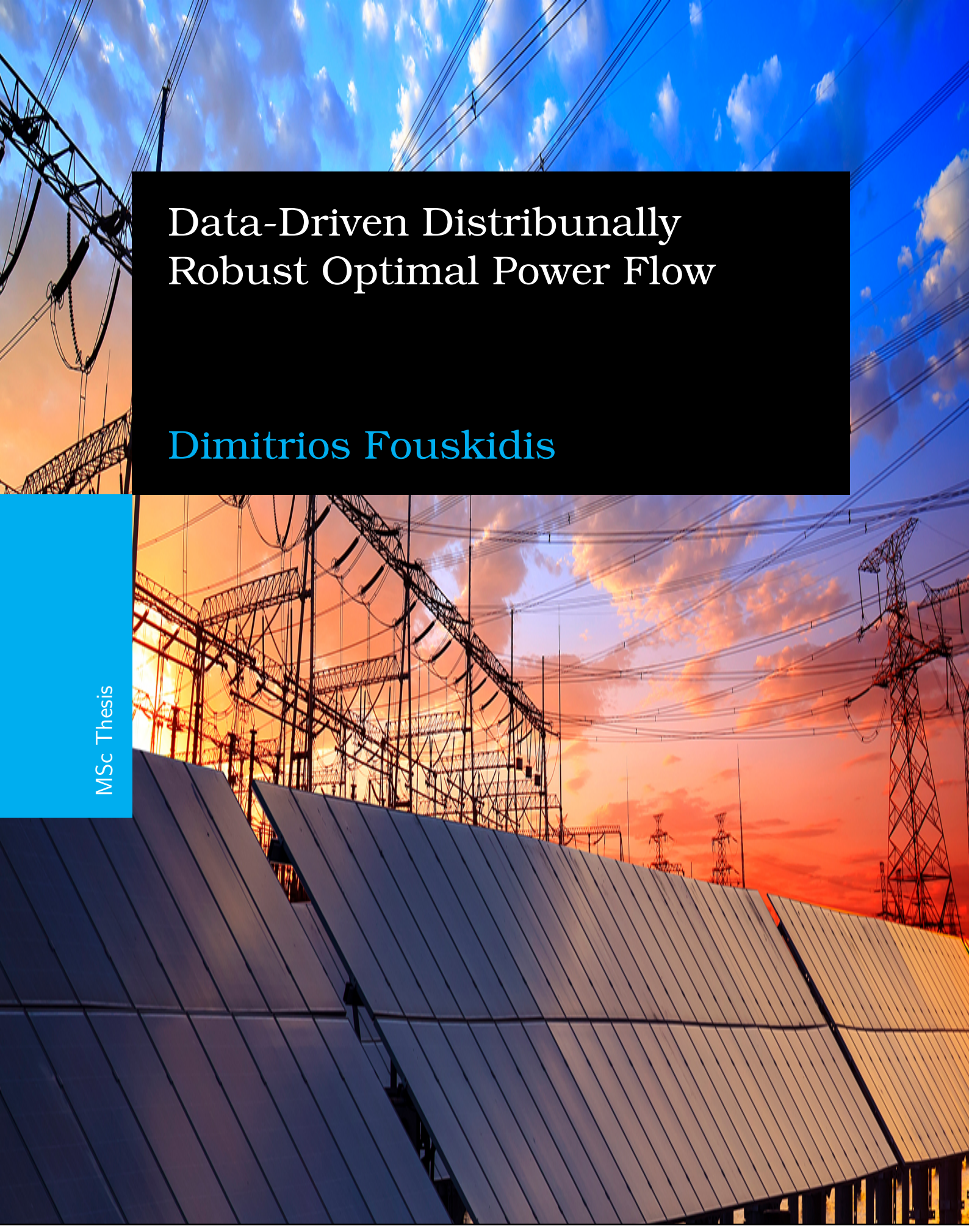


Data-Driven Distribunally Robust Optimal Power Flow

Dimitrios Fouskidis

MSc Thesis

TU Delft Delft University of Technology

Delft Center for Systems and Control

Data-Driven Distribunally Robust Optimal Power Flow

MSc Thesis

Dimitrios Fouskidis

Supervisors: Asst. Prof. Dimitris Boskos
Asst. Prof. Pedro Vergara Barrios

August 19, 2022

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of Technology

Table of Contents

List of Figures

Abstract

The increased diffusion of Renewable Energy Sources (RES) into energy distribution systems gives rise to a number of issues that the Distribution Network Operators (DNOs) need to face. The dependency of RES on weather, along with their intermittent and non-dispatchable nature, urges us to develop frameworks that can guarantee a stable network, despite the power fluctuations.

In this thesis, we develop a Data-Driven Optimal Power Flow (OPF) formulation in order to achieve voltage regulation against operational problems that may occur because of the high PV penetration into energy distribution systems. The unpredictable PV generation deteriorates the system's reliability and introduces instability. Thus, we develop appropriate formulations to define a limit on exports from residential PV owners to the grid. For this purpose, we employ Distribunally Robust Chance Constraint Programming (DRCCP), as a method that can handle constraints that depend on the uncertain PV generation and residential demand. We capture the distributional uncertainties with an ambiguity set and we utilize the Wasserstein metric to parameterize the range of this set.

We divide the thesis into two case studies. In the first case study, we perform the DRCCP optimization with only few recorded data. We evaluate the results of our algorithm in a dataset of 500 days of recorded data and we achieve to reduce significantly the overvoltage instances. We tune the DRCCP in a way that we achieve at least 95% satisfaction of constraints that emerge from the DRCCP, avoiding overconservative solutions and high generation cuts that could harm the PV owners.

In the second case study, we intend to exploit plenty of historical data, but the computational burden hinders us from using them in the DRCCP. Therefore, we employ Wasserstein Barycenters, and we utilize the Wasserstein distance in order to cluster the data. With Wasserstein Barycenters we reduce significantly the running time and we provide an efficient and robust output, for at least 95% satisfaction of the model's constraints.

Acknowledgements

This work consists the final stage of my Master Studies in Systems and Control at TU Delft university. I would like to thank my supervisors Prof. D. Dimitris Boskos and Prof. Pedro Vergara Barrios for their assistance during the writing of this thesis. I would also like to thank my family and my friends for their support during these two years of my studies.

Delft, University of Technology
August 19, 2022

Dimitrios Fouskidis

Chapter 1

Introduction

1-1 Motivation

One major objective of public authorities and people around the globe concerns the reduction of carbon dioxides that conventional fossil fuels emit. Climate change, peoples' awareness and high dependency on energy imports [1] urge European governments to prioritize green energy transition from fossil fuels to sustainable and Renewable Energy Sources (RES). The global impact of energy transition can be addressed by international conferences and agreements, such as the Kyoto protocol in 1997, the Paris agreement in 2015, the European Green Deal and the United Nations Framework Conventions on Climate Change (UNFCCC), where measures and long term targets have been set for the future of the planet.

Solar photovoltaic (PV) units are one of the most promising renewable energy sources, and their use is growing at an exponential rate. In particular, the global PV capacity from 480 GW in 2018 increased to 580 GW in 2019. If the same growth trends continues, PV contributions are expected to reach 2840 GW by 2030 and 8500 GW by 2050 [2, 3]. The main advantage of PV units, compared to other RES, is their easy installation in urban areas and buildings. As a result, the cost of investment is affordable to individuals and not only to private investing funds [4]. However, the gradual raise of PV penetration into residential areas and their connection to the grid created new requirements for the network operators, regarding the system's balance. After the installation, the residential PVs are connected to the distribution energy network and Distribution Network Operators (DNOs) need to face challenges that concern technical and security issues.

One significant feature of PV units, is their uncertain generation. The PV generation is intermittent and weather dependant. The difficulty to specify the net load and to determine the most cost-efficient dispatch of the conventional generators introduces new problems for the network operators. Overvoltage, congestion and ramping issues occur with high likelihood and deteriorate the reliability of the grid. Except from these technical problems, the storage of energy during peak generation and low demand may not be profitable for investors [5, 6]. Since the historical data are limited and the probability distribution concerning the PV generation

and the residential demand is not known, we develop distributionally robust optimization methods that can guarantee a stable network, despite the power deviations.

1-2 Research question and intended outcomes

In this report, we focus on data-driven control via Distribunally Robust Chance Constraint Programming (DRCCP). We apply DRCCP for an optimal power flow (OPF) formulation in a distribution energy network. Within this OPF formulation, we incorporate the uncertainties of PV generation and residential demand and we evaluate ways to define the minimum curtailment of excess power during time periods of low demand and high irradiation. Through power curtailment, we achieve voltage regulation in a way that the voltages in the entire network lie within the safety limits. The approach of PV power curtailment and optimization problems that we develop to allocate the power demands, lead us to formulate the main research question of this report :

How to achieve voltage regulation in distribution energy networks, by curtailing appropriately the power exports from residential PV installations to the energy grid?

We conduct the analysis of this research question in stages by answering the following sub-questions:

What are the available optimization methods to incorporate the uncertain PV generation and residential demand?

With this sub-question we investigate the optimization methods that can handle the uncertain PV generation and residential demand.

How to exploit data from previous years, in order to mitigate the power mismatch?

With this sub-question we examine the mismatch between PV power exports and residential demand.

How to use data efficiently and reduce the computational effort?

With this sub-question we intend to evaluate the available options to decrease the computational effort of the problem.

The intended outcomes from this thesis are:

- To identify the problems that occur from PV diffusion into energy distribution systems.
- To achieve voltage regulation through curtailment of PV power exports during high sun irradiation and low demand periods.
- To model the unknown distribution of PV generation and residential demand in a robust way.
- To appropriately cluster the recorded data and obtain more efficient optimization algorithms.

1-3 Methodology

The methodology that we employ in order to address the research question is the following:

Data-driven Wasserstein ambiguity sets.

In order to model the uncertain distribution of PV generation and residential demand, we construct an uncertainty set. We choose this set to capture both the deviation of PV generation and the deviation of the residential demand. We construct this set with the assistance of the Wasserstein distance metric, that is derived from the Optimal Transport theory.

DRCCP over Wasserstein ambiguity sets on a Dutch energy distribution system with high PV penetration.

In order to implement our research, we utilize a Dutch distribution energy network with high PV penetration. We employ this network in order to define the curtailed PV output to the power grid during a summer day with high mismatch between generation and demand. We linearize the network, we insert real data from previous years and we simulate our experiments on MATLAB 2020b environment. All simulations run on AMD Ryzen 4800h Processor with 8 GB of RAM.

Clustering data for efficient DRCCP

We develop two distinctive case studies. One scenario indicates that we form the DRCCP with 20 days of recorded data for consecutive time-instances. In the second scenario, we use the whole dataset of 500 recorded days to form the DRCCP. However, due to high computational effort, we cluster the data. Then, we evaluate the performance of these two case studies in terms of robustness and computational effort.

1-4 Contributions

The solution of DRCCP over Wasserstein ambiguity sets is still limited for energy distribution systems with high RES penetration. To the best of our knowledge, only [7] examines a relative topic, but for simplicity they consider the risk of every constraint individually. Moreover, other works exist, such as [8] and [9], that address either to energy transmission systems, or they make use of moment ambiguity set. In our case, we leverage for the first time in literature the DRCCP method with Wasserstein ambiguity sets for energy distribution systems, where the violation probability is considered for all constraints simultaneously. A second contribution of this thesis concerns the utilization of Wasserstein Barycenters in energy systems. Wasserstein Barycenters correspond clusters of samples under optimal Wasserstein distance. We find applications of Wasserstein Barycenters in machine learning and image processing tasks [10], but until now there are no applications in energy systems. Since many OPF formulations that incorporate RES face computational burdens due to model complexity [11], the sample clustering is crucial to reduce the computational time without significant loss of information.

1-5 Thesis Outline

We divide the thesis into 7 distinctive chapters. We organize the chapters in a way that we answer the sub-questions step by step until we reach consequently to the answer of the main research question. The remainder of the thesis is organized as follows.

Chapter 2 analyzes the optimization methods that we can employ in problems with uncertain constraints. We specify various optimization methods and we choose the approach that fits to our problem the best.

Chapter 3 presents the DRCCP principles and builds the Wasserstein ambiguity set step by step. In this chapter we also formulate the structure of the optimization problem that we will use in our implementation.

Chapter 4 engages the power systems and the OPF in the problem. In this chapter, we analyze the different types of networks in the energy grid, we introduce the benefits of power curtailment from PVs and we present the linearized version of distribution energy networks.

Chapter 5 includes the implementation part of the DRCCP with a Wasserstein ambiguity set into a Dutch energy distribution system. We formulate the case-study, we insert the data and we present the results of our simulation.

Chapter 6 analyzes Wasserstein Barycenters. In this chapter, we examine the literature behind Wasserstein Barycenters, we present the corresponding algorithm and we solve again the OPF problem with the clustered dataset.

Chapter 7 is the final chapter of the thesis, where we place the conclusion by answering the research question. Moreover, we provide recommendations for future work.

Chapter 2

Optimization problems with uncertain constraints

The diffusion of PVs into modern society has inserted numerous challenges that DNOs need to face. DNOs are in charge of operating the distribution networks in a reliable way, by enforcing operational constraints, such as voltage limits and thermal limits. In particular, the high variability of PV generation affects the stability of the network, since it influences the voltage of the network. To ensure that residential PV units do not violate the operational constraints of the network, we analyze optimization methods that are able to handle problems with uncertain parameters in constraints. The purpose of this chapter is to examine these methods and then to determine the one that fits to our problem the best.

The remainder of this chapter is organized as follows. In Sections 2-1, 2-2 and 2-3 we investigate the available methods in literature. We introduce Robust Optimization, Chance Constraints Programs and Scenario Approach and we present their theoretical principles with their mathematical representations. However, in Section 2-4 we justify why the aforementioned methods do not suit to our case study and we introduce Distribunally Robust Chance Constraint Programs (DRCCP) as an alternative.

2-1 Robust Optimization and Chance Constraint Optimization

Robust optimization (RO) and Chance Constrain Programming (CCP) are two methods that we use widely to take decisions under constraints dependent on uncertain parameters. On the one hand, the goal of RO is to determine the decision variables in a way that the result is feasible in all possible realizations of the uncertainties in the constraints. In this approach we assume the support of the uncertainties known and the main disadvantage is the conservative solution that we may get [12, 13, 14, 15]. The mathematical representation of RO problems is as follows [16]:

$$\begin{aligned} \min_{x\in X,c} \quad & c \\ \text{s.t.} \quad & \max_{\xi\in\Xi} f(x,\xi) \leq c \qquad \forall \xi \in \Xi \end{aligned} \tag{2-1.1}$$

We know that $\max_{\xi\in\Xi} f(x,\xi) \leq c$ equals to $f(x,\xi) \leq c$, so we set $F(x,\xi) = f(x,\xi) - c$ and we rewrite the optimization problem (2-1.1) as:

$$\begin{aligned} \min_{x\in X,c} \quad & c \\ \text{s.t.} \quad & F(x,\xi) \leq 0 \qquad \forall \xi \in \Xi \end{aligned}$$

where F is the constraints function and $\xi \in \Xi$ represents the uncertainty parameter in the constraints.

Likewise, if we set $\max_{\xi\in\Xi} f(x,\xi) = c$, then we get the min-max problem:

$$\min_{x\in X} \max_{\xi\in\Xi} f(x,\xi)$$

On the other hand, we classify Chance Constrained Programs (CCP) as a specific type of Stochastic Optimization (SO) problems, and we intend to find a solution that satisfies the constraints with high probability. While in RO we assumed only the support of the uncertainties known, now in CCP we assume that the uncertainties belong to a known probability distribution. The mathematical representation of the CCP problems is as follows:

$$\begin{aligned} \min_{x\in X} \quad & c^T x \\ \text{s.t.} \quad & \mathbb{P}(F(x,\xi) \leq 0) \geq 1 - a \end{aligned} \tag{2-1.2}$$

where F is the constraints function, $\xi \in \Xi$ is the uncertain parameter and $a \in (0,1)$ is a predefined parameter, namely the *violation parameter*, which determines the probability that the constraints are violated. Consequently, $(1-a)$ indicates the probability that the constraints are feasible. We could see the CCP method as a way to solve problems whenever the validity of the constraints $F(x,\xi) \leq 0$ for all values of $\xi \in \Xi$ is too costly or unfeasible. [8, 12, 17, 18].

When we perform CCP, we may face some difficulties. In particular, we may need to use Monte Carlo simulation in order to validate that a chance constraint is fulfilled at a specific point x. As a result, this would increase the computational effort, especially when we select very small violation parameter a. In addition to that, we may have scalability problems because the feasibility set of a CCP may be nonconvex, even though our constraints function F is convex [18]. In the following section we examine alternatives, in order to turn the possibly non-convex CCP into a convex approximation.

2-2 Value at Risk and Conditional Value at Risk

Value at Risk (VaR) and Conditional Value at Risk (CVaR) are two risk measures that are used in financial sector by investment companies and brokerage firms. We define VaR and CVaR for a random variable Y with distribution $\mathbb{P}$ at the risk-sensitivity level $a \in (0,1)$ as below:

$$\mathrm{VaR}_a^{\mathbb{P}}(Y) := \inf_{t\in\mathbb{R}} \left(t \in \mathbb{R} \middle| \mathbb{P}\left(Y \leq t\right) \geq 1-\alpha \right) \tag{2-2.1}$$

$$\mathrm{CVaR}_a^{\mathbb{P}}(Y) := \inf_{t\in\mathbb{R}} \left(a^{-1}\mathbb{E}_{\mathbb{P}}\left((Y-t)_+\right) + t \right) \tag{2-2.2}$$

where $[t]_+ = \begin{cases} t & \text{when } t > 0, \\ 0 & \text{when } t \leq 0. \end{cases}$

Equation (2-2.1) describes VaR_α as the lowest value t, where the probability of the random variable Y under the distribution $\mathbb{P}$ to be less than a threshold t is greater or equal to $(1-a)$. We observe that the constraint of the classical CCPs of Equation (2-1.2) corresponds to $\mathrm{VaR}_a\left(F\left(x,\xi\right)\right) \leq 0$. On the other hand, Equation (2-2.2) indicates the conditional expectation when the random variable Y exceeds this value t [12, 19, 20]. Figure 2-1 illustrates the aforementioned risk measures:

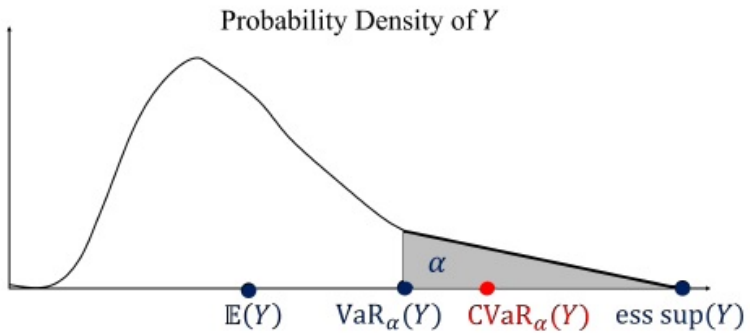


Figure 2-1: VaR_a and CVaR_a illustration. With VaR_α we find the lowest value t where the probability of the random variable Y under the distribution $\mathbb{P}$ to be less than a threshold t is greater or equal to $(1-a)$. The CVaR_a indicates the conditional expectation when the random variable Y exceeds this value t. The figure also illustrates the expected value of Y and the essential supremum of Y [21].

Since VaR misses the property of convexity, we use the convex measure CVaR_a for the approximation of CCPs [18, 20, 21, 22]. In this way, we obtain a convex CCP formulation of Equation (2-1.2), by considering the non-negative, non-decreasing, convex function $\psi(z) : \mathbb{R} \to \mathbb{R}$, given as $\psi(z) = \max(z+1, 0)$. For a given $\mathbb{P} \in \mathcal{P}(\Xi)$, we define $\Psi_{\mathbb{P}} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ as

$$\Psi_{\mathbb{P}}(x,t) := t\ \mathbb{E}_{\mathbb{P}}\left(\psi\left(t^{-1}F(x,\xi)\right)\right) \tag{2-2.3}$$

For Equation (2-2.3), we know that if $x \mapsto F(x,\xi)$ is convex $\forall\ \xi \in \Xi$, then $\Psi_{\mathbb{P}}$ is convex in x and t. This is indeed true, because $F(x,\xi)$ is convex, $\psi(z)$ is non-decreasing and convex and

thus $(x,t) \mapsto t\psi\left(t^{-1}F(x,\xi)\right)$ is convex. With the following implication, according to [18], we transform the CCP problem (2-1.2) into a convex conservative approximation.

$$\inf_{t>0}\left(\Psi_{\mathbb{P}}(x,t)-ta\right)\leq 0 \implies \mathbb{P}(F(x,\xi)\leq 0)\geq 1-a$$

This is indeed true because through the implication we get that the inequality

$$\inf_{t>0}\left(\Psi_{\mathbb{P}}(x,t)-ta\right)\leq 0 \tag{2-2.4}$$

acts as a conservative approximation of the constraint of the CCP (2-1.2). Therefore, we can rewrite (2-1.2) as:

$$\begin{aligned}\min_{x\in X,t>0}\quad & c^T x\\ \text{s.t.}\quad & \inf_{t>0}\left(\Psi_{\mathbb{P}}(x,t)-ta\right)\leq 0\end{aligned}$$

If we substitute $\psi(z)=\max(z+1,0)$ to (2-2.4), we get:

$$\inf_{t>0}\left(\mathbb{E}\left(\left(F(x,\xi)+t\right)_+\right)-ta\right)\leq 0$$

Then, we replace $\inf_{t>0}$ with $\inf_{t\in\mathbb{R}}$ without violating the validity of the expression and we get:

$$\inf_{t\in\mathbb{R}}\left(\mathbb{E}\left(-ta\left(F(x,\xi)+t\right)_+\right)\right)\leq 0$$

which is equal to the CVaR_a term that we introduced in (2-2.2). Following the aforementioned steps, we show that the convex conservative approximation of the original CCP (possible non-convex) (2-1.2) is the following problem that employs the CVaR_a measure:

$$\begin{aligned}\min_{x\in X}\quad & c^T x\\ \text{s.t.}\quad & \text{CVaR}_a\left(F\left(x,\xi\right)\right)\leq 0\end{aligned}$$

2-3 Scenario Approach

One alternative method to handle optimization problems with uncertain constraints is the Scenario Approach (SA). In SA we extract randomly a specific number N of instances-scenarios that describe the uncertain situations. The constraints should be feasible only for these specific scenarios that we draw. With this method we assume that there are plenty scenarios available and the more we increase the number of instances that we choose, the more precise the solution is. SA is a convex optimization problem when the original problem is convex and it is computationally efficient in case N is not extremely large. The mathematical representation of SA method for N i.i.d. samples is as follows:

$$\begin{aligned} \min_{x \in X} \quad & c^T x \\ \text{s.t.} \quad & F(x, \xi_i) \leq 0, \quad i = 1, \cdots, N \end{aligned}$$

where $x \in \mathbb{R}^n$, F is the constraints function and the term $\xi_i \in \Xi$ expresses the uncertainty parameter through the scenarios $i = 1, \cdots, N$ that we draw.

The main idea of the SA is to determine appropriately the necessary number N of scenarios that we use in our optimization through the constraints. By selecting the suitable number of scenarios, we can guarantee that our solution satisfies all the rest unseen scenarios, except from a fraction that we decide. This fraction goes to zero as we increase the number of scenarios that we draw. The formula to determine the number of scenarios is as follows:

$$N \geq \frac{2}{a} \ln \frac{1}{\beta} + 2n + \frac{2n}{a} \ln \frac{2}{a} \tag{2-3.1}$$

where $a \in (0, 1)$ is the violation parameter, similar to the a of the CCPs, and $\beta \in (0, 1)$ is the confidence parameter.

In other words, the SA method mentions that our solution x satisfies all the constraints except from an a-fraction with probability no smaller than $1 - \beta$. The more we increase N, the more we can decrease a. Moreover, for β we usually use small values, i.e 10^{-10}, but not 0, since $\beta = 0$ would correspond to $N = \infty$ [12, 16].

2-4 Distributionally Robust Chance Constrained Programs

In the previous section we mentioned RO, CCPs and SA as methods that handle problems that incorporate uncertain parameters in the constraints. The main principles of these methods were that either these uncertainties belong to a known probability distribution or there are plenty of samples available. However, in most cases the available samples are limited and they may not represent the true probability distribution. In other words, the true distribution is unknown to us and with the data that we have available we can only construct an empirical distribution $\hat{\mathbb{P}}$. We define the empirical distribution by $\hat{\mathbb{P}} =: \frac{1}{N} \sum_{i=1}^{N} \delta_{\hat{\xi}_i}$, where $\{\hat{\xi}_i\}_{i \in N}$ are the N observed samples.

These two problems lead us to evaluate Distributionally Robust Chance Constrained Programs (DRCCPs) as an alternative approach. In this method, we focus to find a solution that satisfies the constraints with high probability for the worst case realization of a wider family of distributions, namely the *ambiguity set*. Literature works propose two distinctive methods to form ambiguity sets. One approach is to construct moment-based ambiguity sets with distributions that have the same mean and covariance as the uncertain parameters. With this approach, we need to calculate the mean and the covariance of the uncertain parameter through the empirical distribution. We employ moment-based ambiguity sets in power systems implementations and OPF problems that incorporate RES. The second approach is to construct the ambiguity set by including distributions that are close to our empirical distribution in terms of a specific distance function. These functions can be the Prohorov metric, the Kullback–Leibler divergence, the ϕ-divergence or the Wasserstein metric. In this

thesis we focus on Wasserstein ambiguity sets, since they offer significant out-of-sample performance, asymptotic guarantees and we can also implement them in various OPF problems [8, 12, 14, 19, 23, 24].

Wasserstein ambiguity sets resemble a ball in space of probability distributions. In the center of this ball we place the empirical distribution and we expand the radius by adjusting the aforementioned probability distance metric. If we increase the radius, we capture more distributions that match to the distribution of the center. On the other hand, if the radius is set to zero, then the ambiguity set includes only the empirical distribution that we construct from the available samples. However, there must be a trade-off when we define the ambiguity set. More specifically, the radius should be big enough to include the true distribution with high confidence, but small enough in order exclude pathological distributions [19]. Below we present the mathematical representation of DRCCP:

$$\begin{aligned} \min_{x \in X} \quad & c^T x \\ \text{s.t.} \quad & \sup_{\mathbb{P} \in \mathbb{B}_\theta} \mathbb{P}\left(F(x, \xi) \leq 0\right) \geq 1 - \alpha \end{aligned}$$

where F is the constraints function, $\xi \in \Xi$ is the uncertainty parameter in the constraints, $\mathbb{B}_\theta$ represents the ambiguity set and a is the violation parameter.

Synopsis of Chapter 2

The main objective of this chapter was to get an insight on the different optimization methods that can handle constraints depending on uncertain parameters. We analyzed the Robust and the Chance Constraints optimization methods and we referred also to the Scenario Approach method. However, since the aforementioned optimization techniques are not suitable in case studies where the amount of data is limited and the probability distribution of the uncertain parameters is not known, we focus on Distributionally Robust Chance Constrained Programming. In DRCCP, we search for solutions where the constraints are feasible with high confidence not only for the empirical distribution that we have from our data, but for a wider family of distributions. We call this family of distributions *ambiguity set*. The ambiguity set will assist us in the following chapters to capture the uncertainties that are derived from the unpredictable nature of the PV generation and the residential demand, which lead to overvoltages in the network.

Chapter 3

Wasserstein Distributionally Robust Chance Constraint Programming

In the previous chapter, we explained the DRCCP method in order to solve optimization problems that incorporate constraints with uncertain parameters. During the analysis of the DRCCP, we mentioned the various probability distance functions that we have available in order to construct the ambiguity set. Since in this report we focus mainly on the Wasserstein distance metric, the purpose of this chapter is to present step by step the theoretical parts of the DRCCP over Wasserstein ambiguity sets.

The remainder of this chapter is organized as follows. Initially, in Section 3-1 we introduce the Wasserstein distance theory in order to build our ambiguity set and then in Section 3-2 we examine the way to determine the radius of the ambiguity set. In Section 3-3 we present the DRCCP method, in Section 3-4 we formulate the convex DRCCP and in Section 3-5 we propose the problem reformulation according to the affine class of constraints function.

3-1 Wasserstein ambiguity sets

In this section we intend to construct the ambiguity set that we mentioned for the first time in Section 2-4. The ambiguity set consists of a family of distributions, close to the empirical distribution, that we obtain from the recorded data. We use this ambiguity set to solve the CCP of Section 2-1 and we call this class of optimization problems Distribunally Robust Chance Constraint Programming (DRCCP). In DRCCP, the constraints are feasible with high confidence for the worst-case distribution of the ambiguity set.

Initially, we define the Wasserstein distance as a metric that we utilize to construct the ambiguity set. The Wasserstein distance emerges from the Optimal Transport Theory, where our goal is to move a mass from one probability distribution to another probability distribution with the cheapest possible way. Figure 3-1 illustrates the Optimal Transport concept, where we have two distributions X and Y and our purpose is to move all the mass from distribution X to distribution Y, with the most profitable transportation cost.

Figure 3-1: In Optimal Transport Theory we intend to move all the mass from distribution X to distribution Y with the most economical transportation cost. [25]

Let $\mathcal{B}(\Xi)$ and $\mathcal{P}(\Xi)$ correspond to the Borel σ-algebra and the set of Borel probability measures on the set Ξ, respectively. If we denote by $\mathcal{P}_p(\Xi) \subseteq \mathcal{P}(\Xi)$ the set of Borel probability measures with finite p-th moments, with $p > [1, +\infty)$, we define the type-p Wasserstein distance between two distributions $\mathbb{Q}, \mathbb{Q}^{'}$, supported on the set $\Xi \subseteq \mathbb{R}^d$, as:

$$W_p(\mathbb{Q}, \mathbb{Q}^{'}) := \left(\inf_{\pi \in \Pi(\mathbb{Q}, \mathbb{Q}^{'})} \int_{\Xi^2} ||\xi - \xi^{'}||^p \pi(d\xi, d\xi^{'}) \right)^{\frac{1}{p}} \tag{3-1.1}$$

The term Π is the set of all joint probability distributions of ξ and $\xi^{'}$ with marginals $\mathbb{Q}$ and $\mathbb{Q}^{'}$, respectively and π is a transportation plan for moving a mass from distribution $\mathbb{Q}$ to another distribution $\mathbb{Q}^{'}$. Thus, the Wasserstein distance between $\mathbb{Q}$ and $\mathbb{Q}^{'}$ represents the cost of an optimal mass transportation plan, where the norm $|| \cdot ||$ encodes the transportation costs of moving unit mass from ξ to $\xi^{'}$. The Wasserstein distance is non-negative, symmetric and goes to zero whenever $\mathbb{Q} = \mathbb{Q}^{'}$ [26].

Except from the primal problem that we introduced in (3-1.1), the dual problem of the most efficient transportation cost between two distributions $\mathbb{Q}, \mathbb{Q}^{'}$ is the following:

$$\begin{aligned} &\sup \left(\int_{\Xi} \psi(\xi^{'}) \mathbb{Q}^{'}(d\xi) - \int_{\Xi} \phi(\xi) \mathbb{Q}(d\xi) \right) \\ &\text{s.t.} \quad \phi \text{ and } \psi \text{ are bounded continuous functions on } \mathbb{R}^d \text{ with} \\ &|\psi(\xi) - \phi(\xi^{'})| \leq ||\xi - \xi^{'}||^p \quad \forall \xi, \xi^{'} \in \mathbb{R}^d \end{aligned}$$

In other words, the dual problem corresponds to a profit maximization problem, where the owner of a mass decides to hire a third party company to change the position of a mass from distribution $\mathbb{Q}$ to distribution $\mathbb{Q}^{'}$, instead of doing it on his own. Then, the third party company buys a mass at the origin ξ with cost $\phi(\xi)$ and sells the mass at the final destination place $\xi^{'}$ with cost $\psi(\xi^{'})$. The presence of the constraints reassures us that the owner prefers to use a third party company to relocate the mass from the origin ξ to the destination $\xi^{'}$, instead of moving the mass on his own paying transportation cost $||\xi - \xi^{'}||^p$ [26].

In case $p = 1$, we simplify the dual problem to

$$\sup_{\phi \in \mathcal{L}} \left(\int_{\Xi} \phi(\xi) \mathbb{Q}(d\xi) - \int_{\Xi} \phi(\xi^{'}) \mathbb{Q}^{'}(d\xi^{'}) \right),$$

where $\mathcal{L}$ represents the space of all Lipschitz functions with $|\phi(\xi) - \phi(\xi^{'})| \leq ||\xi - \xi^{'}||$, $\forall \xi, \xi^{'} \in \Xi$.

Now that we defined the Wasserstein distance metric, we create the Wasserstein ambiguity set. The ambiguity set resembles a ball and at the center of the ball we place the empirical distribution that we construct from the available data. We denote the empirical distribution by $\hat{\mathbb{P}}_N := \frac{1}{N}\sum_{i=1}^{N}\delta_{\hat{\xi}_i}$, with N the number of the available data that belong to the set $\hat{\Xi}_N := \{\hat{\xi}_i\}_{i\in N}$. The mathematical representation of the Wasserstein ambiguity set is as follows:

$$\mathbb{B}_\theta(\hat{\mathbb{P}}_N) := \Big\{\mathbb{Q} \ \in \mathcal{P}_p(\Xi) \ : \ W_p(\hat{\mathbb{P}}_N, \mathbb{Q}) \leq \theta\Big\}$$

which expresses the Wasserstein ball of radius θ centered at the empirical distribution, and includes all distributions that are at maximum θ close to $\hat{\mathbb{P}}_N$ [19].

3-2 Determination of the Wasserstein Radius θ

The radius of the Wasserstein ball plays a significant role in the performance of the DRCCP. As we explained in Section 2-4, there must be a balance between how big and how small this radius should be. In particular, the Wasserstein ambiguity set should be big enough to include the true distribution with high confidence, but small enough in order exclude pathological distributions.

Initially, the radius of the Wasserstein ball depends on the number of the available samples N. More information is available to the decision-maker when there are many samples available, which results in smaller radii and less conservative solutions. Therefore, the radius θ, or $\theta(N)$, decreases as we increase the number of available samples and it bounds $W_p(\hat{\mathbb{P}}_N, \mathbb{P})$ from above, where $\mathbb{P}$ the unknown true probability distribution [14].

We assume $\mathbb{P}$ is light-tailed, namely its tail decays at exponential rate with $A := \mathbb{E}_{\mathbb{P}}\left[\exp\left(||\xi||^a\right)\right] < \infty$ and $a > 1$. Then, we know that the probability of the ambiguity set $\mathbb{B}_\theta$ with radius θ to contain the true distribution $\mathbb{P}$ goes exponentially towards unity for N samples. Therefore, we can determine the radius of the Wasserstein ball that includes the distribution $\mathbb{P}$ with confidence $1-\beta$, $\beta \in (0,1)$ as below:

$$\theta_N(\beta) = \begin{cases} \left(\frac{\log\left(c_1\beta^{-1}\right)}{c_2 N}\right)^{1/\max\{m,2\}}, & \text{if } N \geq \frac{\log\left(c_1\beta^{-1}\right)}{c_2} \\ \left(\frac{\log\left(c_1\beta^{-1}\right)}{c_2 N}\right)^{1/a}, & \text{if } N < \frac{\log\left(c_1\beta^{-1}\right)}{c_2} \end{cases}$$

where, $m \neq 2$, $\xi \in \mathbb{R}^m$ and c_1, c_2 are positive constants that depend on a, A and m [19],[23].

3-3 Distributionally robust chance constraint program over Wasserstein ambiguity sets

In this section we combine all the aforementioned knowledge to present the Distributionally Robust Chance Constraint Program (DRCCP) over Wasserstein ambiguity sets. The DRCCP with Wasserstein ambiguity sets are NP-Hard problems and the feasibility set of DRCCPs is non-convex in most cases [12]. This fact results from the possible non-convex feasibility set of CCP. Therefore, we exploit the CVaR approximation of CCP that we defined in Section 2-2, which transforms the original DRCCP into a convex approximation program.

For the remainder of the thesis, the following assumption holds:

Assumption 1. *The set Ξ is a subset of $\mathbb{R}^d$ and . The function F satisfies:*

(i) for every $\xi \in \Xi$, $x \mapsto F(x, \xi)$ is convex on X

(ii) for every $x \in X$, $\xi \mapsto F(x, \xi)$ is bounded on Ξ

If we denote the empirical distribution by $\hat{\mathbb{P}}_N := \frac{1}{N}\sum_{i=1}^{N} \delta_{\hat{\xi}_i}$, with N the number of the available data that belong to the set $\hat{\Xi}_N := \{\hat{\xi}_i\}_{i \in N}$, the mathematical representation of the DRCCP with Wasserstein ambiguity sets problem with is the following:

$$\begin{aligned} &\min_{x} \; c^T x \\ s.t. \quad &\sup_{\mathbb{P} \in \mathbb{B}_\theta} \mathbb{P}\left((F(x,\xi) > 0)\right) \leq \alpha \end{aligned}$$

where $x \in \mathbb{R}^n$ and $F : \mathbb{R}^n \times \Xi \to \mathbb{R}$ the component-wise maximum function of K constraints.

The feasibility set of this optimization problem may be non-convex. Therefore in the next sections we leverage the CVaR reformulation in order to get a tractable representation of the DRCCP with Wasserstein ambiguity set.

3-4 CVaR approximation of DRCCPs over Wasserstein ambiguity sets

The following expression describes the CVaR approximation of the DRCCP over Wasserstein ambiguity set :

$$\begin{aligned} &\min_{x,t} \; c^T x \\ s.t. \quad &\sup_{\mathbb{P} \in \mathbb{B}_\epsilon} \inf_{t \in \mathbb{R}} \left(\mathbb{E}_{\mathbb{P}} \left((F(x,\xi) + t)_+ \right) - t\alpha \right) \leq 0 \end{aligned} \tag{3-4.1}$$

According to [12] (Lemma IV.2) and [27] (Corollary 37.3.2), the inf and sup in the constraint above can interchange and we can write:

$$\sup_{\mathbb{P} \in \mathbb{B}_\epsilon} \inf_{t \in \mathbb{R}} \left(\mathbb{E}_{\mathbb{P}} \left((F(x,\xi) + t)_+ \right) - t\alpha \right) = \inf_{t \in \mathbb{R}} \sup_{\mathbb{P} \in \mathbb{B}_\theta} \left(\mathbb{E}_{\mathbb{P}} \left((F(x,\xi) + t)_+ \right) - t\alpha \right)$$

By using the sup-inf equality above and the dual representation of the Wasserstein distance that we presented in 3-1 and it is further analyzed in [12] (Section II.B), we can construct the convex reformulation of optimization problem (3-4.1) as:

$$\begin{aligned} & \min_{x,\lambda,s_i,t} c^T x \\ s.t. \quad & \lambda\theta^p + \frac{1}{N}\sum_{i=1}^{N} s_i \leq t\alpha \\ & s_i \geq \sup_{\xi\in\Xi} \left(F(x,\xi) + t - \lambda||\xi - \xi^{'}||^p\right), \quad \forall i \ \in \ N, \\ & \lambda \geq 0, t \ \in \ \mathbb{R}, \ x \ \in \ X, \ s_i \geq 0, \ \forall i \ \in \ N \end{aligned} \tag{3-4.2}$$

3-5 Reformulation of DRCCP for F piecewise affine in uncertainty

The minimization problem (3-4.2) is a general formulation which covers a wide variety of classes of the constraints functions F. However, if we assume known the properties of F on ξ, we can obtain more precise reformulations of (3-4.2). In particular, we develop different reformulations, when we assume F to be piecewise affine in ξ, nonconvex in ξ and convex in ξ. In this section we analyze the precise representation of (3-4.2) for piecewise affine F in ξ and we neglect the rest cases, as they are out of the scope of the thesis.

In particular, let $\Xi = \{\xi \in \mathbb{R}^d | \ C\xi \leq h\}$ be compact, with $C \in \mathbb{R}^{q\times d}$ and $h \in \mathbb{R}^q$ for some $q > 0$. Moreover, we write the constraints function F as $F(x,\xi) := \ \max_{k\leq K} \left(x^T A_k\xi + B_k\xi + b_k(x)\right)$ for a $K > 0$, with $A_k \in \mathbb{R}^{n\times d}$ and $b_k : \mathbb{R}^n \to \mathbb{R}$ convex functions for all $k \in K$. We denote the equivalent of (3-4.2) for a piecewise affine F in ξ as:

$$\begin{aligned} & \min_{x,\lambda,s_i,t} c^T x \\ s.t. \quad & \lambda\theta + \frac{1}{N}\sum_{i=1}^{N} s_i \leq t\alpha \\ & \left(b_k(x) + t + \left(x^T A_k + B_k - C^T\eta_{ik}\right)^T \hat{\xi}_i + \eta_{ik}{}^T h\right)_+ \leq s_i, \\ & ||x^T A_k + B_k - C^T\eta_{ik}|| \leq \lambda, \quad \eta_{ik} \geq 0 \\ & x \in \ X, t \in \mathbb{R}, \lambda \geq 0 \end{aligned} \tag{3-5.1}$$

In the aforementioned optimization problem, the inequality involving the set of variables η_{ik} holds for $i \in N$ and $k \in K$. If the support of the uncertain parameters is unknown or unbounded, then we set $C = 0, h = 0$ and we neglect the variables η.

Proof. We base our proof on work [12]. For piecewise maximum of affine functions of the form $F(x,\xi) := \max_{k\leq K} \left(x^T A_k\xi + B_k\xi + b_k(x)\right)$, we have:

$$
\begin{aligned}
s_i &\geq \left(\sup_{\xi \in \Xi} \left(\max_{k \in K} \left(x^T A_k \xi + B_k \xi + b_k(x) \right) + t - \lambda ||\xi - \hat{\xi}_i|| \right) \right)_+ \\
&= \max_{k \in K} \left(b_k(x) + t + \sup_{\xi \in \Xi} \left(x^T A_k \xi + B_k \xi - \lambda ||\xi - \hat{\xi}_i|| \right) \right)_+ \\
&\geq \left(b_k(x) + t + \sup_{\xi \in \Xi} \left(x^T A_k \xi + B_k \xi - \lambda ||\xi - \hat{\xi}_i|| \right) \right)_+ \qquad (3\text{-}5.2)
\end{aligned}
$$

for every $k \in K$. The second equality interchanges the sup and the max. Then we compute

$$
\begin{aligned}
&\sup_{\xi \in \Xi} \left(x^T A_k \xi + B_k \xi - \lambda ||\xi - \hat{\xi}_i|| \right) \\
&\overset{(a)}{=} \sup_{\xi \in \Xi} \left(x^T A_k \xi + B_k \xi - \sup_{||z_{ik}|| \leq \lambda} z_{ik}^T \left(\xi - \hat{\xi}_i \right) \right) \\
&\overset{(b)}{=} \sup_{\xi \in \Xi} \left(x^T A_k \xi + B_k \xi + \inf_{||z_{ik}|| \leq \lambda} z_{ik}^T \left(\hat{\xi}_i - \xi \right) \right) \\
&= \sup_{\xi \in \Xi} \left(\inf_{||z_{ik}|| \leq \lambda} \left(x^T A_k \xi + B_k \xi + z_{ik}^T \left(\hat{\xi}_i - \xi \right) \right) \right) \\
&= \inf_{||z_{ik}|| \leq \lambda} \left(z_{ik}^T \hat{\xi}_i + \sup_{\xi \in \Xi} \left(\left(x^T A_k + B_k - z_{ik}^T \right) \xi \right) \right) \\
&\overset{(c)}{=} \inf_{||z_{ik}|| \leq \lambda} \left(z_{ik}^T \hat{\xi}_i + \inf_{\substack{\eta_{ik} \geq 0, \\ z_{ik} = x^T A_k + B_k - C^T \eta_{ik}}} \eta_{ik}^T h \right) \\
&= \inf_{\substack{||z_{ik}|| \leq \lambda, \eta_{ik} \geq 0 \\ z_{ik} = x^T A_k + B_k - C^T \eta_{ik}}} \left(z_{ik}^T \hat{\xi}_i + \eta_{ik}^T h \right) \\
&= \inf_{\substack{\eta_{ik} \geq 0 \\ ||x^T A_k + B_k - C^T \eta_{ik}|| \leq \lambda}} \left(\left(x^T A_k + B_k - C^T \eta_{ik} \right)^T \hat{\xi}_i + \eta_{ik}^T h \right) \qquad (3\text{-}5.3)
\end{aligned}
$$

Here, (a) uses the definition of the norm, (b) uses the inf-sup interchange according Lemma IV.2 in [12] and the corollary 37.3.2 in [27] and (c) uses the dual form of the inner linear program from $\Xi = \{\xi \in \mathbb{R}^d | C\xi \leq h\}$. We substitute Equation (3-5.3) to Equation (3-5.2) and we get:

$$
s_i \geq \left(b_k(x) + t + \inf_{\substack{\eta_{ik} \geq 0 \\ ||x^T A_k + B_k - C^T \eta_{ik}|| \leq \lambda}} \left(\left(x^T A_k + B_k - C^T \eta_{ik} \right)^T \hat{\xi}_i + \eta_{ik}^T h \right) \right)_+ , \forall k \in K \qquad (3\text{-}5.4)
$$

Equation (3-5.4) holds if and only if there exists $\eta_{ik} \geq 0 \ \forall k \in K$ suth that $\forall k \in K$

$$
\begin{aligned}
&s_i \geq \left(b_k(x) + t + \left(x^T A_k + B_k - C_{ik}^{\eta} \right)^T \hat{\xi}_i + \eta_{ik}^T h \right)_+ , \\
&||x^T A_k + B_k - C^T \eta_{ik}|| \leq \lambda \qquad (3\text{-}5.5)
\end{aligned}
$$

The validation of the "if" part is trivial according to Proposition V.1. of [12]. We examine the "only if" part in two cases $\forall k \in K$, where the inf in (3-5.4) is either fulfilled or not fulfilled. When the inf in (3-5.4) is fulfilled, then (3-5.5) is satisfied. Otherwise the optimal value of inf is $-\infty$, which turns constraint (3-5.4) to be $s_i \geq 0$. As a result, there is a η_{ik} such that the term inside $(\cdot)_+$ is negative in the constraint of (3-5.5) and thus the constraint of (3-5.5) is transformed into $s_i \geq 0$. □

Synopsis of Chapter 3

In this chapter we presented an analysis of the Wasserstein Distributionally Robust Chance Constraint Programming (DRCCP). We explained the principles of the Wasserstein ambiguity sets and we mentioned the procedure to determine the Wasserstein radius according to the available data. We introduced the mathematical perspective of the DRCCP over Wasserstein ambiguity sets, then we analyzed the CVaR approximation of DRCCPs for a convex reformulation and finally we examined the reformulation of DRCCPs according to the affine class of the constraints.

Chapter 4

Power Systems and Optimal Power Flow

The sustainable transition from conventional fuel sources to green energy has been settled as a national objective to many countries around the world. Global warming, environmental awareness and financial incentives encourage many individuals to establish their own large or small scale PV systems. Regarding the peak demand, PV systems prioritize among conventional energy sources, since they their advantages of zero marginal costs and emissions classify them as an efficient and green way of power production [28].

The purpose of this chapter is to get insights into the fundamentals of the energy grid and more specifically into the structure of the distribution energy networks. The remainder is organized as follows. In Section 4-1 we analyze the principles of the energy grid and we specialize in energy distribution systems. In Section 4-2 we explain the reasons for power curtailment of PV units, in Section 4-3 we present the nonlinear mathematical model of distribution energy networks and finally, in Section 4-4 we linearize our model.

4-1 Energy Grid

This section introduces energy systems and the main elements that structure energy networks. We divide energy networks into three levels, namely the transmission network, the subtransmission network and the distribution network. These networks are responsible to deliver electrical energy from power plants to consumers and they are structured with towers, wires, underground cables, substations and transformers. Figure 4-1 illustrates the energy grid.

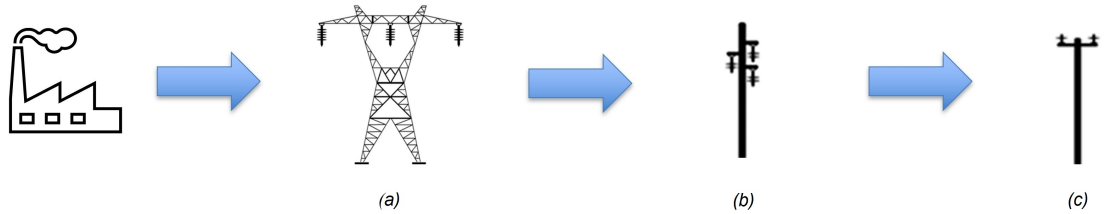


Figure 4-1: The energy grid (a) Transmission, (b) Subtransmission, (c) Distribution networks [29].

More specifically, transmission networks are in charge of transmitting large quantities of power over long distances with minimum power loss. They operate at high voltage (400 kV or 225 kV) and they are responsible to interconnect adjacent countries and to transport electricity from large generators (nuclear, hydro-electric and thermal plants) to subtransmission networks. Then, subtransmission networks provide the electricity they receive to large industrial consumers and to regional distribution networks. The last part of the energy grid consists of the distribution networks that operate at medium or low voltage (20 kV and 400 V), in order to supply with electricity small businesses or residential consumers [30].

Since the scope of this report does not consider transmission or subtransmission networks, we proceed to a further analysis of distribution networks. We sub-divide distribution networks into three structural types, namely radial, loop, and meshed topologies. We illustrate these types in Figure 4-2.

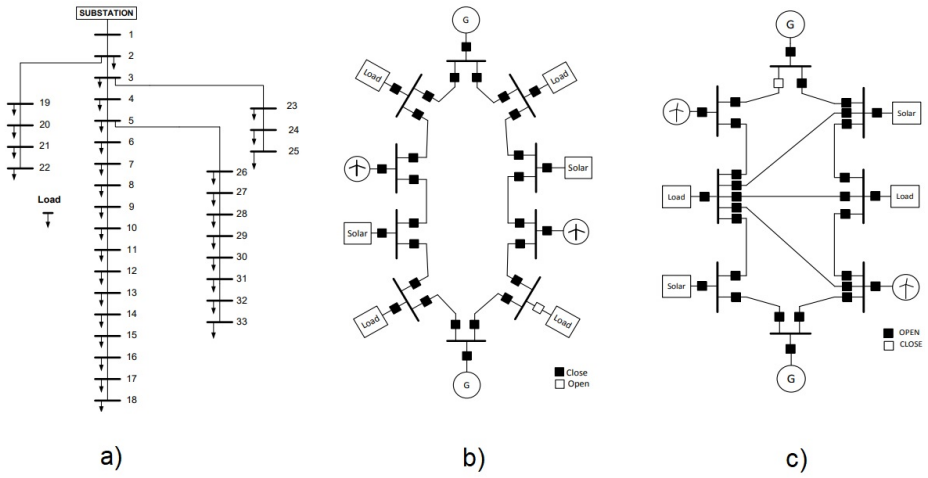


Figure 4-2: Types of Distribution Systems (a) Radial, (b) Loop, (c) Meshed. The radial topology is the most economical type of distribution networks and it has a tree shape. The loop topology is more expensive, but it is more reliable and the meshed topology consists of multiple loop topologies [31].

The first type of distribution systems we examine is the radial distribution system. A radial distribution system has a tree shape, where there are no closed loops and power is delivered via a unique transformer that connects the distribution network with the subtransmission network. When a failure occurs, then the problem affects the whole feeder and the customers connected there. However, it is easier to trace the problem because there is only one path that connects the consumers to the transformer. Moreover, we meet the radial distribution

types almost everywhere due to their affordable cost, and especially in Netherlands, all the low voltage networks are radial.

The second type of distribution systems we examine is the loop structure. In the loop structure, all the nodes are connected to each other in a way that they form a closed loop. Moreover, one or more than one transformers are connected to the network, which supply the nodes with electricity they receive from subtransmission networks. With the loop design of the network, the customers may receive electricity from either directions. This type of systems is more reliable than the radial systems, because the supply will not stop in case of a failure in the lines. Unlike the radial networks, if there is a disruption in the lines, the consumers will continue to receive power from the other direction of the loop. However, these systems are expensive, because they need more switches and conductors during construction.

The third type of distribution networks are the meshed systems. We use this type highly dense areas. Their design corresponds to multiple interconnected loop systems where a customer could receive power by multiple sources. The complexity increases the cost of the construction, but the main advantage of these systems is their high reliability [31, 32, 33].

4-2 Power Curtailment

Power curtailment of PV exports is a method that we investigate in order to mitigate the increased phenomena of voltage rise and congestion. The integration of unpredictable PV units may degrade the network's reliability, since energy distribution systems were primarily designed to perform one-way power flow, from the distribution substation to the consumption users.

The aforementioned problems lead DNOs to take active measures by limiting the residential PV exports to the network. The operators determine this limitation in kilowatts, or as a percentage of the total production and the residential PVs do not export more than a fraction of their installed capacity. Through power limitation, we manage to achieve voltage regulation and consequently to avoid voltage rise in the network [28]. Since this limitation affects the PV owners and their profit, there must be a trade-off during the determination of the power curtailment. This trade-off should consider on the one side the PV investors and on the other side the impacts on the grid during an overvoltage instance [34].

The main problem that power curtailment battles is the mismatch that arises between the demand of the network and the PV generation. This mismatch could be the outcome of two distinctive reasons. Firstly, it could be the result of a power mismatch when the PV generation is high and the capability of the grid to absorb this generation is limited, due to low demand. This phenomenon is very common during midday of summer seasons. Secondly, there could be an energy mismatch due to geographic reasons. In this case, the PV generation is far from the place where the power is consumed. These regions have a significant distance from the load centers and the transmission capacity is poor [28].

Moreover, except from infrastructural reasons, there might be financial incentives for residential PV owners to curtail power of their own establishments. More specifically, PV investors earn money for every MWh they provide to the grid. However, during high PV generation and low demand the price of the MWh is very low, or even negative on the day-ahead market

and on the real-time market Thus, in the future, PV owners may receive incentives to curtail their generation in order to mitigate voltage problems [6].

4-3 Modelling of Distribution Networks

We represent three-phase radial distribution systems as a graph $\Gamma(\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ the set of buses (nodes) and $\mathcal{E}$ the set of lines (edges). The term "bus" in energy systems corresponds to a node of a graph, where a line or several lines are connected and it incorporates several components such as loads and generators. The initial point of the graph is the place where the substation connects the subtransmission system to the distribution system. We call this point *slack node* and we denote it by 0. Moreover, we assume the voltage magnitude at the slack point known and always equal to 1 per unit (p.u.). We denote the rest load nodes by d. In addition, I_0 and I_d translate to the three-phase complex nodal current injection, V_0 and V_d concern the three-phase complex components of the nodal voltages and S_d corresponds to the apparent power. Furthermore, let Y the symmetric three-phase admittance matrix of size $\mathcal{N} \times \mathcal{N}$. In order to construct the admittance matrix, we introduce the impedance of the line between bus i and bus j as z_{ij}. Then the admittance of the line is $y_{ij} = 1/z_{ij} = g_{ij} + jb_{ij}$ and the admittance matrix $Y_{dd} \in \mathbb{C}^{N \times N}$ is as follows:

$$Y_{ij} = \begin{cases} \sum_{l \sim i} y_{il} + y_{ii} & \text{if } i = j \\ -y_{ij} & (i,j) \in \mathcal{E} \\ 0 & (i,j) \notin \mathcal{E} \end{cases}$$

where $l \sim i$ indicates that buses i and j are connected [24]. We use Kirchhoff's laws to describe the relationship between voltage and currents as [35]:

$$\begin{bmatrix} I_0 \\ I_d \end{bmatrix} = \begin{bmatrix} Y_{00} & Y_{0d} \\ Y_{d0} & Y_{dd} \end{bmatrix} \begin{bmatrix} V_0 \\ V_d \end{bmatrix} \tag{4-3.1}$$

Since we can express the current as a function of the apparent power, we use the lower part of Equation (4-3.1) and the apparent power takes the form:

$$S_d = \text{diag}(V_d) I_d^* = \text{diag}(V_d) \left(Y_{dd}^* (V_d)^* + Y_{d0}^* (V_0)^* \right) \tag{4-3.2}$$

We also specify the apparent power as a vector $S_d := [S_1, S_2, \cdots, S_N]$ with rectangular coordinates $S_i = P_i + jQ_i$. Thus, we define the active and reactive power vectors $p = [P_1, P_2, \cdots, P_N]$ and $q = [Q_1, Q_2, \cdots, Q_N]$, respectively.

Equation (4-3.2) is nonconvex in the space of apparent power and bus voltages and therefore we face computation complexity. Hence, in the following section we present an approximation method in order to linearize Equation (4-3.2), such that the voltage magnitude is affine with respect to apparent power.

4-4 Linearization distribution networks

4-4-1 three-phase distribution networks

The purpose of this thesis is to achieve voltage regulation for energy distribution systems with high PV penetration. As a result, we focus to express the limitation in voltage magnitude of each node in a affine representation, with respect to the uncertain terms of PV generation and residential demand. The desired representation has the form of:

$$|V_d| - V_{\max} \leq 0 \leftrightarrow Kx + \mathrm{a} - V_{\max} \leq 0 \tag{4-4.1}$$

Firstly, we examine the three-phase energy distribution systems. We assume $p := \mathfrak{Re}\{S_d\}, q := \mathfrak{Im}\{S_d\}$ and $x := \left(p^T, q^T\right)^T$. We determine the *No Load voltage* as the voltage across the network when the current injections in the buses are zero [36, 37]. The No Load voltage is equal to:

$$w := -Y_{dd}^{-1} Y_{ds} V_0$$

We approximate the voltages in the network as affine equations in the form:

$$V_d = Mx + c$$

where

$$M := \left(Y_{dd}^{-1}\mathrm{diag}(\hat{v}_d)^{*-1}, -jY_{dd}^{-1}\mathrm{diag}(\hat{v}_d)^{*-1}\right)$$
$$c = w$$

Here, the term $\hat{v}_d = |\hat{v}_d|\underline{/\theta}$ corresponds to a predetermined voltage that should be close to nominal voltage of the buses. For the ease of computation, we assume $\hat{v}_d = w$. Moreover, we define $W = \mathrm{diag}(w)$ and we assume that the nominal voltages dominate over the voltage drops. As a result, we express the voltage magnitudes of a three-phase system with the affine approximation of the form [38]:

$$\begin{aligned} & |V_d| = Kx + \mathrm{a} \\ \text{where,} \quad & K := |W|\mathfrak{Re}(W^{-1}M) \\ & x := \left(p^T, q^T\right)^T \\ & \mathrm{a} := |w| \end{aligned}$$

4-4-2 One-phase distribution networks

After analyzing the linearized version of Equation (4-3.2) for a three-phase system, we examine the linearization for one-phase energy distribution systems. For $x := \left(p^T, q^T\right)^T$, our purpose

is again to express the limitation in voltage magnitude of each node in a affine representation in form

$$|V_d| - V_{\max} \leq 0 \leftrightarrow Kx + \mathrm{a} - V_{\max} \leq 0 \leftrightarrow K_1 p + K_2 q + \mathrm{a} - V_{\max} \leq 0 \tag{4-4.2}$$

Let $V_d = \hat{v}_d + \Delta v$, with $\hat{v}_d$ the pre-determined nominal voltage that we mentioned in Subsection 4-4-1 and Δv the deviation from the nominal voltage. We can rewrite Equation (4-4.2) as:

$$S_d = \mathrm{diag}(\hat{v}_d + \Delta v)\,(Y^*_{dd}(\hat{v}_d + \Delta v)^* + Y^*_{ds}(V_s)^*)$$

If we neglect the second-order terms $\mathrm{diag}(\Delta v^t)Y^*(\Delta v^t)^*$ and change the order of terms in Equation (4-3.2), we get:

$$\Lambda \Delta v + \Phi(\Delta v)^* = S_d + \Psi$$

where, $\Lambda := \mathrm{diag}(Y^*_{dd}\hat{v}^*_d + Y^*_{ds}V^*_s)$, $\Phi := \mathrm{diag}(\hat{v}_d)Y^*_{dd}$, $\Psi := -diag(\hat{v}_d)(Y^*_{dd}\overline{v}^* + Y^*_{ds}V^*_s)$.
As nominal voltage we select the no Load Voltage, that we mentioned in Subsection 4-4-1

$$\hat{v}_d = w = -Y^{-1}_{dd}Y_{ds}V_s$$

which gives $\Lambda = 0_{N\times N}$ and $\Psi = 0_N$. Therefore the linearized version of Equation (4-4.2) for one-phase energy distribution systems is:

$$S_d = \mathrm{diag}(\hat{v}_d)Y^*_{dd}(\Delta v)^*$$

and the deviation Δv is:

$$\Delta v = Y^{-1}_{dd}\mathrm{diag}^{-1}(\hat{v}^*_d)(S_d)^*.$$

Then, we define $Y^{-1}_{dd} = (G + jB)^{-1} = Z_R + jZ_I$ and by expanding Δv in rectangular form, we get:

$$\overline{K_1} = \left(Z_R\mathrm{diag}\left(\frac{\cos(\theta)}{|\hat{v}_d|}\right) - Z_I\mathrm{diag}\left(\frac{\sin(\theta)}{|\hat{v}_d|}\right)\right)$$
$$\overline{K_2} = \left(Z_I\mathrm{diag}\left(\frac{\cos(\theta)}{|\hat{v}_d|}\right) + Z_R\mathrm{diag}\left(\frac{\sin(\theta)}{|\hat{v}_d|}\right)\right)$$

which specifies the rectangular matrices $H := \overline{K_1} + j\overline{K_2}$, $J := \overline{K_2} - j\overline{K_1}$ and the coefficient $c = \hat{v}_d$. If we assume again that the nominal voltage dominates over voltage drops, i.e. $\hat{v}_d >> \Delta v$, then we get $V_d = \hat{v}_d + \mathfrak{Re}\{\Delta v\}$ and the linearized coefficients of Equation (4-4.2) for one-phase energy distribution systems become $K_1 := \overline{K_1}$, $K_2 := \overline{K_2}$ and a:= $|\hat{v}_d|$ [24].

Synopsis of Chapter 4

The main goal of this chapter was to present the fundamentals of the energy systems that we use in our OPF formulation. Initially, we introduced the structure of the energy grid and its sub-division into transmission networks, subtransmission networks and then distribution networks. Since the scope of this thesis is engaged with energy distribution networks, we analyzed further the types of distribution networks. Our next step was to examine the reasons of power curtailment, which occur as a result of the high penetration of PVs in energy distribution networks. Finally, we modelled mathematically the energy distribution networks and we presented the affine approximations of voltage magnitudes for three-phase and one-phase systems.

Chapter 5

DRCCP Power Flow

In this chapter we proceed to the simulation part of our work. We describe the steps of our implementation and we present the results of the simulation. The main idea of this chapter is that we do not use the entire dataset that we have available to tune our DRCCP model. From the total 500 scenarios, we employ only 20 scenarios to determine a curtailment factor for every time-instance between 7:00 a.m. and 18:00 p.m with a resolution of 15 minutes. Then, we validate the output of the DRCCP in the whole dataset of 500 days. Hence, we evaluate the DRCCP algorithm on both known days and unseen days and we examine how to construct the ambiguity set in order to incorporate the uncertainty of demand and PV generation.

The remainder of the chapter is organized as follows. Initially in Section 5-1 we present the topology of the distribution energy network that we employ in our case study and we indicate the mathematical representation of our implementation. Next, in Section 5-2 we describe the methods to accelerate the running times of our algorithm by reducing the constraints of the problem. In Section 5-3 we explain thoroughly via a flow chart the steps of the DRCCP algorithm that we implement and finally, in Section 5-4 we present the simulation results.

5-1 Problem formulation

The distribution system that we employ for the purposes of our simulation is a three-phase imbalanced Dutch radial distribution network with high PV penetration. The network consists of 184 buses in total. The 96 buses correspond to connecting network nodes with zero demand and no PV installations and 87 buses are residential consumers with specific demand and PV generation. Additionally, there is one slack bus that we denote it by 0. Figure 5-1 illustrates the topology of the network, where the connecting network nodes have blue color and the houses have green color.

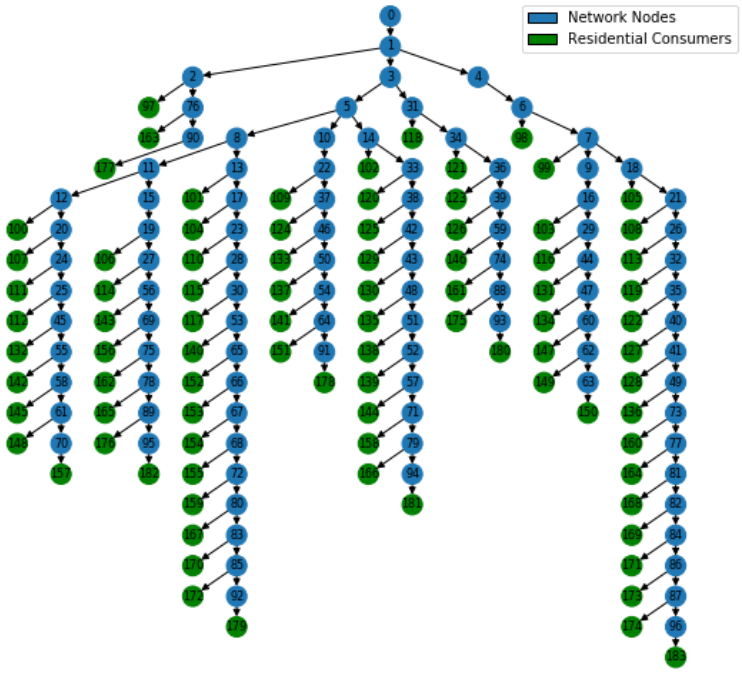


Figure 5-1: The network to implement the DRCCP algorithm. This network is part of the Dutch network and consists of 1 slack bus, 96 connecting buses with no demand and PV generation and 87 residential buses with PV installations.

The distribution energy system operates at 400 Volt, which is also the base to our per unit (p.u.) transformations. The voltage magnitudes in the network should be below the highest operation limit, $\mathrm{V}_{\max} = 1.05\ p.u.$ [39]. The base apparent power that we use is 100 KVA. The power factor $\cos(\phi)$ of the buses, or alternatively the relation between active and apparent power that the residential users consume, is equal to 0.95.

Our goal is to define the minimum curtailment factor, a_c, at every time-instance that limits the active power the PVs export to the network. This curtailment factor will be the same for all PV units of the network, in order to achieve the desired voltage regulation. As a result, we intend to create a stable network without voltage rises caused by the mismatch between PV exports and residential demand. However, since our purpose is to avoid overconservative solutions and high-cuts in the generation, we are flexible to some minor overvoltage instances.

The uncertain parameters of our problem are the active generation of the PVs, p_{PV} and the active demand consumed by the buses, p_d. We denote these two random parameters by ξ_1 and ξ_2, respectively. Additionally, we assume zero reactive power generation by the PV units and we compute the reactive power consumed by the buses as a relation of the active power through the expression $\frac{Q}{P} = \tan(\phi)$. The following table presents the variables that we use in our formulation and their representation during the optimization.

	Variables	Variables in the optimization
Curtailment factor	a_c	x
PV generation	p_{PV}	ξ_1
Active power demand of a house	p_d	ξ_2
Reactive power demand of a house	$q_d = p_d \tan(\phi)$	$q_d = \xi_2 \tan(\phi)$
Active power injected to the network	$p = p_{PV} - p_d$	$p = \xi_1 - \xi_2$

The available data correspond to 500 i.i.d. scenarios of residential consumption and PV generation from the distribution energy network in Figure 5-1. These scenarios are recordings of demand and PV generation during 500 different summer days. The resolution of these data is 15 minutes and they concern the time window from 6:00 a.m. to 18:00 p.m. Since the power generation from the PVs peaks around midday, we limit the time window to our highest interest from 07:00 a.m. to 18:00 p.m. This time-slot corresponds to 45 recorded time instances. We solve one DRCCP problem at every time-instance and thus in the end of the procedure we obtain 45 distinctive curtailment factors of the PV exports to the grid.

We focus to construct our problem according to DRCCP Formulation (3-5.1), where the constraints are piecewise affine with respect to our uncertainties. From Formulation (3-5.1), we set $C = 0, h = 0$ and we neglect the variables η, as we assume the support of the uncertain parameters unbounded. Furthermore, the constraints function should be in the form $F(x,\xi) := \max_{k\leq K}\left(x^T A_k \xi + B_k \xi + b_k(x)\right)$ and we aim to transform the constraints of voltage magnitudes accordingly. In the following optimization problem, we use the linear approximations of the constraints function that we found in Expressions (4-4.1) and (4-4.2).

$$\begin{aligned} \min_{a_c} \quad & a_c \\ \text{s.t.} \quad & |V_{d_k}| \leq V_{\max} \qquad \text{for } k = 1, 2, ..., K \\ & 0 \leq a_c \leq 1 \end{aligned}$$

where,

$$\begin{aligned} |V_{d_k}| &= K_{1_k} p + K_{2_k} q + \mathrm{a}_k \\ &= K_{1_k} \left[(1 - a_c)\, p_{PV} - p_d\right] + K_{2_k} p_d \tan(\phi) + |\hat{v}|_k \\ &= K_{1_k} \left[(1 - a_c)\, \xi_1 - \xi_2\right] + K_{2_k} \xi_2 \tan(\phi) + |\hat{v}|_k \\ &= K_{1_k} (1 - a_c)\, \xi_1 - K_{1_k} \xi_2 + K_{2_k} \xi_2 \tan(\phi) + |\hat{v}|_k \\ &= -x K_{1_k} \xi_1 + K_{1_k} \xi_1 - K_{1_k} \xi_2 + K_{2_k} \xi_2 \tan(\phi) + |\hat{v}|_k \\ &= x \begin{pmatrix} -K_{1_k} & 0 \end{pmatrix} \xi + \begin{pmatrix} K_{1_k} & 0 \end{pmatrix} \xi - \begin{pmatrix} 0 & K_{1_k} \end{pmatrix} \xi + \begin{pmatrix} 0 & K_{2_k} \tan(\phi) \end{pmatrix} \xi + |\hat{v}|_k \\ &= x \underbrace{\begin{pmatrix} -K_{1_k} & 0 \end{pmatrix}}_{\tilde{A}_k} \xi + \underbrace{\begin{pmatrix} K_{1_k} & K_{2_k} \tan(\phi) - M \end{pmatrix}}_{\tilde{B}_k} \xi + \underbrace{|\hat{v}|_k}_{\tilde{b}_k(x)} \\ &= x^T \tilde{A}_k \xi + \tilde{B}_k \xi + \tilde{b}_k(x) \end{aligned}$$

With K nodes we get the following K constraints

$$|V_{d_k}| - V_{\max} \leq 0 \leftrightarrow x^T \underbrace{\tilde{A}_k}_{A_k} \xi + \underbrace{\tilde{B}_k}_{B_k} \xi + \underbrace{\tilde{b}_k(x) - V_{\max}}_{b_k} \leq 0 \qquad \text{for } k = 1, 2, ..., K \tag{5-1.1}$$

We summarize all constraints above and we represent them by only one constraint, which is equal to $\max_{k=1,2,\ldots,K}\left(x^T A_k\xi + B_k\xi + b_k\right) \leq 0$. When we adopt this representation for the constraints, we consider the violation probability for all constraints simultaneously and not individually. Now, we managed to express all the constraints according to the DRCCP formulation (3-5.1).

5-2 Reducing the constraints of the problem

To reduce the number of constraints and the running time of the DRCCP algorithm, we focus on two of the longest feeders of the distribution network of Figure 5-1. In this way, we achieve to reduce the number of buses and to accelerate the running time of the DRCCP algorithm. In particular, we illustrate the sub-network that we use for the DRCCP implementation in Figure 5-2:

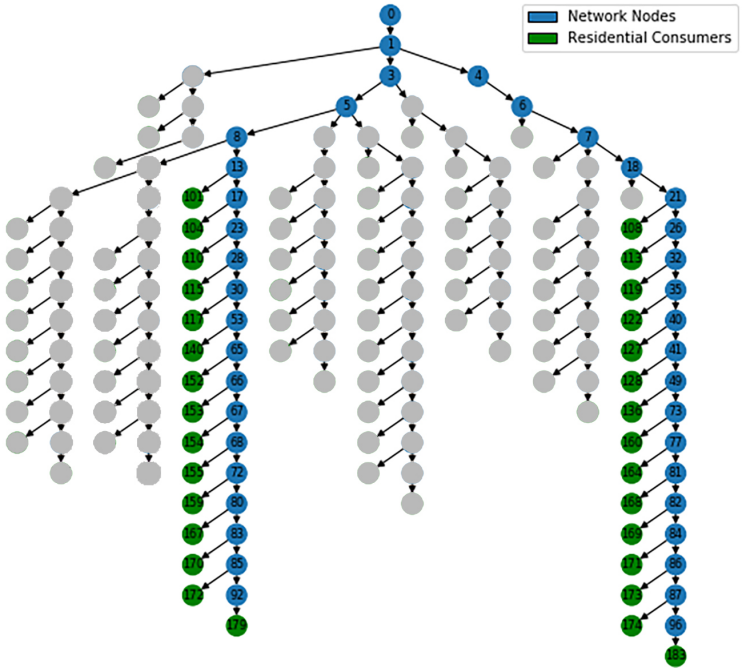


Figure 5-2: The sub-network to implement the DRCCP algorithm. Our sub-network consists now of 1 slack bus, 38 connecting buses with zero demand and PV generation and 30 residential buses with PV installations and demand. We do not consider the part of the original network with grey color.

The sub-network that we apply our formulation on consists of 69 buses. 38 buses correspond to connecting network nodes, 30 are residential consumers with PV installations and one slack bus. We do not consider the buses of the network in Figure 5-2 with grey color. We choose this part of the network, because there are houses relatively far from the slack bus that have higher chance of experiencing overvoltage problems [34]. After focusing only on a part of the network, the empirical distribution lies in space $\mathbb{R}^{204\times500\times45}$ and with less buses compared to the original network of Figure 5-1, our problem includes significantly less constraints. Regarding the empirical distribution, the space $\mathbb{R}^{204\times500\times45}$ is a result of the three-phase

nature of each of the 69 buses, minus the slack bus, the 500 available scenarios, and the 45 time instances of each scenario.

5-3 Data-Driven tuning of the Wasserstein radius

For the implementation of the DRCCP problem we need to tune the Wasserstein radius θ. Our goal is to achieve 95% satisfaction of constraints, or in other words, no more than 5% of our total 500 scenarios to present overvoltages at every time-instance. The 500 samples are a reasonably representative dataset from the distribution. Due to the fact that we solve multiple optimization problems across our time-window, we pick 20 scenarios out of the 500 to form the DRCCP. These 20 scenarios that we draw are the tuning samples of the DRCCP algorithm and they can be considered sufficiently representative, in the sense that some of them will be optimistic sub-samples of the original dataset and some of them will be pessimistic sub-samples of the original dataset. With the term optimistic samples we refer to samples that lead to less curtailment of the PV generation, and as a result, during the validation over the 500 samples we experience high numbers of overvoltage instances. On the other hand, we describe as pessimistic samples the tuning samples that lead to high curtailment of the PV generation through the DRCCP. If we tune the DRCCP algorithm with pessimistic samples, the number of overvoltage violations during the validation process will be small. However, the pessimistic samples will lead us to overconservative solutions. Our purpose is to determine the Wasserstein radius, so that even for the optimistic samples, the system is sufficiently robust to present more than 95% feasibility of constraints when we validate over the entire dataset of 500 scenarios. We pick the minimum such radius to exclude overconservative solutions.

In Section 3-2 we introduced the function $\theta(N)$ to determine the Wasserstein radius of the ambiguity set, regarding the available number of samples N. Considering that choosing apriori the Wasserstein radius through Equation (3-2) will result to conservative solutions and high loss of profit for the PV investors, we follow an alternative heuristic method. We initiate our optimization with $\theta = 0$ and then, we run iteratively the optimization increasing the Wasserstein radius in each step. With this heuristic method, we assume that for the network in Figure 5-2, we can employ the DRCCP algorithm with the same Wasserstein radius θ for different weather conditions and consequently for a different distribution of PV generation. In other words, we only need new recordings of 20 days to perform a DRCCP of the network in Figure 5-2 for a different distribution of PV generation, without having to adjust the Wasserstein radius θ. We base this assumption to the fact that even though the distribution may change, the optimization problem remains the same.

As we mentioned before, we initialize our heuristic algorithm with $\theta = 0$. When we set the Wasserstein radius equal to zero, we solve the DRCCP as if our empirical distribution $\hat{\mathbb{P}}$ is the true distribution and we can rely only on it. However, as previously stated, the true distribution of demand and PV generation is unknown and differs from the empirical distribution $\hat{\mathbb{P}}$. As a result, as we gradually increase the Wasserstein radius, we indicate that we trust less our empirical distribution. We keep increasing the ambiguity set and we validate on the entire set, until we meet the performance goal of the problem, namely the 95% satisfaction of the constraints at every time-instance. The next flow chart of Figure 5-3 depicts all of the preceding steps that we follow when we implement the DRCCP algorithm.

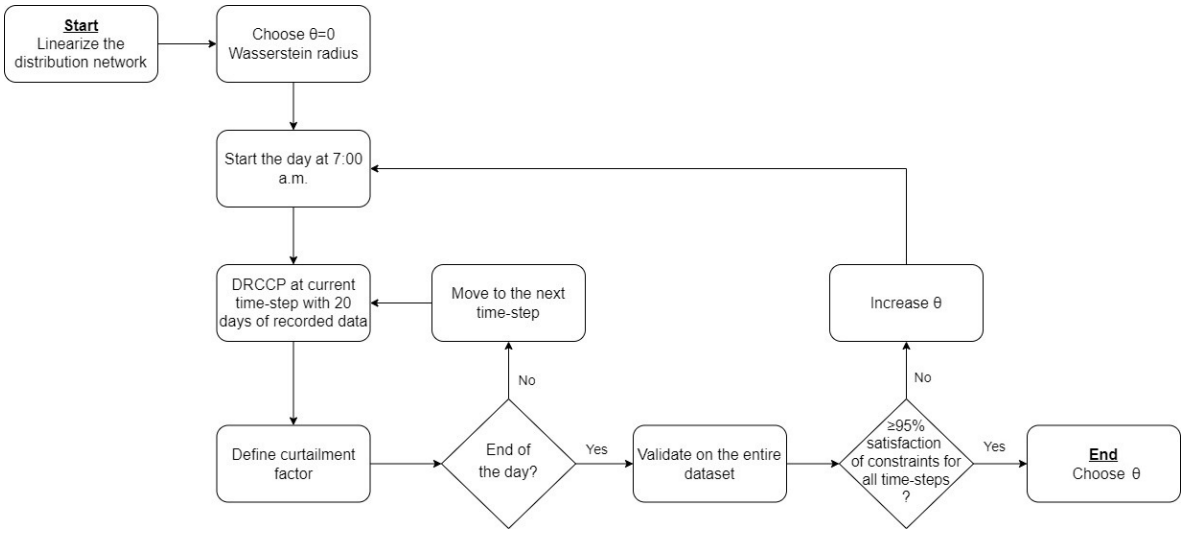
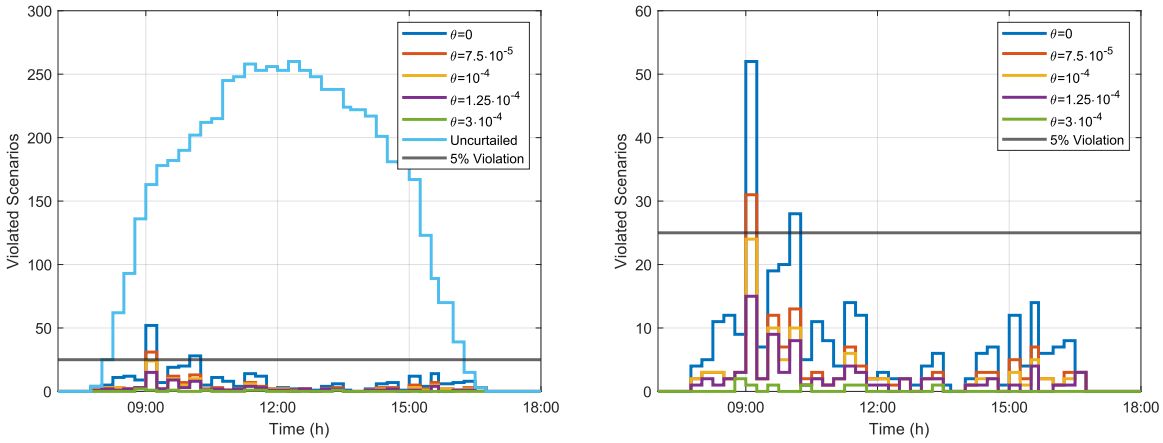


Figure 5-3: Flow chart of the DRCCP implementation. For the DRCCP we need a linearized model of the distribution network. The initial radius of the ambiguity set is $\theta = 0$. We start the DRCCP with $\theta = 0$ at 7:00 a.m. We use 20 days of recorded data for this time-instance and we compute the curtailment factor for that specific time-instance. Then, we move to the next time-instance and we compute again the curtailment factor using 20 days of recorded data. When we have computed the curtailment factor for every time-instance of the day, we validate over the entire dataset of 500 days. If we do not fulfil the design criterion for at least 95% satisfaction of constraints for all time instances, we increase the radius of the ambiguity set and we repeat the whole procedure. Once we fulfil we design criterion and the violated scenarios do not exceed the 5% of the total dataset for all time-instances, we stop the iterations and we do not increase the θ anymore.

5-4 Simulation and Results

5-4-1 Tuning of the Wasserstein radius θ

In this section we present the outcome of our algorithm depicted in the flow chart of Figure 5-3. We set the violation parameter $a = 0.05$ and we investigate the relation of the curtailment factor with the radius of our uncertainty Wasserstein ball. Our design goal is to achieve at least 95% satisfaction of the constraints over the validation data, which corresponds to no more than 25 scenarios out of 500 with overvoltages at each time-step. Figure 5-4 depicts the number of scenarios with overvoltages that occur every time-instance without any curtailment factor and the number of problematic scenarios that preserve after curtailment for different values of θ.



(a) Number of scenarios with overvoltage violations without curtailment and with curtailment for different values of θ.

(b) Number of scenarios with overvoltage violations after curtailment for different values of θ.

Figure 5-4: Violated scenarios without and with curtailment for different values of θ. Without curtailment the number of overvoltage instances is high and hence we need to control the PV output through DRCCP. After curtailment, the number of overvoltage instances drops significantly and depends on the radius of the ambiguity set. The number of violated scenarios decreases monotonically as we increase the radius of our ambiguity set. The black horizontal line indicates the design requirement of maximum 25 scenarios with overvoltage instances. We choose the minimum radius θ that does not exceed the black horizontal line.

Figure 5-4a shows that PV power curtailment is necessary because of the vulnerability of the network to overvoltages. If we do not implement any curtailment, we see in Figure 5-4a that many more than 5% of the 500 scenarios (days) of our dataset are very likely to face overvoltages at many time-instances. Once the curtailment is in place, the days with violation incidents decrease significantly during the entire day and they are highly dependent on the Wasserstein radius, as shown in Figure 5-4b. We indicate with a black horizontal line the allowable 5% of days with overvoltages at all time-instances, which also dictates the minimum Wasserstein radius of the ambiguity set that we choose. We start our iterations with $\theta = 0$ and we observe that we do not fulfil our performance goal, since in two time instances, namely at 9:00 a.m. and at 10:00 a.m., the number of scenarios with overvoltage problems exceeds the black horizontal line. We expected this behaviour, because with $\theta = 0$ we perform the optimization as if the empirical distribution $\hat{\mathbb{P}}$ is the true distribution $\mathbb{P}$. This assumption is not valid and hence we need to increase the Wasserstein radius, in order to incorporate the uncertainty of PV generation and demand. We choose to increase the θ value on each iteration with a step of $1.25 \cdot 10^{-4}$, because with smaller steps the output of the DRCCP algorithm presents only minor changes. On the other hand, adopting a bigger step on each iteration would lead us to miss important information for the intermediate values of θ.

We consider $\theta = 0$ as the lower extreme case for our iterations, regarding the radius of the ambiguity set. In contrast to $\theta = 0$, the DRCCP solution for $\theta = 3 \cdot 10^{-4}$ is the upper extreme case. We observe in Figure 5-4b, that for $\theta = 3 \cdot 10^{-4}$ the violations are almost vanished and we can guarantee a highly stable network with only 20 days of recorded data for the tuning of the DRCCP. As a result, and since we have already rejected $\theta = 0$, all possible values of parameter θ should lie in space $(0, 3 \cdot 10^{-4}]$. In Figure 5-4b, we depict the scenarios

with overvoltages after we implement the DRCCP output for the upper extreme case of θ, the lower extreme case of θ and some intermediate values of θ. The intermediate values are $\theta = 7.5 \cdot 10^{-5}$, $\theta = 10^{-4}$ and $\theta = 1.25 \cdot 10^{-4}$, because the results of these values approach the desired goal the most.

Now, we focus on cases where $\theta = 7.5 \cdot 10^{-5}$, $\theta = 10^{-4}$ and $\theta = 1.25 \cdot 10^{-4}$. In Figure 5-4b, we see that when we assign $\theta = 7.5 \cdot 10^{-5}$, we cannot fulfil the design criterion of maximum 5% violation of the constraints at all time-instances. At 09:00 a.m. the number of violated scenarios exceeds our limit and hence our ambiguity set with radius $\theta = 7.5 \cdot 10^{-5}$ is not adequate to guarantee a network according to our requirements. Therefore, we increase our ambiguity set by one one step and for $\theta = 10^{-4}$ we observe in Figure 5-4b that for the first time we fulfil the design criterion for 95% satisfaction of constraints at every time-instance. According to the flow chart of Figure 5-3, once we meet the design criterion for the first time, we stop the DRCCP iterations. We do not prefer the DRCCP output for $\theta = 1.25 \cdot 10^{-4}$, since it provides a solution with unnecessary PV cuts. Hence, we choose $\theta = 10^{-4}$ as the radius of the ambiguity set, that fulfils the design criterion of our problem, while at the same time it does not provide a conservative solution. If we decrease our θ just for one step, then we violate our design criterion. One the other hand, if we increase θ just for one step and choose $\theta = 1.25 \cdot 10^{-4}$, then we provide a conservative solution. In the following Table 5-1 we provide the exact numbers of the overvoltage scenarios before and after curtailment with different values of θ.

Time instance / Violated Scenarios	7:00	7:15	7:30	7:45	8:00	8:15	8:30	8:45	9:00	9:15	9:30	9:45	10:00	10:15	10:30
Without Curtailment	0	0	0	4	25	62	93	136	163	178	182	190	202	212	21
With Curtailment $\theta = 0$	0	0	0	4	5	11	12	9	**52**	7	19	20	**28**	5	11
With Curtailment $\theta = 7.5 \cdot 10^{-5}$	0	0	0	2	3	3	2	3	**31**	2	12	7	13	2	3
With Curtailment $\theta = 10^{-4}$	0	0	0	2	3	3	2	3	24	2	10	5	10	1	2
With Curtailment $\theta = 1.25 \cdot 10^{-4}$	0	0	0	1	2	1	2	3	15	2	9	3	8	1	2
With Curtailment $\theta = 3 \cdot 10^{-4}$	0	0	0	0	0	0	0	2	1	0	1	0	0	1	0

Time instance / Violated Scenarios	10:45	11:00	11:15	11:30	11:45	12:00	12:15	12:30	12:45	13:00	13:15	13:30	13:45	14:00	14:15
Without Curtailment	245	248	258	253	256	253	260	253	248	238	238	224	222	217	201
With Curtailment $\theta = 0$	8	4	14	12	2	3	2	2	1	4	6	1	0	2	6
With Curtailment $\theta = 7.5 \cdot 10^{-5}$	1	2	7	4	2	2	1	2	0	2	3	0	0	1	3
With Curtailment $\theta = 10^{-4}$	1	2	6	3	2	2	0	2	0	2	2	0	0	1	2
With Curtailment $\theta = 1.25 \cdot 10^{-4}$	1	2	4	3	1	1	0	2	0	2	2	0	0	1	1
With Curtailment $\theta = 3 \cdot 10^{-4}$	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0

Time instance / Violated Scenarios	14:30	14:45	15:00	15:15	15:30	15:45	16:00	16:15	16:30	16:45	17:00	17:15	17:30	17:45	18:00
Without Curtailment	181	180	167	123	89	70	39	15	3	0	0	0	0	0	0
With Curtailment $\theta = 0$	7	1	12	4	14	6	7	8	3	0	0	0	0	0	0
With Curtailment $\theta = 7.5 \cdot 10^{-5}$	3	0	5	2	7	2	1	3	3	0	0	0	0	0	0
With Curtailment $\theta = 10^{-4}$	2	0	3	1	5	2	1	2	3	0	0	0	0	0	0
With Curtailment $\theta = 1.25 \cdot 10^{-4}$	2	0	1	0	4	0	1	1	3	0	0	0	0	0	0
With Curtailment $\theta = 3 \cdot 10^{-4}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 5-1: Violated scenarios during the day before and after the optimized curtailment factor for different radii of the ambiguity set.

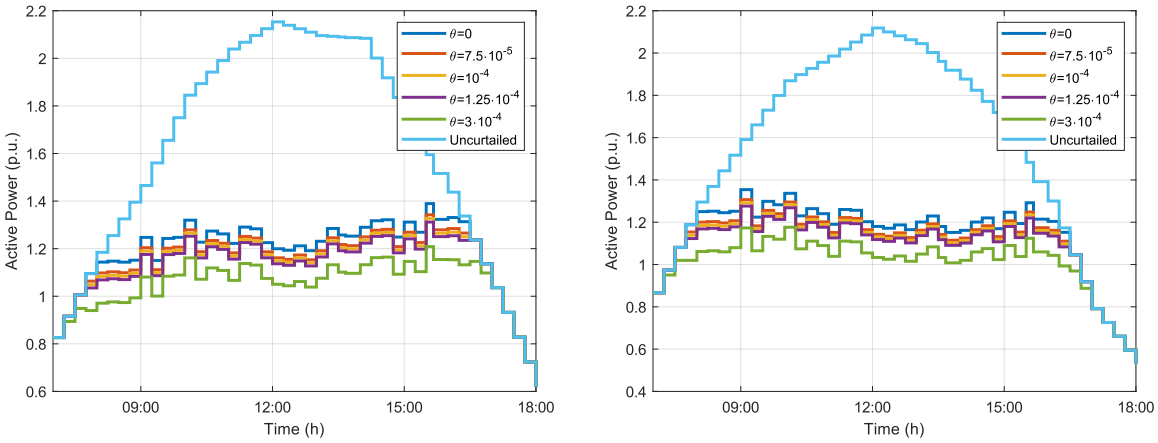
In Table 5-1, we indicate with bold the number of scenarios with overvoltages that preserve after the curtailment and exceed the desired 5% performance limit. If we do not implement any control on the output of the PVs, then the overvoltage instances initiate at 7:45 a.m. and end at 16:30 p.m. The curtailment of PV exports is significant for the stability of the network, because without any control there are 62 overvoltage instances even from 8:15 a.m.

that violate our performance criterion of maximum 25 scenarios with violations at every time-instance.

Once the curtailment is in place, the number of violated scenarios drops. However, we need to tune appropriately the Wasserstein radius that we use in our DRCCP. For the lower extreme case of $\theta = 0$, the overvoltage instances are significantly less at every time-instance, compared to the power flow when no PV control is active. However, with an ambiguity set of radius $\theta = 0$, we cannot capture distributions close to the empirical distribution of the 20 days of recordings and as a result there are two time instances, at 9:00 a.m. and at 10:00 a.m., that violate our design criterion of maximum 25 violated scenarios. Therefore, we need a bigger uncertainty ball to incorporate more uncertainty regarding the empirical distribution. By increasing the Wasserstein radius to $\theta = 7.5 \cdot 10^{-5}$, the violated scenarios at every time step decrease even more. The time-instances that we do not fulfil the design criterion reduce to only one, namely at 9:00 a.m. Furthermore, Figure 5-4b indicates that the number of violated scenarios drops monotonically, as we increase the radius of the ambiguity set. As a result, we increase θ by one step and for $\theta = 10^{-4}$, we observe that for the first time we do not experience more than 25 overvoltage instances at any time-instance. With $\theta = 1.25 \cdot 10^{-4}$ and $\theta = 3 \cdot 10^{-4}$ the overvoltage scenarios decrease even more during the whole day, but the solutions they provide are conservative.

Next, in Figure 5-5 we illustrate how the radius of the ambiguity set affects the PV exports to the network. We use Scenarios 1 and 2 to illustrate the output of our DRCCP optimization, so in Figure 5-5 we examine the behaviour of the algorithm in two days that we used to tune the DRCCP. For the DRCCP algorithm, we use the same Wasserstein radii as in Figure 5-4. For lack of space, we provide the analysis for only two scenarios, but the same principles apply to the remaining 18 known scenarios. We see from Scenarios 1 and 2 of Figure 5-5 that the PV exports after curtailment decrease monotonically, as we increase the radius of our ambiguity set. Thus, a large increase in the Wasserstein radius leads to very conservative solutions and significant PV cuts.

Furthermore, in Figure 5-4b we noticed that with $\theta = 3 \cdot 10^{-4}$ we obtain a highly robust network with a small number of violated scenarios. As a trade-off to the small number of violations, we observe in Figure 5-5 that for that specific value of θ, we curtail a significant portion of PV generation. Moreover, from Scenario 1 and Scenario 2 of Figure 5-5, we notice that without the PV control, the active power that the PVs export peaks around midday, as expected. On the other hand, when we implement curtailment, there is not an explicit peak in the active power during the whole day. The deviations of the active power that the PVs export across the day after the curtailment are small, as they depend on the residential demand. Since in our problem the demand remains low with small deviations between 7:00 a.m. and 18:00 p.m., the PV contribution to the network after curtailment presents similarly only small fluctuations in order to prevent overvoltages.



(a) Scenario 1 **(b)** Scenario 2

Figure 5-5: Uncurtailed PV contribution compared to PV contribution under curtailment with different Wasserstein radii θ for two scenarios. The available PV exports after curtailment do not present any high peaks during the whole day and they decrease monotonically as we increase the Wasserstein radius θ.

Time instance	7:00	7:15	7:30	7:45	8:00	8:15	8:30	8:45	9:00	9:15	9:30	9:45	10:00	10:15	10:30
$\theta = 0$	0	0	0	0	0.035	0.086	0.137	0.175	0.149	0.263	0.249	0.287	0.285	0.352	0.344
$\theta = 7.5 \cdot 10^{-5}$	0	0	0	0.029	0.073	0.121	0.171	0.204	0.178	0.288	0.274	0.311	0.307	0.373	0.364
$\theta = 10^{-4}$	0	0	0	0.042	0.086	0.133	0.182	0.214	0.188	0.296	0.282	0.318	0.314	0.380	0.371
$\theta = 1.25 \cdot 10^{-4}$	0	0	0	0.055	0.098	0.145	0.193	0.224	0.198	0.304	0.290	0.326	0.321	0.387	0.378
$\theta = 3 \cdot 10^{-4}$	0	0.024	0.057	0.142	0.181	0.222	0.265	0.288	0.263	0.359	0.345	0.378	0.371	0.434	0.424

Time instance	10:45	11:00	11:15	11:30	11:45	12:00	12:15	12:30	12:45	13:00	13:15	13:30	13:45	14:00	14:15
$\theta = 0$	0.366	0.401	0.376	0.388	0.423	0.443	0.442	0.430	0.435	0.413	0.386	0.398	0.400	0.381	0.342
$\theta = 7.5 \cdot 10^{-5}$	0.386	0.420	0.395	0.406	0.441	0.461	0.460	0.448	0.453	0.432	0.405	0.417	0.419	0.400	0.362
$\theta = 10^{-4}$	0.392	0.427	0.401	0.412	0.447	0.466	0.466	0.454	0.460	0.438	0.411	0.424	0.425	0.407	0.368
$\theta = 1.25 \cdot 10^{-4}$	0.399	0.433	0.407	0.419	0.453	0.472	0.472	0.459	0.466	0.444	0.417	0.430	0.432	0.413	0.375
$\theta = 3 \cdot 10^{-4}$	0.443	0.477	0.450	0.460	0.494	0.512	0.512	0.500	0.508	0.487	0.459	0.473	0.475	0.457	0.419

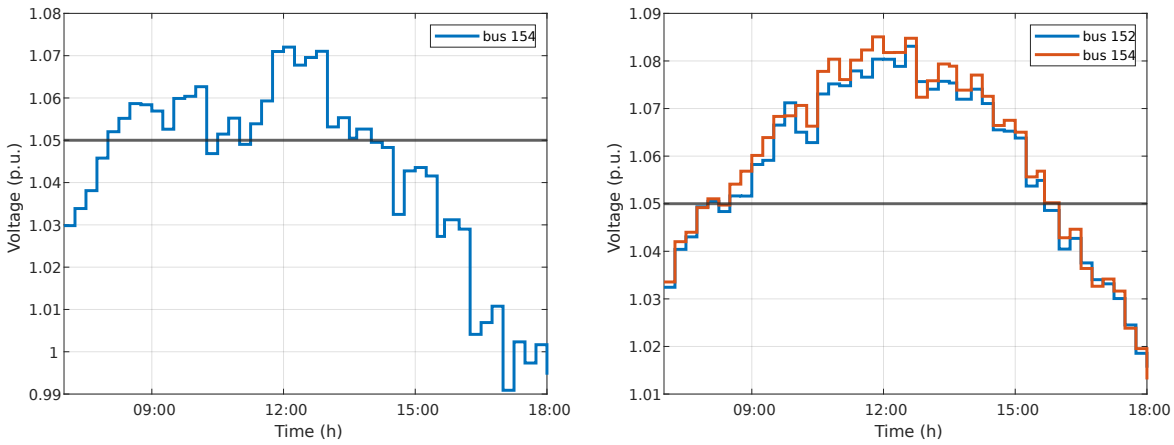
Time instance	14:30	14:45	15:00	15:15	15:30	15:45	16:00	16:15	16:30	16:45	17:00	17:15	17:30	17:45	18:00
$\theta = 0$	0.311	0.320	0.251	0.242	0.129	0.127	0.074	0.018	0	0	0	0	0	0	0
$\theta = 7.5 \cdot 10^{-5}$	0.332	0.341	0.276	0.267	0.158	0.156	0.106	0.053	0	0	0	0	0	0	0
$\theta = 10^{-4}$	0.339	0.349	0.284	0.276	0.168	0.166	0.117	0.065	0	0	0	0	0	0	0
$\theta = 1.25 \cdot 10^{-4}$	0.346	0.356	0.292	0.284	0.178	0.175	0.127	0.077	0	0	0	0	0	0	0
$\theta = 3 \cdot 10^{-4}$	0.393	0.405	0.346	0.341	0.242	0.239	0.198	0.153	0.073	0.034	0	0	0	0	0

Table 5-2: Curtailment factors of PV generation for different radii of the ambiguity set

In Table 5-2 we quantify Figure 5-5 and we observe once more that the more we increase the radius of the ambiguity set, the more we limit the PV exports. Considering that the curtailment factor a_c takes values in space $[0, 1]$, a "0" in Table 5-2 corresponds to no curtailment, while a "1" corresponds to 100% curtailment of the PV contribution to the grid. Moreover, in Table 5-2 we observe that the radius of the ambiguity set determines the first and the last time-instance that the curtailment is active. More specifically, in Table 5-1 we noticed that the first and the last overvoltage instances when we do not apply any curtailment are at 7:45 a.m. and at 16:30 p.m., respectively. Therefore, we expect zero curtailment factor a_c before 7:45 a.m. and after 16:30 p.m. However, when the radius of the ambiguity set is $\theta = 3 \cdot 10^{-4}$,

the curtailment ignites at 7:15 a.m. and ends at 16:45 p.m., which means that due to a big ambiguity set, our overconservative solution starts limiting the PV output too early and stops too late. On the other hand, when we rely only on our empirical data and we set $\theta = 0$, the curtailment starts after the first violation instances at 8:00 a.m. For the intermediate steps of $\theta = 7.5 \cdot 10^{-5}$, $\theta = 10^{-4}$ and $\theta = 1.25 \cdot 10^{-4}$, the curtailment starts when the first overvoltage instances appear, namely at 7:45 a.m. and stops at 16:15 p.m. As a result, we see that for higher values of θ the curtailment process lasts longer, limiting the PV exports when it might not be necessary. On the other hand, for smaller values of θ, the curtailment process initiates in later time-instances.

After choosing $\theta = 10^{-4}$ as the most appropriate radius of our ambiguity set, we mentioned that the power curtailment starts at 7:45 a.m. and ends at 16:15 p.m. according to Table 5-2. We expected this behaviour, since in the 20 recorded days that we insert to the DRCCP algorithm, we face two distinctive scenarios with overvoltage instances at 8:00 a.m., when we do not impose any curtailment. Figure 5-6 depicts the voltage magnitudes of specific buses in Scenario 8 and Scenario 10 and illustrates how overvoltages that take place at 8:00 a.m. evolve over the day when no curtailment is in place.



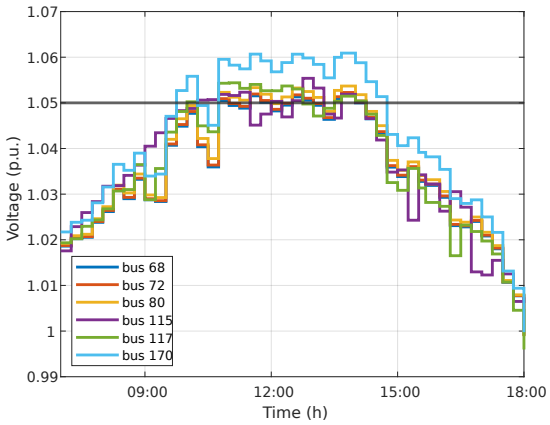
(a) Scenario 8 **(b)** Scenario 10

Figure 5-6: Two scenarios of the tuning dataset that indicate when the first voltage violation occurs, when there is no curtailment. The time-instance of the first overvoltage violation determines the starting time of the curtailment. In scenario 8, the first overvoltage occurs at 8:00 a.m. in bus 154 and in Scenario 10 the first overvoltages occur at 8:00 a.m. in bus 152 and bus 154.

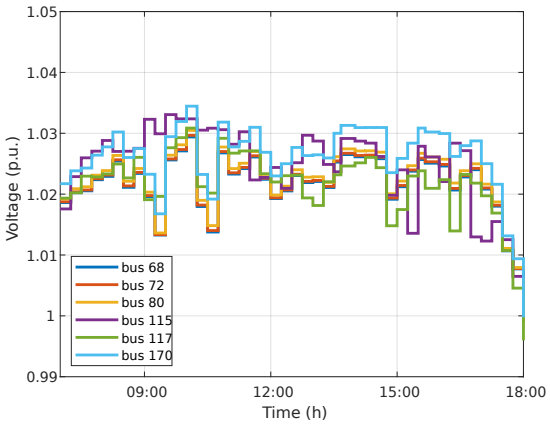
The presence of Scenario 8 and Scenario 10 in the tuning dataset of the DRCCP is significant, since they indicate when the curtailment in the network should begin. In Figure 5-6 we observe that the first overvoltages in the training set of the DRCCP happen at 8:00 a.m. in bus 154 for Scenario 8 and at 8:00 a.m. in buses 152 and 154 for Scenario 10. If we would choose $\theta = 0$ for our ambiguity set, then we would assume that the empirical distribution $\hat{\mathbb{P}}$ of the 20 scenarios represents the true distribution $\mathbb{P}$ of PV generation and residential demand. In that case, the data would dictate the DRCCP to start curtailing at 8:00 a.m., when the first overvoltages occur in bus 152 of Scenario 8 and in buses 152 and 154 of Scenario 10. We confirm our assumption from the first row in Table 5-2 that concerns $\theta = 0$, where the curtailment starts indeed at 8:00 a.m. Since in our case we chose $\theta = 10^{-4}$ for the radius of our ambiguity set,

we increase the uncertainty that we have regarding the empirical distribution $\hat{\mathbb{P}}$, we assume that there are overvoltages even earlier than 8:00 a.m. and we start the curtailment at 7:45 a.m.

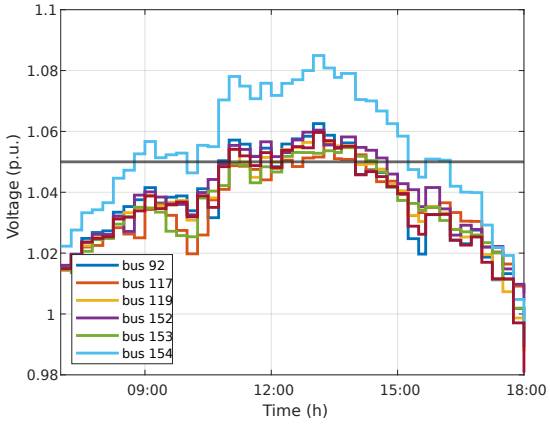
The next step is to examine how the curtailment of PV exports affects the voltage magnitudes to our network. For this purpose, we run the DRCCP algorithm with $\theta = 10^{-4}$ and we depict in Figure 5-7 Scenarios 1 and 2 as two training days to the DRCCP, and Scenarios 21 and 22 in Figure 5-8 as two unknown days to the DRCCP. For lack of space, we provide the analysis for only two scenarios of known days and two scenarios for unknown days. The rest of the scenarios exhibit the same behavior.



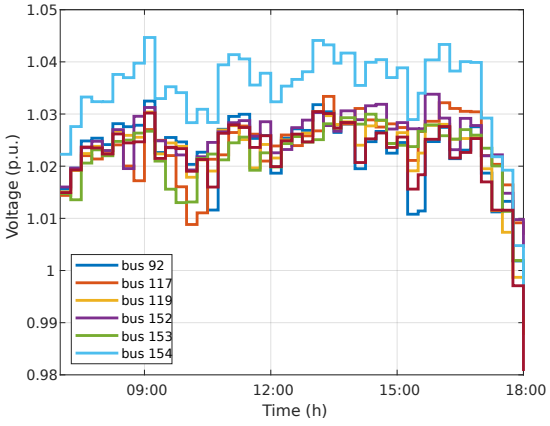
(a) Voltage violations in Scenario 1



(b) Elimination of voltage violations in Scenario 1 after curtailment

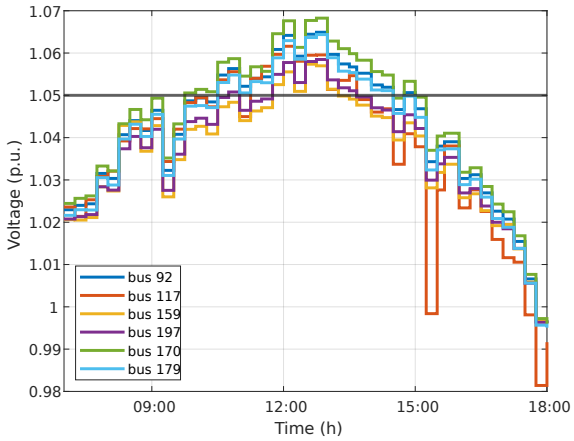


(c) Voltage violations in Scenario 2

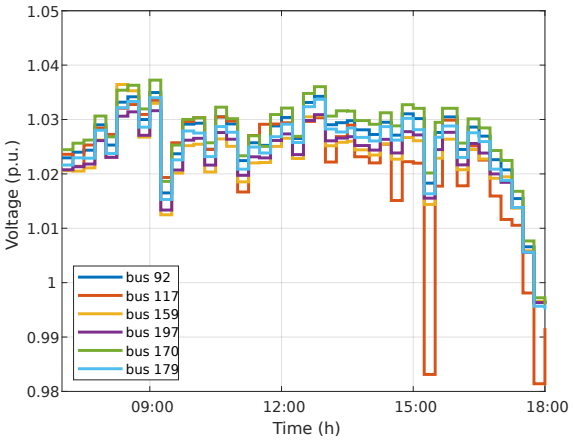


(d) Elimination of voltage violations in Scenario 2 after curtailment

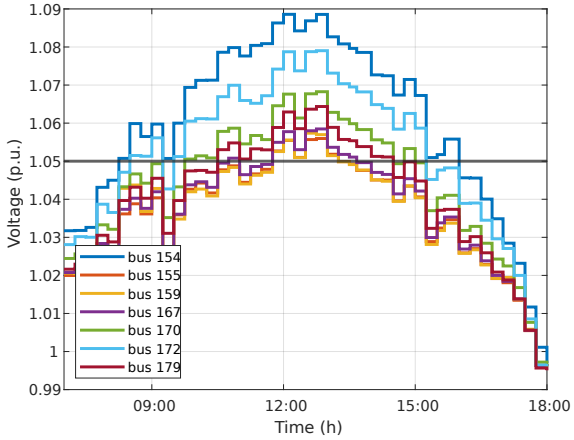
Figure 5-7: Voltage magnitudes before and after curtailment in two known scenarios. On the left side, the voltage magnitudes of 6 buses in Scenario 1 and 6 buses of Scenario 2 that present overvoltage issues, when we do not implement PV curtailment. On the right side, the same buses of Scenario 1 and Scenario 2 when we run the DRCCP algorithm with $\theta = 10^{-4}$. After implementing curtailment, the voltage magnitudes of the buses lie below the limitation of 1.05 p.u.



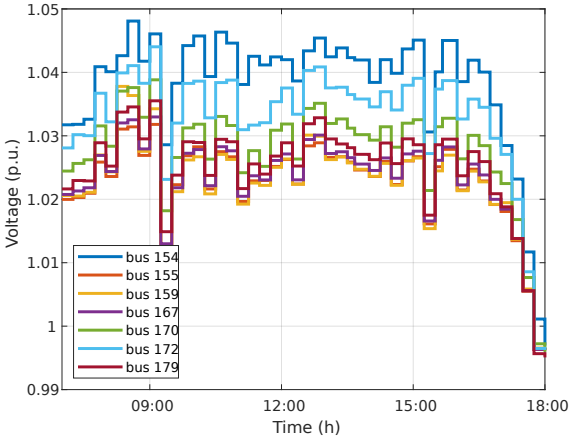
(a) Voltage violations in Scenario 21



(b) Elimination of voltage violations in Scenario 21 after curtailment



(c) Voltage violations in Scenario 22



(d) Elimination of voltage violations in Scenario 22 after curtailment

Figure 5-8: Voltage magnitudes before and after curtailment in two unseen scenarios. On the left side, the voltage magnitudes of 6 buses in Scenario 21 and 7 buses of Scenario 22 that present overvoltage issues, when we do not implement PV curtailment. On the right side, the same buses of Scenario 21 and Scenario 22 when we run the DRCCP algorithm with $\theta = 10^{-4}$. After implementing curtailment, the voltage magnitudes of the buses lie below the limitation of 1.05 p.u.

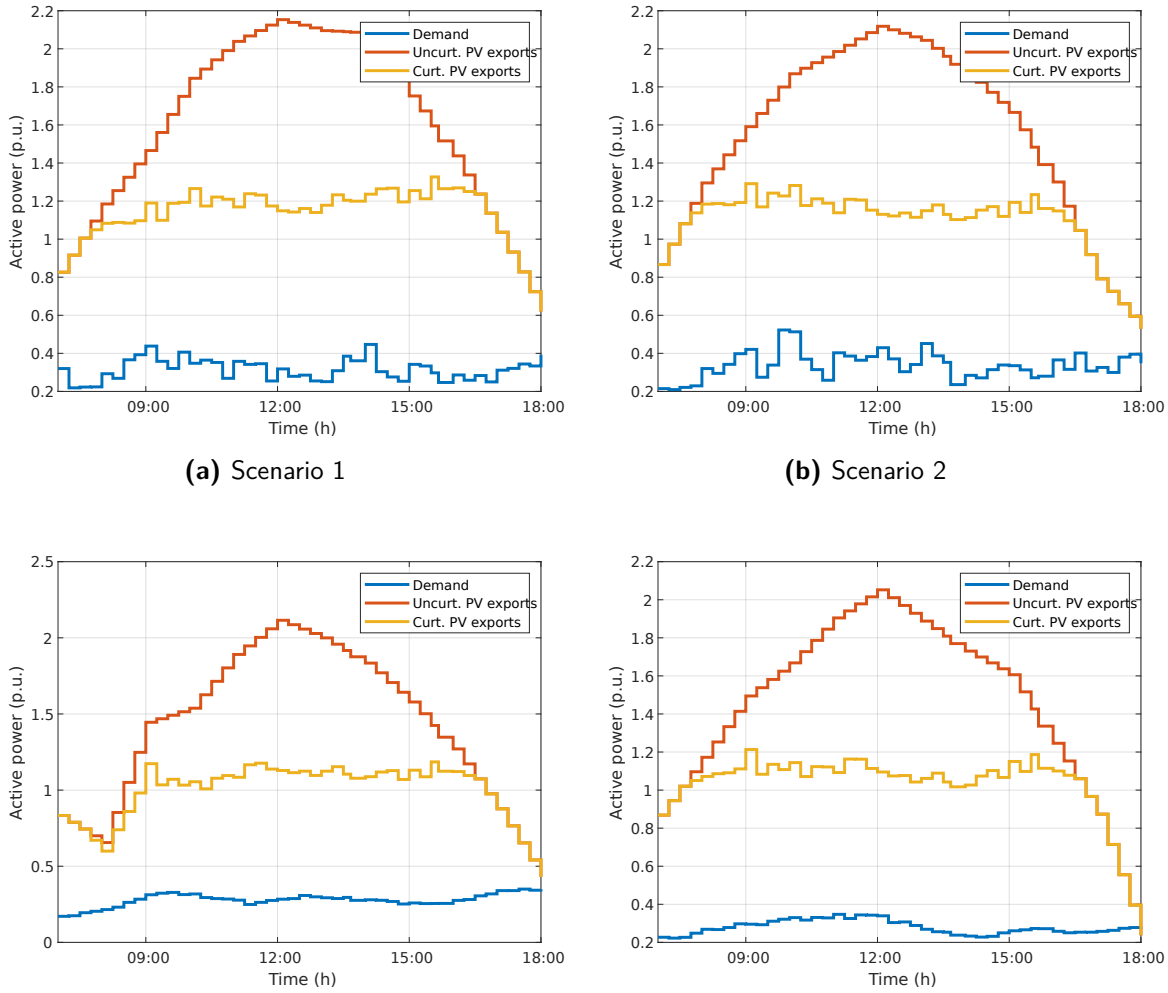
In Figure 5-7 we depict voltage magnitudes of specific buses of Scenarios 1 and 2 before and after curtailment. We set the safety limit to 1.05 p.u. and we indicate it by a black horizontal line. In Figures 5-7a and 5-7c we observe six buses in Scenario 1 and six buses in Scenario 2 exceeding this limit. Then, we implement the DRCCP algorithm with $\theta = 10^{-4}$ and we compute the appropriate curtailment factor in each time-step, according to Table 5-2. Figure 5-7b corresponds to Scenario 1 after the curtailment and likewise, Figure 5-7d corresponds to Scenario 2 after the curtailment. The results from Figures 5-7b and 5-7d indicate that the DRCCP algorithm achieves voltage regulation and the voltage magnitudes of all buses for both Scenario 1 and Scenario 2 lie below the safety limit for all time-instances. However,

Scenarios 1 and 2 of Figure 5-7 belong to the known dataset that we use to tune the DRCCP model. Our purpose is to check the model's robustness during the validation over the 500 scenarios and therefore, in Figure 5-8 we depict our implementation in two days that are completely unknown to the DRCCP model.

Figure 5-8 depicts the voltage magnitudes of Scenarios 21 and 22. These two scenarios are not involved in the tuning of the DRCCP algorithm, so we run the DRCCP algorithm with $\theta = 10^{-4}$, we determine the curtailment factor at every time-instance and we validate the results on Scenarios 21 and 22. In Figure 5-8a we observe six buses of Scenario 21 that present overvoltage issues, mainly in midday, when we do not control the output of the PVs to the network. Likewise, Figure 5-8c depicts 7 buses of Scenario 22 that exceed the voltage limit of 1.05 p.u., when there is no curtailment. Even though that Scenarios 21 and 22 do not belong to the training dataset, we achieve to eliminate the overvoltages when we implement the output of the DRCCP algorithm. In particular, in Figures 5-8b and 5-8d we observe that the voltage magnitudes of the buses in both scenarios lie below the voltage limit for all time-instances.

Finally, in Figure 5-9 we illustrate the mismatch between uncurtailed PV exports and residential demand. This mismatch can lead the network to instability and thus we curtail the PV exports to the grid, in order to achieve voltage regulation. For this purpose we depict Scenarios 1 and 2 from the dataset that we use to tune the DRCCP algorithm with $\theta = 10^{-4}$ and Scenarios 21 and 22 from the recorded days that we validate the DRCCP output.

When we do not implement any control on the PV exports, we notice in Figure 5-9 that the PV contribution to the grid peaks around midday for all scenarios. The residential demand remains low throughout the day, with small deviations for all scenarios of Figure 5-9. This low demand determines the curtailed PV output to the grid. Before 7:45 a.m., the shape of curtailed PV exports follows the shape of uncurtailed PV exports, since the curtailment factor a_c is 0, according to Table 5-2. Then, the curtailment ignites and PV exports do not present any peak exclusively, compared to uncurtailed PV exports. The shape of the curtailed PV contribution to the network follows the shape of the residential demand and it presents small deviations, until 16:30 p.m., where the curtailment factor a_c is again 0 and the curve of curtailed PV exports coincide with the curve of uncurtailed PV exports.



(a) Scenario 1

(b) Scenario 2

(c) Scenario 21

(d) Scenario 22

Figure 5-9: Residential demand compared to uncurtailed PV exports and curtailed PV exports after DRCCP with $\theta = 10^{-4}$ for 4 scenarios. The low residential demand without high fluctuations determines the shape of the curtailed PV exports.

Synopsis of Chapter 5

In this chapter we conducted a step by step analysis of our DRCCP algorithm and we presented the results of our implementation. In order to form the DRCCP, we used 20 days of our recorded data during summer days. We modelled the problem by utilizing a part of the Dutch energy distribution system with high PV penetration. However, the big size of the network and the high number of constraints urged us to considered only two branches of it. We focused our the attention on two of the longest feeders of the distribution network, since due to their long nature, they have higher chance to present voltage problems.

The next step was to simulate the model and to determine one curtailment factor every 15 minutes for the time-instances between 7:00 a.m. and 18:00 p.m. One of the inputs to the optimization problem was the empirical distribution of the PV generation and the demand

of each house. The other input was the Wasserstein distance θ that controls the size of the ambiguity set. We performed our iterations by tuning accordingly the radius of our ambiguity set and we investigated the behavior of the network .

Table 5-3 indicates the curtailment factor a_c of each time-step and the scenarios with overvoltages that preserved after the DRCCP output with $\theta = 10^{-4}$. Considering that the curtailment factor a_c takes values in space $[0, 1]$, a zero in Table 5-3 corresponds to no curtailment, while a one corresponds to 100% curtailment of the PV contribution to the grid.

Time instance / Violated Scenarios	7:00	7:15	7:30	7:45	8:00	8:15	8:30	8:45	9:00	9:15	9:30	9:45	10:00	10:15	10:30
Without Curtailment	0	0	0	4	25	62	93	136	163	178	182	190	202	212	215
With Curtailment	0	0	0	2	3	3	2	3	24	2	10	5	10	1	2
Curtailment factor	0	0	0	0.042	0.086	0.133	0.182	0.214	0.188	0.296	0.282	0.318	0.314	0.380	0.371

Time instance / Violated Scenarios	10:45	11:00	11:15	11:30	11:45	12:00	12:15	12:30	12:45	13:00	13:15	13:30	13:45	14:00	14:15
Without Curtailment	245	248	258	253	256	253	260	253	248	238	238	224	222	217	201
With Curtailment	2	0	3	1	5	2	1	2	3	0	0	0	0	0	0
Curtailment factor	0.392	0.427	0.401	0.412	0.447	0.466	0.466	0.454	0.460	0.438	0.411	0.424	0.425	0.407	0.368

Time instance / Violated Scenarios	14:30	14:45	15:00	15:15	15:30	15:45	16:00	16:15	16:30	16:45	17:00	17:15	17:30	17:45	18:00
Without Curtailment	181	180	167	123	89	70	39	15	3	0	0	0	0	0	0
With Curtailment	3	0	1	0	1	0	0	2	3	0	0	0	0	0	0
Curtailment factor	0.339	0.349	0.284	0.276	0.168	0.166	0.117	0.065	0	0	0	0	0	0	0

Table 5-3: Violated scenarios during the day before and after the optimized curtailment factor. The goal of less than 25 scenarios with overvoltages at every time-instance has been achieved with an ambiguity set with radius $\theta = 10^{-4}$.

In Table 5-3 we notice that the curtailment ignites at 7:45 a.m. and it lasts until 16:15 p.m. We see from the results on the validation data that when we do not impose any curtailment, the network faces overvoltage problems in many days of our dataset. Therefore it is clear that we need to control the PV exports. The goal of our algorithm was not to eliminate completely all potential disruptions. We designed the DRCCP algorithm in a way to battle at least 95% of problematic scenarios at every time-instance. We achieved this purpose and we observe from Table 5-3 that there are not time-instances where the violated scenarios exceed the number 25. In that way we showed that we can design a network that fulfills our design criterion without making overly conservative decisions due to the limited data that we use to tune the DRCCP.

Chapter 6

Wasserstein Barycenters

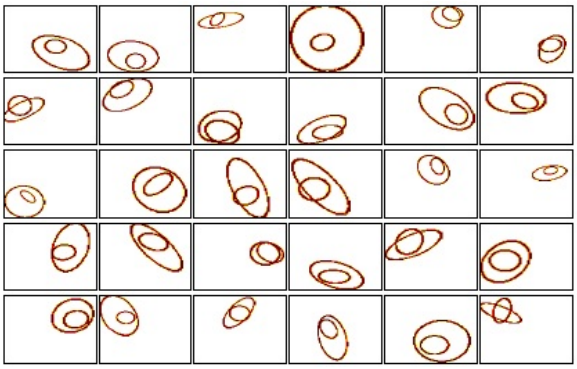
In this chapter we leverage a second case study and we intend to solve the same OPF formulation, using more data in the DRCCP algorithm. While in the previous chapter we used only 20 days of recorded data to tune the DRCCP algorithm, now we exploit the entire dataset of 500 days for the DRCCP tuning. One could assume that the high availability of data could assist us to find easily and precisely the appropriate curtailment factor. However, the high computational effort that arises from the bulk data lead us to seek alternatives in order to incorporate efficiently the entire information.

For this purpose, in this chapter we use the theory of Wasserstein Barycenters for a single distribution to cluster the samples, based on the Wasserstein distance. To the best of our knowledge, this is the first application of Wasserstein Barycenters in energy systems, and through this chapter we evaluate their performance in terms of scalability and robustness.

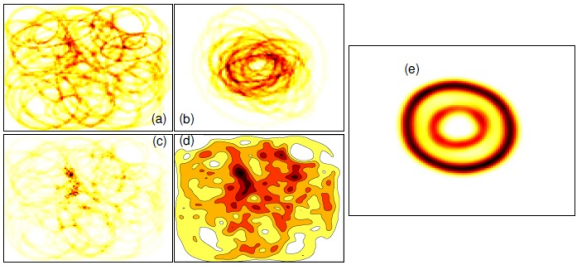
The remainder of the chapter is organized as follows. In Section 6-1 we introduce the main ideas behind Wasserstein Barycenters and some application examples. In Section 6-2 we present the stages to compute the Wasserstein Barycenter through three interconnected algorithms and n Section 6-3 we compare the DRCCP from the Wasserstein Barycenter with both the SA method and the DRCCP with the empirical distribution. Section 6-4 concerns the computation of the Wasserstein Barycenter and finally, in Section 6-5 we utilize the Wasserstein Barycenter to the DRCCP algorithm.

6-1 Introduction to Wasserstein Barycenters

With the term *Wasserstein Barycenters* we describe clusters of empirical data that are formed by optimal transport distances. In other words, we intend to get a more economical discrete distribution than the empirical distribution and we classify the samples from the empirical distribution to weighted clusters according to their most efficient transport cost [10, 40, 41, 42].



(a) Various nested random ellipses



(b) Barycenters using the (a) Euclidean distance (b) Euclidean after re-centering images (c) Jeffrey centroid (d) RKHS distance (Gaussian kernel, $\sigma = 0.002$) (e) 2-Wasserstein distance

Figure 6-1: Barycenters of 30 random nested ellipses with different distances. The 2-Wasserstein Barycenters achieve to cluster the random ellipses capturing also the structure, while for other distances the structure is not preserved. [10].

Wasserstein Barycenters are mainly used in machine learning problems and image processing tasks, since they are able to incorporate the geometric structure of objects [43]. Figure 6-1 from [10] indicates how Wasserstein Barycenters are able to preserve the geometric structure of the data, compared with Barycenters defined for other distances. In Chapter 2, we analyzed the advantages of the Wasserstein distance when we create the uncertainty set of PV generation and residential demand, and now our purpose is to utilize again the Wasserstein distance in order to group effectively our empirical data before using them to the DRCCP.

6-2 Computation of the Wasserstein Barycenter

For the construction of the Wasserstein Barycenter we consider, as in Section 3-1, the set $\Xi \in \mathbb{R}^d, d \geq 1$. Moreover, $\mathcal{B}(\Xi)$ and $\mathcal{P}(\Xi)$ correspond to the Borel σ-algebra and the set of Borel probability measures on the set Ξ, respectively. We also consider two probability measures $\mu, \nu \in P(\Xi)$. If we denote two families of points in Ξ by $X = (x_1, \cdots, x_n)$ and $Y = (y_1, \cdots, y_n)$, the aforementioned measures are on the form $\mu = \sum_{i=1}^{n} a_i \delta_{x_i}$, $X \in \mathbb{R}^{d \times n}$ and $\nu = \sum_{i=1}^{n} b_i \delta_{y_i}$, $Y \in \mathbb{R}^{d \times m}$. The terms a_i belongs to the probability simplex Σ_n, where

$$\Sigma_n \stackrel{\text{def}}{=} \left\{ u \in \mathbb{R}^n \middle| \forall i \leq n, u_i \geq 0, \sum_{i=1}^{\nu} u_i = 1 \right\}. \tag{6-2.1}$$

Likewise, the same holds for the terms b_i.

We also consider, as in Section 3-1, the Wasserstein distance $W_p(\mu, \nu)$ between the two probability measures μ and ν, as given by Equation (3-1.1). In our case we consider $p = 2$ and we define the Wasserstein Barycenter μ with points X and weights a the minimizer of f, where f is:

$$f(\mu) = W_2^2(\mu, \nu) \tag{6-2.2}$$

Given an empirical distribution of n samples, we intend to find a discrete distribution supported on n points, where typically $n < m$, which is as close as possible to the empirical distribution with respect to the Wasserstein distance. For this discrete distribution, or alternatively the Wasserstein barycenter, we take into consideration firstly, the cost $M_{XY} \in \mathbb{R}^{n \times m}$ corresponding to pairwise euclidean distances between points X and Y and secondly, the transportation polytope $\Pi(a, b) = \left\{\pi \in \mathbb{R}_+^{n \times m} \middle| \pi \mathbb{1}_m = a, \pi^T \mathbb{1}_n = b\right\}$. Here, $\mathbb{1}_m$ represents a column vector with m entries of ones.

Given the matrix $M_{XY} \in \mathbb{R}^{n \times m}$, we construct the dual linear program of the primal minimization problem of Equation (6-2.2). The dual optimal transport problem is as follows:

$$\max_{\mathrm{a}, \beta \in \mathcal{C}} \mathrm{a}^T a + \beta^T b, \tag{6-2.3}$$

where the polyhedron $\mathcal{C}_M$ of dual variable is as follows:

$$\mathcal{C}_M = \left\{(\mathrm{a}, \beta) \in \mathbb{R}^{n+m} \middle| \mathrm{a}_i + \beta_j \leq m_{ij}\right\}.$$

According to Proposition 1 of [10], any optimal dual vector "a" of the optimization problem (6-2.3), is a subgradient of the primal optimization problem (6-2.2) as a function of a.

In order to find the Wasserstein Barycenter, we divide the procedure into two interconnected sub-algorithms. In the first algorithm we assume that the points of the barycenter X are fixed and we search for the optimal weights a that lie in the simplex Σ_n. We could resemble this procedure as the formation of the cluster when the center of the cluster is known. We present the first algorithm below, where for the convergence of the algorithm we make use of the subgradient that is derived from the solution of the dual problem of Equation (6-2.3):

Algorithm 1

Inputs: $X \in \mathbb{R}^{d \times n}$, $Y \in \mathbb{R}^{d \times m}$, $b \in \Sigma_m$, tol $= 10^{-4}$
Create M_{XY}
Set $\hat{a} = \tilde{a} = \mathbb{1}_n / n$
while not converged **do**
    $\beta = (t + 1)/2$, $\alpha \leftarrow (1 - \beta^{-1})\hat{a} + \beta^{-1}\tilde{a}$
    Form the subgradient of the primal problem
    $\tilde{a} \leftarrow \tilde{a} \circ e^{-\beta \mathbf{a}}$
    $\tilde{a} \leftarrow \tilde{a} / \tilde{a}^T \mathbb{1}_n$
    $\hat{a} \leftarrow (1 - \beta^{-1})\hat{a} + \beta^{-1}\tilde{a}$, $t \leftarrow t + 1$
Return $\hat{a}$

Then, we move to the second algorithm where we are free to choose the support of the barycenter. More specifically, we iterate in order to recalculate the center of the clusters X.

In order to achieve convergence for the center of the cluster, we make use of the optimal transport π that corresponds to the primal problem of Equation 6-2.2. The second algorithm is as follows:

Algorithm 2

Inputs: $Y \in \mathbb{R}^{d \times m}$, $b \in \Sigma_m$, tol $= 10^{-4}$
Initialize $X \in \mathbb{R}^{d \times n}$ and $a \in \Sigma_n$
while X and a have not converged **do**
    $a \leftarrow \hat{a}$ from Algorithm 1
    find the optimal solution of the primal problem π
    $X \leftarrow (1-\theta)X + \theta Y \pi^{*T} \text{diag}(a^{-1})$, setting $\theta \in [0,1]$ with line-search or a preset value
Return X

We repeat these two interconnected algorithms until they both converge. The whole procedure is non-convex and our optimum is a local minimum. Assuming that there are no constraints regarding the weights, this approach is equivalent to Lloyd's k-means algorithm [10]. However, we choose this alternative implementation instead of Lloyd's algorithm, because we can directly calculate the Wasserstein distance to move the samples from the empirical distribution to the Wasserstein Barycenter, using Equation (3-1.1). The same procedure with the Lloyd's algorithm would require post-processing calculations.

In Algorithm 1 and Algorithm 2 we utilize explicitly the primal and dual optimal transports problems through the terms π and a, respectively. It terms of computational time, these two steps add extra burden to the scalability of the problem. In particular, in Algorithm 1 we need to calculate the subgradient from the optimum of the dual transportation problem in each iteration, which corresponds to a problem with $n + m$ variables and $n \times m$ inequality constraints. The computational cost of Algorithm 2 is even higher, because we compute Algorithm 1 in every iteration and additionally we solve the primal problem to update X.

In order to accelerate the procedure, we present the alternative of the smoothed primal and dual problems' solutions, π_λ and a_λ, respectively. If we define $K = e^{-\lambda M_{XY}}$, with $\lambda > 0$ the smoothing parameter, we can compute the optimal solutions of the primal and dual problems by a pair of vectors $(u, v) \in \mathbb{R}_+^n \times \mathbb{R}_+^m$ as follows:

$$\pi_\lambda = \text{diag}(u) K \text{diag}(v)$$
$$\text{a}_\lambda = \frac{\log(u)}{\lambda} - \frac{\log(u)^T \mathbb{1}_n}{\lambda n}$$

Below we present the steps to define the aforementioned pair of vectors (u, v) and consequently the regularized terms π_λ and a_λ that we will use in Algorithm 1 and Algorithm 2 instead of π and a.

Algorithm 3

Inputs: $M, \lambda, a, b, \text{tol} = 10^{-5}$
$K = e^{-\lambda M}$
$\tilde{K} = \text{diag}(a^{-1})K$
$u = \text{ones}(n, 1)/n$
while u changes **do**
    $u = 1./\left(\tilde{K}\left(b./\left(K^T u\right)\right)\right)$
$v = b./(K^T u)$
$\text{a}_\lambda = \frac{1}{\lambda}\log(u) - \frac{\log(u)^T \mathbb{1}_n}{\lambda n}\mathbb{1}_n$
Return $\text{a}_\lambda, \pi_\lambda$
$\pi_\lambda = \text{diag}(u)K\text{diag}(u)$

6-3 Motivation behind Wasserstein Barycenters

The presence of a high amount of i.i.d. data has a positive impact on our formulation. In our case study, we inspect a data-driven problem that most of the times the lack of data hinders us from getting efficient solutions. Now that we have plenty of data available, we can formulate a more precise model and decrease the radius of the ambiguity set. The 500 available scenarios can provide us with significant information about the real probability distribution of the PV generation and residential demand. The presence of this amount of data gives us incentives to implement the SA approach that we explained in Section 2-3. The SA approach guarantees that the constraints are feasible and, at the same time, we exploit the absence of slack variables and additional constraints that a DRCCP inserts. These slack variables and extra constraints put additional effort to the optimization procedure and hence, the SA becomes significantly faster.

The main disadvantage of high bulk of data is the computational effort. As we mentioned in Section 5-2, we intend to reduce the constraints of the problem as much as possible. Now, the integration if 500 scenarios for precise solutions, comes with the cost of extremely high computational time.

To mitigate this drawback, we utilize Wasserstein Barycenters. With Algorithm 1, Algorithm 2 and Algorithm 3 that we presented in the previous section, we can efficiently incorporate the information of 500 scenarios into few clusters of data, accompanied with their corresponding weights. In particular, we compare the robustness and the computational time that we need for the SA approach and the computational time that we need for the DRCCP with the Wasserstein Barycenter. In this comparison we make use of the entire available dataset, namely the 500 scenarios and we choose 10 points for the Wasserstein Barycenter.

In Section 2-3, we mentioned the Formula (2-3.1) for the SA to determine the appropriate number of scenarios that we need, in order to guarantee that our solution satisfies the constraints with the probability level that we decide. For a reminder, the formula is as follows:

$$N \geq \frac{2}{a}\ln\frac{1}{\beta} + 2n + \frac{2n}{a}\ln\frac{2}{a},$$

where the decision variable $\in \mathbb{R}^n$, a is the violation parameter and β is the confidence parameter. Since in DRCCP we chose the violation parameter to be equal to $a = 0.05$, we assign

correspondingly $a = 0.05$ for 95% satisfaction set in the SA method. With the term $(1 - a)$ in the SA, we express the probability that the constraints are feasible. Moreover, we assign the values $n = 1$, because our decision variable (curtailment factor) is scalar, and $\beta = 10^{-10}$ for the confidence parameter.

Filling in the values of a, n and β in (2-3.1), we get $N \geq 1070.59$, or in other words, we see that we need at least 1071 scenarios to achieve 95% satisfaction of the constraints with the SA. Therefore, it is impossible to guarantee at least 95% satisfaction of the constraints with only 500 available scenarios when we implement the SA method. Hence, the DRCCP with the Wasserstein Barycenter outperforms in terms of robustness, because with the appropriate tuning of the Wasserstein radius we can achieve the desired 95% satisfaction of constraints.

However, we want to prove that the DRCCP with the Wasserstein Barycenter outperforms in computational time as well. For this reason, we compare the running time of the original DRCCP when we use 500 scenarios, with the running time of the DRCCP when we use Algorithm 1, Algorithm 2 and Algorithm 3 for the Wasserstein Barycenter. We perform this comparison only for one time-instance of our problem. While for the computation of the DRCCP with the presence of the Wassserstein Barycenter we need 117 seconds, for the original DRCCP we interrupted the execution of the algorithm after 6770 seconds without finishing. Finally, we check also the execution time of the SA, even though that the performance is limited, and we see that we need 288 seconds. With these results, we conclude that the approach of the Wasserstein Barycenter with the DRCCP outperforms in terms of robustness compared to the SA approach and in terms of computational time compared to SA approach and to the original DRCCP for 500 scenarios.

6-4 Wasserstein Barycenters for DRCCP power flow

In order to find the Wasserstein Barycenter, we need to specify the parameters that appear in Algorithm 1, Algorithm 2 and Algorithm 3. From Table 5-3 we observe that there are no overvoltage violations before 7:45 a.m. and after 16:30 p.m. and therefore, we limit our time-window between 7:45 a.m. and 16:30 p.m. We assume $\Xi \in \mathbb{R}^{(408\times 36)}$ and $Y \in \mathbb{R}^{(408\times 36)\times 500}$. The dimensions of Y emerge from the recorded three phase PV generation of 68 houses and the three phase demand of 68 houses. We divide our time-window into quarters of an hour, and as a result, we obtain 36 time-instances with recordings of PV generation and residential demand. Our purpose is to cluster the recordings of 500 scenarios. The Barycenter that we intend to construct reduces the 500 points of distribution Y to only 10 in distribution X, so $X \in \mathbb{R}^{(408\times 36)\times 10}$. The parameter b is a vector and equals to $b = \{\mathbb{1}_m/m\} = \{\mathbb{1}_{500}/500\}$. We set $\lambda = 60/\text{median}(M_{XY})$ and for the parameter t, which represents the gradient steps, we assign the value $t = 60$ according to [10].

After the successful implementation of Algorithm 1, Algorithm 2 and Algorithm 3, we specify the upper bounds of the Wasserstein distances in each time-step. We refer to upper bounds, because the solution of the Wasserstein Barycenter is a local minimum and because we regularize the solutions of the primal and dual transportation problems through Algorithm 3. In Table 6-1 we present the Wasserstein distances θ_{wb} that we compute in each time-step, which represent the upper bound of the cost to move the data of 500 scenarios from distribution Y to 10 points in distribution X:

Time instance	7:45	8:00	8:15	8:30	8:45	9:00	9:15	9:30	9:45	10:00	10:15	10:30
Wasserstein Distance	9.48e-4	1.23e-3	1.55e-3	1.74e-3	2.41e-3	3.22e-3	3.62e-3	3.65e-3	4.45e-3	5.38e-3	5.05e-3	4.41e-3

Time instance	10:45	11:00	11:15	11:30	11:45	12:00	12:15	12:30	12:45	13:00	13:15	13:30
Wasserstein Distance	3.95e-3	4.21e-3	3.93e-3	3.79e-3	3.64e-3	3.89e-3	3.84e-3	3.44e-3	4.37e-3	5.24e-3	4.59e-3	4.68e-3

Time instance	13:45	14:00	14:15	14:30	14:45	15:00	15:15	15:30	15:45	16:00	16:15	16:30
Wasserstein Distance	3.70e-3	3.69e-3	4.13e-3	3.64e-3	2.75e-3	3.68e-3	4.02e-3	3.45e-3	2.66e-3	2.95e-3	2.87e-3	2.25e-3

Table 6-1: Wasserstein distance computed for the Wasserstein Barycenter with 10 points.

According to our experience, if we reduce the points of the Wasserstein Barycenter, then the Wasserstein distances θ_{wb} at each time instance will increase. On the other hand, in the extreme case of constructing a Wasserstein Barycenter with 500 points for this problem, then we will experience 0 Wasserstein distances θ_{wb} at each time-step. This fact emerges from the fact that the samples will not have to move from Y to X. We confirm our expectations in Table 6-2, where we present the Wasserstein distances for a Wasserstein Barycenter of 20 points. When we compare Wasserstein distances in Tables 6-1 with their corresponding Wasserstein distances in Table 6-2, we notice that the Wasserstein distances for a Barycenter of 20 points are smaller than those for a Barycenter of 10 points. On the other hand, through this comparison, we observe that the order of magnitude of the Wasserstein distances between Tables 6-1 and 6-2 remains the same. In other words, the travel cost to cluster a distribution of 500 samples to a Barycenter of 20 points is lower than clustering the same distribution to a Barycenter of 10 points, but it does not change significantly. We choose to construct our Wasserstein Barycenter with 10 points, because we accelerate the running time of the procedure compared to a Barycenter of 20 points without losing important information.

Time instance	7:45	8:00	8:15	8:30	8:45	9:00	9:15	9:30	9:45	10:00	10:15	10:30
Wasserstein Distance	8.76e-4	1.08e-3	1.33e-3	1.55e-3	2.29e-3	2.29e-3	3.28e-3	3.29e-3	3.94e-3	4.90e-3	4.61e-3	4.04e-3

Time instance	10:45	11:00	11:15	11:30	11:45	12:00	12:15	12:30	12:45	13:00	13:15	13:30
Wasserstein Distance	3.64e-3	3.79e-3	3.57e-3	3.42e-3	3.23e-3	3.57e-3	3.37e-3	3.17e-3	3.93e-3	4.67e-3	4.27e-3	4.27e-3

Time instance	13:45	14:00	14:15	14:30	14:45	15:00	15:15	15:30	15:45	16:00	16:15	16:30
Wasserstein Distance	3.27e-3	3.29e-3	3.83e-3	3.29e-3	2.46e-3	3.26e-3	3.60e-3	3.12e-3	2.32e-3	2.59e-3	2.57e-3	2.05e-3

Table 6-2: Wasserstein distance computed for the Wasserstein Barycenter with 20 points.

6-5 DRCCP using the Wasserstein Barycenter

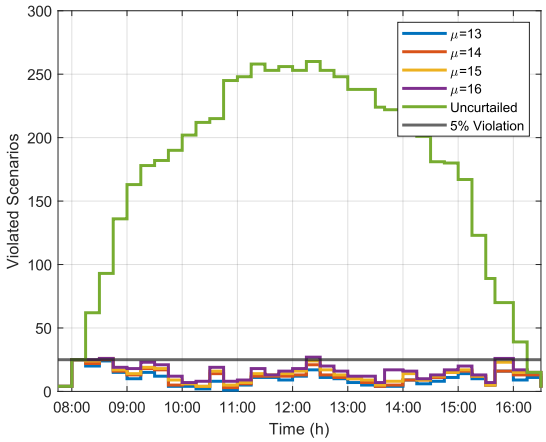
In this section we perform the DRCCP with the Wasserstein Barycenter that we computed in the previous section. The optimization algorithm (3-5.1) now turns to:

$$\begin{aligned} & \min_{x,\lambda,s_i,t} c^T x \\ s.t. \quad & \lambda\left(\theta + \frac{\theta_{wb}}{\mu}\right) + \sum_{i=1}^{N} a_{wb_i} s_i \leq t\alpha \\ & \left(b_k(x) + t + \left(x^T A_k + B_k - C^T \eta_{ik}\right)^T \hat{\xi}_i + {\eta_{ik}}^T h\right)_+ \leq s_i, \\ & ||x^T A_k + B_k - C^T \eta_{ik}|| \leq \lambda, \quad \eta_{ik} \geq 0 \\ & x \in \ X, t \in \mathbb{R}, \lambda \geq 0 \end{aligned} \tag{6-5.1}$$

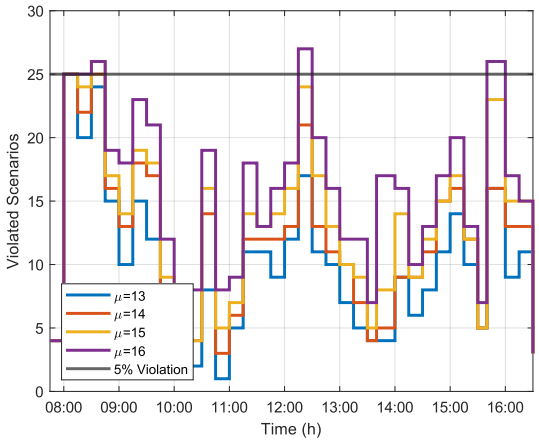
In this new DRCCP reformulation, we add the distance of the Wasserstein Barycenter θ_{wb} that we found in Table 6-1 to the radius of the ambiguity set θ. Moreover, we substitute the term $1/N$ with the term a_{wb}, which is the weight of every point of the Barycenter. With the Wasserstein Barycenters each point has a different weight from others, so we need to assign the points with their corresponding weight. Once more, we assume the support of the uncertain parameters unbounded and we set $C = 0, h = 0$ and we neglect the variables η.

Furthermore, from Table 6-1 we observe that the distances from Wasserstein Barycenter θ_{wb} at each time instance are one order of magnitude greater than the ambiguity radius $\theta = 10^{-4}$ that we defined in Chapter 5. Due to that fact, θ_{wb} dominates over θ and we suspect that the output of the OPF algorithm will be conservative, with high generation cuts for the PV owners. Therefore, we use an empirical parameter $\mu > 1$ that scales down the distance of the Barycenter. The more we increase μ the less conservative solution we get. We implement the new DRCCP formulation of Equation (6-5.1) and in Figure 6-2 we illustrate the number of scenarios in our dataset that present overvoltage violations. In this case we examine the performance before and after applying the curtailment factor that we compute from the Barycenter with $n = 10$ points instead of 500 scenarios.

Figure 6-2a depicts the vulnerability of the network to overvoltage instances when we do not control the PV output. Without PV power curtailment, the number of scenarios with overvoltage instances exceeds the design requirement of maximum 25 violations per time-instance for almost the entire time-window of 7:45 a.m. to 16:30 p.m. When PV curtailment is active, we observe that the number of violated scenarios drops significantly. Figure 6-2b illustrates the violated scenarios when we implement the solution of the DRCCP with the Wasserstein Barycenter for four different values of μ. After multiple iterations with different values of μ, we choose to depict the solutions of the DRCCP with the Wasserstein Barycenter for those μ that approach the desired performance the most. The iterations that we present in Figure 6-2b initiate with $\mu = 13$. Our purpose is to define the maximum μ that does not violate the performance goal of 95% feasibility of constraints at every time-instance. In our case, μ takes integer values and the step of μ at each iteration is 1. With smaller steps of μ, the output of the DRCCP algorithm with the Wasserstein Barycenter presents minor changes, while for bigger steps of μ we lose significant information.



(a) Number of scenarios with overvoltage violations without curtailment and with curtailment for different values of μ and $\theta = 10^{-4}$



(b) Number of scenarios with overvoltage violations after curtailment for different values of μ and $\theta = 10^{-4}$

Figure 6-2: Violated scenarios without and with curtailment for different values of μ and $\theta = 10^{-4}$ after DRCCP with the Wasserstein Barycenters. Without curtailment the number of overvoltage instances is high and hence we need to control the PV output through the optimized curtailment factor. After curtailment, the number of overvoltage instances drops significantly and depends on the scaling factor μ. The number of violated scenarios decreases monotonically as we decrease μ. The black horizontal line indicates the design requirement of maximum 25 scenarios with overvoltage instances. We choose the maximum μ that does not exceed the black horizontal line.

In Figure 6-2b, we notice that for higher values of μ the overvoltage instances increase and the output is less conservative. This fact confirms our expectation for conservative solutions without the scaling parameter μ and as a result, the presence of a $\mu > 1$ in the optimization problem of Equation (6-5.1) contributes to prevent high PV cuts. Furthermore, the black horizontal line dictates which values of μ we should avoid in order to meet the design criterion of 95% satisfaction of the constraints at every time-instance. With $\mu = 16$ and $\mu = 15$ our DRCCP implementation exceeds the limit three times and a single time, respectively. This fact indicates that we need to decrease the μ parameter, in order to increase our uncertainty of PV generation and residential demand. With $\mu = 14$, we fulfil our design criterion for the first time and hence, we choose $\mu = 14$ as the most profitable solution for the PV owners that guarantees at least 95% satisfaction of constraints at every time-instance. For $\mu = 13$ and even lower values, our solutions satisfy the design goal, but they perform overconservative PV cuts.

We quantify the aforementioned results in Table 6-3 and Table 6-4. Initially, in Table 6-3 we present how the μ parameter affects the number of violated scenarios at each iteration. Furthermore, in Table we examine 6-4 how the μ parameter affects curtailment factor that emerges from the DRCCP algorithm at every time-instance.

Time instance / Violated Scenarios	7:45	8:00	8:15	8:30	8:45	9:00	9:15	9:30	9:45	10:00	10:15	10:30
Without Curtailment	4	25	62	93	136	163	178	182	190	202	212	215
With Curtailment $\mu = 13$	4	25	20	24	15	10	15	12	4	4	2	8
With Curtailment $\mu = 14$	4	25	22	25	16	13	18	17	5	6	4	14
With Curtailment $\mu = 15$	4	25	24	**26**	17	14	19	18	9	6	4	16
With Curtailment $\mu = 16$	4	25	25	**26**	19	18	23	21	12	7	8	19

Time instance / Violated Scenarios	10:45	11:00	11:15	11:30	11:45	12:00	12:15	12:30	12:45	13:00	13:15	13:30
Without Curtailment	245	248	258	253	256	253	260	253	248	238	238	224
With Curtailment $\mu = 13$	1	5	11	11	9	12	17	11	10	7	5	4
With Curtailment $\mu = 14$	3	6	12	12	12	13	21	13	11	10	7	4
With Curtailment $\mu = 15$	5	7	14	13	14	16	24	17	13	10	9	5
With Curtailment $\mu = 16$	8	9	18	13	16	18	**27**	20	16	12	12	7

Time instance / Violated Scenarios	13:45	14:00	14:15	14:30	14:45	15:00	15:15	15:30	15:45	16:00	16:15	16:30
Without Curtailment	222	217	201	181	180	167	123	89	70	39	15	3
With Curtailment $\mu = 13$	4	9	6	8	11	14	10	5	16	9	11	3
With Curtailment $\mu = 14$	5	9	9	11	15	16	12	5	16	13	13	3
With Curtailment $\mu = 15$	8	14	9	12	15	17	12	5	23	15	15	3
With Curtailment $\mu = 16$	17	16	10	13	17	20	13	7	**26**	17	15	3

Table 6-3: Violated scenarios during the day before and after the optimized curtailment factor for different values of μ and $\theta = 10^{-4}$.

In Table 6-3 we specify with bold the number of problematic scenarios that preserve after the curtailment and exceed the desired 5% violation limit. The first row in Table 6-3 indicates the number of violated scenarios without curtailment and it is the same as in Table 5-1, since in both cases we have not implemented any PV control yet. Once more, we observe the susceptibility of the network to overvoltages, because without any control there are 62 instances even from 8:15 a.m. that violate our performance criterion of maximum 25 scenarios with overvoltages at every time-instance.

Then, the next rows of Table 6-3 present the number of overvoltage instances that remain after we impose the optimized curtailment factor for different values of μ. The values of μ that we use in our iterations to construct Table 6-3 are the same as in Figure 6-2. Once the curtailment is in place, the number of violated scenarios drops. Since Figure 6-2 indicates that the number of violated scenarios increase monotonically as we increase μ, we search for the maximum value of μ that fulfils our design criterion of at least 95% satisfaction of constraints at every time-instance. We initiate our iterations with $\mu = 13$. For $\mu = 13$, we observe that the performance criterion is not violated throughout the day. Then, we increase μ by one step and we notice that even though the number of scenarios with overvoltages increase at all time-instances, we still satisfy the performance goal. Especially at 8:30 a.m., we experience 25 violated scenarios, which is a sign that $\mu = 14$ is probably the maximum μ that fulfils the criterion. This is indeed the case, because when we increase the μ parameter by one more step, we observe that there is one time-instance, at 8:30 a.m., where the number of violated scenarios exceeds 25. Likewise, for $\mu = 16$, the time-instances where there are more than 25 violated scenarios increase from one to three, namely at 8:30 a.m., at 12:15 p.m. and at 15:45 p.m. Therefore, we confirm that the appropriate value for the μ parameter is $\mu = 14$.

Time instance	7:45	8:00	8:15	8:30	8:45	9:00	9:15	9:30	9:45	10:00	10:15	10:30
$\mu = 13$	0	0	0.072	0.107	0.165	0.213	0.239	0.265	0.321	0.348	0.371	0.349
$\mu = 14$	0	0	0.068	0.102	0.159	0.206	0.232	0.258	0.313	0.339	0.363	0.342
$\mu = 15$	0	0	0.064	0.099	0.155	0.201	0.226	0.252	0.307	0.332	0.356	0.336
$\mu = 16$	0	0	0.061	0.096	0.150	0.196	0.221	0.247	0.301	0.325	0.350	0.331

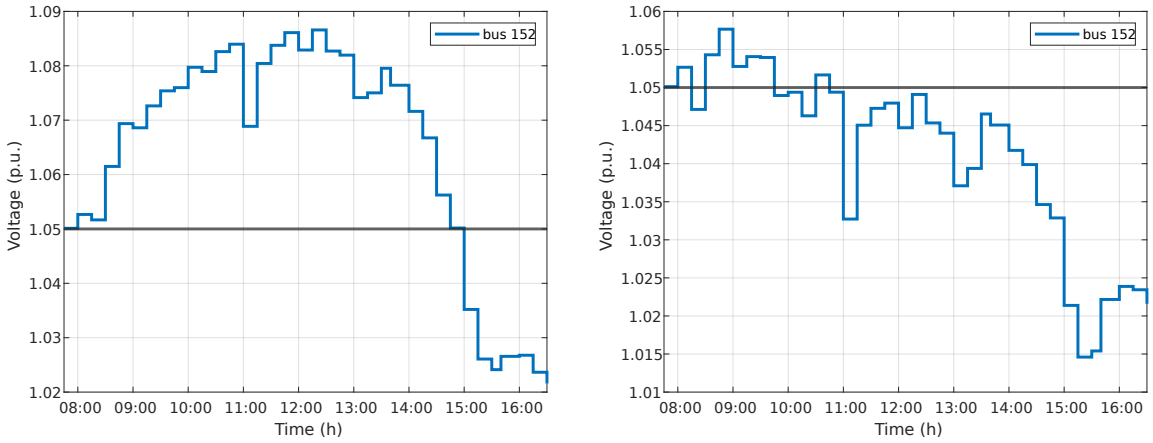
Time instance	10:45	11:00	11:15	11:30	11:45	12:00	12:15	12:30	12:45	13:00	13:15	13:30
$\mu = 13$	0.385	0.400	0.387	0.395	0.408	0.404	0.398	0.397	0.406	0.400	0.391	0.371
$\mu = 14$	0.380	0.394	0.382	0.389	0.402	0.398	0.393	0.392	0.400	0.392	0.385	0.363
$\mu = 15$	0.374	0.389	0.377	0.385	0.398	0.394	0.388	0.388	0.395	0.385	0.379	0.357
$\mu = 16$	0.370	0.384	0.372	0.380	0.394	0.389	0.383	0.384	0.390	0.380	0.373	0.352

Time instance	13:45	14:00	14:15	14:30	14:45	15:00	15:15	15:30	15:45	16:00	16:15	16:30
$\mu = 13$	0.358	0.349	0.342	0.303	0.268	0.242	0.215	0.174	0.096	0.071	0.013	0
$\mu = 14$	0.352	0.343	0.336	0.296	0.263	0.235	0.207	0.167	0.090	0.063	0.005	0
$\mu = 15$	0.347	0.338	0.330	0.291	0.258	0.229	0.200	0.161	0.084	0.057	0	0
$\mu = 16$	0.343	0.333	0.324	0.286	0.254	0.224	0.194	0.155	0.079	0.051	0	0

Table 6-4: Curtailment factors of PV generation for different values of μ and $\theta = 10^{-4}$.

In Table 6-4 we introduce the curtailment factor a_c of every time-instance that concerns the time-window between 7:45 a.m. and 16:30 p.m. In this table, we present the output of the DRCCP algorithm with the Wasserstein Barycenter for four different values of the μ parameter. The values of μ that we use in our iterations to construct Table 6-4 are the same as in Figure 6-2 and Table 6-3. Considering that the curtailment factor a_c takes values in space $[0, 1]$, a "0" in Table 6-4 corresponds to no curtailment, while a "1" corresponds to 100% curtailment of PV exports to the network. The curtailment process ignites simultaneously for all values μ, namely at 8:15 a.m. However, the ending time is not the same for all μ. In particular, for $\mu = 15$ and $\mu = 16$, the curtailment process stops at 16:15 p.m., while for $\mu = 13$ and $\mu = 14$ the curtailment process ends at 16:00. We justify that result, since the μ parameter in the minimization problem of Equation (6-5.1) is in the denominator, and for smaller values of μ the parameter θ_{wb} dominates more over the radius of the ambiguity set θ. As a result, for smaller values of μ the uncertainty increases and we experience more conservative solutions with longer curtailment periods and bigger curtailment factors. On the other hand, if we choose higher values of μ, the domination of θ_{wb} over θ is mitigated, the curtailment factor decrease and we experience higher probability of violating the design goal of 95% satisfaction of constraints at every time-instance.

Figure 6-2 and Table 6-3 assist us to choose $\mu = 14$ in order to fulfil the design goal without overconservative PV cuts. Furthermore, Table 6-3 indicates that our dataset includes 4 cases with overvoltage instances at 7:45 a.m., when we do not implement any control over the PV contribution to the grid. Figure 6-3a depicts the voltage magnitude of bus 152 in Scenario 154 and illustrates how the overvoltage that occurs at 7:45 a.m. evolves over the day without curtailment. In Figure 6-3b, we depict how the overvoltage evolves after the implementation of the optimized curtailment factor with $\mu = 14$ and $\theta = 10^{-4}$.

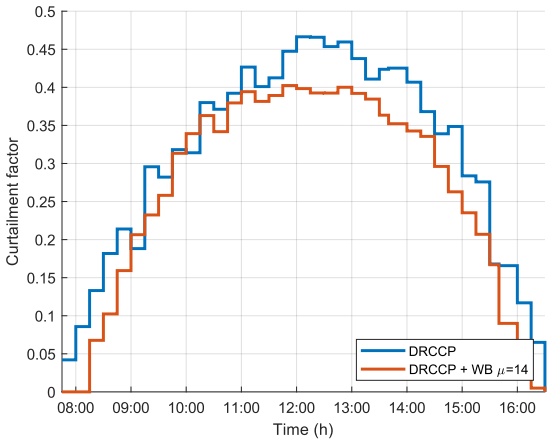


(a) Voltage magnitude of bus 154 before curtailment **(b)** Voltage magnitude of bus 154 after curtailment

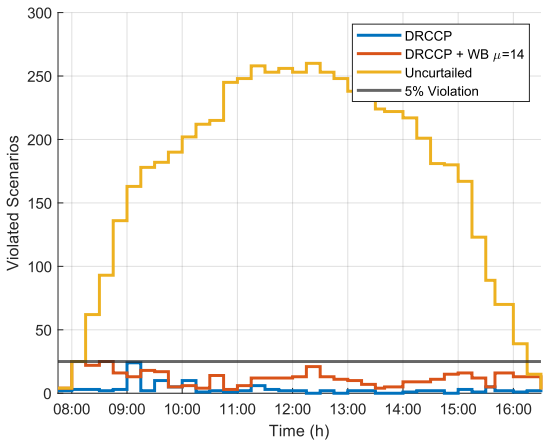
Figure 6-3: Voltage magnitudes of bus 152 in Scenario 154 before and after curtailment. The optimized curtailment factor from the DRCCP algorithm with the Wasserstein Barycenters and $\mu = 14$ and $\theta = 10^{-4}$ cannot eliminate the overvoltage instances of that bus.

From Figure 6-3 we observe that our algorithm cannot regulate the voltage of the node in the initial time-instances. Hence, in this case, we have a scenario where the overvoltage persist despite the application of the curtailment factor. However, this is a result that lies within our compromise of maximum 5% violation of constraints at every time-instance and it is not our purpose to increase the curtailment factor in order to eliminate this specific overvoltage.

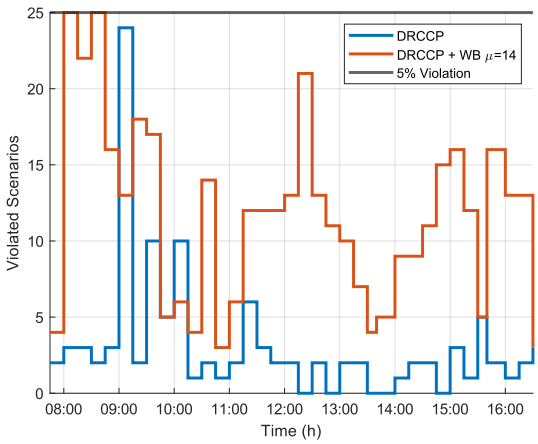
Finally, in Figure 6-4 we present in three sub-figures the results of our two case studies. The first case study concerns the DRCCP approach with 20 tuning scenarios of Chapter 5 and the second case study concerns the approach of the DRCCP with the Wasserstein Barycenter. For the DRCCP method of Chapter 5, we set $\theta = 10^{-4}$ and for approach of the DRCCP with the Wasserstein Barycenter, we set $\mu = 14$ and $\theta = 10^{-4}$. However, these two approaches address to different case studies and we cannot extract a fair comparison. The purpose of Figure 6-4 is to indicate that in both cases, through the DRCCP and the Wasserstein distance metric, we succeed to regulate the voltage within the network and guarantee 95% satisfaction of constraints during the whole day.



(a) Comparison of curtailment factors at every time-instance of the DRCCP approach with 20 tuning scenarios with the DRCCP approach with the Wasserstein Barycenters. The curtailment factors of the DRCCP approach with the Wasserstein Barycenters are mainly smaller than the DRCCP approach with 20 tuning scenarios.



(b) Number of scenarios with overvoltage violations without curtailment and with curtailment for the DRCCP approach with 20 tuning scenarios with the DRCCP approach with the Wasserstein Barycenters. The violated scenarios dominate when there is no PV control.



(c) Number of scenarios with overvoltage violations after curtailment for the DRCCP approach with 20 tuning scenarios with the DRCCP approach with the Wasserstein Barycenters. In both approaches the performance goal is achieved.

Figure 6-4: Concentrated results of 1. DRCCP approach with 20 tuning scenarios and $\theta = 10^{-4}$ and 2. DRCCP approach with the Wasserstein Barycenter of $\mu = 14$ and $\theta = 10^{-4}$.

In particular, in Figure 6-4a, we see that in most cases the curtailment factor of the DRCCP with the Wasserstein Barycenter lies below the curtailment factor of the DRCCP of Chapter 5. We expected this result, since in the DRCCP with the Barycenter we have much more information available and we can get more precise solutions. On the other hand, with the original DRCCP with the 20 tuning scenarios, we intend to cover many unseen days, and therefore the curtailment factor is higher and more conservative. In Figure 6-4b we observe the number of scenarios with overvoltages that occur at every time-instance without any curtailment factor, the number of problematic scenarios that preserve after curtailment with the DRCCP method and the 20 tuning scenarios of Chapter 5 and the number of problematic scenarios that preserve after curtailment with the DRCCP and the Wasserstein Barycenter approach. It is clear that the network undergoes high number of overvoltage instances if none of our approaches is active. However, we notice in Figure 6-4c, that when we implement any of our two approaches, then the curtailment ignites and the overvoltage instances drop. Moreover, for both case studies, we achieve to fulfil our design criterion of at least 95% scenarios without any violations, since in Figure 6-4c the number of problematic scenarios lies below the limit that we have set.

Synopsis of chapter 6

In this chapter we examined the application of the Wasserstein Barycenters in order to handle appropriately the large amount of available data. With the Wasserstein Barycenter we achieved to cluster the data with respect to Wasserstein distance metric and to assign them to clusters with the appropriate weight. We analyzed the algorithms of the Wasserstein Barycenter and then we inspected their efficiency. In terms of robustness, the DRCCP with the Wasserstein Barycenter achieves higher percentage of constraints' feasibility than the SA approach. Moreover, with this approach we manage to exploit the whole information by accelerating the optimization process compared to the DRCCP without the Wasserstein Barycenter. Finally, we performed the DRCCP algorithm with the Wasserstein Barycenter instead of the empirical distribution of 20 scenarios and we presented the corresponding results.

Table 6-5 concentrates the results of our DRCCP reformulation with the Wasserstein Barycenter. In this table we observe that, during the entire time-window, we do not exceed the 25 scenarios with overvoltage violations, which corresponds to no more than 5% violation of constraints. At the same time, we choose the parameter μ of the problem in order to provide the PV owners with the highest profit without overly conservative solutions and high PV cuts.

Time instance / Violated Scenarios	7:45	8:00	8:15	8:30	8:45	9:00	9:15	9:30	9:45	10:00	10:15	10:30
Without Curtailment	4	25	62	93	136	163	178	182	190	202	212	215
With Curtailment	4	25	22	25	16	13	18	17	5	6	4	14
Curtailment factor	0	0	0.068	0.102	0.159	0.206	0.232	0.258	0.313	0.339	0.363	0.342

Time instance / Violated Scenarios	10:45	11:00	11:15	11:30	11:45	12:00	12:15	12:30	12:45	13:00	13:15	13:30
Without Curtailment	245	248	258	253	256	253	260	253	248	238	238	224
With Curtailment	3	6	12	12	12	13	21	13	11	10	7	4
Curtailment factor	0.380	0.394	0.382	0.389	0.402	0.398	0.393	0.392	0.400	0.392	0.385	0.363

Time instance / Violated Scenarios	13:45	14:00	14:15	14:30	14:45	15:00	15:15	15:30	15:45	16:00	16:15	16:30
Without Curtailment	222	217	201	181	180	167	123	89	70	39	15	3
With Curtailment	5	9	9	11	15	16	12	5	16	13	13	3
Curtailment factor	0.352	0.343	0.336	0.296	0.263	0.235	0.207	0.167	0.090	0.063	0.005	0

Table 6-5: Violated scenarios during the day before and after the optimized curtailment factor

Chapter 7

Conclusion

In this chapter, we conclude the findings of our thesis and we answer the sub-research questions, and consequently the main research question. Moreover, we present our recommendations for future research.

7-1 Conclusions

In this section we answer the three sub-questions that we introduced in the introduction.

What are the available optimization methods to incorporate the uncertain PV generation and residential demand?

In order to achieve voltage regulation we needed to configure the optimization method in a way that the uncertain PV generation does not lead the voltage rises in the network. Through the PV curtailment our purpose was to keep the voltage magnitudes below the infrastructural limit. For this reason we leveraged RO, CCP, SA and DRCCP as methods that can handle constraints that depend on uncertain parameters. We analyzed the principles and the assumptions of each parameter and we examined which one was the most suitable for our problem.

After our analysis, we concluded that the most appropriate optimization method to to incorporate the uncertain PV generation and residential demand is the DRCCP approach. In CCP, we assume that the whole probability distribution of the PV generation and residential demand is known. In RO we assume known only the support probability distribution of the PV generation and residential demand. However, a fact that hinders us from using these two approaches is that the true probability distribution of the PV generation and residential demand is unknown. From the available data we can construct only an empirical distribution $\hat{\mathbb{P}}$, but the true distribution remains unknown. On the other hand, for the SA approach we assume plenty of scenarios available. We examined the implementation of the SA, since in our dataset there are 500 scenarios available. The high computational time and the limited feasibility of the constraints that SA guarantees with this number of available scenarios, lead to the conclusion that the most appropriate method for us is the DRCCP method.

How to exploit data from previous years, in order to mitigate the power mismatch?

When we implement the DRCCP, we focus to find a solution that satisfies the constraints with high probability for the worst case realization of a wider family of distributions that are close to the empirical distribution. We call this family of distributions *ambiguity set*. The fact that the implementation of the DRCCP requires the construction of the ambiguity set, leads us to the conclusion that the recorded data from previous years were significant in order to create a Wasserstein uncertainty ball. The center of this ball was the empirical distribution from recorded data of previous years and the radius of the ball changed by adjustments in the Wasserstein distance. The more we increased the radius, the more uncertainty we incorporated to our problem.

How to use data efficiently and reduce the computational effort?

In our thesis we examined two distinctive case studies. In the first case study, we used 20 scenarios to tune the DRCCP algorithm. In this case study we did not experience any computational effort and we managed to reduce significantly the overvoltages, despite the limited data that we had to train the model.

In the second case study we intended to use the whole dataset of 500 scenarios to form the DRCCP. The 500 scenarios imposed significant computation burden and the problem was unable to converge even after two hours for one time-instance. The high computational effort, in case we use high bulk of data in the DRCCP, leads us to the conclusion that we need to examine the method of the Wasserstein Barycenter. With the Wasserstein Barycenter we achieved to cluster the 500 days of recording into only 10 points with their according weights. After the implementation of the Wasserstein Barycenters, the running time of the DRCCP dropped to 117 seconds for one time-instance.

These three sub-questions lead us step by step to answer the main research question of the thesis:

How to achieve voltage regulation in distribution energy networks, by curtailing appropriately the power exports from residential PV installations to the energy grid??

7-2 Recommendations for future work

In this section we present recommendations for future work that could extend the research of this thesis.

Use cross-validation during the tuning of DRCCP algorithm.

Our approach in Chapter 5 was to utilize the first 20 scenarios of our dataset in order to tune the Wasserstein distance of our ambiguity set. In a future work, we could use cross-validation during this tuning. With cross-validation, we divide the dataset into folds of 20 samples and we use a different fold to tune the radius of the ambiguity set and different folds to validate the results at each iteration. In this way, we will get wider insights into the size of our ambiguity set, since we will consider more optimistic and pessimistic samples of the dataset.

Add electric vehicles (EV) to the network.

In this thesis we examined the network failures that emerged from the excess power of the PV exports. More research can be conducted in distribution networks that include PV units and EVs. The intended outcome of this future research is to extend the empirical distribution, in order to incorporate the charging part of EVs. In this case, part of the excess power would be consumed by EVs, and thus the curtailment factor would be reduced.

One other aspect that this recommendation should consider is the undervoltage instances. The undervoltage instances occur mainly in the night, during high demand and low generation. Regarding the PVs, the generation during the night is zero. A feature research should consider the uncertain power needs of EV charging, by incorporating Battery Energy Storage (BES) systems.

Consider overcurrent instances and limitations of the grid-connected devices.

In this thesis we intend to achieve voltage regulation. In feature, we could also consider overcurrent constraints and constraints that concern the connected devices, such as SOC limitations for storage devices, ramping limits and reserve needs. These constraints are necessary to examine further the stability of the network. However, they may increase the computational time of the DRCCP algorithm.

Develop a techno-economic analysis in the network.

In the thesis we mentioned the trade-offs that we need to consider when we curtail the PV exports. We should not take overconservative decisions, otherwise the investors would endure unnecessary loss of profit due to high cuts of PV exports. In a future work, this trade-off could be modelled by a parameter that takes into consideration the risk level when we do not curtail the excess power combined with the economic impact of the curtailment to PV owners.

Consider individual power curtailment for each PV unit.

In this work we adopted a common curtailment factor for all the PV installations of the network. A feature work could extend the research to individual PV limitations. With this approach the role of the PV owners is upgraded and they have the chance to determine on their own their generation.

Bibliography

[1] F. Gökgöz and M. T. Güvercin, "Energy security and renewable energy efficiency in eu," *Renewable and Sustainable Energy Reviews*, vol. 96, pp. 226–239, 2018.

[2] M. A. Franco and S. N. Groesser, "A systematic literature review of the solar photovoltaic value chain for a circular economy," *Sustainability*, vol. 13, no. 17, 2021.

[3] A. Allouhi, S. Rehman, M. S. Buker, and Z. Said, "Up-to-date literature review on solar pv systems: Technology progress, market status and r&d," *Journal of Cleaner Production*, p. 132339, 2022.

[4] A. Vulkan, I. Kloog, M. Dorman, and E. Erell, "Modeling the potential for pv installation in residential buildings in dense urban areas," *Energy and Buildings*, vol. 169, pp. 97–109, 2018.

[5] K. Petrou, A. Procopiou, L. Gutierrez-Lagos, M. Liu, L. Ochoa, T. Langstaff, and J. Theunissen, "Ensuring distribution network integrity using dynamic operating limits for prosumers," *IEEE Transactions on Smart Grid*, vol. 12, pp. 3877–3888, 09 2021.

[6] N. W. Stauffer, "Researchers find benefits of solar photovoltaics outweigh costs," 2020.

[7] Y. Guo, K. Baker, E. Dall'Anese, Z. Hu, and T. H. Summers, "Data-based distributionally robust stochastic optimal power flow—part ii: Case studies," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1493–1503, 2019.

[8] A. Arrigo, C. Ordoudis, J. Kazempour, Z. De Grève, J.-F. Toubeau, and F. Vallée, "Wasserstein distributionally robust chance-constrained optimization for energy and reserve dispatch: An exact and physically-bounded formulation," *European Journal of Operational Research*, vol. 296, no. 1, pp. 304–322, 2022.

[9] R. Mieth and Y. Dvorkin, "Data-driven distributionally robust optimal power flow for distribution systems," *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 363–368, 2018.

[10] M. Cuturi and A. Doucet, "Fast computation of wasserstein barycenters," in *Proceedings of the 31st International Conference on Machine Learning* (E. P. Xing and T. Jebara, eds.), vol. 32 of *Proceedings of Machine Learning Research*, (Bejing, China), pp. 685–693, PMLR, 22–24 Jun 2014.

[11] L. Kotzur, L. Nolting, M. Hoffmann, T. Groß, A. Smolenko, J. Priesmann, H. Büsing, R. Beer, F. Kullmann, B. Singh, A. Praktiknjo, D. Stolten, and M. Robinius, "A modeler's guide to handle complexity in energy systems optimization," *Advances in Applied Energy*, vol. 4, p. 100063, 2021.

[12] A. Hota, A. Cherukuri, and J. Lygeros, "Data-driven chance constrained optimization under wasserstein ambiguity sets," pp. 1501–1506, 07 2019.

[13] A. Zhou, M. Yang, M. Wang, and Y. Zhang, "A linear programming approximation of distributionally robust chance-constrained dispatch with wasserstein distance," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 3366–3377, 2020.

[14] C. Duan, W. Fang, L. Jiang, L. Yao, and J. Liu, "Distributionally robust chance-constrained approximate ac-opf with wasserstein metric," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4924–4936, 2018.

[15] X. Geng and L. Xie, "Data-driven decision making with probabilistic guarantees (part 1): A schematic overview of chance-constrained optimization," *arXiv: Optimization and Control*, 2019.

[16] M. C. Campi, S. Garatti, and M. Prandini, "The scenario approach for systems and control design," *Annual Reviews in Control*, vol. 33, no. 2, pp. 149–157, 2009.

[17] C. Ordoudis, V. A. Nguyen, D. Kuhn, and P. Pinson, "Energy and reserve dispatch with distributionally robust joint chance constraints," *Operations Research Letters*, vol. 49, no. 3, pp. 291–299, 2021.

[18] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM Journal on Optimization*, vol. 17, pp. 969–996, 01 2006.

[19] P. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations," *Mathematical Programming*, vol. 171, 05 2015.

[20] R. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of risk*, vol. 2, pp. 21–42, 01 2000.

[21] M. Chapman, R. Bonalli, K. Smith, I. Yang, M. Pavone, and C. Tomlin, "Risk-sensitive safety analysis using conditional value-at-risk," 01 2021.

[22] S. Sarykalin, G. Serraino, and S. Uryasev, "Value- at-risk vs conditional value-at-risk in risk management and optimization," 09 2008.

[23] B. K. Poolla, A. R. Hota, S. Bolognani, D. S. Callaway, and A. Cherukuri, "Wasserstein distributionally robust look-ahead economic dispatch," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2010–2022, 2021.

[24] Y. Guo, K. Baker, E. Dall'Anese, Z. Hu, and T. H. Summers, "Data-based distributionally robust stochastic optimal power flow—part i: Methodologies," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1483–1492, 2019.

[25] F. Cazals, T. Dreyfus, and D. Mazauric, "Earth mover distance," Structural Bioinformatics Library, 2016.

[26] D. Kuhn, P. Esfahani, V. Nguyen, and S. Shafieezadeh-Abadeh, *Wasserstein Distributionally Robust Optimization: Theory and Applications in Machine Learning*, pp. 130–166. 10 2019.

[27] R. T. Rockafellar, *Convex analysis*. Princeton Mathematical Series, Princeton, N. J.: Princeton University Press, 1970.

[28] E. O'Shaughnessy, J. R. Cruce, and K. Xu, "Too much of a good thing? global trends in the curtailment of solar pv," *Solar Energy*, vol. 208, pp. 1068–1077, 2020.

[29] United States Department of Labour, ""electric power generation, transmission, and distribution : Illustrated glossary - transmission lines"," 2022.

[30] Commission de régulation de l'énergie, "Electricity networks," République française, 2018.

[31] F. Islam, K. Prakash, K. Mamun, A. Lallu, and H. Pota, "Aromatic network: A novel structure for power distribution system," *IEEE Access*, vol. PP, pp. 1–1, 10 2017.

[32] A. Siddiqui, "Determination of size and location of capacitors for placement on a radial distribution system using fuzzy technique," *International Journal on Power System Optimization and Control*, vol. 3, pp. 1–8, 06 2011.

[33] K. Prakash, A. Lallu, F. Islam, and K. Mamun, "Review of power system distribution network architecture," pp. 124–130, 12 2016.

[34] M. Z. Liu, A. T. Procopiou, K. Petrou, L. F. Ochoa, T. Langstaff, J. Harding, and J. Theunissen, "On the fairness of pv curtailment schemes in residential distribution networks," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4502–4512, 2020.

[35] J. S. Giraldo, O. D. Montoya, P. P. Vergara, and F. Milano, "A fixed-point current injection power flow for electric distribution systems using laurent series," *22nd Power Systems Computation Conference*, 01 2022.

[36] S. V. Dhople, S. S. Guggilam, and Y. C. Chen, "Linear approximations to ac power flow in rectangular coordinates," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 211–217, 2015.

[37] S. S. Guggilam, E. Dall'Anese, Y. C. Chen, S. V. Dhople, and G. B. Giannakis, "Scalable optimization methods for distribution networks with high pv integration," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 2061–2070, 2016.

[38] A. Bernstein and E. Dall'Anese, "Linear power-flow models in multiphase distribution networks," in *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pp. 1–6, 2017.

[39] L. Yu, D. Czarkowski, and F. de Leon, "Optimal distributed voltage regulation for secondary networks with dgs," *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 959–967, 2012.

[40] N. Ho, X. Nguyen, M. Yurochkin, H. Bui, V. Huynh, and D. Phung, "Multilevel clustering via wasserstein means," 06 2017.

[41] J. M. Altschuler and E. Boix-Adserà, "Wasserstein barycenters are np-hard to compute," *SIAM Journal on Mathematics of Data Science*, vol. 4, no. 1, pp. 179–203, 2022.

[42] J. M. Altschuler and E. Boix-Adserà, "Wasserstein barycenters can be computed in polynomial time in fixed dimension," *J. Mach. Learn. Res.*, vol. 22, pp. 44:1–44:19, 2021.

[43] P. Dvurechenskii, D. Dvinskikh, A. Gasnikov, C. Uribe, and A. Nedich, "Decentralize and randomize: Faster algorithm for wasserstein barycenters," in *Advances in Neural Information Processing Systems* (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, eds.), vol. 31, Curran Associates, Inc., 2018.

Glossary

List of Acronyms

3mE	Mechanical, Maritime and Materials Engineering
RES	Renewable Energy Sources
DCSC	Delft Center for Systems and Control
UNFCCC	United Nations Framework Convention on Climate Change
PV	Photovoltaics
DNO	Distribution Network Operators
DRCCP	Distribunally Robust Chance Constraint Programming
OPF	Optimal Power Flow
WB	Wasserstein Barycenters
RO	Robust optimization
CCP	Chance Constraint Programming
SA	Scenario Approach
BES	Battery Energy Storage
EV	Electric Vehicles

3mE RES DCSC UNFCCC PV DNO DRCCP OPF WB RO CCP SA BES EV

List of Symbols

β	Confidence parameter
$\cos(\phi)$	Power factor
$\mathbb{E}$	Expectation
$\mathbb{P}$	Probability distribution
$\mathbb{R}$	Real numbers
θ	Wasserstein radius
Ξ	Uncertain set
ξ	Uncertain parameter
ξ_1	Uncertain PV generation
ξ_2	Uncertain residential demand
a	Violation parameter
a_c	Curtailment factor
a_{wb}	Weight that corresponds to a point of a Wasserstein Barycenter
p	Vector of active powers P_i
q	Vector of reactive active powers Q_i
S_d	Apparent power of a node
V_d	Voltage of a node
$V_{\max}$	Maximum voltage magnitude
W_p	Wasserstein distance
y_{ij}	Admittance of a line

ξ ξ_1 ξ_2 Ξ β a a_c $\mathbb{P}$ $\mathbb{R}$ $\mathbb{E}$ W_p θ a_{wb} y_{ij} S_d p q V_d $V_{\max}$ $\cos(\phi)$