equation B.3.3d, Equation B.3.3e, and Equation B.3.3f specify the allowable transitions for the electrolyzer between its different states.

4) *Electrolyzer hydrogen production:* Next we describe how the electrolyzer can produce hydrogen.

$$h_t \leq A_s \hat{p}_t^e + B_s z_t^{\text{on}} \quad \forall t \in \mathcal{T}, \forall s \in \mathcal{S} \quad (\text{B.3.4a})$$

$$p_t^e = \hat{p}_t^e + P^{\text{sb}} z_t^{\text{sb}} \quad \forall t \in \mathcal{T} \quad (\text{B.3.4b})$$

Equation B.3.4a characterizes the hydrogen production, h_t , which relies on the electricity consumption of the electrolyzer. In this formulation, we approximate the relationship using a hypograph composed of linear functions with slopes, A_s , and intercepts, B_s . The rationale behind this approach is explained in detail in Appendix B. It is important to note that this constraint differs from the formulation used in the original work, Baumhof et al., 2023.

The motivation for employing this formulation lies in its scalability. Unlike the use of multiple binary variables as in the original paper, each line segment introduces only one new linear constraint. This scalability aspect is crucial and allows for improved computational

efficiency. For further insights into convex tricks and optimization techniques, refer to references such as Boyd and Vandenberghe, 2004.

Furthermore, Equation B.3.4b computes the total electricity consumption of the electrolyzer, considering its various states. Here, \hat{p}_t^e represents the power consumption, and $P^{sb}z_t^{sb}$ represents the electricity consumption during standby mode.

5) *Hydrogen storage:* We now define how to store hydrogen.

$$h_t = h_t^d + s_t^{\text{in}} \quad \forall t \in \mathcal{T} \quad (\text{B.3.5a})$$

$$d_t = h_t^d + s_t^{\text{out}} \quad \forall t \in \mathcal{T} \quad (\text{B.3.5b})$$

$$s_t^{\text{out}} \leq S^{\text{out}} \quad \forall t \in \mathcal{T} \quad (\text{B.3.5c})$$

$$p_t^c = K^c s_t^{\text{in}} \quad \forall t \in \mathcal{T} \quad (\text{B.3.5d})$$

$$s_{t=1} = S^{\text{ini}} + s_{t=1}^{\text{in}} - s_{t=1}^{\text{out}} \quad (\text{B.3.5e})$$

$$s_t = s_{t-1} + s_t^{\text{in}} - s_t^{\text{out}} \quad \forall t \in \mathcal{T} \setminus \{1\} \quad (\text{B.3.5f})$$

$$s_t \leq C^s \quad \forall t \in \mathcal{T} \quad (\text{B.3.5g})$$

The electrolyzer's hydrogen production, h_t , can either be delivered directly to meet the demand, h_t^d , or stored in the hydrogen storage, s_t^{in} , as per Equation B.3.5a. The hydrogen sold at time t , d_t , is the sum of the hydrogen produced for the market, h_t^d , and the hydrogen released from the storage, s_t^{out} , as enforced by Equation B.3.5b. The maximum hydrogen output from the storage per hour is constrained by the output flow capacity, S^{out} , in Equation B.3.5c. The hydrogen compressor consumes power, p^c , to compress the hydrogen before storage, using the compression coefficient, K^c . The compressor's power consumption is constrained by Equation B.3.5d. The storage in the initial and subsequent hours is handled by Equation B.3.5e and Equation B.3.5f, with the initial storage, S^{ini} , at the start of the time horizon \mathcal{T} . The storage's maximum hydrogen capacity is constrained by Equation B.3.5g.

6) *Hydrogen demand and limit on p_t^{in} :* Lastly, the hydrogen demand is defined such that at certain specified times, $t \in \mathcal{H}_n$, a minimum amount D_n^{\min} must be delivered (we use daily contracts).

$$\sum_{t \in \mathcal{H}_n} d_t \geq D_n^{\min} \quad \forall n \in 1, \dots, N \quad (\text{B.3.6})$$

Additionally, the paper, Baumhof et al., 2023, restricts electricity utilization from the grid. This limitation allows grid electricity usage solely for maintaining the electrolyzer in standby mode. However, we decided to remove this constraint in our model. The rationale behind this modification is that incorporating the original limitation rendered subsequent models infeasible. Therefore, by removing this constraint, we ensure the feasibility of our models while still effectively addressing the hydrogen demand.

7) *Variable declaration:* The model introduced consists of various variables with specific characteristics and limitations. The continuous variables, which are confined to the non-negative orthant, include:

$$d_t, h_t, h_t^d, p_t^e, p_t^c, \hat{p}_t^e, s_t, s_t^{\text{in}}, s_t^{\text{out}}, p_t^{\text{DA+}}, p_t^{\text{DA-}}, p_t^{\text{B+}}, p_t^{\text{B-}}, p_t^w, z_t^{\text{su}} \in \mathbb{R}^+$$

Additionally, there are two continuous variables without restrictions:

$$p_t^{\text{DA}}, p_t^{\text{B}} \in \mathbb{R}.$$

The model also incorporates two binary variables:

$$z_t^{\text{on}}, z_t^{\text{sb}} \in \{0, 1\}.$$

Collectively, these variables form the set \mathbf{x} :

$$\mathbf{x} = \left\{ d_t, h_t, h_t^d, p_t^c, p_t^e, p_t^w, p_t^{\text{DA+}}, p_t^{\text{DA-}}, p_t^{\text{B+}}, p_t^{\text{B-}}, p_t^{\text{DA}}, p_t^{\text{B}}, \hat{p}_t, \right. \\ \left. s_t^{\text{in}}, s_t, s_t^{\text{out}}, z_t^{\text{su}}, z_t^{\text{on}}, z_t^{\text{sb}} \right\}. \quad (\text{B.3.7})$$

This set comprises $17|\mathcal{T}|$ continuous variables and $2|\mathcal{T}|$ binary variables, where the number of time steps, denoted by $|\mathcal{T}|$, may vary depending on the specific model under consideration. In subsequent models, to simplify notation, we will collectively refer to all the constraints mentioned in this section as \mathcal{X} without explicitly repeating them.

C | Models for Wind/Hydrogen Management Under Uncertainty

The model for optimal operation of a wind/hydrogen power plant introduced in the previous chapter does not consider uncertainty. This gap highlights the need to investigate how uncertainty can be introduced into the model. We can capture the inherent risks and fluctuations associated with wind power and market prices by incorporating uncertainty. In this chapter, we will explore methodologies proposed in the literature to address optimization under uncertainty.

C.1 Optimization Under Uncertainty: A Brief Introduction

For any optimization problem, the objective can be given on the form

$$\min_{x \in \mathcal{X}} f(x, \xi) \quad (\text{C.1.1})$$

where x is the decision variable (chosen by the engineer) and $\xi \in \Xi$ is uncontrollable randomness. Notice that for some problems, the uncontrollable element is non-existent while it plays an important role for others.

There are many ways to deal with the uncertainty in optimization problems. A simple approach – and widely used – is to *forget* about the uncertainty. This is done by only considering the expectation of the uncertainty and treating the problem as deterministic, i.e., without any influence of uncertainty. We call this approach the *deterministic* approach. Mathematically, this corresponds to

$$\min_{x \in \mathcal{X}} f(x, \mathbb{E}^{\mathbb{P}}[\xi]) \quad (\text{C.1.2})$$

where $\mathbb{P} : \Xi \rightarrow \mathbb{R}^+$ is some predetermined probability measure.

An alternative approach is to look at the expectation of the objective function. This method is possibly the most commonly used method for dealing with uncertainties and is taught in intro-level courses for decision-making under uncertainty. This method is called *stochastic optimization* (SO); for further information, see, e.g., Birge and Louveaux, 2011. The stochastic optimization problem is given by

$$\min_{x \in \mathcal{X}} \mathbb{E}^{\mathbb{P}}[f(x, \xi)]. \quad (\text{C.1.3})$$

Notice that the expectation is of the entire objective function this time, forcing the engineer to consider multiple realizations of the uncertainty. In practice, SO is often solved by

selecting a finite number of samples or scenarios for the uncertainty and choosing the probability measure, \mathbb{P} , accordingly.

Another approach is to look at the worst-case scenarios for the uncertainty. This approach is popularized under the name *robust optimization* (RO), see, e.g., Ben-Tal et al., 2009, and Bertsimas et al., 2011. The RO objective function is given by

$$\min_{x \in \mathcal{X}} \max_{z \in \mathcal{Z}} f(x, z) \quad (\text{C.1.4})$$

where $\mathcal{Z} \subseteq \Xi$, is called the *uncertainty set*. Notice that for uncertainties with bounded support, it is possible to use the support set as the uncertainty set, while for uncertainties without a bounded support, e.g., normal distributed variables, the engineer has to decide on some proper subset.

There are many advantages to using RO. Generally, it is computationally easy to solve and includes information about the uncertainty that is not included in the deterministic method. Additionally, the method does not require any specific details on the probability measure, only its support. However, it can be difficult to select suitable uncertainty sets, which often leads to overly pessimistic results.

In later years, there has been a growing interest in combining SO and RO in what is known as *distributionally robust optimization* (DRO), see, e.g., Delage and Ye, 2010b, Esfahani and Kuhn, 2015, and Gao and Kleywegt, 2016. In short, the idea is to perform SO while being *robust* in the choice of probability measure, i.e., the selection of measure is not made beforehand. Specifically, we write this as

$$\min_{x \in \mathcal{X}} \max_{\mathbb{P} \in \mathcal{A}} \mathbb{E}^{\mathbb{P}}[f(x, \xi)] \quad (\text{C.1.5})$$

where the *ambiguity set*, \mathcal{A} , is a set containing all possible measures for our uncertain variable, ξ . The question of how to define this ambiguity set will be addressed in detail later in this chapter.

Many optimization problems do not only contain uncertainty in the objective function; some also contain some uncertainty in the constraints. Without loss of generality, we can write this as

$$\min_{x \in \mathcal{X}} f(x, \xi) \quad (\text{C.1.6a})$$

$$\text{s.t. } g(x, \xi) \leq 0 \quad (\text{C.1.6b})$$

where $g : \mathcal{X} \times \Xi$ defines the constraints with uncertainty. As with the objective function, the uncertainty in the constraints can be handled in many ways. Analogous to before, it is possible to *forget* about the uncertainty, i.e.

$$\min_{x \in \mathcal{X}} f(x, \xi) \quad (\text{C.1.7a})$$

$$\text{s.t. } g(x, \mathbb{E}^{\mathbb{P}}[\xi]) \leq 0. \quad (\text{C.1.7b})$$

However, since this may lead to frequent infeasible solutions, this can come at a cost to the engineer upon realization. Therefore, a natural extension to this is to *robustify* the solution by ensuring feasibility for a large subset of all possible realization.

$$\min_{x \in \mathcal{X}} f(x, \xi) \quad (\text{C.1.8a})$$

$$\text{s.t. } g(x, z) \leq 0 \quad \forall z \in \mathcal{Z} \quad (\text{C.1.8b})$$

where $\mathcal{Z} \subseteq \Xi$ again defines the chosen uncertainty set, i.e., the set for which all solutions must be feasible.

A more probabilistic approach ensures the problem is feasible with probability $1 - \epsilon$, where $\epsilon \in [0, 1]$ (generally small). This is handled by a *chance constraint* (CC). In fact, Rockafellar, Uryasev, et al., 2000 showed that Value-at-Risk (VaR) and chance constraints are equivalent problems. We write this as

$$\min_{x \in \mathcal{X}} f(x, \xi) \quad (\text{C.1.9a})$$

$$\text{s.t. } \mathbb{P}[g(x, \xi) \leq 0] \geq 1 - \epsilon \quad (\text{C.1.9b})$$

where ϵ controls how often the problem is allowed to be infeasible.

C.2 A Wind/Hydrogen Model with Uncertainty

Section B.3 assumes that the values of wind power potential, P_t^w , day-ahead price, λ_t^{DA} , and imbalance price, λ_t^{B} , are known. However, this assumption does not align with reality as these values are typically derived from prediction models with inherent uncertainty. Therefore, it is necessary to reformulate the relevant parts of the model presented in section B.3 to incorporate the uncertainty associated with P_t^w , λ_t^{DA} , and λ_t^{B} :

$$\max_{\mathbf{x}} \sum_{t \in \mathcal{T}} \left(p_t^{\text{DA}}(\hat{\lambda}_t^{\text{DA}} + \xi_t^{\text{DA}}) + p_t^{\text{B}}(\hat{\lambda}_t^{\text{B}} + \xi_t^{\text{B}}) \dots \right) \quad (\text{C.2.1a})$$

$$\text{s.t. } p_t^w \leq \hat{P}_t^w + \xi_t^w \quad \forall t \in \mathcal{T} \quad (\text{C.2.1b})$$

$$\mathbf{x} \in \mathcal{X} \quad (\text{C.2.1c})$$

where \mathcal{X} represents all the remaining constraints from the model presented in section B.3, and ξ_t^w , ξ_t^{DA} , and ξ_t^{B} are random variables representing the prediction errors. We observe that the uncertainty in the objective function can be addressed through an expectation operator using some probability measure, \mathbb{P} . Additionally, the uncertainty in Equation C.2.1b can be accommodated by a chance constraint using some other probability measure, \mathbb{P}' . This leads us to the following program:

$$\max_{\mathbf{x}} \sum_{t \in \mathcal{T}} \left(\mathbb{E}^{\mathbb{P}} [p_t^{\text{DA}} \xi_t^{\text{DA}} + p_t^{\text{B}} \xi_t^{\text{B}}] + p_t^{\text{DA}} \hat{\lambda}_t^{\text{DA}} \dots \right) \quad (\text{C.2.2a})$$

$$\text{s.t. } \mathbb{P}' [p_t^w \leq \hat{P}_t^w + \xi_t^w] \geq 1 - \epsilon \quad \forall t \in \mathcal{T} \quad (\text{C.2.2b})$$

$$\mathbf{x} \in \mathcal{X}. \quad (\text{C.2.2c})$$

The model now includes uncertainties, but how to solve the model is not completely clear. The following sections will present different approaches on how to reformulate the model into a solvable form.

C.3 Sample Average Approximation

The first approach is a stochastic optimization method known as sample average approximation (SAA). In SAA, the empirical measure obtained from a set \mathcal{N} of $|\mathcal{N}|$ realizations of the random variables is chosen as the probability measure, represented as

$$\mathbb{P}_{\mathcal{N}}(A) = \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} I_A(X_i) = \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} \delta_{X_i}(A),$$

where A is an event of interest. Hence it is quite obvious that the expectation operator in the objective function becomes the empirical mean

$$\begin{aligned} \sum_{t \in \mathcal{T}} \mathbb{E}^{\mathbb{P}_{\mathcal{N}}} [p_t^{\text{DA}} \xi_t^{\text{DA}} + p_t^{\text{B}} \xi_t^{\text{B}}] &= \\ \sum_{t \in \mathcal{T}} \frac{1}{|\mathcal{N}_t|} \sum_{i \in \mathcal{N}_t} p_t^{\text{DA}} \xi_{t,i}^{\text{DA}} + p_t^{\text{B}} \xi_{t,i}^{\text{B}}. \end{aligned} \quad (\text{C.3.1})$$

The reformulation is also presented in Birge and Louveaux, 2011. Next, we address the reformulation of the chance constraint:

$$\mathbb{P}_{\mathcal{N}} \left[p_t^{\text{w}} \leq \hat{P}_t^{\text{w}} + \xi_t^{\text{w}} \right] \geq 1 - \epsilon. \quad (\text{C.3.2})$$

To tackle this, Pagnoncelli et al., 2009 proposes the following reformulation of Equation C.3.2 using the empirical measure:

$$p_t^{\text{w}} \leq \hat{P}_t^{\text{w}} + \xi_{t,i}^{\text{w}} z_{t,i}^{\text{SAA}} \quad \forall i \in \mathcal{N}_t, \forall t \in \mathcal{T} \quad (\text{C.3.3a})$$

$$\frac{1}{|\mathcal{N}_t|} \sum_{i \in \mathcal{N}_t} z_{t,i}^{\text{SAA}} \geq 1 - \epsilon \quad \forall t \in \mathcal{T} \quad (\text{C.3.3b})$$

$$z_{t,i}^{\text{SAA}} \in \{0, 1\} \quad \forall i \in \mathcal{N}_t, \forall t \in \mathcal{T}. \quad (\text{C.3.3c})$$

However, we identify a potential computational issue with the current reformulation. Using $\sum_{t \in \mathcal{T}} |\mathcal{N}_t|$ binary variables may become computationally challenging when fitting parameters such as ϵ . Therefore, alternative forms of the chance constraint, Equation C.3.2, need to be explored. In Appendix C, we demonstrate that Equation C.3.2 can be reformulated as a VaR constraint, which can be conservatively approximated by a conditional value-at-risk (CVaR) constraint. We define the CVaR as presented in Equation 17.4 in Cornuejols and Tütüncü, 2006:

$$\text{CVaR}_{\epsilon}(x) := \frac{1}{\epsilon} \int f(x, \xi) \geq \text{VaR}_{\epsilon}(x) f(x, \xi) p(\xi) d\xi. \quad (\text{C.3.4})$$

The CVaR definition explicitly involves the VaR function, making it challenging to work with. However, Rockafellar, Uryasev, et al., 2000 introduces the function:

$$F_{\epsilon}(x, \tau) = \tau + \frac{1}{\epsilon} \int_{\xi \in \mathbb{R}^m} [f(x, \xi) - \tau]^+ p(\xi) d\xi, \quad (\text{C.3.5})$$

where $[.]^+ := \max(0, \cdot)$. In Theorem 2 of Rockafellar, Uryasev, et al., 2000, they show that:

$$\min_{x \in X} \text{CVaR}_\epsilon(x) = \min_{x \in X, \tau} F_\epsilon(x, \tau). \quad (\text{C.3.6})$$

Hence, we can conservatively approximate Equation C.3.2 as:

$$\mathbb{P}_{\mathcal{N}} \left[p_t^w \leq \hat{P}_t^w + \xi_t^w \right] \geq 1 - \epsilon \Leftrightarrow \quad (\text{C.3.7})$$

$$\text{CVaR}_\epsilon \left(p_t^w - \hat{P}_t^w - \xi_t^w \right) \leq 0 \Leftrightarrow \quad (\text{C.3.8})$$

$$\min_{\tau} \tau + \frac{1}{\epsilon} \mathbb{E}^{\mathbb{P}_{\mathcal{N}}} [(p_t^w - \hat{P}_t^w - \xi_t^w)^+] \leq 0 \Leftrightarrow \quad (\text{C.3.9})$$

$$\begin{aligned} \min_{\tau, y_i} & \tau + \frac{1}{\epsilon |\mathcal{N}_t|} \sum_{i \in \mathcal{N}_t} y_i \leq 0 & \forall t \in \mathcal{T} \\ \text{s.t.} & y_i \geq 0 & \forall i \in \mathcal{N}_t \\ & y_i \geq p_t^w - \hat{P}_t^w - \xi_t^w - \tau & \forall i \in \mathcal{N}_t \end{aligned} \quad (\text{C.3.10})$$

C.3.1 A SAA Wind/Hydrogen Model

To formulate the wind/hydrogen model as a SAA wind/hydrogen model we need to change it from a maximization to a minimization problem. We do this by multiplying the objective with -1 . We restate the reformulated model below:

$$\min_{\mathbf{x}, W_t, \tau_t, y_{t,i}} - \left(\sum_{t \in \mathcal{T}} W_t + \frac{1}{|\mathcal{N}_t|} \sum_{i \in \mathcal{N}_t} p_t^{\text{DA}} \xi_{t,i}^{\text{DA}} + p_t^{\text{B}} \xi_{t,i}^{\text{B}} \right) \quad (\text{C.3.11a})$$

$$\text{s.t. } W_t = \begin{pmatrix} p_t^{\text{DA}} \hat{\lambda}_t^{\text{DA}} + p_t^{\text{B}} \hat{\lambda}_t^{\text{B}} + d_t \lambda^{\text{h}} \dots \\ -(p_t^{\text{DA}-} + p_t^{\text{B}-}) \lambda^{\text{TSO}} - z_t^{\text{su}} \lambda^{\text{su}} \end{pmatrix} \quad \forall t \in \mathcal{T} \quad (\text{C.3.11b})$$

$$\tau_t + \frac{1}{\epsilon |\mathcal{N}_t|} \sum_{i \in \mathcal{N}_t} y_{t,i} \leq 0 \quad \forall t \in \mathcal{T} \quad (\text{C.3.11c})$$

$$y_{t,i} \geq 0 \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.3.11d})$$

$$y_{t,i} \geq p_t^w - \hat{P}_t^w - \xi_{t,i}^w - \tau_t \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.3.11e})$$

$$y_{t,i}, \tau_t \in \mathbb{R} \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.3.11f})$$

$$W_t \in \mathbb{R} \quad \forall t \in \mathcal{T} \quad (\text{C.3.11g})$$

$$\mathbf{x} \in \mathcal{X}. \quad (\text{C.3.11h})$$

Now we have a solvable form of Equation C.2.2, where the probability measure is the empirical measure. Hence, a natural question to ask is whether Equation C.3.11 converges to Equation C.2.2 as $N \rightarrow \infty$. From proposition 2.1 in Kleywegt et al., 2002, we know that the SAA reformulation will converge for mixed-integer programs. However, proposition 2.4 of the same paper states that the rate of convergence is influenced by the program's variance, making it difficult to estimate. So, if the SAA performs poorly when tested out-of-sample and we find ourselves unable to obtain additional data, we face a challenging situation. Unfortunately, we cannot take any immediate steps to enhance the performance of the SAA. However, an alternative method can address this limitation: Distributionally Robust Optimization.

C.4 Distributionally Robust Optimization

The underlying concern with SAA lies in the possibility that the observed data may not fully capture all relevant information, including potential risks that could negatively impact our decisions. The idea behind DRO is to address this concern by adjusting the probability measure to incorporate these potential risks. In Equation C.4.1 the DRO problem is given in its most general form. We see that the worst probability measure, \mathbb{P} , is chosen within an ambiguity set, \mathcal{A} , which is described in section C.4.1.

$$\min_{x \in \mathcal{X}} \max_{\mathbb{P} \in \mathcal{A}} \mathbb{E}^{\mathbb{P}}[f(x, \xi)] \quad (\text{C.4.1a})$$

$$\text{s.t. } \min_{\mathbb{P}' \in \mathcal{A}'} \mathbb{P}'(g(x, \xi) \leq 0) \geq 1 - \epsilon. \quad (\text{C.4.1b})$$

With the general form established, our objective is to rephrase the stochastic program, Equation C.2.2, into a DRO program. Notably, Equation C.4.1 is formulated as a *min max* program, which aligns with the existing theory in the field. Consequently, we proceed by transforming Equation C.2.2 into the *min max* form to be able to apply established theory and methodologies directly.

$$\min_{\mathbf{x}, W_t} \sum_{t \in \mathcal{T}} -W_t + \max_{\mathbb{P} \in \mathcal{A}} \mathbb{E}^{\mathbb{P}} [p_t^{\text{DA}} \xi_t^{\text{DA}} + p_t^{\text{B}} \xi_t^{\text{B}}] \quad (\text{C.4.2a})$$

$$\text{s.t. } W_t = p_t^{\text{DA}} \hat{\lambda}_t^{\text{DA}} + p_t^{\text{B}} \hat{\lambda}_t^{\text{B}} + d_t \lambda^{\text{h}} - (p_t^{\text{DA}-} + p_t^{\text{B}-}) \lambda^{\text{TSO}} - z_t^{\text{su}} \lambda^{\text{su}} \quad \forall t \in \mathcal{T} \quad (\text{C.4.2b})$$

$$\min_{\mathbb{P}' \in \mathcal{A}'} \mathbb{P}' \left[p_t^{\text{w}} \leq \hat{P}_t^{\text{w}} + \xi_t^{\text{w}} \right] \geq 1 - \epsilon \quad \forall t \in \mathcal{T} \quad (\text{C.4.2c})$$

$$W_t \in \mathbb{R} \quad \forall t \in \mathcal{T} \quad (\text{C.4.2d})$$

$$\mathbf{x} \in \mathcal{X}. \quad (\text{C.4.2e})$$

Now that we have introduced the general DRO formulation for our problem, the first step is to define the structure of the ambiguity set, \mathcal{A} . Once this is done, we can proceed with reformulating the problem into a solvable form.

C.4.1 Selection of Ambiguity Set

Within the field of DRO, two main trains of thought exist when it comes to the selection of the ambiguity set, \mathcal{A} . The first approach is to define the ambiguity set as the set of all distributions satisfying constraints on the moments. An example is to define the ambiguity set from the empirical expectation and variance. Formally, the ambiguity set can be defined as

$$\mathcal{A} = \left\{ \mathbb{P} \in \mathcal{M}(\Xi) \mid \mathbb{E}^{\mathbb{P}}[\xi] = \hat{\mu}, \mathbb{V}^{\mathbb{P}}[\xi] = \hat{\Sigma} \right\} \quad (\text{C.4.3})$$

where Ξ is the uncertainty set, $\xi \in \Xi$ is some realization of uncertainty, $\mathcal{M}(\cdot)$ is the set containing all probability measures with the same support as the input set, $\hat{\mu}$ is the empirical expectation, and $\hat{\Sigma}$ is the empirical covariance. This approach is known as *moment based* DRO, and has been used with good results, see e.g., Delage and Ye, 2010b,

Mehrotra and Zhang, 2014, and Wiesemann et al., 2014. However, as shown by Z. Wang et al., 2014, moment-based DRO can lead to very conservative solutions.

In recent years there has been a growing popularity in the use of *metric based* DRO. The idea behind metric-based DRO is to define the ambiguity set as a ball using some metric in probability measures rather than all measures satisfying some moment restrictions. Generally, the ambiguity sets are given as

$$\mathcal{A} = \{\mathbb{P} \in \mathcal{M}(\Xi) \mid d(\mathbb{P}, \mathbb{P}_N) \leq \theta\} \quad (\text{C.4.4})$$

where \mathbb{P}_N is the empirical measure, and $d(\cdot, \cdot)$ is some suitably selected distance measure. Popular choices of distance measures include different types of ϕ -divergences such as *Kullback–Leibler*, *Hellinger distance* etc. These divergences are – as the name suggests – not metrics which can lead to strange results as shown in Gao and Kleywegt, 2017a. Perhaps as a result of this, there has been a surge in work done using the *Wasserstein* (also known as *Kantorovich–Rubinstein* or *earth mover’s*) distance within the field of DRO, originally started by Esfahani and Kuhn, 2015. The Wasserstein distance is a metric in probability measures, and as such it yields more natural results. Additionally, as shown in Z. Chen et al., 2018, using the Wasserstein distance can lead to more reliable results using fewer training samples compared to ϕ -divergences.

For these reasons, we will throughout this work focus on metric-based DRO, specifically with the use of the Wasserstein distance. We denote our ambiguity set by

$$\mathcal{A}_{W_\infty} = \left\{ \mathbb{P} \in \mathcal{M}(\Xi) \mid d_W^{(\infty)}(\mathbb{P}, \mathbb{P}_N) \leq \theta \right\} \quad (\text{C.4.5})$$

where $d_W^{(\infty)}$ defines the Wasserstein distance using the infinity norm.

The Wasserstein distance, as given in e.g. Villani, 2003, is defined on a metric space, (M, d) , as

$$d_W(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu, \nu)} \int d(x, y) d\gamma(x, y) \quad (\text{C.4.6})$$

where Γ is the set of all copulas between the probability measure for X , μ , and the probability measure for Y , ν , and $(x, y) \in M$. Notice that for the Wasserstein distance with the 1-norm, the distance function, $d(x, y)$, is simply the absolute difference of x and y , $|x - y|$, and for the infinity norm, $d(x, y) = \max(|x - y|)$. These are the only two norms that will result in a linear Wasserstein-DRO problem as we later will see. In Appendix D we have given an example of how to calculate the Wasserstein distance between two discrete measures.

C.4.2 A Wasserstein DRO Wind/Hydrogen Model

Once the ambiguity set is determined as Equation C.4.5, our next objective is to cast Equation C.4.2 into a solvable form using the chosen ambiguity set. However, we encounter two main challenges in this process. First, we need to address the presence of a maximum over an expectation operator in the objective function. Secondly, we must find a solution to effectively minimize the chance constraint, which requires a separate approach to ensure a solvable formulation.

The Objective Function

To reformulate the inner maximization problem in the objective function we use Corollary 5.1 of Esfahani and Kuhn, 2015:

$$\max_{\mathbb{P} \in \mathcal{A}_{W_\infty}} \mathbb{E}^{\mathbb{P}}[f(x, \xi)] = \min_{\psi \geq 0, \sigma_i, \gamma_i \geq 0} \psi \theta + \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} \sigma_i \quad (\text{C.4.7a})$$

$$\text{s.t. } c^\top \hat{\xi}_i + \gamma_i(h - Q\hat{\xi}_i) \leq \sigma_i \quad \forall i \in \mathcal{N} \quad (\text{C.4.7b})$$

$$\|c^\top - Q^\top \gamma_i\|_1 \leq \psi, \quad \forall i \in \mathcal{N} \quad (\text{C.4.7c})$$

where Q and h defines the support of $\hat{\xi}_i$, and ψ , σ_i and γ_i are dual variables coming from the reformulation. While it may be possible to apply the formula without a complete understanding of its underlying principles, a more satisfactory approach from a mathematical standpoint would involve understanding the complete derivation. For this reason, we have provided the detailed derivation in Appendix E.

The Distributionally Robust Chance Constraint

The second challenge is how to reformulate the distributionally robust chance constraint which is given as

$$\min_{\mathbb{P} \in \mathcal{A}_{W_\infty}} \mathbb{P} \left[a(x)^\top \xi \leq b(x) \right] \geq 1 - \epsilon. \quad (\text{C.4.8})$$

Through the utilization of binary variables, we were able to convert the chance constraint, Equation C.3.2, into an equivalent solvable form. This technique can be extended to distributionally robust chance constraints, as exemplified in Z. Chen et al., 2022. However, the dependence on binary variables persists in the reformulation, yielding computational problems. Consequently, we will once more approximate Equation C.4.8 by reformulating it as a CVaR constraint, following the methodology outlined in Zymler et al., 2013:

$$\max_{\mathbb{P} \in \mathcal{A}_{W_\infty}} \mathbb{P}\text{-CVaR}_\epsilon \left[a^\top(x) \xi - b(x) \right] \leq 0 \Rightarrow \min_{\mathbb{P} \in \mathcal{A}} \mathbb{P} \left[a^\top(x) \xi \leq b(x) \right] \geq 1 - \epsilon. \quad (\text{C.4.9})$$

As earlier demonstrated, we apply Theorem 2 in Rockafellar, Uryasev, et al., 2000 to reformulate the left side of Equation C.4.9.

$$\max_{\mathbb{P} \in \mathcal{A}_{W_\infty}} \min_{\tau \in \mathbb{R}} \tau + \frac{1}{\epsilon} \mathbb{E}^{\mathbb{P}} \left[[a^\top(x) \xi - b(x) - \tau]^+ \right] \leq 0 \Leftrightarrow \quad (\text{C.4.10})$$

$$\min_{\tau \in \mathbb{R}} \tau + \frac{1}{\epsilon} \max_{\mathbb{P} \in \mathcal{A}_{W_\infty}} \mathbb{E}^{\mathbb{P}} \left[[a^\top(x) \xi - b(x) - \tau]^+ \right] \leq 0 \quad (\text{C.4.11})$$

where the interchange of the maximization and minimization operations is justified by a stochastic saddle point theorem due to Shapiro and Kleywegt, 2002. We observe that $\max_{\mathbb{P} \in \mathcal{A}_{W_\infty}} \mathbb{E}^{\mathbb{P}} \left[[a^\top(x) \xi - b(x) - \tau]^+ \right]$, can be reformulated by using Equation C.4.7. Consequently, we can reformulate it as a minimization problem and incorporate it into the outer minimization. This gives us the following final form:

$$\tau_t + \frac{1}{\epsilon} \left(\psi_t^{\text{CVaR}} \theta^{\text{CVaR}} + \frac{1}{|\mathcal{N}_t|} \sum_{i \in \mathcal{N}_t} \sigma_{i,t}^{\text{CVaR}} \right) \leq 0 \quad (\text{C.4.12a})$$

$$-\hat{\xi}_{t,i}^w - \left(\hat{P}_t^w - p_t^w \right) - \tau_t + \gamma_{i,t,1}^\top \left(h_t^w - Q_t^w \hat{\xi}_{t,i}^w \right) \leq \sigma_{i,t}^{\text{CVaR}} \quad (\text{C.4.12b})$$

$$\gamma_{i,t,2}^\top \left(h_t^w - Q_t^w \hat{\xi}_{t,i}^w \right) \leq \sigma_{i,t}^{\text{CVaR}} \quad (\text{C.4.12c})$$

$$\left\| (Q_t^w)^\top \gamma_{i,t,1} + 1 \right\|_1 \leq \psi_t^{\text{CVaR}} \quad (\text{C.4.12d})$$

$$\left\| (Q_t^w)^\top \gamma_{i,t,2} \right\|_1 \leq \psi_t^{\text{CVaR}}. \quad (\text{C.4.12e})$$

where Q^w and h^w defines the support of $\hat{\xi}_{t,i}^w$, and the rest of the variables are dual variables¹.

Reformulating the Wind/Hydrogen Model with Uncertainty

By utilizing the general reformulation of the objective Equation C.4.7 and the chance constraint Equation C.4.12, we can transform Equation C.4.2 into a solvable form. To achieve this, we need to determine a support for the three random variables: Wind power error, day-ahead price error, and imbalance price error. While the wind power variable has a natural restriction due to its physical limitations, the two price variables are not inherently bound and can theoretically take any value in the real numbers. Therefore, we choose a support that assumes a range of the forecast plus or minus 10,000, effectively rendering it unbounded. Additionally, to simplify the formulation, we opt for a common Q and h for the imbalance and day-ahead price since they appear in the same expectation operator and share the same forecast.

$$Q^{\text{Price}} := \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix}, \quad h^{\text{Price}} := \begin{bmatrix} 10,000 \\ 10,000 \\ 10,000 \\ 10,000 \end{bmatrix} \quad (\text{C.4.13})$$

which gives the following support

$$Q^{\text{Price}} \begin{bmatrix} \xi^{\text{DA}} \\ \xi^{\text{B}} \end{bmatrix}_t \leq h^{\text{Price}}. \quad (\text{C.4.14})$$

Next, we consider the wind variable, which has a meaningful support due to the limitations of the wind turbine's production capacity. The wind turbine can generate power between 0 and its maximum capacity denoted as C^w . Therefore, if \hat{P}_t^w represents the forecasted wind turbine production in hour t , the maximum undershoot is $C^w - \hat{P}_t^w$, and the maximum overshoot is \hat{P}_t^w . To define the support in terms of Q and h , we can express it as:

$$Q^w := \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad h^w := \begin{bmatrix} C^w - \hat{P}_t^w \\ \hat{P}_t^w \end{bmatrix} \quad (\text{C.4.15})$$

which gives the following support

¹See Appendix E.

$$Q^w \xi_t^w \leq h^w. \quad (\text{C.4.16})$$

With the ambiguity set defined as Equation C.4.5, we can now represent Equation C.4.2 as follows:

$$\min_{\mathbf{x}, W_t, \Upsilon_1} \sum_{t \in \mathcal{T}} -W_t + \psi_t \theta + \frac{1}{|\mathcal{N}_t|} \sum_{i \in \mathcal{N}_t} \sigma_{i,t} \quad (\text{C.4.17a})$$

$$\text{s.t. } W_t = p_t^{\text{DA}} \hat{\lambda}_t^{\text{DA}} + p_t^{\text{B}} \hat{\lambda}_t^{\text{B}} + d_t \lambda^{\text{h}} - (p_t^{\text{DA}-} + p_t^{\text{B}-}) \lambda^{\text{TSO}} - z_t^{\text{su}} \lambda^{\text{su}} \quad \forall t \in \mathcal{T} \quad (\text{C.4.17b})$$

$$\begin{bmatrix} p^{\text{DA}} & p^{\text{B}} \end{bmatrix}_t \begin{bmatrix} \xi^{\text{DA}} \\ \xi^{\text{B}} \end{bmatrix}_{i,t} + \gamma_{i,t}^{\top} \left(h_t^w - Q_t^w \begin{bmatrix} \xi^{\text{DA}} \\ \xi^{\text{B}} \end{bmatrix}_{i,t} \right) \leq \sigma_{i,t} \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.4.17c})$$

$$\|(Q_t^w)^{\top} \gamma_{i,t} - \begin{bmatrix} p^{\text{DA}} & p^{\text{B}} \end{bmatrix}_t\|_1 \leq \psi_t \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.4.17d})$$

$$\tau_t + \frac{1}{\epsilon} \left(\psi_t^{\text{CVaR}} \theta^{\text{CVaR}} + \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_{i,t}^{\text{CVaR}} \right) \leq 0 \quad \forall t \in \mathcal{T} \quad (\text{C.4.17e})$$

$$-\hat{\xi}_{t,i}^w - (\hat{P}_t^w - p_t^w) - \tau_t + \gamma_{i,t,1}^{\top} (h_t^w - Q_t^w \hat{\xi}_{t,i}^w) \leq \sigma_{i,t}^{\text{CVaR}} \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.4.17f})$$

$$\gamma_{i,t,2}^{\top} (h_t^w - Q_t^w \hat{\xi}_{t,i}^w) \leq \sigma_{i,t}^{\text{CVaR}} \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.4.17g})$$

$$\|(Q_t^w)^{\top} \gamma_{i,t,1} + 1\|_1 \leq \psi_t^{\text{CVaR}} \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.4.17h})$$

$$\|(Q_t^w)^{\top} \gamma_{i,t,2}\|_1 \leq \psi_t^{\text{CVaR}} \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.4.17i})$$

$$\gamma_{i,t}, \gamma_{i,t,1}, \gamma_{i,t,2} \geq 0 \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.4.17j})$$

$$v \in \mathbb{R} \quad \forall v \in \Upsilon_1 \quad (\text{C.4.17k})$$

$$W_t \in \mathbb{R} \quad \forall t \in \mathcal{T} \quad (\text{C.4.17l})$$

$$\mathbf{x} \in \mathcal{X} \quad (\text{C.4.17m})$$

where

$$\Upsilon_1 = \{\psi_t, \psi_t^{\text{CVaR}}, \sigma_{i,t}, \sigma_{i,t}^{\text{CVaR}}, \gamma_{i,t}, \gamma_{i,t,1}, \gamma_{i,t,2}, \tau_t \mid \forall t \in \mathcal{T} \wedge \forall i \in \mathcal{N}_t\}. \quad (\text{C.4.18})$$

The problem remains a mixed-integer linear program and can be solved using standard solvers such as Gurobi and CPLEX, as mentioned in Gurobi Optimization, LLC, 2023 and Cplex, 2012.

While the proposed framework incorporates the consideration of forecast errors, it does not guarantee the distributions within the ambiguity set to be realistic, as discussed in Arrigo et al., 2022. To address this limitation, the next section will expand the ambiguity set by incorporating linear dependence information between the day-ahead and imbalance price errors. This extension aims at enforcing more realistic distributions.

C.4.3 A Wasserstein DRO with Correlation Wind/Hydrogen Model

In the previous section, we discussed that Arrigo et al., 2022 found that using only the Wasserstein metric to parameterize ambiguity can lead to conservative and unrealistic distributions. To address this, we introduce a linear dependence structure to the ambiguity set. We denote the set \mathcal{A}_{Σ} , which is given in Equation C.4.19.

$$\mathcal{A}_\Sigma = \left\{ \mathbb{P} \in \mathcal{M}(\Xi) \mid d_W^{(\infty)}(\mathbb{P}, \mathbb{P}_N) \leq \theta, \mathbb{E}^\mathbb{P}[(\xi - \hat{\mu})(\xi - \hat{\mu})^\top] \preceq \hat{\Sigma} \right\} \quad (\text{C.4.19})$$

Theorem 1 in Gao and Kleywegt, 2017b provides a general reformulation for this set. However, since we have a linear objective, are using the empirical distribution as the reference distribution, and using a Wasserstein metric with the infinity norm, we can use Corollary 1 in Gao and Kleywegt, 2017b, which provides a semidefinite reformulation.

$$\max_{\mathbb{P} \in \mathcal{A}_\Sigma} \mathbb{E}^\mathbb{P}[f(x, \xi)] = \min_{\psi \geq 0, \Psi \succeq 0, \sigma_i, \alpha_i} \psi \theta + \langle \Psi, \hat{\Sigma} \rangle_F + \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} \sigma_i \quad (\text{C.4.20a})$$

$$\text{s.t. } \begin{bmatrix} \Psi & -\frac{1}{2}(c - \alpha_i + 2\Psi\hat{\mu}) \\ -\frac{1}{2}(c - \alpha_i + 2\Psi\hat{\mu})^\top & -\alpha_i^\top \hat{\xi}_i + \hat{\mu}\Psi\hat{\mu}^\top + \sigma_i \end{bmatrix} \succeq 0 \quad \forall i \in \mathcal{N} \quad (\text{C.4.20b})$$

$$\|\alpha_i\|_1 \leq \psi \quad \forall i \in \mathcal{N} \quad (\text{C.4.20c})$$

The mathematical derivation required to go from $\max_{\mathbb{P} \in \mathcal{A}_\Sigma} \mathbb{E}^\mathbb{P}[f(x, \xi)]$ to the reformulation presented in Equation C.4.20 is not a trivial process. It necessitates a comprehensive understanding of the underlying concepts. Attempting to apply the reformulation without a proper understanding of its meaning can result in nontrivial errors that may be difficult to detect. As a result, we have provided a detailed derivation of the reformulation in Appendix G to ensure clarity and avoid potential mistakes.

Uncertainty in the Second-Order Moment Estimate

The ambiguity set, Equation C.4.19, includes a constraint that guarantees the positive semidefiniteness of $\hat{\Sigma} - \mathbb{E}^\mathbb{P}[(\xi - \hat{\mu})(\xi - \hat{\mu})^\top]$. This constraint assesses the likelihood of ξ being close to $\hat{\mu}$ by considering the correlations expressed in $\hat{\Sigma}$. However, a potential challenge of this formulation is that it blindly trusts the second-order estimate $\hat{\Sigma}$. To address this issue, Delage and Ye, 2010a propose a trust parameter η as:

$$\mathbb{E}^\mathbb{P}[(\xi - \hat{\mu})(\xi - \hat{\mu})^\top] \preceq \eta \hat{\Sigma}, \quad (\text{C.4.21})$$

where $\eta \geq 1$. The range of η from 1 to ∞ allows for controlling the level of trust in $\hat{\Sigma}$, from 100% to 0%. Hence when $\eta = \infty$, the trust in the estimate, $\hat{\Sigma}$, is 0% and the resulting program is equivalent to the initial DRO program discussed in section C.4.2.

To our knowledge, the combination of a Wasserstein constraint and a second-order moment constraint with a trust parameter in the ambiguity set has not been explored previously. To understand how to incorporate the trust parameter, η , into Equation C.4.20, we refer to Appendix G. We give the updated general reformulation below:

$$\min_{\psi \geq 0, \Psi \succeq 0, \sigma_i, \alpha_i} \psi \theta + \langle \Psi, \eta \hat{\Sigma} \rangle_F + \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} \sigma_i \quad (\text{C.4.22a})$$

$$\text{s.t. } \begin{bmatrix} \Psi & -\frac{1}{2}(c - \alpha_i + 2\Psi\hat{\mu}) \\ -\frac{1}{2}(c - \alpha_i + 2\Psi\hat{\mu})^\top & -\alpha_i^\top \hat{\xi}_i + \hat{\mu}\Psi\hat{\mu}^\top + \sigma_i \end{bmatrix} \succeq 0 \quad \forall i \in \mathcal{N} \quad (\text{C.4.22b})$$

$$\|\alpha_i\|_1 \leq \psi \quad \forall i \in \mathcal{N}. \quad (\text{C.4.22c})$$

Reformulating the Wind/Hydrogen Model with Uncertainty

In Equation C.4.22, it is apparent that the price variable has support across the entire real number line, whereas the reformulation of the wind chance constraint using Equation C.4.12 still exhibits a polyhedral support. We can therefore reuse the support provided in Equation C.4.15. It is important to emphasize that the covariance estimate in Equation G.1c is based on the likelihood estimate rather than the unbiased estimate commonly used in standard statistics packages. Using the unbiased covariance matrix can result in incorrect results² and, in some cases, an infeasible primal formulation, as shown by certain dual variables in Equation C.4.22 diverging to $-\infty$. The reformulation of the wind/hydrogen model with uncertainty, Equation C.2.2, is presented below.

$$\min_{\mathbf{x}, W_t, \Upsilon_2, \Psi_t} \sum_{t \in \mathcal{T}} -W_t + \psi_t \theta + \langle \Psi_t, \eta \hat{\Sigma}_t \rangle_F + \frac{1}{|\mathcal{N}_t|} \sum_{i \in \mathcal{N}_t} \sigma_{i,t} \quad (\text{C.4.23a})$$

$$\text{s.t. } W_t = p_t^{\text{DA}} \hat{\lambda}_t^{\text{DA}} + p_t^{\text{B}} \hat{\lambda}_t^{\text{B}} + d_t \lambda^{\text{h}} - (p_t^{\text{DA}-} + p_t^{\text{B}-}) \lambda^{\text{TSO}} - z_t^{\text{su}} \lambda^{\text{su}} \quad \forall t \in \mathcal{T} \quad (\text{C.4.23b})$$

$$\begin{bmatrix} \Psi_t & -\frac{1}{2} \left(\begin{bmatrix} p^{\text{DA}} \\ p^{\text{B}} \end{bmatrix}_t - \alpha_{i,t} + 2\Psi_t \hat{\mu}_t \right) \\ -\frac{1}{2} \left(\begin{bmatrix} p^{\text{DA}} \\ p^{\text{B}} \end{bmatrix}_t - \alpha_{i,t} + 2\Psi_t \hat{\mu}_t \right)^{\top} & -\alpha_{i,t}^{\top} \begin{bmatrix} \xi^{\text{DA}} \\ \xi^{\text{B}} \end{bmatrix}_{i,t} + \hat{\mu}_t \Psi_t \hat{\mu}_t^{\top} + \sigma_{i,t} \end{bmatrix} \succeq 0 \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.4.23c})$$

$$\|\alpha_{i,t}\|_1 \leq \psi_t \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.4.23d})$$

$$\tau_t + \frac{1}{\epsilon} \left(\psi_t^{\text{CVaR}} \theta^{\text{CVaR}} + \frac{1}{|\mathcal{N}_t|} \sum_{i \in \mathcal{N}_t} \sigma_{i,t}^{\text{CVaR}} \right) \leq 0 \quad \forall t \in \mathcal{T} \quad (\text{C.4.23e})$$

$$-\hat{\xi}_{t,i}^{\text{w}} - \left(\hat{P}_t^{\text{w}} - p_t^{\text{w}} \right) - \tau_t + \gamma_{i,t,1}^{\top} \left(h_t^{\text{w}} - Q_t^{\text{w}} \hat{\xi}_{t,i}^{\text{w}} \right) \leq \sigma_{i,t}^{\text{CVaR}} \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.4.23f})$$

$$\gamma_{i,t,2}^{\top} \left(h_t^{\text{w}} - Q_t^{\text{w}} \hat{\xi}_{t,i}^{\text{w}} \right) \leq \sigma_{i,t}^{\text{CVaR}} \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.4.23g})$$

$$\left\| (Q_t^{\text{w}})^{\top} \gamma_{i,t,1} + 1 \right\|_1 \leq \psi_t^{\text{CVaR}} \quad \forall t \in \mathcal{T} \quad (\text{C.4.23h})$$

$$\left\| (Q_t^{\text{w}})^{\top} \gamma_{i,t,2} \right\|_1 \leq \psi_t^{\text{CVaR}} \quad \forall t \in \mathcal{T} \quad (\text{C.4.23i})$$

$$\gamma_{i,t,1}, \gamma_{i,t,2} \geq 0 \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}_t \quad (\text{C.4.23j})$$

$$\Psi_t \in S_+^{2 \times 2} \quad \forall t \in \mathcal{T} \quad (\text{C.4.23k})$$

$$v \in \mathbb{R} \quad \forall v \in \Upsilon_2 \quad (\text{C.4.23l})$$

$$W_t \in \mathbb{R} \quad \forall t \in \mathcal{T} \quad (\text{C.4.23m})$$

$$\mathbf{x} \in \mathcal{X} \quad (\text{C.4.23n})$$

where $S_+^{2 \times 2}$ is the set of positive semidefinite matrices and

$$\Upsilon_2 = \{ \psi_t, \psi_t^{\text{CVaR}}, \sigma_{i,t}, \sigma_{i,t}^{\text{CVaR}}, \gamma_{i,t,1}, \gamma_{i,t,2}, \tau_t, \alpha_{i,t} \mid \forall t \in \mathcal{T} \wedge \forall i \in \mathcal{N}_t \}. \quad (\text{C.4.24})$$

We observe that the problem we are dealing with is a mixed-integer semidefinite program, which is not solvable using standard solvers such as Gurobi and CPLEX. These solvers are limited to handling mixed-integer linear programs and mixed-integer second-order cone programs, as mentioned in Gurobi Optimization, LLC, 2023 and Cplex, 2012. However, an open-source solver called Pajarito³ has been developed by a group at MIT for the general class of mixed-integer convex programs, Coey et al., 2020. Pajarito uses a branch-and-bound LP outer approximation algorithm that combines a MILP solver like Gurobi or CPLEX with a convex solver like MOSEK, ApS, 2023.

²We encountered this issue ourselves, and identifying and rectifying the error was quite challenging! Avoid making the same mistake :)

³We recommend this YouTube video for a thorough introduction.

C.5 Multi Stage Programs

We have explored various one-stage programs under uncertainty, but another approach to handling uncertainty exists. If some of the random variables in the problem are realized earlier than others, it is possible to formulate a multi-stage program. The advantage of such an approach is that instead of working with unconditional expectations of the random variables, we can utilize conditional expectations for variables realized later in the process. This means certain expectations inherently possess more information due to their access to data from earlier realized variables. This theoretical aspect is intriguing; however, multi-stage programs become significantly more complex in practice as the number of stages increases, making them challenging to solve. Due to this increased complexity, we have not been able to generate satisfactory results and, therefore, have not included them in the main report. Nevertheless, an analysis of the multi-stage approach is provided in Appendix K.

D | Evaluation of Energy Trading Models

D.1 Data

The data used in this study was obtained through a collaboration with Siemens Gamesa Renewable Energy (SGRE) as part of the HOMEY project. The dataset comprises a range of key variables related to a real-life wind farm, including:

- Price forecasts for day-ahead electricity markets, provided by the collaborator.
- Actual day-ahead electricity prices, provided by the collaborator.
- Imbalance prices sourced from the European Network of Transmission System Operators for Electricity (ENTSOE) and confirmed with data from Nordpool.
- Wind speed measurements obtained from on-site sensors.
- The manufacturer's power curve for the wind turbines.
- Numerical weather predictions (NWP) of wind speed, provided by the collaborator.

All data were collected between May 1st, 2022, and March 6th, 2023. Using real-world data from a functioning wind farm provides a robust and realistic basis for the analysis conducted in this thesis.

The data source did not provide wind power forecasts. As such, the forecasts were generated using the statistical methods described in section B.2 based on the available wind speed measurements and wind speed predictions.

We first generated pseudo wind power measurements by parsing the wind speed measurements through the manufacturer's power curve. Next, we fit a statistical model on the pseudo measurements using historical wind speed predictions as predictors. Finally, we used the statistical model and the wind speed predictions to produce wind power forecasts. Figure B.2 shows the manufacturer's power curve together with the statistical model; notice that noise was added to the measurement shown in the plot for confidentiality.

Although the wind power forecasts were not based on actual wind power measurements, we do not expect this to have any significant impact on the conclusions of this thesis.

To ensure the quality of the analysis and the reliability of the results, we have chosen only to include days for which we have access to all relevant data points. This consists of the wind speed measurements, the day-ahead price forecast, the realized day-ahead prices, the imbalance prices, and the wind speed predictions. We removed a total of 5 days spread across the time period.

Additionally, we assume that the day-ahead price forecast accurately represents prices in the imbalance market; that is, we use the day-ahead price forecast as the forecast for the price in the imbalance market. While this assumption is not entirely without risk, it is the best available information for predicting the price in the imbalance market.

Table D.1 shows a summary of data used in this thesis. The abbreviation for each data type is given together with an explanation in natural language and the origin of the data.

Table D.1 – Overview of available data with their meaning in natural language.

Abbreviation	Meaning	Origin
λ_t^{DA}	Realized day-ahead price at time t	Measurement
$\hat{\lambda}_t^{\text{DA}}$	Day-ahead price forecast for time t	Forecast
λ_t^{B}	Realized imbalance price at time t	Measurement
$\hat{\lambda}_t^{\text{B}}$	Imbalance price forecast for time t	Forecast
P_t^{W}	Wind power production at time t	Estimate
\hat{P}_t^{W}	Wind power production forecast for time t	Forecast

Hydrogen Capabilities

We incorporate the hydrogen capabilities of our site. For hydrogen production, we utilize an alkaline electrolyzer, operating at a cell pressure of 30 bar and a temperature of 90 degrees Celsius. We refer to the efficiency curves depicted in Figure B.1 to characterize its efficiency.

The electrolyzer is designed to handle a maximum load of 50% of the wind turbine's capacity, ensuring optimal utilization of available resources. Additionally, it has a standby load of only 1% and a minimum operating load of 15%. Notably, a starting cost of 25 EUR is associated with initiating the electrolyzer from its "off" state. Regarding procurement, we consider the TSO tariff, which represents the additional tariff on top of the electricity price at which electricity is purchased. The TSO tariff stands at 15.06 EUR/MWh.

We have set a maximum capacity of 210.7 kg for hydrogen storage. In addition, we use a compressor with an efficiency of 1.204 kWh/kg.

In our analysis, we assume that all generated hydrogen is sold on a daily contract basis, with a fixed price of 2.1 EUR/kg. The contractual agreement necessitates a minimum daily requirement of 35.1 kg of hydrogen, ensuring a consistent supply. There are no explicit upper bounds set for the contract except for the implicit limitation imposed by the capacity of the electrolyzer.

Confidentiality Disclaimer

Due to confidentiality reasons, all non-public data and results derived thereof presented in this thesis, including performance metrics, tables, graphs, and any other data, have been modified for privacy protection. This is done through aggregation, re-scaling, and/or adding noise (for plots) to ensure that the original data cannot be reconstructed. This modification process does not affect the validity of the results presented in this thesis. Additionally, the location of the wind farm will remain confidential.

D.2 Testing

In order to evaluate the effectiveness of a trading strategy/model, it is important to test it using appropriate performance metrics. These metrics provide insights into the profitability, risk, and overall performance of a given strategy. Additionally, it is necessary to have a clear understanding of how the earnings of a specific model are realized, as this impacts the calculation of performance metrics. Finally, to optimize and compare different trading strategies, it is important to have a method for fitting their parameters so they can be evaluated on a test dataset.

D.2.1 Performance Metrics

To assess the performance of trading strategies, it is crucial to employ multiple metrics that measures various aspects of performance, such as risk and return. Therefore, we use six commonly used performance metrics to evaluate our trading strategies: Average daily return, volatility, Sharpe ratio, information ratio, worst single day (WSD), and percentage of profitable trades.

The average daily return of a trading strategy is given as

$$\bar{R}^{(s)} = \frac{1}{N} \sum_{i=1}^N R_i^{(s)}, \quad (\text{D.2.1})$$

where $(\cdot)^{(s)}$ refers to a strategy s . It provides a straightforward and simple way of measuring the profitability of a strategy. However, it does not consider the risk involved in generating those returns.

Volatility is a measure of the degree of variation in the daily return of a trading strategy over a specified period. The volatility is given by the standard deviation of the returns,

$$\hat{\sigma}^{(s)} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (R_i^{(s)} - \bar{R}^{(s)})^2}. \quad (\text{D.2.2})$$

High volatility is typically associated with higher levels of risk, while low volatility is generally considered less risky. By incorporating volatility as a performance metric, we can gain insights into the risk-return trade-off of a trading strategy.

The Sharpe ratio, proposed by William Sharpe in his papers, Sharpe, 1966 and Sharpe, 1975, is a more comprehensive measure of risk-adjusted return. It quantifies a trading strategy's performance by evaluating its excess reward in relation to its variability. We let $r^{(f)}$ denote the risk-free rate of return and $R^{(d)}$ the excess return. Then the performance of the trading strategy, s , relative to the risk-free rate is given by:

$$R^{(d)} := R^{(s)} - r^{(f)}. \quad (\text{D.2.3})$$

Since there is no natural risk-free rate of return, we assume it to be zero. As a result, the return of the strategy, s , and the excess return are equivalent. The ex-post Sharpe ratio or historical Sharpe ratio can be calculated using the following formula:

$$\hat{S}^{(s)} := \frac{\mathbb{E}[R^{(s)}]}{\sqrt{\mathbb{V}[R^{(s)}]}} = \frac{\bar{R}^{(s)}}{\hat{\sigma}^{(s)}}. \quad (\text{D.2.4})$$

We see that a higher Sharpe ratio indicates a better risk-adjusted performance.

As stated, we do not have a natural risk-free rate, but we are interested in comparing strategies to a benchmark strategy, $R^{(b)}$, given by the deterministic trading strategy. This can be quantified using the information ratio (IR), Grinold and Kahn, 2000, which is the Sharpe ratio generalized by replacing the risk-free rate with the return of a benchmark strategy:

$$\text{IR}^{(s)} := \frac{\mathbb{E}[R^{(s)} - R^{(b)}]}{\sqrt{\text{V}[R^{(s)} - R^{(b)}]}} \quad (\text{D.2.5})$$

We see that the interpretation of the information ratio is similar to the Sharpe ratio.

The last two metrics are the percentage of profitable trades and WSD. The percentage of profitable trades is calculated as the number of profitable trades divided by the total number of trades executed, which can be written as,

$$\%R_+^{(s)} = \frac{\sum_{i=1}^N \mathbb{1}(R_i^{(s)} \geq 0)}{N}, \quad (\text{D.2.6})$$

where $\mathbb{1}$ is an indicator function. This metric provides information on the success rate of a trading strategy, but it does not consider the magnitude of the gains or losses. To also capture the magnitude, we use WSD, which measures the largest nominal decline in one day. The WSD can be calculated as

$$\text{WSD} = \min_{i \in \{1, \dots, N\}} R_i^{(s)}. \quad (\text{D.2.7})$$

D.2.2 Realization

Chapter C presents models for determining the optimal day-ahead bid. However, to calculate how much each model earns, we must consider the price realizations in both the day-ahead and imbalance market. In the imbalance market, a price is assigned for each hour, and the engineer can adjust the planned hydrogen production accordingly. To reflect this reality, an ideal model would have a stage for every hour to allow for rescheduling. Unfortunately, this is computationally infeasible, as explained in Appendix K.1. As a result, we realize earnings passively and actively, creating a lower and upper bound of what can be achieved in the real world given a fixed day-ahead bid.

Passive Realization

The passive realization method follows a predetermined hydrogen production plan, regardless of the market conditions in the imbalance market, which emerge after the day-ahead price is established. Assuming that rescheduling will not have a negative impact on average, this approach results in the least possible earnings. It thus creates a lower bound on the potential earnings given a particular day-ahead bid. The passive realization is illustrated in Figure D.1.



Figure D.1 – Passive realization of the hydrogen production plan for the found day-ahead bid at time -12.

Active Realization

In contrast, the active realization method involves rescheduling the hydrogen production after the day-ahead price realization. This rescheduling is based on complete information about all the uncertainties in the imbalance market and thus assumes perfect information, which is impossible to achieve in practice. However, this approach sets an upper bound on the potential earnings given a day-ahead bid. In Figure D.2, the active realization is illustrated.



Figure D.2 – Active realization of the hydrogen production plan for the found day-ahead bid at time -12.

D.3 Parameter Fitting

Each model described in Chapter C incorporates several parameters that need to be assigned a value. To find the best parameter values, we partition the dataset into a training and test set and then optimize the parameters using the training data. We use six months of data (from May 1 to October 31) as the training set and four months (from November 1 to March 6) as the test set. We optimize the model parameters based on the average daily return ($\bar{R}^{(s)}$) using the active realization method. However, a challenge arises because the models in Chapter C require some data on the first day of deciding the day-ahead bid. Consequently, we use a burn-in period of 6 weeks (42 days) before making any predictions, effectively reducing the length of the training set to a little under five months. It is important to note that if a model has more than one parameter, we will optimize bilinearly. This approach provides an effective solution but also means that the optimum found is not guaranteed to be a global optimum. After determining the optimal parameters based on the training data, we test the models in the test set to evaluate their ability to generalize to new data. This evaluation is critical to ensure that the models are appropriately calibrated and can make good decisions in practice.

E | Results for the Wind/Hydrogen Models

E.1 Sample Selection

When optimizing under uncertainty, a standard assumption is stationarity of the underlying distribution, Besbes et al., 2015. This assumption is challenged by two aspects of our problem.

The first challenge relates to temporal stationarity. Since we are working with time series data, the underlying distribution can change over time. To tackle this challenge, we must choose a set of recent samples that strike a balance between having sufficient data and ensuring a certain degree of temporal stationarity. While we aim to incorporate 40 samples, the nature of our problem prevents us from simply selecting the 40 most recent ones. For instance, if we are currently at day 50, the day-ahead bid needs to be made at 12 PM on day 49. Hence, we specifically choose the most recent samples that are at least 12-36 hours old, taking into consideration the timing of the day-ahead bid. We have illustrated the temporal sampling restriction in Figure E.1.

This brings us to the second challenge, namely stationarity with respect to the price and wind power forecasts. The underlying distribution may depend on the price and wind power forecast and hence a day-ahead forecast of 600 EUR and a power forecast of 0.1 MW may come from a totally different distribution than forecasts of 30 EUR and 0.8 MW.

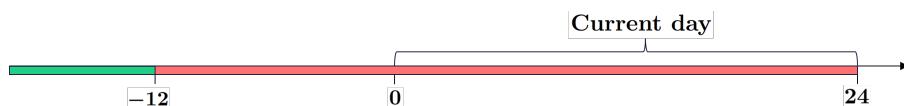


Figure E.1 – Hours that are available for the sample selection on some given day. Green indicates available while red is not available.

How to select similar forecasts?

In the following, we will explain how we select similar forecasts and to assist the reader we provide a graphical overview in Figure E.2.

We employ the K-medoids¹ algorithm with $K = 20$, to cluster forecasts into similar groups. The clustering process considers imbalance prices, day-ahead forecasts, wind power forecasts, and load forecasts. However, a direct clustering method faces a significant limitation: We lack access to the last 36 hours of imbalance prices needed to cluster with the forecasted values.

To overcome this limitation, we develop our own forecasting model for imbalance prices. Due to the high variability in the imbalance market, we adjust the historical imbalance

¹The K-medoids algorithm is explained in Appendix H.

prices to ensure they deviate by at most 100 EUR from the day-ahead prices. To forecast the next 36 hours of imbalance prices, we utilize an XGBoost model, which is an implementation of a gradient boosting model developed by T. Chen and Guestrin, 2016. This model is trained on historical data, using the following input features: Wind power forecasts, day-ahead price forecasts, load forecasts, scheduled flow², and wind speed forecasts.

We re-train the XGBoost model and re-cluster the data from the first day in our training data to the last day in our test data. This is done to ensure that the model is updated when new data becomes available.

It is important to note that we have not conducted a proper hyperparameter tuning for the XGBoost model and the K-medoids algorithm with a cross validation setup. Due to the computational time demands of our models, we have manually adjusted the hyperparameters and evaluated their performance solely on the training data. For detailed implementation steps, please refer to Appendix I, which provides a pseudo-code outlining the process of training the XGBoost model, adjusting the imbalance prices, and incorporating the predicted imbalance prices into the clustering process.

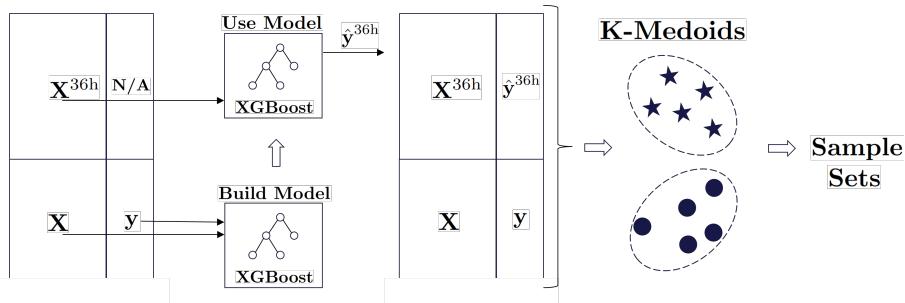


Figure E.2 – An overview of how we first forecast and then cluster to select relevant samples.

E.2 Results

In the following, results for all the models will be presented. We will abbreviate how we refer to the different models as follows:

- **The Deterministic Wind/Hydrogen Model:** The deterministic model
- **The SAA Wind/Hydrogen Model:** The SAA model
- **The Wasserstein DRO Wind/Hydrogen Model:** The DRO model
- **The Wasserstein DRO with Correlation Wind/Hydrogen Model:** The DRO model with correlation

E.2.1 The Deterministic Wind/Hydrogen Model

The first model we discuss is the deterministic model, which is explained in detail in section B.3. We multiply the imbalance price by 0.9999 to ensure that the expected power production is sold in the day-ahead market, with any remaining surplus or deficit corrected in the imbalance market. The deterministic model serves as the baseline for all subsequent models. Table E.1 displays the results obtained from the deterministic model on the training data. It is important to note that the reported daily average return of 1,000

²Load forecasts and scheduled flow are sourced from ENTSOE.

EUR is a normalized value due to confidentiality reasons. All other returns have been normalized using this constant, and the remaining metrics have been computed based on these normalized returns.

Table E.2 presents the results obtained by applying the deterministic model to the test data. We notice that the returns decrease by 150 EUR for the active realization and by around 80 EUR for the passive realization. However, the volatility decreases significantly, and the percentage of days which shows a profit increases from the training to the test data. These changes suggest a shift in market dynamics, which we will discuss later in this chapter. Figure L.1 in Appendix L shows the accumulated returns, highlighting the relatively low volatility in the test data.

Table E.1 – Performance metrics for the deterministic model for the training data. Values have been scaled for confidentiality. The deterministic model is used as the benchmark, therefore IR is not defined.

	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Passive	894	954	0.94	NA	-545	87.9%
Active	1,000	969	1.03	NA	-250	95.0%

Table E.2 – Performance metrics for the deterministic model for the test data. Values have been scaled for confidentiality. The deterministic model is used as the benchmark, therefore IR is not defined.

	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Passive	713	543	1.31	NA	-345	94.3 %
Active	745	543	1.37	NA	-249	95.1 %

E.2.2 The SAA Wind/Hydrogen Model

The second model we tested is the SAA model presented in section C.3. It includes a chance constraint that limits the model's belief about how much wind power it can produce. This constraint has a parameter, ϵ , which we fitted to the training data. In Table L.1 in Appendix L these results are given. From the table, we observed that the variation in ϵ has a minimal impact. However, with a greater number of decimal places, $\epsilon = 0.5$ yielded the highest return for the active realization. For this value of ϵ , the return is lower than the deterministic model and the volatility is higher which results in a lower Sharp ratio. This is due to the model's high level of speculation, which is reflected in a lower percentage of profitable days. In other words, the model engages in many high-risk trades, where too many are unsuccessful. For $\epsilon = 0.5$ the results from the training data are shown in Table E.3. The SAA model was evaluated on the test data using $\epsilon = 0.5$, and the results are presented in Table E.4. With the deterministic model as a baseline, we found that the SAA model also underperformed on the test data, generating less profit, with higher volatility and a lower percentage of profitable days. The accumulated returns for the passive and active realizations for the test data are shown in Figure L.2, Appendix L. Based on these findings, we conclude that incorporating uncertainty through the SAA method is a poor representation that exposes the model to excessive risk.

Table E.3 – Performance metrics for the SAA model with optimal ϵ on the training data.

	ϵ	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Passive	0.50	860	1,277	0.67	-0.045	-4,804	80.7 %
Active	0.50	986	1,290	0.76	-0.018	-4,721	85.0 %

Table E.4 – Performance metrics for the SAA model with optimal ϵ on the test data.

	ϵ	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Passive	0.50	676	591	1.14	-0.199	-1,438	92.6 %
Active	0.50	715	588	1.22	-0.164	-1,336	94.3 %

E.2.3 The Wasserstein DRO Wind/Hydrogen Model

The third model is the DRO model presented in subsection C.4.2. There are three parameters in this model, namely ϵ for the distributionally robust chance constraint, θ^{CVaR} for the ambiguity set related to the distributionally robust chance constraint, and θ for the ambiguity set related to the expectation operator in the objective, which controls the belief on the historical errors of the price forecasts.

As explained in section D.3, we fit the parameters bilinearly and present the results in Table L.4, Table L.3, and Table L.2 found in Appendix L. Our findings indicate that the DRO model performs better than the deterministic model in terms of return, with the highest return achieved when $\epsilon = 0.5$, $\theta^{\text{CVaR}} = 0.0$, and $\theta = 22$ for the active realization, but the DRO model exhibits higher volatility, resulting in a lower Sharp ratio. Increasing θ can improve the Sharp ratio and come closer to the Sharp ratio of the deterministic model. However, this comes with a decrease in the expected daily return. We observe that ϵ has little effect, as seen with the SAA model. Increasing θ^{CVaR} does decrease the volatility, but it also results in a significant decrease in return resulting in an overall lower Sharp ratio. The result in the training data for the best parameters is shown in Table E.5.

Next, we test the optimal parameters on the test data with results presented in Table E.6. Similar to the training data, the DRO model outperforms the deterministic model regarding the return. It also has a lower standard deviation and a better Sharp ratio for both passive and active realizations. The accumulated returns for the passive and active realizations for the test data are shown in Figure L.3, Appendix L.

Table E.5 – Performance metrics for the DRO model with optimal parameters on the training data.

	θ	θ^{CVaR}	ϵ	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Passive	22	0	0.50	893	993	0.90	-0.003	-754	86.4 %
Active	22	0	0.50	1,010	1,012	1.00	0.040	-742	93.6 %

Table E.6 – Performance metrics for the DRO model with optimal parameters on the test data.

	θ	θ^{CVaR}	ϵ	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Passive	22	0	0.50	714	534	1.34	0.004	-263	94.3 %
Active	22	0	0.50	757	539	1.41	0.129	-179	95.9 %

E.2.4 The Wasserstein DRO with Correlation Wind/Hydrogen Model

The DRO model with correlation is the fourth model. In this model, we introduce dependence between the errors of the forecasted prices as explained in subsection C.4.3.

The DRO model with correlation involves four key parameters. The first two parameters are ϵ and θ^{CVaR} , which relates to the distributionally robust chance constraint. The third parameter, θ , determines the Wasserstein radius for the ambiguity set associated with the expectation operator in the objective, influencing the level of confidence in the errors of historical price forecasts. Lastly, the trust parameter, η , determines our level of trust in the estimated correlation structure, with 1 representing complete trust and ∞ representing no trust.

As explained in section D.3, we fit the parameters bilinearly and present the results in Table L.6 and Table L.5 found in Appendix L. Notice that we do not present the dependence on ϵ and θ^{CVaR} since these parameters proved to make little difference for the DRO model and hence assumed to be the same. Our findings indicate that the DRO model with correlation performs better than the deterministic model in terms of return, with the highest return achieved when $\theta = 25$ and $\eta = 1.5$ for the active realization. However, the DRO model with correlation exhibits higher volatility, resulting in a lower Sharpe ratio. Increasing θ can improve the Sharpe ratio and come closer to the Sharpe ratio of the deterministic model, but this comes with a decrease in the expected daily return. The result in the training data for the best parameters is shown in Table E.7.

Next, we test the optimal parameters on the test data with results presented in Table E.8. Similar to the training data, the DRO model with correlation outperforms the deterministic model regarding the return. It also has a lower standard deviation and a better Sharpe ratio for passive and active realizations. The accumulated returns for the passive and active realizations for the test data are shown in Figure L.4, Appendix L.

Table E.7 – Performance metrics for the DRO model with correlation with optimal parameters on the training data.

	θ	η	θ^{CVaR}	ϵ	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	% R_+
Passive	25	1.5	0	0.50	894	993	0.90	0.000	-761	87.9%
Active	25	1.5	0	0.50	1,014	1,014	1.00	0.049	-749	95.0%

Table E.8 – Performance metrics for the DRO model with correlation with optimal parameters on the test data.

	θ	η	θ^{CVaR}	ϵ	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	% R_+
Passive	25	1.5	0	0.50	712	535	1.33	-0.010	-263	94.3%
Active	25	1.5	0	0.50	761	539	1.41	0.176	-174	95.9%

E.3 An Adaptive Wasserstein Radius for the Price

Upon analyzing the performance of the models, we observed that the DRO models outperformed the deterministic model, while the SAA model underperformed, especially in the test data.

The SAA model and DRO models use their understanding of the uncertainty to make better decisions. However, we saw that the SAA model had very high volatility in its returns which led to an overall loss compared to other models, indicating over-speculation.

This can be attributed to the naive approach employed by the SAA model, which fully trusts the historical realization of uncertainty.

Furthermore, it is worth noting that the market was generally highly-priced in the training period with high volatility, as shown in Table E.9. Such conditions are suitable for speculation and may have contributed to the higher returns generally seen in the training data. In contrast, the test period had less volatility, making predicting the market easier. Although these conditions resulted in the SAA model underperforming significantly, the DRO models still performed better than the deterministic model in the test data.

Table E.9 – Mean values of day-ahead price, imbalance price, wind power, and the standard deviation of the same parameters.

Data period	$\bar{\lambda}^{\text{DA}}$	σ^{DA}	$\bar{\lambda}^{\text{B}}$	σ^{B}	\bar{P}^{w}	σ^{w}
Train	285.1	163.1	273.3	219.3	0.27	0.18
Test	157.7	112.6	156.1	125.0	0.39	0.26

This underpins the fact that the prices and, in particular, uncertainty in price forecasts are not stationary; they change over time. Therefore, we would like to dynamically adjust the risk level by adapting the Wasserstein radius for the price in the DRO models to align with the prevailing market conditions.

The adaptive risk management algorithm, outlined in Appendix J, incorporates the principles of a PI controller. The algorithm assesses the profitability of increasing or decreasing the Wasserstein radius for price on the previous day and adjusts it accordingly. If increasing the radius would have resulted in higher profits, a small step is taken to increase it. Conversely, if decreasing the radius would have been more profitable, a small step is taken to reduce it. This feedback loop allows the algorithm to adapt the risk awareness and align it with the changing dynamics of the market.

To avoid excessive oscillations in risk levels, the algorithm includes integral action, which considers the historical performance of the control parameter. If the radius has been consistently increased or decreased over multiple iterations, the step size gradually becomes larger. This "momentum" effect ensures a smoother adjustment process, allowing the algorithm to change risk awareness to suit market conditions.

The parameters in the controller were fitted on the training data, employing the passive realization as the deciding factor for adjusting the risk awareness. The choice was made to use the passive and not active realization because they have the same day-ahead bid. The passive realization also includes the original hydrogen production plan. Therefore, the passive realization is more sensitive and can pick up on smaller signals.

The results obtained from the training data are presented in Table E.10, highlighting the substantial performance improvement achieved by choosing the risk level adaptively.

Table E.10 – Performance metrics for the DRO model with adaptively chosen Wasserstein radius for the price on the training data. $\varepsilon = 0.5$ and $\theta^{\text{CVaR}} = 0$.

	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Passive	919	1,112	0.83	0.052	-1,103	84.3 %
Active	1,041	1,132	0.92	0.088	-1,091	89.3 %

Table E.11 presents the performance of the DRO model with adaptive risk awareness on the test data. While the controller yielded significantly better results than a constant risk

level in the training data, the adaptive approach demonstrated comparable performance in the test data.

Table E.11 – Performance metrics for the DRO model with adaptively chosen Wasserstein radius for the price on the test data. $\varepsilon = 0.5$ and $\theta^{\text{CVaR}} = 0$.

	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Passive	715	546	1.31	0.022	-260	94.3 %
Active	753	548	1.38	0.090	-179	95.9 %

Figure E.3 illustrates the selection of θ by the algorithm over time. Notably, distinct regimes can be observed where the controller chooses a risk level within a narrow interval. This suggests that while the market may exhibit local stability, market-changing events may occur that shift the underlying dynamics. Specifically, there is a significant increase in θ around day 130, corresponding to the beginning of September 2022. This period was a time when the market experienced panic and subsequently returned to more reasonable price levels.

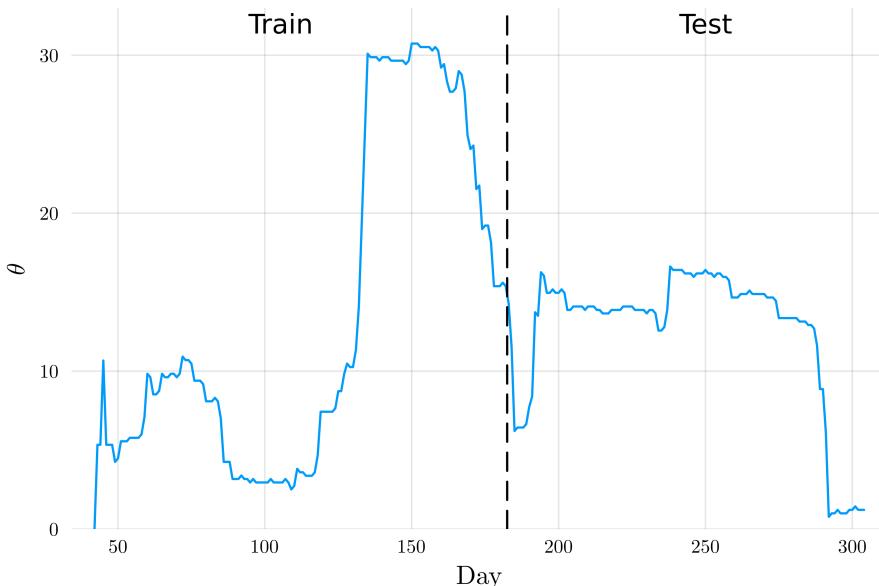


Figure E.3 – Risk awareness w.r.t. price, θ , chosen by the control algorithm over time.

The adaptive approach, with its ability to adjust the risk level based on historical performance and market conditions, shows promise in enhancing the profitability and robustness of the DRO model. However, it is important to note that this is an experimental stage, and further research and refinement are required to validate the effectiveness of the adaptive risk management approach and optimize the controller for different market scenarios.

E.4 Comparison of Models

In this section, we make an in-depth comparison of the performance of the different wind/hydrogen models. To assess the models' performance, we compare returns, which provide a measure of the profitability of each model over the test period, and volatility, which provides insights into the risk appetite of the models.

Figure E.4 shows the accumulated returns for all methods using the active realization. Upon analyzing the results, it becomes evident that the DRO models exhibit similar levels of performance. The deterministic model also performs reasonably well. While it may not leverage the same level of adaptability as the DRO models, its consistent performance suggests a stable investment approach that can yield favorable returns in these conditions. Notably, the SAA model falls significantly behind the DRO models and the deterministic model. The speculative nature of the SAA model introduces higher risks and uncertainties, leading to less consistent performance.

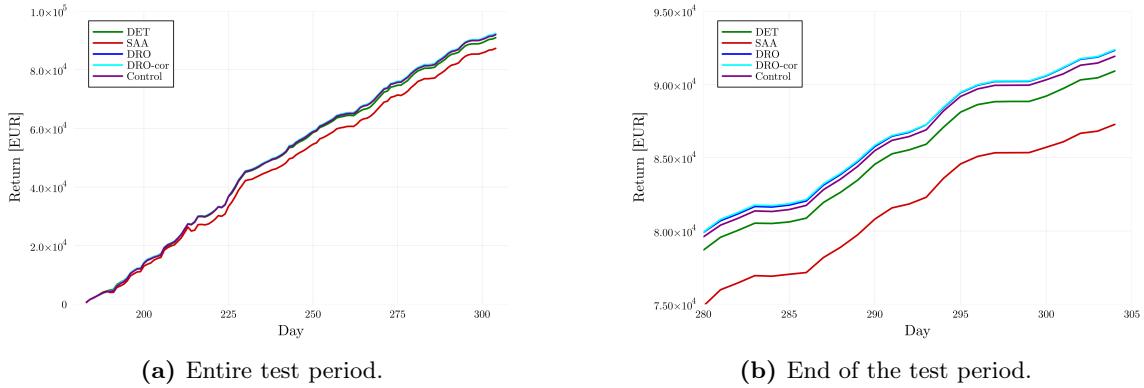


Figure E.4 – Accumulated return for active realizations in the test data.

Figure E.5 shows the accumulated returns for all models using the passive realization. Upon examining the results, it becomes apparent that all the models, except the SAA model, perform almost identically for the passive realization. The SAA model once again falls significantly behind the other models.

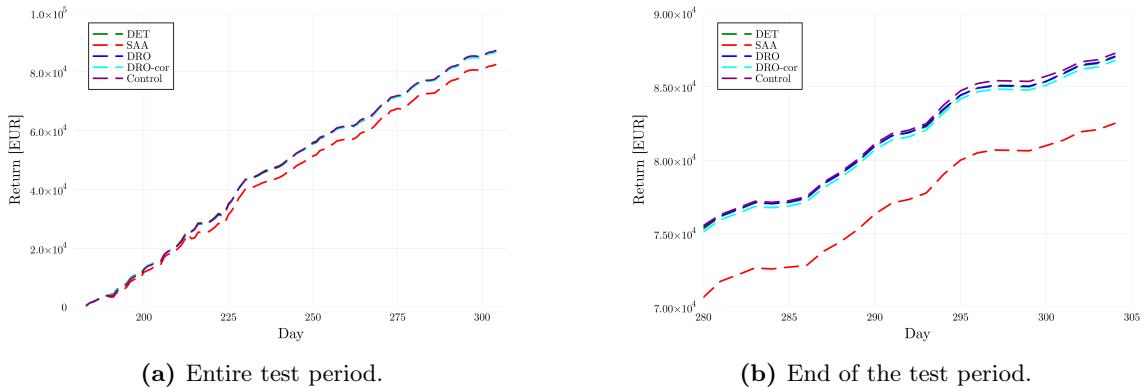


Figure E.5 – Accumulated return for passive realizations in the test data.

Table E.12 presents the performance metrics of the wind/hydrogen models for both the training and test datasets, with percentages given relative to the deterministic model for the same data period and realization.

When comparing to the deterministic model in the training data, the SAA model shows a slight decrease in average daily return (-1.4% for active, -3.8% for passive) and a significant increase in volatility (33.1% for active, 33.9% for passive). The DRO model achieves a slight increase in average daily return (1.0% for active, -0.1% for passive) but also increased volatility (4.4% for active, 4.1% for passive). The DRO model with correlation also achieves a slight increase in average daily return (1.4% for active and 0.0% for passive) but again

with an increase in volatility (4.6% for active, 4.1% for passive). The DRO model with adaptive Wasserstein radius for the price outperforms the deterministic model in terms of average daily return (4.1% increase for active, 2.8% increase for passive), but once again with a volatility increase (16.8% for active, 16.6% for passive). These results indicate that the DRO model with adaptive Wasserstein radius for the price shows promise in achieving better performance for both active and passive realizations.

Comparing the results to the deterministic model in the test data, we observe that the SAA model exhibits a decrease in average return (-4.0% for active, -5.2% for passive) and an increase in volatility (8.3% for active, 8.8% for passive). The DRO model achieves better average daily return (1.6% for active, 0.1% for passive) while reducing volatility (-0.7% for active, -1.7% for passive). The DRO model with correlation also gets better average daily return (2.1% increase for active, -0.1% for passive) and lower volatility (-0.7% for active, -1.5% for passive). The DRO model with adaptive Wasserstein radius for the price demonstrates a slight increase in average return (1.1% for active, 0.3% for passive) but an increase in volatility (0.9% for active, 0.6% for passive).

Table E.12 – Average daily return, \bar{R} , and volatility, $\hat{\sigma}$, for the deterministic model, SAA, DRO, DRO with correlation, and Control (DRO with adaptive Wasserstein radius for price), with percentages relative to the deterministic model.

		Training				Test			
		Model	\bar{R}	% $\Delta\bar{R}$	$\hat{\sigma}$	% $\Delta\hat{\sigma}$	\bar{R}	% $\Delta\bar{R}$	$\hat{\sigma}$
Passive	DET	894	0.0%	954	0.0%	713	0.0%	543	0.0%
	SAA	860	-3.8%	1,277	33.9%	676	-5.2%	591	8.8%
	DRO	893	-0.1%	993	4.1%	714	0.1%	534	-1.7%
	DRO-cor	894	0.0%	993	4.1%	712	-0.1%	535	-1.5%
	Control	919	2.8%	1,112	16.6%	715	0.3%	546	0.6%
Active	DET	1,000	0.0%	969	0.0%	745	0.0%	543	0.0%
	SAA	986	-1.4%	1,290	33.1%	715	-4.0%	588	8.3%
	DRO	1,010	1.0%	1,012	4.4%	757	1.6%	539	-0.7%
	DRO-cor	1,014	1.4%	1,014	4.6%	761	2.1%	539	-0.7%
	Control	1,041	4.1%	1,132	16.8%	753	1.1%	548	0.9%

To gain further insights into the performance of different wind/hydrogen models, we examine Figure E.6 that displays the volatility of daily return (standard deviation) on the x-axis and the expected daily return on the y-axis. This plot considers all the models, including the deterministic model, the SAA model, the DRO model, the DRO model with correlation, and the DRO model with adaptive Wasserstein radius for the price (Control). Additionally, for the DRO models, multiple θ values outline their performance at various levels of trust in the empirical distribution of errors in the price forecast.

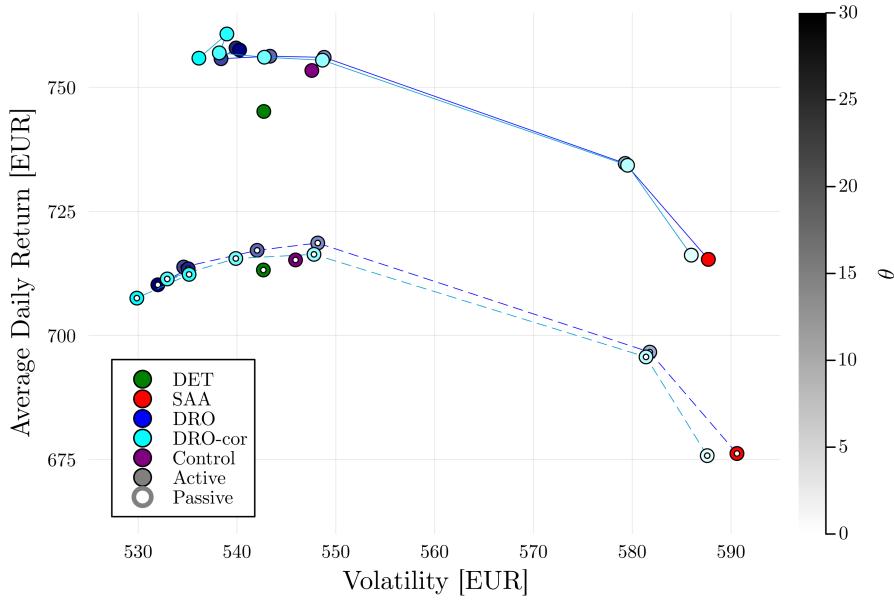


Figure E.6 – Average daily return and volatility of the different models for varying θ in the test data. θ takes the following values: $\{0, 5, 10, 15, 20, 25, 30\}$.

The plot reveals valuable information about each model's trade-off between volatility and expected return. First and foremost, it becomes evident that the SAA model exhibits poor performance in both volatility and return.

Moving on to the DRO models with fixed θ , which demonstrate similar performance characteristics when compared to each other. These models outline the *efficient frontier*, which represents the curve at which we cannot achieve higher return without accepting higher volatility. Along the efficient frontier, these models balance volatility and expected return. We note that the efficient frontier is outlined when $10 \leq \theta$, that is, a θ below 10 is sub-optimal in the test data. Additionally, notice that increasing θ has a larger effect on the passive than the active realization. This implies that the passive realization is more sensitive to the value of θ . In practice, we see that the passive realization spreads the hydrogen production before significant changes in the day-ahead bids occur, hence the difference. The deterministic model lies below the efficient frontier. Its position indicates that achieving higher profits with the same level of volatility is possible by utilizing the DRO models instead of relying solely on expected values.

Finally, the DRO model with adaptive Wasserstein radius for the price lies almost on the efficient frontier. The control algorithm is trained to maximize return. Therefore, it is close to the most speculative part of the efficient frontier. The model's positioning highlights its suitability for investors with a higher risk appetite, explicitly targeting maximum return. If an investor has other objectives, these should be reflected in how the control algorithm works.

F | Sensitivity Analysis

In this section, we investigate two key factors: The importance of sample selection and the effect of having access to electrolysis. It is important to note that we have omitted the DRO model with correlation from this analysis due to its high computational cost. By examining these aspects, we aim to gain a deeper understanding of our models' performance and limitations, enhancing our study's overall insights and applicability.

F.1 Importance of Sample Selection

In Chapter E, we saw that the DRO models outperformed the other models; however, these results are largely determined by how well we are able to select the errors. If the errors were better selected, we would expect that the SAA model would improve. To test how the DRO model compares with the SAA model and the deterministic model, when the selection of errors improves, we synthetically improve the selected errors; that is, we pretend we did a better job clustering the historical realizations.

Figure F.1a illustrates the selected errors for the imbalance price for all hours in a day. Single errors are illustrated as dots, their mean as a dark blue line, and one standard deviation of the errors as a light blue band around the mean. The realized error is given as a dashed black line, from which the mean of our selected errors seems to deviate quite a bit. To synthetically improve our selected errors, we move the mean of our selected errors towards the realized error by some percentage. For an improvement of 50%, this means:

$$\hat{\xi}_{t,i}^{(50\%)} = \hat{\xi}_{t,i} + 0.5(\xi_t - \hat{\mu}_t) \quad \forall i, \quad (\text{F.1.1})$$

where $\hat{\xi}_{t,i}$ is the i 'th error our clustering algorithm selects for time t , ξ_t is the realized error at time t , and $\hat{\mu}_t$ is the mean of all the errors selected by our clustering algorithm at time t . Figure F.1b shows the selected errors for a day with an improvement in the mean of 50%.

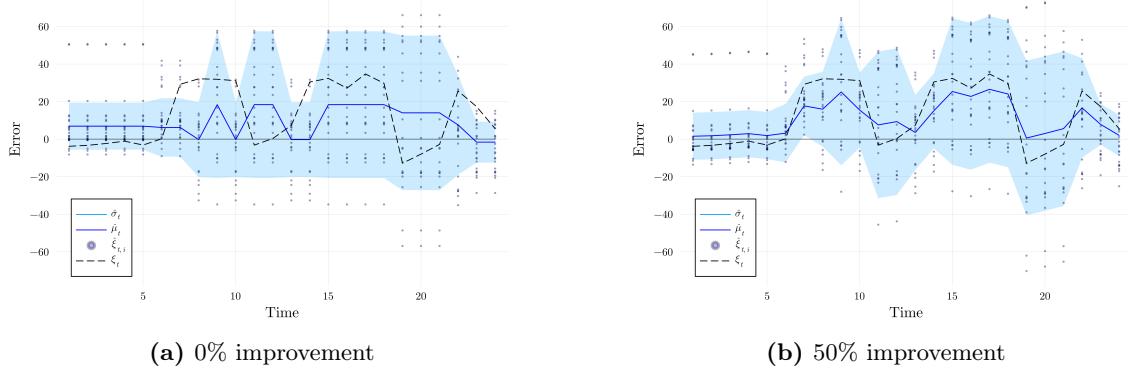


Figure F.1 – Empirical distributions for errors in the imbalance forecast determined by our clustering algorithm for a day with and without synthetic improvement to the mean. Dots are chosen historical realizations; the dashed line is the realization for the day in question.

Figure F.2 depicts the outcomes of solving the optimization problem for improvements of 0% to 18% to the mean when used on the test data. The considered models include the deterministic model, the SAA model, the DRO model with $\theta = 22$, and the DRO model with an adaptive Wasserstein radius for the price.

Upon examination, it becomes apparent that the deterministic model remains unaffected as it does not incorporate knowledge about the historical errors that describe the future realization. The DRO model with $\theta = 22$ gradually progresses as the percentage improvement increases. The SAA model starts at a lower performance level, but as the errors improve it surpasses the DRO model in both the active and passive realizations at an improvement of 8%.

The DRO model with adaptive Wasserstein radius for the price adjusts its level of risk aversion based on the improved error estimation. It consistently outperforms the SAA model until the improvement reaches 10% and remains nearly on par with the SAA model all the way up to an improvement of 18%.

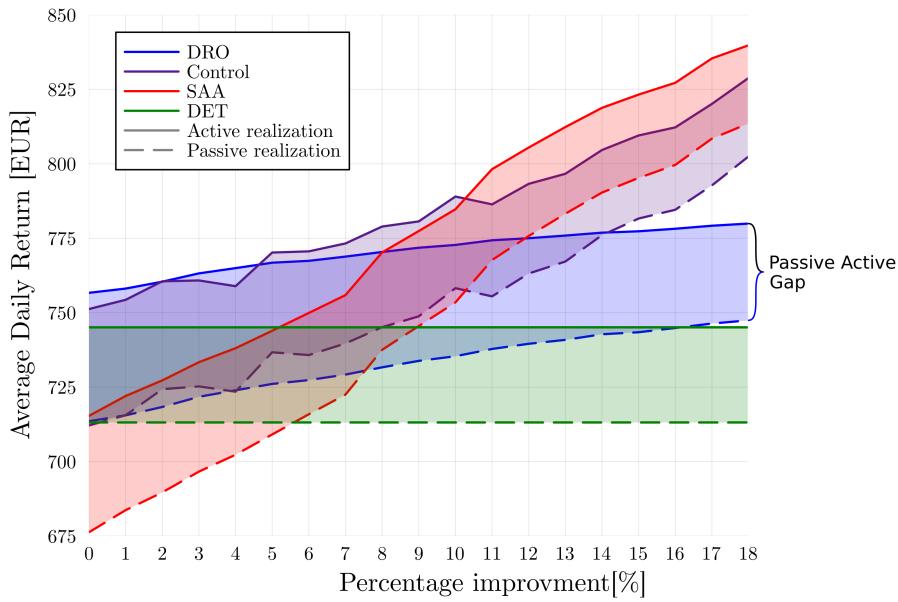


Figure F.2 – Performance of the deterministic model, the SAA model, the DRO model with $\theta = 22$, and the DRO model with adaptive Wasserstein radius for the price (Control) when the mean of the selected errors is synthetically improved.

However, the analysis has not provided any insight into the volatility aspects related to distributional robustness. This dimension is explored in Figure F.3.

Firstly, we observe that the deterministic model remains unchanged when improving the selected errors. Consequently, there is only one data point for the passive realization and one for the active realization.

Next, we examine the SAA model, which exhibits oscillations in volatility between 580 and 590 until an improvement of 16%. Beyond that point, the volatility increases to 600. The passive realization demonstrates slightly higher volatility compared to the active realization, although they are nearly equal.

Moving on to the DRO model with a fixed θ value of 22. This model displays the lowest volatility, and as the error estimation improves, both the volatility and return exhibit gradual and consistent growth. The passive realization for the DRO model with a fixed θ experiences lower volatility compared to the active realization.

Lastly, we analyze the DRO model with adaptive Wasserstein radius for the price. Initially, it starts with a volatility close to that of the DRO model with a fixed θ . However, as the error estimation improves, the volatility approaches that of the SAA model. At the highest level of improvement (18%), the model's volatility remains just below 580, which is approximately 20 EUR better than the SAA model. Additionally, in the initial stages, the passive realization exhibits lower volatility than the active realization, but as the error estimation improves, the volatility of the two realizations becomes nearly equal.

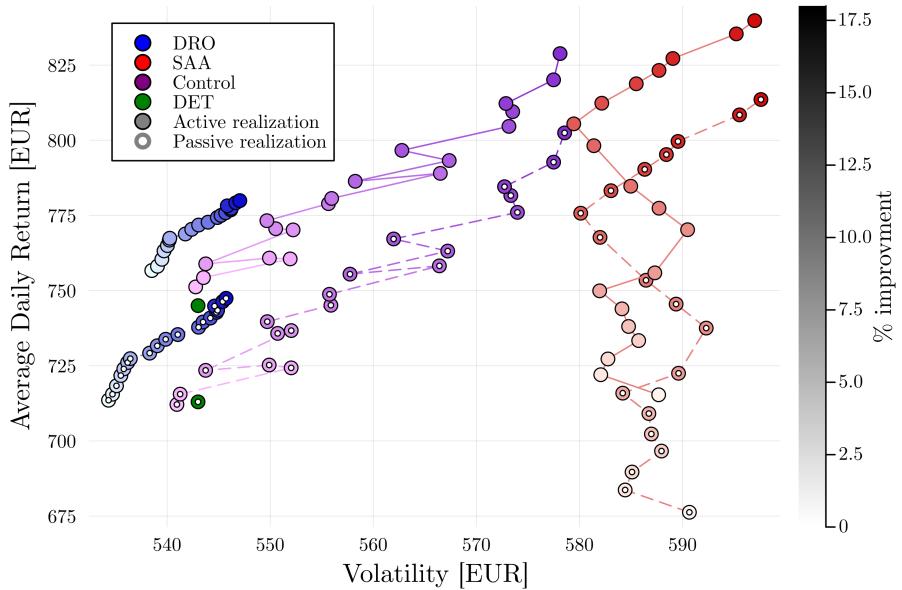


Figure F.3 – Return and volatility for the deterministic model, the SAA model, the DRO model with $\theta = 22$, and the DRO model with adaptive Wasserstein radius for the price. For each point on the outlined curves, the improvement increase by 1%, starting at 0% and ending at 18%.

F.2 Importance of the Hydrogen

So far, our analysis has focused on the overall performance of the entire wind/hydrogen power plant. However, we also want to understand the specific value that the electrolyzer brings compared to having only wind power. Figure F.4 shows the average daily return when varying the price of hydrogen (λ^H). The data is from the test period, and the DRO model is solved with $\theta = 22$.

The light blue band represents the DRO model, where daily hydrogen demand contracts of 35.1 kg are considered. For the dark blue band, we have removed these daily contracts, and the model determines the optimal amount of hydrogen to produce each day. We compare these two scenarios to a scenario without electrolysis altogether, but only the wind turbine (represented by the black line).

For the scenario without an electrolyzer, the passive and active realization is the same because no hydrogen production can be adjusted in the imbalance market. Therefore, the day-ahead bid completely determines the operation.

Figure F.4 also shows that when there are no daily contracts, the passive-active band outperforms the model without an electrolyzer entirely when the price of hydrogen exceeds 2 EUR/kg. However, this does not include investment costs, so the actual price may need to be higher than 2 EUR/kg to turn a profit.

In the scenario with daily minimum contracts on hydrogen, the price of hydrogen that needs to be charged is higher compared to the scenario without daily contracts for two reasons. Firstly, when there is no wind, and the electricity price is high, hydrogen production is still necessary, increasing the costs. Secondly, on some days when the wind is blowing and the electricity price is high, some electricity must be used for hydrogen production instead of being sold directly. As a result, the price of hydrogen must compensate for this missed opportunity for return. Consequently, we observe that the profitability of this scenario surpasses the profitability of the scenario without an electrolyzer at a much later

point compared to the scenario without daily minimum contracts. The profit surpasses the profit of the scenario without an electrolyzer only when the price of hydrogen exceeds 3.8 EUR/kg for the active realization and 4.8 EUR/kg for the passive realization.

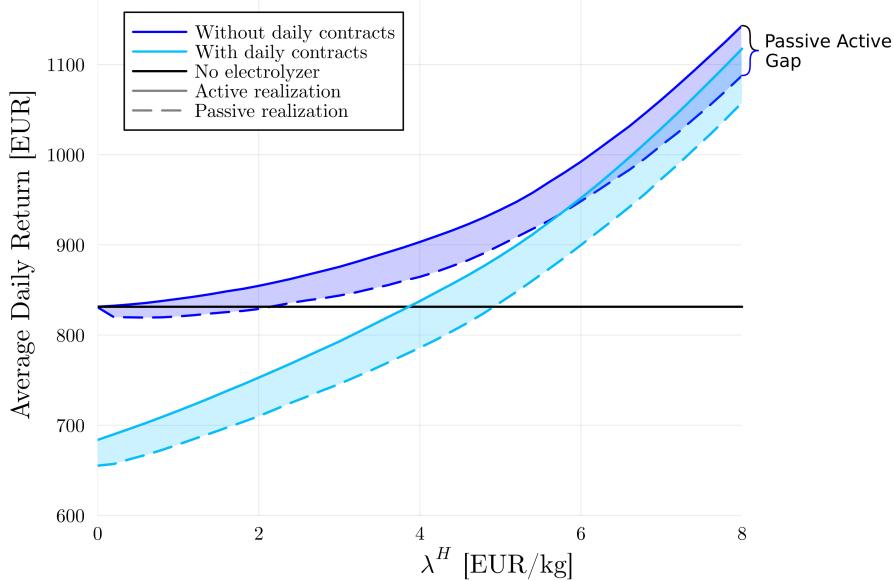


Figure F.4 – The average daily return for the test period. The DRO model is optimized with $\theta = 22$. Both with and without daily hydrogen contracts and a scenario without access to an electrolyzer are shown.

The information provided in Figure F.4 does not address the specific quantity of hydrogen produced in the different scenarios. To explore this aspect, we refer to Figure F.5, which illustrates the hydrogen production for the same scenarios as in Figure F.4.

Upon examination, we observe that as the price increases, the average daily amount of hydrogen produced with and without daily contracts seems to converge. Additionally, in the case without daily contracts, we notice that when the price exceeds 2.8 EUR/kg, the passive realization exhibits a higher average daily hydrogen production than what is required by the daily contracts. This happens at a hydrogen price of 3.2 EUR/kg for the active realization. Consequently, we can conclude that having a daily contract is suboptimal because it enforces production even on unfavorable days. If production were allowed to vary freely, the average daily production would exceed the amount specified in the daily contracts at a much lower cost in the test data.

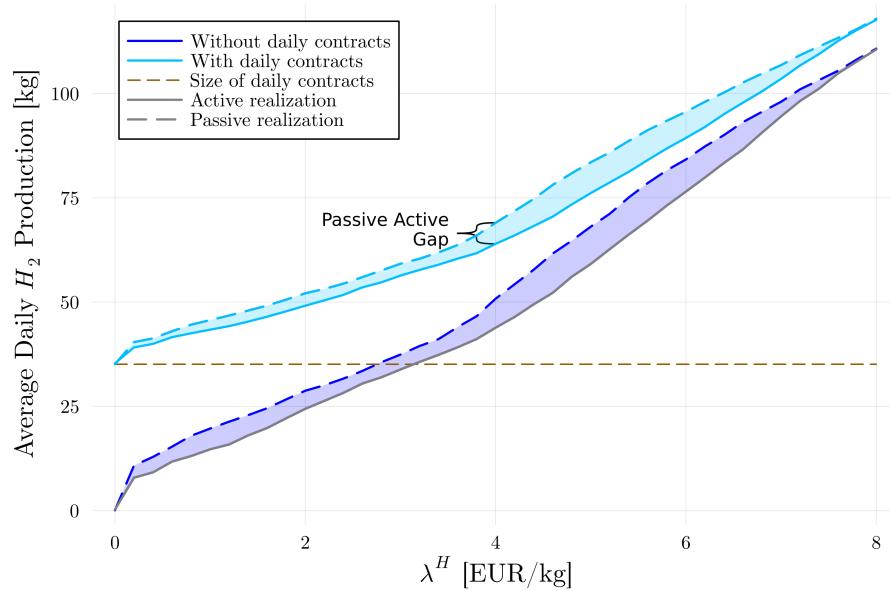


Figure F.5 – The hydrogen production for the DRO model with $\theta = 22$ is calculated using test data. Both with and without daily hydrogen contracts are shown. The size of the daily contracts is also shown.

Finally, we will explore the potentially volatility reducing effects of having access to electrolysis at a wind power plant. We will exclude the passive realization from our analysis as the findings are identical to those of the active realization. If you wish to view the corresponding plots for the passive realization, they are provided in Appendix M.

We begin by examining the scenario with daily contracts. Figure F.6 illustrates return and volatility for the deterministic model, the SAA model, and the DRO model with $\theta = 22$ when considering varying hydrogen prices, λ^H . Additionally, the scenario without access to an electrolyzer is shown.

Our observations reveal that increasing the price of hydrogen initially reduces volatility until a point after which it starts to increase. We note that for all models, having access to an electrolyzer can decrease the volatility. The volatility is reduced by approximately 20 EUR for the deterministic model and the DRO model and 10 EUR for the SAA model. The DRO model outperforms the other two models for all tested hydrogen prices, providing a slightly higher return with lower volatility. In Figure F.6, the lowest volatility is indicated by a gold circle. For the DRO model, this is at 3.2 EUR/kg; for both the deterministic model and the SAA model, this happens at 3.4 EUR/kg. Prices below this will yield both lower returns and higher volatility. Additionally, we note not having access to an electrolyzer gives a higher volatility compared to points with a similar return.

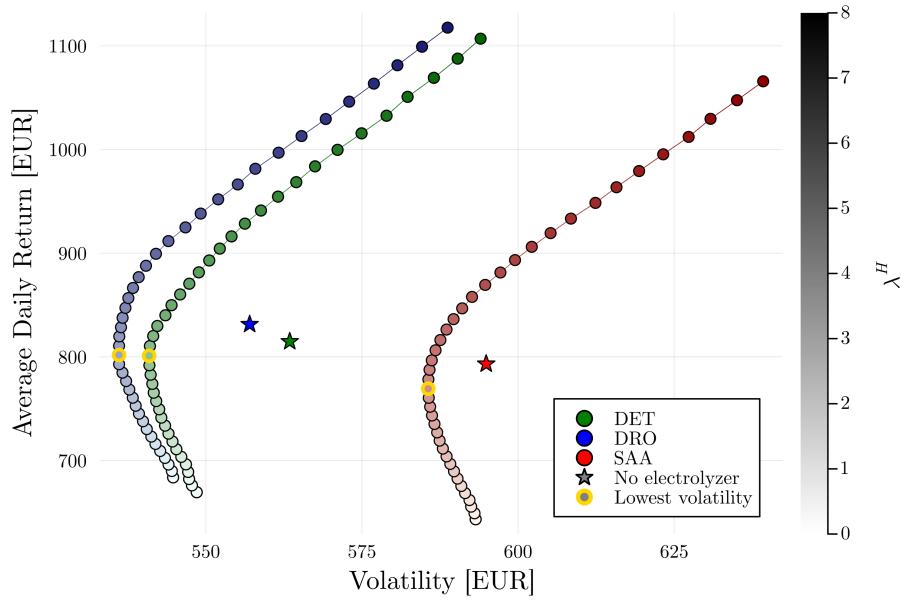


Figure F.6 – Active realization for the scenario with daily hydrogen contracts. For each point on the outlined curves, the hydrogen price increase by 0.2 EUR/kg, starting at 0.0 EUR/kg and ending at 8.0 EUR/kg.

Continuing our analysis, we also examined the scenario without daily contracts. The corresponding plot can be found in Figure F.7.

As seen in Figure F.4, the scenario without an electrolyzer is almost the same as having no contracts and a hydrogen price of zero. Hence, we also see that these two scenarios coincide for all models in Figure F.7.

In contrast to the results obtained in the scenario involving daily contracts, where the electrolyzer had a relatively smaller stabilizing effect on the SAA model compared to the other two models, our current findings reveal an equal stabilizing effect of the electrolyzer on all three models. Furthermore, when the electrolyzer is most effective, it reduces volatility by approximately 20 EUR.

Consistent with the previous scenario, the DRO model outperforms the two other models, delivering slightly higher returns and lower volatility for all tested hydrogen prices.

The lowest volatility for the deterministic model is found for a price of 3.6 EUR/kg, 3.8 EUR/kg for the DRO model, and 4.0 EUR/kg for the SAA model. These values are quite close to the findings in the scenario with daily contracts. This suggests that the stabilizing range remains the same with and without contracts, albeit with variations in magnitude for the SAA model.

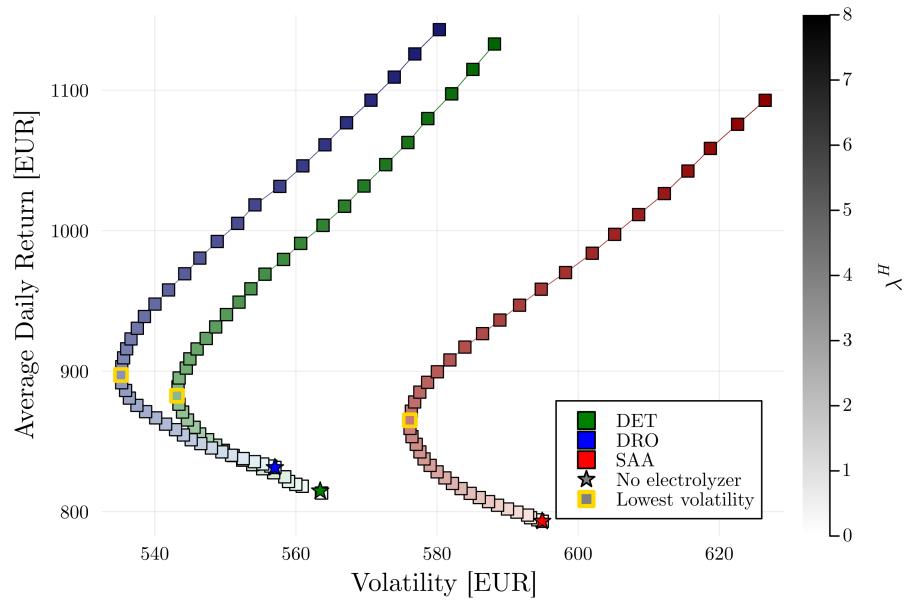


Figure F.7 – Active realization for the scenario without daily hydrogen contracts. For each point on the outlined curves, the hydrogen price increase by 0.2 EUR/kg, starting at 0.0 EUR/kg and ending at 8.0 EUR/kg.

G | Discussion

G.1 What have we seen?

Our study explored different aspects of wind/hydrogen power plant optimization models and uncovered some interesting findings.

Compared to the deterministic model the SAA model exhibited a decrease in returns between -5.2% and -4.0% and an increase in volatility between 8.8% and 8.3%. In contrast, the DRO model demonstrated an increase in returns between 0.1% and 2.1% and a decrease in volatility between -1.7% and 0.7%. Furthermore, incorporating correlation structures in the DRO model yielded a further increase in returns between -0.1% and 2.1% and a decrease in volatility between -1.5% and 0.7%. The DRO model with an adaptive Wasserstein radius exhibited an increase in returns between 0.3% and 1.1% and a slight increase in volatility between 0.6% and 0.9%. Note that all intervals are given by the passive and active realization.

One important factor that significantly affected the model's performance was the selection of historical errors for representing future uncertainty. We put a lot of effort into finding an efficient clustering algorithm, but we found that the SAA model did not perform well because it relied too much on the selected errors, leading to excessive speculation. On the other hand, the DRO models outperformed both the SAA model and the deterministic model.

We also conducted experiments where we synthetically improved the selected errors. As expected, when the errors improved – 8%, to be specific – the SAA model outperformed the DRO model. However, this raises an important question regarding the feasibility of developing an error selection algorithm that consistently selects errors so accurately that the SAA model outperforms the DRO model. If it were indeed possible to predict the error distribution with such precision, it would imply that the forecasts themselves could have been constructed in a more effective manner.

We also analyzed the effect of the Wasserstein radius for the price, θ . In the active realization, we found that for θ between 10 and 30, the day-ahead bids remained similar, indicating that θ did not have a strong influence on day-ahead bidding strategies. However, in the passive realization, we noticed that as θ increased beyond the optimal value, the model's return declined, as shown in Figure E.6. This suggests that higher θ values had a greater impact on the hydrogen production strategy before affecting the day-ahead bids. Specifically, increasing θ led to a more evenly spread hydrogen production profile throughout the day.

Furthermore, we observed that the profitability of wind/hydrogen power plants depended on hydrogen prices and contractual agreements. Without a daily contract, profitability was guaranteed when prices exceeded 2.0 EUR/kg, while with a daily contract, prices ranged between 3.8 EUR/kg and 4.8 EUR/kg for active and passive realization respectively. Interestingly, our analysis of the test data showed that without daily contracts, the average

daily production closely matched the specified contractual amount when hydrogen prices ranged from 2.8 EUR/kg to 3.2 EUR/kg for passive and active realization respectively. This suggests that extending the contract duration beyond a single day could enable hydrogen production on economically advantageous days, resulting in a more cost-effective generation of hydrogen.

In terms of computational time, the optimization models showed significant variations. The deterministic model, the SAA model, and the DRO model, which were formulated as mixed integer linear programs, generally had relatively short solution times, typically measured in minutes when back-testing on around 300 days, when solved on a standard laptop. On the other hand, the DRO model with correlation, formulated as a mixed integer semi-definite program, required significantly more computational power and took around 5 hours to back-test on the 300 days. This corresponds to 1 minute per day, so optimizing over a single day would be feasible for the more complex model. However, despite the longer solution time, the marginal improvement achieved by the DRO model with correlation compared to the DRO model without correlation was minimal. This highlights the need to carefully evaluate whether the extended computational time is justified by the small enhancements in results.

G.2 Future Research

This section outlines potential avenues for future research that can enhance the performance and applicability of wind/hydrogen power plant optimization models. Four potential research areas: Error selection/estimation, control of the risk-awareness, the refinement of the two-stage modeling approach, and end-to-end optimization. In the following subsections, we will delve into each area, discussing their potential contributions to the field and presenting opportunities for further advancements.

G.2.1 Error Selection/Estimation

Our study emphasized the significance of error selection/estimation. We chose to use a clustering algorithm to select historical errors to represent future uncertainty.

One way to improve the error selection algorithm is by investigating more advanced forecasting models, instead of the XGBoost model. Such sophisticated architectures enable the identification of complex relationships and subtle patterns within the features, contributing to more accurate results.

In addition, one can also expand the feature set utilized by the clustering algorithm. Incorporating relevant features such as meteorological and additional market data may enhance the algorithm's ability to capture the underlying drivers of uncertainty and improve the accuracy of error selection.

Another path to take is to estimate the future error directly. One way to do this is by exploring advanced probabilistic forecasting techniques. Integrating the probabilistic forecast as distributions directly has the potential to enhance the performance of the models.

G.2.2 Control of the Risk-Awareness

Our study introduced an algorithm for adaptive selection of the risk-awareness parameter, θ . This algorithm draws upon principles from control theory and determines whether to increase or decrease the parameter based on a short period of recent data. We determined the algorithm's parameters (step size, momentum, etc.) by evaluating their impact on the training data. While the algorithm performed exceptionally well during the training period,

it exhibited similar performance to a fixed θ when applied to the test data, suggesting that some of our design choices may not generalize effectively.

For future investigations, it could be valuable to delve into the realm of control theory. We believe this area offers a substantial potential for improvement with relatively modest efforts. One intriguing avenue to explore is the application of model predictive control (MPC), where a model is constructed to predict the risk level instead of relying solely on the most recent realizations.

G.2.3 The 2-Stage Model

We wanted to include a 2-stage DRO model in our thesis; however, we ran into issues regarding its performance which eventually led us to banish it to Appendix K. While the 2-stage formulation remains of great interest to us, and an area we believe has great promise, a number of challenges need to be solved for it to be successful.

Firstly, it is not possible to re-write the 2-stage DRO wind/hydrogen model into a robust form that is – relatively – easily solvable. Therefore, the 2-stage DRO model is computationally slow, and the ambiguity set is limited to a small number of discretization points in its support. In the future, it would be interesting to investigate the possibilities of a suitable wind/hydrogen model that does not include integer variables since this would allow for an efficient reformulation of the 2-stage DRO program that could include continuous support.

Secondly, we faced challenges in determining similar historical days for generating plausible future realizations and estimating the cost of reallocating probability in the DRO program. These issues arose from difficulties in determining dissimilarity in time series data. We chose to use Wasserstein distance as a metric of dissimilarity, and while that may seem like a good idea, it turned out that for our particular problem, it was not. To understand why, let us look at the two simulated time series given in Figure G.1.

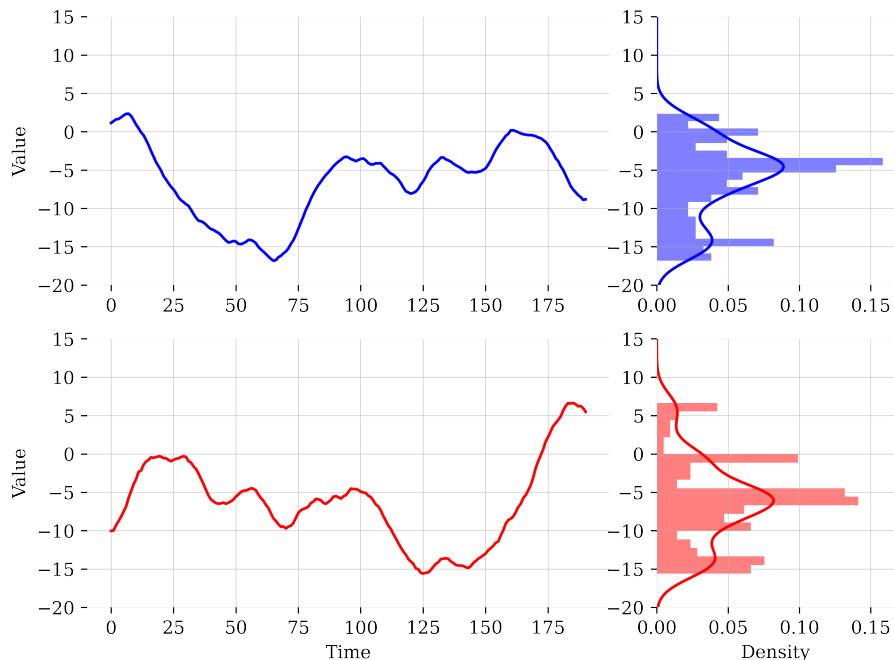
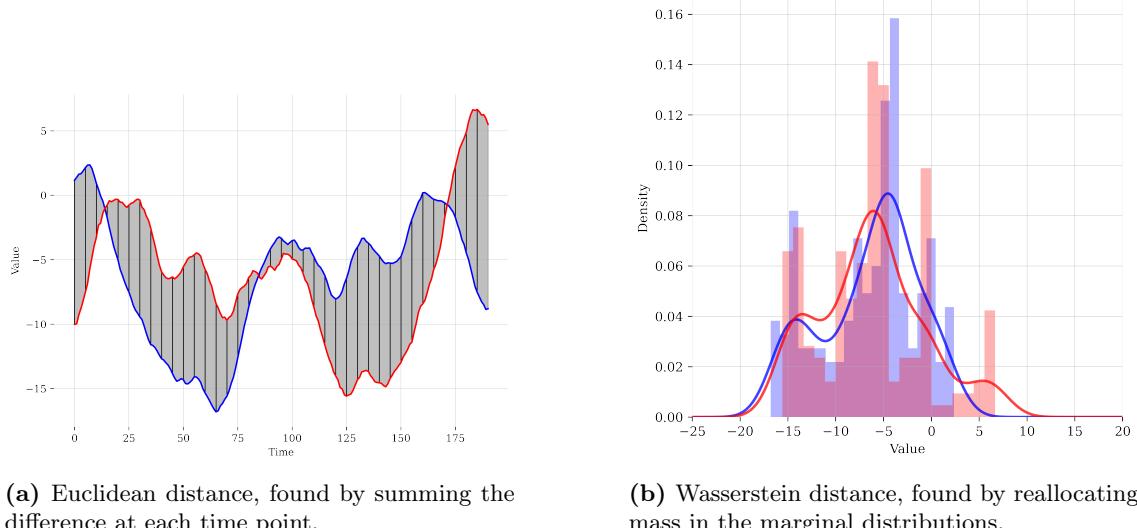


Figure G.1 – Two simulated time series with marginal distributions.

Figure G.2a demonstrates the Euclidean distance between the time series, computed by

summing the differences at each time point. Figure G.2b displays the marginal distributions and the Wasserstein distance, which involves minimal reallocation of probability mass between the distributions. Although the Euclidean distance may appear large in this case, the Wasserstein distance, lacking temporal resolution, seems relatively small. In our wind/hydrogen model, the temporal dimension played a crucial role, rendering the Wasserstein distance unsuitable.



(a) Euclidean distance, found by summing the difference at each time point.

(b) Wasserstein distance, found by reallocating mass in the marginal distributions.

Figure G.2 – Dissimilarity in two time series.

Choosing the Euclidean distance as an alternative is not without its drawbacks, as there are scenarios where it also falls short. Additionally, a naive Euclidian distance views a simple time shift as a dramatic difference and will require huge amounts of data to work, especially for multivariate settings such as ours. Therefore, determining dissimilarity between time series remains an open-ended and problem-specific challenge. To enable the success of the 2-stage DRO model, it is imperative to explore the properties and performance of different dissimilarity metrics within the context of a 2-stage DRO model for wind/hydrogen. A promising avenue to investigate is using "Dynamic Time Warping," as introduced in references such as Tavenard, 2021 or the Fréchet distance.

G.2.4 End-to-End Optimization

A completely different approach to take would be to train a prediction-decision system end to end. This is done by Costa and Iyengar, 2022 where they create an end-to-end distributionally robust system for portfolio construction that integrates the asset return prediction model with a distributionally robust portfolio optimization model. In our case, a similar setup, where the forecasting of wind power and market prices would be integrated with the model for optimal operation of the wind/hydrogen power plant, could be very promising.

However, Costa and Iyengar, 2022 base their model on a methodology from Amos and Kolter, 2017 with a convexity requirement in the original program. The formulation used in this thesis includes integer variables. Therefore, the model needs to be rethought and all integer variables removed before the methodology can be applied.

H | Conclusion

This thesis has shed light on the performance and applicability of Wasserstein DRO models in the context of wind/hydrogen power plants. Furthermore, by addressing the research questions posed at the start of this study, we have deepened the knowledge of the benefits and value offered by these optimization models.

The first research question focused on when Wasserstein DRO is advantageous. We made a clustering algorithm for determining similar historical scenarios to represent future uncertainty. While these distributions did include enough information for DRO to increase performance compared to the deterministic model, we saw that significant improvements were required for SAA to perform better. We believe it is questionable that such improvement of the error selection is possible unless the original forecast is suboptimal.

The second research question focused on the performance of Wasserstein DRO models. Our rigorous analysis and experimentation showed that the DRO models consistently outperformed both the SAA model and the deterministic model. This finding highlights the effectiveness of DRO models in managing uncertainty and optimizing the operation of wind/hydrogen power plants. Furthermore, incorporating distributional robustness provides robust solutions that mitigate the risks associated with the observed distributions, ultimately leading to improved performance in the operation of a wind/hydrogen power plant.

Moving on to the third research question, we investigated the benefits of Wasserstein distributional robustness in the context of wind/hydrogen power plants. Our study revealed that Wasserstein DRO allowed us to adjust the risk appetite explicitly by adjusting the radius of the Wasserstein ball. This enabled us to develop a control algorithm to dynamically adjust the risk to a suitable level in changing markets.

Lastly, we examined the value of having access to electrolysis at a wind power plant, which constituted the fourth research question. As discussed in section F.2, our findings showed that having access to electrolysis resulted in significant reductions in volatility, up to 20 EUR, regardless of the presence of daily contracts. Furthermore, our investigation into returns demonstrated that when daily contracts were in place, electrolysis enabled higher returns when the hydrogen price exceeded 4.8 EUR. Without daily contracts, we observed higher returns when the hydrogen price exceeded 2.0 EUR.

In summary, our research has shown that Wasserstein DRO models perform well in the context of wind/hydrogen power plants. Their benefits lie not only in their ability to provide robust solutions but also in their ability to adjust the risk level. Furthermore, our analysis has indicated that electrolysis at wind power plants offers multiple benefits. Specifically, it has been observed that electrolysis contributes to reduced volatility and, importantly, has the potential to increase returns. The extent to which returns are improved depends on the contractual framework and hydrogen price. These findings contribute to the knowledge of how to use Wasserstein DRO to manage hybrid power plants.

List of Figures

B.1	Efficiency curve for a 52.25MW alkaline electrolyzer working at 90°C and 30 bar and linear approximation thereof. The figure is taken from Baumhof et al., 2023.	7
B.2	Fitted power curve and power curve from the manufacturer. Please notice that random noise has been added to measurements for confidentiality purposes.	8
B.3	An overview of the wind/hydrogen power plant. The figure is taken from figure 2 in Baumhof et al., 2023	9
D.1	Passive realization of the hydrogen production plan for the found day-ahead bid at time -12.	29
D.2	Active realization of the hydrogen production plan for the found day-ahead bid at time -12.	30
E.1	Hours that are available for the sample selection on some given day. Green indicates available while red is not available.	31
E.2	An overview of how we first forecast and then cluster to select relevant samples.	32
E.3	Risk awareness w.r.t. price, θ , chosen by the control algorithm over time.	37
E.4	Accumulated return for active realizations in the test data.	38
E.5	Accumulated return for passive realizations in the test data.	38
E.6	Average daily return and volatility of the different models for varying θ in the test data. θ takes the following values: $\{0, 5, 10, 15, 20, 25, 30\}$.	40
F.1	Empirical distributions for errors in the imbalance forecast determined by our clustering algorithm for a day with and without synthetic improvement to the mean. Dots are chosen historical realizations; the dashed line is the realization for the day in question.	42
F.2	Performance of the deterministic model, the SAA model, the DRO model with $\theta = 22$, and the DRO model with adaptive Wasserstein radius for the price (Control) when the mean of the selected errors is synthetically improved.	43
F.3	Return and volatility for the deterministic model, the SAA model, the DRO model with $\theta = 22$, and the DRO model with adaptive Wasserstein radius for the price. For each point on the outlined curves, the improvement increase by 1%, starting at 0% and ending at 18%.	44
F.4	The average daily return for the test period. The DRO model is optimized with $\theta = 22$. Both with and without daily hydrogen contracts and a scenario without access to an electrolyzer are shown.	45

F.5	The hydrogen production for the DRO model with $\theta = 22$ is calculated using test data. Both with and without daily hydrogen contracts are shown. The size of the daily contracts is also shown.	46
F.6	Active realization for the scenario with daily hydrogen contracts. For each point on the outlined curves, the hydrogen price increase by 0.2 EUR/kg, starting at 0.0 EUR/kg and ending at 8.0 EUR/kg.	47
F.7	Active realization for the scenario without daily hydrogen contracts. For each point on the outlined curves, the hydrogen price increase by 0.2 EUR/kg, starting at 0.0 EUR/kg and ending at 8.0 EUR/kg.	48
G.1	Two simulated time series with marginal distributions.	51
G.2	Dissimilarity in two time series.	52
B.1	Plot of $f(x)$ together with $g(x)$ and $g'(x)$	64
C.1	Illustration of CVaR and VaR.	65
H.1	Illustration of how the mean and medoid of a cluster is selected. The figure is taken from Figure 1 in Entezami et al., 2020.	74
K.1	Accumulated return for active realizations in the test data.	83
L.1	Accumulated returns over time for test data. Rescaled wrt expected active return from the training data.	85
L.2	Accumulated returns over time for test data for SAA.	86
L.3	Accumulated returns over time for test data for regular DRO.	88
L.4	Accumulated returns over time for test data for DRO with correlation.	89
M.1	Passive realization for the scenario with daily hydrogen contracts. For each point on the outlined curves, the hydrogen price increase by 0.2 EUR, starting at 0.0 EUR and ending at 8.0 EUR.	90
M.2	Passive realization for the scenario without daily hydrogen contracts. For each point on the outlined curves, the hydrogen price increase by 0.2 EUR, starting at 0.0 EUR and ending at 8.0 EUR.	90

List of Tables

A.1	All the papers use Wasserstein distributionally robust optimization to model Wind/P2X power plants.	3
D.1	Overview of available data with their meaning in natural language.	27
E.1	Performance metrics for the deterministic model for the training data. Values have been scaled for confidentiality. The deterministic model is used as the benchmark, therefore IR is not defined.	33
E.2	Performance metrics for the deterministic model for the test data. Values have been scaled for confidentiality. The deterministic model is used as the benchmark, therefore IR is not defined.	33
E.3	Performance metrics for the SAA model with optimal ϵ on the training data.	34
E.4	Performance metrics for the SAA model with optimal ϵ on the test data.	34
E.5	Performance metrics for the DRO model with optimal parameters on the training data.	34
E.6	Performance metrics for the DRO model with optimal parameters on the test data.	34
E.7	Performance metrics for the DRO model with correlation with optimal parameters on the training data.	35
E.8	Performance metrics for the DRO model with correlation with optimal parameters on the test data.	35
E.9	Mean values of day-ahead price, imbalance price, wind power, and the standard deviation of the same parameters.	36
E.10	Performance metrics for the DRO model with adaptively chosen Wasserstein radius for the price on the training data. $\epsilon = 0.5$ and $\theta^{\text{CVaR}} = 0$	36
E.11	Performance metrics for the DRO model with adaptively chosen Wasserstein radius for the price on the test data. $\epsilon = 0.5$ and $\theta^{\text{CVaR}} = 0$	37
E.12	Average daily return, \bar{R} , and volatility, $\hat{\sigma}$, for the deterministic model, SAA, DRO, DRO with correlation, and Control (DRO with adaptive Wasserstein radius for price), with percentages relative to the deterministic model.	39
K.1	Performance metrics for the deterministic model for an 8-hour day with an active realization.	82
K.2	Performance metrics for the 2-stage DRO model with varying θ for training data.	82
K.3	Performance metrics for the 2-stage DRO model with varying θ for the test data. Notice that $\theta = 0$ was optimal in the training data.	83

K.4	Average daily return, \bar{R} , and volatility, $\hat{\sigma}$, for the deterministic model, the 1-stage SAA model, the 1-stage DRO model, and the 2-stage DRO model, for an 8-hour day and active realization with percentages relative to the deterministic model.	84
L.1	Performance metrics for SAA with varying ϵ for training data.	85
L.2	Performance metrics for regular DRO with varying θ , $\theta^{\text{CVaR}} = 0$, and $\epsilon = 0.5$ for training data.	86
L.3	Performance metrics for regular DRO with $\theta = 22$, varying θ^{CVaR} , and $\epsilon = 0.5$ for training data.	87
L.4	Performance metrics for regular DRO with $\theta = 22$, $\theta^{\text{CVaR}} = 0$, and varying ϵ for training data.	87
L.5	Performance metrics for regular DRO with correlation with varying θ , $\eta = 1$, $\theta^{\text{CVaR}} = 0$, and $\epsilon = 0.5$ for training data.	88
L.6	Performance metrics for regular DRO with correlation with $\theta = 25$, varying η , $\theta^{\text{CVaR}} = 0$, and $\epsilon = 0.5$ for training data.	89

Bibliography

- Alvarez, J., Ponambalam, K., & Quintana, V. H. (2007). Electricity markets under uncertainty. *2007 IEEE Lausanne Power Tech*, 966–970.
- Amos, B., & Kolter, J. Z. (2017). Optnet: Differentiable optimization as a layer in neural networks. *International Conference on Machine Learning*, 136–145.
- ApS, M. (2023). *The mosek optimization toolbox for matlab manual. version 10.0.* <http://docs.mosek.com/9.0/toolbox/index.html>
- Arrigo, A., Kazempour, J., De Grève, Z., Toubeau, J.-F., & Vallée, F. (2022). Embedding dependencies between wind farms in distributionally robust optimal power flow. *IEEE Transactions on Power Systems*.
- Åström, K., & Hägglund, T. (1995). *Pid controllers*. International Society for Measurement; Control. <https://aiecp.files.wordpress.com/2012/07/1-0-1-k-j-astrom-pid-controllers-theory-design-and-tuning-2ed.pdf>
- Baumhof, M. T., Raheli, E., Johnsen, A. G., & Kazempour, J. (2023). Optimization of hybrid power plants: When is a detailed electrolyzer model necessary? <https://arxiv.org/abs/2301.05310>
- Ben-Tal, A., Ghaoui, L. E., & Nemirovski, A. (2009). *Robust optimization*. Princeton University Press. <https://www2.isye.gatech.edu/~nemirovs/ROBook.pdf>
- Bertsimas, D., Brown, D. B., & Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Review*, 53(3), 464–501. <https://arxiv.org/abs/1010.5445>
- Bertsimas, D., Gupta, V., & Kallus, N. (2016). Robust sample average approximation. <https://arxiv.org/abs/1408.4445>
- Bertsimas, D., Sim, M., & Zhang, M. (2018). Adaptive distributionally robust optimization. *Management Science*, 65. <https://doi.org/10.1287/mnsc.2017.2952>
- Besbes, O., Gur, Y., & Zeevi, A. (2015). Non-stationary stochastic optimization. *Operations research*, 63(5), 1227–1244.
- Birge, J. R., & Louveaux, F. (2011). *Introduction to stochastic programming*. Springer.
- Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press. https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf
- Chen, R., & Paschalidis, I. C. (2020). Distributionally robust learning. *Foundations and Trends® in Optimization*, 4(1-2), 1–243. <https://arxiv.org/pdf/2108.08993.pdf>
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 785–794.
- Chen, Z., Kuhn, D., & Wiesemann, W. (2018). Data-driven chance constrained programs over wasserstein balls. <https://arxiv.org/abs/1809.00210>
- Chen, Z., Kuhn, D., & Wiesemann, W. (2022). Data-driven chance constrained programs over wasserstein balls. *Operations Research*. <https://arxiv.org/abs/1809.00210>

- Coey, C., Lubin, M., & Vielma, J. P. (2020). Outer approximation with conic certificates for mixed-integer convex problems. *Mathematical Programming Computation*, 12(2), 249–293. <https://arxiv.org/pdf/1808.05290.pdf>
- Cornuejols, G., & Tütüncü, R. (2006). *Optimization methods in finance* (Vol. 5). Cambridge University Press. https://web.math.ku.dk/~rolf/CT_FinOpt.pdf
- Costa, G., & Iyengar, G. N. (2022). Distributionally robust end-to-end portfolio construction. *arXiv preprint arXiv:2206.05134*.
- Cplex, I. I. (2012). V12. 8: User's manual for cplex. https://www.ibm.com/docs/en/SSSA5_P_12.8.0/ilog.odms.studio.help/pdf/usrcplex.pdf
- Daneshvar, M., Mohammadi-Ivatloo, B., Zare, K., & Asadi, S. (2020). Two-stage stochastic programming model for optimal scheduling of the wind-thermal-hydropower-pumped storage system considering the flexibility assessment. *Energy*, 193, 116657.
- Danish-Energy-Agency. (2020). Electricity market design: Integrating renewables at low cost. https://ens.dk/sites/ens.dk/files/Globalcooperation/layouted_electricity_market_design_v3.pdf
- Danish-Energy-Agency & Energinet. (2017). Technology data – renewable fuels. https://ens.dk/sites/ens.dk/files/Analyser/version_09_-_technology_data_for_renewable_fuels.pdf
- Delage, E., & Ye, Y. (2010a). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3), 595–612.
- Delage, E., & Ye, Y. (2010b). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58, 595–612. https://www.researchgate.net/publication/220244490_Distributionally_Robust_Optimization_Under_Moment_Uncertainty_with_Application_to_Data-Driven_Problems
- Dong, W., Sun, H., Mei, C., Li, Z., Zhang, J., & Yang, H. (2023). Forecast-driven stochastic optimization scheduling of an energy management system for an isolated hydrogen microgrid. *Energy Conversion and Management*, 277, 116640.
- Eghbali, N., Hakimi, S. M., Hasankhani, A., Derakhshan, G., & Abdi, B. (2022). Stochastic energy management for a renewable energy based microgrid considering battery, hydrogen storage, and demand response. *Sustainable Energy, Grids and Networks*, 30, 100652. <https://doi.org/https://doi.org/10.1016/j.segan.2022.100652>
- Energinet. (2022). Ancillary services to be delivered in denmark. <https://en.energinet.dk/electricity/ancillary-services/access-to-the-ancillary-service-markets/>
- Energinet. (2023). *Status på den grønne omstilling*. <https://energinet.dk/energidata/status-pa-gron-energi/nogletal-om-den-gronne-omstilling/>
- Entezami, A., Sarmadi, H., & Saeedi Razavi, B. (2020). An innovative hybrid strategy for structural health monitoring by modal flexibility and clustering methods. *Journal of Civil Structural Health Monitoring*, 10(5), 845–859.
- Esfahani, P. M., & Kuhn, D. (2015). Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. [http://arxiv.org/abs/1505.05116](https://arxiv.org/abs/1505.05116)
- Fathima, A. H., & Palanisamy, K. (2015). Optimization in microgrids with hybrid energy systems—a review. *Renewable and Sustainable Energy Reviews*, 45, 431–446.
- Gao, R., & Kleywegt, A. J. (2016). Distributionally robust stochastic optimization with wasserstein distance. <https://arxiv.org/abs/1604.02199>
- Gao, R., & Kleywegt, A. J. (2017a). Data-driven robust optimization with known marginal distributions. <https://faculty.mccombs.utexas.edu/rui.gao/copula.pdf>

- Gao, R., & Kleywegt, A. J. (2017b). Distributionally robust stochastic optimization with dependence structure. *arXiv preprint arXiv:1701.04200*. <https://arxiv.org/pdf/1701.04200.pdf>
- Giebel, G., & Kariniotakis, G. (2017). Wind power forecasting—a review of the state of the art. In G. Kariniotakis (Ed.), *Renewable energy forecasting: From models to applications* (pp. 59–109). Woodhead Publishing.
- Grinold, R. C., & Kahn, R. N. (2000). *Active portfolio management*. McGraw Hill New York.
- Guo, J., Zheng, Y., Hu, Z., Zheng, C., Mao, J., Du, K., Jaroniec, M., Qiao, S.-Z., & Ling, T. (2023). Direct seawater electrolysis by adjusting the local reaction environment of a catalyst. <https://doi.org/10.1038/s41560-023-01195-x>
- Gurobi Optimization, LLC. (2023). Gurobi Optimizer Reference Manual. https://www.gurobi.com/wp-content/plugins/hd_documentations/documentation/9.0/refman.pdf
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Vol. 2). Springer.
- Kempler, P., & Nielander, A. (2023). Reliable reporting of faradaic efficiencies for electrocatalysis research. *Nature communications*, 14, 1158. <https://doi.org/10.1038/s41467-023-36880-8>
- Kim, K. (2020). Dual decomposition of two-stage distributionally robust mixed-integer programming under the wasserstein ambiguity set. <https://optimization-online.org/?p=16363>
- Kim, S., Pasupathy, R., & Henderson, S. (2015). A guide to sample average approximation. <https://people.orie.cornell.edu/shane/pubs/SAAGuide.pdf>
- Kleywegt, A. J., Shapiro, A., & Homem-de-Mello, T. (2002). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on optimization*, 12(2), 479–502. <https://bpb-us-w2.wpmucdn.com/sites.gatech.edu/dist/8/1346/files/2020/09/stochdiscreteopt.pdf>
- Kong, F., Mi, J., & Wang, Y. (2022). A two-stage distributionally robust optimization model for optimizing water-hydrogen complementary operation under multiple uncertainties. *Journal of Cleaner Production*, 378, 134538.
- Lange, M., & Focken, U. (2006). *Physical approach to short-term wind power prediction* (1st ed.). Springer, Berlin, Heidelberg. [https://doi.org/https://doi.org/10.1007/3-540-31106-8](https://doi.org/10.1007/3-540-31106-8)
- Madsen, H., & Holst, J. (2006). *Modelling non-linear and non-stationary time series*. Department of Applied Mathematics; Computer Science, Technical University of Denmark. https://www.researchgate.net/profile/Henrik-Madsen-2/publication/233400364_Lecture_notes_for_Advanced_Time_Series_Analysis/links/5bbfbf3a458515a7a9e29a65/Lecture-notes-for-Advanced-Time-Series-Analysis.pdf
- Matoušek, J., & Gärtner, B. (2007). *Understanding and using linear programming* (Vol. 33). Springer. <https://blogs.epfl.ch/extrema/documents/Maison%5C%2020.05.10.pdf>
- Mehrotra, S., & Zhang, H. (2014). Models and algorithms for distributionally robust least squares problems. *Mathematical Programming*, 146, 123–141. <https://optimization-online.org/wp-content/uploads/2011/02/2925.pdf>
- Nielsen, T. S., Madsen, H., Nielsen, H. A., Landberg, L., & Giebel, G. Zephyr - the prediction models. In: In *European wind energy conference 2001, copenhagen*. WIP-Renewable Energies/ETA, 2001. <http://www2.compute.dtu.dk/pubdb/pubs/683-full.html>

- Nikaidô, H., et al. (1954). On von neumann's minimax theorem. *Pacific J. Math*, 4(1), 65–72.
- Nordpool. (2018). Xbid launch information package. <https://www.nordpoolgroup.com/globalassets/download-center/xbid/xbid-launch-information-package.pdf>
- Ortiz, M. M., & Nørstebø, V. (2021). *Business case analysis for hydrogen in european wind farms*. EU. <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5d8de3f5b&appId=PPGMS>
- Pagnoncelli, B. K., Ahmed, S., & Shapiro, A. (2009). Sample average approximation method for chance constrained programming: Theory and applications. *Journal of optimization theory and applications*, 142(2), 399–416. https://www.researchgate.net/publication/225490846_Sample_Average_Approximation_Method_for_Chance_Constrained_Programming_Theory_and_Applications
- Pan, L., Xu, X., Liu, J., & Hu, W. (2023). Adaptive robust scheduling of a hydro/photovoltaic/pumped-storage hybrid system in day-ahead electricity and hydrogen markets. *Sustainable Cities and Society*, 104571.
- Petersen, K. B., & Pedersen, M. S. (2012). The matrix cookbook [Version 20121115]. <http://www2.compute.dtu.dk/pubdb/pubs/3274-full.html>
- Pinson, P., Madsen, H., Nielsen, H. A., Papaefthymiou, G., & Klöckl, B. (2009). From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12(1), 51–62. [https://doi.org/https://doi.org/10.1002/we.284](https://doi.org/10.1002/we.284)
- Pinson, P., & Tastu, J. (2014). Discussion of “prediction intervals for short-term wind farm generation forecasts” and “combined nonparametric prediction intervals for wind power generation”. *IEEE Transactions on Sustainable Energy*, 5, 1019–1020. https://backend.orbit.dtu.dk/ws/files/93684466/Discussion_of_Prediction_intervals.pdf
- Roald, L. A., Pozo, D., Papavasiliou, A., Molzahn, D. K., Kazempour, J., & Conejo, A. (2023). Power systems optimization under uncertainty: A review of methods and applications. *Electric Power Systems Research*, 214, 108725.
- Rockafellar, R. T., Uryasev, S., et al. (2000). Optimization of conditional value-at-risk. *Journal of risk*, 2, 21–42. https://www.ise.ufl.edu/uryasev/files/2011/11/CVaR1_JOR.pdf
- Shapiro, A., & Kleywegt, A. (2002). Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17(3), 523–542. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=d55381d7237c026389b228c77ee047a1637e5573>
- Sharpe, W. F. (1966). Mutual fund performance. *The Journal of business*, 39(1), 119–138.
- Sharpe, W. F. (1975). Adjusting for risk in portfolio performance measurement. *The Journal of Portfolio Management*, 1(2), 29–34.
- Soini, V. (2021). Wind power intermittency and the balancing power market: Evidence from denmark. *Energy Economics*, 100, 105381. <https://doi.org/https://doi.org/10.1016/j.eneco.2021.105381>
- Tavenard, R. (2021). An introduction to dynamic time warping. <https://rtavenar.github.io/blog/dtw.html>
- Tonbari, M. A. E.-M. E. (2021). Decomposition methods in column generation and data-driven stochastic optimization. <http://hdl.handle.net/1853/66134>
- Villani, C. (2003). *Topics in optimal transportation*. American Mathematical Soc. <https://www.math.ucla.edu/~wgangbo/Cedric-Villani.pdf>
- Wang, Y., Kazemi, M., Nojavan, S., & Jermsittiparsert, K. (2020). Robust design of off-grid solar-powered charging station for hydrogen and electric vehicles via robust

- optimization approach. *International Journal of Hydrogen Energy*, 45(38), 18995–19006.
- Wang, Z., Glynn, P., & Ye, Y. (2014). Likelihood robust optimization for data-driven problems. <https://arxiv.org/abs/1307.6279>
- Wiesemann, W., Kuhn, D., & Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, 62, 1358–1376. <https://optimization-online.org/?p=12302>
- Xu, M., Pinson, P., Lu, Z., Qiao, Y., & Min, Y. (2016). Adaptive robust polynomial regression for power curve modeling with application to wind power forecasting. *Wind Energy*, 19(12), 2321–2336. <https://onlinelibrary.wiley.com/doi/epdf/10.1002/we.1985>
- Yang, J., Yang, M., Wang, M., Du, P., & Yu, Y. (2020). A deep reinforcement learning method for managing wind farm uncertainties through energy storage system control and external reserve purchasing. *International Journal of Electrical Power & Energy Systems*, 119, 105928.
- Zheng, Y., Wang, J., You, S., Li, X., Bindner, H. W., & Münster, M. (2023). Data-driven scheme for optimal day-ahead operation of a wind/hydrogen system under multiple uncertainties. *Applied Energy*, 329, 120201.
- Zheng, Y., You, S., Li, X., Bindner, H. W., & Münster, M. (2022). Data-driven robust optimization for optimal scheduling of power to methanol. *Energy Conversion and Management*, 256, 115338.
- Zhu, Z., Weng, Z., & Zheng, H. (2022). Optimal operation of a microgrid with hydrogen storage based on deep reinforcement learning. *Electronics*, 11(2), 196.
- Zymler, S., Kuhn, D., & Rustem, B. (2013). Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137. <https://cyberleninka.org/article/n/264405.pdf>

Appendices

A Relevant Extra Material

- Stochastic Integer Programming
- DRO intuition (Kuhn lecture)
- DRO derivations
- Mixed-Integer Convex Programs
- Integer Linear Programming
- Wasserstein distance
- Trajectory Similarity

B Hypograph Formulation of Concave Function

We want to show that the maximum of the hypograph given by a set of affine functions defined by extended line segments is equivalent to the piece-wise function of the same line segments for concave functions.

Let $f : [a, b] \rightarrow \mathbb{R}$ denote some concave function. Further, let X denote a partition such that

$$X = \{x_1, x_2, \dots\} \quad (\text{B.1a})$$

$$a = x_1 \leq x_2 \leq \dots = b. \quad (\text{B.1b})$$

Let $g : [a, b] \rightarrow \mathbb{R}$ define a piece-wise linear function and be given by

$$g(x) = \left\{ y \mid y = \left(1 - \frac{x - x_n}{x_{n+1} - x_n}\right) f(x_n) + \frac{x - x_n}{x_{n+1} - x_n} f(x_{n+1}), \quad x_n \leq x \leq x_{n+1} \right\} \quad (\text{B.2})$$

that is, the piece-wise function generated by line segments from the surrounding partition points. Let an alternative function, $g' : [a, b] \rightarrow \mathbb{R}$ be defined by

$$g'(x) = \min \left\{ y \mid y = \left(1 - \frac{x - x_n}{x_{n+1} - x_n}\right) f(x_n) + \frac{x - x_n}{x_{n+1} - x_n} f(x_{n+1}), \quad \forall n \in \{1, 2, \dots\} \right\} \quad (\text{B.3})$$

Notice that f can be approximated by the piece-wise linear function given by $\{g_1, g_2, \dots\}$. Now suppose that for some x' we have that $g'(x') < g(x')$, then there must exist a line segment defined outside the surrounding partition of x' for which its extension taken at x' is greater. Since we have made no assumption of the partition, this can be infinitely fine at x' . Therefore it must hold that

$$\exists y \in \{y \mid \alpha f(x_n) + (1 - \alpha) f(x_{n+1}), \quad \alpha \in \mathbb{R}\} \quad (\text{B.4a})$$

$$s.t. \quad x' \notin [x_n, x_{n+1}] \quad \wedge \quad y < f(x'). \quad (\text{B.4b})$$

This contradicts the fact that f is concave. Therefore we conclude that it must hold that $g'(x') \leq g(x')$ in general. Now let us assume that

$$\exists x' \Rightarrow g'(x') > g(x'). \quad (\text{B.5})$$

We immediately notice that this is a clear violation of the definition of g' since it, by definition, is the minimum of all affine functions defined by the extended line segments from the same partition. Hence we conclude that

$$g'(x) = g(x) \quad \forall x \in [a, b]. \quad (\text{B.6})$$

Since $g'(x)$ is defined as the minimum of all extended line segments, it can equally be defined at the maximum of the hypograph of all extended line segments, which is what we wanted to show. Figure B.1 shows a graphical representation of f , g , and g' , additionally, it also shows the hypograph of g in the shaded area.

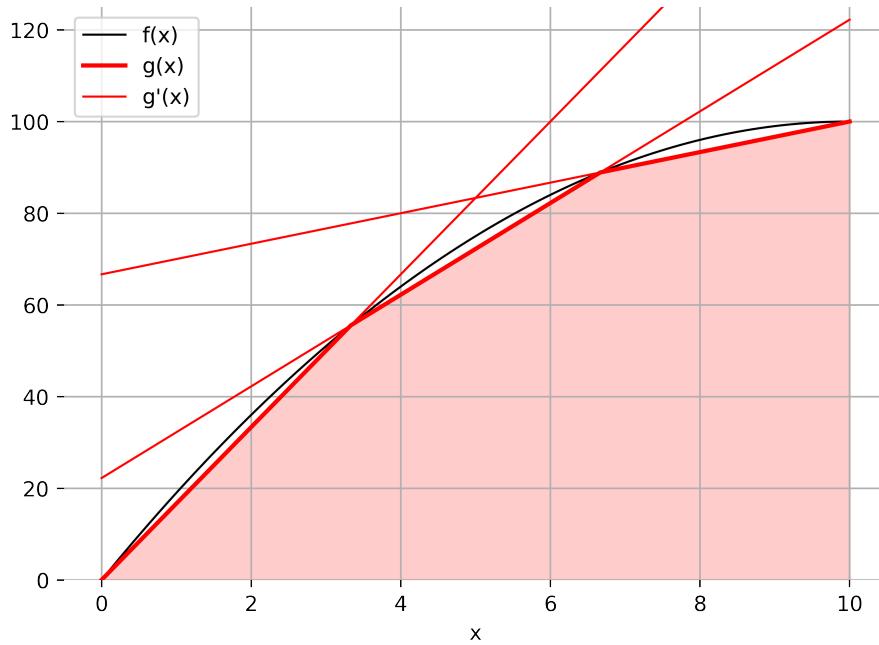


Figure B.1 – Plot of $f(x)$ together with $g(x)$ and $g'(x)$.

C Chance Constraints, VaR and CVaR

Value-at-Risk (VaR) is a measure used to estimate the maximum loss given a certain confidence level, ϵ , considering a decision variable, x , a random variable, ξ , and a loss function, $f(x, \xi)$. Mathematically, VaR is defined as:

$$\text{VaR}_\epsilon(x) := \min\{\tau \in \mathbb{R} : \mathbb{P}[\tau \geq f(x, \xi)] \geq 1 - \epsilon\}. \quad (\text{C.1})$$

To understand the equivalence between VaR and a chance constraint, we can observe that τ serves as an auxiliary variable in the definition of VaR. This allows us to reframe VaR

as a chance constraint by setting τ to some arbitrary number, here we choose $\tau = 0$. The equivalent formulation can be expressed as:

$$\mathbb{P}[0 \geq f(x, \xi)] \geq 1 - \epsilon \Leftrightarrow \quad (\text{C.2})$$

$$\begin{aligned} \min_x & 0 \\ \text{s.t. } & \mathbb{P}[0 \geq f(x, \xi)] \geq 1 - \epsilon \end{aligned} \Leftrightarrow \quad (\text{C.3})$$

$$\min\{0 : \mathbb{P}[0 \geq f(x, \xi)] \geq 1 - \epsilon\} \quad (\text{C.4})$$

Next, we demonstrate that the Conditional Value-at-Risk (CVaR) of a decision variable, x , with respect to a loss function, $f(x, \xi)$, always serves as an upper bound for VaR. To establish this relationship, we utilize the definition of CVaR as provided in Cornuejols and Tütüncü, 2006.

$$\text{CVaR}_\epsilon(x) := \frac{1}{\epsilon} \int_{f(x, \xi) \geq \text{VaR}_\epsilon(x)} f(x, \xi) p(\xi) d\xi \quad (\text{C.5})$$

We can now derive the following bound:

$$\text{CVaR}_\epsilon(x) = \frac{1}{\epsilon} \int_{f(x, \xi) \geq \text{VaR}_\epsilon(x)} f(x, \xi) p(\xi) d\xi \quad (\text{C.6})$$

$$\geq \frac{1}{\epsilon} \int_{f(x, \xi) \geq \text{VaR}_\epsilon(x)} \text{VaR}_\epsilon(x) p(\xi) d\xi \quad (\text{C.7})$$

$$= \frac{\text{VaR}_\epsilon(x)}{\epsilon} \int_{f(x, \xi) \geq \text{VaR}_\epsilon(x)} p(\xi) d\xi \quad (\text{C.8})$$

$$\geq \text{VaR}_\epsilon(x), \quad (\text{C.9})$$

where $p(\xi)$ is the probability density/mass function. We illustrate VaR and CVaR in Figure C.1 to obtain an intuition for why this bound must hold. VaR is the value at the ' $1 - \epsilon$ '-quantile, while CVaR is the expectation for all the probability mass above the ' $1 - \epsilon$ '-quantile.

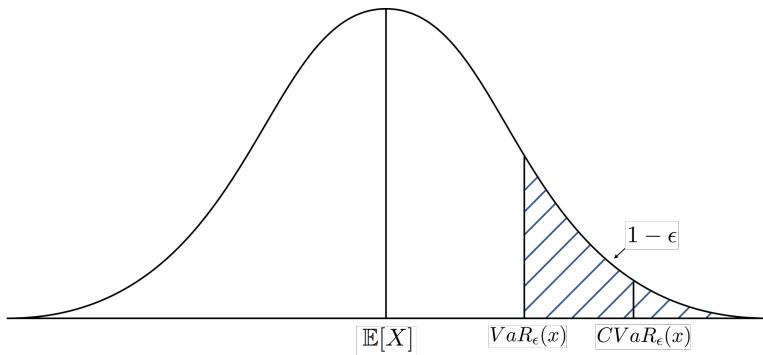


Figure C.1 – Illustration of CVaR and VaR.

D Wasserstein Metric for Two Discrete Measures

We have given the following example to assist the understanding of how the Wasserstein distance between two probability measures works.

Let two discrete spaces be given by X and Y , where each point in the space has equal probability mass. Hence, the measure for X , μ , and the measure for Y , ν , are given by

$$\mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \quad (\text{D.1})$$

$$\nu = \frac{1}{N} \sum_{j=1}^N \delta_{y_j} \quad (\text{D.2})$$

where δ is the Dirac mass. For any copula in the set of all copulas between μ and ν , $\pi \in \Pi(\mu, \nu)$, it holds that

$$\sum_{i=1}^N \pi_{ij} = 1 \quad \forall j \in \{1, \dots, N\} \quad (\text{D.3})$$

$$\sum_{j=1}^N \pi_{ij} = 1 \quad \forall i \in \{1, \dots, N\}. \quad (\text{D.4})$$

In turn, the Wasserstein distance for discrete measures is given by

$$d_W(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \pi_{ij} \|x_i - y_j\|. \quad (\text{D.5})$$

An alternative formulation is given by

$$d_W(\mu, \nu) = \inf_{\sigma} \frac{1}{N} \sum_{i=1}^N \|x_i - y_{\sigma(i)}\| \quad (\text{D.6})$$

where σ is a permutation, that is, a map from $\{1, \dots, N\}$ to $\{1, \dots, N\}$ that reorders the elements. This problem is equivalent to Monge's problem (from transportation theory). Additionally, R. Chen and Paschalidis, 2020 show that the Wasserstein distance can be found by solving the linear program

$$d_W(\mu, \nu) = \min_{\pi} \sum_{i=1}^N \sum_{j=1}^N \pi_{ij} \|x_i - y_j\| \quad (\text{D.7a})$$

$$\text{s.t. } \sum_{i=1}^N \pi_{ij} = \nu_j \quad \forall j \in \{1, \dots, N\} \quad (\text{D.7b})$$

$$\sum_{j=1}^N \pi_{ij} = \mu_i \quad \forall i \in \{1, \dots, N\} \quad (\text{D.7c})$$

$$\pi_{ij} \geq 0 \quad \forall i, j \in \{1, \dots, N\}. \quad (\text{D.7d})$$

E Objective Function for the Wasserstein DRO Wind/Hydrogen Model

To give a thorough understanding of Equation C.4.7, we here give the full derivation. Our starting point is the general formulation,

$$\max_{\mathbb{P} \geq 0} \mathbb{E}^{\mathbb{P}}[f(x, \xi)] \quad (\text{E.1a})$$

$$\text{s.t. } d_W(\mathbb{P}, \mathbb{P}_N) \leq \theta, \quad (\text{E.1b})$$

where $d_W(\mathbb{P}, \mathbb{P}_N)$ is the Wasserstein distance between the empirical distribution and the distribution chosen by nature. To simplify things, we assume nature's distribution, \mathbb{P} , lies in a finite set, Ξ_K , given by:

$$\Xi_K = \{\xi_1, \dots, \xi_K\}. \quad (\text{E.2})$$

Further, we can write out Equation E.1 where we suppress the input x :

$$\max_{\mathbb{P} \geq 0} \sum_{k=1}^K f(\xi_k) \mathbb{P}_k \quad (\text{E.3a})$$

$$\left. \begin{array}{l} \text{s.t. } \min_{\pi \geq 0} \sum_{k=1}^K \sum_{l=1}^K \|\xi_k - \xi_l\| \pi_{kl} \\ \text{s.t. } \sum_{k=1}^K \pi_{kl} = (\mathbb{P}_N)_l \quad \forall l \in \mathcal{K} \\ \sum_{l=1}^K \pi_{kl} = \mathbb{P}_k \quad \forall k \in \mathcal{K} \end{array} \right\} \leq \theta. \quad (\text{E.3b})$$

We know that the probability for each support point in the empirical distribution has probability $\frac{1}{N}$, and all other points have probability 0. Hence we can rewrite Equation E.3 as

$$\max_{\mathbb{P} \geq 0} \sum_{k=1}^K f(\xi_k) \mathbb{P}_k \quad (\text{E.4a})$$

$$\left. \begin{array}{l} \text{s.t. } \min_{\pi \geq 0} \sum_{k=1}^K \sum_{i=1}^N \|\xi_k - \hat{\xi}_i\| \pi_{ki} \\ \text{s.t. } \sum_{k=1}^K \pi_{ki} = \frac{1}{N} \quad \forall i \in \mathcal{N} \\ \sum_{i=1}^N \pi_{ki} = \mathbb{P}_k \quad \forall k \in \mathcal{K} \end{array} \right\} \leq \theta \quad (\text{E.4b})$$

Suppose the minimum value of Equation E.4b is less than or equal to a certain θ . In that case, it is always equivalent to stating that there exists a feasible solution to the optimization problem whose objective value is less than or equal to θ . Therefore, we can eliminate the minimum operator and include π_{ki} as a variable in the outer maximization problem. Additionally, we can replace all instances of \mathbb{P}_k with $\sum_{i=1}^N \pi_{ki}$ due to the last constraint in Equation E.4b:

$$\max_{\pi \geq 0} \sum_{k=1}^K \sum_{i=1}^N f(\xi_k) \pi_{ki} \quad (\text{E.5a})$$

$$\text{s.t. } \sum_{k=1}^K \sum_{i=1}^N \|\xi_k - \hat{\xi}_i\| \pi_{ki} \leq \theta \quad (\text{E.5b})$$

$$\sum_{k=1}^K \pi_{ki} = \frac{1}{N} \quad \forall i \in \mathcal{N}. \quad (\text{E.5c})$$

We see that Equation E.5 has $N+1$ constraints and $K \cdot N$ variables which essentially means ∞ because later we will let $K \rightarrow \infty$. We now take the dual of Equation E.5 where we name the dual variables of Equation E.5b ψ and σ_i for each of the constraints in Equation E.5c.

$$\min_{\psi \geq 0, \sigma_i} \psi \theta + \frac{1}{N} \sum_{i=1}^N \sigma_i \quad (\text{E.6a})$$

$$\text{s.t. } \sigma_i + \psi \|\xi_k - \hat{\xi}_i\| \geq f(\xi_k) \quad \forall i \in \mathcal{N}, \forall k \in \mathcal{K} \quad (\text{E.6b})$$

In Appendix F, we have included a concise explanation aimed at facilitating understanding of the concept of taking the dual of an optimization problem. This explanation is intended to provide readers with an intuitive understanding of the relationship between Equation E.6 and Equation E.5.

We now let $K \rightarrow \infty$ by adding a maximization problem in Equation E.6b.

$$\min_{\psi \geq 0, \sigma_i} \psi \theta + \frac{1}{N} \sum_{i=1}^N \sigma_i \quad (\text{E.7a})$$

$$\text{s.t. } \max_{\xi \in \Xi} f(\xi) - \psi \|\xi - \hat{\xi}_i\| \leq \sigma_i \quad \forall i \in \mathcal{N} \quad (\text{E.7b})$$

To proceed, it is important to introduce the concept of a dual norm and discuss two assumptions. According to Appendix A.1.6 in Boyd and Vandenberghe, 2004, the dual norm of a given norm can be expressed as follows:

$$|v|* = \max |s| \leq 1 v^\top s. \quad (\text{E.8})$$

Furthermore, it is worth noting that the dual of the dual norm is equivalent to the original norm:

$$|v|_{**} = |v|. \quad (\text{E.9})$$

Moving on, the first assumption we make is that the set Ξ can be represented as a polyhedron:

$$\Xi = \{\xi \mid Q\xi \leq h\}. \quad (\text{E.10})$$

The second assumption is that the function f is linear:

$$f(x, \xi) = c(x)^\top \xi. \quad (\text{E.11})$$

We will suppress the input of $c(x)$, x , for the remainder of this section. We can now substitute the norm in Equation E.7b with a maximization problem using Equation E.8 and Equation E.9.

$$\min_{\psi \geq 0, \sigma_i} \psi \theta + \frac{1}{N} \sum_{i=1}^N \sigma_i \quad (\text{E.12a})$$

$$\text{s.t. } \max_{\xi \in \Xi} c^\top \xi - \max_{\|\alpha\|_* \leq \psi} \alpha_i^\top (\xi - \hat{\xi}_i) \leq \sigma_i \quad \forall i \in \mathcal{N} \quad (\text{E.12b})$$

We can again rewrite by the fact that $-max$ is equivalent to min :

$$\min_{\psi \geq 0, \sigma_i} \psi \theta + \frac{1}{N} \sum_{i=1}^N \sigma_i \quad (\text{E.13a})$$

$$\text{s.t. } \max_{\xi \in \Xi} \min_{\|\alpha\|_* \leq \psi} c^\top \xi - \alpha_i^\top (\xi - \hat{\xi}_i) \leq \sigma_i \quad \forall i \in \mathcal{N}. \quad (\text{E.13b})$$

By observing the bilinearity of $c^\top \xi - \alpha_i^\top (\xi - \hat{\xi}_i)$ in terms of α and ξ , we conclude that they form a saddle point. This observation allows us to interchange the order of the "min" and "max" operations, which is commonly referred to as the Von Neumann-Fan minimax theorem, Nikaidô et al., 1954. Consequently, we can eliminate the inner minimization problem and include α as part of the variable set in the outer minimization problem:

$$\min_{\psi \geq 0, \sigma_i, \alpha_i} \psi \theta + \frac{1}{N} \sum_{i=1}^N \sigma_i \quad (\text{E.14a})$$

$$\text{s.t. } \max_{\xi \in \Xi} c^\top \xi - \alpha_i^\top (\xi - \hat{\xi}_i) \leq \sigma_i \quad \forall i \in \mathcal{N} \quad (\text{E.14b})$$

$$\|\alpha_i\|_* \leq \psi \quad \forall i \in \mathcal{N}. \quad (\text{E.14c})$$

We almost have Equation E.14 in a solvable form, but we still need to deal with the inner maximization problem. We write out the problem:

$$\max_{\xi} c^\top \xi - \alpha_i^\top \xi \quad (\text{E.15a})$$

$$\text{s.t. } Q\xi \leq h. \quad (\text{E.15b})$$

We dualize Equation E.15 and name the dual variable of Equation E.15b γ :

$$\min_{\gamma_i \geq 0} h\gamma_i \quad (\text{E.16a})$$

$$\text{s.t. } Q^\top \gamma_i = c^\top - \alpha_i^\top. \quad (\text{E.16b})$$

We insert into Equation E.14 and remove α by substitution using $c^\top - Q^\top \gamma_i = \alpha_i^\top$:

$$\min_{\psi \geq 0, \sigma_i, \gamma_i \geq 0} \psi \theta + \frac{1}{N} \sum_{i=1}^N \sigma_i \quad (\text{E.17a})$$

$$\text{s.t. } c^\top \hat{\xi}_i + \gamma_i(h - Q^\top \hat{\xi}_i) \leq \sigma_i \quad \forall i \in \mathcal{N} \quad (\text{E.17b})$$

$$\|c^\top - Q^\top \gamma_i\|_* \leq \psi \quad \forall i \in \mathcal{N}. \quad (\text{E.17c})$$

We see that this is the same as given in Corollary 5.1 in Esfahani and Kuhn, 2015.

F Dual of a Linear Program

Numerous resources, such as chapter 5 of Boyd and Vandenberghe, 2004, explain how to determine the dual of a mathematical program in detail. Therefore, in this section, we aim at providing readers with an intuitive understanding of the concept of the dual program. We will utilize problem Equation F.1 from Matoušek and Gärtner, 2007 to achieve this:

$$\max_{x_1, x_2} 2x_1 + 3x_2 \quad (\text{F.1a})$$

$$\text{s.t. } 4x_1 + 8x_2 \leq 12 \quad (\text{F.1b})$$

$$2x_1 + x_2 \leq 3 \quad (\text{F.1c})$$

$$3x_1 + 2x_2 \leq 4 \quad (\text{F.1d})$$

$$x_1, x_2 \in \mathbb{R}_+. \quad (\text{F.1e})$$

Instead of solving Equation F.1, we want to derive an upper bound of the form:

$$d_1 x_1 + d_2 x_2 \leq h, \quad (\text{F.2})$$

where $d_1 \geq 2$ and $d_2 \geq 3$. To find such a bound, we can divide Equation F.7b by two and obtain the following bound:

$$2x_1 + 4x_2 \leq 6. \quad (\text{F.3})$$

We can obtain an even better bound by adding Equation F.7b and Equation F.1c and then divide by 3,

$$\frac{1}{3} (4x_1 + 8x_2 + 2x_1 + x_2) \leq \frac{1}{3} (12 + 3) = 5. \quad (\text{F.4})$$

Instead of relying on random guessing to determine the optimal boundary, we can approach the task more methodically. We combine the constraints of Equation F.1 as follows

$$y_1(4x_1 + 8x_2) + y_2(2x_1 + x_2) + y_3(3x_1 + 2x_2) \leq 12y_1 + 3y_2 + 4y_3 \Rightarrow \quad (\text{F.5})$$

$$\underbrace{(4y_1 + 2y_2 + 3y_3)}_{d_1} x_1 + \underbrace{(8y_1 + y_2 + 2y_3)}_{d_2} x_2 \leq \underbrace{12y_1 + 3y_2 + 4y_3}_{h} \quad (\text{F.6})$$

Constructing a linear program that identifies the best effective combination of constraints that minimizes the upper bound of the objective function is now possible:

$$\max_{y_1, y_2, y_3} \quad h = 12y_1 + 3y_2 + 4y_3 \quad (\text{F.7a})$$

$$\text{s.t.} \quad d_1 = 4y_1 + 2y_2 + 3y_3 \geq 2 \quad (\text{F.7b})$$

$$d_2 = 8y_1 + y_2 + 2y_3 \geq 3 \quad (\text{F.7c})$$

$$y_1, y_2, y_3 \in \mathbb{R}_+. \quad (\text{F.7d})$$

To prevent the \leq -constraints of Equation F.1 from being reversed, it is necessary for y_1 , y_2 , and y_3 to be non-negative. The dual program, Equation F.7, is a complementary approach to the primal program. While the primal program seeks to optimize the objective function by finding the best combination of variables, the dual program aims to optimize the same objective function by identifying the best combination of constraints. It is crucial to remember that the values of the objective function for both programs will be equivalent only if strong duality is satisfied. It should be noted that strong duality does not hold in general for convex programs. Nevertheless, for the programs addressed in this thesis, we can rely on the assurance that strong duality holds.

G Objective Function for the Wasserstein DRO with Correlation Wind/Hydrogen Model

To give a thorough understanding of Equation C.4.20, we here provide the full derivation. We begin by applying Theorem 4.4 in Esfahani and Kuhn, 2015 and Remark 1 in Gao and Kleywegt, 2017b to obtain a finite program:

$$\max_{\mathbb{P} \in \mathcal{A}_{\Sigma}} \mathbb{E}^{\mathbb{P}}[f(\xi)] = \max_{\xi_i \in \Xi} \quad \frac{1}{N} \sum_{i=1}^N c^T \xi_i \quad (\text{G.1a})$$

$$\text{s.t.} \quad \frac{1}{N} \sum_{i=1}^N \|\xi_i - \hat{\xi}_i\| \leq \theta \quad (\text{G.1b})$$

$$\frac{1}{N} \sum_{i=1}^N (\xi_i - \hat{\mu}) (\xi_i - \hat{\mu})^T \preceq \hat{\Sigma}. \quad (\text{G.1c})$$

To express Equation G.1 as a minimization problem, we start by constructing the Lagrangian:

$$\begin{aligned} \mathcal{L}(\xi_i, \psi, \Psi) &= \frac{1}{N} \sum_{i=1}^N c^T \xi_i + \psi \left(\theta - \frac{1}{N} \sum_{i=1}^N \|\xi_i - \hat{\xi}_i\| \right) + \langle \Psi, \hat{\Sigma} - \frac{1}{N} \sum_{i=1}^N (\xi_i - \hat{\mu}) (\xi_i - \hat{\mu})^T \rangle_F \\ &\quad (\text{G.2}) \end{aligned}$$

$$= \psi \theta + \langle \Psi, \hat{\Sigma} \rangle_F + \frac{1}{N} \sum_{i=1}^N \left(c^T \xi_i - \psi \|\xi_i - \hat{\xi}_i\| - (\xi_i - \hat{\mu}) \Psi (\xi_i - \hat{\mu})^T \right) \quad (\text{G.3})$$

Here, $\langle \cdot, \cdot \rangle_F$ represents the Frobenius inner product. We then derive the Lagrange dual function:

$$g(\psi, \Psi) = \psi\theta + \langle \Psi, \hat{\Sigma} \rangle_F + \frac{1}{N} \sum_{i=1}^N \max_{\xi_i \in \Xi} \left(c^\top \xi_i - \psi \|\xi_i - \hat{\xi}_i\| - (\xi_i - \hat{\mu}) \Psi (\xi_i - \hat{\mu})^\top \right). \quad (\text{G.4})$$

We again use the dual norm given in Equation E.8 and Equation E.9 to get rid of the norm in Equation G.4.

$$g(\psi, \Psi) = \psi\theta + \langle \Psi, \hat{\Sigma} \rangle_F + \frac{1}{N} \sum_{i=1}^N \min_{\|\alpha\|_* \leq \psi} \max_{\xi_i \in \Xi} \left(c^\top \xi_i - \alpha^\top (\xi_i - \hat{\xi}_i) - (\xi_i - \hat{\mu}) \Psi (\xi_i - \hat{\mu})^\top \right). \quad (\text{G.5})$$

We can write up the dual program using section 5.3 in Boyd and Vandenberghe, 2004:

$$\min_{\psi \geq 0, \Psi \succeq 0, \sigma_i, \alpha_i} \psi\theta + \langle \Psi, \hat{\Sigma} \rangle_F + \frac{1}{N} \sum_{i=1}^N \sigma_i \quad (\text{G.6a})$$

$$\text{s.t. } \sigma_i \geq \max_{\xi_i \in \Xi} c^\top \xi_i - \alpha^\top (\xi_i - \hat{\xi}_i) - (\xi_i - \hat{\mu}) \Psi (\xi_i - \hat{\mu})^\top \quad \forall i \in \mathcal{N} \quad (\text{G.6b})$$

$$\|\alpha_i\|_* \leq \psi \quad \forall i \in \mathcal{N}. \quad (\text{G.6c})$$

To solve Equation G.6, we must first reformulate the constraint, Equation G.6b. Assuming that $\xi \in \mathbb{R}$, we can differentiate with respect to ξ_i , set the derivative equal to zero, and solve for ξ_i :

$$\frac{d}{d\xi_i} (c^\top \xi_i - \alpha^\top (\xi_i - \hat{\xi}_i) - (\xi_i - \hat{\mu}) \Psi (\xi_i - \hat{\mu})^\top) = 0 \quad (\text{G.7})$$

$$c^\top - \alpha^\top - 2(\xi_i - \hat{\mu}) \Psi = 0 \quad (\text{G.8})$$

$$\xi_i = -\frac{1}{2}(\alpha^\top - c^\top - 2\hat{\mu}\Psi)\Psi^{-1}. \quad (\text{G.9})$$

For the above, it is worth noting that Ψ^{-1} is not guaranteed to exist because Ψ is only restricted to lie in the positive semidefinite cone and not the positive definite cone. However, if the inverse does not exist, one can just use the generalized inverse. We now reinsert into Equation G.6b, and reorganize the variables:

$$-\frac{1}{4}(c - \alpha_i + 2\Psi\hat{\mu})\Psi^{-1}(c - \alpha_i + 2\Psi\hat{\mu})^\top \leq -\alpha_i^\top \hat{\xi}_i + \hat{\mu}\Psi\hat{\mu}^\top + \sigma_i. \quad (\text{G.10})$$

From Petersen and Pedersen, 2012 we know that the Schur complement is given by,

$$b^\top A^{-1} b \leq c \iff \begin{bmatrix} A & b \\ b^\top & c \end{bmatrix} \succeq 0, \quad (\text{G.11})$$

and hence we can rewrite Equation G.10 as

$$\begin{bmatrix} \Psi & -\frac{1}{2}(c - \alpha_i + 2\Psi\hat{\mu}) \\ -\frac{1}{2}(c - \alpha_i + 2\Psi\hat{\mu})^\top & -\alpha_i^\top \hat{\xi}_i + \hat{\mu}\Psi\hat{\mu}^\top + \sigma_i \end{bmatrix} \succeq 0 \quad (\text{G.12})$$

We can now reinsert Equation G.12 in Equation G.6 and arrive at Equation G.13 which is also given in the main text as Equation C.4.20. We see that it corresponds to what is given in Collary 1 in Gao and Kleywegt, 2017b:

$$\min_{\psi \geq 0, \Psi \succeq 0, \sigma_i, \alpha_i} \psi\theta + \langle \Psi, \hat{\Sigma} \rangle_F + \frac{1}{N} \sum_{i=1}^N \sigma_i \quad (\text{G.13a})$$

$$\text{s.t. } \begin{bmatrix} \Psi & -\frac{1}{2}(c - \alpha_i + 2\Psi\hat{\mu}) \\ -\frac{1}{2}(c - \alpha_i + 2\Psi\hat{\mu})^\top & -\alpha_i^\top \hat{\xi}_i + \hat{\mu}\Psi\hat{\mu}^\top + \sigma_i \end{bmatrix} \succeq 0 \quad \forall i \in \mathcal{N} \quad (\text{G.13b})$$

$$\|\alpha_i\|_* \leq \psi \quad \forall i \in \mathcal{N}. \quad (\text{G.13c})$$

It is now easy to see that if we want to add a trust parameter, η , as discussed in Equation C.4.3, we just add it to the objective function and get:

$$\min_{\psi \geq 0, \Psi \succeq 0, \sigma_i, \alpha_i} \psi\theta + \langle \Psi, \eta\hat{\Sigma} \rangle_F + \frac{1}{N} \sum_{i=1}^N \sigma_i \quad (\text{G.14a})$$

$$\text{s.t. } \begin{bmatrix} \Psi & -\frac{1}{2}(c - \alpha_i + 2\Psi\hat{\mu}) \\ -\frac{1}{2}(c - \alpha_i + 2\Psi\hat{\mu})^\top & -\alpha_i^\top \hat{\xi}_i + \hat{\mu}\Psi\hat{\mu}^\top + \sigma_i \end{bmatrix} \succeq 0 \quad \forall i \in \mathcal{N} \quad (\text{G.14b})$$

$$\|\alpha_i\|_* \leq \psi \quad \forall i \in \mathcal{N}. \quad (\text{G.14c})$$

H K-Medoids

K-medoids is an algorithm that partitions a dataset into K distinct clusters, where each cluster is represented by a medoid, which is the data point closest to the center of the cluster. The algorithm aims to minimize the dissimilarity between data points within a cluster and maximize the dissimilarity between clusters. Based on Hastie et al., 2009, the steps of the algorithm can be summarized as follows:

1. Randomly select K data points from the dataset to be the initial medoids.
2. Assign each data point to its closest medoid based on a chosen distance metric, such as Euclidean distance. Let c_i denote the cluster assignment of data point i , and let m_k denote the medoid of cluster k .
3. For each cluster k , calculate the total dissimilarity between each data point i in the cluster and its medoid m_k , defined as:

$$D_k = \sum_{i \in k} d(i, m_k)$$

where $d(i, j)$ is the chosen distance metric between data points i and j .

4. For each cluster, k , select a non-medoid data point, i , in the cluster that minimizes the total dissimilarity to all other points in the same cluster:

$$i^* = \operatorname{argmin}(D_k - d(i, m_k) + d(i, i'))$$

where i' is a non-medoid data point in cluster k .

5. Swap the current medoid, m_k , with the selected data point, i^* , and update the cluster assignment, c_i , for all data points affected by the swap.
6. Repeat steps 2 to 5 until the medoids no longer change or a maximum number of iterations is reached.

In Figure H.1, the optimal mean and medoid of respectively a K-means and a K-medoids algorithm is shown.

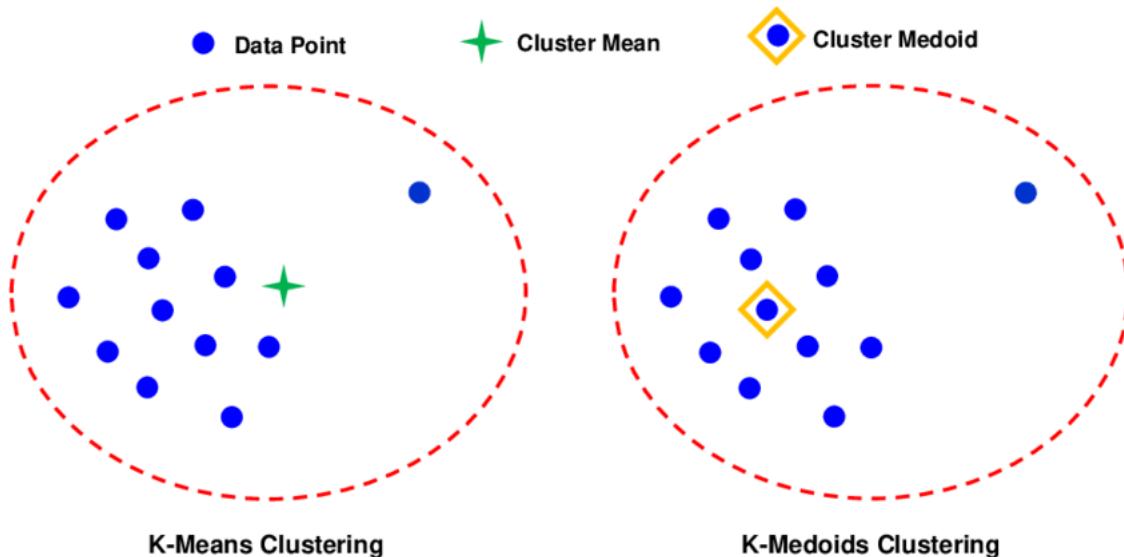


Figure H.1 – Illustration of how the mean and medoid of a cluster is selected. The figure is taken from Figure 1 in Entezami et al., 2020.

I Pseudo Code for the Sample Selection Clustering

The following pseudo-code outlines the implementation steps of the clustering algorithm used in this study. This algorithm employs a combination of the XGBoost model for forecasting and the K-medoids algorithm for clustering. The XGBoost model is trained on historical data to predict the imbalance prices, which are then used as input variables in the K-medoids algorithm to cluster the forecasts into similar groups. It is important to note that the hyperparameters for both models have been manually adjusted and evaluated solely on the training data. Please refer to the provided pseudo code for a detailed overview of the process involved in training the XGBoost model, adjusting the imbalance prices, and incorporating the predicted imbalance prices into the clustering process.

Algorithm 1 Clustering and Forecasting Algorithm

Require: Data: $clustering_data$, $forecast_data$

Ensure: Clusters: $clusters$

```

1:  $clustering\_data \leftarrow \text{standardize\_cols}([\text{adjusted\_balancing\_price}, \text{power\_forecast},$ 
    $\text{price\_forecast}, \text{load\_forecast}])$ 
2:  $forecast\_data \leftarrow \text{standardize\_cols}([\text{adjusted\_balancing\_price}, \text{power\_forecast},$ 
    $\text{price\_forecast}, \text{load\_forecast}, \text{scheduled\_flow}, \text{wind\_speed\_forecast}])$ 
3:  $num\_shifts \leftarrow 8$ ,  $horizon \leftarrow 36$ ,  $end\_date \leftarrow 303$ ,  $start\_date \leftarrow 42$ ,  $no\_clusters \leftarrow 20$ 
4:  $end\_date\_adj \leftarrow end\_date - start\_date$ 
5:  $clusters \leftarrow \text{alloc}(end\_date\_adj)$ 
6:  $j \leftarrow 1$ 
7: for  $i \leftarrow start\_date$  to  $end\_date$  do
8:    $clustering\_data\_i \leftarrow clustering\_data[:, 1 : (i + 1) * 24]$ ,  $forecast\_data\_i \leftarrow forecast\_data[:, 1 : (i + 1) * 24]$ 
9:    $features, targets \leftarrow \text{split\_dataset}(forecast\_data\_i)$ 
10:   $shifted\_features, shifted\_targets \leftarrow \text{shift\_dataset}(features, targets, num\_shifts)$ 
11:   $features\_train, targets\_train \leftarrow \text{train\_split}(shifted\_features, shifted\_targets, 1 : end\_horizon)$ 
12:   $features\_test, targets\_test \leftarrow \text{test\_split}(shifted\_features, shifted\_targets, end\_horizon + 1 : end)$ 
13:  if  $i = start\_date$  then
14:     $model \leftarrow \text{XGBoost}(features\_train, num\_round = 50, label = targets\_train, eta$ 
    $= 0.05, max\_depth = 12)$ 
15:  else
16:     $\text{XGBoost.update}(model, features\_train, num\_round = 20, label = tar$ 
    $gets\_train)$ 
17:  end if
18:   $balancing\_prices \leftarrow \text{XGBoost.predict}(model, features\_test)$ 
19:   $clustering\_data\_i[1, (i + 1) * 24 - horizon + 1 : (i + 1) * 24] .= balancing\_prices$ 
20:   $distanceMatrix \leftarrow \text{Distances.pairwise}(\text{Distances.SqEuclidean}(), clustering\_data\_i)$ 
21:   $clusters[j] \leftarrow \text{kmedoids}(distanceMatrix, no\_clusters; maxiter = 200)$ 
22:   $j \leftarrow j + 1$ 
23: end for
24: return  $clusters$ 
```

J A PI Controller

A PI (Proportional-Integral) controller is a control mechanism commonly used in optimization algorithms to adjust a parameter, θ , iteratively, Åström and Hägglund, 1995. It aims to optimize the algorithm's performance by dynamically updating θ based on some metric; here the return. The controller consists of two components: the proportional component, which adjusts θ based on the current return, and the integral component, which accounts for the historical behavior of the return. By continuously monitoring and adapting θ , the PI controller seeks to find an optimal value that maximizes the return. The following algorithm outlines the steps in implementing a PI controller to control the parameter, θ .

Algorithm 2 PI Controller Algorithm

```

1: Parameters:
2:  $\theta_0 = 0.0$ 
3:  $\theta_\Delta = 0.22$ 
4:  $\theta_{\text{search}} = 1.0$ 
5:  $\text{momentum}_{\text{max}} = 4.0$ 
6:  $\text{momentum}_u = \text{momentum}_{\text{max}}$ 
7:  $\text{momentum}_d = \text{momentum}_{\text{max}}$ 
8: for  $block = t1$  to  $t2$  do
9:    $\theta_{\text{candidates}} \leftarrow [\max(\theta_i - \theta_{\text{search}}, 0), \theta_i, \theta_i + \theta_{\text{search}}, \max(\theta_i - 5\theta_{\text{search}}, 0), \theta_i + 5\theta_{\text{search}}]$ 
10:  for  $k = 1$  to  $5$  do
11:     $\theta \leftarrow \theta_{\text{candidates}}[k]$ 
12:     $\text{obj}[k] \leftarrow f(\theta)$ 
13:  end for
14:  if  $\text{obj}[2] = \max(\text{obj})$  then
15:     $\theta_N \leftarrow \theta_{\text{candidates}}[2]$ 
16:     $\text{momentum}_u \leftarrow \max(1, \text{momentum}_u - 1)$ 
17:     $\text{momentum}_d \leftarrow \max(1, \text{momentum}_d - 1)$ 
18:  else
19:    if  $\arg \max(\text{obj}) = 1 \vee \arg \max(\text{obj}) = 5$  then
20:      if  $\arg \max(\text{obj}) = 1$  then
21:         $\theta_N \leftarrow \max(0, \theta_N - \theta_\Delta \times \text{momentum}_d)$ 
22:      else
23:         $\theta_N \leftarrow \max(0, \theta_N - \theta_\Delta \times \text{momentum}_d^{2.3})$ 
24:      end if
25:       $\text{momentum}_d \leftarrow \min(\text{momentum}_{\text{max}}, \text{momentum}_d + 1)$ 
26:       $\text{momentum}_u \leftarrow \max(1, \text{momentum}_u - 1)$ 
27:    else
28:      if  $\arg \max(\text{obj}) = 3$  then
29:         $\theta_N \leftarrow \theta_N + \theta_\Delta \times \text{momentum}_u$ 
30:      else
31:         $\theta_N \leftarrow \theta_N + \theta_\Delta \times \text{momentum}_u^{2.3}$ 
32:      end if
33:       $\text{momentum}_d \leftarrow \max(1, \text{momentum}_d - 1)$ 
34:       $\text{momentum}_u \leftarrow \min(\text{momentum}_{\text{max}}, \text{momentum}_u + 1)$ 
35:    end if
36:  end if
37: end for

```

K 2-Stage Models

K.1 2-stage Program Theory

In this appendix, we will explore what is referred to as *multi-stage* optimization. Generally speaking, multistage programs grow quite dramatically in complexity when the number of stages grows, therefore, we will in the following only consider a program for which decisions are made at two instances, that is, a two stage program.

A two stage program is given on the generic form

$$\min_x \quad f(x) + \mathbb{E}^{\mathbb{P}}[Q(x, \xi)] \quad (\text{K.1a})$$

$$\text{s.t.} \quad g(x) \leq 0 \quad (\text{K.1b})$$

where the inner program, Q , is given by

$$Q(x, \xi) = \min_y \quad \phi(x, y, \xi) \quad (\text{K.2a})$$

$$\text{s.t.} \quad \gamma(x, y, \xi) \leq 0. \quad (\text{K.2b})$$

We call x the *first stage variables* and y the *second stage variables*, additionally, we may refer to the first stage program as the *outer* program and the second stage program as the *inner* program. A problem, where there is some potential for alleviating some of the consequences or – equivalently – improving the objective by making a change in a second stage, we call a *recourse* program. Where the *recourse* is the decision made in the second stage. Since the completely generic problem can be very difficult to solve we restrict our attention to problems that are linear in the first stage.

There exist many possibilities when it comes to solving two stage stochastic programs. These include (but are not limited to) scenario generation, two stage robust optimization, and sample average approximation. If we go through them one by one, the idea of scenario generation is to produce probable future realizations and optimize over their expectation under an appropriately selected measure. For two stage robust optimization, the engineer expects the worst-case second stage realization selected within some predefined sample space. Finally, for sample average approximation the engineer expects the future behaviour of the second stage realizations to be well explained by historic realizations. This leads to an optimization problem over the expectation of selected historic realizations under the empirical measure. For further information about two stage stochastic programs in general see, e.g., Birge and Louveaux, 2011, S. Kim et al., 2015.

2-stage Sample Average Approximation

The first approach we consider is the sample average approximation. This is given by

$$\min_x \quad c^\top x + \mathbb{E}^{\mathbb{P}}[Q(x, \xi)] \quad (\text{K.3a})$$

$$\text{s.t.} \quad Ax \leq b \quad (\text{K.3b})$$

where \mathbb{P} is the empirical measure. For our particular model, this corresponds to the program:

$$\min_x \quad - \left(\sum_{t \in \mathcal{T}} p_t^{\text{DA}} \hat{\lambda}_t^{\text{DA}} - p_t^{\text{DA}-} \lambda^{\text{TSO}} \right) + \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} Q(x, \xi_s) \quad (\text{K.4a})$$

$$\text{s.t.} \quad p_t^{\text{DA}} = p_t^{\text{DA}+} - p_t^{\text{DA}-} \quad t \in \mathcal{T} \quad (\text{K.4b})$$

$$p_t^{\text{DA}+}, p_t^{\text{DA}-} \geq 0 \quad t \in \mathcal{T}. \quad (\text{K.4c})$$

where \mathcal{S} is some predetermined set of historical realizations and x is the set of 1-stage variables given by:

$$x = \left\{ p_t^{\text{DA}}, p_t^{\text{DA}+}, p_t^{\text{DA}-} \right\}. \quad (\text{K.5})$$

For the inner program, Q , we define the set of constraints, \mathcal{Y} , to be all the constraints in the set \mathcal{X} just without the 1-stage constraints given in Equation K.4. The inner problem is hence given by:

$$\min_y - \left(\sum_{t \in \mathcal{T}} p_t^{\text{DA}} \xi_t^{\text{DA}} + p_t^B (\hat{\lambda}_t^B + \xi_t^B) - p_t^{\text{B}-} \lambda^{\text{TSO}} + d_t \lambda^h - z_t^{\text{su}} \lambda^{\text{su}} \right) \quad (\text{K.6a})$$

$$\text{s.t. } \hat{P}_t^w + \xi_t^w = p_t^{\text{DA}} + p_t^B + p_t^e + p_t^c \quad t \in \mathcal{T} \quad (\text{K.6b})$$

$$p_t^B = p_t^{\text{B}+} - p_t^{\text{B}-} \quad t \in \mathcal{T} \quad (\text{K.6c})$$

$$p_t^{\text{B}+}, p_t^{\text{B}-} \geq 0 \quad t \in \mathcal{T} \quad (\text{K.6d})$$

$$y_s \in \mathcal{Y} \quad s \in \mathcal{S}. \quad (\text{K.6e})$$

where y is the set of second-stage variables given by all the variables related to the imbalance price, wind, and hydrogen. We can combine these two, and get the SAA model on complete form:

$$\min_{x,y} - \left(\sum_{t \in \mathcal{T}} p_t^{\text{DA}} \hat{\lambda}_t^{\text{DA}} - p_t^{\text{DA}-} \lambda^{\text{TSO}} \right) + \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} Q_s \quad (\text{K.7a})$$

$$\text{s.t. } Q_s = - \sum_{t \in \mathcal{T}} \begin{pmatrix} p_t^{\text{DA}} \xi_{s,t}^{\text{DA}} + p_{s,t}^B (\hat{\lambda}_t^B + \xi_{s,t}^B) & \dots \\ -p_{s,t}^{\text{B}-} \lambda^{\text{TSO}} + d_{s,t} \lambda^h - z_{s,t}^{\text{su}} \lambda^{\text{su}} \end{pmatrix} \quad s \in \mathcal{S} \quad (\text{K.7b})$$

$$\hat{P}_t^w + \xi_{s,t}^w = p_t^{\text{DA}} + p_{s,t}^B + p_{s,t}^e + p_{s,t}^c \quad t \in \mathcal{T}, s \in \mathcal{S} \quad (\text{K.7c})$$

$$p_t^{\text{DA}+}, p_t^{\text{DA}-} \geq 0 \quad t \in \mathcal{T} \quad (\text{K.7d})$$

$$p_{s,t}^{\text{B}+}, p_{s,t}^{\text{B}-} \geq 0 \quad t \in \mathcal{T}, s \in \mathcal{S} \quad (\text{K.7e})$$

$$y_s \in \mathcal{Y} \quad s \in \mathcal{S}. \quad (\text{K.7f})$$

Bertsimas et al., 2016 discuss the use of SAA extensively. One advantage of SAA is that it is guaranteed to converge towards optimum asymptotically. In other words, for a sufficiently large sample size, the SAA deliveres excellent results under the assumption of stationarity in the underlying distributions. However, it can be very difficult to estimate how large "sufficiently" large is. Therefore, it is also difficult to estimate the cost of realization, which in many test cases is seen to be substantial. This leads to a natural desire to be robust towards small changes in the underlying distribution, that is, small changes to the empirical measure.

2-stage Distributionally Robust Optimization

In two stage DRO, we once again consider a type of stochastic optimization made over the worst-case measure within some θ -ball of the empirical distribution. We formulate this as:

$$\min_x c^\top x + \max_{\mathbb{P} \in \mathcal{A}} \mathbb{E}^{\mathbb{P}}[Q(x, \xi)] \quad (\text{K.8a})$$

$$\text{s.t. } Ax \leq b. \quad (\text{K.8b})$$

We could be tempted to just select the ambiguity set, \mathcal{A} , as the one given in Equation C.4.5 but we are now working with time series and therefore need to be careful what metric we use. We propose the ambiguity set:

$$\mathcal{A}_{2s} = \left\{ \mathbb{P} \in \mathcal{M}(\Xi) \mid d_{\text{W}}^{(W_1)}(\mathbb{P}, \mathbb{P}_N) \leq \theta \right\}. \quad (\text{K.9})$$

Notice that d now represents a distance between two multivariate stochastic processes. We have chosen to use the Wasserstein distance with the 1-norm as the metric within our Wasserstein ball for the DRO program.

As we did for one stage DRO, we start by handling the inner maximization problem $\max_{\mathbb{P} \in \mathcal{A}} \mathbb{E}^{\mathbb{P}}[Q(x, \xi)]$. Since we use the Wasserstein metric, this can be defined by

$$\max_{\mathbb{P} \in \mathcal{A}_{2s}} \sum_{s \in S} \mathbb{P}_s Q(x, \xi_s) \quad (\text{K.10a})$$

$$\begin{aligned} \min_{\pi \geq 0} & \quad \sum_{s \in S} \sum_{n \in \mathcal{N}} d_{\text{W}}^{(1)}(\xi_s, \hat{\xi}_n) \pi_{sn} \\ \text{s.t.} & \quad \begin{aligned} & \sum_{s \in S} \pi_{sn} = \frac{1}{N}, \quad \forall n \in \mathcal{N} \\ & \sum_{n \in \mathcal{N}} \pi_{sn} = \mathbb{P}_s, \quad \forall s \in S \end{aligned} \end{aligned} \quad (\text{K.10b})$$

where \mathcal{N} is a set of historical realization, \mathcal{S} is a set of historical and synthetic (generated) realizations, and θ is the chosen Wasserstein radius. Since the inner program is a minimization program and the constraints is a *less than or equal to*, the program is equivalent to:

$$\max_{\pi \geq 0} \sum_{s \in S} \sum_{n \in \mathcal{N}} \pi_{sn} Q(x, \xi_s) \quad (\text{K.11a})$$

$$\begin{aligned} \text{s.t.} & \quad \sum_{s \in S} \sum_{n \in \mathcal{N}} d_{\text{W}}^{(1)}(\xi_s, \hat{\xi}_n) \pi_{sn} \leq \theta \\ & \quad \sum_{s \in S} \pi_{sn} = \frac{1}{N}, \quad \forall n \in \mathcal{N}. \end{aligned} \quad (\text{K.11b})$$

$$\quad (\text{K.11c})$$

We want to make the maximization problem into a minimization problem, therefore we take the dual (for intuition about duals, see Appendix F) and get:

$$\min_{\gamma, \alpha} \theta \alpha + \frac{1}{N} \sum_{n \in \mathcal{N}} \gamma_n \quad (\text{K.12a})$$

$$\text{s.t.} \quad \gamma_n \geq Q(x, \xi_s) - \alpha d_{\text{W}}^{(1)}(\xi_s, \hat{\xi}_n), \quad \forall s \in \mathcal{S}, \forall n \in \mathcal{N} \quad (\text{K.12b})$$

$$\alpha \geq 0. \quad (\text{K.12c})$$

C.f. Theorem 1 in Gao and Kleywegt, 2016, there is no duality gap. We can now take this result and put it back into the original program given in Equation K.8. Since they are now both minimization problems, we can simplify it to one single program:

$$\min_{x, \gamma, \alpha} c^\top x + \theta \alpha + \frac{1}{N} \sum_{n \in \mathcal{N}} \gamma_n \quad (\text{K.13a})$$

$$\text{s.t.} \quad Ax \leq b \quad (\text{K.13b})$$

$$\gamma_n \geq Q(x, \xi_s) - \alpha d_{\text{W}}^{(1)}(\xi_s, \hat{\xi}_n) \quad \forall s \in \mathcal{S}, \forall n \in \mathcal{N} \quad (\text{K.13c})$$

$$\alpha \geq 0. \quad (\text{K.13d})$$

This result is consistent with Theorem 1. in Tonbari, 2021. As done for 1-stage DRO and by Bertsimas et al., 2018, it may be beneficial to further dualize this twice over and thereby re-writing the problem as an equivalent robust optimization problem. Notice that this could ease the computational burden since we would not be required to solve all the inner programs. However, in our specific case, the inner program contains integer variables; hence, it is very difficult to determine what consequences this dualization would have because we would expect some duality gap. We will therefore omit these potential steps, as they are infeasible in our particular case. This results in the particular formulation:

$$\min_{x,y,\gamma,\alpha} - \left(\sum_{t \in \mathcal{T}} p_t^{\text{DA}} \hat{\lambda}_t^{\text{DA}} - p_t^{\text{DA}-} \lambda^{\text{TSO}} \right) + \theta\alpha + \frac{1}{N} \sum_{n \in \mathcal{N}} \gamma_n \quad (\text{K.14a})$$

$$\text{s.t. } Q_s = - \sum_{t \in \mathcal{T}} \begin{pmatrix} p_t^{\text{DA}} \xi_{s,t}^{\text{DA}} + p_{s,t}^B (\hat{\lambda}_t^B + \xi_{s,t}^B) & \dots \\ -p_{s,t}^B \lambda^{\text{TSO}} + d_{s,t} \lambda^h - z_{s,t}^{\text{su}} \lambda^{\text{su}} \end{pmatrix} \quad s \in \mathcal{S} \quad (\text{K.14b})$$

$$\gamma_n \geq Q_s - \alpha d_W^{(1)}(\xi_s, \hat{\xi}_n), \quad \forall s \in \mathcal{S}, \forall n \in \mathcal{N} \quad (\text{K.14c})$$

$$\hat{P}_t^w + \xi_{s,t}^w = p_t^{\text{DA}} + p_{s,t}^B + p_{s,t}^e + p_{s,t}^c \quad t \in \mathcal{T}, s \in \mathcal{S} \quad (\text{K.14d})$$

$$p_t^{\text{DA}+}, p_t^{\text{DA}-} \geq 0 \quad t \in \mathcal{T} \quad (\text{K.14e})$$

$$p_{s,t}^{B+}, p_{s,t}^{B-} \geq 0 \quad t \in \mathcal{T}, s \in \mathcal{S} \quad (\text{K.14f})$$

$$\alpha \geq 0 \quad (\text{K.14g})$$

$$y_s \in \mathcal{Y} \quad s \in \mathcal{S}. \quad (\text{K.14h})$$

Much work is being done to improve the computational burden our particular type of problem. K. Kim, 2020 presents a dual decomposition algorithm for two stage distributionally robust mixed-integer programs using the Wasserstein ambiguity set. *Argonne-National-Laboratory* is spearheading work to produce efficient solvers for these types of problems, see, e.g., their relevant git. For additional information about improved computational complexity, Tonbari, 2021 also discuss different decomposition methods for two stage programs with integer variables in the inner program.

Generation of Synthetic Samples

The solution to the two-stage DRO problem given above requires some synthetic samples. A naive approach to generating these would be to generate all permutations of some finite discretization of the sample space. However, in our case, any sample is a daily time series and even with a low resolution in the discretization will lead to an infeasible number of synthetic samples. Say we want to discretize any time point by 2 possible values, then the naive approach would require 2^{24} synthetic samples.

It is obvious that we cannot simply discretize over the sample space, instead we should try to produce a finite amount of reasonable samples. Notice that the samples do not have to approximate the expected distribution, since the samples will be allocated probability mass depending on their Wasserstein distance to the selected historical samples. It is therefore of greater importance that the synthetic samples covers as much as the sample space as possible. The issue of probable scenario generation is discussed by Pinson et al., 2009; we will go through the main ideas here.

Let X and Y denote two stochastic processes, which may be correlated and auto-correlated. Additionally, they need not follow any parametric distribution. Let their empirical distribution function be given by

$$EDP_X(t) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}(X_i \leq t) \quad (\text{K.15})$$

$$EDP_Y(t) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}(Y_i \leq t), \quad (\text{K.16})$$

where $\mathbb{1}$ is the indicator function. Then let

$$X_{unif} = EDP_X(X) \quad (\text{K.17})$$

$$Y_{unif} = EDP_Y(X) \quad (\text{K.18})$$

where X_{unif} and Y_{unif} are now both uniformly distributed variables on $[0, 1]$. Now we let

$$X_{norm} = \Phi^{-1}(X_{unif}) \quad (\text{K.19})$$

$$Y_{norm} = \Phi^{-1}(Y_{unif}) \quad (\text{K.20})$$

where Φ is the cumulative distribution function for a standard Gaussian distribution. X_{norm} and Y_{norm} are now standard normal variables. We now estimate the multivariate covariance structure

$$\hat{\Sigma} = Cov([X_{norm}, Y_{norm}]). \quad (\text{K.21})$$

This allows us to generate normal samples with covariance, $\hat{\Sigma}$, using standard statistical software packages. Let a sample be given by

$$[X_{norm}^s, Y_{norm}^s] \quad (\text{K.22})$$

then it can be transformed back by

$$X^s = EDP_X^{-1}(\Phi(X_{norm}^s)) \quad (\text{K.23})$$

$$Y^s = EDP_Y^{-1}(\Phi(Y_{norm}^s)) \quad (\text{K.24})$$

and like this, we have generated a synthetic sample with the same linear dependence structure and moments as the original samples.

K.2 Sample Selection

We use the Wasserstein distance to select similar historical days for each specific day under consideration. The Wasserstein distance measures the dissimilarity between probability distributions. In our case, we calculate the Wasserstein distance between the forecasted values (wind power and day-ahead prices) of the historical days and the corresponding forecasted values of the day in question.

We aim to identify historical days with the smallest Wasserstein distance, indicating a high degree of similarity in the forecasted values. This selection process ensures that the historical days selected closely resemble the characteristics of the day being optimized,

enabling us to make informed decisions based on past observations and providing valuable insights into the likely outcomes and patterns in the future.

It is important to note that selecting similar historical days based on the Wasserstein distance is subject to certain constraints and considerations. The availability of historical data plays a crucial role, and we must ensure that sufficient data points are present for accurate comparisons. This is why we have chosen to down-scale to an 8-hour day such that every day is split into three. By employing this approach, we capture the underlying patterns and dependencies presented in the forecasted values, facilitating the construction of an accurate 2-stage optimization model.

Please refer to Appendix D for a more detailed explanation of the Wasserstein distance calculation.

K.3 Results

Section K.1 presents two 2-stage models: The 2-stage SAA model and the 2-stage DRO model. However, in this section, we only report results for the 2-stage DRO model since the 2-stage SAA model is equivalent to the 2-stage DRO model with $\theta = 0$. Additionally, we only consider the active realization as the 2-stage models do not come with a hydrogen plan, making it impossible to realize the model passively. For comparison, the active realization of the deterministic model for both the training and test data in an 8-hour day setup (as used in the 2-stage DRO model) is given in Table K.1.

Table K.1 – Performance metrics for the deterministic model for an 8-hour day with an active realization.

Data	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Training	1,000	1,140	0.88	NA	-834	85.1 %
Test	756	707	1.07	NA	-654	89.9 %

Table K.2 shows the performance metrics for the 2-stage DRO model with varying values of θ on the training data. We observe that the optimal value of θ is 0, which yields a return of 953 and a volatility of 1,427. This parameter achieves the highest average daily return.

Table K.2 – Performance metrics for the 2-stage DRO model with varying θ for training data.

	θ	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Active	0	953	1,427	0.67	-0.061	-7,474	80.6%
	2	950	1,151	0.83	-0.117	-2,126	86.3%
	4	937	1,063	0.88	-0.182	-1,340	87.3%
	8	923	1,049	0.88	-0.202	-794	87.1%
	16	921	1,056	0.87	-0.191	-794	87.3%

Moving on to the test data shown in Table K.3, we see that the 2-stage DRO model achieves a return of 637 and volatility of 820, resulting in a Sharpe ratio of 0.82 for the θ value determined in the training data. The return is lower than the deterministic model, and the volatility is higher. We see that for other choices of θ , the 2-stage DRO model performs quite well compared to the deterministic models in terms of return and volatility. This

indicates that something goes wrong with how we select the Wasserstein radius in training data.

Table K.3 – Performance metrics for the 2-stage DRO model with varying θ for the test data. Notice that $\theta = 0$ was optimal in the training data.

	θ	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Active	0	673	820	0.82	-0.082	-2,976	85.8%
	2	720	730	0.99	-0.038	-1,611	90.2%
	4	747	703	1.06	-0.001	-1,367	90.7%
	8	757	689	1.10	0.000	-664	91.0%
	16	758	691	1.10	0.002	-631	91.0%

K.4 Comparison of Models

Figure K.1 shows the accumulated returns for all methods using the active realization. Upon analyzing the results, we see that the 1-stage DRO model is the only method that outperforms the deterministic model.

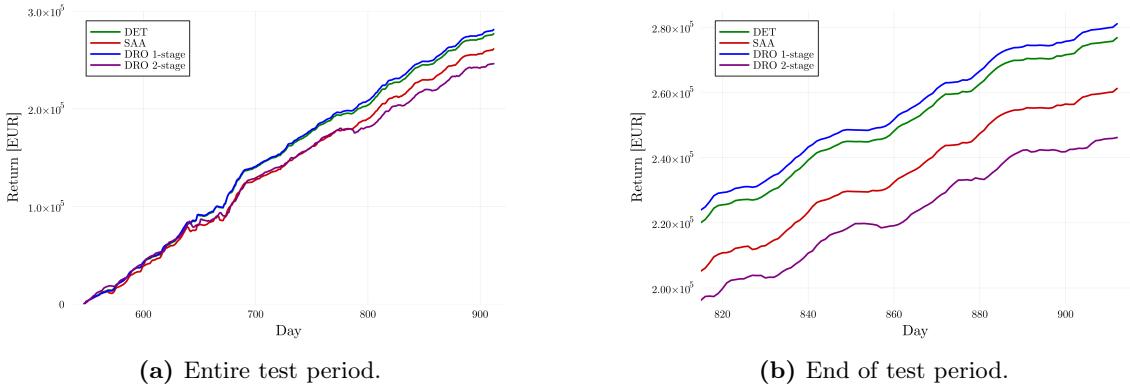


Figure K.1 – Accumulated return for active realizations in the test data.

Table K.4 presents the performance metrics of different models for both the training and test datasets, with percentages given relative to the deterministic model. The metrics evaluated include the average daily return, \bar{R} , and the volatility, $\hat{\sigma}$.

When compared to the deterministic model in the training data, the 1-stage SAA model shows a slight decrease in average daily return (-2.4%) and a significant increase in volatility (56.8%). The 1-stage DRO model achieves a slight increase in average daily return (0.6%) but also an increase in volatility (6.1%). The 2-stage DRO model sees a decrease in return (-4.7%) and an increase in volatility (25.2%).

Comparing the results to the deterministic model in the test data, we observe that the 1-stage SAA model exhibits larger a decrease in average return (-5.6%) and an increase in volatility (10.7%). The 1-stage DRO model achieves better average daily return (1.6%) while also reducing volatility (-1.3%). The 2-stage DRO model sees an even larger decrease in return than before (-11.0%) and higher volatility (16.0%).

Table K.4 – Average daily return, \bar{R} , and volatility, $\hat{\sigma}$, for the deterministic model, the 1-stage SAA model, the 1-stage DRO model, and the 2-stage DRO model, for an 8-hour day and active realization with percentages relative to the deterministic model.

Model	Training				Test			
	\bar{R}	% $\Delta\bar{R}$	$\hat{\sigma}$	% $\Delta\hat{\sigma}$	\bar{R}	% $\Delta\bar{R}$	$\hat{\sigma}$	% $\Delta\hat{\sigma}$
DET	1,000	0.0%	1,140	0.0%	756	0.0%	707	0.0%
SAA-1stage	976	-2.4%	1,788	56.8%	714	-5.6%	783	10.7%
DRO-1stage	1,006	0.6%	1,210	6.1%	768	1.6%	698	-1.3%
DRO-2stage	953	-4.7%	1,427	25.2%	673	-11.0%	820	16.0%

K.5 Discussion

The 2-stage DRO model presented here has revealed some inherent issues that warrant further consideration. While using the Wasserstein distance for sample selection initially appeared to be a sensible approach, it is important to acknowledge its major flaw: The loss of temporal information in the compared distributions.

The absence of temporal elements within the compared distributions can result in days with significantly different forecasts for specific hours being considered similar days. Consequently, the model may erroneously identify these days as similar, despite the notable variations in forecasts for certain hours. Additionally, the 2-stage DRO model's treatment of synthetic samples becomes problematic. The model may assign a disproportionately low cost to transfer probability from the selected historical samples to synthetic samples that exhibit substantial discrepancies in specific hours for the same reasons. This is demonstrated in the dramatic change in the optimal Wasserstein radius for the training- and test data; the poor choice in metric leads to a model that generalizes very poorly.

These observations emphasize the challenges of determining similarities between days with multiple variables and the importance of choosing a suitable metric. The complexity of comparing times series makes it inherently difficult to capture and integrate all relevant factors, including forecasted values, temporal patterns, and contextual information. Despite efforts to address this issue, satisfactory solutions remain elusive, and ongoing research is needed to further explore and develop effective methodologies in this area.

It is worth noting that while the model performed relatively better with higher θ values in the test data, indicating a more risk-averse approach, it does not necessarily imply that the model performed well. Instead, it highlights that the model relies on a conservative strategy to limit speculation and mitigate risks by turning to a long-term "good" strategy that works reasonably well for any forecast.

L Training Results

We have given all the results from fitting parameters on the training data in the following. Plots show accumulated returns over time for test data with the found optimal parameters.

L.1 Training Results for the Deterministic Wind/Hydrogen Model

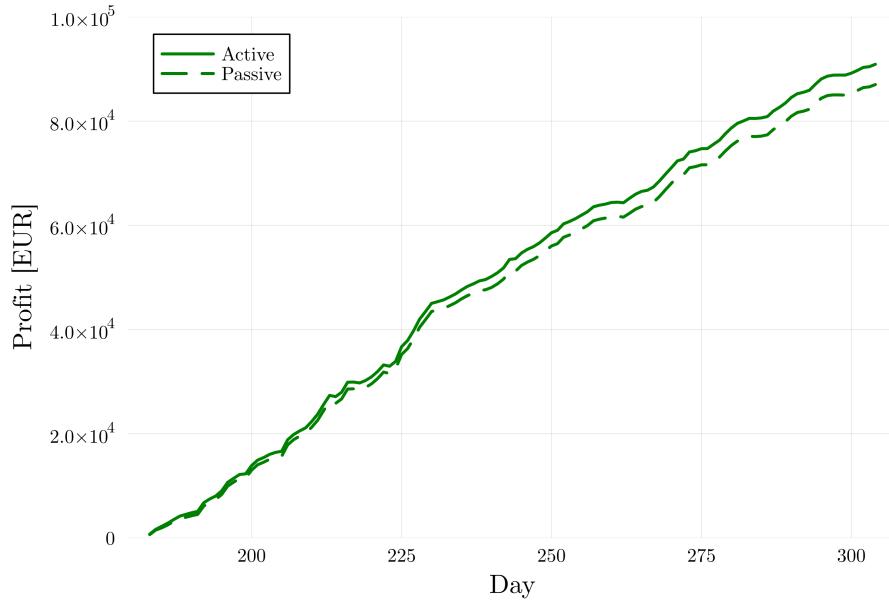


Figure L.1 – Accumulated returns over time for test data. Rescaled wrt expected active return from the training data.

L.2 Training Results for the SAA Wind/Hydrogen Model

Table L.1 – Performance metrics for SAA with varying ϵ for training data.

	ϵ	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	% R_+
Passive	0.01	858	1,275	0.67	-0.047	-4,804	80.7 %
	0.05	858	1,276	0.67	-0.047	-4,804	80.7 %
	0.15	859	1,276	0.67	-0.046	-4,804	80.7 %
	0.30	859	1,277	0.67	-0.046	-4,804	80.7 %
	0.50	860	1,277	0.67	-0.045	-4,804	80.7 %
Active	0.01	984	1,288	0.76	-0.021	-4,721	85.7 %
	0.05	985	1,288	0.76	-0.021	-4,721	85.7 %
	0.15	985	1,289	0.76	-0.020	-4,721	85.7 %
	0.30	985	1,289	0.76	-0.019	-4,721	85.0 %
	0.50	986	1,290	0.76	-0.018	-4,721	85.0 %

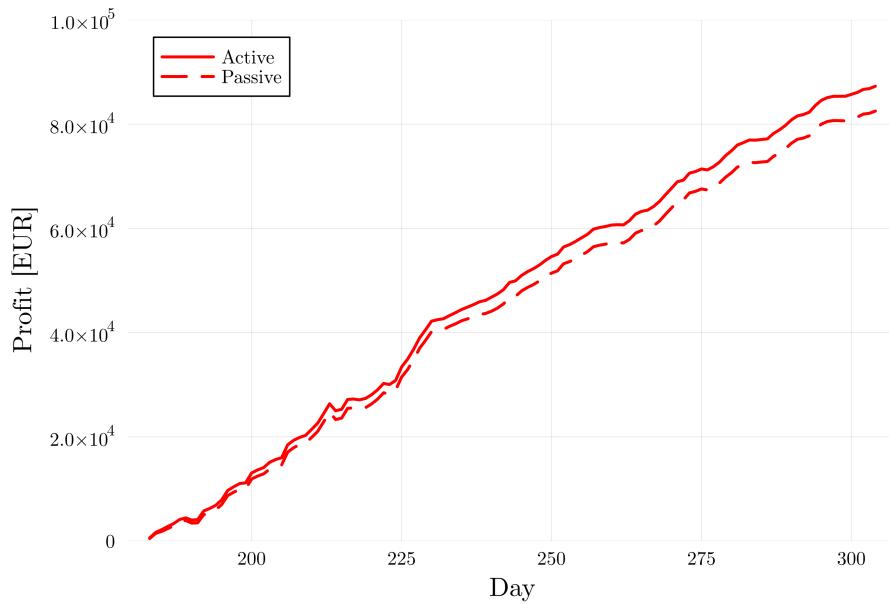


Figure L.2 – Accumulated returns over time for test data for SAA.

L.3 Training Results for the Wasserstein DRO Wind/Hydrogen Model

Table L.2 – Performance metrics for regular DRO with varying θ , $\theta^{\text{CVaR}} = 0$, and $\epsilon = 0.5$ for training data.

	θ	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Passive	0	860	1,277	0.67	-0.045	-4,804	80.7 %
	10	885	1,111	0.80	-0.018	-1,900	80.7 %
	20	889	991	0.90	-0.017	-754	86.4 %
	22	893	993	0.90	-0.003	-754	86.4 %
	30	881	951	0.93	-0.062	-495	87.9 %
Active	0	986	1,290	0.76	-0.018	-4,721	85.0 %
	10	1,006	1,125	0.89	0.013	-1,540	85.0 %
	20	1,006	1,008	1.00	0.023	-742	92.9 %
	22	1,010	1,012	1.00	0.040	-742	93.6 %
	30	999	971	1.03	-0.006	-198	95.0 %

Table L.3 – Performance metrics for regular DRO with $\theta = 22$, varying θ^{CVaR} , and $\epsilon = 0.5$ for training data.

	θ^{CVaR}	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Passive	0.00	893	993	0.90	-0.003	-754	86.4 %
	0.02	893	991	0.90	-0.005	-750	87.1 %
	0.04	892	988	0.90	-0.007	-745	87.1 %
	0.08	889	985	0.90	-0.017	-739	87.1 %
	0.16	883	980	0.90	-0.041	-739	87.1 %
Active	0.00	1,010	1,012	1.00	0.040	-742	93.6 %
	0.02	1,010	1,010	1.00	0.039	-738	93.6 %
	0.04	1,010	1,008	1.02	0.039	-734	93.6 %
	0.08	1,009	1,004	1.00	0.035	-728	93.6 %
	0.16	1,005	999	1.01	0.020	-728	93.6 %

Table L.4 – Performance metrics for regular DRO with $\theta = 22$, $\theta^{\text{CVaR}} = 0$, and varying ϵ for training data.

	ϵ	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Passive	0.01	888	983	0.90	-0.022	-743	87.1 %
	0.05	889	985	0.90	-0.020	-744	87.1 %
	0.15	890	988	0.90	-0.013	-747	87.1 %
	0.30	892	991	0.90	-0.009	-751	87.1 %
	0.50	893	993	0.90	-0.003	-754	86.4 %
Active	0.01	1,006	1,003	1.00	0.023	-731	93.6 %
	0.05	1,007	1,005	1.00	0.026	-732	93.6 %
	0.15	1,008	1,008	1.00	0.031	-736	93.6 %
	0.30	1,009	1,010	1.00	0.036	-739	93.6 %
	0.50	1,010	1,012	1.00	0.040	-742	93.6 %

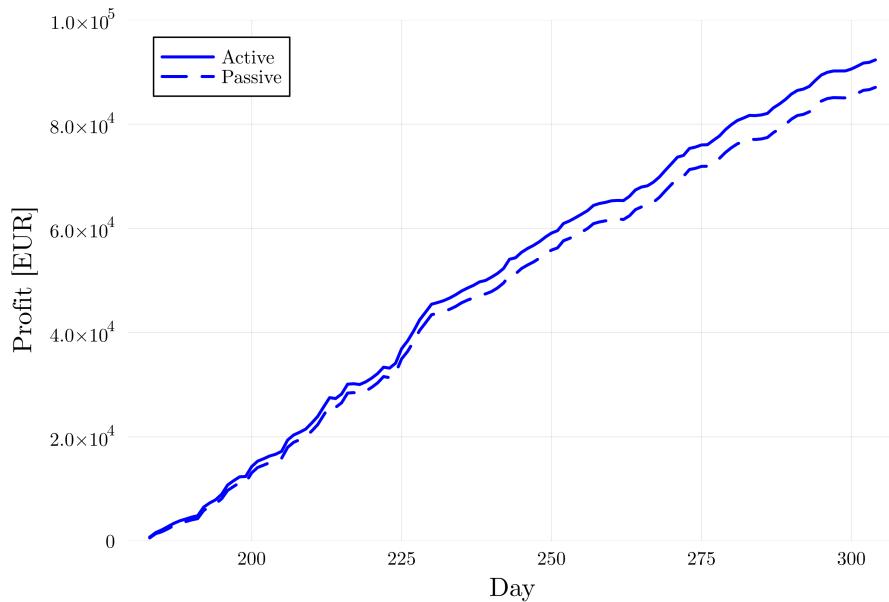


Figure L.3 – Accumulated returns over time for test data for regular DRO.

L.4 Training Results for the Wasserstein DRO with Correlation Wind/Hydrogen Model

Table L.5 – Performance metrics for regular DRO with correlation with varying θ , $\eta = 1$, $\theta^{\text{CVaR}} = 0$, and $\epsilon = 0.5$ for training data.

	θ	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	$\%R_+$
Passive	0	853	1,274	0.67	-0.054	-4,804	80.7%
	10	880	1,109	0.79	-0.029	-1,966	80.7%
	20	890	989	0.90	-0.016	-754	85.7%
	25	891	988	0.90	-0.010	-754	87.1%
	30	890	958	0.93	-0.020	-495	87.1%
Active	0	986	1,290	0.76	-0.018	-4,721	85.0%
	10	1,006	1,125	0.89	0.014	-1,540	85.0%
	20	1,010	1,008	1.00	0.038	-742	92.9%
	25	1,011	1,010	1.00	0.043	-742	94.3%
	30	1,009	978	1.03	0.046	-290	94.3%

Table L.6 – Performance metrics for regular DRO with correlation with $\theta = 25$, varying η , $\theta^{\text{CVaR}} = 0$, and $\epsilon = 0.5$ for training data.

	η	\bar{R}	$\hat{\sigma}$	\hat{S}	IR	WSD	% R_+
Passive	1.0	891	988	0.90	-0.010	-754	87.1%
	1.5	894	993	0.90	0.000	-761	87.9%
	2.0	886	977	0.91	-0.033	-756	87.9%
	4.0	889	991	0.90	-0.015	-754	87.9%
	1,000	887	982	0.90	-0.030	-754	87.9%
Active	1.0	1,011	1,010	1.00	0.043	-742	94.3%
	1.5	1,014	1,014	1.00	0.049	-749	95.0%
	2.0	1,006	999	1.01	0.026	-744	95.0%
	4.0	1,011	1,013	1.00	0.040	-742	95.0%
	1,000	1,003	1,003	1.00	0.014	-742	95.0%

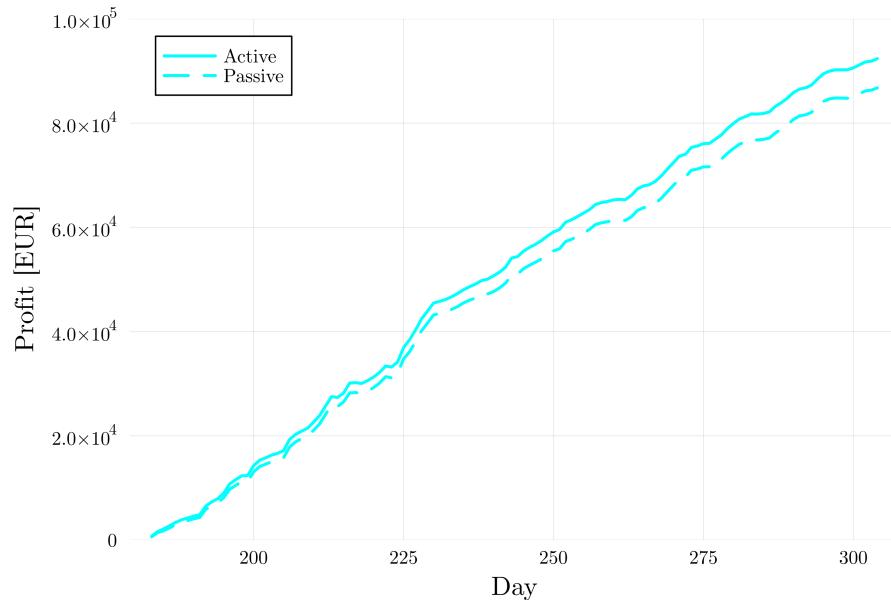


Figure L.4 – Accumulated returns over time for test data for DRO with correlation.

M Passive Realization for Varying Hydrogen Prices

In section F.2, we analyzed how access to an electrolyzer affected the returns of a hybrid power plant. Only active realizations were analyzed because the conclusion is the same for passive realization, but the passive realizations are given here for completeness.

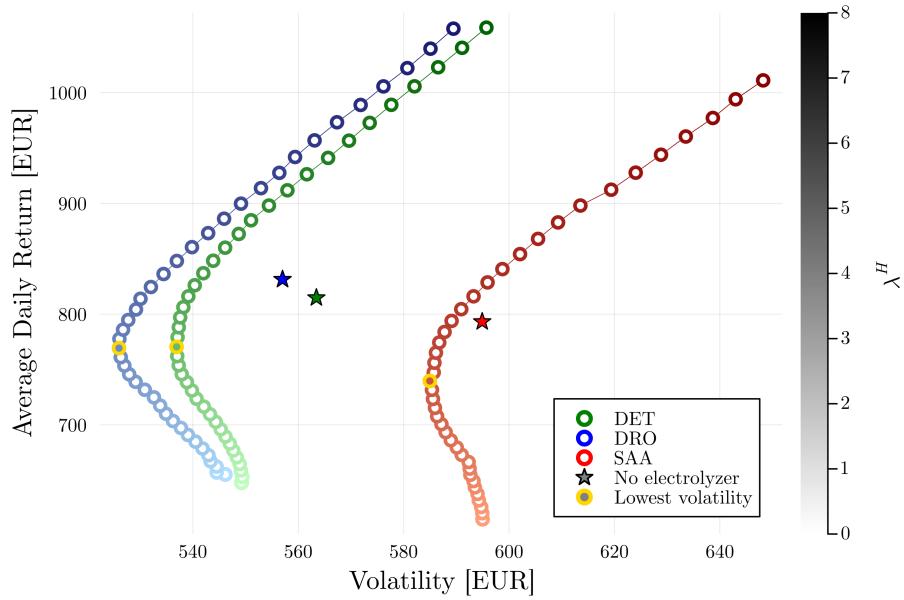


Figure M.1 – Passive realization for the scenario with daily hydrogen contracts. For each point on the outlined curves, the hydrogen price increase by 0.2 EUR, starting at 0.0 EUR and ending at 8.0 EUR.

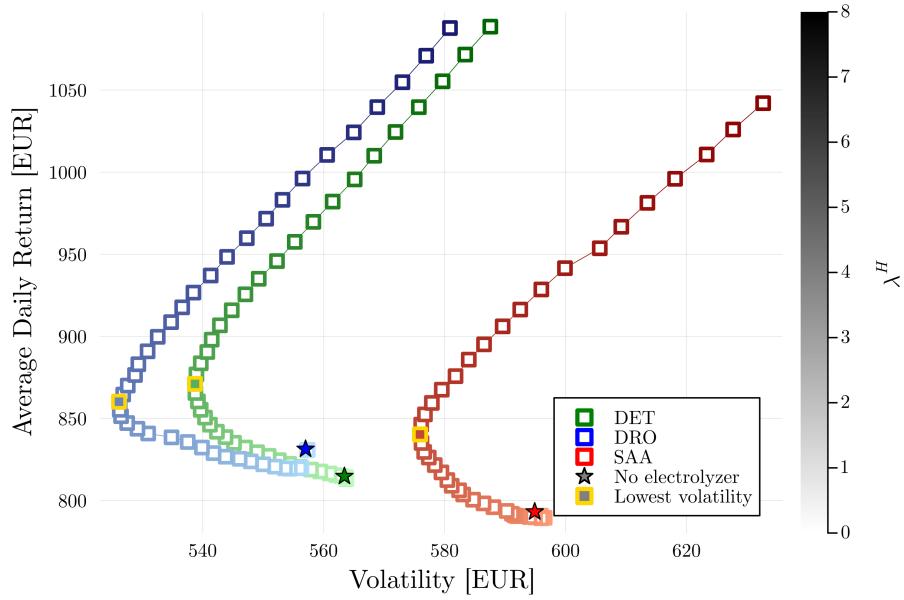


Figure M.2 – Passive realization for the scenario without daily hydrogen contracts. For each point on the outlined curves, the hydrogen price increase by 0.2 EUR, starting at 0.0 EUR and ending at 8.0 EUR.